

A Web Portal for Regional Projection of Weather Forecast Using GRID Middleware

A.S. Cofiño, D. San-Martín, and J.M. Gutiérrez

Dept. of Applied Mathematics and Computer Science,
University of Cantabria. Santander, 39005, Spain

{antonio.cofino,daniel.sanmartin,manuel.gutierrez}@unican.es
<http://www.meteo.unican.es>

Abstract. Weather forecast is a complex multi-disciplinary problem which requires a cascade of different scientific tools, from differential equation solvers to high-dimensional statistical and data-mining algorithms. The demand for high-resolution predictions is continuously increasing due to the multiple applications in hydrology, agronomy, etc., which require regional meteorological inputs. To fill the gap between the coarse-resolution lattices used by global weather models and the regional needs of applications, a number of statistical downscaling techniques have been proposed. In this paper we describe a Web portal which integrates the necessary tools with Grid middleware allowing for distributed data access and computing. The portal is part of the ENSEMBLES EU-funded project and allows end users to interactively downscale weather predictions using a web browser. Both the architecture and the usage of the portal are described in this paper.

Keywords: Grid computing, weather forecast, statistical downscaling, problem-solving environments, data mining, Web portals.

1 Introduction

Weather forecast is a complex interdisciplinary problem which involves several scientific disciplines to model and simulate the atmosphere dynamics in order to predict its future evolution. During the last four decades, several atmospheric circulation models have been developed and successively improved to include new interactions with other main climate components: hydrosphere, cryosphere, lithosphere and biosphere. Nowadays, the resulting Global Circulation Models (GCMs) are numerically solved on coarse-grained resolution grids covering the world, providing a low-resolution description of the future state of the atmosphere (temperature, humidity, geopotential, wind components, etc.) for a particular future period of time (e.g., next week, month, or season). These models are also forced with different future green-house gas emission scenarios to obtain a future estimate of the climate change for the next century. The computational requirements of this problem have recently increased with the adoption of the “ensemble forecast” methodology to deal with uncertainty. Ensemble methods

compute a probabilistic prediction of the atmospheric variables, based on several integrations of the circulation models started from perturbed initial conditions, or perturbed models (see [1] for more details).

The main shortcoming of GCMs is that their spatial resolution is currently constrained by both computational and physical considerations to a scale of hundreds of kilometers; for instance, Fig. 1(a) shows the land-sea mask used by these models with a resolution of one degree (aprox. 100 km at middle latitudes). However, meteorological phenomena such as rainfall, vary on much more local scales, as it can be observed analyzing high-resolution observation databases. For instance, Fig. 1(b) shows the location of 2600 stations in Spain, with records of daily temperature and precipitation during the last 60 years (Fig. 1(c) shows the accumulated precipitation pattern for February 1945, exhibiting high spatial variability). This spatial and temporal high-resolution information is required in different application domains to feed impact models with meteorological information in order to characterize the future evolution of key socio-economic sectors, such as water resources, crop yields, and power consumption.

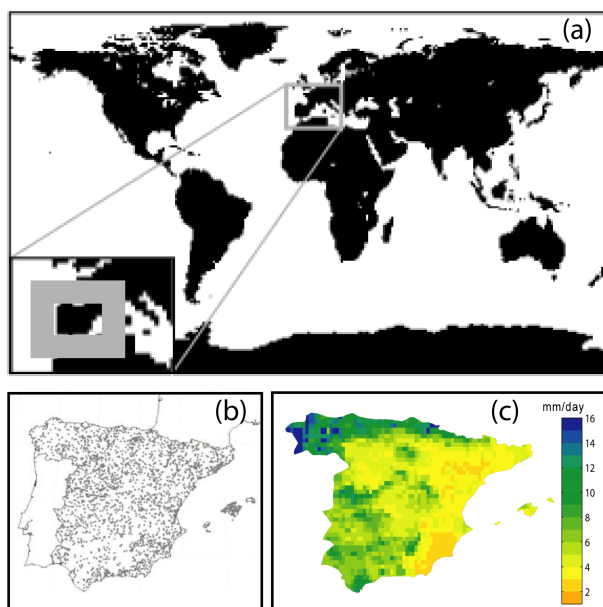


Fig. 1. (a) Land-ocean mask with one degree resolution; the inset shows a magnification of the Iberian peninsula. (b) Network of 2650 stations in Spain with daily records. (c) High-resolution interpolated precipitation (0.2 degrees) for February 1949.

