

A novel approach to statistical-dynamical downscaling for long-term wind resource predictions

Article

Accepted Version

Chavez-Arroyo, R., Fernades-Correia, P., Lozano-Galiana, S., Sanz-Rodrigo, J., Amezcua, J. and Probst, O. (2018) A novel approach to statistical-dynamical downscaling for long-term wind resource predictions. *Meteorological Applications*, 25 (2). pp. 171-183. ISSN 1469-8080 doi: <https://doi.org/10.1002/met.1678> Available at <https://centaur.reading.ac.uk/70280/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/met.1678>

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



A novel approach to statistical-dynamical downscaling for long-term wind resource predictions

Journal:	<i>Meteorological Applications</i>
Manuscript ID	MET-16-0155.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Chávez-Arroyo, Roberto; Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias; Centro Nacional de Energías Renovables, Energía Eólica Fernandes-Correia, Pedro; Centro Nacional de Energías Renovables, Energía Eólica Lozano-Galiana, Sergio; Centro Nacional de Energías Renovables, Energía Eólica Sanz-Rodrigo, Javier; CENER, Wind energy department Amezcuca, Javier; University of Reading, Department of Meteorology Probst, Oliver; Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias
Keywords:	Energy < Climate change impacts, NWP < Modelling, Sensitivity Analysis < Modelling, Statistical Models < Modelling, Reliability < Verification, Ensemble < Climate change impacts
Manuscript keywords:	Long-term wind resource, statistical-dynamical downscaling, stratified sampling, mean sea level maps, image processing, reanalysis data

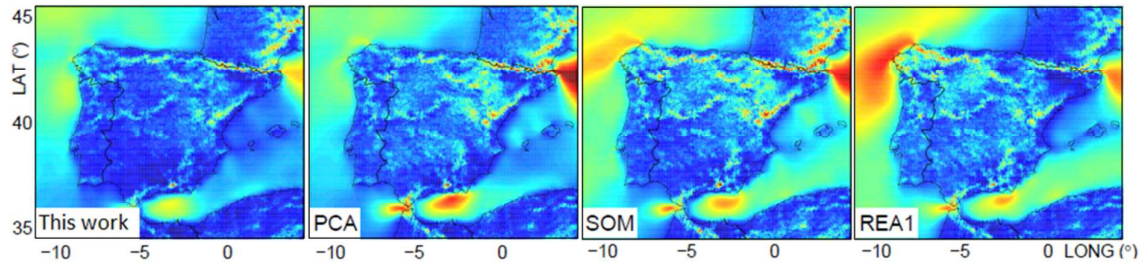
SCHOLARONE™
Manuscripts



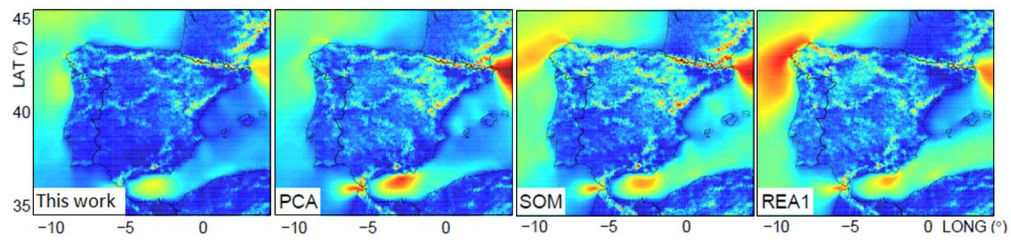
Graphical table of contents

A novel approach to statistical-dynamical downscaling for long-term wind resource predictions

Roberto Chávez-Arroyo, Pedro Fernandes-Correia, Sergio Lozano-Galiana, Javier Sanz-Rodrigo, Javier Amezcua, *Oliver Probst



A novel method to determine representative periods (typically a year) for the estimation of the long-term mesoscale wind resource has been proposed and compared to other recently published techniques. It provides a computationally lean while accurate solution of the problem of constructing long-term mesoscale wind maps through downscaling without having to go through a *brute force* procedure. Applications include a wider dissemination of mesoscale wind maps because of faster and cheaper execution, as well as greater flexibility for sensitivity analyses.



781x184mm (96 x 96 DPI)

For Peer Review

1 A novel approach to statistical-dynamical downscaling for long-term 2 wind resource predictions

3 ^{1,2}Roberto Chávez-Arroyo, ²Pedro Fernandes-Correia, ²Sergio Lozano-Galiana, ²Javier Sanz-Rodrigo,
4 ³Javier Amezcua, ^{1,*}Oliver Probst

5 ¹School of Engineering and Sciences, Tecnológico de Monterrey, Eugenio Garza Sada 2501 Sur,
6 Monterrey, NL, CP 64849, Mexico; ²Centro Nacional de Energías Renovables, Av. Ciudad de la
7 Innovación, 7, 31621 Sarriguren, Navarra, Spain; ³Department of Meteorology, University of Reading,
8 Earley Gate, PO Box 243, Reading, RG6 6BB, UK

9 *Corresponding author: oprobst@itesm.mx

10 Abstract

11 A new method for the long-term prediction of the wind resource based on the concept of statistical-
12 dynamical downscaling is presented. This new approach uses mean sea-level pressure (MSLP) maps from
13 global reanalysis data (NCEP-DOE AMIP-II) and image processing techniques to identify a synthetic
14 reference period which optimally matches the corresponding long-term maps. Four different image
15 processing techniques, averaged into one image similarity error index, are used to evaluate image
16 similarity. A representative set of days is selected by requiring the error index to be minimal. Validation
17 of representativeness in terms of the wind resource for the Iberian domain is performed against ten years
18 of measured wind data from Navarra (Spain), as well as mesoscale simulations of the Iberian Peninsula.
19 The new approach is shown to outperform not only the industry-standard method but also other recently
20 proposed methods in its capability to achieve mesoscale level representativeness. A particular advantage
21 of the new method is its capability of simultaneously providing a representative period for all potential
22 wind farm sites located within large regional domains without requiring re-running the method for
23 different candidate sites.

24 **Keywords:** Long-term wind resource; statistical-dynamical downscaling; stratified sampling; mean sea
25 level maps; reanalysis data; image processing

26 1. Introduction

27 The accurate representation of large-scale circulation systems and their associated local wind field is
28 essential to the development and financing of wind power plants. Consequently, the interest in methods
29 providing insights into the representativeness of on-site measured wind data has sparked in recent years.
30 To date, most industry-standard analyses of the long-term wind resource of a wind project rely on
31 statistical relationships (Romo Perea *et al.*, 2011), often in the form of a linear regression, between the on-
32 site towers and suitable reference sites, typically either Automatic Surface Observation Stations (ASOS)
33 or virtual wind towers from numerical climate data bases such as NCEP/NARR (North American
34 Regional Reanalysis) (Mesinger *et al.*, 2006), NASA/MERRA (Modern-Era Retrospective Analysis for
35 Research and Applications) (Rienecker *et al.*, 2011), ECMWF/ERA-Interim among others (Liléo *et al.*,
36 2013). While the predictions derived from statistical relationships may be quite accurate if the long-term
37 station is located at a site with similar wind climate and exposure conditions, and long-term
38 measurements have been conducted at a height level similar to the projected turbine hub height, these
39 idealized conditions are only seldom met. The construction of statistical associations between long-term

40 atmospheric data with local variables is widely known as *statistical downscaling*. In these methods, a
41 statistical model is trained to use empirical relationships between local observed variables, predictands,
42 and large scale atmospheric fields known as predictors (e.g. García-Bustamante *et al.* (2012)). This
43 approach is very computationally efficient; however, similar to the virtual series method, its reliability
44 relies on the strength of the statistical links between atmospheric circulations and regional observations.

45 On the opposite side are approaches that solve the equations of the atmosphere dynamics for a period
46 of interest. Such approaches are generally referred to as *dynamical downscaling* and provide solutions to
47 the atmospheric dynamics in the region of interest based on the full set of balance equations at the chosen
48 simulation grid and semi-empirical relationships for physical processes occurring at the sub-grid level
49 (Pielke, 2002). A possible approach to assessing the long-term wind resource in a region of interest is to
50 apply the dynamical downscaling methodology for periods long enough to capture the intra- and inter-
51 annual, and if possible, decadal variability. Despite the significant increase in computing power, the need
52 for solutions at finer-resolution scales and the continuous development of more complex physics such as
53 the implementation of higher-order models for the solution of sub-grid processes (e.g. Barranger & Kallos
54 (2012)) requires substantially higher computational resources than those typically available at the desktop
55 level or in small research clusters.

56 Practical solutions to the dilemma outlined above are *hybrid methods*, often termed *statistical*
57 *dynamical downscaling* (SDD), which combine the deterministic approach of dynamical downscaling
58 with statistical techniques. SDD methods can be roughly grouped into three categories: (1) approaches
59 based on the classification of weather types (Frey-Buness *et al.*, 1995), (2) algorithmic methods with
60 suitable evaluation metrics where almost no pre-classification by the user is required (Fuentes &
61 Heimann, 2000a; Cutler *et al.*, 2006; Hagemann, 2008; Rife *et al.*, 2013; Tammelin *et al.*, 2013; Chávez-
62 Arroyo *et al.*, 2013; Martínez *et al.*, 2013), and (3) construction of statistical relations between long-term
63 and short-term dynamical downscaled data (Hahmann *et al.*, 2012). Weather classification approaches are
64 rather typical in the atmospheric sciences, where algorithmic methods have made a more recent
65 appearance. In both cases the underlying assumption is that any regional climate can be associated with a
66 specific frequency distribution of classified large-scale weather situations (Frey-Buness *et al.*, 1995).

67 While methods based on the synoptic-scale classification of stationary weather regimes are generally
68 physically meaningful and often lead to a reasonably small number of classes, there are several downsides
69 which limit their usefulness for comprehensive wind prospecting purposes. First, there is generally a loss
70 of the time-dependent phenomena (Fuentes & Heimann, 2000b). Secondly, the simulation of daily or
71 seasonal phenomena is difficult because of the large number of classes required for such a distinction.
72 Thirdly, classification is rather case-specific and somewhat subjective as it requires site-specific
73 knowledge (Cutler *et al.*, 2006).

74 Some of these issues can in principle be tackled by algorithmic approaches such as those described
75 below. Martínez *et al.* (2013) disaggregated the large-scale climate into their Empirical Orthogonal
76 Functions (EOF) which were then dynamically downscaled. They were able to account for the temporal
77 variability by applying the Principal Components time series to the downscaled version of the
78 corresponding EOF which also served for the comparisons with observations. Fuentes and Heimann
79 (2000) proposed the dynamical downscaling of the most representative multi-day episodes of quasi-
80 stationary circulations. These episodes were created by hierarchical cluster analysis combined with a
81 spatio-temporal metric to collect consecutive dates with similar patterns described by their first Principal
82 Components of the geopotential at 500 hPa. Hagemann (2008) proposed the use of Self-Organizing Maps

83 (SOM) to select a representative continuous set of 365 days which were afterwards dynamically
84 downscaled.

85 Rife et al. (2013) based their selection on a 365-day sample with equal representation for all days. The
86 sample resulted from testing very large number sets generated by random stratified sampling and
87 selecting the best rated sample's wind speed and direction distributions according to their proposed
88 distance metric based on the χ^2 statistic. Their selection was made from daily averages of wind speed and
89 direction from the first native vertical level from the MERRA global reanalysis data point that best
90 correlated with the measurements used for validation. The ideas of Hageman (2008), Tammelin *et al.*
91 (2013) and Rife *et al.* (2013) were also addressed by Chavez-Arroyo *et al.* (2013) and Chávez-Arroyo *et al.*
92 (2015). They employed the stratified sampling technique (Rife *et al.*, 2013) to generate a large pool of
93 candidate years that were ranked according to their similarity to the long-term climate. Chávez-Arroyo *et al.*
94 (2013), Chávez-Arroyo *et al.* (2015), and (Hagemann (2008) used large-scale atmospheric circulation
95 predictors, which probe representativeness at a mesoscale level. Recently, Vanvyve *et al.* (2015)
96 presented another interesting methodology suitable for statistical-dynamical downscaling based on an
97 analogue ensemble approach.

