

User Guide of the ENSEMBLES Downscaling Portal (version 2)

J.M. Gutiérrez¹, D. San Martín^{1,2}, A.S. Cofiño³, S. Herrera^{1,2}, R. Manzanas¹, M.D. Frías³

¹ *Instituto de Física de Cantabria, CSIC-Universidad de Cantabria, Santander, Spain.*

² *Predictia Intelligent Data Solutions, Santander, Spain.*

³ *Dpto. Matemática Aplicada y C.C. Universidad de Cantabria. Santander, Spain*

correspondence: gutierjm@ifca.unican.es, daniel@predictia.es, cofinoa@unican.es

version: v3–November 2012

Abstract

<http://www.meteo.unican.es/ensembles>. This report describes the structure and usage of the statistical downscaling portal developed by the *Santander Meteorology Group* (<http://www.meteo.unican.es>) with the technical assistance of *Predictia* (<http://www.predictia.es>) as part of the activities of the EU-funded ENSEMBLES project (2004–2009, see <http://ensembles-eu.metoffice.com>). The current operational version (version 2) is a complete reimplementa-tion of the portal, allowing particular adaptations and views for supporting projects and institutions (see the acknowledgements at the end). The three main actions necessary to create a downscaling method (defining the predictors, choosing the local/regional target variable to be downscaled and creating the downscaling method) are described step by step, illustrating the different options available from the portal. Afterwards, the application of the method to downscale GCM climate scenarios is described and some information about validating and interpreting the results is provided. Therefore, this document is intended to be a brief user guide for the downscaling portal and requires additional “good practice documents” to learn about the optimum regions and predictors for the statistical downscaling process.

1 Introduction

Statistical downscaling is a sound and mature field which allows adapting the coarse-resolution (typically 250 km) global climate change scenarios provided by the Global Climate Models (GCMs) to regional or local scale. These methods link the large scale outputs of GCMs (typically large-scale fields such as 500 mb geopotential height) with simultaneous local historical observations (typically surface variables such as precipitation or temperature) in the region of interest. Therefore, these techniques allow filling the gap between the low-resolution GCM outputs and the models used in different impact sectors —such as agriculture, energy or health— which require daily meteorological inputs in special high-resolution grids, or gauge networks.

Statistical downscaling is nowadays a complex multi-disciplinary discipline involving a cascade of different scientific applications to access and process large amounts of heterogeneous data. Therefore, interactive user-friendly tools are necessary in order to ease the downscaling process for end users, thus maximizing the exploitation of the available climate projections. The *ENSEMBLES Downscaling Portal* described in this document was initially developed within the EU-funded ENSEMBLES project (2004–2009) following an end-to-end approach. Afterwards, a complete reimplementa-tion (version 2) was performed to ensure the appropriate adaptation of the portal (different views for different users) to the needs of future supporting projects and institutions (see the acknowledgements at the end for the current list of supporting projects and institutions).

This user guide is intended for end-users with some basic knowledge on statistical downscaling and focus on the steps to be followed to undertake a particular downscal-

ing experiment using the downscaling portal. As an illustrative example, the portal includes a “demo” experiment *Iberia_demo*, which focuses on maximum temperature in five locations/cities for the 2091–2100 decade. This experiment is available for all users (in write-protect mode) and can be followed step to step through the different panels of the portal in order to see a typical application.

The know-how information about selecting appropriate predictors, calibrating/validating the downscaling method, selecting the appropriate GCMs and scenarios, assumptions of the statistical downscaling methodology, etc., is not dealt with in this document. Thus, before using the portal, we strongly recommend the user to read the *Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods*¹ (which constitutes “supporting material” of the Intergovernmental Panel on Climate Change, IPCC).

Finally, we want to remark that **this portal should not be used as a black-box input-output tool** since, otherwise, the obtained regional projections could be misleading or even wrong. Therefore, some background knowledge about the meteorological conditions in the area of interest and the main large scale drivers influencing the climate is needed to appropriately use the downscaling tool and to obtain meaningful results. Moreover, the results obtained from the *ENSEMBLES Downscaling Portal* should not be directly used in impact applications without the necessary knowledge about the assumptions and limitations of this methodology. Thus, we strongly advise end-users to work in collaboration with downscaling groups, or at least have some support from them, in order to define the experiments and to appropriately analyze and use the results.

¹http://www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf

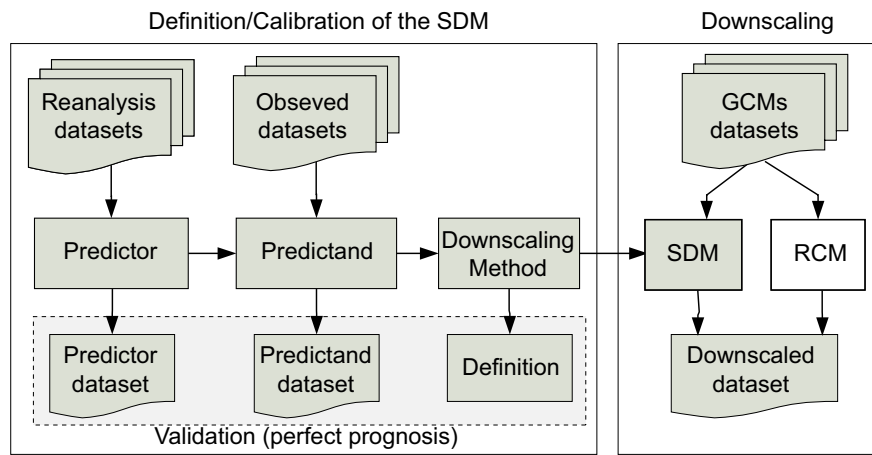


Figure 1: Scheme of the downscaling process using either Statistical Downscaling Methods (SDM) or Regional Climate Models (RCM); in the former case, besides the Global Circulation Model (GCM) scenarios, reanalysis and observed local data are necessary to perform the downscaling. Details of the definition/calibration of the statistical downscaling approach are shown.

2 Downscaling Elements

To fill the gap between the coarse-scale GCM outputs and the local/regional needs of end-users, a number of dynamical models (Regional Climate Models, RCMs) and statistical methods (Statistical Downscaling Methods, SDMs) have been developed. On the one hand, RCMs are directly coupled to the outputs of the GCMs (GCMs datasets) and provide high-resolution (typically 25 km) gridded downscaled datasets for the variables of interest, as simulated from the physical equations and parameterizations included in the RCM (see the scheme of this downscaling process in Fig. 1, left panel). On the other hand, SDMs combine the information of retrospective GCM analysis/forecasts databases (Reanalysis datasets) with simultaneous historical observations of the variables of interest (Observed datasets, either station networks or grids of interpolated observations) to infer appropriate statistical transfer models. Therefore, besides the GCM datasets, two basic ingredients of the statistical downscaling methodology are the Reanalysis and Observations datasets, which are required to define and calibrate the statistical downscaling methods.

The diagram in Fig. 1 (right panel) shows how the different ingredients of the statistical downscaling process are used to define a SDM for a particular application. A particular subset (geographical region, variables and historical temporal window) of the reanalysis constitutes the predictor dataset, whereas the historical records (for the same temporal window) from a goal variable on a number of stations over the region of interest forms the predictand dataset. These data are used to calibrate and validate a particular downscaling method before using it for downscaling purposes (i.e. for projecting GCM datasets). These three basic ingredients are the basis of the portal workflow, as described in the following sections.

The skill of the downscaling methods depends on the variable, season and region of interest, with the latter variation dominating. Thus, for each particular application and

case study, an ensemble of statistical downscaling methods needs to be tested and validated to achieve the maximum skill and a proper representation of uncertainties. Thus, validation is a key issue in the *ENSEMBLES downscaling portal* and, as we will show later, it is automatically performed when a downscaling method is defined.

3 Structure of the Portal

The portal has been organized in different windows (tabs) to gradually access the information necessary to define a downscaling task: (1) Predictor, (2) Predictand, (3) Downscaling Method and (4) Downscale. (1-3) correspond to the calibration/validation of a particular downscaling method, whereas (4) corresponds to the actual downscaling process, applying the calibrated method to different GCMs and scenarios. These windows can be accessed from the corresponding upper tabs of the portal, as shown in Fig. 2 (1).

A first window (*My home*) is shown after login to the portal (see Fig. 2) and provides information about the existing downscaling experiments (2) and the status of the submitted jobs (3), as well as the user's account profile (4). The *Experiment manager* panel shows the details of the experiments already created by the user—a unique experiment, “Iberia_demo”, in this case; see Fig. 2 (5)—, each including a set of predictors (6) defined in a particular region from a reanalysis project—MSLP, T850,Q850 and Z500 from ERA40, in this case—and one, or several, predictands—maximum temperature in five stations in the Iberian peninsula from GSOD_Europe database, labeled as “Tmax_cities” as shown in Fig. 2 (7)—. Each of the predictands may have one or several associated downscaling methods—in this case, only the default *analog* method (8)—. The user can browse the information and navigate through the panel by clicking in the different components.

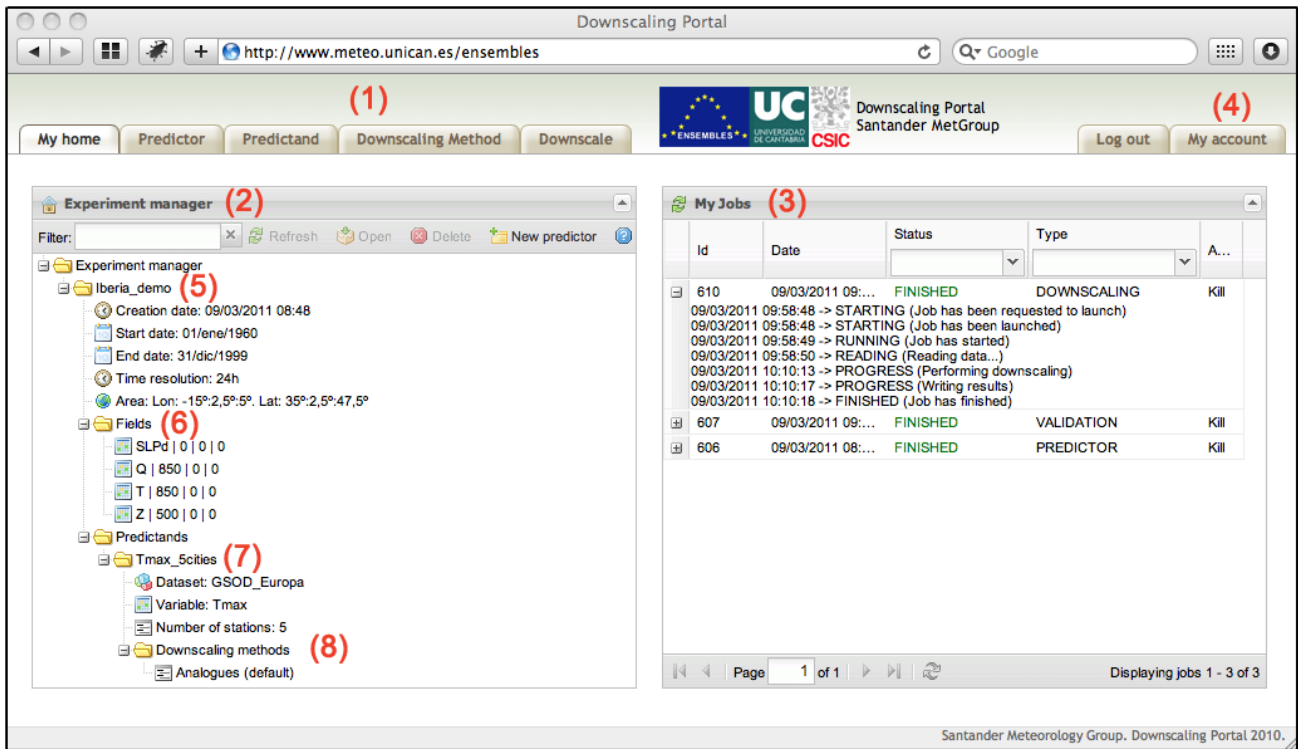


Figure 2: Main window of the downscaling portal. Management of the experiments (left) and the jobs/tasks (right).

The panel *MyJobs* allows monitoring the status (starting, reading, running, finished, etc.) and type (predictors, validation, downscaling) of the jobs, which are run in parallel by the portal through a queue of computational resources, which allows to handle and monitor simultaneously several requests². Moreover, a thread with the different executions stages (reading, performing downscaling, writing results, etc.) and the corresponding execution times can be displayed for each job. Finally, a job can be killed during its execution when it is taking longer than expected and when the user needs extra computational slots. The information about the account details, including the restrictions holding on the resources (number of simultaneous jobs, etc.), can be consulted at any time in the “My Account” tab (see figure 3) in the upper-right corner of the window. It also gives information about the databases available for the current user.

Each downscaling experiment contains all the information needed for the downscaling process: a unique set of predictors, a number of predictands and a number of downscaling methods. To define an experiment the following three sequential steps must be followed, each of them corresponding to each of the tabs shown in Fig. 2 (1):

1. **Predictors:** Definition of the geographical region and predictors to be used in the experiment.
2. **Predictands:** Definition of one or several predictands of interest to be downscaled in this experiment (i.e. with this particular configuration of predictors).

²The current version of the portal runs in the computing infrastructure of the Santander Meteorology Group; <http://www.meteo.unican.es/computing>

3. **Downscaling Method:** Definition and validation of one or several downscaling methods to be applied in the experiment.

Once the *Predictor* → *Predictand* → *Downscaling Method* chain of tasks has been completed, the downscaling methods will be ready to downscale the control and future scenarios of any of the available GCMs (see the scheme in Fig. 1). This final task is done in the **Downscale** window.

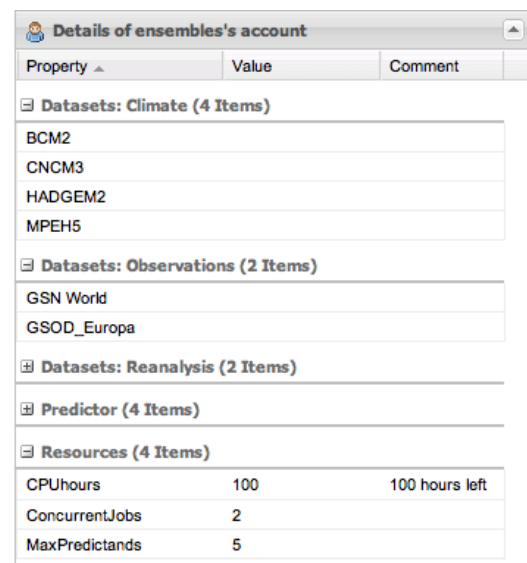


Figure 3: “My Account” tab. It gives information about the databases and resources available for the current user.

4 Selecting the Predictors

Each particular experiment (shown in the “Experiment manager” panel from “MyHome” window) is based on a single predictor dataset defined from reanalysis data over a particular region with a particular resolution. Therefore, a one-to-one correspondence is established in the portal between an “experiment” and the particular predictor dataset used³. New predictors (i.e. new experiments) can be defined from the “Experiment manager” window (“New predictor” button) or from the “Predictor” window (second tab of the portal) by specifying a reanalysis (ERA40 by default), a geographical area, a grid resolution (the original reanalysis resolution by default) and a set of large-scale variables (variable-level pairs).

In order to manage a homogeneous basic set of parameters for the different GCM outputs (reanalysis and climate change projections), a dataset of commonly-used predictor variables at a daily basis has been defined (see Table 1).

Variable (Code)	Levels (mb)	Time
Geopotential (Z)	1000,850,700,500,300	00
V velocity (V)	850,700,500,300	00
U velocity (U)	850,700,500,300	00
Temperature (T)	850,700,500,300	00
Specific humidity (Q)	850,700,500,300	00
Relative Vorticity (VO)	850,700,500,300	00
Divergence (D)	850,700,500,300	00
MSLP (MSL)	surface	daily
2m Temperature (2T)	surface	00

Table 1: Description of the variables, height levels and times (UTC) of the common set of parameters used in the portal. Time values *daily* refer to daily mean values, whereas times 00 refer to instantaneous values.

As a compromise among the different native horizontal resolutions of the models that will be used to project future climate, a common $2.5^\circ \times 2.5^\circ$ grid was considered. Reanalysis and models are interpolated to this grid using standard bilinear interpolation. In particular, we have downloaded, post-processed and stored data for the ERA40 ECMWF reanalysis, the NCEP/NCAR Reanalysis1 (see Brands et al., 2012, for a comparison of these two reanalysis for downscaling purposes) and from different GCMs from the ENSEMBLES project⁴ both in control (20c3m, for 1961-2000) and

³Note that this restriction could be problematic for a friendly use of the portal, since running a downscaling method for a given predictand with different predictors would imply defining a new experiment. However, the flexibility to freely combine predictors, predictands and downscaling techniques lead to data-compatibility problems which can not be solved in a user-friendly form. This restriction may change in future versions of the portal, if the development team find out a solution to overcome these problems.

⁴Both the IPCC-AR4 simulations (ENSEMBLES Stream1) and the new simulations done in the project (Stream2); see <http://cera-www.dkrz.de/WDCC/ui/BrowseExperiments.jsp?proj=ENSEMBLES>

future (B1, A1B and A2, for 2001-2100) scenarios; these models will be described later in the downscaling section. As shown in Fig. 1, predictor datasets are defined based on reanalysis data (since day-to-day correspondence with observations is required in order to establish the statistical transfer functions used for downscaling).