To fill this gap between model outputs and end user requirements, a number of different statistical and machine learning techniques have been developed (down-scaling methods). These techniques use the information contained in meteorological databases of GCM simulations in combination with simultaneous historical

high-resolution observations to infer statistical models relating predicted circulation patterns to observations at local points (see [2,3] for an overview). Some of these methods are more efficient than others for different regions and variables. For instance, the STARDEX project intercompared over 20 different statistical downscaling methods. The skill was found to vary depending on the particular method, variable, season and region, with the latter variation dominating [4]. Thus, for each particular application a range of the better statistical downscaling methods should be tested and validated to achieve the maximum skill and a proper representation of uncertainties. This is a complex task for end-users from applied knowledge domains, since it requires an appropriate statistical background to deal with the downscaling methods and certain computing expertise to deal with geographically distributed data in different formats.

The Statistical Downscaling Portal (SDP) described in this paper has been designed and developed in order to facilitate the downscaling task to end-users through a user-friendly Web portal. In this form, users can obtain their down-scaled data testing and validating different methods in a transparent form not worrying about the details of the downscaling techniques and the data formats and access. This paper describes both the portal architecture and its application to practical problems. In Sec. 2 we give a brief overview of the ENSEMBLES project and describe the planned climate simulations which will be available from the portal for downscaling purposes. In Sec. 3 we illustrate the use of the portal considering an example of seasonal forecast in the North of Spain. Finally, in Sec. 4 we describe the implementation of the portal using Grid middleware.

2 The ENSEMBLES Project. Seasonal Forecast

The statistical downscaling portal is being developed as a part of the EU-funded ENSEMBLES project (for more details see [5] 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 global circulation models developed in Europe. These models will be used in different of experiments to provide seasonal forecasts, decadal forecasts, and climate change scenarios. In particular, the seasonal simulations will provide the scientific community with huge amounts of high-dimensional information to explore and validate the skill of these predictions in different regions of the world. This task will require intensive computational effort due to the volume of information and the complexity of the algorithms involved in the analysis. Each seasonal simulation runs for six months, providing the forecasted daily state of the atmosphere for this period.

One of the ENSEMBLES project's aims is maximizing the exploitation of the results by linking the outputs of the ensemble prediction system to a range of applications, including agronomy, energy, water resources, insurance and weather risk management. This requires efficient downscaling tools to fill the gap between model output resolution and user needs. The portal described in this paper was designed for this purpose and provides user-friendly web access to statistical downscaling techniques and simulations produced in ENSEMBLES.

3 The Downscaling Web Portal

Downscaling methods work by obtaining a statistical model which relates predictors (gridded values of atmospheric variables) to predictands (high-resolution surface observations of goal variables; mainly temperature and precipitation). In the so-called Perfect Prog approach the parameters of these models are fitted using long historical data records. For this purpose, there is a number of model simulation databases (called reanalysis) covering long periods (for instance, the ERA40 reanalysis contains daily information for the period 1957-2002) and simultaneous observation networks (for instance, there is a network of over 2600 stations covering Spain with daily information since 1950, see Fig. 1(b)). With these two sources of information a number of different downscaling algorithms can be calibrated for a region and variable of interest. The resulting models can be later applied to regionally project outputs from ENSEMBLES seasonal models. According to this, the downscaling portal has been organized in three sequential steps: selecting the region of interest and the predictors, selecting the predictands and selecting and applying the downscaling method.

