Informes Técnicos Ciemat

1404

Junio, 2017

An Approach about Uncertainty on Statistical Downscaling

E. E. Lucio-Eceiza
J. Navarro
E. García-Bustamante
J. F. González-Rouco
C. Rojas-Labanda

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD

Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

Informes Técnicos Ciemat **1404**
Junio, 2017

An Approach about Uncertainty on Statistical Downscaling

E. E. Lucio-Eceiza[1]
J. Navarro[2,3]
E. García-Bustamante[2,3]
J. F. González-Rouco[1,3]
C. Rojas-Labanda[1]

1 *Universidad Complutense de Madrid*

2 *CIEMAT*

1 *Unidad Mixta CIEMAT-UCM*

Departamento de Energía

Toda correspondencia en relación con este trabajo debe dirigirse al Servicio de Información y Documentación, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Ciudad Universitaria, 28040-MADRID, ESPAÑA.

Las solicitudes de ejemplares deben dirigirse a este mismo Servicio.

Los descriptores se han seleccionado del Thesauro del DOE para describir las materias que contiene este informe con vistas a su recuperación. La catalogación se ha hecho utilizando el documento DOE/TIC-4602 (Rev. 1) Descriptive Cataloguing On-Line, y la clasificación de acuerdo con el documento DOE/TIC.4584-R7 Subject Categories and Scope publicados por el Office of Scientific and Technical Information del Departamento de Energía de los Estados Unidos.

Catálogo general de publicaciones oficiales
http://www.060.es

Depósito Legal: M-26385-2011
ISSN: 1135-9420
NIPO: 058-17-019-0

Editorial CIEMAT

CLASIFICACIÓN DOE Y DESCRIPTORES

S17
DATA COVARIANCES; STATISTICAL MODELS; CLIMATE MODELS; WIND; ATMOSPHERIC CIRCULATION

An Approach about Uncertainty on Statistical Downscaling

Lucio-Eceiza, E. E.; Navarro, J.; García-Bustamante, E.; González-Rouco, J. F.; Rojas-Labanda,C.
44 pp. 48 refs. 17 figs. 3 tablas

Abstract:

This report deals with statistical downscaling strategies for estimating local surface wind with application to wind resource assessment. In particular, it summarizes the downscaling strategy used in the High Performance Computing for Energy (HPC4E) European project Task 4.2 (statistical downscaling) by The Center for Energy, Environmental and Technological Research (CIEMAT).

This report aims at providing an understanding of the regional/local climatological variability of the wind through the identification of relationships between the most important large-scale circulations and their associated local wind patterns or surface wind. The analysis of the variability of wind at regional scales and for monthly resolutions is conducted via a statistical method mainly based on Empirical Orthogonal Function (EOF) and Canonical Correlation Analysis (CCA). This is a multivariate statistical technique that isolates sets of predictor and predictand variables that are optimally correlated. In addition, the sensitivity associated to various methodological aspects is also explored, based on the argument that a single selection of the model set up cannot account, by itself, for the uncertainty in estimations that arise in the downscaling step. The evaluation of this uncertainty provides confidence in the robustness of the model skill in reproducing the observed wind and helps discriminating whether certain parameters have a decisive influence in the quality of estimates.

Un Ejercicio de Incertidumbres Enmetodologías de downscalingestadístico

Lucio-Eceiza, E. E.; Navarro, J.; García-Bustamante, E.; González-Rouco, J. F.; Rojas-Labanda,C.
44 pp. 48 refs. 17 figs. 3 tables

Resumen:

Este informe detalla la utilización de estrategias de dowscaling estadístico para la estimación del viento en superficie local aplicada a la evaluación de recurso eólico. En particular, resume la estrategia de downscaling estadístico utilizada para el High Performance Computing for Energy (HPC4E) European project Task 4.2 (statistical downscaling) por el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT).

Este informe pretende ayudar a comprender la variabilidad climatológica regional/local del viento a través de la identificación de relaciones entre circulaciones predominantes a gran escala y sus patrones locales asociados de viento local en superficie. El estudio de la variabilidad del viento a escalas regionales y en resoluciones mensuales se realiza mediante métodos estadísticos basados principalmente en Funciones Ortogonales Empíricas (EOF) y Análisis de Correlación Canónica (CCA). Ésta es una técnica estadística multivariante que aísla conjuntos de variables predictoras y predictandas que están óptimamente correlacionadas. Adicionalmente, también se explora la sensibilidad asociada a varios aspectos metodológicos. Para ello nos basamos en el argumento de que una única selección en los parámetros

An Approach about uncertainty on statistical downscaling

Etor E. Lucio-Eceiza[1], Jorge Navarro[2,3], Elena García-Bustamante[2,3], J. Fidel Gonzalez-Rouco[1,3] y Cristina Rojas-Labanda[1]

Final report
Versión: v01

7 de Junio de 2017

[1] Universidad Complutense de Madrid
[2] CIEMAT
[3] Unidad Mixta CIEMAT-UCM

Table of Contents

An approach about uncertainty on statistical downscaling 3
Un ejercicio de incertidumbres en metodologías de downscaling estadístico................. 3
Executive Summary.. 5
1. Introduction 6
2. Data 9
2.1 Observational datasets, predictands... 9
2.2 Large scale variables, predictors. 17
3. Statistical Downscaling Methodology. 19
4. Downscaling Experiment: reference case 21
4.1 Large and local scale coupled dynamics 21
4.2 Estimations derived from the reference case 28
5. Uncertainty assessment. 29
5.1 Uncertainties related to parameter configuration. 29
5.1.1 Spatial uncertainty 30
5.1.2 Temporal uncertainty... 32
5.2 Uncertainties related to predictor source 34
6. Conclusions.. 37
Acknowledgements 40
References 41

An approach about uncertainty on statistical downscaling

This report deals with statistical downscaling strategies for estimating local surface wind with application to wind resource assessment. In particular, it summarizes the downscaling strategy used in the High Performance Computing for Energy (HPC4E) European project Task 4.2 (statistical downscaling) by The Center for Energy, Environmental and Technological Research (CIEMAT).

This report aims at providing an understanding of the regional/local climatological variability of the wind through the identification of relationships between the most important large-scale circulations and their associated local wind patterns or surface wind. The analysis of the variability of wind at regional scales and for monthly resolutions is conducted via a statistical method mainly based on Empirical Orthogonal Function (EOF) and Canonical Correlation Analysis (CCA). This is a multivariate statistical technique that isolates sets of predictor and predictand variables that are optimally correlated. In addition, the sensitivity associated to various methodological aspects is also explored, based on the argument that a single selection of the model set up cannot account, by itself, for the uncertainty in estimations that arise in the downscaling step. The evaluation of this uncertainty provides confidence in the robustness of the model skill in reproducing the observed wind and helps discriminating whether certain parameters have a decisive influence in the quality of estimates.

Un ejercicio de incertidumbres en metodologías de downscaling estadístico

Este informe detalla la utilización de estrategias de dowscaling estadístico para la estimación del viento en superficie local aplicada a la evaluación de recurso eólico. En particular, resume la estrategia de downscaling estadístico utilizada para el High Performance Computing for Energy (HPC4E) European project Task 4.2 (statistical downscaling) por el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT).

Este informe pretende ayudar a comprender la variabilidad climatológica regional/local del viento a través de la identificación de relaciones entre circulaciones predominantes a gran escala y sus patrones locales asociados de viento local en superficie. El estudio de la variabilidad del viento a escalas regionales y en resoluciones mensuales se realiza mediante métodos estadísticos basados principalmente en Funciones Ortogonales Empíricas (EOF) y Análisis de Correlación Canónica (CCA). Ésta es una técnica estadística multivariante que aísla conjuntos de variables predictoras y predictandas que están óptimamente correlacionadas. Adicionalmente, también se explora la sensibilidad asociada a varios aspectos metodológicos. Para ello nos basamos en el argumento de que una única selección en los parámetros del modelo no puede, por sí misma, considerar la incertidumbre en las estimaciones que se dan durante el proceso de downscaling. La evaluación de esta incertidumbre nos ofrece confianza en la robustez de la capacidad del modelo en reproducir el viento observado y nos ayuda a discriminar entre parámetros que puedan o no tener una influencia decisiva en la calidad de dichas estimaciones.

Executive Summary

This executive summary compiles the most relevant results and conclusions achieved within the development of this report. These points are reported in bullets and referred to the corresponding section or chapter within this report where a more detailed description is provided.

An approach about uncertainty on statistical downscaling:

- The downscaling method based on EOF and CCA is able to identify the major large and local scale coupled circulation patterns;
- This methodology also offers predictive skill over the local observations, although it underestimates the variability, a common downside in linear methods;
- It is a robust statistical method as it remains basically unaffected by the use of different crossvalidation periods;
- It is, however, sensible to model configuration changes, specially in sites/wind components with larger variability;
- It is not equally sensible to every parameter, either: those with largest impact are the retained number of EOF/CCA modes and window size;
- The predictor source (i.e. used reanalysis model) on the other hand has little impact on the estimations;
- The uncertainty width can vary through time, to the point of even altering the regional dynamics.

An Approach about uncertainty on statistical downscaling

1. Introduction

Surface winds are not only governed by the radiative and rotational mechanisms that drive large scale circulation, but also by processes of different nature and lifetime. The aggregated influence of all of them shapes the complicated behaviour of the wind field at smaller temporal and spatial scales. One of the main contributors to the complexity of wind variability is the orography that induces substantial variations to the geostrophic flow. The presence of extensive geographical attributes like oceans, large mountain ranges and desserts or smaller scale terrain features like hills, valleys or urban settlements are responsible for thermally driven flows, momentum transport circulations caused by gravity waves or turbulent mixing, and forced channellings of the wind. As the local scale becomes important, more physical processes are involved in the wind circulation. This is the case, for instance, of boundary layer dynamics, which enhance or hamper local flows depending on the effectiveness of the turbulent mixing.

The generalized definition of regional scale refers to a sub-continental scale with high heterogeneity in climatic features that are the product of interactions of phenomena at multiple timescales, from intra daily to multi centennial, combining mesoscale circulations and local forcings. The need of assessing the regional climate variability can be thought in terms of two different viewpoints. On one hand, the understanding of the multiple interactions of physical mechanisms playing a role in generating climatic variability at regional spatial scales can be still considered a challenge. On the other hand, mankind is subjected to countless sociological, political, economical, environmental and cultural aspects that are very sensitive to climate variability.

The wind field can be loosely considered as the local response to the large scale circulation. This response includes, and is sometimes overridden by, the effect of the orography and a variety of additional factors such as vegetation, land-sea interactions or other thermally-driven phenomena. This combination of large and smaller scale forcings imposes a high spatial as well as temporal variability on the surface wind field. The large variability and the vectorial nature of the wind field introduces an additional complexity to its diagnosis and prediction, providing a valuable scientific interest on this topic. Exploring its variability at the regional scale involves practical applications that range from the short term wind forecasts to the assessment of climate change.

The complexity and multiplicity of the mechanisms involved at the regional scales calls for a wide spectrum of techniques and strategies that may be applied to gain insight into the regional climate problem. In practice no strategy can totally compensate the need for measurements to accurately describe the climate fields and their variations (Trenberth, 2008) and indeed many studies pursue an assessment of the surface wind field variability providing statistical descriptions of wind related variables based solely on observed records typically at the regional/local scale (García-Bustamante et al., 2012, 2013, GB12 and GB13 hereafter; and Lucio-Eceiza et al., 2017c; LE17c hereafter). The quality of observations and the scarceness, both in space and time, of measurements are two factors that hamper the informative power of such assessments. The quality of observations is usually bounded by the presence in records of inhomogeneities, gaps, missing data or errors associated with measurement errors and data management issues (Jimenez et al, 2010 and Lucio-Eceiza et al, 2017a, 2017b, LE17a,b here after). The limited coverage of observations is often a problem in many regions that difficults or even impedes variability studies on lower frequencies, for which longer records are needed. Such regional/local approaches can be complemented by analyses with a broader spatial perspective in which the regional/local variability is studied in terms of changes in the large scale circulation of the atmosphere. In fact, specific strategies can be designed to capture the interactions between large scale dynamics and the regional/local scale variability. These so called downscaling techniques (von Storch and Zwiers, 1999) can be exploited to deliver estimations and/or predictions of regional variability for different purposes.

The downscaling approaches employ large scale atmospheric circulation information to obtain estimations of variables at the regional/local scale by identifying the main statistical associations between both spatial scales (statistical downscaling) or explicitly resolving the physics involved using mesoscale models (dynamical downscaling). The large scale atmospheric state is provided by gridded observations, reanalysis data or general circulation model (GCM) outputs. One of the assets of the downscaling strategies is that they allow overcoming GCMs deficiencies in simulating the regional climate. This lack of reliability arises because of their coarse spatial resolution (ca. 100 to 300 km) which does not allow adequately resolving sub-grid scale processes that need to be parameterized. The concept of across-scales or downscaling approach is already applied in the early 1960s when methods were designed to establish classifications of the large scale atmospheric states and then relate them to the local observed features of the climate. During the subsequent decades dynamical and statistical downscaling strategies were adopted in order to satisfy the needs at regional and local scales.

The dynamical downscaling is based on the use of regional circulation models (RCMs) that solve the fundamental equations of the atmosphere yielding finer time-space simulations. These models that originally were used for numerical weather prediction issues are also called limited area models (LAMs) and evolved as sophisticated versions of GCMs over a confined geographical area with typical spatial resolutions that range from the 50 km to 10 km or even higher (Jiménez et al., 2010b).

The use of RCMs may imply however relatively high computational resources. Thus, the empirical or statistical approaches stand as a practical procedure to explore connections between the large scale forcings or predictors and the regional/local response of a climatic variable or predictands (García-Bustamante et al., 2012, 2013 and Lucio-Eceiza et al., 2017c). The computational demand is lower than in the case of RCMs and the implementation of the statistical model, although dependant on the selected strategy is to a great extent more straightforward than that of RCMs. Notwithstanding, statistical downscaling methods can provide also an understanding of the physical mechanisms responsible for the variability of regional fields. From a different perspective, they can also be applied within GCM simulations for validation purposes by assessing the ability of the models in generating consistent large scale forcings in different regions of the globe. The statistical downscaling methods require training historical data of both predictand and predictor variables in order to identify the relationships between them.