Figure 4 shows the *view*(top) and *create* (bottom) panels from the “Predictor” window allowing to visualize predictor datasets of the existing experiments, or to create new ones, respectively. Note that online help (label 1 in the figure) is provided in all windows to give relevant information about the different tasks to be performed. For a particular experiment selected from the pop-up menu (2) the “view” panel shows the following information (3): Dataset (reanalysis used), Dates (time period), Time resolution (24h for daily data), Lon and Lat (geographical domain), Resolution (horizontal and vertical resolution) and, finally, Predictors (variables used as predictors for the experiment). In this example (*Iberia_demo*) we have considered an area of interest covering the Iberian peninsula and included basic predictor parameters covering the period 1960-1999; this information constitutes the *predictor dataset* (as shown in Fig. 1).

The “create” panel allows defining new experiments by defining the associated predictor dataset (see Figure 4 bottom). In particular in the following we illustrate the definition of the *Iberia_demo* predictor dataset above described. First, a reanalysis must be chosen (4), and the time window and grid (longitude/latitude area and grid resolution) to be used in the experiment must be specified; the map (5) shows the resulting grid. Alternatively, the region to be used can be graphically selected by shift-clicking and dragging in this window, and the resolution can be manually configured in (4). Afterwards, the particular predictors must be selected (6) by choosing the variable, level (when required) and the base hour (by default 00 UTC), for instantaneous variables (see Table 1); in the example shown, the selected predictors are Z at 500 mb, T and Q at 850 mb and SLPd (‘d’ denotes daily mean). Moreover, since the GCM models to be used later for downscaling may lack some of these predictors, a panel (7) indicates the GCMs (among those available for the user) compatible with the selected set of predictors, i.e. the models with the scenario data required to be downscaled within the current predictor dataset (i.e. within the current experiment). Once the information has been defined, a name can be given (8) and the “create new predictor” button can be clicked to define the new experiment. Note that the name of the experiment can be changed afterwards from the “My Home” window.

Note that the definition of a predictor dataset involves several calculations to prepare the data in order to speed-up the downscaling process; for instance, PCs explaining a 99% of the variance are computed (and stored) for the selected period. Thus, when creating a new predictor/experiment, a job is launched to the portal (labeled as ‘PREDICTOR’) and the execution can be followed in the ‘Jobs panel’ until termination. For instance, the ‘My Jobs’ panel shown in Fig. 2 shows a job with ID code 606 ran on 9th March 2011 to define a PREDICTOR dataset (*Iberia_demo* in this case) which lasted 5 minutes (not shown in the figure).

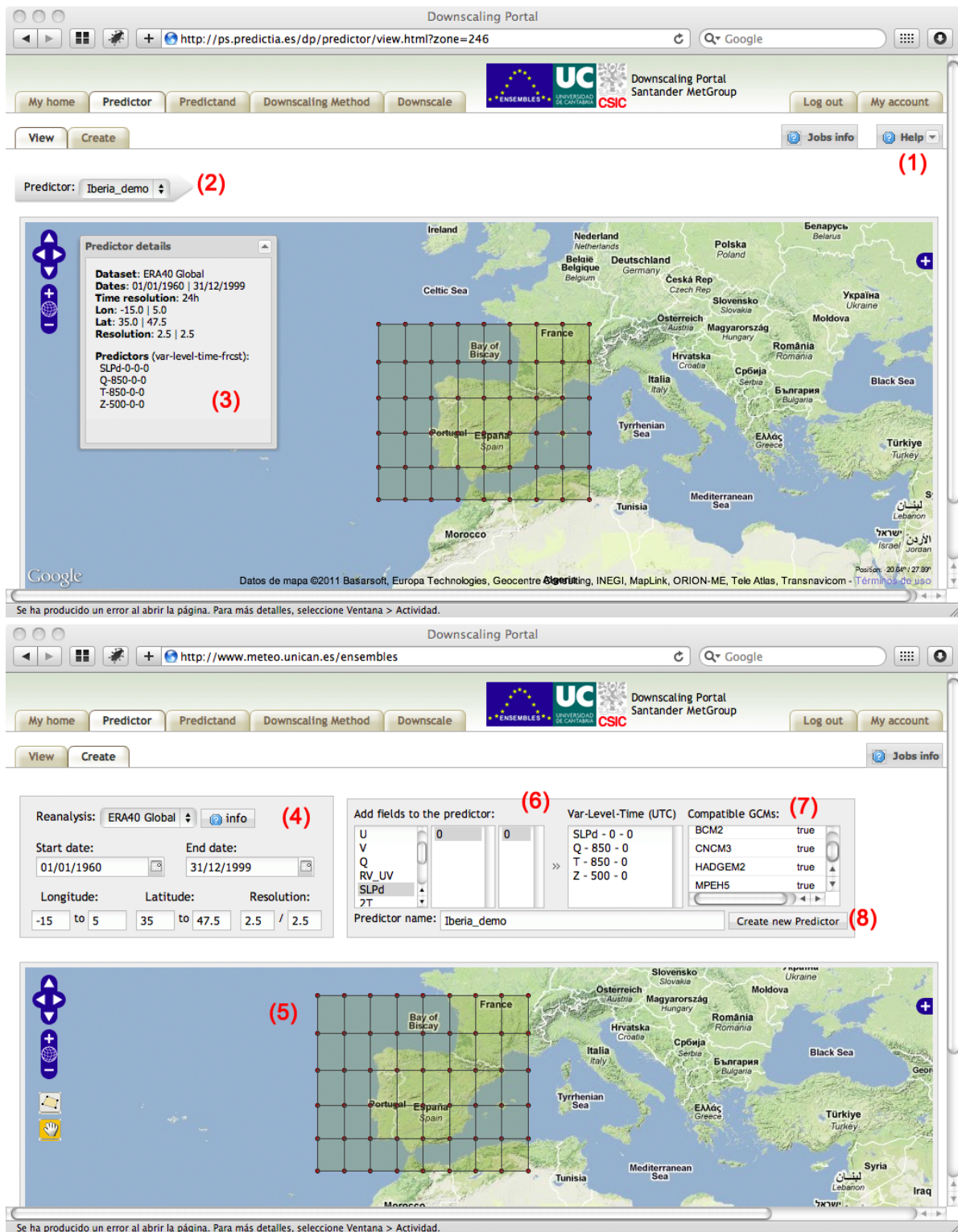


Figure 4: Windows to visualize an existing predictor (above) and to create new ones (below). Numbers refer to the different elements of the windows and are explained in the running text.

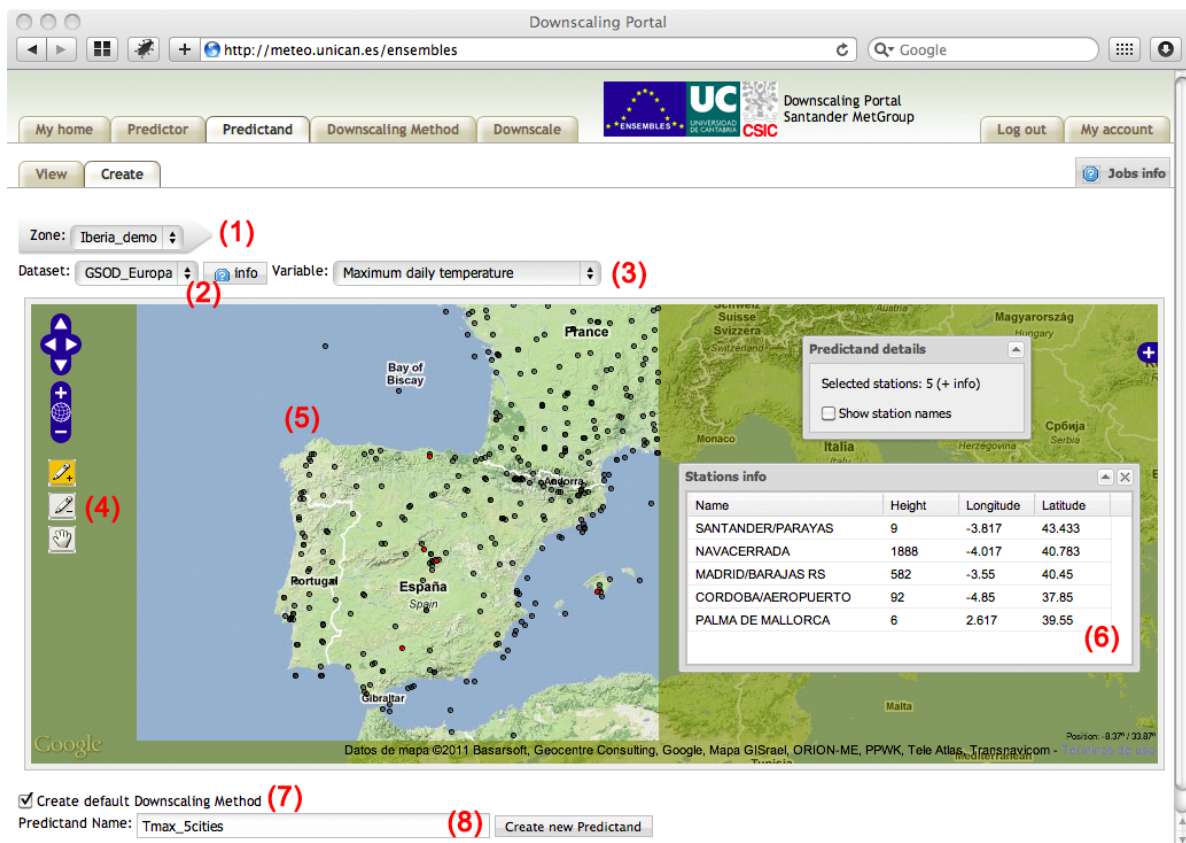


Figure 5: Window to create a predictand for a particular experiment (*Iberia_demo* in this case).