98 The present work is a new contribution to the field of statistical dynamical downscaling and has two
99 main objectives: (1) introduction of an improved algorithmic approach for the determination of
100 representative periods, and (2) a systematic comparison of the new proposal against published methods.
101 The new method is based on the assumption that regional wind flow is driven by large-scale systems
102 which can be suitably analysed in terms of synoptic patterns. While the previously published methods
103 require a relatively modest computing time compared to the one required for downscaling large regional
104 areas the computational effort is still considerable, which motivated the search for a computationally
105 leaner method which should ideally retain or improve the accuracy of the previously proposed methods
106 for identifying representative periods. As will be shown in the following, the novel approach proposed in
107 this work does indeed comply with these expectations.

108

109 2. Methods and data

110 A total of five different methods for the determination of representative periods were evaluated. One
111 technique, termed the *Best Annual Mean and Standard Deviation* or BAMS method, is new and will be
112 described at some detail in the present paper. It will be shown that this novel approach consistently
113 outperforms all other methods. Another important method implemented and evaluated in this work is the
114 one put forward by Rife *et al.* (2013) discussed above. The Rife method was implemented in two variants
115 (REA1 and REA2) described in greater detail below. Two other techniques recently published and used
116 for comparison in the present work are the PCA (Principal Component Analysis)-based approach
117 proposed by Chávez-Arroyo *et al.* (2013) and the SOM (Self-Organizing Maps) method used by Chávez-
118 Arroyo *et al.* (2015). The last method used for comparison and termed the traditional (TRA) approach in
119 this work is the wind industry-standard procedure for statistical downscaling of the long-term wind
120 resource. The traditional method (TRA) is based on the construction of one synthetic year (365 days)
121 where each calendar day of the representative year is determined from random draws among the
122 corresponding set of repetitions of that calendar day within the long-term period.

123

124 2.1 The Best Annual Mean and Standard Deviation (BAMS) method

125 The novel approach introduced in this work incorporates several elements explored successfully in
126 published work. One key ingredient is the use of mean sea-level pressure (MSLP) mesoscale maps for a

127 region of interest as predictors; this approach was found to outperform the industry-standard practice in
 128 Chávez-Arroyo *et al.* (2013) and Chávez-Arroyo *et al.* (2015). Another element is the use of the stratified
 129 sampling technique put forward by Rife *et al.* (2013) in which a large number of monthly data sets (10^5
 130 samples/month) are generated by randomly sampling the long-term SLP data, each sample containing 28,
 131 30, or 31-days per month, depending on the calendar month. The new ingredient in the BAMS method is
 132 the way the similarity between the long-term and the representative period is established. Whereas Rife *et al.*
 133 *et al.* (2013) use comparisons between wind resource at a point location the BAMS method uses a
 134 mesoscale approach to establish this similarity by using the complete information contained in the
 135 regional MSLP maps. Such an approach seems more appropriate for mesoscale downscaling applications,
 136 though it was clear from the outset that the Rife approach might work better at the specific location used
 137 for tuning the method.

138 Evidently, a tool is required to establish the similarity between mesoscale maps. The novelty in the
 139 present work consists in its departure from traditional statistical methods used in the atmospheric
 140 sciences, both linear (such as PCA) and non-linear (such as SOM), by using image processing techniques
 141 successfully employed in other areas of science and engineering. Four different image similarity detection
 142 techniques are used and combined into one similarity index after applying a linear normalization operator
 143 to the sub-index associated with each method. An optimal or representative set of days is selected by
 144 requiring the error index to be minimal.

145 In the following we will consider two types of maps, $L(s)$ (long-term) and $R(s)$ (candidate
 146 representative period), which have been obtained from temporally averaging over their respective
 147 observational periods; $s \in [1, Q]$ with Q being the number of grid points. Both the temporal mean (e.g.
 148 $L(s) = \sum L(s, t)/T$) and the standard deviation (e.g. $L(s) = (\sum(L(s, t) - \bar{L})^2)^{1/2}/(T - 1)$) are
 149 considered for both of the similarity indices considered. We can then define the *first similarity index*
 150 SI_1 as the root mean square error of the point-to-point comparison between the L and R map as

$$SI_1 \equiv \langle \varepsilon \rangle = \left(\frac{1}{Q} \sum_{r=1}^Q (L_r - R_r)^2 \right)^{1/2} \quad (1)$$

151 The *second similarity index* SI_2 is an average Pearson correlation coefficient $\rho(x, y)$ between rows and
 152 columns from each pair of maps. In order to account for small variations in the sea level pressure (SLP)
 153 patterns which are due to small shifts between maps but still represent almost identical spatial patterns,
 154 the correlations are computed not only between corresponding rows (i.e., $\rho(R_a, L_a)$, a being the row
 155 index, $a \in [1, A]$, A being the number of rows) and columns (i.e. $\rho(R_b, L_b)$, b being the column index,
 156 $b \in [1, B]$), B being the number of columns), but also from certain neighbouring rows and columns as
 157 defined by $\rho(R_a, L_{a \pm 0.03A})$ and $\rho(R_b, L_{b \pm 0.03B})$, respectively, i.e.

$$SI_2 \equiv \bar{\rho} = \frac{1}{3A} \sum_{k=-1}^1 \sum_{a=1}^A \rho(R_a, L_{a+k \cdot 0.03A}) + \frac{1}{3B} \sum_{k=-1}^1 \sum_{b=1}^B \rho(R_b, L_{b+k \cdot 0.03B}) \quad (2)$$

158
 159 The third similarity index $SI_3 = \gamma$, is based on a comparison between distributions of the SLP values
 160 contained in both rows and columns of the images representing the long-term and the candidate
 161 representative period, respectively. As before, the analysis is conducted on both the average SLP and its
 162 standard deviation. Specifically, the following calculations are carried out: a histogram f_a of the SLP

163 values of the a th row and, similarly, a histogram g_b of the SLP values of the b th column in each of the
 164 SLP mean or standard deviation map is first calculated using 30 bins. Then, the Pearson correlation for all
 165 pairs of distributions is calculated from the rows with the same index, i.e. $\text{corr}(f_a^{\text{LT}}, f_a^{\text{RP}})$ and, similarly,
 166 from columns with the same index, i.e. $\text{corr}(g_b^{\text{LT}}, g_b^{\text{RP}})$, where in both cases the superscript LT refers to
 167 *long-term* and RP to *representative period*. Finally the row and column correlations are averaged to
 168 compute the third similarity index γ according to

$$\text{SI}_3 \equiv \gamma = \frac{1}{A+B} \left(\sum_{a=1}^A \text{corr}(f_a^{\text{LT}}, f_a^{\text{RP}}) + \sum_{b=1}^B \text{corr}(g_b^{\text{LT}}, g_b^{\text{RP}}) \right) \quad (3)$$

169 The fourth metric or similarity index SI_4 is a structural similarity index proposed by Wang *et al.*
 170 (2004). This technique is designed for the quality assessment based on the degradation of structural
 171 information of an image. This technique attempts to quantify the visibility of errors (differences) between
 172 a distorted image and a reference image. The structural similarity index SSIM for two images x and y is
 173 computed as:

$$\text{SI}_4 \equiv \text{SSIM}(x, y) = [l(x, y)^\alpha c(x, y)^\beta s(x, y)^\Gamma], \quad (4a)$$

174 where $l(x, y)$ is the luminosity function, $c(x, y)$ the contrast comparison function, and $s(x, y)$ the
 175 structure comparison function, respectively defined by

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (4a)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (4b)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (4c)$$

176 The mean μ and the standard deviation σ of an image have their usual meanings, $\sigma_{xy} = \text{cov}(x, y)$ is
 177 the covariance between images x and y , and C_1, C_2 , and C_3 are small constants ($C_1 \ll \mu_x^2$, etc.). α, β , and
 178 Γ are free parameters used to adjust the relative importance of the three image comparison functions
 179 introduced above; in the present work $\alpha = \beta = \Gamma = 1$.

180 In the BAMS method, the four image similarity components described above are linearly transformed
 181 using the lower and upper bounds of each statistic calculated from the N samples making up the candidate
 182 set of representative periods i . This transformation scales the ranges of each of the four similarity indices
 183 to $[0, 1]$ in order to allow them to be averaged into one overall similarity measure. The linearly scaled
 184 version $\phi_{i,j}^*$ of any of the four metrics ($\text{SI}_k, k = 1 \dots 4$) is then obtained from

$$\phi_{i,j}^* = \frac{\phi_{i,j} - \min_k\{\phi_{k,j}\}}{\max_k\{\phi_{k,j}\} - \min_k\{\phi_{k,j}\}}, \quad (5)$$

185 for the i th candidate and a given month j . All four normalized metrics are then averaged into one
 186 performance index $\tau_{i,j}$. In the present work, equal weights were assigned to the each of the four indices,
 187 although generalization to non-equal weights is evidently straightforward:

$$\tau_{i,j} = \frac{1}{4} \{ \langle \varepsilon \rangle_{i,j}^* + (1 - \rho_{i,j}^*) + (1 - \gamma_{i,j}^*) + (1 - \text{SSIM}_{i,j}^*) \}, \quad (6)$$

188 where $\langle \varepsilon \rangle_{i,j}^*$, $\rho_{i,j}^*$, $\gamma_{i,j}^*$, and $\text{SSIM}_{i,j}^*$ are the linearly scaled versions of the four similarity indices
 189 according to equation (5). Please note that the complements of $\rho_{i,j}^*$, $\gamma_{i,j}^*$, and $\text{SSIM}_{i,j}^*$ are used in equation
 190 (6) are used for consistency with the first similarity metric $\langle \varepsilon \rangle_{i,j}^*$. The BAMS index is now constructed by
 191 averaging the performance indices $\tau_{i,j}$ for the monthly mean μ_{SLP} of the sea level pressure (SLP) and its
 192 standard deviation σ_{SLP} .

$$\text{BAMS}_{ij} = \frac{1}{2} \{ \tau_{ij}(\mu_{\text{SLP}}) + \tau_{ij}(\sigma_{\text{SLP}}) \} \quad (7)$$

193 The representative year is then selected by requiring $\text{BAMS}_{i,j}$ to be at a minimum for each month, i.e.

$$\text{BAMS}_{\text{RP},j} = \min_k \{ \text{BAMS}_{k,j} \} \quad (8)$$

194

195 2.2 Observational data base used for validation

196 The surface observational data employed here consist of wind speed and direction measurements from 22
 197 automatic weather observation stations (ASOS) in *Comunidad Foral de Navarra*, Spain. These stations
 198 are managed and maintained by the *Sección de Evaluación de Recursos Agrarios del Departamento de*
 199 *Agricultura, Ganadería y Alimentación* at the regional Government of Navarra. Figure 1 shows their
 200 location together with the topography information which outline the complex topographic elements
 201 dominating the region: a complex system of mountain ranges in the centre, on the eastern and western
 202 side the mountain lines which form the last foothills of the Pyrenees to the north, and the less complex
 203 low lands of the Ebro valley which characterizes the south of the region. This figure also shows the
 204 position of the MERRA grid points over the area which is used for further analysis below.

205 Prior to the analyses related to the main topic of this paper the wind speed and wind direction records
 206 of the 22 stations were subjected to an extensive quality assurance analysis in order to remove both rough
 207 and systematic errors by using the methodology described in Chávez-Arroyo & Probst (2015). The data
 208 period from 1 January 2001 through 31 December 2012 was used for the purposes of the present work.
 209 The total data recovery after quality assurance was 94% global, with station data recovery ranging from
 210 82% (NM20) to 98% (NM2) with a standard deviation of 5 percentage points.

