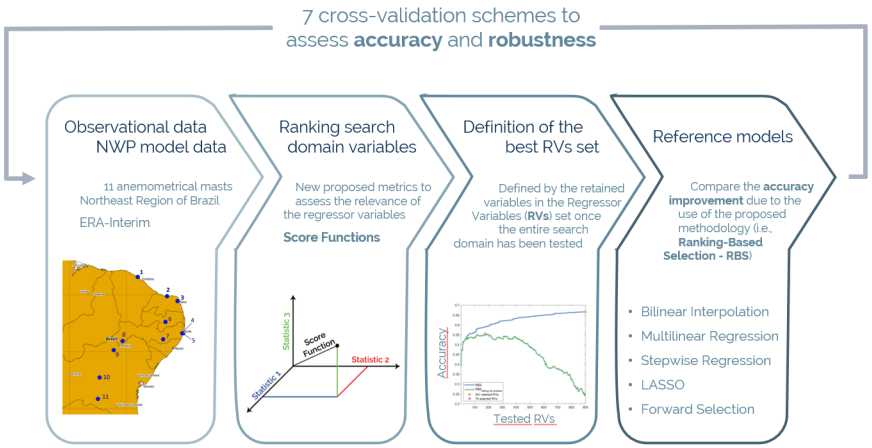
Graphical Abstract

Statistical downscaling of local wind speed based on objective definition of the set of regressor variables

Gabriel Dantas, Alexandre Costa, Olga Vilela, Pierre Pinson



Highlights

Statistical downscaling of local wind speed based on objective definition of the set of regressor variables

Gabriel Dantas, Alexandre Costa, Olga Vilela, Pierre Pinson

- Improving statistical downscaling employing objective selection of GCM grid points
- Low computational effort approach which can be applied to any regressive model
- Benchmarking of different approaches to select GCM grid points
- New proposed metrics to assess the relevance of the regressor variables
- Assessing the methodology robustness concerning various cross-validation schemes

Statistical downscaling of local wind speed based on objective definition of the set of regressor variables

Gabriel Dantas[a], Alexandre Costa[a,b], Olga Vilela[a,b], Pierre Pinson[c]

[a] *Center for Renewable Energy from the Federal University of Pernambuco, Av. da Arquitetura, s/n, CDU, Recife, 50740-550, Pernambuco, Brazil*
[b] *Department of Nuclear Energy from the Federal University of Pernambuco, Av. Prof. Luiz Freire, 1000, CDU, Recife, 50740-545, Pernambuco, Brazil*
[c] *Department of Technology, Management and Economics from Technical University of Denmark, Akademivej, 358, Kgs. Lyngby, Denmark*

Abstract

Statistical downscaling techniques play a fundamental role in addressing the increasing demand for accurate description of the wind behavior oriented to wind farms feasibility studies and short-term wind power forecasting. However, only a few studies have focused on developing or applying objective (automatic) techniques to select which General Circulation Model grid points (i.e., which regressor variables) lead to a higher statistical downscaling accuracy. In this context, this work brings an innovative methodology for selecting the set of regressor variables for the downscaling of the local surface wind speed. The methodology demands low computational effort, being very flexible to be applied with any regressive model at any region of interest. Furthermore, the methodology can employ reanalysis data (aiming for wind resource assessment) or forecasting data (aiming for wind farm dispatch). This work adopts Multilinear Regression to check the methodology concerning accuracy and robustness. The results for 11 case studies in the Northeast Region of Brazil are exciting. For 90% of the cases, the proposed methodology presented equivalent or better results than the reference models that perform regressor variable selection. Furthermore, the methodology significantly reduced computational effort compared to general regressors variable selection approaches such as Forward Selection.

Keywords: Wind Energy, Statistical Downscaling, Predictor Selection, Cross-validation Approaches

1. Introduction

Statistical downscaling models, mainly those based on regressive models, have played a fundamental role in describing atmospheric variables. In the field of Wind Energy, these models are primarily applied for resource assessment (using large-scale reanalysis data as inputs) [1, 2, 3, 4] or for operational forecasting (using large-scale numerical forecasting data as inputs) [5, 6, 7, 8]. The common approaches for selecting regressor variables (RVs) to atmospheric statistical downscaling models subjectively adopt the General Circulation Model (GCM) grid points closest to the site of interest [9, 10, 11]. However, some studies indicate that GCM grid points far from the site of interest may have a high capability to describe local behavior of some atmospheric variables [12, 13, 10, 14, 15, 16]. Thus, one way to enhance the downscaling approaches is by extending the set of macroscale RVs. Nonetheless, an arbitrary increase of the RVs' set may cause a decrease in the overall accuracy as some combinations of RVs may negatively contribute to the model's fit. Such impasse can be addressed by the objective (automatic) RVs selection method. To date, few publications concern the development or application of objective RVs selection methods to atmospheric variables downscaling. Such methods can be divided into two groups: regional selection and individual selection.

The regional selection methods are those in which search domain groups[1] are evaluated as possible RVs, taking into account characteristics of the entire group rather than the individual characteristics of each domain variable. The most usual way to perform this type of RV selection is to test the regressive model performance for different areas of the search domain, either by choosing these areas subjectively [e.g., 17, 18] or by selecting them objectively [e.g., 16]. There are also other feasible ways to make this type of selection, such as the one proposed by Hofer *et al.* [12] or Guo *et al.* [14]. As this class of methods cannot select variables individually, its selection ability is usually quite restricted to areas around the site of interest, neglecting distant variables that could positively contribute to the model's accuracy. Also, some of these regional selection methods depend on the definition of subjective thresholds [e.g., 17, 18, 12, 14], reducing their generalization capability.

[1] The search domain stands for the set of variables used as inputs for the RVs' selection methods. Thus, search domain groups are subsets of the whole search domain.

The individual selection methods[2] are those in which the characteristics of search domain variables are analyzed individually. Many publications that use individual selection methods focus their attention on the nature of the GCM variable to be used (e.g., choosing between temperature, wind speed, and atmospheric pressure) [e.g., 19, 20, 21, 11]. However, they do not focus on selecting the best GCM grid points to be used as RVs. Such selection approaches can occur in two ways: through embedded methods or by means of wrapper methods [22]. The embedded methods are inherent to the training process of the regressive model parameters. Therefore, each embedded method concerns a specific regression model and cannot be applied to other models [23]. Stepwise Regression (SWR) [24, 25, 26, 20] and Lasso Regression [27, 28, 29] as two of the most employed embedded methods, both referring to Multilinear Regression (MLR). On the other hand, wrapper methods address the "task of selecting the best set of RVs" as an "np-complete problem", using the regressive model as a black box to select subsets of variables according to the accuracy they obtain [30, 22]. For this reason, wrapper methods are more general and can be applied in association with any regressive model. However, very few publications make use of wrapper methods to select the best GCM grid points [16], being the most representative that by Sauter and Venema [13] about Simulated Annealing [31]. The individual selection using wrapper methods, such as Simulated Annealing [13] or FS [32], requires a tremendous computational effort. This fact makes them unfeasible for operational applications with large search domains. Also, simulated annealing has a limited generalization capability as it requires the definition of some parameters that may vary according to the site of interest and the specific atmospheric variable.

In this context, this article proposes an objective (automatic) methodology (called as Ranking-Based Selection, RBS) for the individual selection of the GCM grid points to be adopted as RVs applied for the local surface wind speed downscaling. RBS methodology fits into the group of wrapper selection approaches. It tests each of the variables in the search domain following the order established by the score function (that relates each search domain variable with the target). One domain variable is evaluated at each iteration, i.e., the decision is made to add or not the evaluated variable to the set of the best

[2]The individual selection methods are also found in the literature as Screening Regression

regressor variables. In this way, RBS is very general and can be applied with any regressive model for any site of interest. However, unlike most wrapper methods, this methodology requires a low computational effort, being able even for operational applications and on-line learning approaches. In this work, RBS was applied with MLR [33] due to its low computational effort and easiness regarding assessing the results. MLR is also widely applied in the statistical downscaling of atmospheric variables, obtaining satisfactory results in several studies [34, 35, 36, 37, 38].

The following sections of this article deal with: 2. the proposed methodology for the RVs selection; 3. Description of the dataset concerning 11 case studies in the Northeast Region of Brazil, results, and discussion; 4. conclusions and perspective regarding future works.

2. Methodology

The methodology proposed in this work is based on ranking search domain variables according to a score function that relates each domain variable with the target. Then, following the order established by the score function, each of the variables in the search domain is tested to assess whether or not it should be added to the set of best regressor variables. After all domain variables have been tested, the best set of regressor variables is established.

2.1. Score Functions

Here, the score function stands for the norm of an n-dimensional vector. Each coordinates axis projection is given by a statistic that relates a given search domain variable with the target (i.e., observations) (see Figure 1). Each statistic scales from zero (the worst contribution to the target description) to one (the best contribution to the target description). Such normalization aims at establishing a common framework to gather all statistics needed to define a given score function. In this work, 11 different score functions were checked: five one-dimensional, four two-dimensional, one three-dimensional, and the skill score defined by Taylor [39].

2.1.1. One-dimensional score functions

Here, the absolute value of the so-called Pearson's correlation coefficient (Co) (Equation 1) and the Normalized Mutual Information (Mi) (Equation 2) have been checked as one-dimensional score functions.

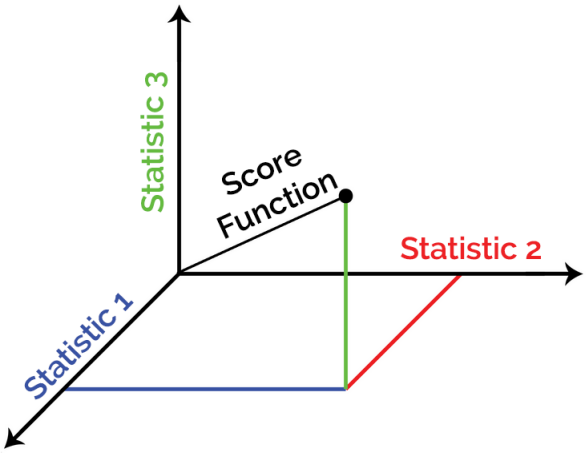


Figure 1: Geometrical description of a three-dimensional score function

$$Co = \left\| \frac{cov(d,g)}{\sigma_d \sigma_g} \right\| \tag{1}$$

$$Mi = \frac{2I_{d,g}}{H_d + H_g} \tag{2}$$

where d stands for a domain variable; g means target variable; $I_{d,g}$ means the mutual information between d and g; H_d means the entropy of d; H_g means the entropy of g; σ_d means the standard deviation of d; σ_g means the standard deviation of g.

Furthermore, three innovative functions are here proposed: a) the score function De, based on the standard deviation ratio (Equation 3);

$$De = 1 - \left[\frac{\left\| 1 - (\frac{\sigma_d}{\sigma_g}) \right\|}{max \left\| 1 - (\frac{\sigma_d}{\sigma_g}) \right\|_{Domain}} \right] \tag{3}$$

b) the score function Ma, based on the root mean squared error between the magnitudes of the Fourier transform of the target and domain variable (Equation 4);

$$Ma = \frac{RMSE_{magnitude}}{\max(RMSE_{magnitude})_{Domain}} \tag{4}$$

$$RMSE_{magnitude} = \sqrt{\frac{1}{N}\sum_{f=1}^{N}\left(\frac{|D_f|}{\sigma_d} - \frac{|G_f|}{\sigma_g}\right)^2} \tag{5}$$

where D stands for the Fourier transform of the domain variable, and G means the Fourier transform of the target.

c) the score function Ph, based on the root mean squared error between the phases of the Fourier transform of the target and domain variable (Equation 6).

$$Ph = \frac{RMSE_{sin} + RMSE_{cos}}{\max(RMSE_{sin} + RMSE_{cos})_{Domain}} \tag{6}$$

$$RMSE_{sin} = \sqrt{\frac{1}{N}\sum_{f=1}^{N}\left(\sin(\theta_{D_f}) - \sin(\theta_{G_f})\right)^2} \tag{7}$$

$$RMSE_{cos} = \sqrt{\frac{1}{N}\sum_{f=1}^{N}\left(\cos(\theta_{D_f}) - \cos(\theta_{G_f})\right)^2} \tag{8}$$

where θ_D stands for the phase angle of D, and θ_G means the phase angle of G.

Note that these last three score functions (De, Ma, and Ph) are conveniently scaled to range from 0 (for the worst domain variable, i.e., that one with the worst ability to describe the corresponding target feature) to 1 (the best domain variable, i.e., that one with the best ability to describe the corresponding target feature).

2.1.2. Two-dimensional and three-dimensional score functions

The four two-dimensional score functions and the single three-dimensional score function used in this study (Equations 9 to 13) are combinations of the following one-dimensional score functions: Co, De, Ma, and Ph.

$$CoDe = \frac{1}{2}\sqrt{Co^2 + De^2} \tag{9}$$

$$DeMa = \frac{1}{2}\sqrt{De^2 + Ma^2} \tag{10}$$

$$DePh = \frac{1}{2}\sqrt{De^2 + Ph^2} \tag{11}$$

$$MaPh = \frac{1}{2}\sqrt{Ma^2 + Ph^2} \tag{12}$$

$$DePh = \frac{1}{3}\sqrt{De^2 + Ma^2 + Ph^2} \tag{13}$$

128

129

130 However, no score function was composed simultaneously by Co and Ma
131 or Co and Ph. This was done to avoid redundancy due to the intrinsic
132 relationship between the correlation coefficient and the signal's frequency
133 structure and phase. Please, note that the coefficients (1/2, 1/3) are adopted
134 to scale the functions within the range [0,1].

135 *2.1.3. Taylor's skill score as an extra one-dimensional score function*

136 For this study, the skill score (SS4) described by Taylor [39] was also
137 checked as a score function (Equation 14). According to Taylor's definition,
138 note that it also ranges from zero to one.

$$SS4 = \frac{\left(1 + \frac{cov(d,g)}{\sigma_d \sigma_g}\right)^4}{4\left(\left(\frac{\sigma_d}{\sigma_g}\right) + \left(\frac{\sigma_g}{\sigma_d}\right)\right)^2} \tag{14}$$

139 *2.2. Finding the best RVs set*

140 The domain variables are sorted in descending order according to the
141 value of the employed score function. After that, successive regressions are
142 performed, starting with the best domain variable (the one with the highest
143 score function value) as a single regressor variable. The RVs set includes
144 a new domain variable following the score function values descending order
145 at each new iteration. However, the added variable is only kept in the RVs
146 set if the model's accuracy improves after its inclusion. It is important to
147 note that the domain variable will not be tested anymore in case of rejection.

Therefore, the best set of RVs is finally defined by the retained variables in the RVs set once the entire search domain has been tested.

In this work, SS4 (Taylor's skill score) plays a dual role. On the one hand, SS4 is one of the one-dimensional score functions. On the other hand, it measures the overall model's accuracy mainly because of its ability to assess the similarity between estimates and observations from the point of view of the amplitude of variations (variance) and phase and frequency structure (Pearson correlation). At last, regressions are here made by MLR because of its simplicity and good accuracy for atmospheric downscaling purposes. However, note that any other regressive method could be employed.

2.3. Cross-validations

In this work, k-fold cross-validation is adopted due to the low computational effort and guarantees independence between the temporal blocks adopted for parameters inference and the model's accuracy assessment [40]. In this sense, time series are divided into three distinct periods: calibration, validation, and test. Score functions and regression coefficients regarding accumulated domain variables have been defined over the calibration set. After that, the validation set is used to identify the best set of RVs, i.e., to evaluate the performance improvement provided by each of the domain variables and decide which ones will be retained in the RVs' best set. At last, the final overall performance of the methodology is independently checked over the test set, without any other parameterization nor choice.

A sensitivity analysis checked the independence of the model's accuracy concerning the period of data used in the parameters' calibration (i.e., model's robustness). The first part of the time series (corresponding to two-thirds of the overall length) has been split into calibration and validation sets with different dimensions, being fixed the test set comprising the last third of the time series. In this sense, Table 1 shows seven schemes for the division between the calibration and validation sets.

2.4. Reference models

As a baseline, five reference models are here adopted to an intercomparison with the proposed methodology (i.e., RBS). The first reference model is the Bilinear Interpolation of the four GCM grid points closest to the studied site (hereafter, called BLI4) [41]. The second reference model stands for the multilinear regression based on the four closest GCM grid points (hereafter,

Table 1: Cross-validation schemes

Cross-validation scheme	Calibration	Validation
Scheme 1	"Odd days" in the first two thirds of the time series	"Even days" in the first two thirds of the time series
Scheme 2	"Odd days" in the first two thirds of the time series	First two thirds of the time series
Scheme 3	First third of the time series	Second third of the time series
Scheme 4	Second third of the time series	First third of the time series
Scheme 5	First third of the time series	First two thirds of the time series
Scheme 6	Second third of the time series	First two thirds of the time series
Scheme 7	First two thirds of the time series	Second third of the time series

called MLR4) [10, 11]. This model is widely used for Atmospheric Sciences and represents the simplest way to get a regression without RVs selection.

The following three reference models make use of multilinear regression and different approaches for RVs selection. Two embedded models exclusively designed for multilinear regression with a computational effort similar to RBS: the SWR [24, 25, 26, 20] and the Lasso Regression (Lasso) [27, 28, 29]; at last, a wrapper model: the Forward Selection [32], a computationally intensive approach.

Note that the last three reference models adopt (as input) the same search domain applied to the RBS methodology, and no corrections for bias and standard deviation were applied to the outputs of the RBS and reference models. This was done to establish a common framework for the intercomparison. In addition, the $Improvement_{SS4}$ (see Equation 15) was used to compare the accuracy improvement due to the use of RBS concerning each of the reference models. It is a dimensionless metric, which presents positive values when the RBS's accuracy exceeds the reference model's accuracy, and negative if the opposite occurs.

$$Improvement_{SS4_{RBSoverRef}} = \frac{SS4_{RBS} - SS4_{Ref}}{SS4_{Ref}} \quad (15)$$

where SS_{RBS} stands for the SS4 between the RBS and the target; and SS_{Ref} stands for the SS4 between the reference model and the target.

3. Results and discussion

The discussions on the results are here divided into five main sections. The first describes the observational data used (i.e., the target). The second section describes the macro-scale data used (i.e., GCM output). The third section addresses the different cross-validation schemes. It should be noted that this section points out the best cross-validation scheme, i.e., the following sections refer to results concerning that best cross-validation scheme. The fourth section discusses the decision concerning the best set of RVs. Furthermore, the fifth discusses the final results and comparison with reference models.

3.1. Observational data

The case studies refer to 11 anemometrical masts in the Northeast Region of Brazil, which comprises around 85% of the Brazilian wind power capacity [42]. Figure 2 shows the locations of such masts. Note that sites 10 and 11 are located in a complex orography area. Furthermore, site 4 is located near buildings that acts as an obstacle depending on the wind direction. These characteristics directly affect the wind flow and significantly increase the microscale information present in the observed signal.

All stations provide observational (horizontal) wind speed and wind direction integrated at 10-minute intervals. In addition, cup anemometers are installed at heights varying from 50 to 100 m above ground level. Table 2 lists the observational data sources, Table 3 lists the descriptive statistics of the observational data, and Figure 3 shows the duration of the measurement campaigns. An objective quality assurance procedure was applied to observational data to identify outliers over time series [43, 44]. After performing the quality assurance procedure, the time series integration interval was set to 6-hour to match the macroscale data temporal resolution.

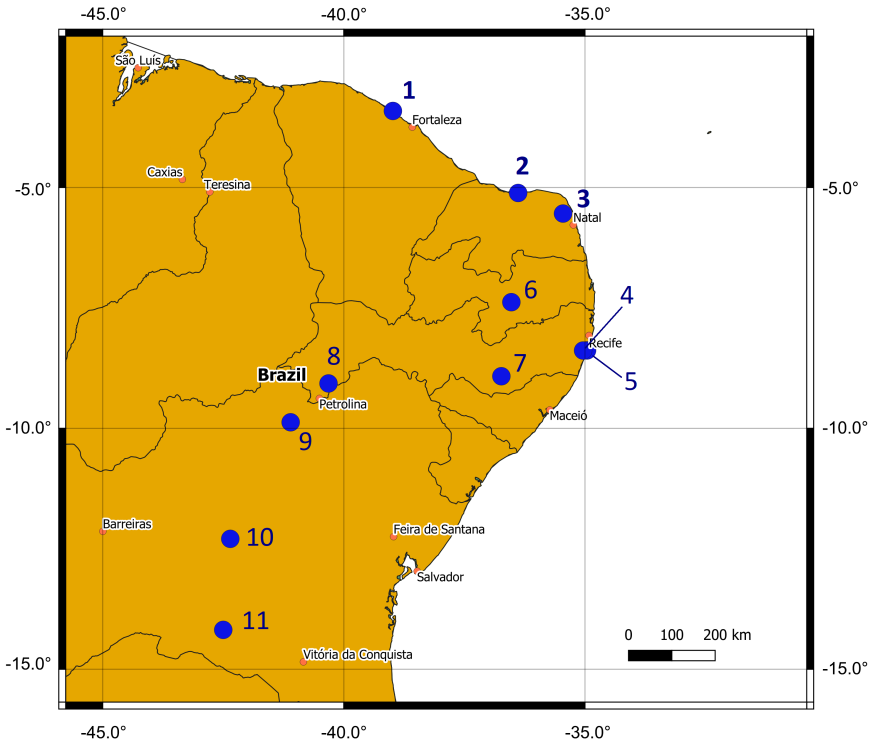


Figure 2: The geographical position of the studied sites

Table 2: Observational data sources

Sites	**Data Provided by**
1	SEINFRA-CE [45]
6 and 8	SONDA Project [46]
4 and 5	Group of Environmental Fluid Mechanics [47] (data granted under the PILACAS project)
2, 3, 7, 9, 10 and 11	National System Operator [48] (data granted under the HPC4E project)

3.2. Macro-scale data

RVs are here magnitudes of horizontal wind speed based on zonal and meridional wind speed components. The ECMWF's ERA-Interim reanalysis program provides them with a spatial resolution of 0.75° at 00 06 12 18 UTC

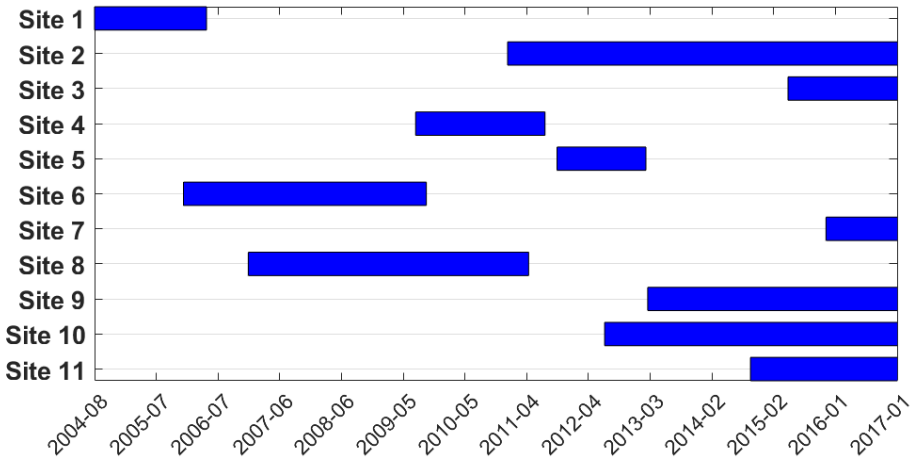


Figure 3: Observational data measurement campaign

Table 3: Descriptive statistics of the observational data

Sites	**μ [m/s]**	**σ [m/s]**	**c [m/s]**	**k [.]**
1	8,90	2,43	9,89	3,49
2	7,49	2,07	8,34	3,41
3	7,49	1,88	8,31	3,61
4	5,95	1,67	6,62	3,38
5	4,68	1,55	5,23	3,06
6	5,51	1,81	6,16	3,10
7	7,82	1,93	8,67	3,63
8	4,81	1,43	5,37	3,30
9	7,61	2,47	8,51	3,10
10	8,71	2,90	9,74	3,08
11	7,88	2,88	8,84	2,86

Where **μ** stands for the average of the observational time series; **σ** stands for the standard deviation of the series; and c and k stand for the Weibull scale and shape parameters estimated for the time series, respectively. Note that the shape and scale Weibull parameters were estimated using the energy pattern factor method [49, 50].

[51]. Here, the search domain comprises 30x30 horizontal grid points centered on each anemometric mast location. The best model level was selected among the ten vertical model levels closest to the surface for each mast location and

each model.

3.3. Cross-validation schemes

Seven cross-validation schemes (CV schemes) were tested to assess the independence of the model's accuracy concerning the period of data used in the parameters' calibration (i.e., model's robustness). To minimize computational effort, only MLR4 was used apart from RBS. Note that the reference model MLR4 does not perform RVs selection. Thus, the validation period does not affect the parameter inference. Because of this, the cross-validation schemes "1 and 2", "3 and 5", and "4 and 6" have, respectively, the same results. That is, effectively, MLR4 uses four distinct cross-validation schemes.

Notice that the MLR4 has very similar accuracy for all cross-validation schemes (Figure 4). This behavior can be viewed through the salmon-colored thin band (for a given site, the lowest and highest band values stand for the poorest and richest CV performance, respectively). However, it is noteworthy that schemes 4 and 6 have slightly higher accuracy than the others.

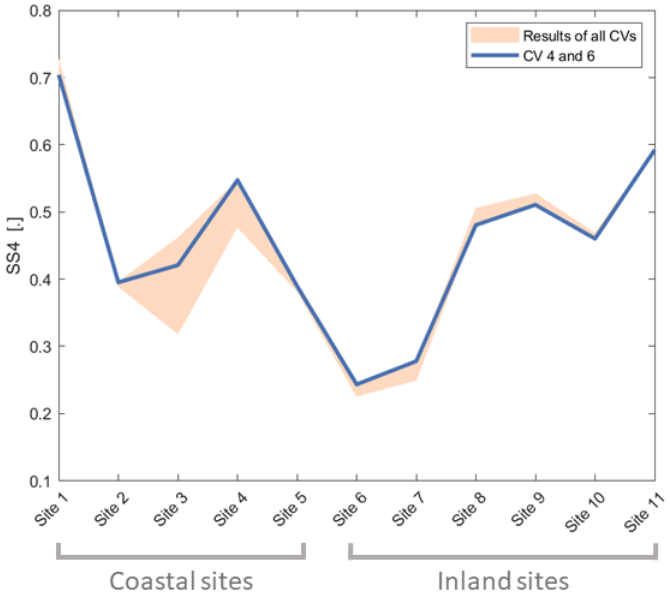


Figure 4: MLR4 best accuracy for all cross-validation schemes and studied sites

The RBS is less robust than the MLR4, i.e., there is a more significant fluctuation in the RBS results as a function of the different cross-validation

schemes (see Figure 5). Note that the greater the number of distinct cross-validation schemes, the greater the fluctuation in the results tends to be. In this case, RBS uses seven distinct cross-validation schemes. However, neglecting CV7, the RBS robustness gets close to that presented by MLR4 (see Figure 5). The low accuracy concerning CV7 is because the calibration period completely contains the validation period, making poor model decision-making over the validation period. On the other hand, the best results for RBS correspond to CV4 and CV1, respectively (see Figure 5). These two cross-validation schemes give more accuracy than the others because their calibration periods comprise data close to the test period, making the regression coefficients appropriate for describing the target behavior over the test period. Also, such schemes do not use concomitant data between calibration and validation periods, which improves the RBS generalization capability. Finally, CV4 was adopted here because of its high accuracy for both MLR4 and RBS models. Therefore, the following sections (about results and discussions) only consider results for that cross-validation scheme (CV4).

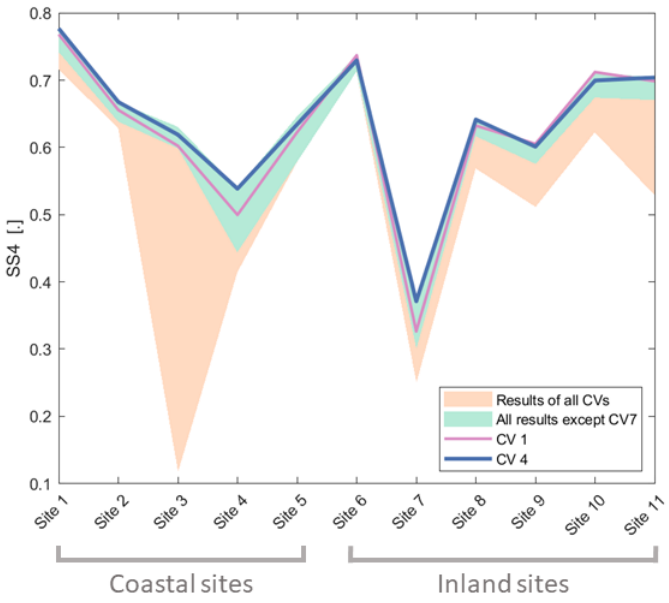


Figure 5: RBS best accuracy for all cross-validation schemes and studied sites

3.4. The decision concerning the best set of RVs

In order to define the best set of RVs for a given site and a given score function, search domain variables are sorted according to the descending order of the corresponding score function values. After, successive multilinear regressions are performed, starting with the best domain variable (the one with the highest score function value) as a single regressor variable. At each iteration, the addition of a new search domain variable is tested following the score function values in descending order (see section 2.2). In this way, Figure 6 shows the accuracy (SS4 in the validation period) for of each of the iterations. The blue curve shows the result of the proposed methodology. The green curve shows the addition of domain variables to the RVs set without eliminating those that do not promote improvements. Such figure shows the result for site 9 and the score function Ma (the best score function for site 9). After checking the 10 GCM model levels, the third level closest to the surface presented the best results.

In this regard, the first important aspect of being observed concerns improved performance due to the rejection of variables that do not improve performance when added to the VRs set (difference between the blue and green curves in Figure 6). This is mainly due to the reduction of redundant information present in the set of VRs. This excess of redundant information is mainly because the score functions consider the similarity of the signals in the frequency domain (except for De). As a result, the variables with the highest score function values tend to be quite similar. The reduction of redundant information due to the rejection of variables allows the model to select variables that can assist in describing the observed signal without a solid tendency to overfitting despite the low value of the score function. With this, RBS can select variables with low similarity with the target. This fact is only possible in embedded methods or high computational effort wrapper methods, such as forward selection. However, RBS, despite being a wrapper methodology, can accomplish this by requiring a low computational effort since the number of regressions performed in the selection process remains numerically equal to the number of variables in the search domain.

3.5. Performance of the Score Functions

Among the 11 score functions studied here, 9 are innovative contributions of this work. Figure 7 shows RBS' accuracy for all studied sites according to the different score functions. It should be noted that the accuracy can vary considerably, depending on the score function. Such variability demonstrates

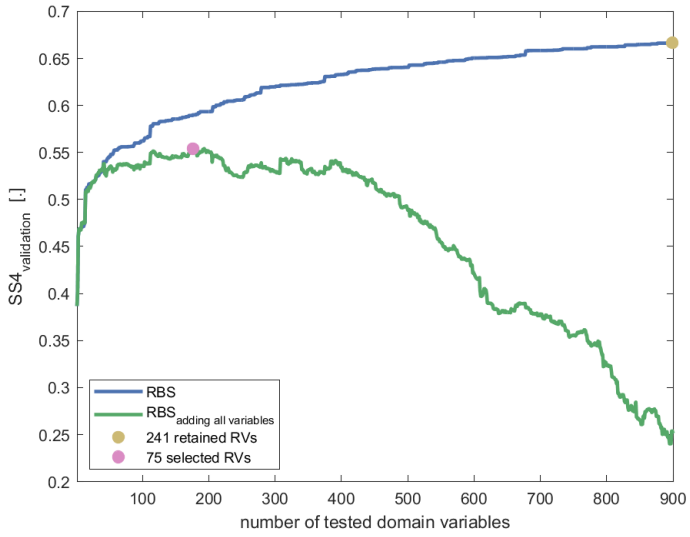


Figure 6: Sorted search domain variables accuracy curve. The abscissa axis represents how many variables of the search domain were tested as input for the regression. The ordinate axis reports the accuracy obtained by the regression. Note that, in the blue curve, the model's performance remains the same when the domain variable is rejected

the importance of the score function, as its choice will considerably affect the downscaling result. Here, results from 5 score functions are emphasized (Ph, Ma, Co, Mi, and De). The worst results come with De (Figure 7b), based on the ratio between the observed signal standard deviation and the "search domain variable" standard deviation (see Equation 3). Such function does not take into account the discrepancies between both signals in the frequency domain. On the other hand, score functions that look at such discrepancies (like Ph, Ma, Co, and Mi) present exciting results. Ph presents the best overall (SS4) results for the coastal sites, except for site 3 (see Figure 7b). For inland sites, Co, Ma, and Mi present similar results (see Figure 7a). Indicating that, for the studied sites, the orderings created from Pearson's correlation and normalized mutual information are more influenced by characteristics arising from discrepancies in the magnitude of the series in the frequency domain than by the phase discrepancies. Thus, due to the improvement in performance obtained by Ma on sites 3, 7, and 8, it is the

one that promotes the best complementarity concerning Ph for the studied sites. So, enabling good performances in all the studied sites using only two score functions.

In addition to the exciting results presented by the one-dimensional score functions, the combination of score functions proposed in this paper (i.e., two- and three-dimensional score functions, see section 2.1) also presented interesting results. It was able to capture the quality of the one-dimensional score functions being combined. For example, MaPh was able to assimilate characteristics of the score functions Ma (which produces good results at inland sites), and Ph (which produces good results at coastal sites). Thus, MaPh presented a more homogenous behavior for all sites (see Figure 8a). However, although the combination promotes greater accuracy homogeneity in the results, it could not match or surpass the best results obtained by the score functions that are being combined. For instance, MaPh could not match or surpass the best results obtained by Ma and Ph individually (see Figure 8b). In this sense, for future work, more in-depth studies are suggested concerning strategies for combining one-dimensional score functions. Given the results presented by the different score functions, the next section (about results and discussions) will only consider the best results obtained by Ma and Ph (as shown in Figure 8b).

3.6. Comparison with reference models

RBS showed substantial improvements in accuracy compared to those reference models that do not perform RVs selection (i.e., BLI4 and MLR4). As shown by Figure 9a, RBS outperforms BLI4 and MLR4 at 10 of the 11 sites studied (i.e., all sites except site 4 – see Tables 4 and 5)[3], obtaining accuracy improvements in the range of 10% to 70% concerning BLI4 and MLR4. Such improvement concerning MLR4 and BLI4 is mainly due to the ability of RBS to make a better description of the phase and frequency structure of the observed signal, as can be seen through the Pearson correlation coefficient (cf. Figure 9b). In addition, a better description of the observed

[3]This work adopted an objective criterion to define situations in which the RBS presented accuracy "lower than", "similar to", and "higher than" the reference model. If $|Improvement_{SS4}| \leq 1\%$, then RBS achieved accuracy similar to the reference model; if $Improvement_{SS4} < -1\%$, then RBS achieved lower accuracy than the reference model; if $Improvement_{SS4} > 1\%$, then RBS achieved higher accuracy than the reference model.

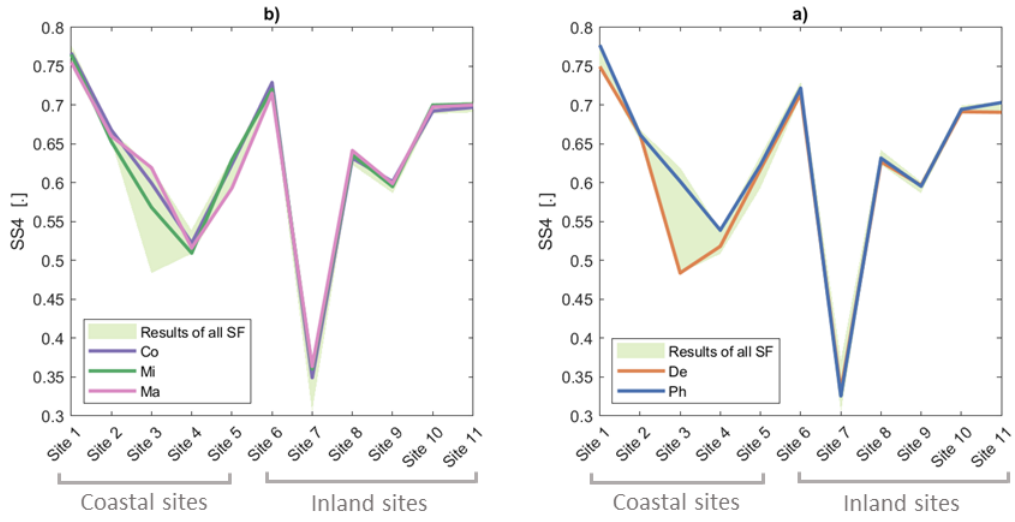


Figure 7: The green area represents the range of results obtained using the best levels of the GCM model for each of the score functions and studied sites. The purple, green, pink, orange, and blue lines show the Co, Mi, Ma, De, and Ph score functions' results, respectively

signal variance also contributed to the improved performance compared to MLR4 (cf. Figure 9c).

Compared to the embedded reference models (i.e., SWR and Lasso), RBS showed good accuracy improvement, mainly in coastal sites (cf. Figure 10a). In inland sites (sites where the best score function is mostly Ma), the performance obtained by RBS is very similar to that obtained by the embedded models. Even so, RBS achieved considerable improvement concerning these models in two of the inland sites (i.e., sites 7, and 8) and was not surpassed by them in any location. RBS showed a substantial improvement over the embedded models in coastal sites, with improvements between 4% and 8% concerning SWR and between 5% and 17% about Lasso (see Table 4 and Figure 10a). Thus, it is noticeable that in coastal sites, the improvement of the RBS over the Lasso is more expressive than over the SWR. Overall, RBS effectively surpassed the accuracy obtained by SWR at 6 of the 11 studied sites, obtaining similar accuracy at 4 sites. Regarding Lasso, RBS was more accurate at 5 of the 11 studied sites, obtaining similar accuracy at 6 sites. This improvement obtained by the RBS about these models is

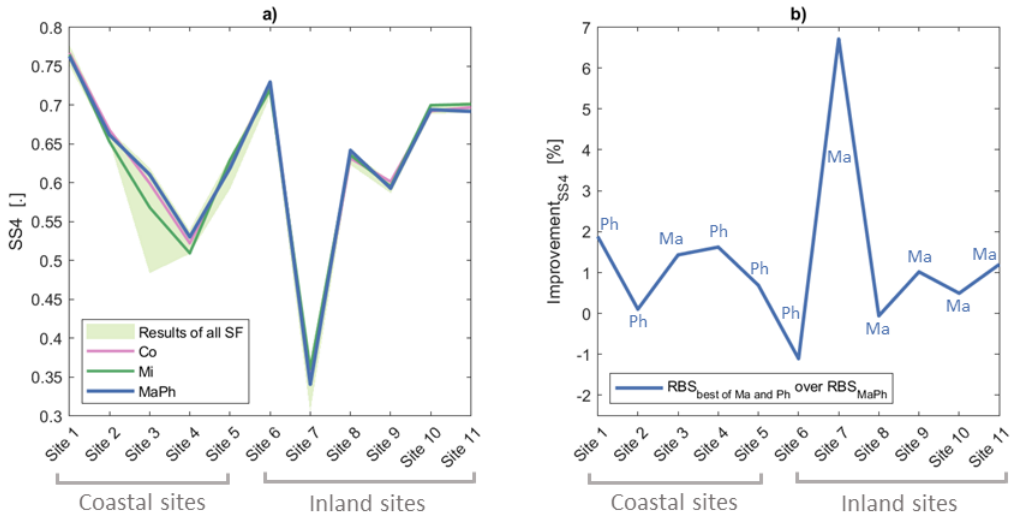


Figure 8: a) The green area represents the range of results obtained using the best GCM model levels for each of the score functions in each of the studied sites and the pink, green, and blue lines represent the results of Co, Mi and MaPh, respectively. b) $Improvement_{SS4}$ of the "Ma and Ph best results" over the MaPh, emphasizing which of the one-dimension score functions (i.e., Ph or Ma) were selected in each of the studied sites

due to a better description of the observed signal in the frequency domain (see Figure 10b) and a better description of the variance of the observed signal (see Figure 10c). Concerning the computational effort, RBS demanded slightly more than SWR and the same as Lasso. Therefore, the trade-off between accuracy and computational effort demonstrates the feasibility of RBS compared to RVs selection models explicitly developed to optimize the multilinear regression (i.e., SWR and Lasso).

RBS tends to outperform FS at coastal sites more efficiently and obtain similar results at inland sites (except at site 7) as in embedded reference models. Overall, RBS effectively outperforms FS at 6 of the 11 studied sites, showing similar accuracy at 3 sites (see Table 4 and Figure 11a). Furthermore, the improvement of RBS over FS reaches 22%, whereas the improvements of FS over RBS do not exceed 2.3%. The improvement of the RBS over the FS is mainly due to a better description of the phase and frequency structure of the observed signal (see Figure 11b), as the description of the

variance of the observed signal is quite similar in both models (see Figure 11c). From the point of view of computational effort, RBS shows a good advantage. To get the best set of RVs, the FS needs to perform about 61 to 288 times more regressions than the RBS (see Table 6). This reduction in the computational effort is because, unlike the FS, the number of regressions performed by the RBS varies linearly with the number of variables in the search domain. Therefore, when considering accuracy and computational effort, RBS becomes much more interesting than FS.

It should be emphasized that RBS showed higher accuracy than the other reference models at coastal sites (see Table 7). This result demonstrates the methodology's great applicability, given the large number of wind power plants located close to the coast, for instance, in the Northeast Region of Brazil. The methodology is also an excellent candidate to present interesting results for offshore applications.

In summary, for 90% of the cases, the proposed methodology presented equivalent or better results than the reference models that perform regressor variable selection (i.e., SWR, Lasso and FS – See green and yellow cells in Table 4).

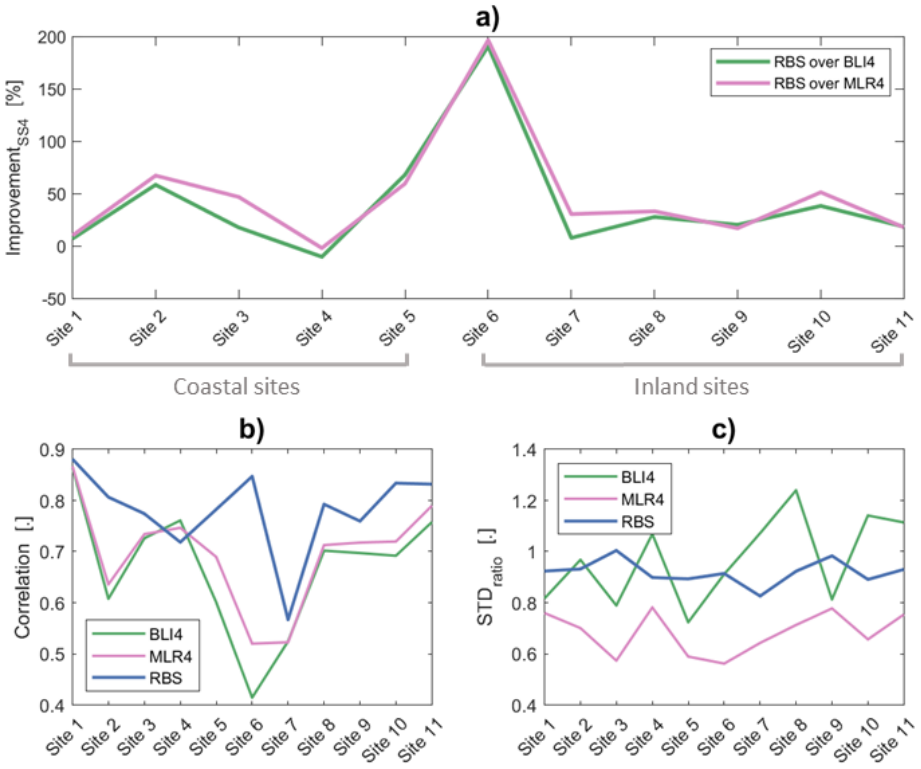


Figure 9: Comparing the accuracy obtained by BLI4, MLR4, and RBS, taking into account the GCM model levels that achieved the best results for each model and each studied site. a) $Improvement_{SS4}$ of the RBS over BLI4 and MLR4; b) Correlation coefficient; c) Standard deviation ratio

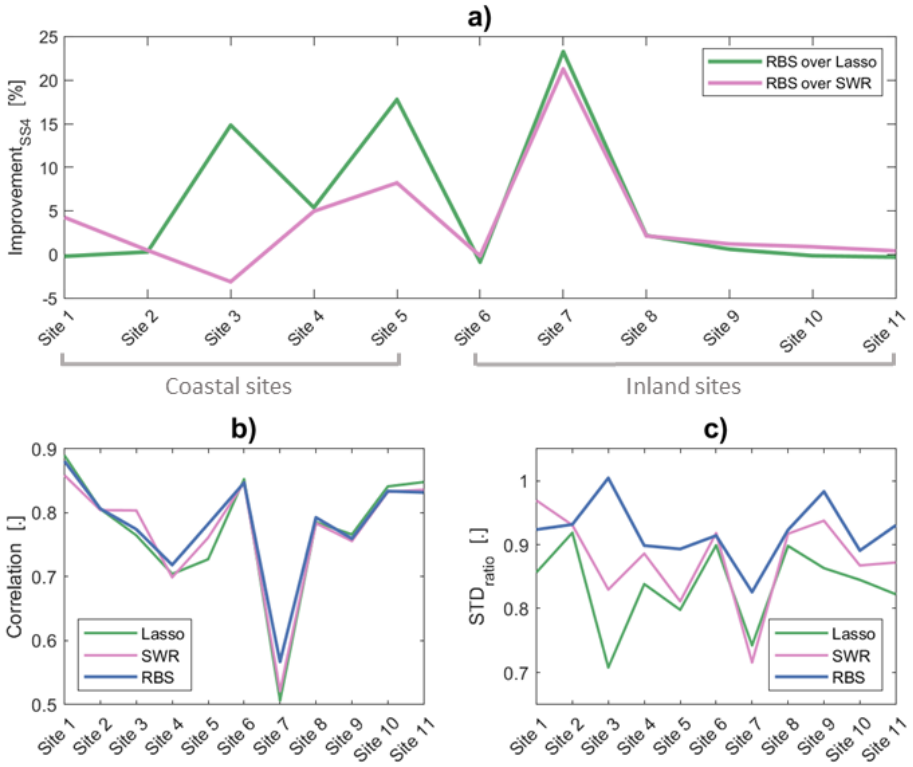


Figure 10: Comparing the accuracy obtained by Lasso, SWR, and RBS, taking into account the GCM model levels that achieved the best results for each model and each studied site. a) $Improvement_{SS4}$ of the RBS over Lasso and SWR; b) Correlation coefficient; c) Standard deviation ratio

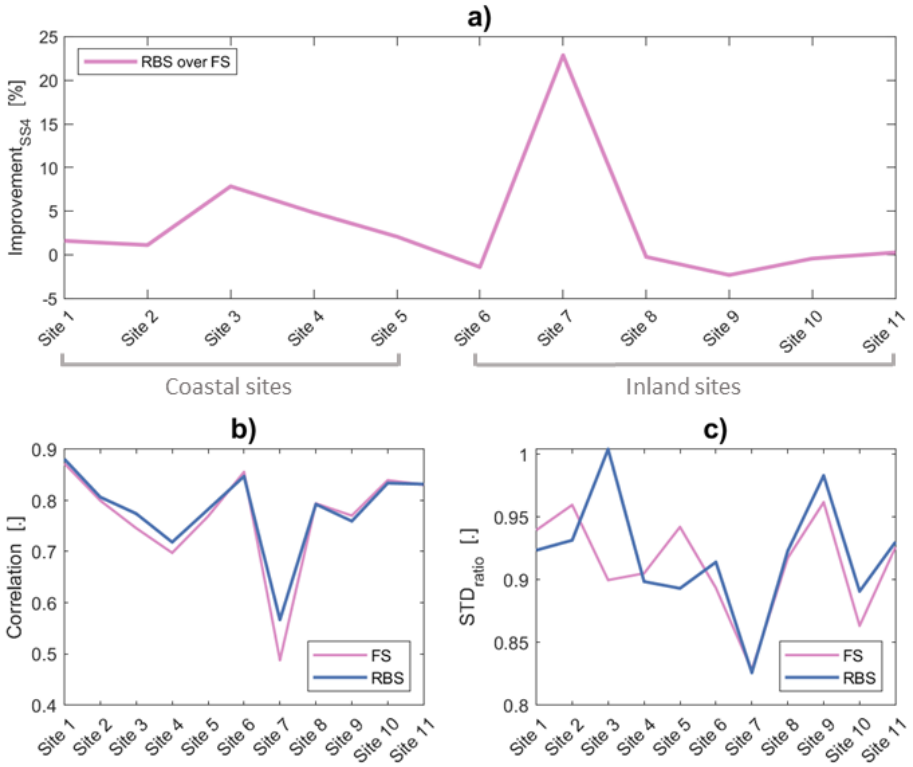


Figure 11: Comparing the accuracy obtained by FS and RBS, taking into account the GCM model levels that achieved the best results for each model and each studied site. a) $Improvement_{SS4}$ of the RBS over FS; b) Correlation coefficient; c) Standard deviation ratio

4. Conclusions and Perspectives

The main highlights of the methodology proposed here are: a) different cross-validations schemes were checked over eleven case studies in order to assess the robustness and accuracy of the proposed methodology; b) eleven score functions were assessed, being eight originally proposed here; c) an objective (automatic) wrapper method to select RVs; d) a benchmarking with different approaches for the statistical downscaling of the surface wind speed.

Assessing different cross-validation strategies is of fundamental importance to guarantee the generalization capability of regressive models. In this

Table 4: Accuracy improvement achieved by RBS regarding the reference models (overall benchmarking)

	$\textbf{Improvement}_{\textbf{SS4}_{\textbf{RBSoverRef}}}$ [%]					
Sites	**IBL4**	**MLR4**	**SWR**	**Lasso**	**FS**	
1	7,28	10,35	4,29	-0,17	1,63	**Coastal sites**
2	58,71	67,48	0,52	0,34	1,15	
3	17,98	47,11	-3,08	14,84	7,86	
4	-9,95	-1,61	4,99	5,41	4,85	
5	68,25	59,83	8,24	17,78	2,12	
6	190,83	196,58	-0,13	-0,83	-1,36	**Inland sites**
7	8,08	30,74	21,27	23,26	22,86	
8	28,05	33,40	2,17	2,22	-0,21	
9	20,52	17,23	1,24	0,64	-2,29	
10	38,57	51,55	0,92	-0,10	-0,40	
11	18,72	18,16	0,45	-0,26	0,29	

Table cells in red color indicate cases where RBS achieved lower accuracy than the reference model; yellow ones indicate where RBS achieved accuracy similar to the reference model; the greens indicate where RBS achieved higher accuracy than the reference model.

Table 5: Overall benchmarking of the reference models regarding the RBS methodology

	Number of sites where RBS accuracy was:		
	Lower than	**Similar to**	**Higher than**
BLI4	1	0	10
MLR4	1	0	10
SWR	1	4	6
Lasso	0	6	5
FS	2	3	6

sense, the sensitivity analysis concerning the seven cross-validation schemes indicates that the best results over the test period are achieved when the calibration period includes data close to the test period without intersection between the calibration and validation periods.

Regarding the score functions, it should be emphasized that eight of the eleven score functions are innovative contributions of this work. The one-dimensional score functions play a prominent role in this paper, mainly Ma and Ph, based on the similarity of the signals in the frequency domain. Ph

Table 6: The number of regressions to achieve the best set of RVs

	Number of performed regressions	
	Forward Selection	**Ranking-Based Selection**
Site 1	95,931	900
Site 2	259,735	900
Site 3	55,456	900
Site 4	117,505	900
Site 5	100,620	900
Site 6	119,016	900
Site 7	115,990	900
Site 8	170,065	900
Site 9	182,871	900
Site 10	150,535	900
Site 11	209,275	900

Table 7: Overall benchmarking of the reference models regarding the RBS methodology considering only the coastal sites

	Number of coastal sites where RBS accuracy was:		
	Lower than	**Similar to**	**Higher than**
BLI4	1	0	4
MLR4	1	0	4
SWR	1	1	3
Lasso	0	2	3
FS	0	0	5

and Ma present the best results at coastal and inland sites, respectively. However, the combined MaPh two-dimensional score function exhibits more robustness regarding the studied sites.

Concerning the mechanism to decide on the best set of RVs, it is clear that the best practice concerns retaining only those search domain variables that improve the model's accuracy. In this way, even variables with less similarity with the target could be retained by the proposed RBS methodology. However, the authors' preliminary results indicate that for short calibration/validation periods, accumulated RVs could lead to better results than the practice of retaining only those variables that improve the model's accuracy.

At last, the benchmark performed with different statistical downscaling approaches reveals the virtues of the methodology proposed here. RBS overcomes or matches all reference models' accuracy in most studied sites, even the FS accuracy. However, RBS has a significantly lower computational cost than FS, and it needs 100 times fewer regressions to achieve results similar to FS.

Furthermore, due to the excellent results obtained by RBS in coastal locations, it is a strong candidate to present interesting results for offshore applications.

Despite the encouraging results presented in this paper, some issues should be addressed to a better understanding of the results, leading to an improvement of the proposed methodology. The first is to study the dependence of the results regarding the duration of the calibration and validation periods. It is also necessary to understand why the best results for Ph (based on the phase similarity of the signals) and Ma (based on the similarity of the magnitude of the signals in the frequency domain) occur in coastal and inland locations, respectively. Furthermore, exploring new strategies to combine one-dimensional score functions and test the proposed methodology for non-linear regression models is necessary.

Acknowledgement

The authors would like to express their gratitude by the R&D projects "High Performance Computing for Energy (HPC4E), Workpackage 4. H2020 EU Programme, Grant Agreement No 689772" and "IBITU.INTELIPREV, in the framework of the ANEEL's R&D Programme (`https://www.aneel.gov.br`)" as well as the fellowship by FACEPE (`http://www.facepe.br`).

References

[1] S. C. Pryor, J. T. Schoof, R. J. Barthelmie, Empirical downscaling of wind speed probability distributions, Journal of Geophysical Research: Atmospheres 110 (D19) (2005). doi:https://doi.org/10.1029/2005JD005899.

[2] J. Badger, H. Frank, A. N. Hahmann, G. Giebel, Wind-climate estimation based on mesoscale and microscale modeling: Statistical-dynamical

downscaling for wind energy applications, Journal of Applied Meteorology and Climatology 53 (8) (2014) 1901 – 1919. doi:10.1175/JAMC-D-13-0147.1.

[3] F. Veronesi, S. Grassi, M. Raubal, Statistical learning approach for wind resource assessment, Renewable and Sustainable Energy Reviews 56 (2016) 836–850. doi:https://doi.org/10.1016/j.rser.2015.11.099.

[4] I. González-Aparicio, F. Monforti, P. Volker, A. Zucker, F. Careri, T. Huld, J. Badger, Simulating european wind power generation applying statistical downscaling to reanalysis data, Applied Energy 199 (2017) 155–168. doi:https://doi.org/10.1016/j.apenergy.2017.04.066.

[5] T. Nielsen, H. Madsen, H. A. Nielsen, P. Pinson, G. Kariniotakis, N. Siebert, I. Marti, M. Lange, U. Focken, L. V. Bremen, G. Louka, G. Kallos, G. Galanis, Short-term Wind Power Forecasting Using Advanced Statistical Methods, in: The European Wind Energy Conference, EWEC 2006, Athènes, Greece, 2006, p. 9 pages.

[6] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, E. Feitosa, A review on the young history of the wind power short-term prediction, Renewable and Sustainable Energy Reviews 12 (6) (2008) 1725–1744. doi:https://doi.org/10.1016/j.rser.2007.01.015.

[7] S. Salcedo-Sanz, Ángel M. Pérez-Bellido, E. G. Ortiz-García, A. Portilla-Figueras, L. Prieto, D. Paredes, Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction, Renewable Energy 34 (6) (2009) 1451–1457. doi:https://doi.org/10.1016/j.renene.2008.10.017.

[8] E. Carlini, G. Giannuzzi, C. Pisani, A. Vaccaro, D. Villacci, Physical and statistical downscaling for wind power forecasting, in: 2016 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), 2016, pp. 338–342. doi:10.1109/SPEEDAM.2016.7526019.

[9] B. Hewitson, R. Crane, Climate downscaling: techniques and application, Climate Research 07 (2) (1996) 85–95. doi:10.3354/cr007085.

[10] C. L. Curry, D. van der Kamp, A. H. Monahan, Statistical downscaling of historical monthly mean winds over a coastal region of complex

terrain. i. predicting wind speed, Climate Dynamics 38 (7) (2012) 1281–1299. doi:10.1007/s00382-011-1173-3.

[11] A. Dupré, P. Drobinski, B. Alonzo, J. Badosa, C. Briard, R. Plougonven, Sub-hourly forecasting of wind speed and wind energy, Renewable Energy 145 (2020) 2373–2379. doi:https://doi.org/10.1016/j.renene.2019.07.161.

[12] M. Hofer, T. Mölg, B. Marzeion, G. Kaser, Empirical-statistical downscaling of reanalysis data to high-resolution air temperature and specific humidity above a glacier surface (cordillera blanca, peru), Journal of Geophysical Research: Atmospheres 115 (D12) (2010). doi:https://doi.org/10.1029/2009JD012556.

[13] T. Sauter, V. Venema, Natural three-dimensional predictor domains for statistical precipitation downscaling, Journal of Climate 24 (23) (2011) 6132 – 6145. doi:10.1175/2011JCLI4155.1.

[14] Y. Guo, J. Li, Y. Li, A time-scale decomposition approach to statistically downscale summer rainfall over north china, Journal of Climate 25 (2) (2012) 572–591. doi:10.1175/JCLI-D-11-00014.1.

[15] P. Horton, M. Jaboyedoff, R. Metzger, C. Obled, R. Marty, Spatial relationship between the atmospheric circulation and the precipitation measured in the western swiss alps by means of the analogue method, Natural Hazards and Earth System Sciences 12 (3) (2012) 777–784. doi:10.5194/nhess-12-777-2012.

[16] S. Radanovics, J.-P. Vidal, E. Sauquet, A. Ben Daoud, G. Bontron, Optimising predictor domains for spatially coherent precipitation downscaling, Hydrology and Earth System Sciences 17 (10) (2013) 4189–4208. doi:10.5194/hess-17-4189-2013.

[17] B. Timbal, B. J. McAvaney, An analogue-based method to downscale surface air temperature: application for australia, Climate Dynamics 17 (12) (2001) 947–963. doi:10.1007/s003820100156.

[18] A. D'onofrio, J.-P. Boulanger, E. C. Segura, Chac: a weather pattern classification system for regional climate downscaling of daily precipitation, Climatic Change 98 (3) (2010) 405–427. doi:10.1007/s10584-009-9738-4.

[19] M. K. Goyal, C. S. P. Ojha, Evaluation of linear regression methods as downscaling tools in temperature projections over the pichola lake basin in india, Hydrological Processes 25 (9) (2011) 1453–1465. doi:https://doi.org/10.1002/hyp.7911.

[20] C. Yang, N. Wang, S. Wang, A comparison of three predictor selection methods for statistical downscaling, International Journal of Climatology 37 (3) (2017) 1238–1249. doi:https://doi.org/10.1002/joc.4772.

[21] R. S. V. Teegavarapu, A. Goly, Optimal selection of predictor variables in statistical downscaling models of precipitation, Water Resources Management 32 (6) (2018) 1969–1992. doi:10.1007/s11269-017-1887-z.

[22] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (null) (2003) 1157–1182.

[23] O. García-Hinde, V. Gómez-Verdejo, M. Martínez-Ramón, C. Casanova-Mateo, J. Sanz-Justo, S. Jiménez-Fernández, S. Salcedo-Sanz, Feature selection in solar radiation prediction using bootstrapped svrs, in: 2016 IEEE Congress on Evolutionary Computation (CEC), 2016, pp. 3638–3645. doi:10.1109/CEC.2016.7744250.

[24] R. Huth, Statistical downscaling of daily temperature in central europe, Journal of Climate 15 (13) (2002) 1731 – 1742. doi:10.1175/1520-0442(2002)015¡1731:SDODTI¿2.0.CO;2.

[25] C. Harpham, R. L. Wilby, Multi-site downscaling of heavy daily precipitation occurrence and amounts, Journal of Hydrology 312 (1) (2005) 235–255. doi:https://doi.org/10.1016/j.jhydrol.2005.02.020.

[26] M. Hessami, P. Gachon, T. B. Ouarda, A. St-Hilaire, Automated regression-based statistical downscaling tool, Environmental Modelling & Software 23 (6) (2008) 813–834. doi:https://doi.org/10.1016/j.envsoft.2007.10.004.

[27] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society: Series B (Methodological) 58 (1) (1996) 267–288. doi:https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[28] D. Hammami, T. S. Lee, T. B. M. J. Ouarda, J. Lee, Predictor selection for downscaling gcm data with lasso, Journal of Geophysical Research: Atmospheres 117 (D17) (2012). doi:https://doi.org/10.1029/2012JD017864.

[29] L. Gao, K. Schulz, M. Bernhardt, Statistical downscaling of era-interim forecast precipitation data in complex terrain using lasso algorithm, Advances in Meteorology 2014 (2014) 472741. doi:10.1155/2014/472741.

[30] R. Kohavi, G. H. John, Wrappers for feature subset selection, Artificial Intelligence 97 (1) (1997) 273–324, relevance. doi:https://doi.org/10.1016/S0004-3702(97)00043-X.

[31] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by simulated annealing, Science 220 (4598) (1983) 671–680. doi:10.1126/science.220.4598.671.

[32] M. Ranaboldo, G. Giebel, B. Codina, Implementation of a model output statistics based on meteorological variable screening for short-term wind power forecast, Wind Energy 16 (6) (2013) 811–826. doi:https://doi.org/10.1002/we.1506.

[33] D. S. Wilks, Statistical methods in the atmospheric sciences, 3rd Edition, Elsevier Academic Press, 2011.

[34] R. Huth, Downscaling of humidity variables: a search for suitable predictors and predictands, International Journal of Climatology 25 (2) (2005) 243–250. doi:https://doi.org/10.1002/joc.1122.

[35] C. S. Cheng, G. Li, Q. Li, H. Auld, Statistical downscaling of hourly and daily climate scenarios for various meteorological variables in south-central canada, Theoretical and Applied Climatology 91 (1) (2008) 129–147. doi:10.1007/s00704-007-0302-8.

[36] J. T. Chu, J. Xia, C.-Y. Xu, V. P. Singh, Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in haihe river, china, Theoretical and Applied Climatology 99 (1) (2010) 149–161. doi:10.1007/s00704-009-0129-6.

[37] D. A. Sachindra, F. Huang, A. Barton, B. J. C. Perera, Least square support vector and multi-linear regression for statistically

downscaling general circulation model outputs to catchment streamflows, International Journal of Climatology 33 (5) (2013) 1087–1106. doi:https://doi.org/10.1002/joc.3493.

[38] X. Li, Z. Li, W. Huang, P. Zhou, Performance of statistical and machine learning ensembles for daily temperature downscaling, Theoretical and Applied Climatology 140 (1) (2020) 571–588. doi:10.1007/s00704-020-03098-3.

[39] K. E. Taylor, Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres 106 (D7) (2001) 7183–7192. doi:https://doi.org/10.1029/2000JD900719.

[40] J. W. Messner, P. Pinson, J. Browell, M. B. Bjerregård, I. Schicker, Evaluation of wind power forecasts—an up-to-date view, Wind Energy 23 (6) (2020) 1461–1481. doi:https://doi.org/10.1002/we.2497.

[41] C. Accadia, S. Mariani, M. Casaioli, A. Lavagnini, A. Speranza, Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids, Weather and Forecasting 18 (5) (2003) 918 – 932. doi:10.1175/1520-0434(2003)018¡0918:SOPFSS¿2.0.CO;2.

[42] ABEEólica, Annual wind energy report 2019, Tech. rep., Brazilian Wind Power Association (2020).

[43] C. Moraes, Procedimento objetivo para a garantia de qualidade de dados observacionais de vento em superfície no litoral do Rio Grande do Norte, Procedimento objetivo para a garantia de qualidade de dados observacionais de vento em superfície no litoral do Rio Grande do Norte, Master's thesis, Federal University of Pernambuco, Recife, Pernambuco, Brazil (2013).

[44] G. Pedrosa, Detecção e Diagnóstico de Falhas na Performance de Aerogeradores, Master's thesis, Federal University of Pernambuco, Recife, Pernambuco, Brazil (2016).

[45] Infrastructure secretariat of the state of ceará, `https://www.seinfra.ce.gov.br/`, accessed: 2021-09-30.

[46] National environmental data organization system (sonda, http://sonda.ccst.inpe.br/, accessed: 2021-09-30.

[47] Group of environmental fluid mechanics, https://www.ufpe.br/mecfluamb/, accessed: 2021-09-30.

[48] Brazilian transmission system operator, http://www.ons.org.br/, accessed: 2021-09-30.

[49] S. A. Akdağ, A. Dinler, A new method to estimate weibull parameters for wind energy applications, Energy Conversion and Management 50 (7) (2009) 1761–1766. doi:https://doi.org/10.1016/j.enconman.2009.03.020.
URL https://www.sciencedirect.com/science/article/pii/S0196890409000934

[50] P. A. Costa Rocha, R. C. de Sousa, C. F. de Andrade, M. E. V. da Silva, Comparison of seven numerical methods for determining weibull parameters for wind energy generation in the northeast region of brazil, Applied Energy 89 (1) (2012) 395–400, special issue on Thermal Energy Management in the Process Industries. doi:https://doi.org/10.1016/j.apenergy.2011.08.003.
URL https://www.sciencedirect.com/science/article/pii/S0306261911004983

[51] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, F. Vitart, The era-interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society 137 (656) (2011) 553–597. doi:https://doi.org/10.1002/qj.828.