3.1 Selecting the Region of Interest

After login as a guest (restricted functionality) or as a registered user, the first step in the downscaling portal is selecting 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 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. 2a) 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 can be visualized in the lower panels of the window. The defined regions and predictors can be stored in the portal to be used in future sessions. The region manager (see Fig. 2b) allows the user to visualize and load the stored regions.

3.2 Selecting the Stations and the Local Variable of Interest

Once the region of interest and the predictors have been selected, the user can move to the second window (“predictand” tab) and select the network and stations within the region where local forecasts are to be computed. The portal includes some public information such as the Global Station Network (GSN) data, which includes daily observations of precipitation, temperature, etc., for some thousand stations in Europe during the last 50 years. The user can also upload private data to the portal which will be only available for authorized users. The selection of the desired network and the particular stations and the variable to be used in the study is also visually done in a browser window. For

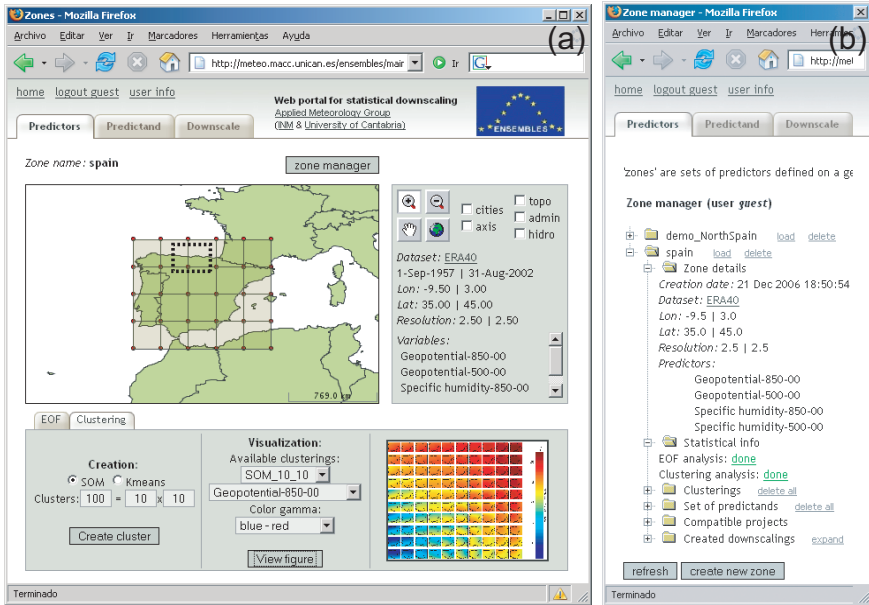


Fig. 2. (a) “Predictors” tab to define the region of interest and the predictors to be used in the downscaling process. (b) Region and predictors manager.

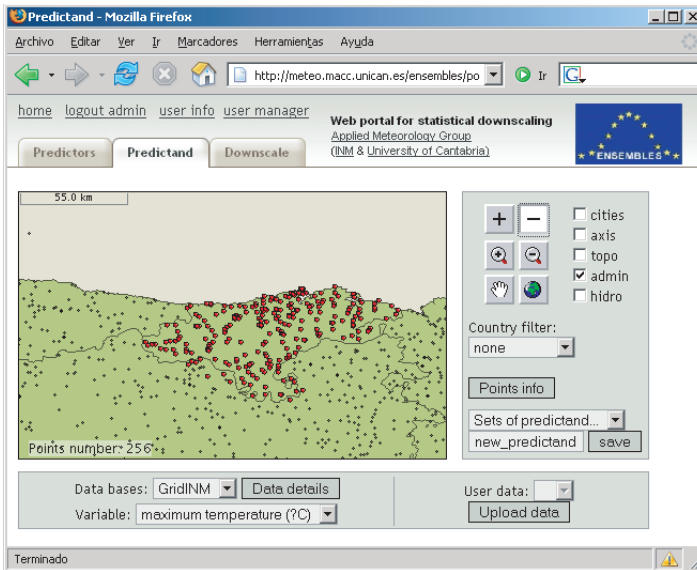


Fig. 3. “Predictand” selector for the desired network and variable

instance, Fig. 3 shows a set of stations in the small region of the North of Spain shown with the dashed box in Fig. 2(a). In this case, maximum temperature is the variable of interest.

