The ENSEMBLES Statistical Downscaling Portal. 
An End-to-End Tool for Regional Impact Studies

D. San-Martín a A.S. Cofiño a S. Herrera a J.M. Gutiérrez b

a University of Cantabria. Santander, Spain
b Instituto de Física de Cantabria. CSIC. Santander, Spain

Abstract
The demand for high-resolution seasonal and climate change forecasts is continuously increasing in a variety of socio-economic impact sectors, including agriculture, energy, health, and insurance. To fill the gap between the coarse-resolution outputs available from Global Circulation Models (GCMs) and the regional needs of the impact applications used in the above sectors, a number of statistical downscaling techniques have been developed. Statistical downscaling is nowadays a mature and 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 predictions.

The Statistical Downscaling Portal (SD Portal) described in this paper has been designed following an end-to-end approach in order to transparently connect data providers and end users. To this aim, Internet and distributed computing technologies have been combined together with statistical tools to directly downscale GCM outputs to the regional or local scale required by impact applications. Thus, users can test and validate online different methods (regression, neural networks, analogs, weather typing, etc.) using a Web browser, not worrying about the details of the techniques used or the different formats of the data accessed. The portal is part of the ENSEMBLES EU-funded project.

Key words: climate change scenarios, seasonal forecast, data mining, statistical downscaling, GRID computing, Web tools

1 Introduction

The Statistical Downscaling Portal (SD Portal) has been developed as part of the EU-funded ENSEMBLES project with the aim of maximizing the exploitation of the multi-model seasonal and climate change ensemble predictions produced by different modelling centres [for more details see http://www.ensembles-eu.org and Hewitt and Griggs, 2004]. These predictions, which are mainly based on Global Circulation Models (GCMs), are needed by end-users from different socio-economic sectors such as agriculture, energy or health, in order to run their impact models with appropriate climatic input [see, e.g., Thomson et al., 2006]. The practical interest of seasonal to interannual predictions lies in their potential economic benefits in planning the future (e.g., adopting protection or adaptation measures) according to the forecasted events (e.g., a drought warning).

The main limitation for the application of these predictions in impact studies is the coarse spatial resolution of GCMs (for instance, the models used in the IPCC fourth assessment report had a typical resolution of 110 km). This clearly contrasts with the regional or local meteorological inputs needed by the impact applications run by end-users (crop yield models, energy demand models, etc.). For instance, Fig. 1(a) shows the land-sea mask used by the ECHAM5/MPI-OM model, with a horizontal resolution of 200 km (T63); on the other hand, Fig. 1(b) shows the maximum daily surface temperature observed in Spain for the period 1960-2000, which varies on a much more local scale. Panels (c)-(f) show the histograms of the daily maximum temperatures in four different locations for a five years period (1995-2000): Navacerrada, Madrid, Córdoba and Barcelona, respectively, revealing a typical summer-winter bimodality in all cases, but with different mean values (panels c-e), and variability (panel f). This high-resolution spatial and temporal information is required by end-users.

To fill this gap a number of dynamical [see, e.g. Christensen et al., 2007, and other papers in the same special issue] and statistical [see Wilby and Wigley, 1997, Zorita and von Storch, 1999, for an overview] downscaling techniques have been developed. Statistical tech-
Fig. 1. (a) Land-ocean mask of a T63 model (200 km horizontal resolution); the box shows the Iberian peninsula. (b) Maximum temperature observations interpolated in a high-resolution grid (20 km) in Spain for 1960-2000 (daily averages of the whole period are shown). (c)-(f) Histograms of the daily maximum temperatures in four different locations (indicated in panel b) for the period 1995-2000.

The SD Portal described in this paper has been designed and developed following and end-to-end approach to link end-users to data providers by facilitating the downscaling task through a user-friendly Web portal. In this form, users can easily upload their observation grids or networks with the variables of interest (e.g., evapotranspiration, rainfall and temperature for crop-yield models) and downscale different seasonal and climate change model predictions, testing and validating online a range of statistical downscaling methods. This is possible due to the use of distributed data access and computation technologies such as GRID [Foster and Kesselman, 2007] operating behind the portal and allowing distributed resources to operate collectively. A registered user is required in order to log into the portal with full access to data and computational resources. However, a “guest” account with limited functionality is available to test the portal (www.meteo.unican.es/ensembles).