211

212 2.3 Numerical data base

213 The Regional Weather Forecasting Model SKIRON was used to generate the numerical data set of
 214 surface wind speed employed in this study and includes 9 years of dynamically downscaled wind fields
 215 for the Iberian Peninsula for the period 2004-2012. The details of the SKIRON regional model are
 216 described in Kallos *et al.* (2005). Initial and boundary conditions were obtained from the NCEP Global
 217 Forecast System (GFS) global Numerical Weather Prediction model. Each SKIRON run is initialized as a
 218 cold start with the GFS analysis at 12 UTC of each day, and is run with a forecast horizon of 36 hours,
 219 updating the boundary conditions with GFS data every three hours. The first 12 hours are discarded as
 220 part of the spin-up time of the model (Gastón *et al.*, 2008) while the next 24 forecast hours are stored

221 every hour. These outputs are post-processed in order to obtain the wind speed at 80 m above surface
222 from the native η vertical levels by performing a power-law interpolation to account for wind shear. The
223 static data for representing vegetation, topography and soil texture are introduced at a resolution of 30",
224 30" and 2' respectively. In addition, a buffer zone of 100 km is considered by discarding this distance
225 from the borders to avoid possible dynamical and numerical inconsistencies at the interface between the
226 GFS and SKIRON boundaries.

227

228 *2.4 The PCA and the SOM methods*

229 The methods based on Principal Component Analysis (PCA) and Self-Organizing Maps (SOM) have
230 been introduced by Chávez-Arroyo *et al.* (2013) and Chávez-Arroyo *et al.* (2015), respectively. As the
231 BAMS method introduced in the present work both the PCA and the SOM method use mean sea level
232 pressure maps from NCEP-DOE AMIP-II reanalysis (Kanamitsu *et al.*, 2002) as input. In the
233 implementation of the PCA method the first Empirical Orthogonal Functions (EOFs) representing 70% of
234 the variability were used. In the case of the SOM method, the number of nodes (C) used to describe the
235 distribution function of the original data set was defined by three rectangular arrays with dimensions
236 $[2 \times 3]$, $[4 \times 5]$ and $[6 \times 6]$ equivalent to 6, 20 and 36 nodes, respectively. The reader is referred to the
237 references above for a detailed description of both methods.

238

239 *2.5 Implementations of the Rife method*

240 The method described in Rife *et al.* (2013) requires data from a reanalysis location, taken to be the
241 MERRA grid point closest to the point of interest in the original paper. A distance metric involving the
242 squared differences between both the wind speed and wind direction distributions of the long-term and
243 candidate representative year is used to determine the most representative (synthetic) year. Two versions
244 of the Rife method (termed REA1 and REA2) were coded to allow for a fair comparison in the case of the
245 validations against observational data (obtained from 22 met stations at the Navarra province). In the
246 REA1 implementation the MERRA grid point showing the highest average correlation of the daily wind
247 speed averages with all 22 Navarra ASOS data was selected; conversely, in the REA2 implementation the
248 MERRA point with the lowest average correlation with all 22 Navarra weather stations was used.

249 As the Rife method in its original version was not designed to achieve mesoscale representativeness, a
250 fair comparison called for an appropriate selection of the MERRA reference point to be used. Six
251 locations on the Iberian Peninsula were selected randomly (shown in the inset of Figure 1) and the
252 method was repeated for each of the reference points. The results of this sensitivity analysis are shown in
253 Figure 6. While the reference point was found to have no statistically significant impact on the results, the
254 results for two implementations (called R-I and R-II), chosen to be the ones with the "best" and "worst"
255 error metrics, are shown below.

256

257 *2.6 Validation metrics*

258 The following metrics were calculated for each of the time series associated with each of the 22 weather
259 stations (ASOS) at Navarra as well as each of the 99050 virtual wind speed records at 80 m above sea
260 level obtained from the mesoscale model SKIRON. The first validation metric is the Relative Absolute
261 Error (RAE) of a particular statistic such as the average ($x = \langle U \rangle$) or standard deviation ($x = \sigma(U)$) of
262 the wind speed between the long-term (LT) and the representative (RP) time series:

$$|\varepsilon_x| = \left| \frac{x^{LT} - x^{RP}}{x^{LT}} \right| \times 100 [\%] \quad (9)$$

263 The second metric used for validation is the Mean Absolute Error (MAE) of either the hourly ($T = 24$
 264 hours) or monthly ($T = 12$ months) wind speed time series averages:

$$|\varepsilon_T| = \frac{1}{T} \sum_{t=1}^T |U_t^{LT} - U_t^{RP}| \quad (10)$$

265 The third validation metric is the relative wind speed frequency difference

$$D = \sum_{g=1}^{n_{bins}} w_g \left| \frac{f_g^{LT} - f_g^{RP}}{f_g^{LT}} \right| \times 100 [\%] \quad (11)$$

266 where f_g^{LT} and f_g^{RP} are the relative wind speed frequencies corresponding to the long-term and
 267 representative period respectively, computed for a total number of bins given by $n_{bins} = 20$. The
 268 weighting factor w_g is taken as the long-term frequency ($w_g = f_g^{LT}$) in order to assign more importance
 269 to those bins with higher frequency of occurrence.

270 The last metric is based on the two-sample Kolmogorov-Smirnov (KS) test. This non-parametric
 271 hypothesis test evaluates the difference between the empirical cumulative wind speed distribution
 272 functions F of each pair of long-term and representative data sets (F_g^{LT} and F_g^{RP}). The KS metric is
 273 defined as

$$KS = \max(|F^{LT} - F^{RP}|) \quad (12)$$

274 2.7 Analysis of statistical significance

275 In order to allow for a meaningful statistical comparison among the six methods described above a
 276 consistent procedure had to be devised to deal with the random structure associated with the Monte Carlo
 277 approach common to all methods. For this purpose, 100 realizations of each method were conducted, with
 278 each realization implying the generation of 10^5 trial cases, with the exception of the industry-standard or
 279 *traditional* method (TRA) which only generates one trial case. Subsequently, the results were subjected to
 280 a one-way ANOVA (Analysis of Variance) test (Wilks, 2006), as well as a non-parametric version, the
 281 Kruskal-Wallis (KW) test (Kruskal & Wallis, 1952). The Kruskal-Wallis test was run in addition to
 282 ANOVA as it had not been possible from the outset to ensure homoscedasticity (i.e. similar variance
 283 among group), a relatively stringent requirement for ANOVA. The KW test starts by ranking all data
 284 from all groups i together (in the present case, values of a given error metric for the different methods),
 285 ignoring group membership. The KW metric H then measures the deviation of the average group ranks \bar{r}_i
 286 from the overall average rank \bar{r} of the sample compared to the scatter of individual ranks r_{ij} of all data
 287 around \bar{r} , where j identifies data points within group i . A suitable p -value is defined in order to decide
 288 whether at least one group is significantly different from the others.

- 289 a. Each method was run 100 times and all error statistics (section 2.6) are gathered for either the
 290 Navarra ASOS or the mesoscale SKIRON wind field.
- 291 b. The average and the standard deviation of the error metrics of each trial were computed in the
 292 space dimension, i.e. for $Q = 22$ stations in the case of the Navarra ASOS network and for $Q =$

293 99050 SKIRON grid points in the case of the mesoscale simulations. One group (with 100 sample
294 elements) was generated for each method.

295 c. Thereafter, a group comparison was conducted through the ANOVA test, in which the groups
296 correspond to the six methods discussed above: BAMS, PCA, SOM, REA1, REA2, and TRA.
297 The null hypothesis for the ANOVA is “the means for all the groups are equal”, while the
298 alternative hypothesis is “at least one mean is different”. The hypotheses for the KW test are the
299 same with the median instead of the mean. If the null hypothesis was rejected (p -value $\leq \alpha =$
300 0.05) in either the ANOVA or the KW test a Tukey-Kramer test (Tukey, 1949) was conducted in
301 order to establish if the difference between any two of the means or medians (in the case of KW)
302 was significant and to construct confidence limits (see e.g. (Hochberg & Tamhane, 1987) for a
303 review of multiple comparison techniques). Tukey confidence limits are shown as bars in Figures
304 4, 5, 6, and 9.

305 The multiple trials of all methods were performed in identical manner for both data sets used for
306 validation, i.e. the Navarra observational data and the numerical simulations conducted with SKIRON.
307 The long-term reference periods are 1 January 2001 through 31 December 2012 (12 years) for the
308 Navarra (ASOS) records and 1 January 2004 through 31 December 2012 (9 years) for the SKIRON-
309 derived wind data.

310

311 **3. Results and discussion**

312 *3.1 BAMS predictions for the Navarra network*

313 The results of the BAMS method for a single run are illustrated in Figure 2; as in the following sections –
314 with the exception of section 3.5 where the effect of the length of the representative period is assessed –
315 the representative period is taken to be one year. Figure 2 (a) shows a histogram of the relative
316 frequencies of the wind speed time series of both the long term and the best-matched representative year,
317 together with their corresponding Weibull probability density functions (PDF). It is conspicuous that the
318 PDF is very accurately predicted. Similarly, the daily profile (Figure 2 (b)) and the wind rose (Figure 2
319 (d)), two important metrics for annual energy production (AEP) estimate of wind farms, are almost
320 identically reproduced by the representative. In the case of the seasonal profile (Figure 2 (c)) the majority
321 of the monthly wind speeds are accurately predicted but occasional variations of up to 10% are also
322 observed. It should be noted that at the ensemble level the BAMS actually predicts the seasonal profile
323 consistently better than the other methods studied for comparison (see section 3.3).

324 *3.2 BAMS predictions for the SKIRON mesoscale map of the Iberian Peninsula*

325 Figure 3 shows the spatial distribution of the error of the wind speed average and energy density in the
326 Iberian domain, obtained from one realization of the BAMS method. It can be noticed that, with the
327 exception of a small area in the eastern part of the Iberian Peninsula, the wind speed error for most part of
328 the domain is very low and homogeneous, with a mean of the map (i.e. mean of $\langle \epsilon \rangle (U)$ for the 99050 time
329 series) of 1.2% and a median of 0.9% for the absolute relative error of the wind speed prediction (left
330 figure). Little dependence on geographic feature can be noticed in the map, which illustrates the
331 robustness of the method. Given the cubic dependence of the wind energy density on wind speed the error
332 in this variable is evidently higher, with an average error of 4% and a median of 3.1% compared to the
333 long-term results, and the spatial variation is somewhat higher than in the case of the wind speed (right
334 figure).

335 3.3 Comparative performance of different methods

336 (a) Kruskal-Wallis / ANOVA – Tukey-Kramer analysis for the Navarra network

337 A comparative study of the performance of the six methods discussed in this paper was conducted
338 according to the methodology described in section 2.7. As ANOVA has more stringent requirements
339 regarding the statistical properties of the groups of data to be compared (homoscedasticity, normality)
340 than the Kruskal-Wallis (KW) method, only the results obtained with the latter are shown for brevity.
341 However, it should be mentioned that the qualitative findings of the ANOVA approach (combined with
342 the Tukey-Kramer pairwise group comparison as in the KW case) are essentially identical to the ones
343 obtained with KW.

344 As shown in Figure 4 for the case of the Navarra met station network the p -value for the group
345 comparisons of all six metrics shown in the figure is $\ll 0.05$, so it can be safely stated that at least one of
346 the method is statistically different from the others. As illustrated by the Tukey-Kramer error bars the new
347 method (BAMS) and the two implementations of the Rife method all outperform the traditional method
348 (TRA) in a statistically significant sense for most of the error metrics but are indistinguishable among
349 each other. The PCA and SOM method, on the other hand, were found to be indistinguishable from the
350 TRA method in this study case. The Rife method, in its REA1 implementation, outperforms the PCA and
351 SOM methods in some of the metrics ($\varepsilon_{(U)}$, ε_{24} , D , KS) while being indistinguishable from them in others
352 ($\varepsilon_{\sigma(U)}$, ε_{12}); the REA2 implementation is always indistinguishable from SOM and PCA. A noteworthy
353 feature is the fact the BAMS methods clearly outperforms all other methods in the case of the seasonal
354 profile (as measured by ε_{12}). We will see further below that this feature was also confirmed in the case of
355 the mesoscale simulations for the Iberian Peninsula. While the REA1 method is indistinguishable from
356 BAMS in five of the metrics and outperformed by BAMS in one (ε_{12}) it does show the lowest rank in
357 three of the metrics, insinuating a consistently good performance. It should be stated that this good result
358 can at least in part traced back to the fact that the MERRA reference used for the REA analysis was
359 selected within the region of the met stations against which the methods are validated, and that on top of
360 that the MERRA point with the highest average correlation with the Navarra network had been chosen.
361 This situation is somewhat different for the cases where the training region is different from the validation
362 region, as in the case of the mesoscale simulation discussed next.

363

364 (b) Kruskal-Wallis (KW) / ANOVA – Tukey-Kramer (TK) analysis for SKIRON mesoscale simulations

365 Evidently, validation of the different methods against mesoscale simulation is much more relevant to the
366 general topic of the present work (statistical-dynamical downscaling) than point comparisons, as the
367 rationale behind this approach is precisely the reduction of the computational effort required to perform a
368 long-term mesoscale simulation for a large area, such as the Iberian Peninsula. Figure 5 shows the results
369 obtained with the six methods.

370 It is clear from these graphs that in the case of the mesoscale simulation the new method (BAMS) now
371 dramatically outperforms all other methods. In all metrics BAMS has by far the lowest error ranks and is
372 clearly distinguishable (in a KW-TK sense) from all methods but the PCA approach which is statistically
373 indistinguishable in the case of three metrics, and the REA implementation R-I (the one providing the
374 best ranks in the KW method) whose TK error bounds overlap with those of BAMS in the case of $\varepsilon_{\sigma(U)}$.
375 Not unexpectedly, the runner-up to BAMS is also a mesoscale method, and the local method proposed by
376 Rife *et al.* (2013) (in its implementations R-I and R-II, the ones providing the best and worst ranks in the
377 KW analysis, respectively, out of the six implementations R1 through R6, see Figure 1) is not working as

378 well on a larger scale as on the local scale with data from the region which was used to train it in the first
379 place. As seen from Figure 8 the REA implementations are now essentially indistinguishable from the
380 traditional method (TRA) used for reference. It is interesting to note that the method based on a linear
381 pattern analysis (PCA) is working somewhat better than its non-linear counterpart (SOM), which
382 performs a complex (and computationally demanding) detection of non-linear patterns. This is of course a
383 preliminary verdict, and other studies will have to show if this conclusion can be sustained on other data
384 sets.

385 In order to treat the REA method in all fairness and to avoid that the selection of the reference point
386 used for training of the method would unduly influence the results, six versions were implemented where
387 the reference point was changed to six random locations (R1 through R6) distributed evenly over the
388 Iberian Peninsula (Figure 1). The results are shown in Figure 6, where both a boxplot for the individual
389 results (providing error statistics for the 99050 locations/time series of the SKIRON mesoscale wind map)
390 and a group comparison plot have been provided. For brevity, only the metrics D (measuring wind speed
391 probability distributions) and $\varepsilon_{(U)}$ (measuring the global error in the wind speed prediction) have been
392 shown. ANOVA/Tukey-Kramer results are shown, but as in the previous cases, the Kruskal-Wallis
393 analysis provided essentially the same results. It can be seen from the figure that the results obtained with
394 either of the REA implementation were found to be statistically indistinguishable in an ANOVA sense, as
395 indicated by the high value of the p -value (0.72 and 0.84, respectively) and the overlap of the TK error
396 bounds. This is of course not all that surprising as in a local method like REA an increase in accuracy in
397 one location (the training location) is likely to come at the expense of a lesser accuracy in the rest of the
398 simulation region, resulting in a similar overall error. It has of course to be recognized that the REA was
399 not originally designed for mesoscale assessments for regions as large as the Iberian Peninsula, and that
400 the extension of it by the authors of the present work was for reference purposes only.

401 *3.4 Repeatability of the methods*

402 In order to obtain further insights into repeatability of the results the variance for each of the annual
403 ensemble average wind speed values of the 99050 locations modelled in the SKIRON mesh was
404 calculated, where - as before - the ensemble consisted of $M = 100$ realizations of each method. The
405 results are shown in Figure 7 as maps of the wind speed variance for each method. It is conspicuous from
406 the colour scales of the maps that the traditional method by far has the largest variability among
407 realizations and that the latter is relatively uniform over the simulation region, with the exception of a few
408 regions of higher variability in the mountain ranges in North-eastern Spain. This results was somewhat
409 expected, given the simple design of the method. On the other end of the range, the BAMS method is
410 again seen to outperform all other methods in terms of its high repeatability and a very high homogeneity,
411 i.e. very low dependence of the variability on location. It therefore provided what is expected from a
412 model suitable for mesoscale modelling. The runner-up, as before, as the PCA-based method which
413 shows a total variance only some 10% higher than BAMS and also a relatively good spatial homogeneity.
414 SOM and REA1 have a similar total variance and spatial homogeneity over the continental part of the
415 simulation domain, though REA1 shows a higher variability over the seas, particular the Atlantic Ocean
416 adjacent to Portugal and Galicia (North-western Spain).

417 *3.5 The role of the length of the representative period*

418 While it is intuitive to assume that a representative year should be the natural choice for the representative
419 (synthetic) period of a long-term period to be simulated under a mesoscale approach, it can also be
420 expected that there will be cost-benefit trade-off which has to be assessed for an optimal use of

421 computational resources while ensuring accuracy. Periods significantly shorter than one year are prone to
422 create biases associated with a lack of seasonal representativeness. Periods composed of several years
423 might be able to better assess inter-annual fluctuations but come at a higher computational cost. Figure 8
424 has the evolution of four error metrics as a function of the duration of the representative period; note the
425 non-linear time axis. It is evident from the figure that a consistent reduction both in average and in
426 variability among Monte Carlo realizations occurs as the length of the synthetic period is increased.

427 A power law was found to be a very good fit ($R^2 > 0.99$) to all four metrics shown in Figure 8, with
428 power law exponents in the range of -0.57 (for $\varepsilon_{\langle U \rangle}$) to -0.77 (for $|\varepsilon_{12}|$), indicating that the seasonal
429 profile benefits most from an extension of the representative period which is of course very intuitive. In
430 can be seen from Figure 8 that the average absolute error of the average wind speed $\langle U \rangle$ (i.e. $\varepsilon_{\langle U \rangle}$) is of the
431 order of 1.5% for a one-year representative period, down from about 6% if a 36-day period is chosen.
432 Using two years as a representative period provides only a marginal improvement to about 1%.

433 While Figure 9 strongly suggests a significant improvement in long-term prediction accuracy as a
434 function of the length of the representative period it still remained to be assessed to what extent these
435 improvements are significant in a statistical sense. In order to address this question another round of
436 Kruskal-Wallis (KW) non-parametric group comparisons was conducted where each group consisted of
437 the long-term predictions elaborated with representative periods (RP) of varying length (as before) and a
438 given error metric. As shown by the results in KW ranks do indeed decline consistently as a function of
439 the RP length, increasing the RP length not always guarantees a significant improvement in prediction
440 accuracy, e.g. upon increasing the RP length from 120 to 180 days. However, it is evident from the figure
441 that the expected error in a given error metric can be tuned by selecting the appropriate length of the
442 representative period.

443 In order to further elaborate on this aspect, which directly translates into the computational effort
444 required to achieve a given level of accuracy, an attempt was made to directly relate the BAMS scores (as
445 calculated by equation (7)) with the expected accuracy of the (SKIRON) mesoscale simulation conducted
446 with the reduced synthetic period instead of the full long-term period. Figure 10 shows the corresponding
447 results. It can be seen that in all six error metrics shown the relationship between the error and the BAMS
448 scores is approximately linear, allowing the level of computational effort (length of the period to be
449 simulated) to be adjusted to the level of final accuracy required. Evidently, a reduction of the error
450 associated with the selection of the synthetic period is only worthwhile to the extent that this error is
451 larger than or comparable to other sources of errors associated with the mesoscale modelling process.

452 *3.6 Savings in computational resources*

453 In an effort to quantify the computational cost of the main contribution of the present work, the BAMS
454 method, was run for the Iberian domain on an Intel Xeon processor (using a single core@2.53GHz),
455 requiring 38 hours of execution. However, as the method can be fully parallelized the same task using the
456 full 12-core set of the processor would require less than four hours. With this computational investment
457 the required execution time of a downscaling process for the same domain can be reduced by a factor of
458 20 to 30 (depending on the exact lengths of both the representative and the required long-term period).
459 This translates into a speedup in a cloud computing environment from typically one month to little over
460 one day, saving both considerable amounts of financial resources and allowing for a greater range of
461 sensitivity studies.

462 **4. Summary and conclusions**

463 A new method for the statistical-dynamical downscaling of the resource, termed the BAMS method (*Best*
464 *Annual Mean and Standard Deviation*), has been introduced. This method allows to construct a synthetic
465 representative period with an optimal similarity compared to a long-term period of interest; this reduced
466 period can then be used for dynamical downscaling purposes at a fraction of the time that would have
467 been required if the long-term period had been downscaled directly. The BAMS method is an algorithmic
468 approach which does not require specific knowledge about the wind climate in the region of interest, as
469 opposed to statistical-dynamical downscaling methods based on classifications of wind resource data
470 prior to the construction of the representative data set. It has the distinctive advantage over other
471 algorithmic methods that it does not require the specification of *a priori* information such as a number of
472 retained orthogonal components (as in the case of *Principal Component Analysis*) or prior selection of the
473 number and structure of cluster arrays (as in the case of *Self-Organizing Maps*). Much as many
474 algorithmic but unlike most classification-based methods BAMS retains control of the temporality of the
475 data which is very important in the case of wind resource studies where seasonal and daily profiles are of
476 great interest. The BAMS method is based on the use of mesoscale input information, in the present case
477 mean sea level pressure (MSLP) maps, which allow for the construction of a regionally representative
478 period rather than one tuned to a specific site of interest. It was shown that the new method clearly
479 outperforms the other methods in all error metrics studied, indicative of different characteristics of the
480 wind resource. Even with the continuous rise in computing power available to general and research users
481 long-term mesoscale simulation remain a significant computational challenge, and it is believed that the
482 approach presented in the current work significantly contributes to the progress of the field.

483 **Acknowledgements**

484 Support from Tecnológico de Monterrey through the Energy and Climate Change group, as well as a
485 tuition stipend for one of the authors (R.Ch.) is greatly acknowledged. R.Ch. also appreciates ample
486 support from CONACYT (Mexico) for a Ph.D. scholarship and his research year abroad through the
487 *Becas Mixtas* fund. One of the authors (J.A.) acknowledges the support of the UK National Centre for
488 Earth Observation (NCEO).

489 **References**

- 490 Barranger N., Kallos G. 2012. The use of an atmospheric model to solve wind forcing turbulent flows
491 over complex terrain for wind resource assessment. In: *EMS general assembly*. Łódź.,
- 492 Castro CL. 2005. Dynamical downscaling: Assessment of value retained and added using the Regional
493 Atmospheric Modeling System (RAMS). *J. Geophys. Res.* **110**:1–21.
- 494 Chávez-Arroyo R., Lozano-Galiana S., Sanz-Rodrigo J., Probst O. 2013. On the Application of Principal
495 Component Analysis for Accurate Statistical-dynamical Downscaling of Wind Fields. *Energy*
496 *Procedia* **40**:67–76.
- 497 Chávez-Arroyo R., Lozano-Galiana S., Sanz-Rodrigo J., Probst O. 2015. Statistical-dynamical
498 downscaling of wind fields using self-organizing maps. *Appl. Therm. Eng.* **75**:1201–1209.
- 499 Chávez-Arroyo R., Probst O. 2015. Quality assurance of near-surface wind velocity measurements in
500 Mexico. *Meteorol. Appl.* **22**:165–177.
- 501 Cutler NJ., Jorgensen BH., Ersboll BK., Badger J. 2006. Class generation for numerical wind atlases.
502 *Wind Eng.* **30**:401–415.

- 503 Frank HP., Landberg L. 1997. Modeling the wind climate of Ireland. *Bound. Lay. Meteorol.* **85**:359–378.
- 504 Frey-Buness F., Heimann D., Sausen R. 1995. A Statistical-Dynamical Downscaling Procedure for
505 Global Climate Simulations. *Theor. Appl. Clim.* **50**:117–131.
- 506 Fuentes U., Heimann D. 2000a. An Improved Statistical-Dynamical Downscaling Scheme and its
507 Application to the Alpine Precipitation Climatology. *Theor. Appl. Clim.* **65**:119–135.
- 508 Fuentes U., Heimann D. 2000b. An Improved Statistical-Dynamical Downscaling Scheme and its
509 Application to the Alpine Precipitation Climatology. *Theor. Appl. Clim.* **65**:119–135.
- 510 García-Bustamante E., González-Rouco JF., Navarro J., Xoplaki E., Jiménez P a., Montávez JP. 2012.
511 North Atlantic atmospheric circulation and surface wind in the Northeast of the Iberian Peninsula:
512 uncertainty and long term downscaled variability. *Clim. Dyn.* **38**:141–160.
- 513 Gastón M., Pascal E., Frías L., Martí I., Irigoyen U., Cantero E., Sergio L., Loureiro Y. 2008. Wind
514 resources map of Spain at mesoscale . Methodology and validation on. In: *European Wind Energy*
515 *Conference and Exhibition*. Brussels, Belgium,.
- 516 Hagemann K. 2008. Mesoscale Wind Atlas of South Africa. University of Cape Town.
- 517 Hahmann AN., Lange J., Peña A., Hasager CB. 2012. *The NORSEWInD numerical wind atlas for the*
518 *South Baltic*. Roskilde, Denmark.
- 519 Hochberg Y., Tamhane AC. 1987. *Multiple Comparison Procedures*. NJ: John Wiley & Sons, Inc.
- 520 Kallos G., Katsafados P., Papadopoulos A. 2005. *The Weather Forecasting System SKIRON - Description*
521 *of the model*. Athens, Greece.
- 522 Kanamitsu M., Ebisuzaki W., Woollen J., Yang S-K., Hnilo JJ., Fiorino M., Potter GL. 2002. NCEP–
523 DOE AMIP-II Reanalysis (R-2). *B. Am. Meteorol. Soc.* **83**:1631–1643.
- 524 Kruskal WH., Wallis WA. 1952. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.*
525 **47**:583–621.
- 526 Liléo S., Berge E., Undheim O., Klinkert R., Bredesen RE. 2013. Long-term correction of wind
527 measurements. State-of-the-art, guidelines and future work. In: *EWEA*. 1–10.
- 528 Martínez Y., Yu W., Lin H. 2013. A new Statistical–Dynamical Downscaling procedure based on EOF
529 analysis for regional time series generation. *J. Appl. Meteor. Clim.* **52**:935–952.
- 530 Mesinger F., DiMego G., Kalnay E., Mitchell K., Shafran PC., Ebisuzaki W., Jović D., Woollen J.,
531 Rogers E., Berbery EH., Ek MB., Fan Y., Grumbine R., Higgins W., Li H., Lin Y., Manikin G.,
532 Parrish D., Shi W. 2006. North American Regional Reanalysis. *Bull. Am. Meteorol. Soc.* **87**:343–
533 360.
- 534 Pielke RAS. 2002. *Mesoscale Meteorological Modeling*. Academic Press.
- 535 Rienecker MM., Suarez MJ., Gelaro R., Todling R., Bacmeister J., Liu E., Bosilovich MG., Schubert SD.,
536 Takacs L., Kim G., Bloom S., Chen J., Collins D., Conaty A., da Silva A., Gu W., Joiner J., Koster
537 RD., Lucchesi R., Molod A., Owens T., Pawson S., Pegion P., Redder CR., Reichle R., Robertson
538 F., Ruddick A., Sienkiewicz M., Woollen J. 2011. MERRA - NASA's Modern-Era Retrospective
539 Analysis for Research and Applications. *J. Clim.* **24**:1–41.

- 540 Rife DL., Vanvyve E., Pinto JO., Monaghan AJ., Davis CA., Poulos GS. 2013. Selecting representative
541 days for more efficient dynamical climate downscaling: Application to wind energy. *J. Appl.*
542 *Meteorol. Climatol.* **52**:47–63.
- 543 Romo Perea A., Amezcua J., Probst O. 2011. Validation of three new measure-correlate-predict models
544 for the long-term prospection of the wind resource. *J. Renew. Sustain. Energy* **3**:23105.
- 545 Tammelin B., Vihma T., Atlaskin E., Badger J., Fortelius C., Gregow H., Horttanainen M., Hyvönen R.,
546 Kilpinen J., Latikka J. 2013. Production of the Finnish Wind Atlas. *Wind Energy* **16**:19–35.
- 547 Tukey JW. 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics* **5**:99–114.
- 548 Vanvyve E., Delle Monache L., Monaghan AJ., Pinto JO. 2015. Wind resource estimates with an analog
549 ensemble approach. *Renew. Energy* **74**:761–773.
- 550 Wang Z., Bovik AC., Sheikh HR., Simoncelli EP. 2004. Image quality assessment: From error visibility
551 to structural similarity. *IEEE Trans. Image Process.* **13**:600–612.
- 552 Wilks DS. 2006. Statistical Methods in the Atmospheric Sciences. *Int. Geophys. Ser.* **102**:380–380.
- 553
- 554

555 **Figure captions**

556 **Figure 1** Location of the ASOS stations used in the present work. The MERRA reference locations REA1
557 and REA2 for the REA method in the case of the local validation are indicated in the main graph. *Inset:*
558 Locations of the reference locations R1 through R6 used for the REA method in the case of the mesoscale
559 validation.

560 **Figure 2** Comparison of the basic wind statistics for the long-term period (LT) and the representative year
561 (RY) determined by one realization of the BAMS method for one ASOS (NM19) from the Navarra
562 network. (a) Histograms of the wind speed distributions, (b) daily profiles, (c) seasonal profiles, and (d)
563 wind roses.

564 **Figure 3** (a) Absolute wind speed error for one realization of the BAMS method. (b) Absolute error of the
565 wind energy density.

566 **Figure 4** Results of the Kruskal-Wallis / Tukey-Kramer pairwise group comparison for the six methods
567 under study using the Navarra observational data for validation

568 **Figure 5** Results of the Kruskal-Wallis / Tukey-Kramer pairwise group comparison for the six methods
569 under study using the SKIRON mesoscale data for the Iberian Peninsula for validation

570 **Figure 6** Results of the ANOVA / Tukey-Kramer pairwise group comparison for six implementations of
571 the REA method

572 **Figure 7** Maps of the variance of the wind speed of the representative year obtained from $M=100$
573 realizations of each method

574 **Figure 8** Box-plots with several error statistics as a function of the length of the synthetic period selected
575 for the case of the BAMS method. Each sample in the plots is built with the average of results obtained at
576 the 99050 SKIRON grid points of each metric. Red asterisks denote the mean of each error metric over
577 the 100 trials. Green circles (with corresponding vertical axis on the right of each plot) denote the
578 standard deviation of the trials.

579 **Figure 9** Results of the multiple comparisons test (Kruskall-Wallis / Turkey-Kramer) for SKIRON-
580 derived wind fields and different length of the representative period selected through the BAMS method.
581 Each metric include their mean and their narrow confidence intervals for the join significance (with
582 $\alpha = 0.05$) constructed from the Kruskall-Wallis results.

583 **Figure 10** Correlation between the BAMS scores and six different error metrics for the case of the
584 SKIRON mesoscale simulations

585

586

587

588

589

590

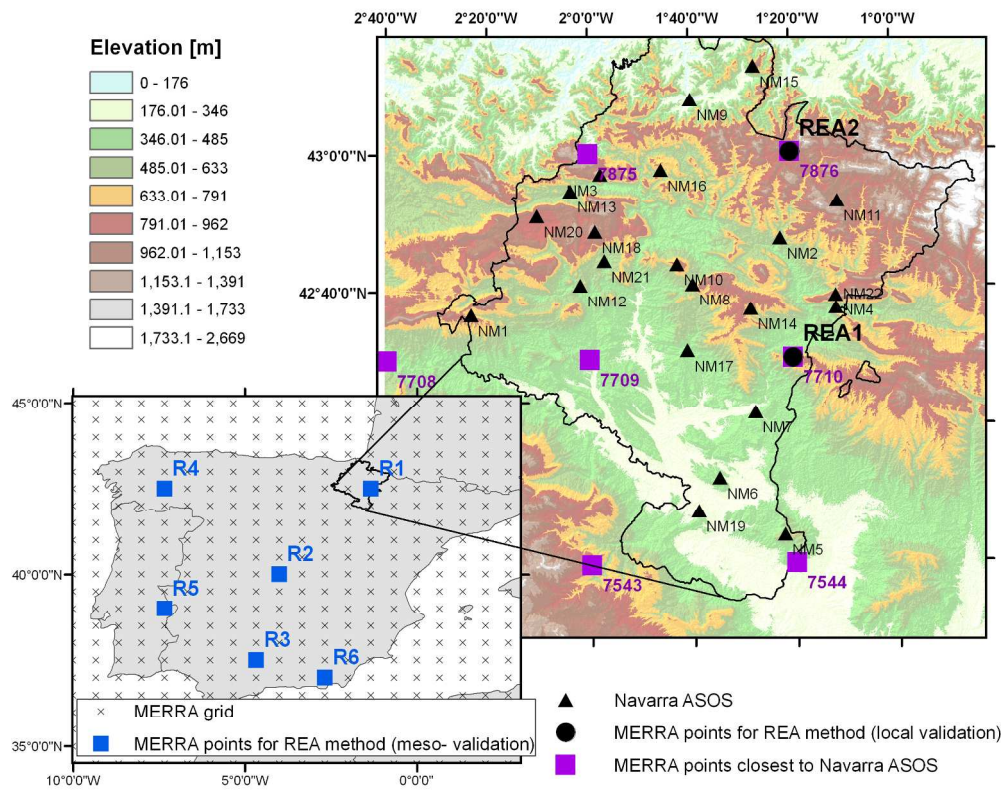
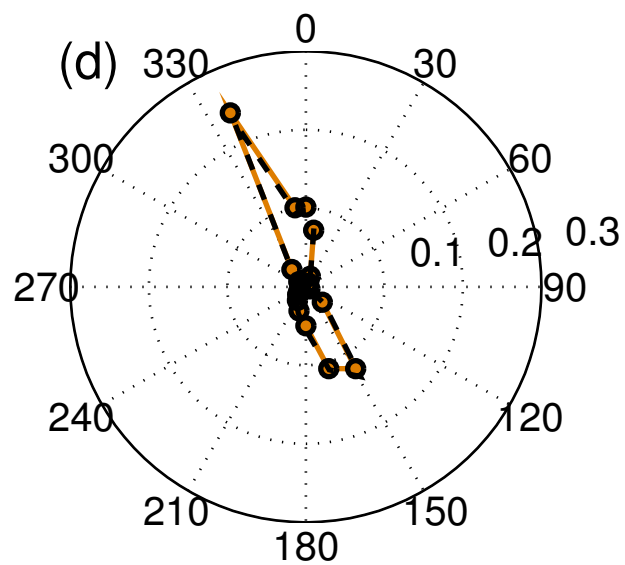
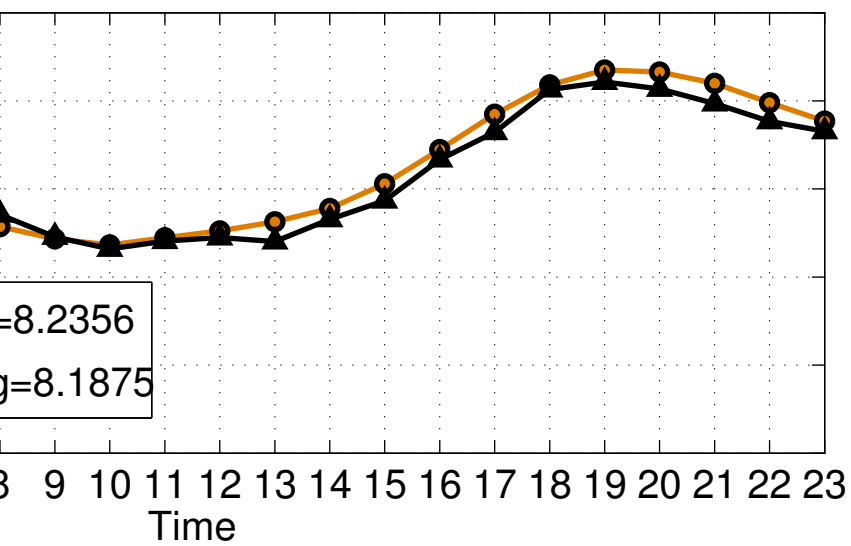
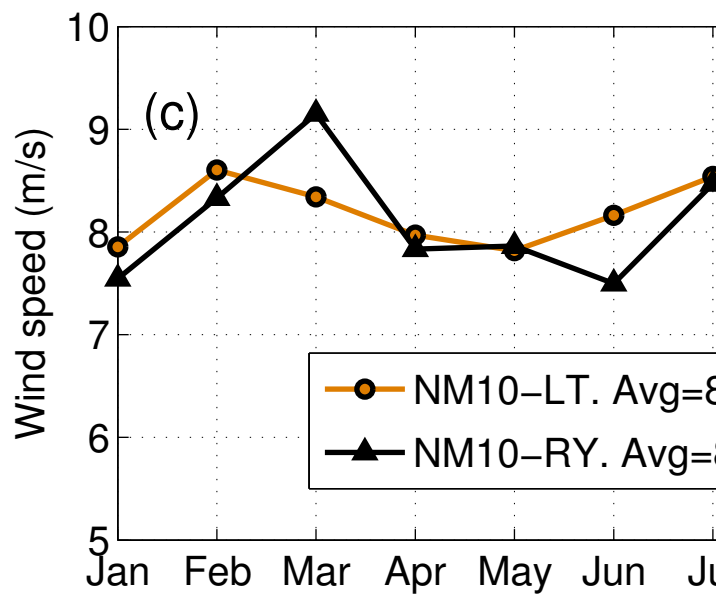
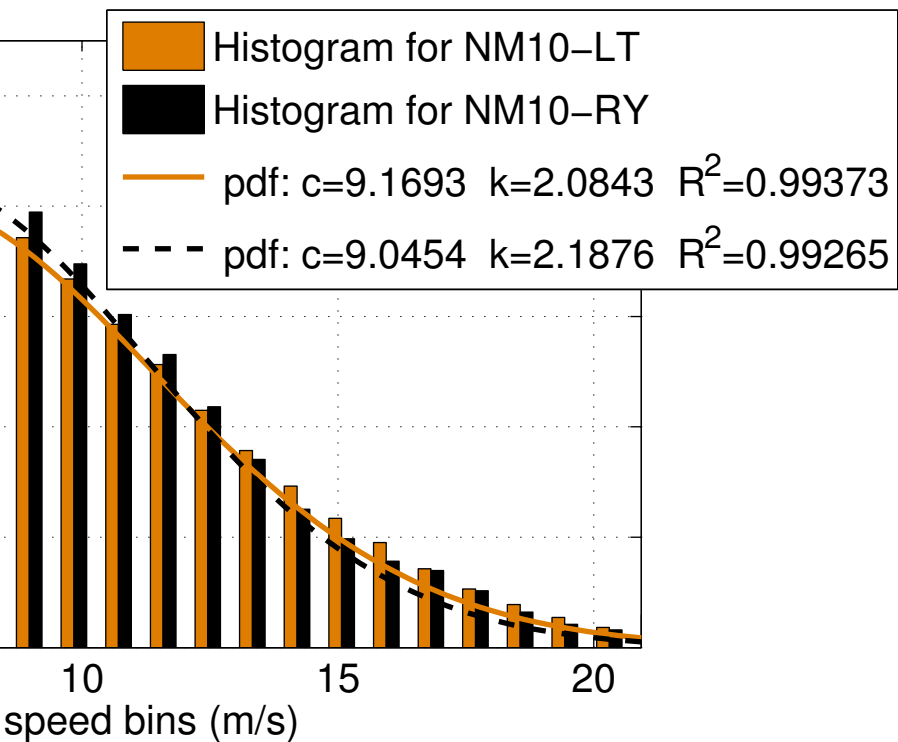


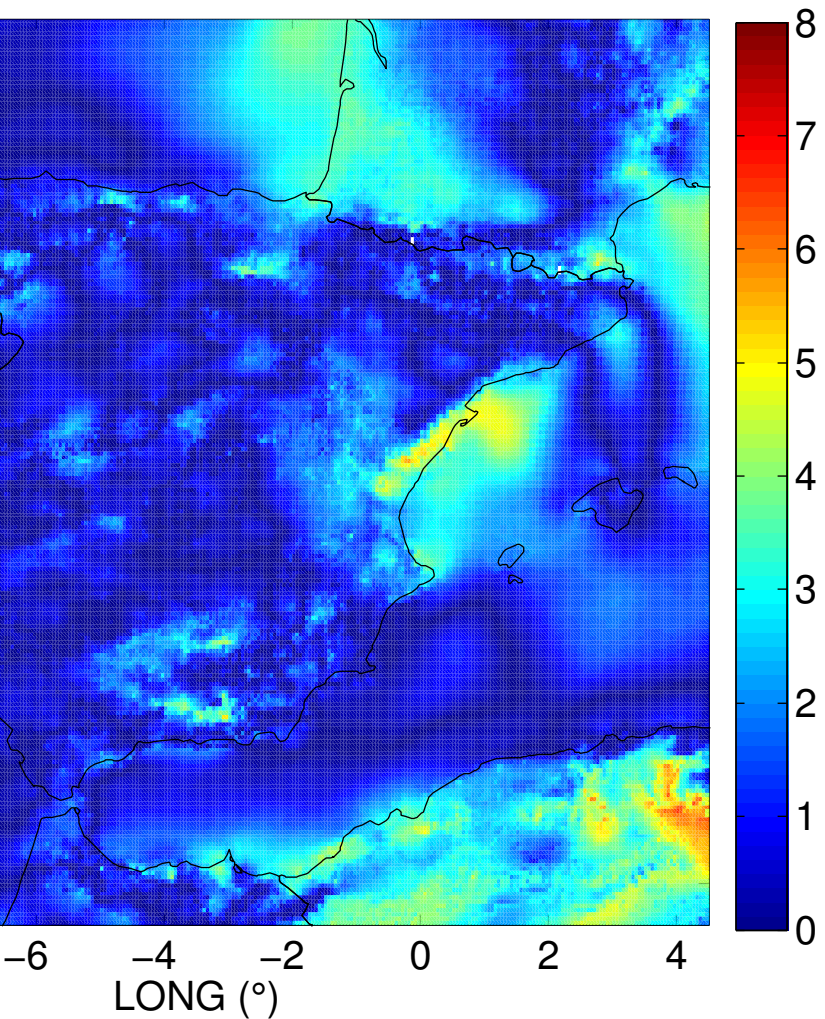
Figure 1 Location of the ASOS stations used in the present work. The MERRA reference locations REA1 and REA2 for the REA method in the case of the local validation are indicated in the main graph. Inset: Locations of the reference locations R1 through R6 used for the REA method in the case of the mesoscale validation.

Figure 1

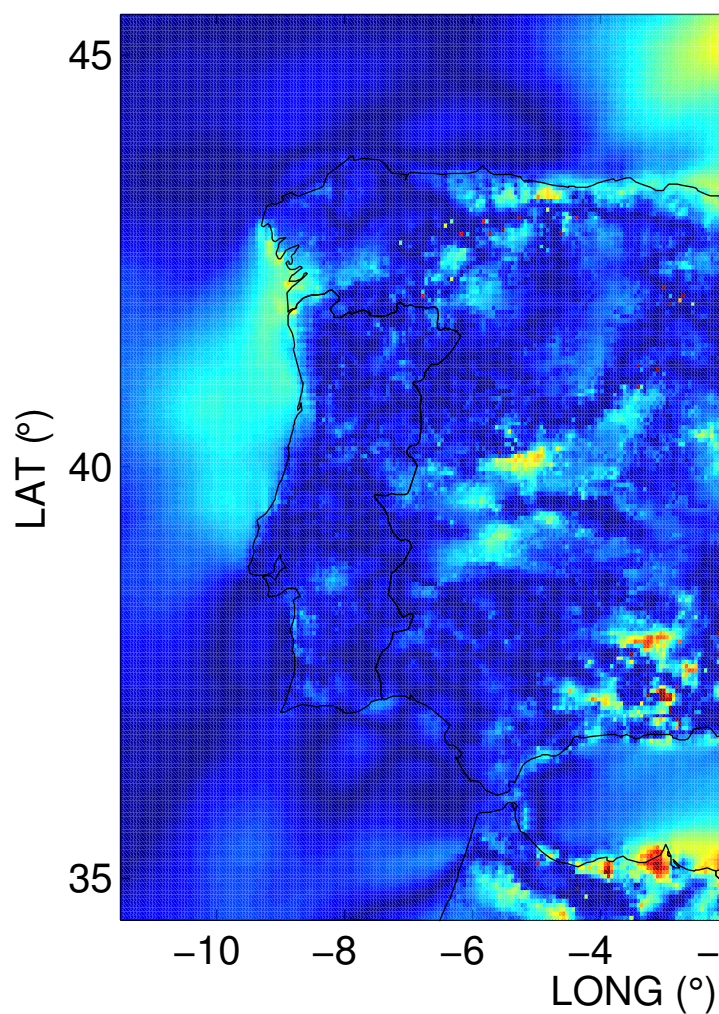
215x177mm (300 x 300 DPI)



speed absolute error, $\epsilon_{\langle U \rangle}$ (%)



Energy density absolute



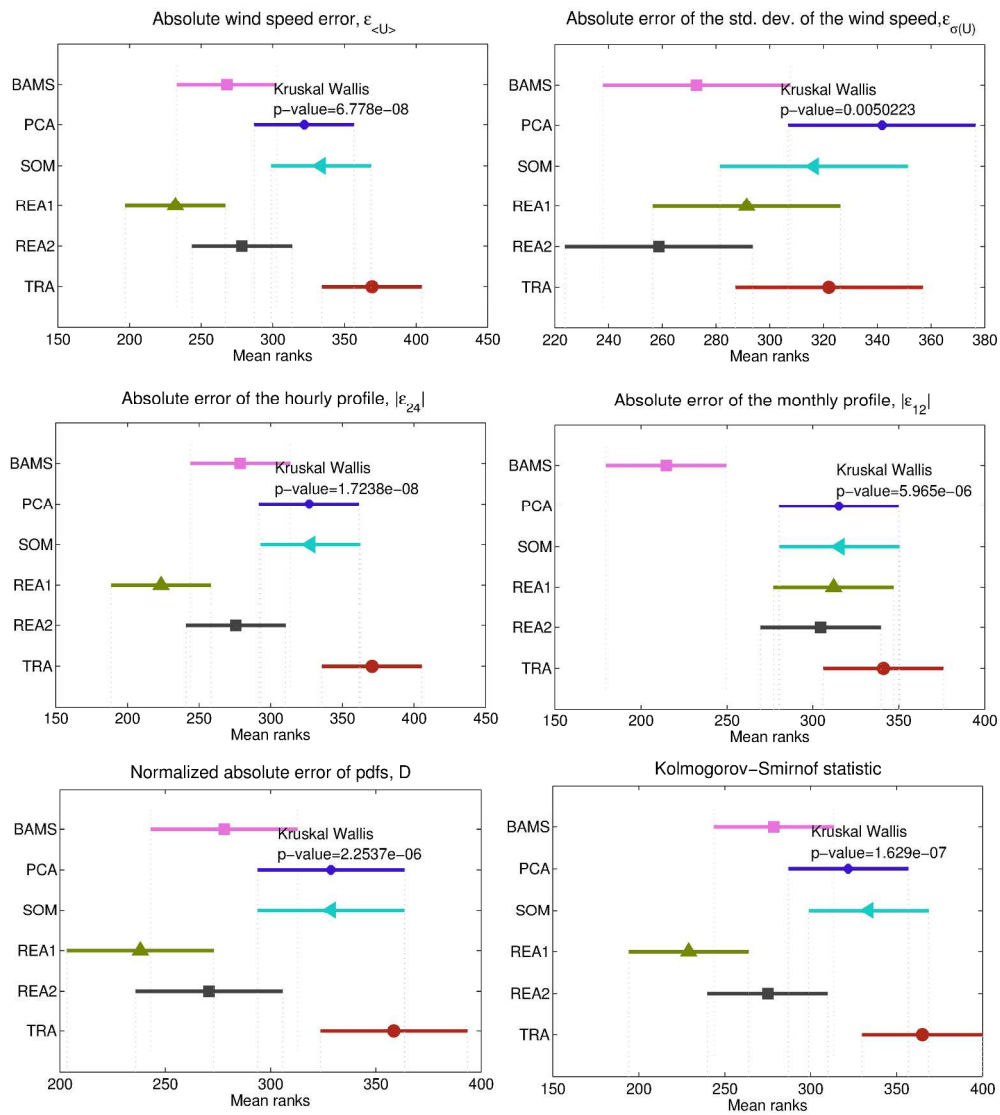


Figure 4 Results of the Kruskal-Wallis / Tukey-Kramer pairwise group comparison for the six methods under study using the Navarra observational data for validation

Figure 4
360x402mm (300 x 300 DPI)

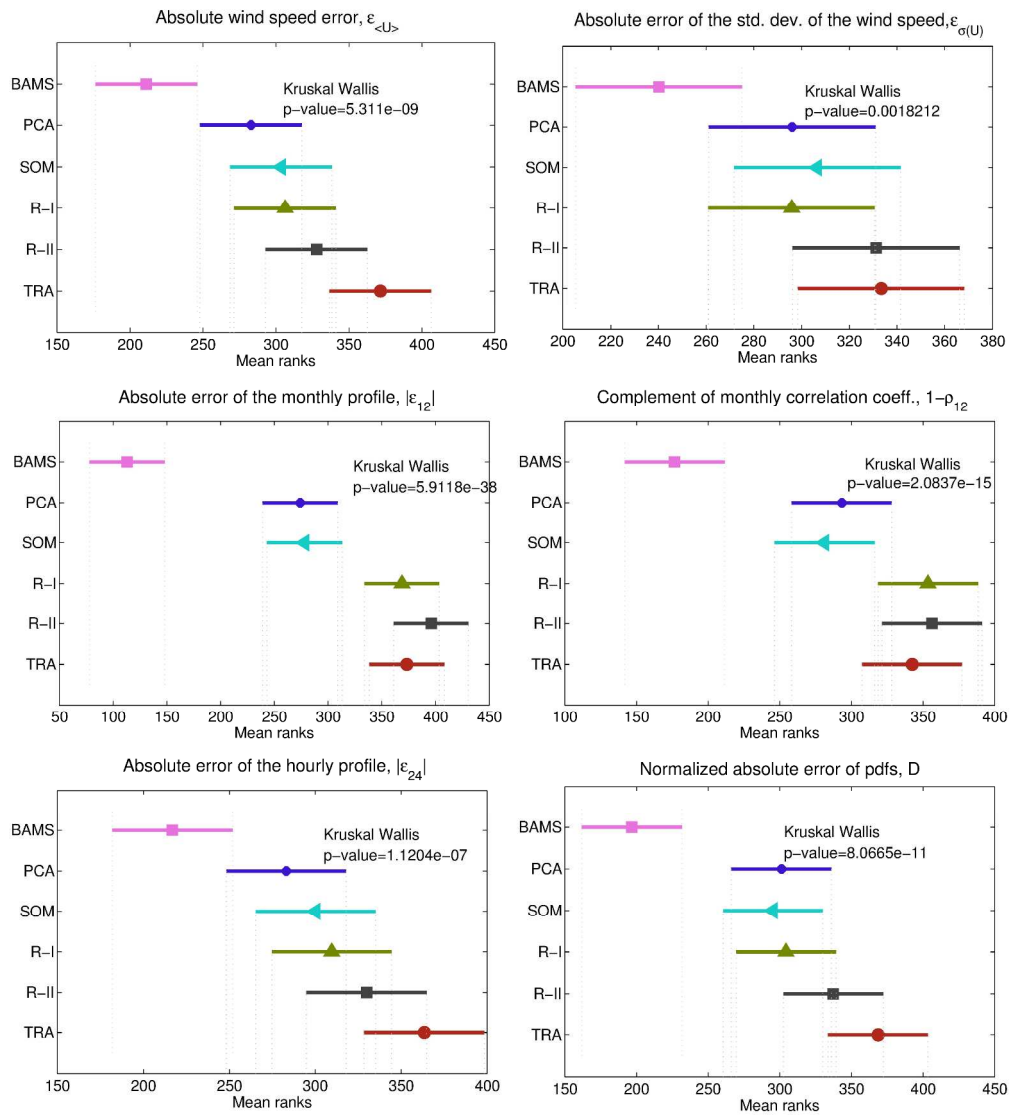


Figure 5 Results of the Kruskal-Wallis / Tukey-Kramer pairwise group comparison for the six methods under study using the SKIRON mesoscale data for the Iberian Peninsula for validation

Figure 5
360x402mm (300 x 300 DPI)

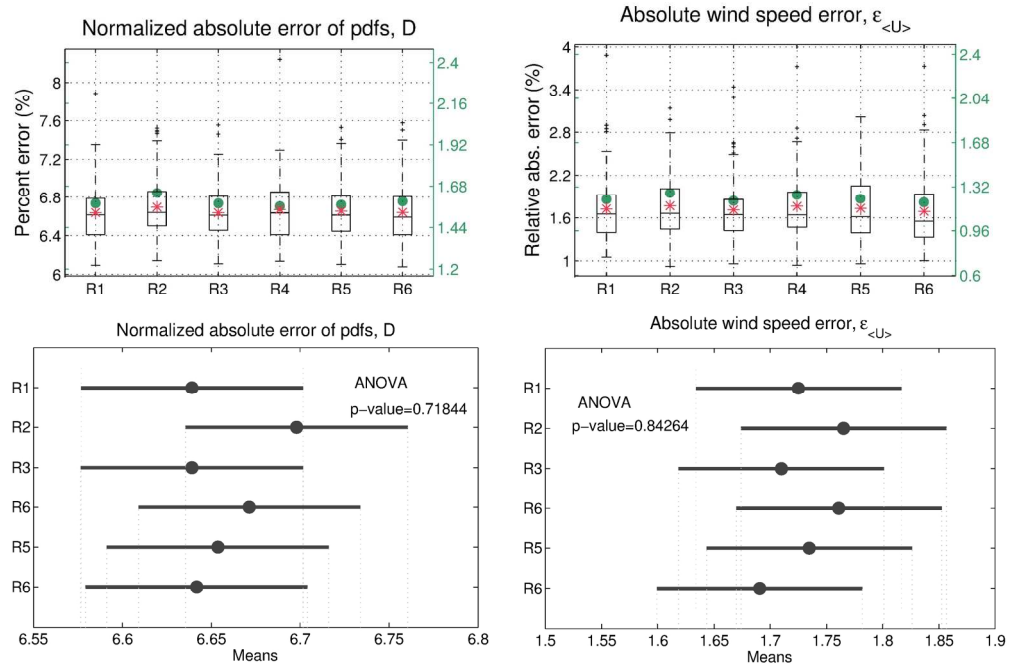
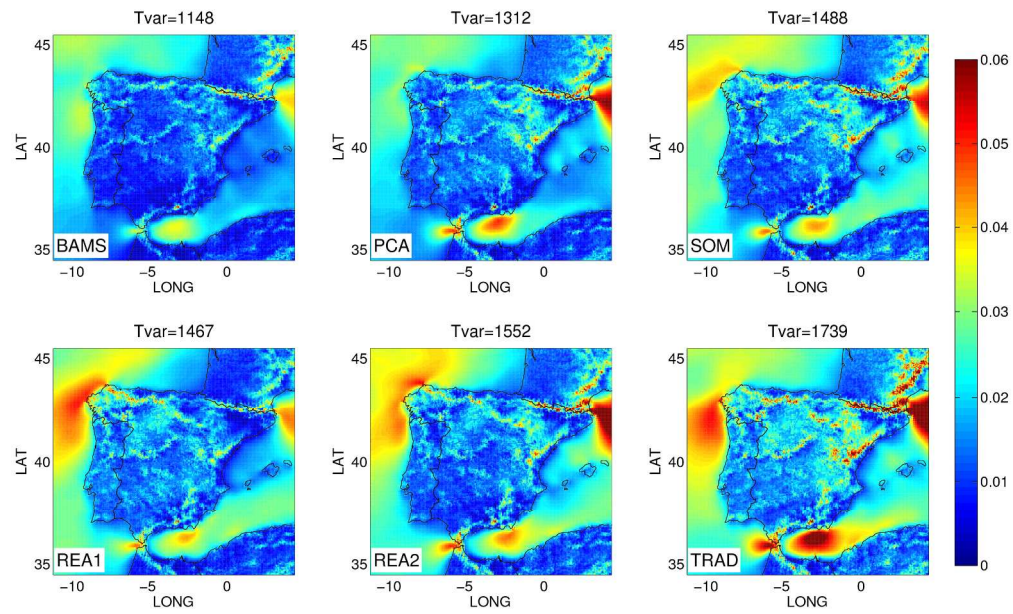


Figure 6 Results of the ANOVA / Tukey-Kramer pairwise group comparison for six implementations of the REA method

Figure 6
187x129mm (300 x 300 DPI)

Review



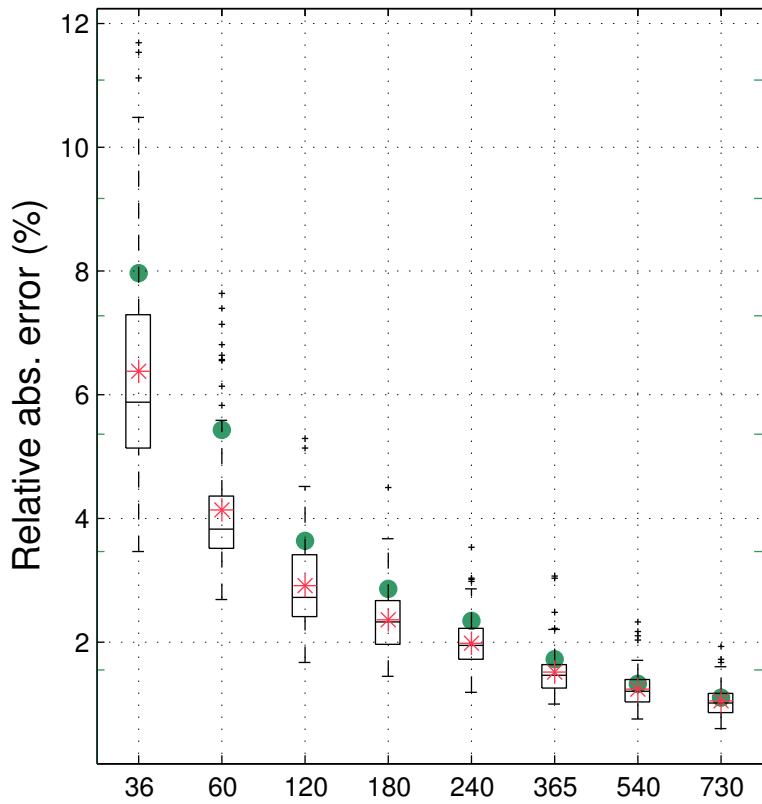
Maps of the variance of the wind speed of the representative year obtained from $M=100$ realizations of each method

Figure 7

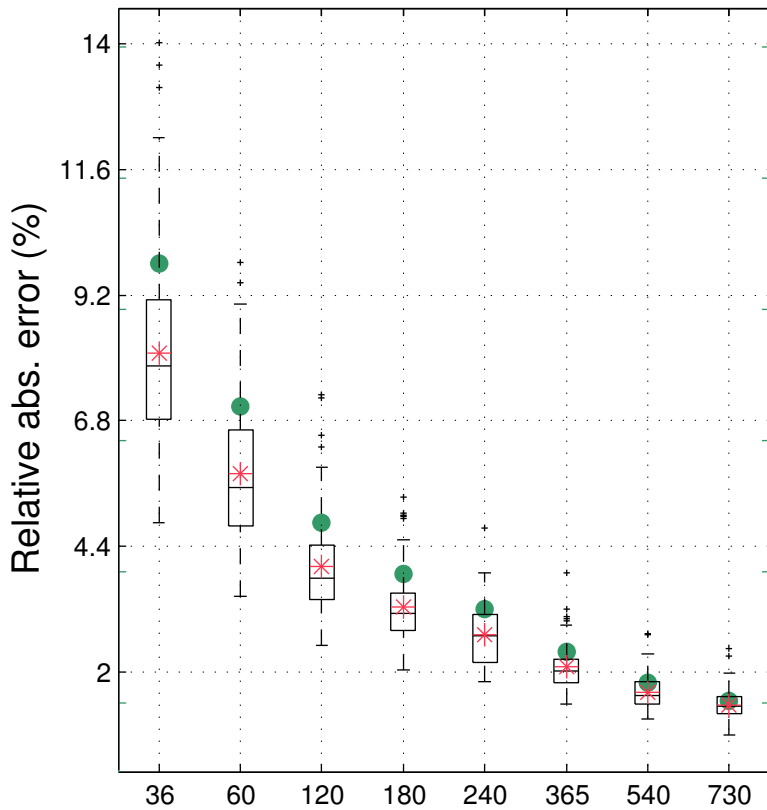
706x433mm (96 x 96 DPI)

Review

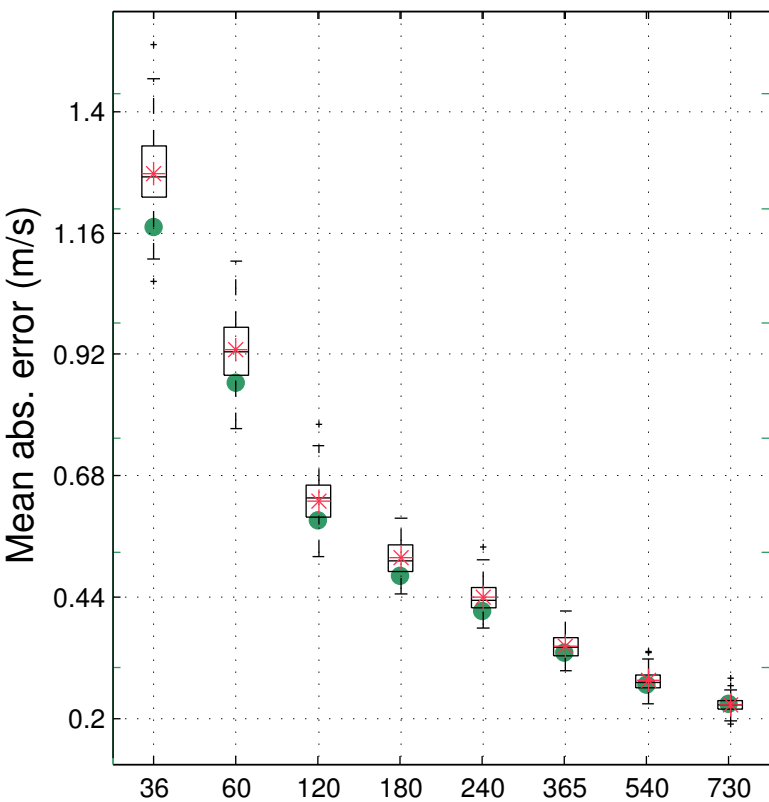
$$\epsilon_{\langle U \rangle}$$



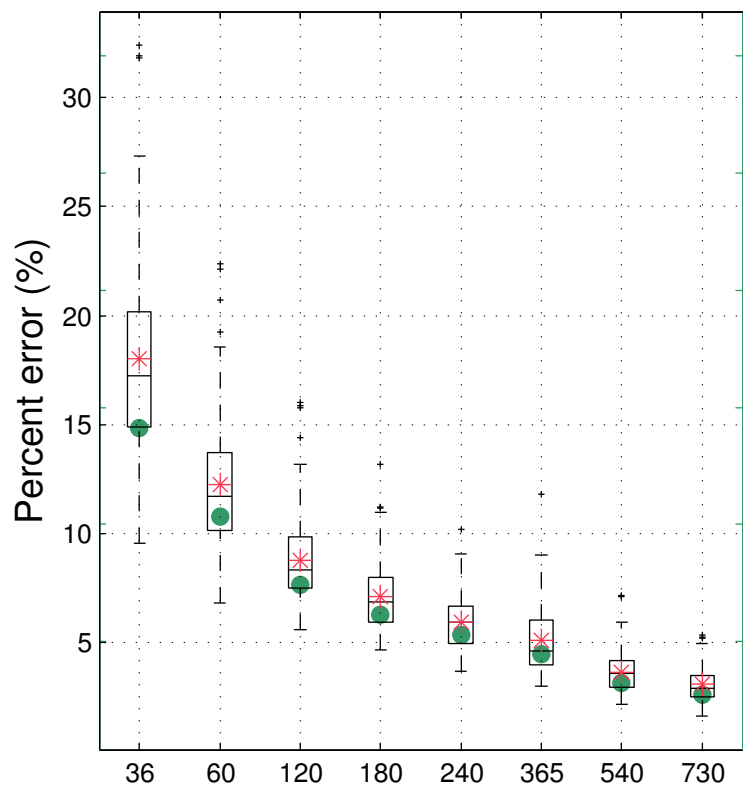
$$\epsilon_{\sigma(U)}$$

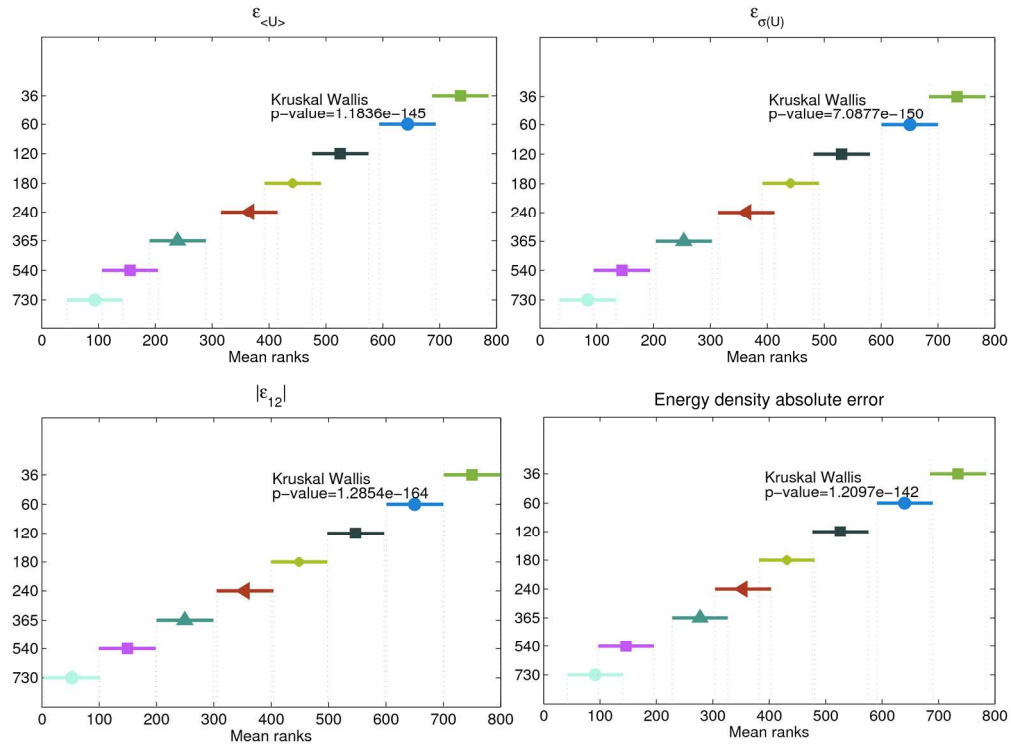


$$|\epsilon_{12}|$$



Energy density absolute error

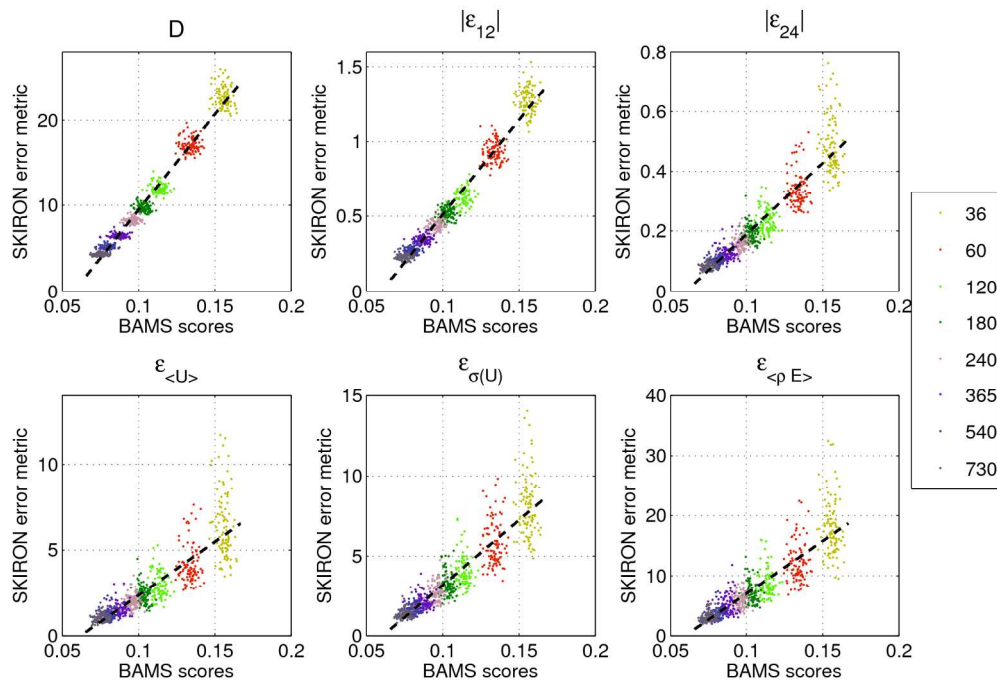




Results of the multiple comparisons test (Kruskal-Wallis / Turkey-Kramer) for SKIRON-derived wind fields and different length of the representative period selected through the BAMS method. Each metric include their mean and their narrow confidence intervals for the join significance (with $\alpha = 0.05$) constructed from the Kruskal-Wallis results

Figure 9
675x500mm (96 x 96 DPI)

view



Correlation between the BAMS scores and six different error metrics for the case of the SKIRON mesoscale simulations

Figure 10

540x372mm (96 x 96 DPI)

Review