5 Selecting the Predictand(s)

The statistical downscaling portal contains different sources of historical data which can be used as predictands (targets) in the downscaling process. For instance, open-access datasets such as GSN (Global Stations Network) or GSOD (Global Summary of the Day) have been included in order to have a minimum set of historical information to test the downscaling methods worldwide (consult the information about these datasets in the portal). Moreover, the user can include new observation datasets into the portal. This option will be available in the new version of the portal⁵.

The “Predictands” window allows viewing and creating predictands for an experiment from the available historical datasets. Each predictand must be defined by considering a single variable of interest (e.g. maximum temperature) and a number of points/stations among the ones lying within the region defined while creating the experiment (e.g. five cities in the Iberian peninsula). Figure 5 illustrates the steps to be followed to create a new predictand for a particular experiment, selected from the list of available experiments (1). First, the historical dataset to be used must be selected (2), in this case the *GSOD_Europe* dataset, and the variable of interest must be chosen among those existing for the dataset

(3), in this case “maximum daily temperature”. Afterwards the points/stations of interest must be graphically selected by adding (or removing) points (4) and shift clicking and dragging on the map to define an inclusion (or exclusion) square (5); the labels of the stations can be optionally displayed on the map to facilitate this task. Moreover, information about the stations currently selected can be consulted at any time (6). According to the restrictions of the user’s account, there is a maximum of stations/points that can be selected for a particular predictand. For instance, users with a basic profile (i.e. those not involved in the supporting projects or institutions) can only select five stations (see Fig. 3 for additional information).

Once the dataset, variable and stations have been defined, a name can be given to the predictand and it can be included in the corresponding experiment by clicking on the “Create new Predictand” button (8). Note that if the “create default downscaling method” checkbox is selected, then a default statistical downscaling method (a pre-defined analog method) will be defined and validated for this predictand (see the next section); in this case a *VALIDATION* job will be run by the portal and a new “default” downscaling method will be automatically associated to the predictand.

Once the predictor and predictand have been defined for a particular experiment, the common historical dataset will be used to calibrate and validate the different downscaling methods, as explained in the next section.

⁵Note that the downscaling portal is compatible with the *MeteoLab* observation datasets format; see <http://www.meteo.unican.es/trac/MLToolbox/wiki/NewObs>

Figure 6: Configuration panels for the different statistical downscaling techniques included in the statistical downscaling portal: (a) analogues, (b) weather typing, (c) linear regression, (d) neural networks. Note that when the predictand is precipitation linear regression is replaced by generalized linear models (GLMs), with the same configuration options.

6 The Downscaling Method(s)

Different statistical methods have been proposed in the literature to adapt the coarse predictions provided by global climate models to the finer scales required by impact studies. These methods usually work in two steps (*perfect prognosis* approach⁶): Firstly, an empirical relationship (a statistical model) is established between the large-scale reanalysis variables (predictors) and the small-scale observed variables of interest (predictands) using data from a common histor-

⁶Thus, systematic model errors are not taken into account with this methodology and it will be a component of the downscaling error. Recently MOS-like approaches have been tested in the climate change context with promising results. These methods will be included as an alternative to Perfect Prognosis in a future version of the downscaling portal.

ical period (the intersection of the reanalysis time-window and the observations availability period, typically between 15 and 30 years). Then, the resulting statistical model is applied to data from different GCM climate change simulations in different scenarios to obtain the projected local forecast (in this case the predictor data is build considering the predictor variables from the GCM outputs).

Usually, the different statistical downscaling methodologies are broadly categorized into three classes (see, e.g. Gutierrez et al., 2012, and references therein):

- *Weather typing (analogues)*, based on nearest neighbors or in a pre-classification of the reanalysis into a finite number of weather types obtained according to their synoptic similarity; these methods are usually non-generative, since they consist of an algorithmic procedure to obtain

the prediction, such as the method of analogs.

- *Transfer functions (regression)*, based on linear regression or nonlinear models (e.g., neural networks) to infer the relationships between predictands and the large-scale predictors; these methods are “generative” in the sense that the projections are derived from a model obtained from data.
- *Weather generators*, which stochastically simulate daily climate values based on the available monthly average projections or in resampling or simulation procedures applied to the daily data. These techniques are temporal disaggregation methods.

The downscaling portal includes techniques from the first two categories (*weather generators* will be also implemented in a future version of the portal), thus allowing to test and compare the performance of several approaches (note that the skill of statistical downscaling methods varies from variable to variable and from region to region). For a particular experiment, a number of methods can be selected and configured from the “Downscaling Method” window, as shown in figure 6. The *default* configuration corresponds to an analog downscaling method, from the *weather-typing* category, considering the closest analog day (Fig. 6a); additional configurations with a different number of analogs/neighbors (1) and inference methods (2) can be selected by the user in this window. A comment can be included in (3) (this is optional) and a name for the particular technique in (4). These text boxes are defined for each downscaling method as shown in Fig. 6. Finally, the button (5) allows creating the defined technique. The status of the downscaling process can be checked at any time in the “My jobs” panel of “My home” window (Fig. 2). From the second tab of the *weather-typing* category, it is also possible to perform a statistical downscaling method based on a pre-classification of the reanalysis data into a finite number of weather types defined according to their synoptic similarity (Fig. 6b). The clustering method implemented in the portal is the k-means approach (1) and it is possible to select the number of clusters or weather types to be considered (2). As in the *Analogs* tab, different approaches are provided (3), in this case, to infer a prediction from the observations within the corresponding cluster.

The portal also includes linear regression methods (from the *transfer functions* category), (Fig. 6c). In its *default* configuration, only the first five PCs of the predictors are considered for downscaling however, the user can modify this number including an arbitrary number of PCs (1). A number of neighbor grid-boxes can be also introduced in the model (2). This option tries to solve the underestimation of the predicted variability taking into account local effects by means of the nearest grid points data selected. However, the raw model output values will be used as predictors and the spatial coherence of the method will be lost, since different local points will have different downscaling models, with different predictors (the same variables, but over different grid points). It is also possible to apply the linear regression method conditioned on clusters which is based on the k-means approach (3). The number of clusters can be indicated in (4), then a regression model is derived for each cluster. A nonlinear transfer function model based on neural networks is also implemented in the portal (Fig. 6d). In the “Neural Network

(ELM)” tab, the user can also define the number of Principal Components of the predictors (1) or the number of nearest neighbors to be considered in the model (2). The default configuration of this *method* takes 100 hidden neurons and applies a sigmoidal function as activation function of the neural network. This configuration can be modified by the user introducing an arbitrary number of hidden neurons (3) or selecting another activation function within those implemented in the portal (4). A clustering method based on k-means can also be defined (4) indicating the number of clusters to be considered (6) in this case. Note that the Generalized Linear Model (GLM) is also provided from the *transfer functions* category. The GLM is a generalization of ordinary linear regression to variables that are not normally distributed (e.g. precipitation). Then, this approach is available in the portal to downscale precipitation. The configuration for this approach is similar to the options available in the linear regression tab.

As mention above, once a downscaling method is defined, a name must be assigned in the corresponding text-box labeled as *Downscaling method name*. Then, the method will be automatically validated by clicking on the “Create new Method” button. Note that every new downscaling method is automatically validated by the portal. Therefore, a job (labeled as **VALIDATION**) will be submitted to the portal and its execution can be followed in the “Jobs panel” until termination.

6.1 Validation of the SDM

Every downscaling method defined in the portal is automatically validated using a train/test validation approach. The common historical period for predictors (reanalysis; note that this validation is done in Perfect Prognosis conditions) and predictands (local observations) is split into *training* (75% of the data) and *test* (the remaining 25%) subsets. In the *training* phase the downscaling method is calibrated using the training data (e.g. the regression coefficients are fitted to the data), whereas in the *test* phase the method is validated on the test data (note that the test data is not used in the calibration phase and, thus, the results can be extrapolated to new datasets).

The validation results are given in the “View” panel, for a particular predictor, predictand and downscaling method of interest (1 in figure 7). A description of the downscaling method is given in (2). The results of the validation are given both as a summary PDF file (3) and in tabular form in the application window (4).

The validation is performed both on a daily and on a 10-day aggregated basis (4). In both cases, basic statistics (mean, standard deviation, minimum and maximum values and percentage of missing data) of the observations (*Obs. stats*) and the downscaled predictions (*Pred. stats*) are calculated and displayed (5). Furthermore, other scores, such as percentiles, are also computed, but they are not shown in the default view for the sake of clarity (this can be configured in

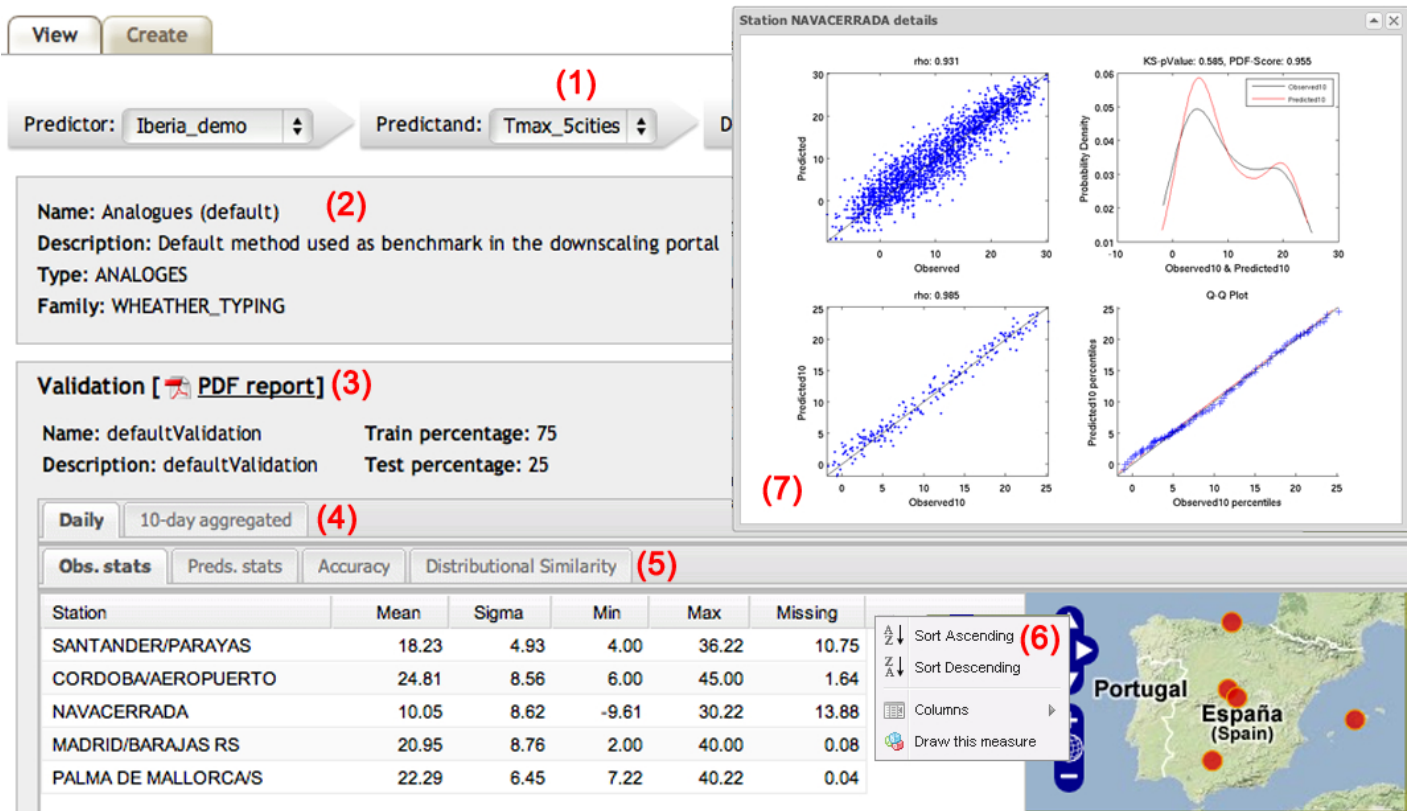


Figure 7: Window to access to the results from the validation of the downscaling method.

the “columns” choice menu in 6). Similarly, the *Accuracy* and *Distributional Similarity* tabs show different validation scores related to the accuracy (the default ones are correlation, MAE, RMSE and normalized RMSE; see Appendix 1) and the reliability (bias, normalized bias, ratio of variances and p-Value of the Kolmogorov-Smirnov test) of the method. In the case of precipitation, some additional scores related to the “occurrence” character of precipitation will also be shown. In particular the ratio of observed and predicted non-precipitation frequencies and the Hit and False Alarm Rates (HIR and FAR, respectively; see Appendix 1 for details).

By clicking on the right arrow in any of the score labels a menu will appear (6). From it, the user can choose which scores (columns) to visualize, the ascending/descending ranking of the stations; moreover, there is the possibility to display the spatial distribution of the score on the right hand side map.

By clicking on any row/station (Navacerrada in this case) a new panel will be displayed (7). This panel shows the basic descriptive and validation scores together with some graphical representations, providing a summary of the performance of the downscaling method. Note that this panel is slightly different for temperature and precipitation due the mixed (discrete/continuous) behavior of the latter variable, that must be taken into account for validation. The two plots on the left show the accuracy of the method in the selected station by displaying a scatterplot of observed vs predicted values for the test data, on daily (top) and 10-day aggregated

(bottom) basis. The more accurate the method, the more linear the plot and the higher the correlation (ρ for temperature and r for precipitation). Note that for precipitation the HIR and the FAR scores are also given in the daily case, thus characterizing the discrete part of the distribution (see Appendix 1 for details on validation scores). The two plots on the right show the distributional similarity of the observed and predicted values, on a 10-day aggregated basis; the upper figure shows the observed and predicted PDFs, including the KS-pValue and PDF-Score (see Appendix 1 for details), whereas the figure in the bottom shows the quantile-quantile plot of the observations and predictions. In the case of precipitation, the plots correspond to rainy days; moreover, the numbers on the top of the figures show the scores for non-rain days and, thus, the combination of both pieces of information gives a general idea of the performance of the method for this mixed (discrete and continuous) variable.

6.2 Downscaling Reanalysis

The validation utilities described in the previous section perform an automatic validation of the downscaling methods by comparing the observations with the corresponding down-scaled values from the reanalysis in the historical period, considering at random 75% of the data for training and the remaining 25% for testing. In order to allow further validation analysis, there is the possibility to apply the statistical down-scaling methods in retrospective perfect-prog mode, by con-

The screenshot shows the 'Downscaling Portal' interface. At the top, there are navigation tabs: 'My home', 'Predictor', 'Predictand', 'Downscaling Method', and 'Downscale'. The 'Downscale' tab is active. Below this, there are sub-tabs for 'Climate Change', 'Seasonal', and 'Hindcast', with 'Hindcast' selected and marked with a red (1). A 'View' button and a 'Create' button (marked with a red (4)) are visible. The configuration section includes: 'Predictor: Iberia_demo', 'Predictand: Tmax_5cities', and 'Downscaling method: Analogues (default)' (marked with a red (3)). Below this, the 'Project' is set to 'ERA40' with an 'info' button (marked with a red (2)).

	January	February	March	April	May	June	July	August	September	October	November	December
1961 - 1970	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1971 - 1980	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1981 - 1990	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1991 - 2000 (5)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

At the bottom, there is a 'Run selected downscalings' button and a footer with 'Developed by: predictia' and 'Santander Meteorology Group. Downscaling Portal 2012.'

Figure 8: Downscaling from reanalysis data (hindcast) for particular time slices.

Considering reanalysis data as input in the statistical downscaling method. In this way the whole predicted series for the historical period can be obtained. Note that, since the training and downscaling periods can overlap in this case, special care is to be taken in the definition of the time-window for the predictors, considering a time-slice of the available reanalysis data. However, in order to avoid problems with the analog method, a one-month temporal exclusion window centered in the downscaling date is considered in this case. Moreover, the sensitivity of the SD methods to reanalysis uncertainty (Brands et al., 2012) can be tested by using different reanalysis datasets as input data at this point.

This option is available in the “Hindcast” tab (1) from the “Downscale” window (see Fig. 8), where different reanalysis datasets (2) can be selected (e.g. ERA40, NCEP). The downscaling method is selected in (3) from those already validated for the selected “Predictor” and “Predictand”. Note that the information of the predictor (“Predictor” tab) includes the particular reanalysis and periods used; this information is to be considered when performing hind cast experiments since the training and downscaling (test) datasets might overlap. Different panels are available for viewing (and downloading) the existing downscalings or for creating new ones (4), such in the present case. The available periods for downscaling are organized in decades, which can be directly selected for downscaling by clicking on the corresponding check-boxes. In the next section, further details are given on the downscaling jobs and the access to the resulting data.

7 Downscaling GCM Scenarios

Once a target predictand to be downscaled has been selected for a particular experiment (predictor set), and the statistical downscaling method has been calibrated and validated using reanalysis data (under Perfect Prognosis conditions), then the downscaling method is ready to be applied to future climate change scenarios, considering GCMs outputs in different control (20c3m, for 1961-2000) and future scenarios (B1, A1B and A2, for 2001-2100). This option is available in the “Downscale” window (the last tab of the application). The portal contains GCM daily data from the following four GCMs: BCM2, CNCM3, MPEH5 (ENSEMBLES Stream1) and HADGEM2 (ENSEMBLES Stream2⁷), which have been validated at a daily basis for the different upper-level fields included as predictors in the portal (Brands et al., 2011). The available variables and scenarios as well as information about the spatial coverage for each particular GCM can be consulted by clicking on the “Info” label for the corresponding model or on the “My Account” tab (as shown in Fig. 3).

Figure 9 shows the “Downscale” window with the “create” tab selected, as shown in (1). This window allows creating new downscalings for a particular predictor, predictand and downscaling method, selected from (2), as well as the scenario of interest (A1B in this case). For the particular

⁷The details of the models are given in <http://cera-www.dkrz.de/WDCC/ui/BrowseExperiments.jsp?proj=ENSEMBLES>. Preferably Stream1 models were selected for this version of the portal; however, HADGEM2 was selected from Stream2 because the availability of daily data for the Stream1 MetOffice models was limited.

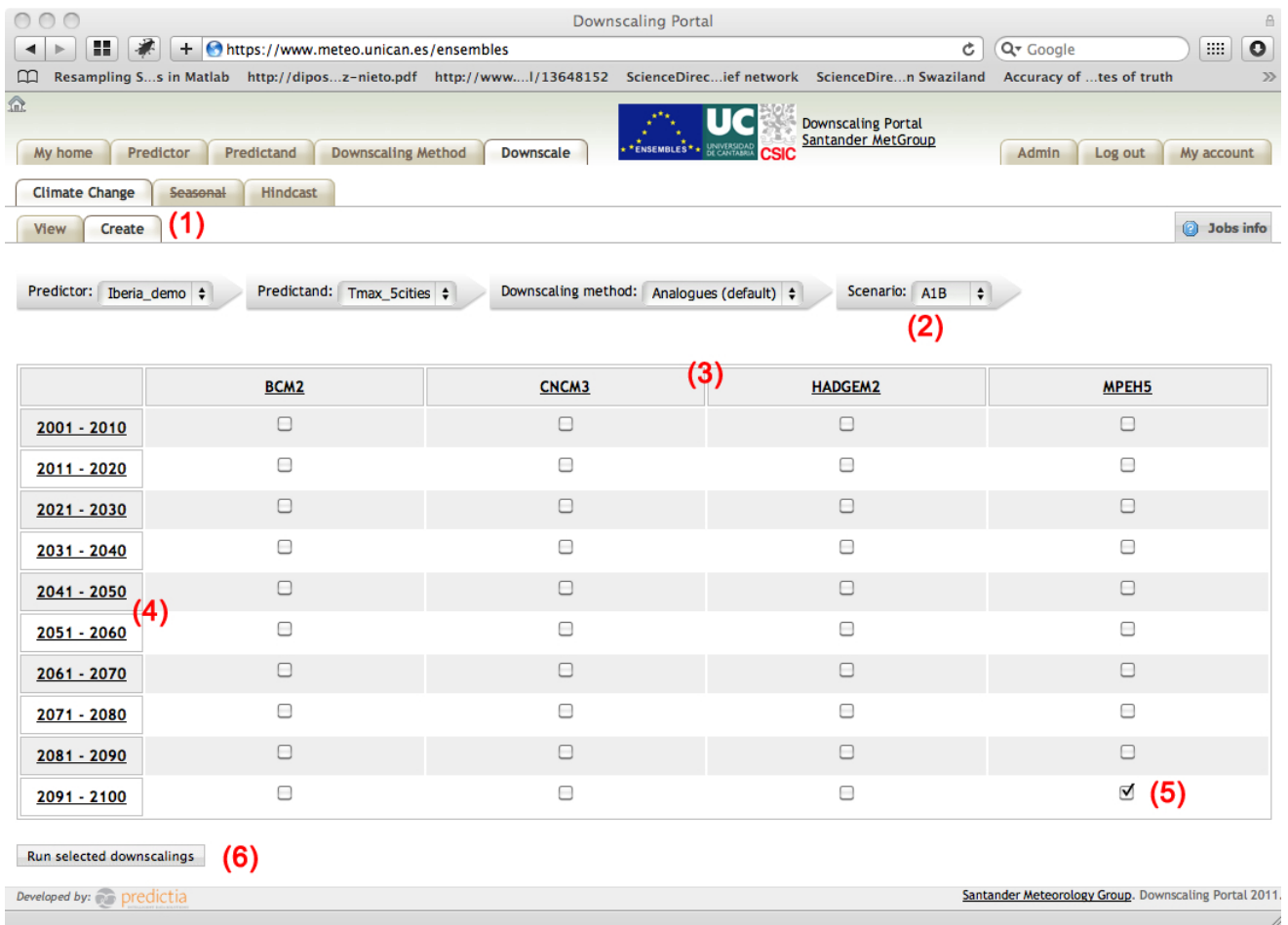


Figure 9: Downscale “create” window to apply downscaling methods using GCM scenario data for particular time slices.



Figure 10: Cells become green once the downscaling is finished.

selection, the window shows a downscaling matrix including the possible combinations of GCMs with available data⁸ (in columns) —as shown in (3)— and the corresponding time periods with available simulations (organized in rows, decade by decade) —as shown in (4).— In this case all the GCM simulations span the whole period of 10 decades but, in general, different models may have different simulated periods (e.g. the models downloaded from the IPCC database which include only certain time-slices, e.g. 2081-2100).

Each of the elements in the matrix (a decade for a particular GCM for a given scenario) is considered a **downscaling cell** and it is run by the portal as an **independent job**. Same criteria is applied in the “hindcast” tab. One or several of these cells (jobs) can be selected by clicking on them, as in

⁸Note that some of the variables included in the predictor definition may be missing for some of the GCMs, e.g. 1000 mb levels in the HADGEM2 model; in those cases, the GCM will not be available for downscaling for this predictor; note that this information is available when creating the predictor as shown in Fig. 4 (7).

Fig. 9(5) —note that by clicking on a decade label or on a GCM label, all the corresponding cells are automatically selected;— afterwards, the corresponding downscaling jobs can be submitted by clicking on the “run” button, as shown in (6); note that the portal will submit one job per cell, so the account’s restrictions will determine the maximum number of cells that can be selected/submitted simultaneously⁹. For instance, users with a basic profile (i.e. those not involved in the supporting projects or institutions) can only run two jobs simultaneously, which include the creation of predictor, predictand (with the basic downscaling method), or downscaling method, as well as the downscaling jobs. Therefore, downscaling the A1B scenario for the whole 2001-2100 period for a particular GCM would require five run steps (two decades each) in the portal (in case that the user is not run-

⁹See Fig. 3 for more information about your account’s restrictions; in particular you may consult the number of simultaneous jobs allowed for your account: *ConcurrentJobs*. These limitations have been considered to keep the downscaling jobs at a reasonable level of complexity, in terms of the memory needed and duration of the task.

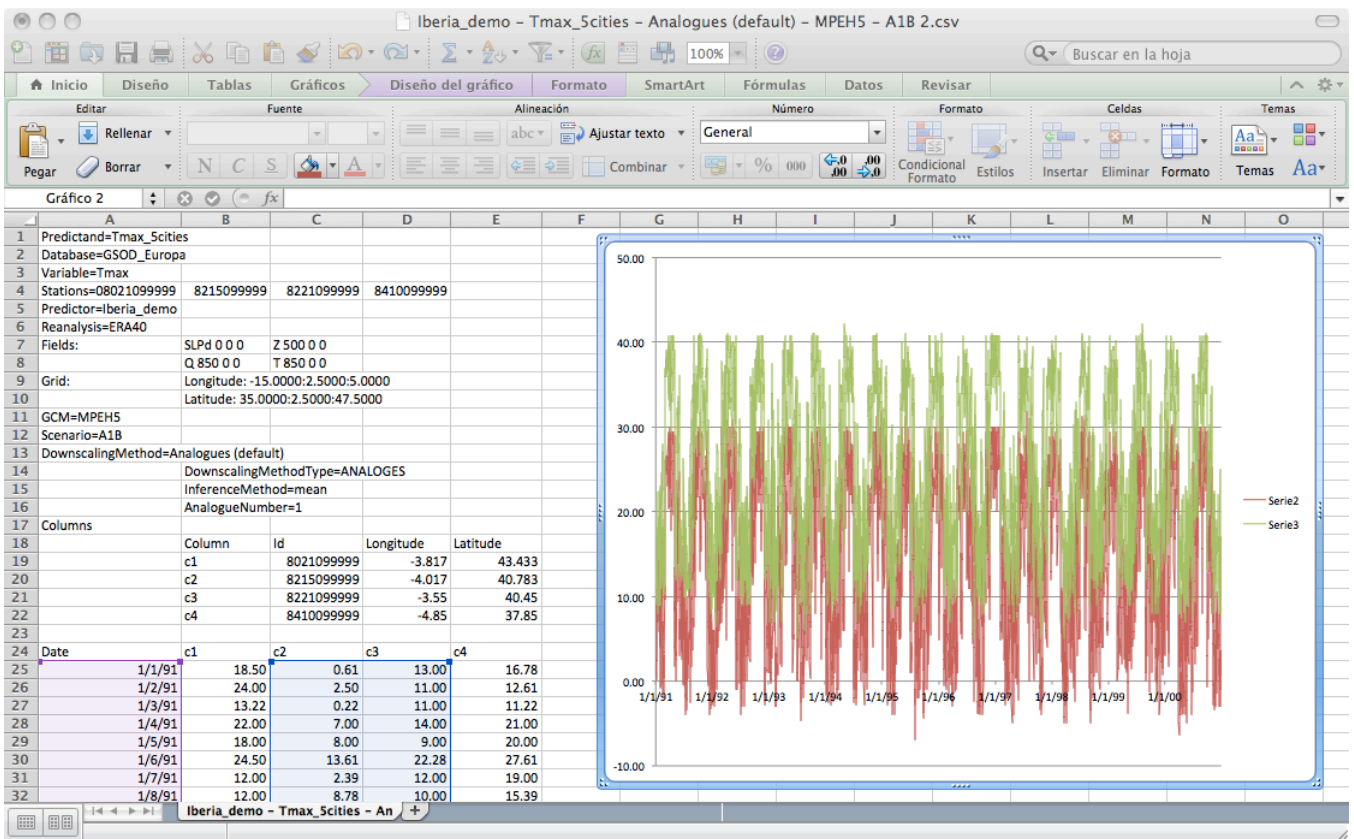


Figure 11: Downscaled projections can be downloaded in a .csv file and loaded in, e.g. Excel.

ning any other task). **We strongly advise** the users to first downscale, download and analyze a single decade before performing more exhaustive downscaling tasks, as we did in the *Iberia_demo* experiment (the 2091-2100 decade for the ECHAM5 A1B scenario).

The status of the jobs can be checked at any time in the “Jobs info” button (in the upper right corner of the window) or in the “My jobs” panel of “My home” window. A typical downscaling job will access the required data (the GCM scenario simulations and the reanalysis and observed data) and apply the downscaling method, producing the local projections for the defined locations/stations and period; this process takes typically some minutes and goes through different stages, which are indicated in the “Jobs info” panel: STARTING, RUNNING, etc., until the job finishes normally (FINISHED), or abnormally (ERROR). The different stages are also indicated with a background color in the corresponding downscaling cell: **yellow** for **STARTING** (i.e., the job is waiting at the execution queue), **blue** for **RUNNING** (i.e. the job is running at the cluster), **green** for **FINISHED** and **red** for **ERROR** (indicating some failure of the process). In this last case, we advise the user to wait a couple of hours and re-submit the job (in order to avoid possible spurious errors in the computing infrastructure) and, if the error persists, contact the portal development team using the email contact form included in the upper left corner of the portal.

The completed downscaled projections (downscaling matrix cells) can be consulted and downloaded through the

“View” panel. By clicking on the existing ones (those with a check box) the user can select those downscaling cells of interest and download them in a “.csv” file (“Download selected downscalings”). This file can be easily converted to a commonly used *Excel* ‘.xls’ in which daily predictions for all the stations (in columns) selected in the “Predictand” window are displayed in rows; note that the dates may not be consecutive and, therefore, you may need to sort the rows by the first column (the date) to obtain a chronological file. This allows the user to easily manipulate the data, drawing projected time series, etc. For instance, Fig. 11 shows the “csv” file downloaded with the projections of the 2091-2100 decade for the ECHAM5 model shown in Fig. 10. A graph of the daily temperatures for two out of the four stations (Madrid and Navacerrada) have been drawn by simply using the drawing facilities in Excel. The “.csv” file includes some header lines (the first 22 lines in Fig. 11) describing the predictors, the GCM and scenario, downscaling method and the predictands/stations (labelled as *c1*, *c2*, etc.) corresponding to the particular downscaling. The remaining rows correspond to the data, including the date in the first column, and the stations in the remaining ones, following the order *c1*, *c2*, etc. defined in the header. Note that the name of the file is also informative of the particular downscaling details (“*Iberia_demo - Tmax_5cities - Analogues (default) - MPEH5 - A1B.csv*” in this case).

8 Acknowledgments

The authors are grateful to the 6th FP EU project ENSEMBLES (GOCE-CT-2003-505539) for partial support for the development of the downscaling portal (see Linden et al., 2009, available at <http://ensembles-eu.metoffice.com>). The authors are also grateful to the 7th FP EU projects FUME (No. 243888), CLIM-RUN (No. 265192), METAFOR (No. 211753), QWECI (No. 243964), and to the MOSAICC initiative by FAO (<http://www.fao.org/climatechange/mosaicc>).

References

- Brands, S., S. Herrera, D. San-Martín, and J.M. Gutiérrez, 2011a: Validation of the ENSEMBLES Global Climate Models over southwestern Europe using probability density functions: A downscaler's perspective. *Climate Research*, 48, 145-161. DOI: 10.3354/cr00995 (open access). http://www.int-res.com/articles/cr_oa/c048p145.pdf
- Brands, S., J.M. Gutiérrez, S. Herrera, and A. Cofiño, 2012: On the Use of Reanalysis Data for Downscaling. *Journal of Climate*, 25, 2517-2526. <http://www.meteo.unican.es/node/73004>
- Gutiérrez, J.M., D. San-Martín, S. Brands, R. Manzanas and S. Herrera, 2012: Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, doi:10.1175/JCLI-D-11-00687.1. <http://www.meteo.unican.es/en/node/73058>
- Kalnay, E., et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77 (3), 437-471.
- Van der Linden P. and Mitchell, JFB, 2009: ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project. *Met Office Hadley Center*, FitzRoy Road, Exeter EX1 3PB, UK. 160pp.
- Perkins, S. E., Pitman, A. J., Holbrook, N. J. and McAneney, J., 2007: Evaluation of the AR4 climate models? simulated daily maximum temperature, minimum temperature and precipitation over Australia using Probability Density Functions, *Journal of Climate*, 20, 4356-4376, doi:10.1175/JCLI4253.1.
- Uppala, S., et al., 2005: The ERA-40 re-analysis. *Quarterly Journal of The Royal Meteorological Society*, 131 (612, Part B), 2961-3012, doi:{10.1256/qj.04.176}.
- Wilks, D., 2006: *Statistical methods in the atmospheric sciences*. Academic Press, 2 ed., Amsterdam, Elsevier.

9 Appendix 1: Validation scores

This section provides a more detailed description of the validation process generated in the downscaling portal. It makes an attempt to help the user to properly analyse the statistical scores calculated for the validation of the downscaling method applied.

Validation is performed at two different time-scales in the downscaling portal: daily and 10-daily aggregated data. Depending on the user's needs, both time-scales might be useful and the downscaling methods may show higher performance on the aggregated one, particularly for precipitation, being more informative for validation purposes. Note that additional validation scores are computed for precipitation, in order to take into account its dual (discrete/continuous) character. These scores will be identified with a "only for precipitation" label in the following description. Labels in bold correspond to the codes used in the downscaling portal (Sec. 6.1). All statistics are computed using the period defined for the particular experiment, so these scores (including descriptive ones), might change from experiment to experiment.

9.1 Descriptive Statistics

Basic descriptive statistics of observed (forecast) series.

- **RR:** *Rainfall Rate (only for precipitation). This score measures the frequency of wet days and it is calculated as the number of wet days divided by the size of the sample, n , expressed in %*

$$RR = \frac{n_{wet}}{n} 100 \quad (1)$$

The threshold considered for defining wet days is 0.1 mm.

- **Mean:** *Arithmetic mean. It measures the central tendency in a sample. It is calculated as the sum of all data points ($x_i, i = 1, \dots, n$) divided by the size of the sample, n .*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

The arithmetic mean is greatly influenced by outliers. For this reason, robust statistics such as the median may provide a better description of central tendency.

- **Median:** *Median. The median is also a measure of location. It is described as the value separating the higher half of the sample from the lower one (50th percentile). It can be found by arranging all the values from the lowest to the highest and picking the middle one. For data symmetrically-distributed, the mean and the median are the same.*
- **Min:** *Minimum. The smallest value in the series.*
- **Max:** *Maximum. The largest value in the series.*
- **Sigma:** *Standard Deviation (also denoted as Std). It shows how much variation or dispersion exists from the*

average. It is defined as the square root of the variance:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

The standard deviation is also greatly influenced by outliers. A useful property of the standard deviation is that, unlike variance, it is expressed in the same units as the data.

- **IQR:** *Interquartile range. It is a robust score that also estimates the dispersion in a sample, but it is not influenced by outliers. It is defined as the difference between the upper (75th percentile) and lower (25th percentile) quartiles, Q_3 and Q_1 respectively.*

$$IQR = Q_3 - Q_1 \quad (4)$$

The interquartile range is commonly used to build box-plots, simple graphical representations that shows with a box the spread of the data falling between the 25th and 75th percentiles.

- **PX:** *X^{th} percentile. Value below which $X\%$ of the data points are found. $X = 5, 10, 90, 95$.*
- **Missing:** *Percentage of missing values within the data: $[0, 100]$.*

9.2 Accuracy

Accuracy is one of the main aspects that must be examined when looking at the quality of a forecast since it measures the level of agreement between forecasts and observed time series. Note that some of the scores are presented in units of some descriptive statistic, what allows for direct comparison among stations and/or seasons, not worrying about their different regimes. In particular, those scores re-scaled by the Mean (Sigma) are named with a n (N) at the beginning of their names.

- **HIR:** *Hit Rate (only for precipitation). It is the probability of occurrences (o) (i.e. wet day) that were correctly forecast (f). This score ranges in $[0, 1]$ being 1 the perfect score.*

$$HIR = P(f = 1 | o = 1) \quad (5)$$

- **FAR:** *False Alarm Rate (only for precipitation). It is the probability of non-occurrences that were incorrectly forecasts. This score ranges in $[0, 1]$ being 0 the perfect score.*

$$FAR = P(f = 1 | o = 0) \quad (6)$$

Note that both scores, HIR and FAR, are only calculated in the portal for the case of the daily precipitation. They are not calculated for the 10-daily validation since aggregated data are considered to be continuous. HIR and FAR must be considered together in order to validate the discrete part precipitation. The threshold considered for defining wet days is 0.1 mm.

- **rho:** *Pearson's Product-Moment Correlation Coefficient. It measures the strength of the linear relationship*

between observations and forecasts. Ranged in $[-1,1]$. Perfect score: 1. The Pearson's correlation coefficient between two variables (x and y or observations (o) and forecasts (f) in our case) is defined as the covariance of the two variables ($Cov(x,y)$) divided by the product of their standard deviations.

$$\rho_{o,f} = \frac{Cov(o, f)}{\sigma_o \sigma_f} = \frac{\sum_{i=1}^n (o_i - \bar{o})(f_i - \bar{f})}{\sqrt{\sum_{i=1}^n (o_i - \bar{o})^2} \sqrt{\sum_{i=1}^n (f_i - \bar{f})^2}} \quad (7)$$

The Pearson's correlation coefficient shows how close the points of a scatter plot (observations against forecasts) are to a straight line. This score ranges in $[-1,1]$. A value of 1 (-1) implies that a linear equation describes the relationship between observations and forecasts perfectly. Thus, all the data points lies on a line indicating that forecasts increase (decrease) as observations increase. A value of 0 implies that there is no linear correlation between the variables. This score does not take bias into account, i.e., it is possible for a forecast with large errors to still have a good Pearson's correlation coefficient respect to the observations. This score is sensitive to outliers.

- **r:** Spearman's Rank Correlation Coefficient. This score measures the dependence, through some monotonic function, between observations and forecasts. The Spearman's correlation coefficient is defined as the Pearson's correlation coefficient considering the ranked variables. The sign of this score shows the direction of association between observations and forecasts. A positive (negative) coefficient indicates that forecasts tend to increase (decrease) as observations increase. Its magnitude increases as observations and forecasts become closer to being perfect monotone functions of each other.

It ranges in $[-1,1]$. A Spearman's correlation coefficient of 1 (-1) results when observations and forecasts keep a perfect monotone relationship, even if their relationship is not linear. Note, that this does not yield to a perfect Pearson's correlation. The Spearman's correlation coefficient is less sensitive than the Pearson's one to outliers that may be in the tails of both observations and/or predictions. This score should be used when validating precipitation rather than the Pearson correlation coefficient.

- **MAE:** Mean Absolute Error. It is an average of the forecast absolute errors. This score ranges in $[0,\infty)$ being 0 the perfect score

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - o_i| \quad (8)$$

- **nMAE:** Mean Absolute Error (MAE), in units of the observed mean. It ranges in $[0,\infty)$. Perfect score: 0.

$$\frac{MAE}{\bar{o}} \quad (9)$$

It has a singularity at $\bar{o} = 0$ (could occur for temperatures, for instance).

- **NMAE:** Mean Absolute Error (MAE), in units of the observed standard deviation. Ranged in $[0,\infty)$. Perfect

score: 0.

$$\frac{MAE}{\sigma_o} \quad (10)$$

It has a singularity at $\sigma_o = 0$, but this is not a realistic situation.

- **RMSE:** Root Mean Square Error. It measures the average magnitude of the forecast errors, weighted according to the square of the error. This score ranges in $[0,\infty)$. Perfect score: 0.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2} \quad (11)$$

The Root Mean Square Error shows high influence on large errors than on smaller ones, which may be appropriate if large errors are especially undesirable. However it may also encourage conservative forecasting.

- **nRMSE:** Root Mean Square Error (RMSE), in units of the observed mean. Ranged in $[0,\infty)$. Perfect score: 0.

$$\frac{RMSE}{\bar{o}} \quad (12)$$

- **NRMSE:** Root Mean Square Error (RMSE), in units of the observed standard deviation. Ranged in $[0,\infty)$. Perfect score: 0.

$$\frac{RMSE}{\sigma_o} \quad (13)$$

9.3 Distributional Similarity

The analysis of the distributional similarity is also a characteristic that describes the quality of a forecast/simulation, particularly at temporal scales where no serial correspondence between observations and predictions is required (e.g. for climate change projections). Thus, these scores measure similarity in climatological terms. Note that distributional similarity must be carefully examined, specially for climate change studies. These are the scores shown by the portal.

- **Ratio:** Ratio of wet days (only for precipitation). Ratio between forecasted and observed frequencies of wet days. It ranges in $[0,\infty)$. Perfect score: 1.

$$Ratio = \frac{P(f = 1)}{P(o = 1)} \quad (14)$$

The threshold considered for defining wet days is 0.1 mm. It presents a singularity when $P(o = 1) = 0$ (no rain occurs).

- **Bias:** Additive Bias. This score measures the average forecast error. It ranges in $(-\infty, +\infty)$. Perfect score: 0.

$$Bias = \frac{1}{n} \sum_{i=1}^n (f_i - o_i) \quad (15)$$

It does not measure the punctual correspondence between forecasts and observations, i.e., it is possible to get a perfect score for a bad forecast if errors are compensated.

- **NBias:** Bias, in units of the observed standard deviation. It ranges in $[0, \infty)$. Perfect score: 0.

$$\frac{Bias}{\sigma_o} \quad (16)$$

It has a singularity at $\sigma_o = 0$.

- **RV:** Ratio of Variances. This score measures the ratio between forecast and observed variances, in units of the observed one. It ranges in $[0, \infty)$. Perfect score: 1.

$$RV = \frac{\sigma_f^2}{\sigma_o^2} \quad (17)$$

It has a singularity at $\sigma_o = 0$.

- **KS-pValue:** pValue from the two-sample Kolmogorov-Smirnov test. This score ranges in $[0, 1]$. The null hypothesis of equality of distributions is rejected when the significance level equals or exceeds this pValue. The Kolmogorov - Smirnov test for two samples of sizes n and n' measures a distance, $D_{n,n'}$, between both cumulative density functions. $D_{n,n'}$ is calculated as:

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)| \quad (18)$$

where $F_{1,n}$ and $F_{2,n'}$ are the empirical cumulative distribution functions of the first and second sample, respectively. This test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions. Therefore, it is a *must* to consider this score for validation, especially when projecting under climate change scenarios.

- **KSX-pVvalue:** pValue from the two-sample Kolmogorov-Smirnov test restricted to observations and forecasts under their respective X^{th} percentiles (here $X = 10, 90$). It ranges in $[0, 1]$. The null hypothesis of equality of distributions is rejected when the significance level equals or exceeds this pValue.
- **PDF Score:** The PDF Score measures the overlap between observed and forecasted empirical probability density functions. It ranges in $[0, 1]$. Perfect score: 1. This score is calculated as in Perkins and McAneney (2007):

$$PDFScore = \sum_{i=1}^{200} (PDF_{f_i} - PDF_{o_i})$$

being PDF_{f_i} the forecast probability density for the i^{th} bin and PDF_{o_i} the observed probability density both for the i^{th} bin. 200 discrete bins (classes) are defined for the whole range of observations and predictions. Then, the probability density for each class is estimated by Kernel Density Smoothing. Observed and forecast probability densities are then compared for each class, retaining each pair minimum. The resulting sample of minima is finally summed up.

In the portal, Gaussian Kernels and a width parameter optimized for normal distributions are considered to estimate the probability densities. Therefore, the user must be aware that this score is more appropriate for validating temperature than for precipitation (see Brands et al., 2012, for a critical analysis of this score).

In addition, the PDF Score is hardly sensitive to failures in the tails of the distributions. Thus, the user should not rely exclusively on this score for validation, especially when projecting under climate change scenarios. We strongly recommend to consider both KS and PDF scores in conjunction.

Note that, for the special case of daily precipitation, and due to the high mass of probability density located at zero, the *KS-pValue*, *KSX-pVvalue* and the *PDF Score* are calculated for the continuous part of the distributions, by considering exclusively the observed and forecasted wet days. The discrete occurrence/non occurrence event is validated through the above explained HIR, FAR and Ratio scores. The latter scores are calculated over the entire observed and forecasted series for the 10-daily precipitation and temperature at both time-scales.