This paper is organized as follows. In Sec. 2 we give a brief overview of the ENSEMBLES project and describe the datasets available in the portal for downscaling purposes. In Sec. 3 we analyze the downscaling problem and introduce the main techniques available for this task. In Sec. 4 we introduce the portal and its main components. In Sec. 5 we illustrate the use of the portal in a illustrative example, considering maximum temperature in Spain. Finally some conclusions and future work are presented in Sec. 6.

2 The ENSEMBLES Project

The SD Portal is being developed as a part of the EU-funded ENSEMBLES project, run from 2004-2009 (for more details see Hewitt and Griggs [2004] and www.ensembles-eu.org). The goal of this project is to develop an ensemble prediction system based on the principal state-of-the-art GCMs developed in Europe, validated against quality controlled, high resolution gridded datasets, to produce an objective probabilistic estimate of uncertainty related to model inadequacy and future forcing scenarios, at the seasonal to decadal and longer timescales [see Troccoli et al., 2008, Mechl et al., 2007, and references therein.

et al., 2007]. 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. This step may be very time-consuming for end-users, since it requires working with different meteorological formats/technologies and with statistical downscaling procedures. Therefore, the apparently simple task of obtaining a simulated data field (e.g. surface maximum temperature) at the required temporal aggregation (e.g. daily) downscaled to a specific location (e.g. to Madrid, Spain) may result very time-consuming for some users.
<table>
<thead>
<tr>
<th>Variable (Code)</th>
<th>Levels</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geopotential (Z)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>V velocity (V)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>U velocity (U)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>Temperature (T)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>Relative humidity (R)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>Specific humidity (Q)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>Potential Vorticity (PV)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>Relative Vorticity (VO)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>Divergence (D)</td>
<td>yes</td>
<td>00,12</td>
</tr>
<tr>
<td>MSLP (MSL)</td>
<td>no</td>
<td>00,12</td>
</tr>
<tr>
<td>2m Temperature (2T)</td>
<td>no</td>
<td>00,12</td>
</tr>
<tr>
<td>10m E-Wind Component (10U)</td>
<td>no</td>
<td>00,12</td>
</tr>
<tr>
<td>10m N-Wind Component (10V)</td>
<td>no</td>
<td>00,12</td>
</tr>
<tr>
<td>Total Precipitation (TP)</td>
<td>no</td>
<td>[00,24]</td>
</tr>
<tr>
<td>Minimum Temperature (Tn)</td>
<td>no</td>
<td>[00,24]</td>
</tr>
<tr>
<td>Maximum Temperature (Tx)</td>
<td>no</td>
<td>[00,24]</td>
</tr>
</tbody>
</table>

Table 1

Description of the variables, height levels and times (UTC) of the common set of parameters used in the portal. Variables labeled by levels stand for 300, 500, 700, 850, 925, 1000 mb where the remaining are only surface 2D outputs. Time intervals [0, 24] refer to accumulated, maximum or minimum values, whereas times 00, 12 refer to instantaneous values.

for an updated status of ensemble prediction at seasonal and climate scales. As a result, this project will produce huge amounts of model outputs to be stored in central repositories: the CERA system in the Max-Plank Institute for climate change simulations, [www.mad.zmaw.de/projects-at-md/ensembles](http://www.mad.zmaw.de/projects-at-md/ensembles) and the ECMWF archiving system for seasonal to decadal predictions, [www.ecmwf.int/research/ensembles](http://www.ecmwf.int/research/ensembles). As we shall see later, this data can be combined with reanalysis databases and historical records to run downscaling experiments from the web portal.

2.1 Predictor Datasets

In order to manage a homogeneous basic set of parameters in the portal for the different GCM outputs (reanalysis, seasonal forecast and climate change predictions), we have considered initially a small dataset of commonly-used predictor variables at a daily basis (see Table 1). Since the main area of interest of the ENSEMBLES project is Europe, these variables have been locally stored for the European region shown in Fig. 1(a).

In particular, we have downloaded and stored data for three different reanalysis projects: NCEP/NCAR ReanalysisI 1948-2007 ([www.cdc.noaa.gov](http://www.cdc.noaa.gov)), ERA40 ECMWF 1957-2002 ([www.ecmwf.int](http://www.ecmwf.int)), and JRA25 Japanese Reanalysis 1979-2004 ([jra.kishou.go.jp](http://jra.kishou.go.jp)).

In the case of seasonal forecast, we have also downloaded and stored data from the DEMETER project [Palmer et al., 2004], a multi-model seasonal prediction experiment including seven models ran for six months four times a year in the ERA40 period using 9 different perturbed initial conditions (9 members). Moreover, data from the ENSEMBLES Stream 1 seasonal forecast (the DEMETER follow on project) has been stored and downloaded for a narrower region (Spain), but there is work in progress to connect the portal to the remote server at ECMWF using OPeNDAP technology (see [www.ecmwf.int](http://www.ecmwf.int) for more details). The goal is to set up this connection for the Stream 2 simulations, which will provide a new dataset of seasonal hindcasts for 1960-2005, including seasonal runs (7 months, four start dates per year for Feb, May, Aug, Nov), and annual runs (14 months, with at least one start date per year: Nov). For more details we refer the reader to [www.ecmwf.int/research/EU_projects/ENSEMBLES](http://www.ecmwf.int/research/EU_projects/ENSEMBLES).

The Anthropogenic Climate Change (ACC) simulations have been obtained from different sources. On the one hand the IPCC data centre ([www.ipcc-data.org](http://www.ipcc-data.org)) provides access to the models contributing the AR4; these datasets are referred to as PCMDI-Model. On the other hand, the CERA system provides datasets with a common list of parameters for different ACC experiments performed in the ENSEMBLES project; these datasets are referred to as CERA-Model. Finally some data is directly obtained from local providers, referred to as Model. So far, the following data is available, but the portal provides updated information of new datasets included, as they become available:

- **PCMDI-CGCM3**: Canadian Centre for Climate Modelling and Analysis, including 20th century (from 1951 to 2000) and scenarios A1B, B1 (daily data in pressure levels is only included for the periods 2046-2065 and 2081-2100).

The portal focuses very much on ENSEMBLES results and, hence, in the near future, the SD Portal will provide access to the data of the seven GCMs (BCM2, CNRM-CM3, ECHAM5_MPI-OM, EGMAM, HadCM3, HadGEM1 and IPSL-CM4) which form the ENSEMBLES’ ACC Stream 1 and 2, run using five different forcings: multicentennial control forcing, historical forcing to 2000, and the B1, A1B and A2 SRES scenarios to 2100.

2.2 Predictands, Historical Observed Records

The ENSEMBLES project also provides different daily historical records which can be used in downscaling experiments, including both raw observations from local stations and interpolated high-resolution grids:
• ECA (European Climate Assessment Dataset project). Daily datasets of precipitation, temperature, pressure, humidity, cloud cover, sunshine and snow depth records since 1900 over networks of 100-1000 stations.
• Ensembles 50km gridded daily observation records of precipitation and surface temperature for the period 1950-2006 (see eca.knmi.nl).

3 The Statistical Downscaling (SD) Approach

Different statistical methods have been proposed 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: First an empirical relationship (a statistical model) is established between the large-scale GCM variables (predictors) and the small-scale observed parameters of interest (predictands) using a historical common period (usually a reanalysis period). Then, the resulting statistical model is applied to future GCM predictions to obtain the estimated local forecast. Usually, the different statistical downscaling methodologies are broadly categorized into three classes [Benestad et al., 2008]:

• **Regression or Transfer functions**, based on linear 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 predictions are derived from a model obtained from data.
• **Weather typing**, based on a pre-classification 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.
• **Weather generators**, which stochastically simulate daily climate values based on the available monthly average predictions. These techniques are temporal disaggregation methods and they have not been included yet in the portal.

The SD Portal includes different techniques from the first two categories (see Fig. 2), thus allowing to test and compare the performance of these approaches for different variables and regions (note that the skill of SD methods varies from variable to variable and from region to region [Schmidli et al., 2007]). Fig. 2 shows the selection panel of the SD Portal, with the default configuration corresponding to an analog downscaling method from the weather-typing category using the mean of 25 neighbors to estimate the forecast (we shall refer to this method as SD1 through this paper). Other alternatives in this category include a two-step analog algorithm and different versions of weather-typing approaches: k-means or Self-Organizing Maps [see Gutiérrez et al., 2004, 2005, for more details]. On the other hand, the portal includes different algorithms from the regression category, including linear regression models and nonlinear neural networks; in both cases the user can choose between PCs of the synoptic fields, or nearest grid-point data, as predictor variables; in the case of neural networks, a feedforward model with two hidden layers (with 5 and 3 neurons, respectively) is trained 10 different times selecting the models with lower test error [see Castillo et al., 1999, for more details on neural networks]. We shall refer to this method as SD2 through this paper.

![Fig. 2. Panel to select a statistical downscaling method from the categories “weather typing” and “regression.”](image)

It is important to remark that there are a number of general recommendations which should be followed in order to get consistent and reliable results from the statistical downscaling process. For instance, the variables used in the downscaling process should ideally be primary model variables, not based on parameterisations [see Wilby et al., 2004, or the ENSEMBLES regional scenario web portal www.cru.uea.ac.uk/projects/ensembles/ScenariosPortal/ for more information].

4 The Statistical Downscaling Portal

The SD Portal has been organized in different windows to gradually access the information to define a downscaling task. The first window, **My home** tab, is the user’s main window and provides information about the existing downscaling experiments, the user’s account profile and the status of the submitted jobs (see Fig. 3). The
Experiment manager panel shows the details of the experiments already created by the user (a unique experiment, “SpainTZ”, in this case), each including a set of predictors defined in a particular region from a reanalysis project (T and Z from ERA40 in Spain) and one, or several, predictands (maximum temperature in four different stations in Spain from the ECA network). Each predictand also displays the information of the downscaling methods which have been defined and applied with this data. The user can browse the information and navigate through the portal by clicking in the different components.

The Profile panel displays the databases available for the current user (for instance, the reanalysis available for the ensemblesfp6 user are shown in Fig. 3), as well as the account information, restrictions (maximum number of simultaneous jobs, maximum size of data request, etc.) and areas of study allowed. Finally, the Jobs panel allows monitoring the status (running, finished, etc.) and type (observations, downscaling, etc.) of the jobs, and also allows downloading the files produced as a result of each downscaling task. Each request is handled by the portal as an independent job, so several requests can be handled and monitored simultaneously.

Each downscaling experiment encodes all the information needed for the downscaling process and can defined in three sequential steps:

1. Definition of the region and predictors to be used.
2. Definition of the predictands of interest.
3. Definition, validation and application of the downscaling method.

Each of this steps can be performed by selecting, in the appropriate order, the corresponding tabs in the application: “predictor”, “predictand”, and “downscaling”. The details are explained in the following sections.

4.1 Precitor: Selecting the region and the predictors

The first step to define a downscaling experiment is to select the region of interest and the desired predictors that
shall be used to fit the downscaling methods. To this aim, the portal allows the user to visually select a lattice with the desired resolution over a geographical area of interest and to include the desired variables from the reanalysis to be used as predictors (4D cubes of reanalysis information). This process is carried out by clicking and dragging in the “predictors” window (see Fig. 4) and entering the information such as region, lattice resolution, variables, etc.

Once the zone and predictors have been defined, several internal processes are computed to obtain statistical information needed at a later stage of the downscaling process (principal components, clustering, etc.). This information is stored in the portal and it can be managed from the experiment manager panel in the My Home window.

4.2 Predictand: Selecting the stations and variable

Once the region of interest and the predictors have been selected, the user can move to the second window (“predictand” tab) and select the stations where local forecasts are to be computed. This process can also be done visually by first selecting the network (ECA in this case), then the variable of interest (maximum temperature) and, finally, adding (or removing) stations until the desired set is selected (in red). For instance, Fig. 5 shows a selection of maximum temperature for the four stations shown in Fig. 1(b).

The user can also upload private data to the portal to be used in the downscaling process.

4.3 Downscaling: Definition and Validation

After selecting the predictors and predictand over the region of interest, the portal allows the user to choose among different downscaling methodologies to create one, or several, downscaling algorithms (see Sec. 3). Moreover, these methods can be validated in a Perfect Prognosis scenario using reanalysis data both for training and testing the method. In this case, the reanalysis period is split in two parts, one for training and one for validating the method, so no overlap is produced (see Fig. 6).

For instance, if we consider the two statistical downscaling methods described in Sec. 3 (an analog method, SD1, and a neural network model, SD2), we can use the portal validation tool to obtain the Root Mean Squared Error (RMSE) of the downscaled values for the four stations of interest. Fig. 7 shows the obtained results; in the left column the observations for the test period are plotted against the corresponding predictions obtained with the SD1 method. In the right column the two downscaling methods are plotted one against the other for the test period. It can be shown that the performance is satisfactory in the four stations considered and both methods have comparable skill. However, both methods seem to overestimate the temperature of Madrid and Córdoba in the “cold” tail (Winter period) and, hence, future results of climate change need to be carefully analyzed and validated in these stations. This validation process allows us to estimate the uncertainty or error attributable to the statistical downscaling technique.
Note that this example is only intended to be an illustrative application of the portal, so we are not interested in developing the optimum downscaling approach but in describing and analyzing the capabilities of this tool; for instance, the above overestimation problem could be addressed by developing different statistical downscaling methods for each of the seasons, thus taking into account this non-stationarity into the downscaling process.

![Graphs](image-url)

Fig. 7. Observations and predictions of the analog (SD1) and neural network (SD2) downscaling methods applied to (a)-(b) Navacerrada, (c)-(d) Madrid, (e)-(f) Córdoba and (g)-(h) Barcelona the period 1995–2000. The numbers within the figure in the right panel indicate the mean Root Mean Square Error (RMSE) of each of the methods.

Once the performance of the methods has been tested, they can be applied to different seasonal or climate change experiments selecting the desired models, years and months/seasons of interest from a matrix containing all the possible combinations to downscale model outputs to local stations (see Fig. 8 for an example with the MPI-ECHAM5 model). This matrix illustrates the complexity of this problem, since each box is a possible downscaling job. For instance, Fig. 8 indicates that there is a single scenario, 20th Century, available to downscale within the year 2000, but there are four different scenarios to downscale in 2001, corresponding to the 20th Century, A1B, A2, and B1 emission scenarios, respectively. A particular downscaling task can be performed by selecting some of these boxes and, then, clicking the “run downscaling” button to submit the corresponding job. The color of the boxes changes from blue to yellow and finally to green when the downscaling is sent, running and finished, respectively, thus allowing to monitor the downscaling process.

Therefore, an efficient design of the computational load is required in order to develop an interactive portal where users can run several jobs simultaneously. The implementation adopted in the portal is described in the following section.

### 4.4 Implementation

The portal has been designed and implemented following an internet-based approach for distributed data-access and computing using Java technology. Fig. 9 shows the design of the portal, which operates using the local cluster using Matlab software (the current kernel of the portal is the open-source MeteoLab toolbox for Matlab; see [www.meteo.unican.es/meteolab](http://www.meteo.unican.es/meteolab)). In addition to the local data, the portal is also prepared to access distributed datasets through OPeNDAP protocol to remote storage servers (e.g., the ECMWF ENSEMBLES' server for seasonal to decadal predictions). OPeNDAP technology ([www.opendap.org](http://www.opendap.org)) allows exposing scientific datasets in the Web (mainly global model outputs) and subsetting this datasets using HTTP protocol. When the necessary information is ready to run a downscaling job requested by a user, the portal send the data and a Matlab script to the local server queue. A monitoring system has been developed and deployed in the portal server, so users can check and control the current state of the submitted jobs.

### 5 Example. Downscaling Tmax in Spain

In this section we describe the application of the SD Portal to an illustrative example, considering the output of the MPI-ECHAM5 model to obtain downscaled values for four locations in Spain exhibiting different climatology (see Fig. 1). To this aim we first applied the portal to downscale the 20th century scenario from 1950–2000 using SD1 and SD2, comparing the results with those obtained from the direct model output and from the observed climatology.
On the other hand, the green curves show the Winter (DJF) and Summer (JJA) PDFs obtained by directly interpolating the MPI-ECHAM5 model outputs to each of the locations. Comparing the results with those corresponding to the observations it can be easily seen that the model exhibits a clear bias and, moreover, it does not account for the different variances shown in both stations (the model output is closer to the climatology of Madrid). The red and light-blue curves show the PDFs corresponding to the statistical downscaling methods in perfect model conditions (i.e., using reanalysis data as input). If we compare these curves with the corresponding green ones (direct model output) we can easily check that the statistical downscaling process overcomes the limitations of the model output, allowing to accurately recover the observed climatology and accounting for the different means and variances shown in both stations. Note that, as previously observed in the validation step (Fig. 7), the statistical downscaling methods used in this example fail to properly reproduce the “cold” Winter tail of the climatology in Madrid and, thus, special care must be taken to analyze the results corresponding to future climate change projections.

Finally, the pink and yellow curves show the PDFs obtained by downscaling the MPI-ECHAM5 model (20th Century scenario). These curves are very close to the ones obtained in perfect-prognosis conditions, thus, indicating a good performance of the climate change model to reproduce the climate of this period. Finally, note
that the Gaussian PDFs have been considered for the sake of visual simplicity; however, in a practical application, the daily downscaled values can be directly used in a practical application, without imposing any special distribution to the data.

that the Gaussian PDFs have been considered for the sake of visual simplicity; however, in a practical application, the daily downscaled values can be directly used in a practical application, without imposing any special distribution to the data.

$$\frac{\mu, \sigma}{F_{\text{req}} (T)}$$

$$\frac{\mu, \sigma}{F_{\text{req}} (T)}$$

$$\frac{\mu, \sigma}{F_{\text{req}} (T)}$$

<table>
<thead>
<tr>
<th></th>
<th>( S_\mu )</th>
<th>( S_\sigma )</th>
<th>( W_\mu )</th>
<th>( W_\sigma )</th>
<th>( S_\mu )</th>
<th>( S_\sigma )</th>
<th>( W_\mu )</th>
<th>( W_\sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBS</td>
<td>20.79</td>
<td>4.75</td>
<td>3.25</td>
<td>4.52</td>
<td>30.12</td>
<td>3.96</td>
<td>11.67</td>
<td>3.32</td>
</tr>
<tr>
<td>MPECH5</td>
<td>26.45</td>
<td>4.30</td>
<td>7.25</td>
<td>3.15</td>
<td>27.20</td>
<td>4.22</td>
<td>7.04</td>
<td>3.25</td>
</tr>
<tr>
<td>SD1 perf</td>
<td>19.78</td>
<td>4.29</td>
<td>3.57</td>
<td>3.75</td>
<td>29.40</td>
<td>3.79</td>
<td>12.10</td>
<td>2.55</td>
</tr>
<tr>
<td>SD2 perf</td>
<td>19.19</td>
<td>4.35</td>
<td>2.84</td>
<td>3.53</td>
<td>29.19</td>
<td>3.61</td>
<td>11.79</td>
<td>2.69</td>
</tr>
<tr>
<td>SD1 mpeh</td>
<td>19.33</td>
<td>5.06</td>
<td>3.16</td>
<td>4.25</td>
<td>29.07</td>
<td>4.41</td>
<td>12.16</td>
<td>2.72</td>
</tr>
<tr>
<td>SD2 mpeh</td>
<td>18.82</td>
<td>5.09</td>
<td>2.26</td>
<td>3.71</td>
<td>28.83</td>
<td>4.27</td>
<td>11.28</td>
<td>2.52</td>
</tr>
</tbody>
</table>

Table 2

Summer \((S)\) and Winter \((W)\) mean \((\mu)\) and standard deviation \((\sigma)\) values for Navacerrada (left) and Madrid (right) corresponding to the observed records, the direct output of the MPEH5 model and two statistical downsampling methods with perfect reanalysis (perf) and simulated climate (mpeh) conditions, corresponding to the period 1995-2000.

Table 2 shows the mean and standard deviations of the distributions shown in Fig. 10. From this table it is also evident the benefits of statistical downsampling procedures for climate change regionalization. Moreover, none of the two statistical techniques used has shown to be better than the other since the best results vary from case to case.

5.2 Future Projections

The main limitation to apply the statistical downscaling approach to future simulations is the stationary assumption (the relationships between predictors and predictands do not vary under future climate conditions), which may be violated under climate change conditions. However, recent studies indicate that the statistical linkages seem to be robust and consistent when sound predictors with a physical basis are used to drive the local predictands in the models [see Frías et al., 2006, Timbal et al., 2008, and reference therein for more details]. In this downsampling experiment we have used temperature and geopotential fields to downscale surface maximum temperature and, hence, the predictor dataset seems to be robust for extrapolation to future climate conditions.

For instance, Figure 12 shows the Gaussian distributions fitted to the future maximum daily Winter temperatures in Navacerrada for different decades in the period 2000-2100, corresponding to the MPI-ECHAM5 model under scenario A1B. Fig. 12(a) shows the distributions corresponding to the interpolated direct model outputs, whereas panels (b) and (c) show the distributions of the downscaled values obtained using \( SD1 \) and \( SD2 \) methods, respectively. From these figures it can be easily shown that future projections are bias corrected and exhibit an increasing variance for future decades. The direct model outputs show an increasing mean but do not seem to capture this important trend in the variance.

Finally, Figure 12 shows the present climate and future decadal climatologies for the Winter maximum daily temperatures in the four stations shown in Fig. 1(b), considering the MPI-ECHAM5 model under scenario A1B. Decadal observed climatologies within the period 1960-2000 are also shown for the sake of comparison. These figures show the box-and-whiskers plots of the daily values for the corresponding periods, thus providing full information about the distribution quartiles, extremes, and outliers (in this case, the distributions are approximately Gaussian, but in general this type of plot is more informative than the one provided in Fig. 11). This figure illustrates the benefits of the statistical downsampling approach which is able to “calibrate” the model output for the different local climatologies, providing also a compatible estimate of the future projections. In some cases (e.g., panel a) the downsampling process leads to an
Fig. 11. Gaussian distributions fitted to the Winter daily maximum temperatures for different decades in Navacerrada for the (a) direct ECHAM5 model output (A1B scenario), (b) SD1 downscaling method and (c) SD2 downscaling method. Increasing variance, not shown in the direct model output. In other cases, the increasing temperature trend is lower for the downscaled values than for the model outputs (e.g., panel b). Finally, in some cases (panel a) both downscaling methods exhibit some disagreement about the future trends, thus introducing an extra uncertainty factor in the analysis.

Finally, we want to remark that the present example has been only shown for illustrative purposes, with the aim of discussing different problems which may arise when using statistical downscaling methods to perform local projection of GCM outputs.

5.3 Extreme events

One of the advantages to work with daily data is the possibility to analyze indices associated with extreme or rare events. For instance, we have studied the trend of “hot” days in Summer and “cold” days in Winter in Navacerrada by considering the 90th (25.5 °C) and 10th (-3 °C) percentile values, respectively, of the respective seasonal reference periods in 1960-2000. The average frequency of hot/cold days is 10% during this period, and the future frequencies can be easily estimated by using the daily predicted or downscaled values analyzed in the previous section, and shown in Fig. 12(a). In each case, the threshold percentiles are computing considering the corresponding outputs or downscaled values in the 1960-2000 period. Figure 13 shows the frequency values obtained for Winter/Summer cold/hot events, indicating a clear decrement/increment in the frequencies of these events, respectively. For instance, the increasing trend of Summer hot events is lower in the downscaled values than in the direct model output. Note that, accord-
ing to Figs. 10 and 11 the tails of the distribution seem to be better represented by the downscaled values and, thus, this example may indicate an overestimation of the hot events given by the model outputs in this location, Navacerrada, which is located at 1890 meters a.s.l.

Fig. 13. Observed and predicted frequencies of (a) cold events in Winter and (b) hot events in Summer in Navacerrada (see text for more information).

6 Conclusions

In this paper we have presented a statistical downscaling portal to fill the gap between weather forecast modelers (data producers) and application developers (data consumers). This portal integrates datasets from GCM outputs and observations and uses statistical modeling tools to project model outputs to local observations. This portal allows to easily and quickly obtain regional seasonal predictions and climate change scenarios but, however, it should not be used as a black-box tool, since this could lead to unreliable outputs or inappropriate use of downscaled data.

Finally, a new version of the portal using GRID computing facilities is being designed [see Cofiño et al., 2007, for a preliminary study] in the framework of the 6th EU FP EELA project (see www.eu-eela.org) based on gLite middleware (see cern.ch/glite). GRID computing is a new paradigm for Internet-based distributed computing which enables the development of interactive problem-solving environments integrating the sharing, selection, and aggregation of geographically distributed autonomous resources, such as computers and databases Foster and Kesselman [2007].

Acknowledgments.

The authors are grateful to the 6th FP EU project ENSEMBLES (GOCE-CT-2003-505539) and to the CICYT project CGI2005-06966-C07-02/CLI for partial support of this work.

References

J. Schmidli, C.M. Goodess, C. Frei, M.R. Haylock,