3.3 Selecting and Running Downscaling Methods

After selecting the predictors and predictand over a region of interest, the portal allows the user to choose among different downscaling methods from the categories “weather typing”, “regression” or “weather generators”. These methods can be applied for different seasonal experiments (including the DEMETER project and ENSEMBLES’s experiments) selecting the desired models, years and seasons from a matrix containing all the possible combinations to downscale model outputs to local stations (see Fig. 4(a)). This matrix illustrates the complexity of this problem, since each box is a possible downscaling job (the completed jobs are shown in green color). For instance, there is a total of 50 (years) \times 12 (seasons) \times 3 (models) \times 6 (lead time) = 1800 possible downscaling jobs for the DEMETER experiment. Therefore, an efficient design of the computational load is required in order to develop an interactive portal where users can run several jobs simultaneously. In the next section we show how this problem is solved using GRID.

In the near future, this portal will be extended to cover regional projection of climate change scenarios to obtain local estimations of future variables of interest (including an estimation of the associated uncertainty).

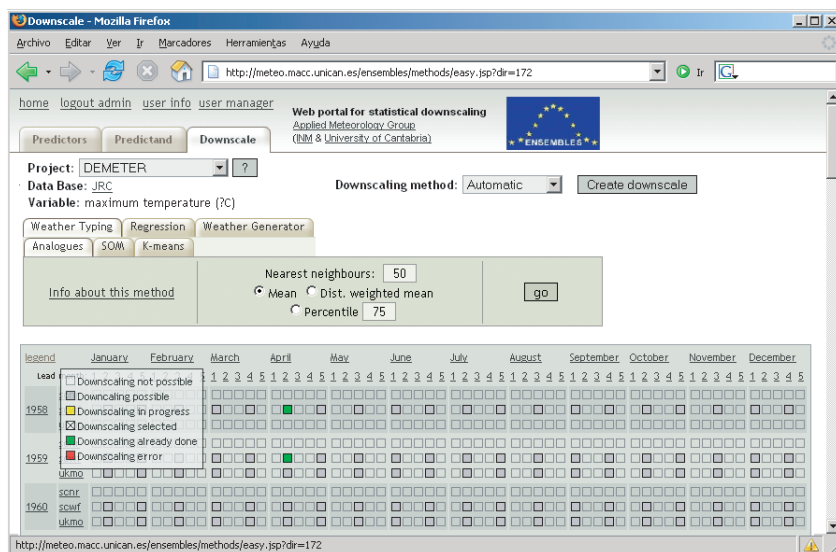


Fig. 4. Downscaling window with the downscaling methods and the matrix of possible downscaling experiments (combining the GCMs, seasons, years and lead times)

4 Implementation with GRID Middleware

GRID computing is a new paradigm for Internet-based distributed computing. It enables the development of interactive problem-solving environments integrating the sharing, selection, and aggregation of geographically distributed autonomous resources, such as computers and databases [6]. The developers of GRIDs applications need to address numerous challenges: security, heterogeneity, resource management, application decomposition, etc. A number of projects around the world are developing technologies (middleware) to run complex applications addressing and testing those challenges. This portal will be implemented in the framework of the 6th EU FP EELA project (see www.eu-eela.org) using its infrastructure based on gLite middleware (see cern.ch/glite).

Fig. 5 shows the design of the portal, which has been implemented using Java technology. On the one hand, the portal can operate using the local `meteo.unican.es` cluster. In this case, in addition to the local data, the portal can access distributed datasets through OPeNDAP protocol to remote storage servers. OPeNDAP technology (see 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. The scripts to run downscaling jobs are based on the open-source toolbox `MeteoLab` developed by our group (see www.meteo.unican.es/MeteoLab for details).