The application of statistical downscaling techniques to variables like precipitation (e.g., González-Rouco et al., 2000) or temperature (e.g., Xoplaki et al., 2003a,b) using the sea level pressure or other alternative large scale predictor fields is widespread in the literature. However studies of the wind field variability based on statistical approaches are relatively scarce (e.g., Kaas et al. 1996; Najac et al. 2009).

The transfer of information between spatial scales involves many sources of uncertainty that propagate from the global to the regional/local scale in downscaling exercises (GB12,13 and LE17c). In the context of future climate projections, the uncertainties associated with the radiative forcing are accounted for by considering a variety of climate change scenarios and by the use of a suite of GCMs to represent the intermodel variability. For instance, Pryor et al. (2006) studied the possible changes of surface winds in northern Europe through the downscaling of several GCM simulations under different scenarios. The uncertainty associated with the use of a specific model for the downscaling step can also be addressed. In the case of RCMs, the effect of changes in the discretization of the equations of motion or the different parameterizations can imply changes that contribute to uncertainty in the simulated regional field. In the case of statistical methods, the effect of applying different methodologies can also be examined (Zorita and von Storch, 1999). Even in the case of using one specific downscaling method, uncertainties arise from a number of somewhat subjective decisions taken in the design of the statistical model. Usually such decisions are founded on good practice and lead to skillful estimations of the target regional variables. However, introducing changes in the model configuration by changing parameter values, spatial domains, etc., would produce somewhat modified, though still skillful estimations (GB12, 13 and LE17c). These additional uncertainties are difficult to estimate, but they can at least be explored considering a variety of configuration designs in the downscaling approach. This path illustrates the sensitivity to changes in the model configuration, what can be regarded as a methodological variability or methodological uncertainty. Nevertheless, this type of studies are rather uncommon in the case of wind related variables, thus an exploration of the methodological sensitivity of statistically downscaled wind field seems pertinent. Furthermore, the uncertainty may also arrive from potential inaccuracies of the GCMs or reanalysis as they provide the large scale information that feeds the downscaling models (LE17c). For this reason, it is also interesting to explore the uncertainty that is associated to the use of different datasets as boundary and initial conditions in the case of the dynamical downscaling or as predictors if dealing with statistical models.

In this work we present three Statistical Downscaling (SD) exercises in which we describe the regional and local behavior of the surface zonal and meridional wind fields, and wind power production. These exercises are carried out over two distinct regions of the world, north of Spain and Northeastern North America, with their own very different domain scales and geographical characteristics. In all the cases, the analysis is conducted over monthly timescales and during wintertime. For these analyses only one SD technique will be used, which is mainly based on Empirical Orthogonal Functions (EOF; Lorenz 1956) and Canonical Correlation Analysis (CCA; Karl 1990). In exchange for that, however, a complete in-depth evaluation of the methodology will be presented: along with the description of the coupled local and large-scale main circulation modes, we also evaluate the predictability of the wind/power production via this methodology and a complete sensitivity evaluation of the predictability to changes on the configuration parameters. Additionally, we will extend the SD outside the calibration period and evaluate this reconstruction exercise and its associated sensitivity.

The description of the observational database and large scale variables is presented in Section 3. A thorough description of the methodology is given in Section 4. The description of the main large scale and local coupled circulation modes are described in Section 5. The predictability of the method and sensitivity to parameter configurations is presented in Section 6. The summary and conclusions are given in Section 7.

2. Data

2.1 Observational datasets, predictands.

For the present work three observational datasets are used, located over two distinct regions. The first two databases are situated along the orographically complex Comunidad Foral de Navarra (CFN, see Fig. 1a), situated over the north-east of Spain and with the Bay of Biscay to the north and the Pyrenees to the north-east as natural barriers. The surface wind speed and direction database used for the current exercise consists on 29 meteorological stations that span from January 1992 to September 2005 (colored circles; GB12), a reduced subset from the original set of 41 observational sites (Jiménez et al. 2010a). The original observations are taken at 10-minutes for wind speed and wind direction. The anemometers are located either at 10 meters (white circles) or 2 meters (orange). From these initial measurements the zonal (u) and meridional (v) components of the wind are computed and then monthly averaged for the specific purposes of the study herein. Due to the short existence of some of the sites, not all of them could be included during the statistical downscaling exercise. The wind power production database consists on the three wind farms, whose power outputs were obtained as the spatial averages of every wind turbine at each farm (pink squares; García-Bustamante et al. 2008). These farms have observational wind speeds recorded at hub height (between 30 and 40 m). The wind and wind power records of the farms span from June 1999 to May 2003. For CFN, the analyzed period is an extended winter season going from September to March (7 months), corresponding to the windiest months in this region (GB12, GB13).

The mean flow in the CFN (Fig. 1b) is directed from NW to SE, channeled from the northern valleys along the Ebro Valley (Jiménez et al. 2008a). This is a characteristic cold and dry wind pattern in the CFN region known as Cierzo. The flow in the opposite direction (SE to NW) is known as Bochorno, resulting from the advection of moist and warmer air from the Mediterranean Sea. The wind farms located at the center of the region show a NE-SW flow in agreement with the mean direction at the surrounding stations while the other, more northerly located, presents a more SE-NW direction. This is explained by the fact that the northern area in the CFN is exposed to a different large scale circulation regime than the central and southern sections of the region (Jiménez et al. 2008b).

The spatially averaged, or regional, zonal and meridional wind components (Fig. 1b) present opposite sign throughout the whole period with a correlation value of -0.77. This suggests a conservation of the surface momentum between both wind components at these timescales (Jiménez et al. 2008b). The regional series present a substantial intra and interannual, of great interest in the context of wind energy assessment. Fig. 1c shows the standardized monthly wind speed (blue) taken at hub height, and wind power production (magenta) at El Perdón wind farm where it is evidenced a linear connection between both variables (García-Bustamante et al. 2009) for the whole observational period with a correlation value of up to 0.94.

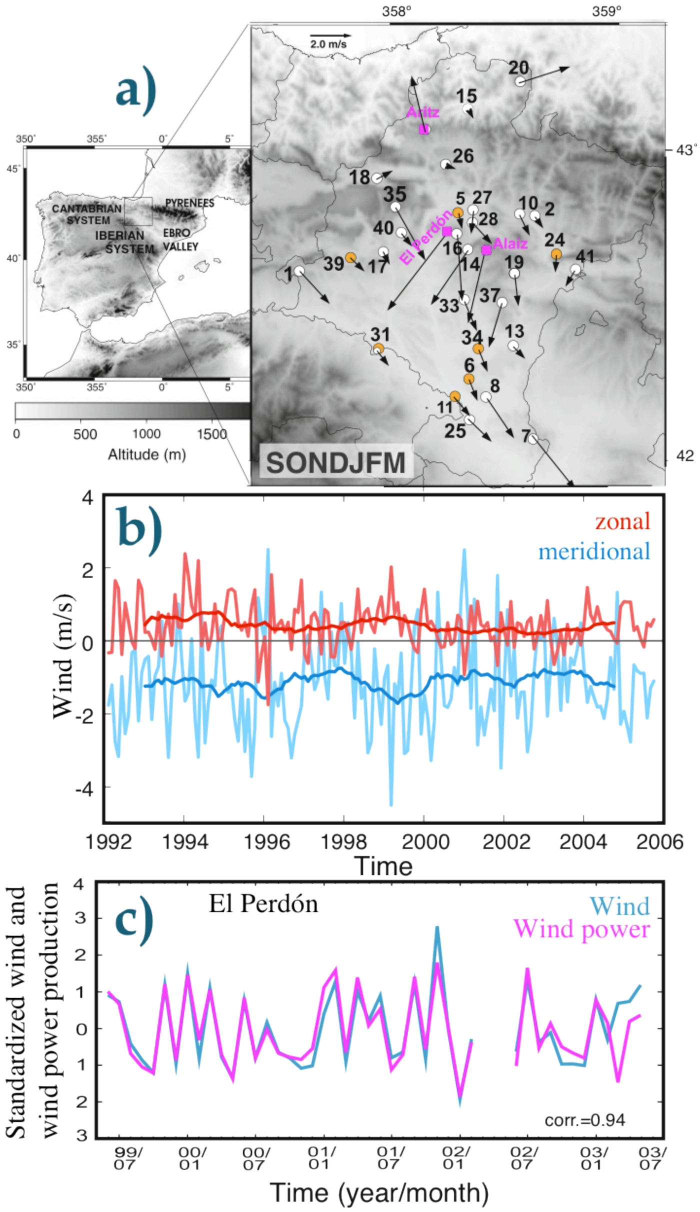

Figure 1: (a) The area under study: the left panel shows the Iberian Peninsula and the main geographical features surrounding the CFN. The right panel amplifies the region of the CFN and its orography (shading). Circles stand for the location of the wind stations and squares correspond to the wind farm locations at Alaiz, Aritz and El Perdón. Orange circles represent those stations with anemometers at 2 m while the rest are located at 10 m height (see Jiménez et al. 2010a, Table 2.4, for sites description).The arrows indicate the mean wind speed (length) and direction (angle) computed for the SONDJFM months. (b) Temporal evolution of the SONDJFM monthly regional zonal (red) and meridional (blue) wind components. The thicker lines correspond to the 3 years moving averages. (c) Standardized wind speed (blue) and wind power production (magenta) for the same months at El Perdón. (a,c) are modified figures from GB13 and (b) from GB12.

The third observational database consists on a surface wind speed and direction database of hourly, 3-hourly and synoptic resolution comprised by 525 sites located across the region of North Eastern North America (NENA, Fig. 2a; LE17a): 434 of these sites span over Atlantic Canada, Ontario and Quebec; 52 are located across 5 northeastern estates of USA; an additional 40 buoys are located over the open sea nearby the eastern Canadian coast and the Canadian Great Lakes. The database spans from January 1953 to January 2010, although it is unevenly distributed over time and space. To avoid problems related with data representativeness, for this SD exercise a subset of 95 sites (Fig. 2a in magenta) has been selected. The calibration period has been also limited to 1980-2010 (LE17c). As with the previous case, the hourly wind speed and direction data have been converted to zonal and meridional wind components and then averaged to monthly resolution. As with the former experiments, this study will be centered during wintertime, during the months from November to March (5 months), which is also the season of highest winds and associated wind variability.

Due to the privileged geographical location of this region, the large scale dynamics foster the transit of most summertime and wintertime cyclonic events that originate all over North America. From Western Canada (Mackenzie and Alberta Lows), through central (Colorado Lows) and southeaster areas (Coastal and Hatteras lows) of the contiguous United States, and from closer origins (Hudson and Great Lakes lows) as well (Conrad 2009). These cyclones can be either from tropical or extratropical origin (Plante et al. 2014), and are more frequent and intense during wintertime (Wang et al. 2006). In combination with topographical effects and channelings such as the *suetes* (Cape Breton) or *wreckhouse winds* (Newfoundland), these cyclonic events make this region prone to extreme events especially during the winter months (Richards and Abuamer 2007). The mean winter winds are predominantly westerlies (Fig. 2a), reaching their maximum values along the coast of Labrador, Newfoundland and open sea. Although this is a relatively flat area, with mountains no higher than 1000m, there are some places of pronunciated heights with a very high associated winds (e.g. Mount Washington, at ~2000m with 16 m s^{-1} winds). Fig. 2b shows the regional zonal and meridional averages for the whole observational period, calculated from the subset used for the SD exercise. The wind components show a pronounced intra and especially interannual variability, mostly concentrated through periods 1953-1975, 1980-1990 and 1995-2005. A prevalence of westerly wind can be observed throughout the whole observational period. For meridional winds, however, a progressive stilling of the northerly flows is observed, a tendency that is reduced for the last decades.

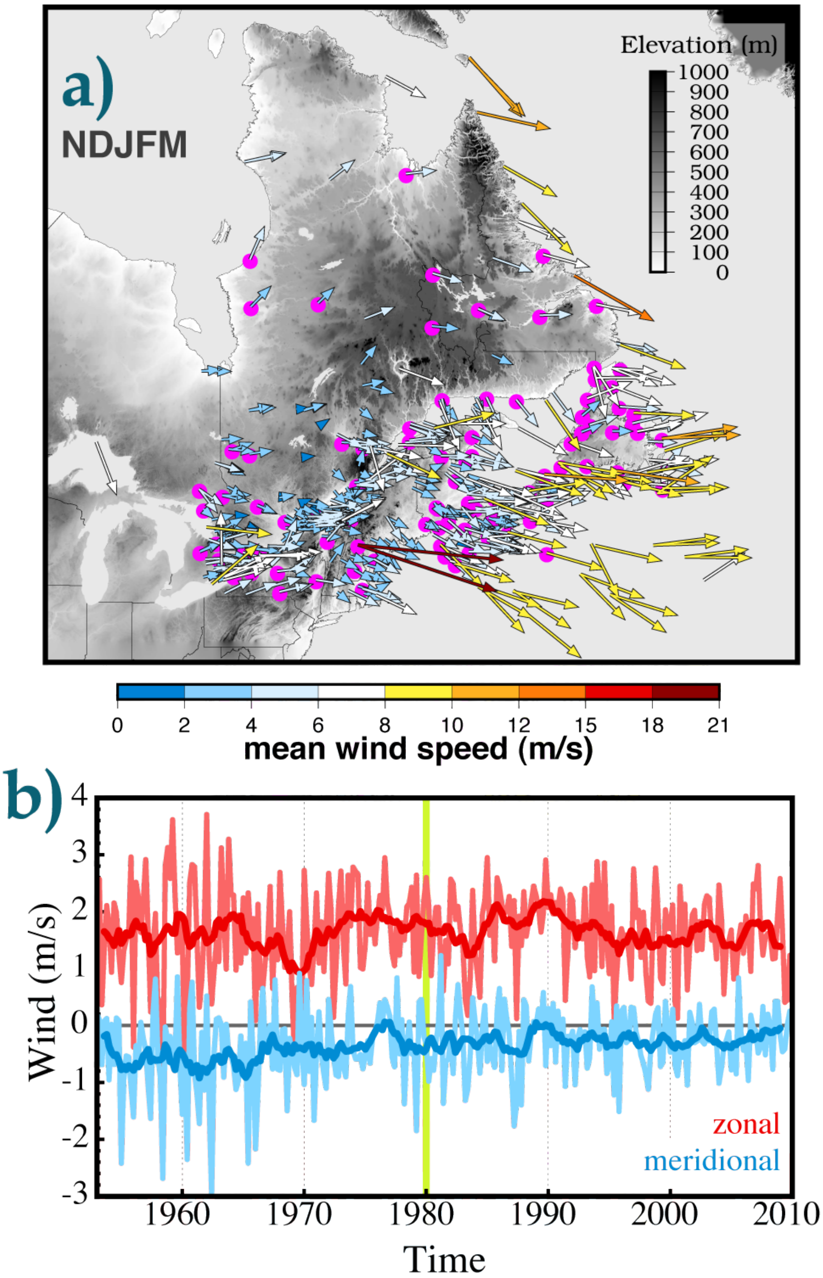

Figure 2: (a) Mean wind speed of the original database computed with a extended winter season consisting on the NDJFM months. The mean wind speed is given by the color code and length of the arrow. The topography is given in grayscale. The circles in magenta correspond to the 95 sites out of the original 525 used in this exercise. (b) Regional zonal and meridional averages obtained averaging the magenta sites. The thicker lines correspond to a 5-month centred running mean. The beginning of the calibration period (1980) is marked with a vertical green line. Modified from LE17c.

Accurate observational records are necessary to any realistic climotological assessment. Unfortunatelly, despite of the care put during the instrumental siting, recording, and data storage and management, a vast arrange of possible errors may emerge, errors that can have a potentially drastic effect on the measured variables. Although some preventive masures can be taken to avoid many of the problems, it is nevertheless paramount to develop an ancirally series of tests or checks that are able to detect the large variety of issues that might still be inadvertently present. These methods are known as Quality Control procedures, and target issues that can be mainly grouped into two categories:

- *Data Management Issues* (LE17a). They target issues related with data transcription and collection or with errors occurred during data manipulation, Additionally they handle with standardization of practices that can vary across institutions like measurement, units or reference times, which can be issues of importance for datasets built using data from various source institutions

- *Measurement Errors* (LE17b). These tests address the temporal or spatial consistency in the data. They are designed to deal with errors often produced at the moment of sampling, due to instrumental malfunction, calibration or exposure problems. They target errors that are generally of local nature and less likely to depend on procedures established by the data source institution.

The aforementioned datasets have been subjected to an exhaustive Quality Control procedure (Jiménez et al. 2010a, LE17a,b). Since an in-depth description of the methods that have been devised and applied is beyond the scope of this work, we will only describe some illustrative cases for each family of errors. found at each datasets. We refer the interested readers to the related papers for further information.

A particularly interesting case among *Data Management Issues* are the periods of data that might have been accidentally duplicated during data retrieval, transmission, and archival. These errors can take place within the same series (intra-site duplications, see Fig. 3a) or to accidentally transfer data from one series to another (inter-site duplications, Fig. 3b). These duplication can entail brief 24-h periods of data or involve up to several years of mismanaged records (LE17a). It is not trivial to sort out between naturally occurring duplications produced by a combination of atmospheric similarities, pure chance and data paucity, and those genuinely erroneous ones. There are, however, some features shared by the erroneous intra- and inter- duplications. The intra-duplications tend to share common date features as they are primarily caused by data resubmission under two different time stamps. Similarly, the inter-site duplications are often caused by one misfiled report under two different stations and are expected to occur simultaneously in time

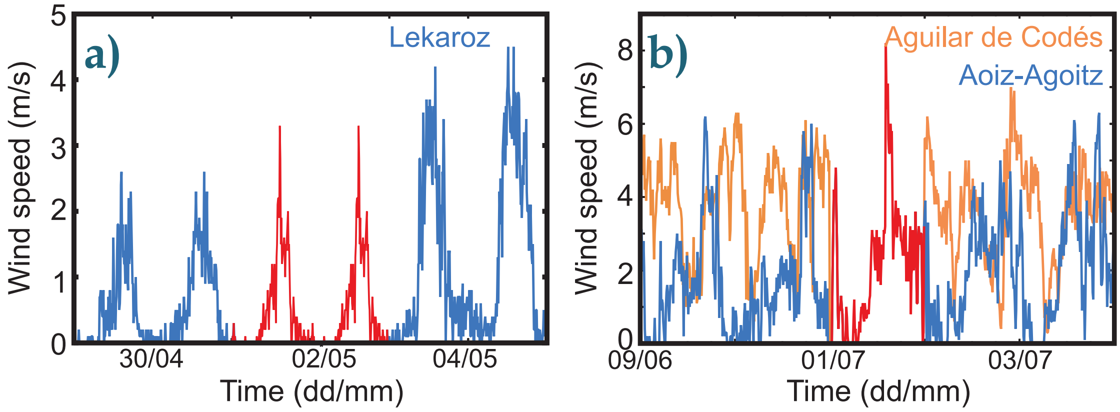

Figure 3: (a) Example of an intra-site duplication error for wind speed records at Lekaroz. The two consecutive 24-h periods in red have coincident observations. (b) Example of an inter-site duplication error for wind speed records between the sites of Aguilar de Codés (in orange) and Aoiz (blue). A whole 24-h period (in red) is simultaneously shared by both sites. Both correspond to sites located at CFN and are a modified version from Jiménez et al. (2010a).

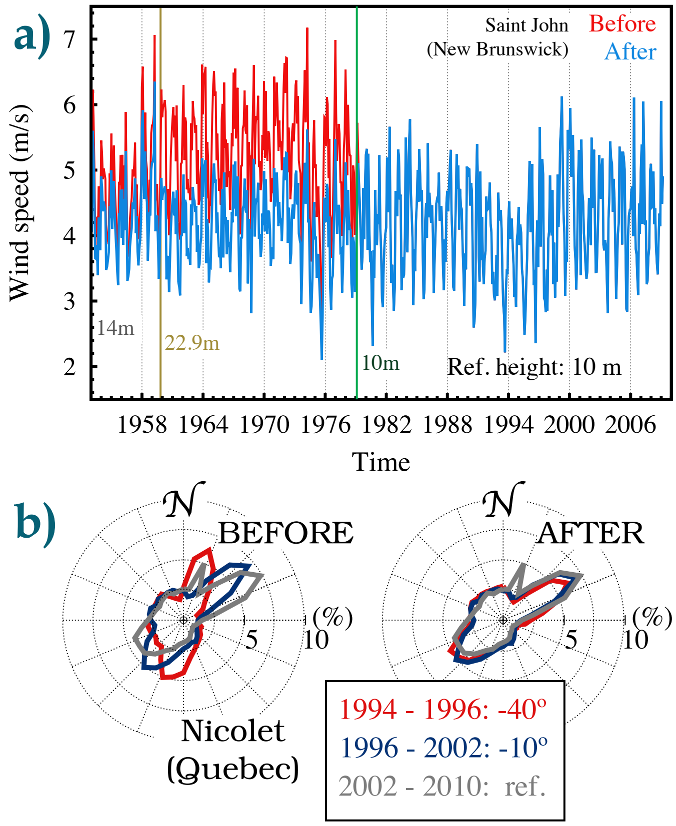

Figure 4: (b) Monthly wind speed of a site located at Saint John (New Brunswick) before (red) and after (blue) standardization to its reference height (10 m). Vertical color bars indicate documented anemometer heights. (b) Wind rose showing wind direction bias before (left) and after (right) correction, corresponding to a site located at Nicolet (Quebec) with 3 changes, each indicated by a line of different color.

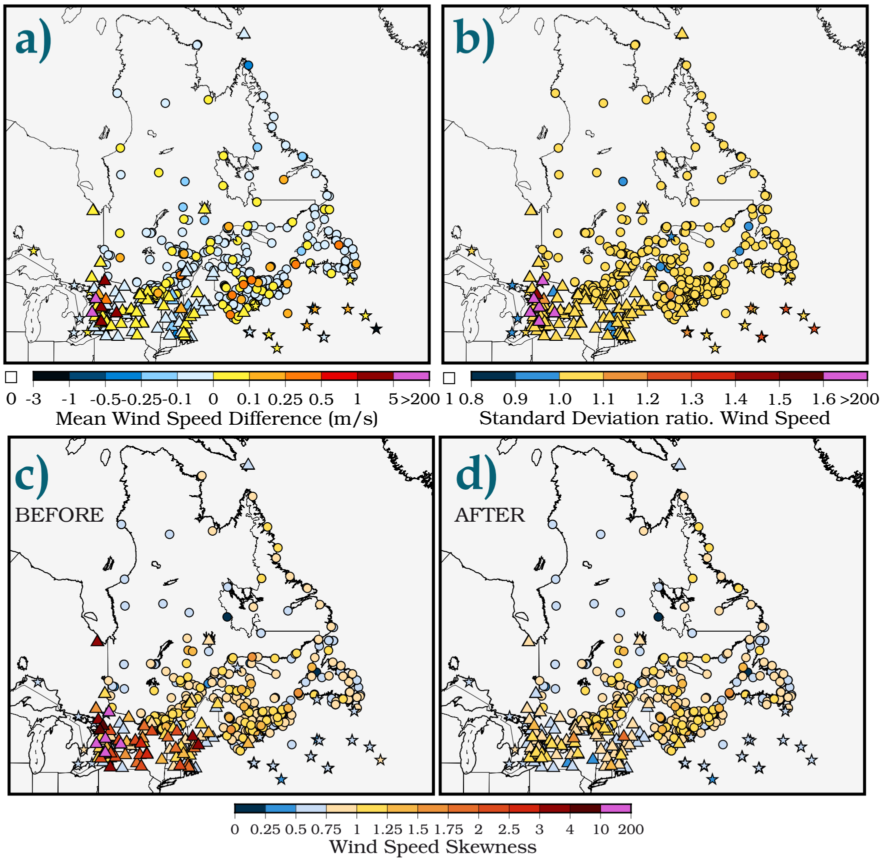

Figure 5: Spatial distribution of (a) differences in mean and (b) standard deviation ratios before and after the QC. Spatial distribution of skewness (c) before and (d) after the QC. Modified from LE17b.

Regarding *Measurement Errors,* those associated with systematic biases might have a great impact on the observational time series, as they can span from weeks to several decades. These errors are related to a great variety of factors such as changes in the measuring devices, different averaging methods, changes in anemometer heights or changes in exposure or site relocation. They can be corrected making extensive use of collected metadata, as in the case of wind speed biases produced by anemometer height changes (Fig. 4a) or detected and corrected thanks to algorithms suited to identify sudden changes in the orientation of wind direction vanes (Fig. 4b). For more information about the detection and correction of these and other problems please refer to LE17a,b.

The detection and eventual removal/correction of these errors affected with varying degrees on the different sites depending on their original quality, but they overall had a clear impact on the general statistics of the datasets. Fig. 5 shows the effects on the QC process on the surface wind database located in NENA (LE17b). This evaluation has been made for mean wind speed (a), standard deviation (b) and skewness (c,d), obtained from the calculation of the 1st to 3rd order moments (von Storch and Zwiers 1999). The majority of the sites remain unaffected or with small changes. In others, however, the changes are evident, with up to 200 m/s differences in the mean (a) and standard deviation ratios (b). A considerable reduction on the skewness, a measure of the asymmetry of the distribution, can also be appreciated (c,d) for several sites.

2.2 Large scale variables, predictors.

For the Statistical Downscaling exercises presented in this work, six variables have been used as predictor fields (see Table 1): sea level pressure (SLP), 850 and 500 hPa geopotential height (Z_{850} and Z_{500}), 10 m height or surface level meridional and zonal wind components, and 500-850 hPa thickness (ΔZ). These variables have been obtained either as products of several reanalysis or from observational gridded databases. The gridded observational or reconstructed SLP datasets were compiled from a variety of data sources (see references). The reanalysis models, on their hand, use different assimilated data, assimilation schemes and model characteristics. Additionally each predictor source has each own horizontal resolution, providing, in principle, different information on the large scale dynamics. All in all, this ensemble of sources allows for an extensive mapping of all the possible sensitivities associated to the data source.

Table 1: Characteristics of the variables employed as predictors: variables considered for the exercises, data source, level and units. Modified from LE17c.

Variable	Source	Level	Units
SLP	Gridded Data/Reanalysis	Surface	hPa
U,V	Reanalysis	10m*	m s^{-1}
Z	Reanalysis	850, 500 hPa	m
ΔZ	Reanalysis	$\Delta Z = Z_{850} - Z_{500}$	m
**for CFSR and 20CRv2c reanalysis models, surface level*			

For the SD experiments of surface wind fields (GB12) and wind power (GB13) conducted in CFN, the variables were obtained from Era-40 reanalysis (Uppala et al. 2005) at a 2.5ºx2.5º resolution (see summary of predictor sources in Table 2). A SLP reconstructed gridded database from Luterbacher et al. (2002) was also used for sensitivity experiments beyond the observational period of the predictand, at 5ºx5º horizontal resolution.

Table 2: Reanalysis models and observational datasets used in this analysis: name, institution, time coverage, used spatial resolution and useful references. Modified from LE17c.

Name	Institution	Time Range	Grid Resolution	Reference
		Reanalysis Models		
Era-40	ECMWF	1958-2001	0.75x0.75/2.5x2.5	Uppala et al. (2005)
Era-Interim	ECMWF	1979-	0.75x0.75	Dee et al. (2001)
Era-20C	EMCWF	1900-2010	0.75x0.75	Poli et al. (2016)
JRA25	JMA	1979-	1.25x1.25	Onogi et al. (2007)
JRA55	JMA	1958-2012	1.25x1.25	Ebita et al (2011)
MERRA	NASA	1979-	0.5x2/3	Rienecker et al. (2011)
NCAR-R1	NCEP-NCAR	1949-	2.5x2.5	Kistler et al. (2001)
DOE-R2	NCEP-DOE	1979-	2.5x2.5	Kanamitsu et al. (2002)
CFSR	NCEP	1979-	0.5x0.5	Saha et al. (2010)
20CRv2c	NOAA et al.	1851-2015	2x2	Compo et al. (2011)
		SLP Gridded Datasets		
Luterbacher		1650-1999	5x5	Luterbacher et al. (2002)

For the SD experiment in NENA (LE17c) the variables were obtained from 10 different reanalysis models (Table 2): 3 from the ECMWF, Era-40, Era-Interim (Dee et al. 2011) and Era-20C (Poli et al. 2016) all of them at 0.75ºx0.75º; 2 from the JMA, JRA25 (Onogi et al. 2017) and JRA55 (Ebita et al. 2011) at 1.25ºx1.25º; MERRA (Rienecker et al. 2011) from NASA, at 0.5ºx2/3º; from NCEP 3 more, NCAR-R1 (Kistler et al. 2001), DOE-R2 (Kanamitsu et al. 2002), both at 2.5ºx2.5º and CFSR (Saha et al. 2010) at 0.5ºx0.5º; and from NOAA, 20CRv2c (Compo et al. 2011), at 2ºx2º.

3. Statistical Downscaling Methodology.

The statistical method used in this work is mainly based on EOF and CCA techniques. Before any further step, however, the annual cycle of the original predictor and predictand fields is removed by subtracting the monthly climatological mean, thus obtaining anomaly fields. These time series are also detrended applying a linear least square fit in order to ensure the long-term stationarity and avoid spurious relationships between data points (Xoplaki et al., 2003b). In order to account for the latitudinal distortions, the detrended anomalies from the large scale fields are weighted at each grid point by multiplying by the square root of the latitude to consider the decreasing size of grid boxes with latitude (North et al., 1982b). Finally, the time series are standardized (dividing by their standard deviation).

These detrended and standardized fields are then independently projected onto their S-mode EOF space (González-Rouco et al. 2000). This step allows to only retaining the modes with maximum explained variance, thus reducing noise, and at the same time drastically diminishing the number of degrees of freedom (Lorentz 1956). These predictor and predictand EOF fields are the information ingested by the CCA technique. This multivariate statistical technique isolates linear associations between sets of predictor and predictand EOFs that are optimally correlated (Glahn, 1968). The linear associations are called canonical component vector or canonical coordinates and they are normalised to unit variance. The number of obtained CCA modes cannot, by definition, exceed the number of introduced EOF modes. The original variables can be expressed as a linear combination of its canonical coordinates and its canonical correlation patterns. The canonical coordinates or scores describe the amplitude and sign of the corresponding patterns at each time instant.

The estimation of the predictand field is obtained with a regression model that presents the predictand variable as a linear combination of the canonical coordinates and patterns, and the predictor fields (Zorita and von Storch 1999).

The skill of the statistical model is verified using a crossvalidation approach. This allows to reduce a possible overfitting of data by the model (Barnett and Preisendorfer, 1987). The crossvalidation is a resampling technique in which a small number of data is dismissed and the model is trained with the retained data subset. The removed values are then estimated with the calibrated model. This procedure is repeated recursively by sampling subsets of the same length along the entire observational record in order to obtain a full set of independent estimates (Michaelsen, 1987).

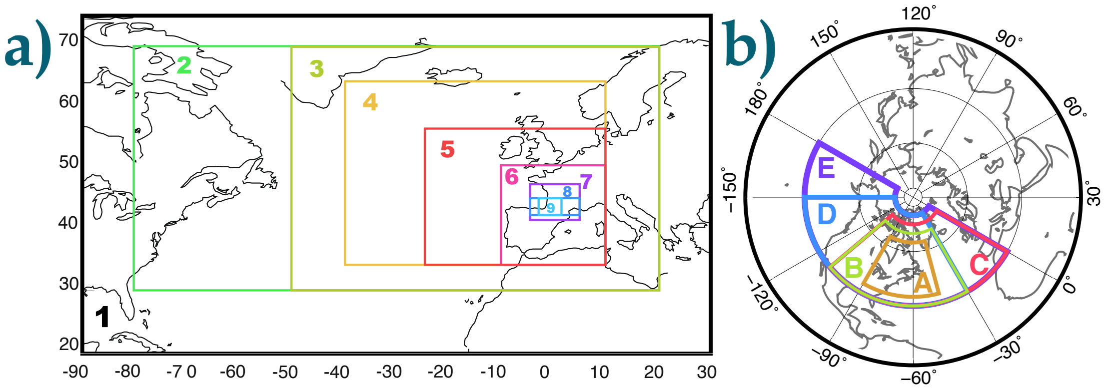

Figure 6: Different large scale domains (boxes) used as predictor windows in GB11 and GB12 (left), and LE17c (right).

Inherent to the downscaling methodologies are the associated uncertainties that are propagated through the established relationships between the large scale and local information. A fundamental level of uncertainties, that will not be addressed in this work, is related to the choice of one SD technique over another. Other uncertainties are, on the other hand, related to the particular choice of some parameters in the statistical model. We have systematically explored the effect that the change on five different parameters have in this statistical model (they are summarized in Table 3):

- The variable chosen as predictor. This parameter evaluates the predictability of the predictand to the dynamics of different surface or upper variables. The predictor can range from a variable used alone .(e.g. SLP, see Table 1) to a combination of multiple variables (e.g. SLP+UV+Z_{850}). GB12,13 use up to 25 possible combinations, and LE17c up to 18 combinations.

- The size of the spatial domain of the predictor fields. As illustrated in Fig. 6, 9 windows have been used for CFN and 5 for NENA. The size of the windows has been selected to go from the smallest possible ones that fit the observational domain, to those that comprehend the large-scale phenomena that could affect the region: Atlantic and Mediterranean for GB12,13 and Atlantic and Pacific Basins for LE17c.

- The number of retained EOF and CCA modes. As it has been introduced before, the CCA method computes linear combinations of predictor and predictand EOF patterns, which means that the method is sensitive to the number of introduced EOF modes: too few EOF could ignore modes with a significant signal, while too many could overfit the model to our particular dataset. Additionally, as the estimations are calculated with the CCA patterns, these are also sensitive to the number of retained CCA modes. There are many methods to select the minimum or maximum number of significant EOF modes, being an easy but effective one the visual Scree test (Cattell 1966). GB12,13 explore a range of $EOF_{predictor}/EOF_{predictand}/CCA$ modes that goes from 2/2/2-6/6/6 ecompassing 31 possible choices. LE17c on the other hand, range from 4/4/2-7/7/7 with 62 possibilites.

- The size of the crossvalidation subset. This parameter evaluates the robustness of the statistical method to changes in the calibrated/estimated sample size. GB12,13 selected between 9 subsets, 1 month and 28 months (7 years), and LE17c between 5 subsets, 1, 2, 5 (1 year), 10 or 20 (4 years) months.

- Source of the predictor variables. This parameter analyzes the sensitivity of the estimations to the Source of the predictor variables. LE17c compare the outcomes between 10 different reanalyses (see Table 2).

Table 3: Parameters selected for the sensitivity analysis. The second and third column correspond to the number of combinations used in each experiment.

Parameter	GB12,13	LE17c
Predictor	25	18
Window	9	5
EOF/CCA	31	62
Crossvalidation	9	5
Predictor Source		10

Prior to this uncertainty analysis, a *reference configuration* has been chosen, which is presented in the next section. Despite not being the optimum configuration, it nevertheless offers a valuable information about the leading coupled predictor and predictand dynamics in the three different study cases, and the predictive skill of the method. It also offers a starting point for the sensitivity analysis presented in Section 5.

4. Downscaling Experiment: reference case

This section analyses the results of the reference configurations at each experiment. First, we describe the first coupled mode at each case, then we proceed to look at the estimated regional predictand obtained from this reference configuration. For a description of the rest of coupled modes in each configuration please refer to GB12,13 and LE17c, respectively.

4.1 Large and local scale coupled dynamics

The chosen reference configuration in CFN for the analysis of the surface wind field (GB12) combines Z_{850} and ΔZ as predictors; it uses the window 4 (see Figure 5a) for the description of the large-scale dynamics; it retains 4 (4) predictor (predictand) EOF modes, accounting for the 81.5% (90.2%) of explained variance; 2 CCA modes are retained for the estimation of the wind. The sampling subset of the crossvalidation is set to 1 month.

The leading pair of canonical patterns (CCA1) is presented in Figs. 7a,b for the predictor and predictand variables. Their associated Canonical Series is presented in Fig. 7c. This first mode accounts for 18% (22%) of the total variance of the Z_{850} (ΔZ) predictor field and around 36% of the total predictand variance. The large scale pattern (Fig. 7a, contours for ΔZ, shaded for Z_{850}) depicts a dipole structure with positive anomalies over the North Atlantic area that reaches the west side of the Iberian Peninsula. Negative anomalies are located northeast of the British Isles, centred over the Scandinavian Peninsula. The coupled CCA1 pattern for the predictand (regional wind) presents NW-SE wind oriented anomalies which can be related to the well known regional 'Cierzo'. This pattern corresponds to the positive phase of the mode, although it should be noticed that the reverse sign of the patterns is also possible, as it is determined by the sign of eigenvalues of the cross-correlation matrix (von Storch and Zwiers, 1999). In contrast, the negative phase of this mode presents a southeasterly flow that advects the relatively warm and moist air from the Mediterranean that can be regarded as the typical 'Bochorno'. As it can be observed, the two patterns are physically meaningful since the large scale mode induces a pressure gradient that favours a NW-SE (SE-NW) direction for the geostrophic wind. The orientation of the Ebro Valley, aligned with this NW-SE direction, also contributes with a strong channeling effect at the surface. The local wind pattern arises as the result of the large scale atmospheric structure modulated by the regional orographic configuration. In agreement with a more meridional large scale circulation, in this canonical mode the largest amount of explained variance (46%) corresponds to the meridional component of the wind while in the case of the zonal one, accounting for a smaller portion (25%) of variance for the zonal component. This large scale pattern is broadly responsible for predominant meridional circulations not only in the region under study, but also in wider areas over the European continent. It has been found that this same large scale regime causes a partial blocking of the westerlies, associated with the Mistral conditions (Buzzi et al., 2003; Burlando, 2009).

The corresponding Canonical Series (Fig. 7c) present a correlation of 0.87 and exhibit considerable intra and inter annual variability with extreme values at the end of 1995 and 1999. It shows a higher variability till 2000 and slightly lower thereafter. This change in the regional wind variance is noticeable in both the predictand and predictor series, suggesting to changes in the large scale circulation variability. The canonical series of the local wind components presents a correlation of 0.92 and -0.83 with the regional time series of the observed zonal and meridional wind components, respectively. This first canonical mode describes the most important changes in regional monthly wind variability for the period studied.

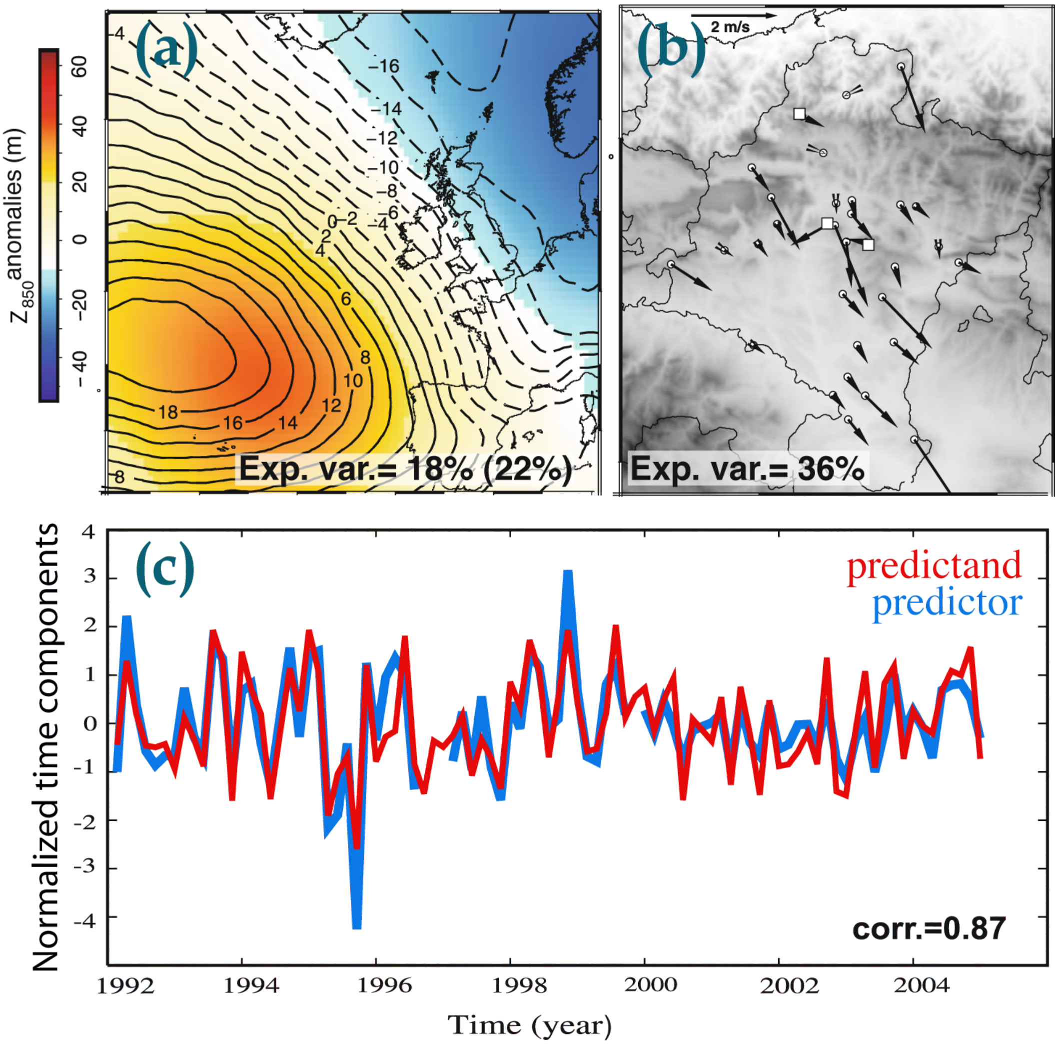

Figure 7: Canonical patterns and series of the first CCA mode (CCA1). (a) Predictor patterns, Z_{850} m (shaded) and ΔZ m (contours); the explained variance of the Z_{859} and ΔZ field is also indicated; (b) predictand (local wind in the CFN) pattern together with the variance accounted for by this mode; wind farms are represented with white squares. c) canonical series for predictor (blue) and predictand (red), their associated correlation value is also indicated. Modified from GB12.

The chosen reference configuration in CFN for the analysis of the wind power production (GB13) also combines Z_{850} and ΔZ as predictors and uses the window 4 (see Figure 5a) for the description of the large-scale dynamics; it retains 4 (2) predictor (predictand) EOF modes, accounting for the 81.5% (97%) of explained variance, and 2 CCA modes. The sampling subset of the crossvalidation is also set to 1 month.

The first pair of canonical patterns (CCA1) are shown in Figs. 8a,b. The variance explained by the first canonical mode is 25 % in the case of the large scale predictors and 63 % for the wind power. The large scale dynamics consist of a negative anomaly center located westward of the British Isles, a configuration connected with anomalous southwesterly flows in the region (Jiménez et al. 2009). The corresponding local pattern shows a dipole with positive anomalies of wind power production to the north and negative ones to the center of the region. This pattern is coherent to that obtained with wind velocities (not shown) as predictand. It shows windy conditions equivalent to positive wind power anomalies in the northern areas and a decelerated flow, equivalent to negative wind power anomalies, to the center of the region. It is worth noting that the large scale pattern resembles the second pattern obtained by GB12 for the zonal and meridional wind field (not shown). This pattern was found to be responsible for anomalous eastward geostrophic flow over the region. The corresponding canonical pattern of the wind therein showed a more zonal orientation of the circulation at the windiest locations over northern and central parts of the CFN, thus, at the wind farm locations. Since the farms are located at higher elevations, are less affected by local effects and exhibit a large influence of quasi-geostrophic circulations. Hence, the variability of both the wind and the wind power at these specific locations is dominated by the same large scale patterns.

The corresponding Canonical time Series present a strong correlation (0.89, Fig. 8c). The correlations between the canonical series and monthly wind power observations reaches ~0.7 at the sites, suggesting that this mode is responsible for the wind power monthly production. The correlations show opposing signs of Aritz respect to El Perdón and Alaiz (not shown), which is consistent with the canonical pattern (Fig. 8b).

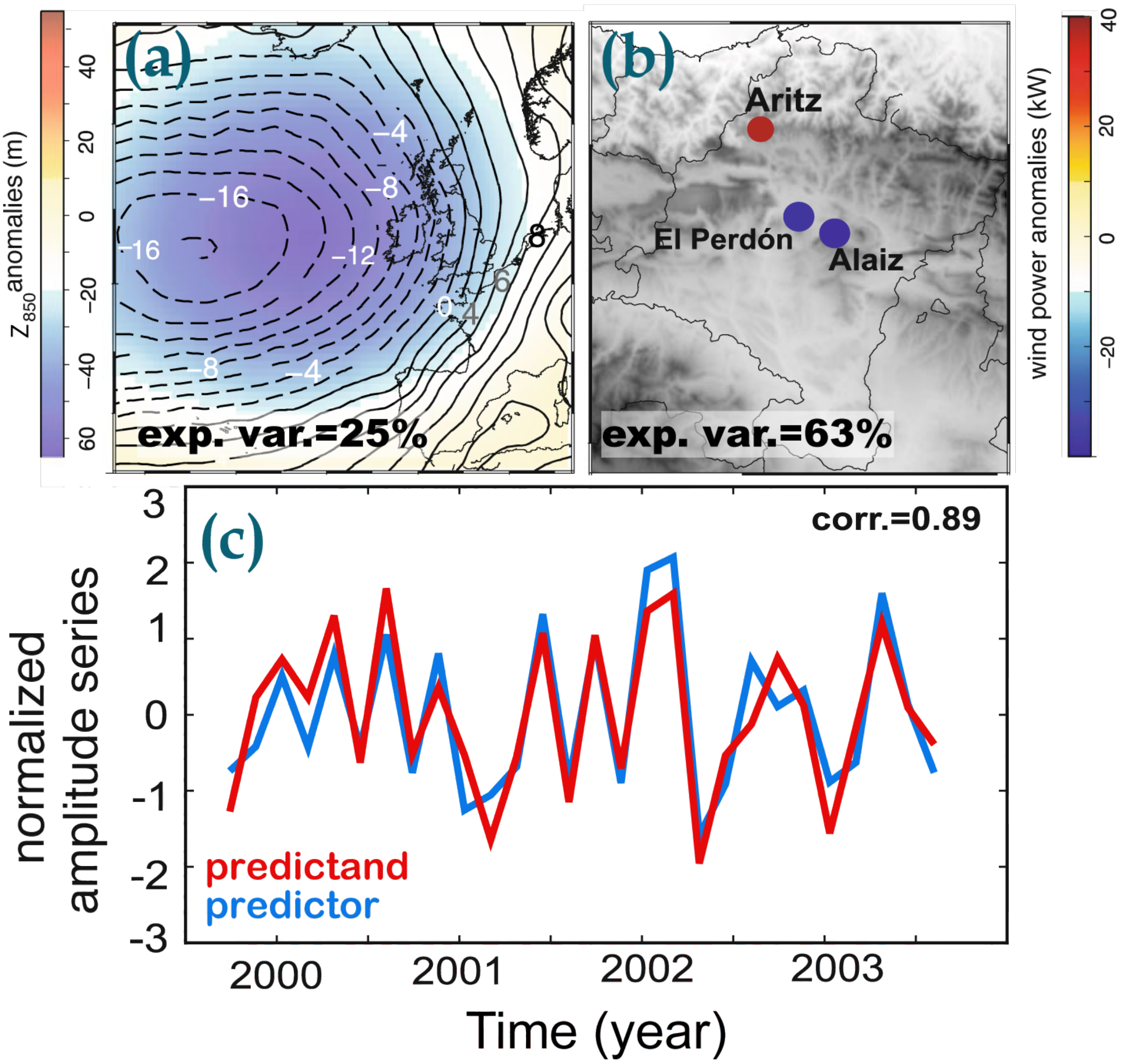

Figure 8: First canonical pair of patterns (CCA1) of (a) the predictor fields (Z_{850} m, shaded, and ΔZ m, in contours); (b) regional wind power pattern field or predictand ;and (c) associated Canonical Series. Modified from GB13.

The reference configuration for NENA (LE17c) uses SLP as predictor variable and window B (Fig. 6b) to the describe the large-scale dynamics. It retains 4 (4) predictor (predictand) EOF modes, accounting for 78.1% (70.9%) of the explained variance. It retains 3 CCA modes for the estimation of the local wind, as the 4th CCA pair bears a very low correlation value (and associated physical meaning). The sampling subset of the crossvalidation is set to 1 month. Although the same configuration has been run for all the reanalysis models (see Table 2), here we will only present the outcomes from Era-Interim.

The first Canonical Pattern is shown in Fig. 9a. This first mode accounts for 30.4% (26.7%) of the total variance for the predictor (predictand). The Predictor field, SLP (isolines), presents a zonal dipole structure with positive anomalies centred over the North Atlantic and negative anomalies centred between Greenland and Iceland, below the Denmark Strait. The associated predictand CCA pattern of surface zonal and meridional wind speed anomalies (arrows), shows a clockwise spinning structure, SW to NE, with higher wind speeds at the East Coast, specially along the coast of Labrador, Newfoundland and Nova Scotia, and a weaker response around Great Lakes and Ontario. The two patterns are physically linked, as the large scale mode induces a pressure gradient that favours SW-NE flows. This flow also tends to fill the negative anomaly pressure system of the north, as expected. The orographic anomaly of Mt. Washington (see Section 2.1) produces the largest wind variability on the whole region. This is offers a practical example of the convenience of using correlations over covariances for EOF and CCA exercises to reduce spurious weightings due to variability disparities within the database. The large scale pattern shows an equivalent barotropic structure in ΔZ (shadings), suggesting a coupling between the large and local scale that extends to the upper air levels.

The associated Canonical Series (CS1, Fig. 9b) show a correlation of 0.85. The first decade shows the largest variability and extreme values. This period is followed by a 5 year interval with the lowest variability, which is regained afterwards, although not as large as before. The 1990-1995 interval shows a persistent tendency for positive anomalies. These changes also manifest in the predictor series, suggesting to variability changes in the large scale dynamics. The meridional flows are the most strongly modulated by this pattern as the meridional component (Fig. 2b) shows significant correlations values of ~0.8 with the canonical series at a 0.05 level ($p<0.05$). The correlation value is lower (~0.5) for the regional zonal component.

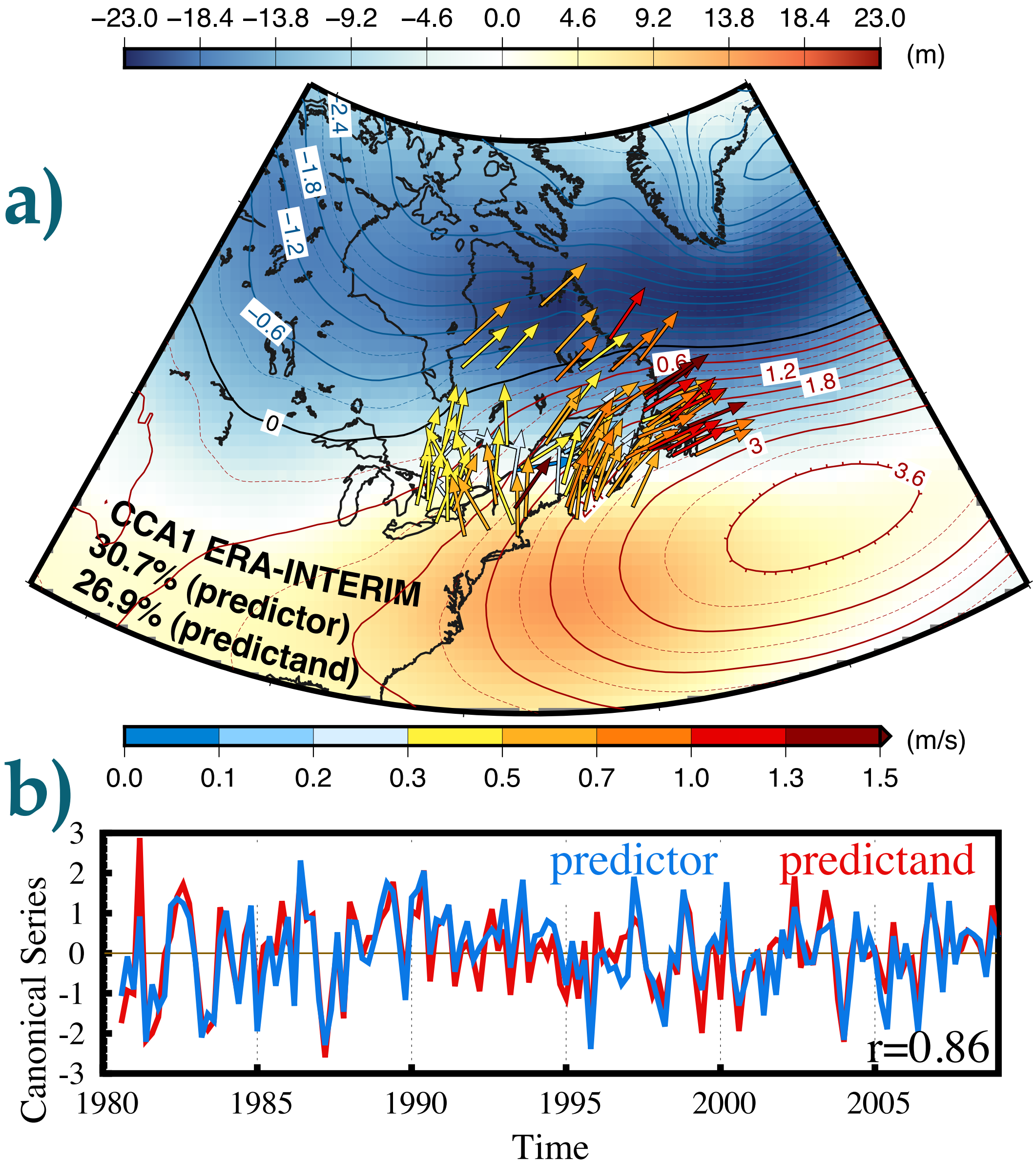

Figure 9: (a) First CCA pattern (CCA1) using SLP as predictor and ERA-Interim as source. The isolines correspond to SLP (predictor, hPa), the shadings to ∆Z (regressed pattern, m) and the vector filed to the predictand (m s^{-1}). The wind speeds are given with the color scheme. (b) First CS of the predictor (blue) and predictand (red). The correlation between both is also given.

4.2 Estimations derived from the reference case

The estimation of the regional surface wind components in CFN (Figure 10; GB12) show a good accordance with the observations, with correlation values of 0.79 and 0.80 for the zonal and meridional component respectively, all significant (p<0.05).

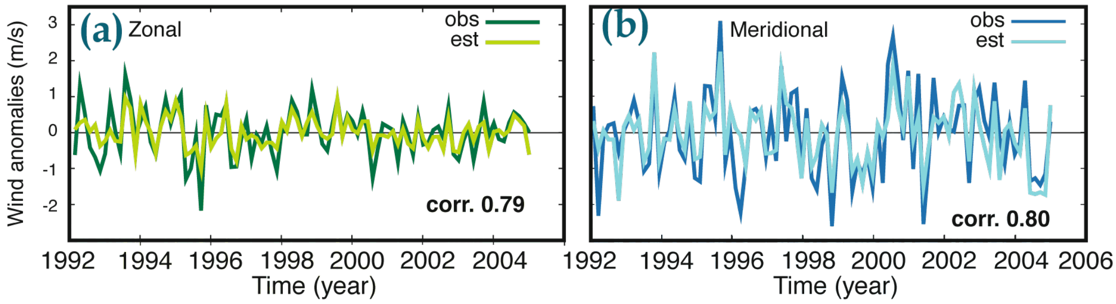

Figure 10: Regional estimated and observed monthly (a) zonal and (b) meridional wind component anomaly series. Their correlation values are also indicated. Modified from GB11.

In contrast with the high correlation values of the wind components the wind power production (Figure 11; GB13) is worse predicted, with correlation values for Aritz with 0.7, El Perdón with 0.54 and Alaiz with 0.33. These values are in accordance with the lower correlation values with wind velocity estimations (GB12, not shown) with which they keep a linear relationship in monthly scales (García-Bustamante et al. 2009 and GB13). The low correlation values with wind velocity can be explained on the basis of the non-linear transformations applied to obtain it from the wind components together with the potential accumulation of the errors from each component.

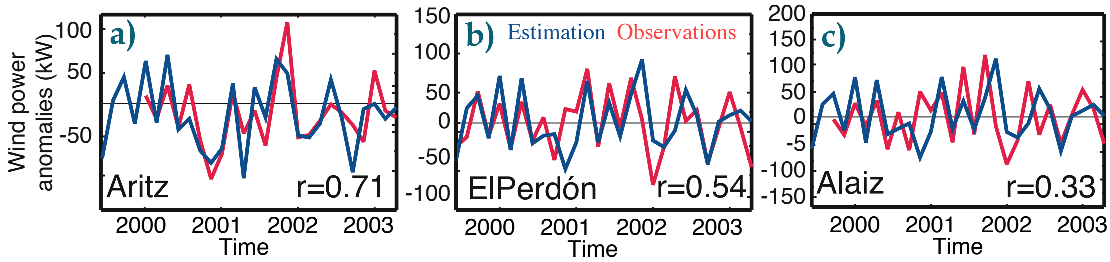

Figure 11: Local estimated and observed monthly wind power anomalies for (a) Aritz, (b) El Perdón and (c) Alaiz. Their correlation values are also indicated. Modified from GB12.

As with CFN, the estimated regional wind anomalies obtained from the reference model also show a great accordance with the observations in NENA (Figure 12; LE17c) with correlation values between the estimated and observed regional zonal (meridional) wind component anomalies of 0.83 (0.73), all significant (p<0.05).

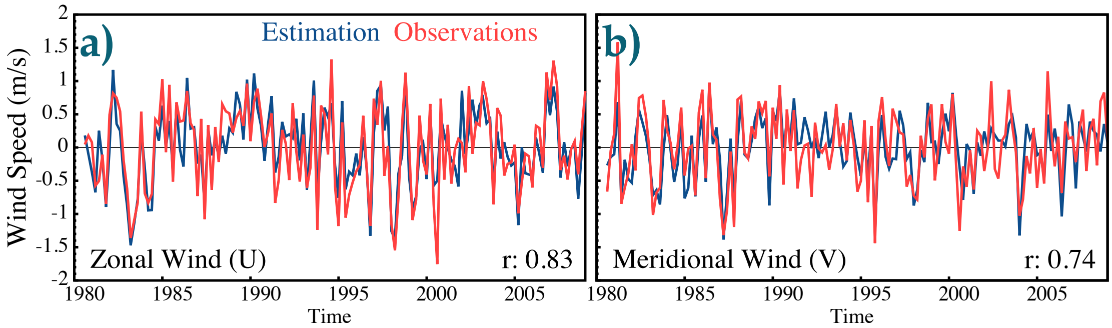

Figure 12: Regional estimated and observed monthly (a) zonal and (b) meridional wind component anomaly series. These estimations have been produced with Era-Interim reanalysis alone. Their correlation values are also indicated

These results evidence a certain underestimation of the variance that can be attributed to the linear constraint imposed by the method in the search of associations between the regional and the synoptic circulations (von Storch and Zwiers 2003).

5. Uncertainty assessment.

In this section we evaluate the uncertainties associated with the statistical downscaling model. The methodological sensitivity is assessed in order to evaluate to what extent a certain choice of parameters in the configuration of the experiment (for example, the reference case) produces an impact in the estimations, thus exploring the robustness of the downscaling strategy. The approach consists in allowing a certain degree of variability in each parameter that is important for the model configuration. This variation of parameters generates an ensemble of estimates that will allow for an assessment of the spatial and temporal variability of the methodological sensitivity. The parameters subjected to change have been previously presented in Section 3. Variations in any of the first four parameters (predictor variable, window size, retained modes and crossvalidation subset) set a family of possible model configurations and will be addressed in subsection 5.1. The fifth parameter, however, is external to the model configuration itself as it is the source of the predictors and its effects will be addressed in section 5.2.

5.1 Uncertainties related to parameter configuration.

This evaluation is conducted in two phases. At first, a simpler approximation on the sensitvity is presented, where, rooted in the *reference configuration* and changing one parameter at a time a small number of combinations can be sampled. This approximation allows for an assessment of the spatial variability of the method sensitivity to changes in one specific parameter (see Table 3). Then, an exhaustive analysis is carried out in GB12, where all the parameters are independently changed, producing a large ensemble of configurations. By doing so the temporal evolution of the methodological sensitivity is studied.

5.1.1 Spatial uncertainty

This uncertainty analysis evaluates the parameter type with the highest sensitivity, fixing one parameter at a time while freely changing the rest. Here the sensitivity is defined as the spread of the ensemble of estimations at each observational site. The range of the configuration parameters is given in Section 3.

Fig. 13 shows the most important parameters for zonal (a) and meridional (b) components in CFN in terms of produced uncertainty. In both cases, the most important parameter is the size of the window used for the large scale information, followed by the number of retained EOF/CCA modes and variable type as predictor. The crossvalidation subset size is irrelevant in every case, indicative that the methodology is statistically robust. The meridional estimations show a greater sensitivity than the zonal ones, related with the natural higher variability of meridional winds (Figs. 13c,d). In fact, the sensitivity vs. variability plot show a linear relationship between these characteristics. Although these analyses are conducted over anomalies, the nature of the wind PDF relates higher variability to higher mean wind speeds, suggesting that in sites with higher wind speeds the estimations are more sensitive to the parameter configurations. This is confirmed by the higher sensitivity values for the sites in the north and those corresponding to the wind farms (Sections 2 and 4).

Similar results have been found in NENA (LE17c; Fig. 14, exercise presented for Era-Interim reanalysis), where the main parameters are either the retained number of EOF/CCA modes or window size (a,b). The sites with higher sensitivity values are distributed close to the coast, and over channelings that foster higher variability (c,d). The outlier in (c,d) corresponds to Mt. Washington, the windiest site on the whole area. This site, incidentally, is more prone to be affected by the predictor variable of choice. Being located in a mountain, is less affected by local particularities and more by the geostrophic flows, that are better captured by upper level predictors. In NENA, contrary to CFN, the variability, and thus the sensitivity values, is higher for zonal winds.

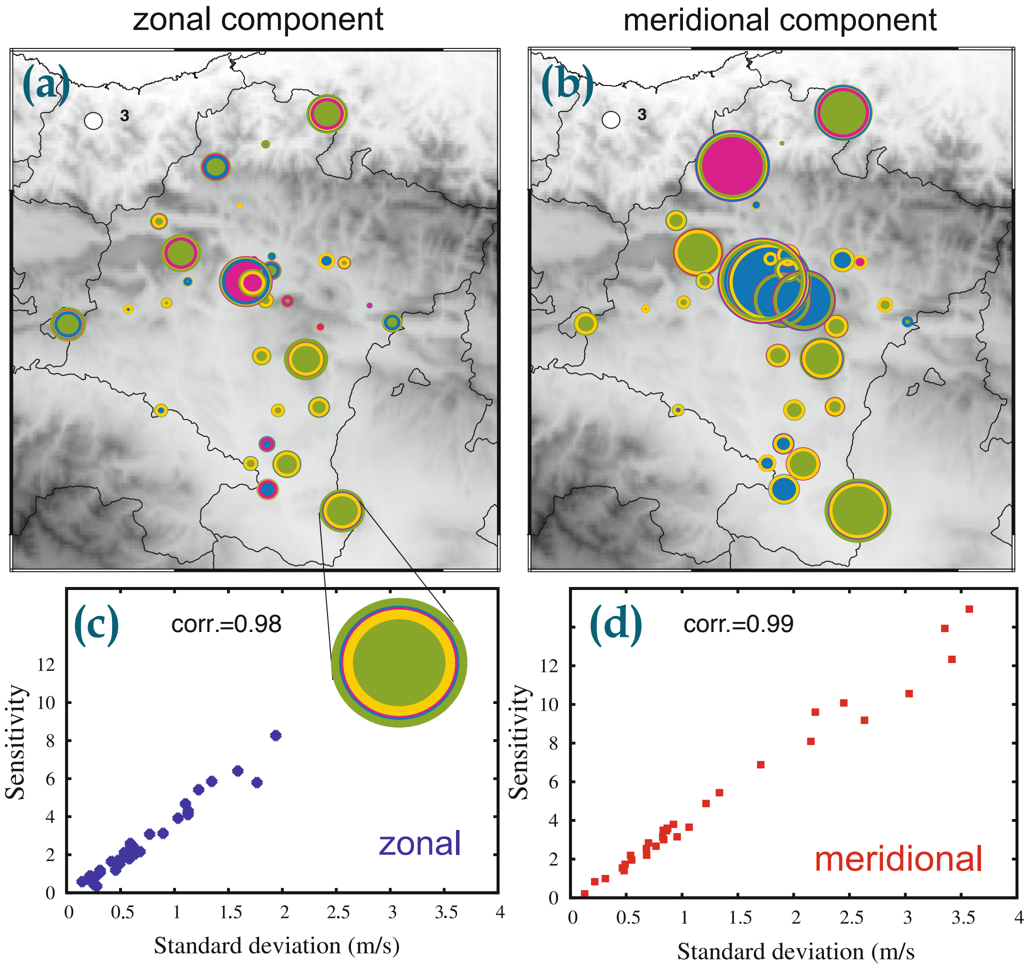

Figure 13: (a,b) Methodological sensitivity to parameter changes for (a) zonal and (b) meridional winds. The sensitivity (in m/s) is indicated with the diameter of the circle. The color of the inner circumference indicates the most relevant parameter at each site (green is related with window size, violet with predictor variable, blue with the number of retained EOF/CCA modes, yellow with the crossvalidation subset size). (c,d) Sensitivity value versus standard deviation at each site for (c) zonal and (d) meridional winds. Modified from GB12.

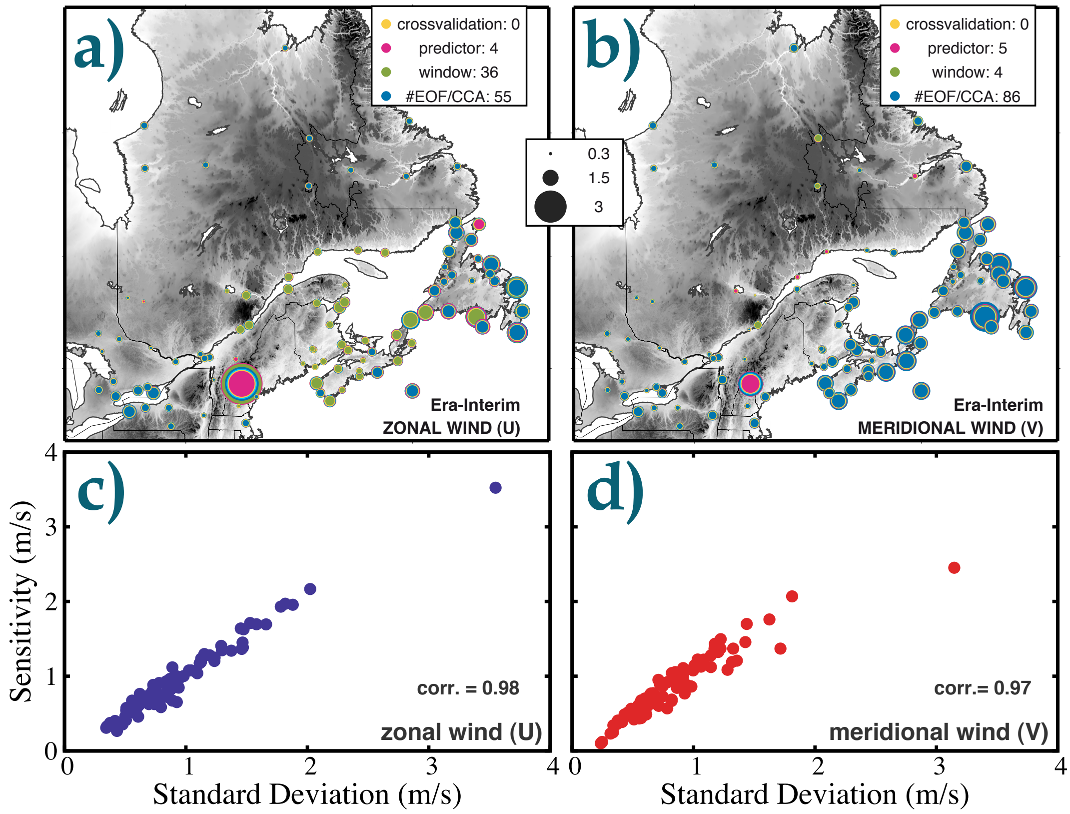

Figure 14: Same as Fig.11 but for North Eastern North America, with Era-Interim reanalysis.

5.1.2 Temporal uncertainty

This analysis allows all the parameters to vary freely, giving rise to a considerably larger number of estimations. In addition to that, the estimation of the wind field has been extended beyond the calibration period used in the previous exercises. This experiment allows to assess the behavior of the ensemble not only in the short timescales that involve the calibration, but also along interdecadal and even multicentenal timescales. To achieve the longest possible reconstruction period, the downscaling method has been applied with Luterbacher et al. (2002) SLP gridded database (see Table 2) as predictor. The reconstruction has been obtained projecting the CCA modes obtained during the calibration period (1993-1999 for this database) over the whole length of the predictor anomalies (1650-1999).

The ensemble of reconstructed regional zonal and meridional wind components is shown in Fig. 15. Some general features regarding the long term past variability can be discussed, observing the overall uncertainty band. The higher variability that the meridional (Fig. 15b) component presented during the calibration period (e.g. Fig. 10) is still maintained for the whole reconstructed period. In the same way, the negative correlation between the two components also exists at longer time scales as a consequence of the momentum conservation. High anomalous winds (vertical blue lines, Fig. 15a,b) can be observed, specially in the meridional component, in the second half of the 17th and 20th centuries, the first half of the 18th century, the beginning of the 20th century and during the observational period (in magenta). No significant trends at multicentennial scales are found for the whole reconstruction period.

A sensitivity analysis of the number of retained canonical patterns on the estimated wind components is also presented. For that, the uncertainties associated with the estimated wind components have been separated according to the inclusion (orange) or exclusion (gray) of the third and fourth canonical modes. The first situation produces a visible bias to positive (negative) zonal (meridional) anomalies in the earlier years of the reconstruction, while the opposite is true for the case with only two canonical modes (gray in Fig. 15). The third large scale canonical pattern (CCA3, not shown) consists in a dipole with positive anomalies over the eastern North Atlantic and a center of opposite sign located to the north of the IP contributing to NE-SW wind anomalies in the region, ideal for Cierzo conditions (de Pedraza 1985). The strength of the association of this pattern with the regional wind in the northeastern Iberian Peninsula is variable and might be related to the time interval considered in the calibration period (GB12). The projection of its corresponding canonical series is presented in Fig. 15c. A negative tendency of the series is apparent, leading to a reversal on the sign of the large scale pattern. This reversal might be responsible of the reverse of sign in the past wind estimations depending on the inclusion or not of the third canonical mode. The apparently tenuous impact of the retained EOF/CCA modes evidenced during the calibration/validation period, turns out to be of importance in the application of the downscaling model outside of the calibration period. This fact stress the need for assessing and understanding of the uncertainties associated to the methodology for obtaining downscaling estimates and illustrates that estimations based on a single configuration of the model must be interpreted with care.

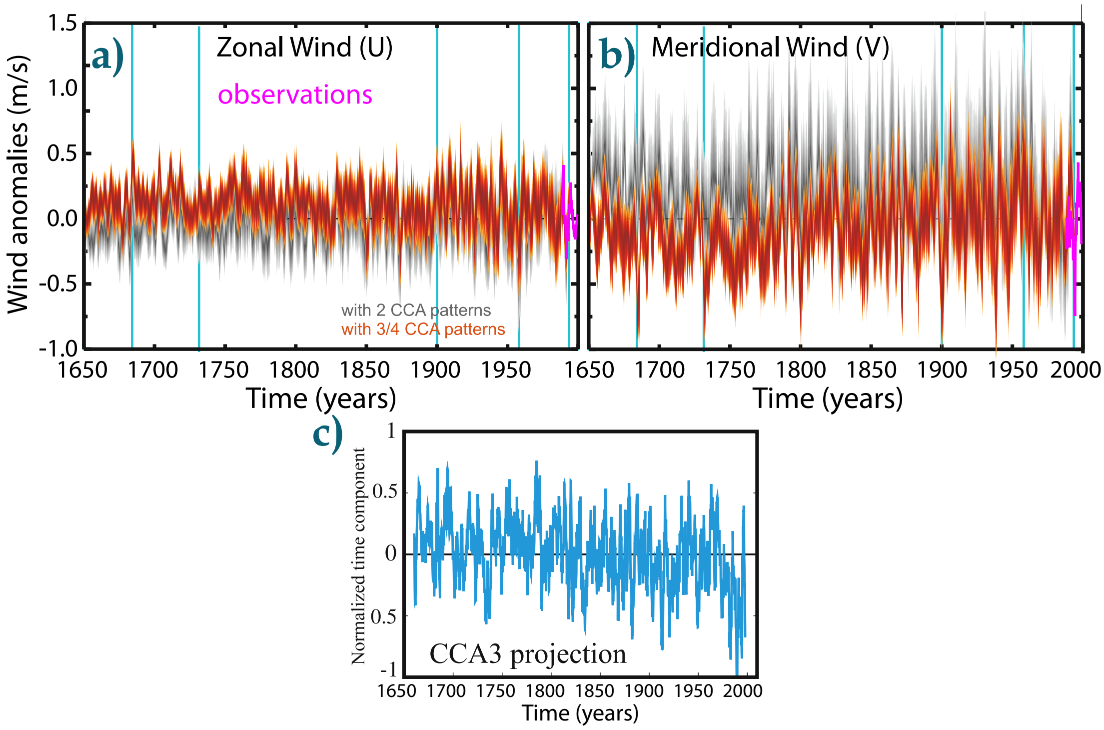

Figure 15 Uncertainty showing the influence of the inclusion of the third and fourth CCA mode obtained using Luterbacher et al. (2002) SLP reconstruction as predictor for zonal (a) and the meridional (b) wind component. Gray areas correspond to the case with only two canonical patterns while orange stands for the cases including the third and fourth patterns. The observations are given in magenta. (c) The projection on the third canonical series from the middle of 17th century till present. All the series present a 2 year moving-average filter. Modified from GB12.

5.2 Uncertainties related to predictor source

In addition to the previous uncertainty analysis a fifth parameter has been also introduced related to the source used for the predictor fields (LE17c). As was previously commented (Section 3.2), the reanalysis models present very different grid resolutions, assimilated data and dynamical and parameterization schemes which could entail, in principle, differences in the described large scale dynamics.

As a first exercise we present the regional and local wind estimations obtained from a particular model configuration. This configuration was run using each of the 10 different reanalysis models in Table 2 as predictor sources. This configuration has been chosen to yeld good estimation results, enabling the assessment of the predictive skills of this method. The configuration chosen for this exercise uses SLP as predictor, the smallest window (window A, Fig. 6b), 7 predictor and predictand EOF modes, retains 6 CCA modes for the estimation and uses a crossvalidation subset of 1 month.

The regional and local estimations derived from the 10 reanalysis models are presented by means of a Taylor diagram (Fig. 16; Taylor 2001) for zonal (a) and meridional (b) components, respectively. This polar diagram integrates the information about correlation, standard deviation ratio and RMSD in a concise way. The correlation value is given with the angle, from worst to best in a clockwise fashion, and the standard deviation ratio is given by the radial coordinate. The RMSE value is indicated in concentric green circles. Each of the small dots represent a site, and the large ones the corresponding regional average. Each reanalysis is represent with a different color.

To address the goodness of the predictive skills in a concise way, we have grouped the results into 3 differentiated regions (r1, r2, r3, Fig. 16). As it can be appreciated, the statistical model offers a good balance between correlations and ratios (r2) for most of the sites, specially for zonal winds (up to 97% of the sites for zonal and 93% for meridional components), regardless of the predictor source. The few sites with bad correlation values (r1) are located in places with complex orography such as mountain ridges, valleys and small islands (not shown). On the other hand, the sites with unrealistic variability representation (r3) are mostly placed to the SW of NENA (not shown), less responsive to the zonal fluxes driven by North Atlantic large scale patterns, the most relevant ones in this region. Zonally channeled narrow straits and sea entrances are pose more problems for meridional estimations, leading to a worse predictive performance. Overall, the method tends to underestimate the variability at the sites, a common problem between linear downscaling techniques (see Section 4.2).

As it can be observed, the regional averages (large dots) show correlations and ratios that depart from the centroid of individual sites, offering much better results. This is due to the fact that the regional averages can cancel out many local effects that in some cases are not well captured by the downscaling model, yielding better results than what are locally able to (GB12). The ratios of the regional estimations are very close to 1, offering a regional realistic variability description.

One of the most remarkable aspects of these results is that they unanimously lead to very similar results irrespective to the reanalysis model used as source. To deepen into this, a sensitivity analysis like the one presented in Section 5.1.2 has been conducted extended to every reanalysis model. As with the former case, the results are hardly dependent on the choice of one reanalysis over another (not shown). This can be explained by the fact that the main large scale patterns that govern this region can be described basically in the same way regardless of the underlying grid resolution (among many other factors). Figure 17 illustrates such a case when SLP variable is used as predictor, but similar outcomes can be observed for other variables (not shown). The cross-correlations between the first 7 Principal Components between all the reanalysis sources (Fig. 17a) present values higher than 0.95 in the best case (PC 1) and 0.7 for the worst case (PC 6), with averages around 0.9. The ranges of their explained variances (inset) are, in the same way, very close, which denotes a similar description of the underlying dynamics. Figs. 17b and 17c show the time evolution of PC1 and PC6. In PC6, although there are some differences in higher frequencies, it can be observed that the long term tendencies are similar for the whole ensemble of reanalysis models.

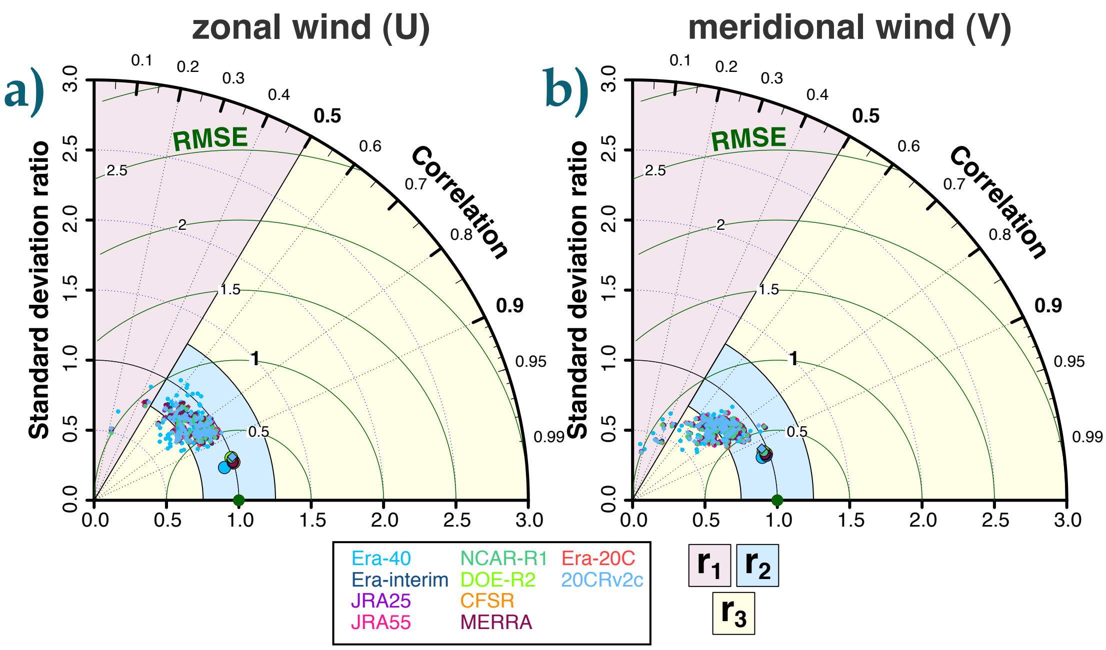

Figure 16: Taylor diagrams constructed with the model estimations using SLP-7/7/6-A-1 configuration in relation with the observations (highlighted sites Fig. 2a) for 10 reanalysis (colors, see Table 2). Each small dot corresponds to a particular site while the large ones depict the regional averages. (a) shows the results for zonal winds and (b) for meridional winds. Each Taylor diagram is divided in 3 regions according to their correlation values and standard deviation ratios: r1 (pink) addresses the estimation with poor temporal correlation values (**ρ**<0.5); r2 (blue) comprises those with best correlation and standard ratio values (**ρ**≥0.5 and 0.75 ≤ σ/σ ≤ 1.25); r3 (yellow) corresponds to sites with good correlation values but bad variability representation.

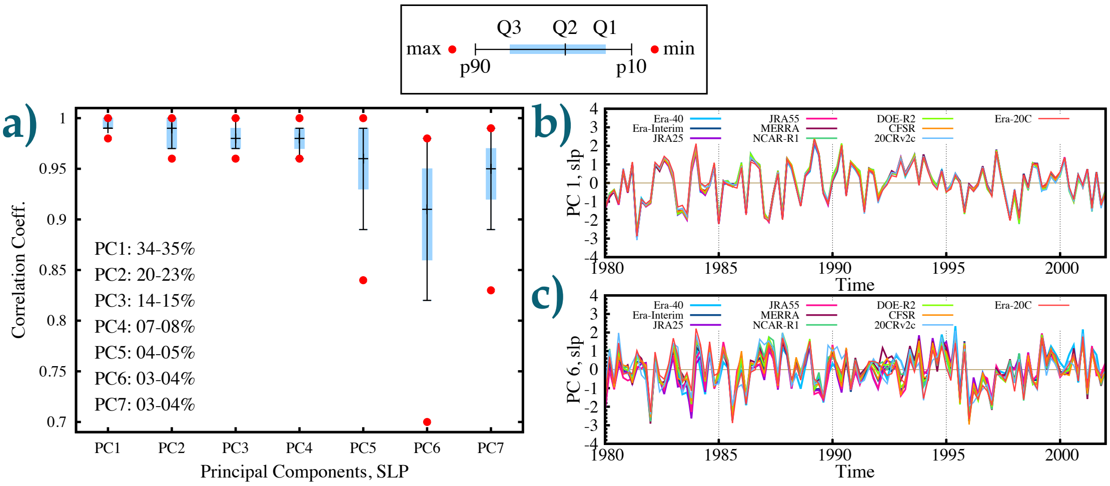

Figure 17: (a) box-plot of cross-correlation values between all the reanalysis sources from PC 1 to 7 for SLP predictor and B window. Inset, range of explained variances is indicated. (b) and (c), temporal plot of PC 1 and 6 for all the reanalysis sources (colors).

6. Conclusions

In this work we analyze the surface wind (GB12, LE17c) and wind power monthly behavior (GB12) via a Statistical Downscaling methodology mainly based in EOF and CCA techniques. The experiments have been conducted over two distinct geographical areas, Comunidad Foral de Navarra (CFN) and North Eastern North America (NENA); and calibration period lengths, 1993-2005/1999-2003 (CFN) and 1980-2010 (NENA), but are centered during the wintertime season in all the cases. We have identified the major large scale circulation patterns (predictor) related to the local variability (predictand) and the predictability of the surface wind via this method. We have also described the spatial and temporal methodological uncertainties to different parameter changes.

Previous to these studies, and to ensure the accuracy of the results, the observational databases (predictand) were subjected to an exhaustive battery of quality checks (Jiménez et al. 2010a; LE17a,b). These tests are focused on the identification and correction/removal/flagging of two families of errors: data management issues (LE17a) and measurement errors (LE17b). The corrections have been proven to have a clear impact on the general statistics of the dataset (e.g., Fig. 5), which underlines the importance of QC procedures to assure any meaningful climatological analysis.

Three major circulation patterns have been described with the help of a reference configuration. In the case of CFN, the leading mode for wind consists on a dipole of positive anomalies over the North Atlantic area and negative anomalies located northeast of the British Isles that favor NW-SE wind flows also known as ‘Cierzo’. For wind power, the leading mode consists on a negative anomaly center located westward of the British Isles that favors southwesterly flows and decouples the behaviour of the northern farms (positive anomalies) from the center (negative anomalies). For NENA, the leading large-scale mode consists on a zonally distributed dipole structure with positive anomalies centred over the North Atlantic and negative anomalies centred over Greenland favouring SW-NE clockwise flows that primarily affects the Eastern area. This reference configuration, despite not being the most skillfull one, still allows us to estimate the local variability with correlation values around ~0.8 for regional wind and ~0.6 for local wind power. The two areas of interest show opposite characteristics as CFN (NENA) shows a greater meridional (zonal) variability and higher predictive skill. The estimations of wind power are related to those of monthly wind speed (GB12), an expected outcome due to the linear relationship between them at monthly timescales (García-Bustamante et al. 2009). In every case the variability is slightly underestimated, a downside of linear dowscaling techniques.

The downscaling uncertainty has been evaluated according to changes in four configuration parameters (predictor variable, window size, retained EOF/CCA modes and crossvalidation periods) and an additional external parameter (predictor source).

For the first case the uncertainty has been assessed either varying one parameter type a time (small ensemble, spatial evaluation) or allowing free independent variations (large ensemble, temporal evaluation). The spatial evaluation has showed that the parameter that exerts a larger uncertainty in the estimations is either the number of retained EOF/CCA modes or by the size of the window for the predictors. On the other hand, the sites with larger variability (typically located over mountains, by the coast or in natural channellings) are more sensitivity to the model configuration, following a linear relationship. The method is also statistically robust, as the crossvalidation subset bears almost no effect. The temporal evaluation has been conducted over multicentenal timescales using the downscaling technique to reconstruct the wind components beyond the calibration period over CFN. The analysis of the ensemble of estimations has shown that the uncertainty ranges obtained during a certain period are not necessarily preserved over other periods, exerting drastic changes in the regional estimation of the wind components. Both perspectives have demonstrated that estimations based on a single configuration must be interpreted with care.

Regarding the uncertainty related to the predictor source. A similar analysis as the spatial evaluation has been conducted over NENA, using 10 different global reanalysis models as sources. The sensitivity analysis shows that the reanalysis models do not play a major role in the uncertainty as their estimations are fairly similar, largely due to the fact that they describe essentially the same large-scale dynamics. Moreover, a local site by site comparison of the estimations given by a certain parameter configuration (that has been repeated for many others as well) delivers very similar results. This configuration, chosen to offer one of the best possible estimations, also shows that the method reproduces in a realistic manner the temporal behavior and variability of most of the sites.

The statistical method appears as a robust approach to estimate the monthly wind and wind production in these regions. Estimations with single configuration must be interpreted with care, specially for reconstruction exercises. The linear predictor-predictand relationship that are established through this method tend to underestimate the variability of the predictand. This suggest that it might be interesting to study alternative approaches that do not emphasize the linearity in the methodology. In addition to that, this analysis filters out non-linear processes that resolve in shorter than monthly timescales that might also contribute to some extent to the wind variability underestimation. This calls for the investigation at, for instance, daily timescales, in order to test their skill in reproducing higher frequency wind or wind power variability.

Acknowledgements

These research results have received funding from the EU H2020 Programme (Nº 689772) and from MCTI/RNP-Brazil under the HPC4E Project, grant agreement nº 689772.

References

Barnett, T. P. and R. W. Preisendorfer, 1987: Origin and levels of monthly and seasonal forecast skill for United States air temperature determined by canonical correlation analysis. Mon. Wea. Rev., 115, 1825–1850.

Burlando M (2009) The synoptic-scale surface wind climate regimes of the Mediterranean sea according to the cluster analysis of ERA-40 wind fields. Theor Appl Climatol 96:69–83.

Buzzi A, D'Isidoro M, Davolio S (2003) A case study of an orographic cyclone south of the Alps during the mAP SOP. Q J R Meteorol Soc 129:1795–1818

Cattell RB (1966) The scree test for the number of factors. Multivariate behavioral research 1(2):245–276

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Besse- moulin P, BroNnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli O, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ (2011) The Twentieth Century Reanalysis Pro ject. Quarterly Journal of the Royal Meteorological Society 137(654):1– 28, DOI 10.1002/qj.776

Conrad CT (2009) Severe and Hazardous Weather in Canada: The Geography of Extreme Events. Oxford University Press

Dee D, Uppala S, Simmons A, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda M, Balsamo G, Bauer P, et al (2011) The era-interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the royal meteorological society 137(656): 553–597, DOI 10.1002/qj.828, URL http://doi.wiley.com/10. 1002/qj.828.

Ebita A, Kobayashi S, Ota Y, Moriya M, Kumabe R, Onogi K, Harada Y, Yasui S, Miyaoka K, Takahashi K, Kama- hori H, Kobayashi C, Endo H, Soma M, Oikawa Y, Ishimizu T (2011) The Japanese 55-year Reanalysis JRA- 55: An Interim Report. Sola 7:149–152, DOI 10.2151/sola. 2011- 038

García-Bustamante, E., J. F. González-Rouco, P. A. Jiménez, J. Navarro, and J. P. Montávez, 2008: The influence of the Weibull assumption in monthly wind. Wind Ener., 11, 483–502.

— 2009: A comparison of methodologies for monthly wind energy estimations. Wind Ener., 12, 640–659.

García-Bustamante, E., J. F. González-Rouco, J. Navarro, E. Xoplaki, P. A. Jiménez, and J. P. Montávez. "North Atlantic atmospheric circulation and surface wind in the Northeast of the Iberian Peninsula: uncertainty and long term downscaled variability." Climate dynamics 38, nº 1-2 (2012): 141-160.

García-Bustamante, E., J. F. González-Rouco, J. Navarro, E. Xoplaki, J. Luterbacher, P. A. Jiménez, J. P. Montávez, A. Hidalgo, and E. E. Lucio-Eceiza. "Relationship between wind power production and North Atlantic atmospheric circulation over the northeastern Iberian Peninsula." Climate dynamics 40, nº 3-4 (2013): 935-949.

Glahn HR (1968) Canonical correlation and its relationship to discriminant analysis and multiple regression. Journal of the atmospheric sciences 25(1):23–31

González-Rouco, J., H. Heyen, E. Zorita, and F. Valero, 2000: Agreement between observed rainfall trends and climate change simulations in the southwest of Europe. J. Climate, 13, 3057–3065.

Graybeal, D. Y., Degaetano, A. T., Eggleston, K. L. Complex quality assurance of historical hourly surface airways meterological data. Journal of Applied Meteorology, 21: 1156-1169, 2004.

Jiménez PA, González-Rouco JF, Montávez JP, García-Bustamante E, Navarro J (2008a) Climatology of wind patterns in the Northeast of the Iberian Peninsula. Int J Climatol 29:501–525

Jiménez PA, González-Rouco JF,Montávez JP, Navarro J, García-Bustamante E, Valero F (2008b) Surface wind regionalization in a complex terrain region. J Appl Meteorol Clim 47:308–325. doi:10.1175/2007JAMC1483.1

Jiménez, P.A., J.F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante. "Quality-control and bias correction of high resolution surface wind observations from automated weather stations". J. Atmosph. Ocean. Techn. 2010a, 27, 1101-1122.

Jiménez, P. A., J. F. González-Rouco, E. García-Bustamante, J. Navarro, J. P. Montávez, J. V.-G. de Arellano, J. Dudhia, and A. Roldán, 2010b: Surface wind regionalization over complex terrain: evaluation and analysis of a high resolution wrf numerical simulation. J. Appl Meteor. Climatol., 49, 268–287.

Kaas E, Li T, Schmith T (1996) Statistical hindcast of wind climatology in the North Atlantic and northwestern European region. Clim Res 7:97–110.

Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL (2002) NCEP-DOE AMIP- II reanalysis (R-2). Bulletin of the American Meteorological Society 83(11):1631–1643+1559, DOI 10.1175/ BAMS- 83- 11-1631.

Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D (1990) A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. Journal of Climate 3(10):1053–1079.

Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, Van Den Dool H, Jenne R, Fiorino M (2001) The NCEP- NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. Bulletin of the American Meteorologi- cal Society 82(2):247–267, DOI 10.1175/1520-0477(2001) 082□0247:TNNYRM□2.3.CO;2

Lorenz EN (1956) Empirical orthogonal functions and statistical weather prediction

Lucio-Eceiza, E. E., J. F. González-Rouco, J. Navarro and H. Beltrami: "Quality Control of surface wind observations in North Eastern North America. Part I: Data Management Issues". Journal of Atmospheric and Oceanic Technology (2017a). In review.

Lucio-Eceiza, E. E., J. F. González-Rouco, J. Navarro, H. Beltrami, A. Hidalgo, and J. Conte: "Quality Control of surface wind observations in North Eastern North America. Part II: Measurement Errors". Journal of Atmospheric and Oceanic Technology (2017b). In review.

Lucio-Eceiza, E. E., J. F. González-Rouco, E. García-Bustamante, J. Navarro and H. Beltrami: "Multidecadal To Centennial Surface Wintertime Wind Variability Over Northeastern North America.". Clim Dyn (2017c). In prep.

Luterbacher J, Xoplaki E, Dietrich D, Rickli R, Jacobeit J, Beck C, Gyalistras D, Schmutz C, Wanner H (2002) Sea level pressure fields over the eastern North Atlantic and Europe back to 1500. Clim Dyn 18:545–561.

Michaelsen J (1987), Cross-Validation in Statistical Climate Forecast Models. Journal of Climate and Applied Meterology 26(11):1589–1600.

Najac J, Boé J, Terray L (2009) A multi-model ensemble approach for assessment of climate change impact on surface winds in france. Clim Dyn 32:615–634

North, G., F. Moeng, T. Bell, and R. Calahan, 1982b: The latitude dependence of the variance of zonally averaged quantities. Mon. Wea. Rev., 110, 319–326.

Onogi K, Tsutsi J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H, Matsumoto T, Yamazaki N, Kamahori H, Takahashi K, Kadokura S, Wada K, Kato K, Oyama R, Ose T, Mannoji N, Taira R (2007) The JRA-25 Reanalysis. Journal of the Meteorological Society of Japan 85(3):369–432, DOI 10.2151/jmsj.85.369.

Plante M, Son SW, Atallah E, Gyakum J, Grise K (2014) Extratropical cyclone climatology across eastern Canada. International Journal of Climatology 35(10):2759–2776.

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vi- tart F, Laloyaux P, Tan DGH, Peubey C, Th ́epaut JN, Tr ́emolet Y, H ́olm EV, Bonavita M, Isaksen L, Fisher M (2016) ERA-20C: An atmospheric reanalysis of the twentieth century. Journal of Climate 29(11):4083–4097, DOI 10.1175/JCLI- D- 15- 0556.1.

Pryor, S., J. Schoof, and R. Barthelmie, 2006: Winds of change?: Projections of near-surface winds under climate change scenarios. Geophys. Res. Lett., 33, L11702, doi:10.1029/2006GL026000.

Richards W, Abuamer Y (2007) Atmospheric Hazards: Extreme Wind Gust Climatology in Atlantic Canada 1955- 2000. Tech. rep., Meteorological Service of Canada Atlantic.

Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs L, Kim GK, Bloom S, Chen J, Collins D, Conaty A, Da Silva A, Gu W, Joiner J, Koster RD, Lucchesi R, Molod A, Owens T, Pawson S, Pegion P, Redder CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J (2011) MERRA: NASA's modern-era retrospective analysis for research and applications. Journal of Climate 24(14):3624–3648, DOI 10.1175/JCLI-D-11-00015.1.

de Rooy W, Kok K (2004) A combined physical-statistical approach for the downscaling of model wind speed. Weather Forecast 19:485–495

Saha S, Moorthi S, Pan HL, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou YT, Chuang HY, Juang HMH, Sela J, Iredell M, Treadon R, Kleist D, Van Delst P, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, Van Den Dool H, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm JK, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou CZ, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M (2010) The NCEP climate forecast sys- tem reanalysis. Bulletin of the American Meteorological Society 91(8):1015–1057, DOI 10.1175/2010BAMS3001.1

Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres, 106(D7): 7183-7192, 2001.

Trenberth, K., 2008: Observational needs for climate prediction and adaptation. WMO Bulletin, 57 (1), 29–35.

Uppala SM, K Allberg PW, Simmons AJ, Andrae U, Bechtold VDC, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Berg LVD, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hage- mann S, H ́olm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, Mcnally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J (2005) The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological

Society 131(612): 2961–3012, DOI 10.1256/qj. 04.176, URL http://doi.wiley.com/10.1256/qj.04.176

von Storch, H. and F. Zwiers, 1999: Statistical analysis in climate research. Cambridge University Press.

Wetterhall, F., Halldin, S., Xu, C.-Y. Statistical precipitation downscaling in central Sweden with the analogue method. Journal of Hydrology, 306(1): 174-190, 2005.

Wang XL, Wan H, Swail VR (2006) Observed changes in cyclone activity in Canada and their relationships to major circulation regimes. Journal of Climate 19:896–915, DOI 10.1175/JCLI3664.1

Xoplaki, E., J. González-Rouco, D. Gyalistras, J. Luterbacher, R. Rickli, and H. Wanner, 2003a: Interannual summer air temperature variability over Greece and its connection to the large-scale atmospheric circulation and Mediter- ranean SSTs 1950-1999. Clim. Dyn., 20, 537–554.

Xoplaki, E., J. González-Rouco, and J. Luterbacher, 2003b: Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs. Clim. Dyn., 20, 723–739.

Yun, W.-T., Stefanova, L. and Krishnamurti, T.N. 2003. Improvement of the super ensemble technique for seasonal forecast. Journal of Climate 16, 3834-3840. DOI:10.1175/1520-0442(2003)016<3834:IOTMST>2.0.CO;2

Zorita, E. and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J. Clim., 12, 2474–2489.

1135- 9420