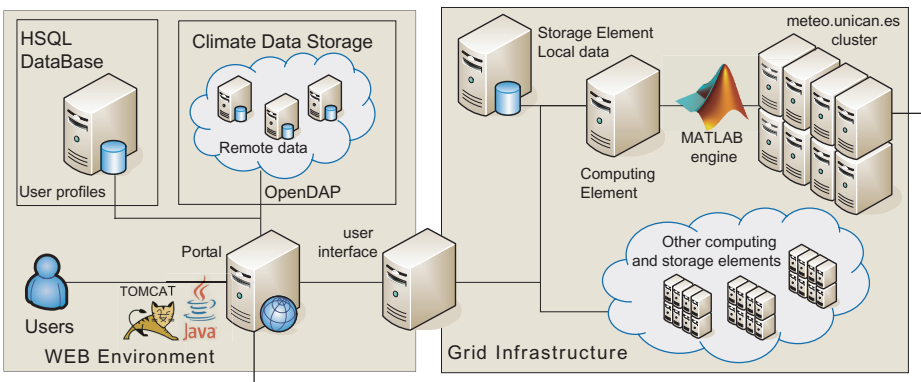


Fig. 5. Scheme of the portal showing the Web and GRID components

On the other hand, the portal will use grid middleware to run the jobs in the GRID using EELA's testbed. In this case, the portal acts as a user interface to the GRID, where a resource broker finds and provides resources to run the job according to the specifications given in the job submission template (in this case, Matlab enabled sites are requested). Apart from the computational resources, the GRID provides distributed data access through the gLite catalog facility

(LFC). Therefore, the datasets stored in the GRID do not need to be submitted, but just conveniently specified the catalog Logical File Name (LFN).

This portal will integrate resources (computing and datasets) from both grid and traditional computing communities, which use heterogeneous authentication methods. Therefore, the portal works with two security alternatives: a local schema based on local authentication and authorization, which is managed by the web portal itself using a database of user profiles, and a second schema which permits the user to configure his profile to use grid security infrastructure (GSI, see www.globus.org/toolkit/docs/4.0/security/) based on X.509 allowing the access to the GRID environment resources. The authorization management is based in Virtual Organization Membership (VOMS) provided by the gLite middleware. This configuration allows to the user make a Single Sign-On (SSO) to store (encrypted) different credentials to access multiple external resources.

5 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 model outputs and observations datasets and uses statistical modeling tools to project model outputs to local observations, which can be geographically distributed in different regions of interest. To this aim, a GRID approach is presented to make available geographically distributed compute and storage resources using the portal. This work is based in gLite middleware used in EELA project and brings together Europe and Latin-America scientific communities.

Acknowledgments. The authors are grateful to the 6th FP EU projects ENSEMBLES (GOCE-CT-2003-505539) and EELA (2004-026409) and to the CLCYT project CGL2005-06966-C07-02/CLI for partial support of this work.

References

1. Gneiting, T., Raftery, A.E.: Weather Forecasting with Ensemble. *Science* **310** (2005) 248-249
2. Wilby, R.L. and T.M.L. Wigley: Downscaling General Circulation Model Output. A Review of Methods and Limitations. *Progress in Physical Geography* **21** (1977), 530-548.
3. Zorita, E. and H. von Storch: The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate* **12** (1999) 2474-2489.
4. Goodess, C. M. *et al.* An intercomparison of statistical downscaling methods for Europe and European regions: Assessing their performance with respect to extreme temperature and precipitation events. *Climate change*, in press.
5. Hewitt, C.D and Griggs, D.J.: Ensembles-based predictions of climate changes and their impacts. *Eos* **85** (2004) 566.
6. Foster, I. and Kesselman, C. (2003). *The Grid 2: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers.