Grant agreement no. 243964

QWeCI

Quantifying Weather and Climate Impacts on Health in Developing Countries

D3.1.b – QWeCI Statistical Downscaling Portal established and open to partners with an initial set of statistical-based seasonal predictions for the target regions with documentation and support on good practises of use

Start date of project: 1st February 2010 Duration: 42 months

Lead contractor:	CSIC	
Coordinator of deliverable:	CSIC	
Evolution of deliverable		
	Due date :	M15
	Date of first draft :	M16
	Start of review :	M18
	Deliverable accepted :	M18

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	RE
CO	Confidential, only for members of the consortium (including the Commission Services)	

Technical Notes
Santander Meteorology Group (CSIC-UC)
SMG:.No.3/2011

Santander Meteorology Group
A multidisciplinary approach for weather & climate

D3.1.b: QWeCI Statistical Downscaling Portal established and open to partners with an initial set of statistical-based seasonal predictions for the target regions, with documentation and support on good practices of use

J.M. Gutiérrez[1], R. Manzanas[1], A.S. Cofiño, D. San-Martín[2], S. Herrera[2]
[1] *Santander Meteorolgy Group, CSIC-Universidad de Cantabria, Spain*
[2] *Predictia Intelligent Data Solutions, Spain*
correspondence: rmanzanas@ifca.unican.es

version:1–*July 2011*

Abstract

Building on the knowledge and experience gained with the ENSEMBLES Downscaling Portal, the QWeCI Statistical Downscaling Portal has been developed and opened to partners at `http://www.meteo.unican.es/downscaling/qweci`. This first version of the portal includes the available local observations in the pilot countries and the state-of-the-art multimodel ENSEMBLES hindcast (predictions from four models, nine members each, for forty years). The portal will be further developed including new methods and datasets in order to provide support to the different activities which need regional seasonal predictions.

Two pre-defined downscaling experiments (one for maximum temperature and the other for precipitation) have been created for each of the pilot countries (Senegal, Ghana and Malawi), thus providing a simple benchmark to further develop optimum downscaling configurations with optimum predictors, etc., in collaboration with local partners. The experiments have been performed in Perfect Prognosis conditions, that is, using ERA40 as predictors, and following a MOS-like approach, i.e., using as unique predictor the same variable to be downscaled (total precipitation and maximum temperature, respectively). Since some key questions as the election of the optimum predictors and domain (region where predictors are defined) have not been treated in depth, results must be taken carefully and considered only as a benchmark to be improved with future work in collaboration with local partners. QWeCI partners will be provided with a personal account which will allow them to perform their own experiments, using all the capabilities of the portal. Moreover, a documentation of the portal has been produced illustration the different steps to be followed in the downscaling process (see `http://www.meteo.unican.es/downscaling/doc/UserGuide.pdf` for a user guide).

1 Introduction

The statistical downscaling portal has been established and open to partners with an initial set of downscaling configurations for the target regions. The portal can be accessed at `www.meteo.unican.es/downscaling/qweci` (Fig. 1 shows the login window to the portal).

Currently, the portal combines information from observations and reanalysis (from Task 1.2a), and from the existing ENSEMBLES multi-model hind cast (seasonal predictions), in order to explore the added value of statistical downscaling in the pilot countries. Fig. 2 shows a schematic diagram showing the validation/calibration and downscaling tasks involved in the process. In the former, an empirical statistical relationship is established between large scale circulation variables and local observations, using the existing reanalysis and historical records, respectively. This relationship can be defined following different methodologies, resulting either in a function or an algorithm, and the particular configuration (variables, parameters, etc.) is iteratively adjusted and calibrated according to the validation results. The final calibrated method is later applied to the seasonal predictions of interest to obtain the local projections (downscaled data).

This document describes the datasets and configurations performed for the initial set of statistical-based seasonal local predictions.

2 Predictands (Observations)

One of the major issues in statistical downscaling is the availability of long records of local observations of good quality (a minimum of 20 years of daily data is desirable). In the case of Africa, where public records are often short and of low quality, this means a serious limitation. Therefore, the first task done for this deliverable was a first quality-control and selection test of the available datasets gathered in the QWeCI atmospheric database (Task 1.2a, available at `http://qweci.uni-koeln.de`). The analysis was conducted in terms of the percentage of missing data and outliers, in order to identify the datasets/stations and periods most suitable for downscaling.

Two daily datasets derived from SYNOP and METAR reports were freely available for the entire African continent from the QWeCI Atmospheric Database: GSOD and

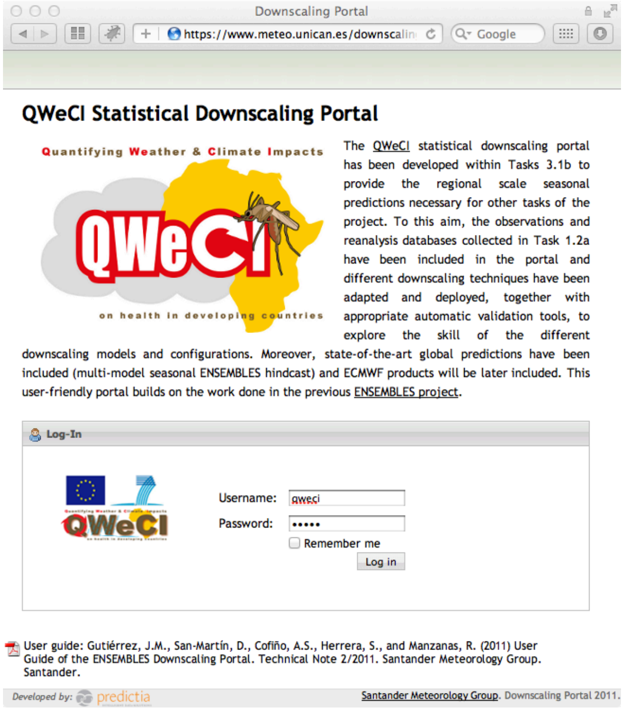

Figure 1: Login page of the QWeCI downscaling portal.

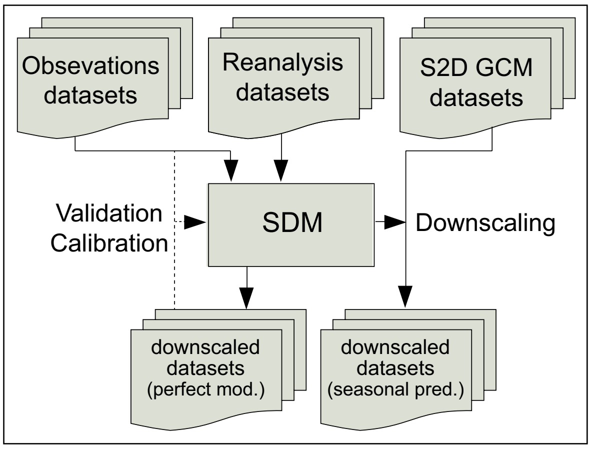

Figure 2: Scheme of the downscaling process using Statistical Downscaling (SD) methods in Perfect Prog. conditions.

MIDAS. The stations from these datasets in the pilot countries have been extracted and, after quality-control have been stored in the portal and labelled as *GSOD_QWeCI* and *MIDAS_QWeCI*, respectively). Moreover, some private datasets owned by local partners were also available. On the one hand, GMet (the Ghanaian meteorological service) provided a dataset with 19 stations for both precipitation and maximum temperature (labelled as *GMet*). On the other hand, Senegal provided CSIC with data of precipitation for 8 stations (labelled as *Senegal*). Some data from Malawi has been also obtained but it is currently being processed and, thus, it is not included in the portal at this stage, nor described in this deliverable.

As a result of the quality-control, important differences among datasets (and among stations within the same dataset) were found. Thus, we considered the optimum dataset in each case and the subset of stations with the smallest mean percentage of missing data over a common period o time. The results for each country are described in the following sections.

2.0.1 Dataset selected for Senegal

GSOD, MIDAS and the Senegalese dataset were intercompared. GSOD has 12 stations in Senegal (11 of them form MIDAS), homogeneously distributed through the entire country. For both variables, the first GSOD records extend back to 1943, but continuous data start in 1973. MIDAS extends from 1985 to 2009. The Senegalese dataset (only precipitation) has only 8 stations, concentrated over a smaller part of the country and covers the period 1950-2000. For precipitation, GSOD and MIDAS show a similar monthly climatology (see Fig. 3). Senegalese dataset reports less precipitation, what could be explained by the different spatial distribution of stations.

The percentage of missing data in GSOD is very high between 1983 and 1998, leading to higher than expected precipitation during this period. In MIDAS, the rate of missing data is clearly smaller in almost all stations for the whole period. Unfortunately, these low rates are reached because missing data were filled with zeros (look at the interannual climatology). This explains the rather high rate of outliers (between 10% and 25% of data) present in all stations. From the previous considerations we recommend the Senegalese dataset (with the exception of Kaolack and Lambaye since both presented suspicious long periods of zeros) for downscaling. Despite a poorer spatial characterization of the country there are neither missing data nor outliers in this dataset. The final period was set to 1961-2000, in order to match with the available reanalysis and the seasonal hindcast (note that this period also avoids the initial decreasing trend of the fifties decade.

For temperature (see Fig. 4), GSOD and MIDAS show a similar monthly climatology, with slightly higher values in MIDAS (probably due to trends, since MIDAS covers a more recent period of time). Although MIDAS presents a rate of missing data smaller than GSOD for the whole period, both are similar when considering only the period from 1975 onwards in GSOD. After selecting those stations with the smallest amount of missing data, we recommend 6/8 stations withing GSOD/MIDAS for downscaling: Saint-Luois, Linguere, Dakar/Yoff, Kaolack, Tambacounda and Ziguinchor for GSOD (1975 onwards) and Diourbel, Podor, Linguere, Ziguinchor, Kaolack, Tambacounda and Matam for MIDAS. Excluding some years presenting problems, both datasets are quite similar in terms of missing data and inter-annual climatologies. In the portal we used the six mentioned GSOD stations: Saint-Luois, Linguere, Dakar/Yoff, Kaolack, Tambacounda and Ziguinchor, and considered the period 1975 onwards.

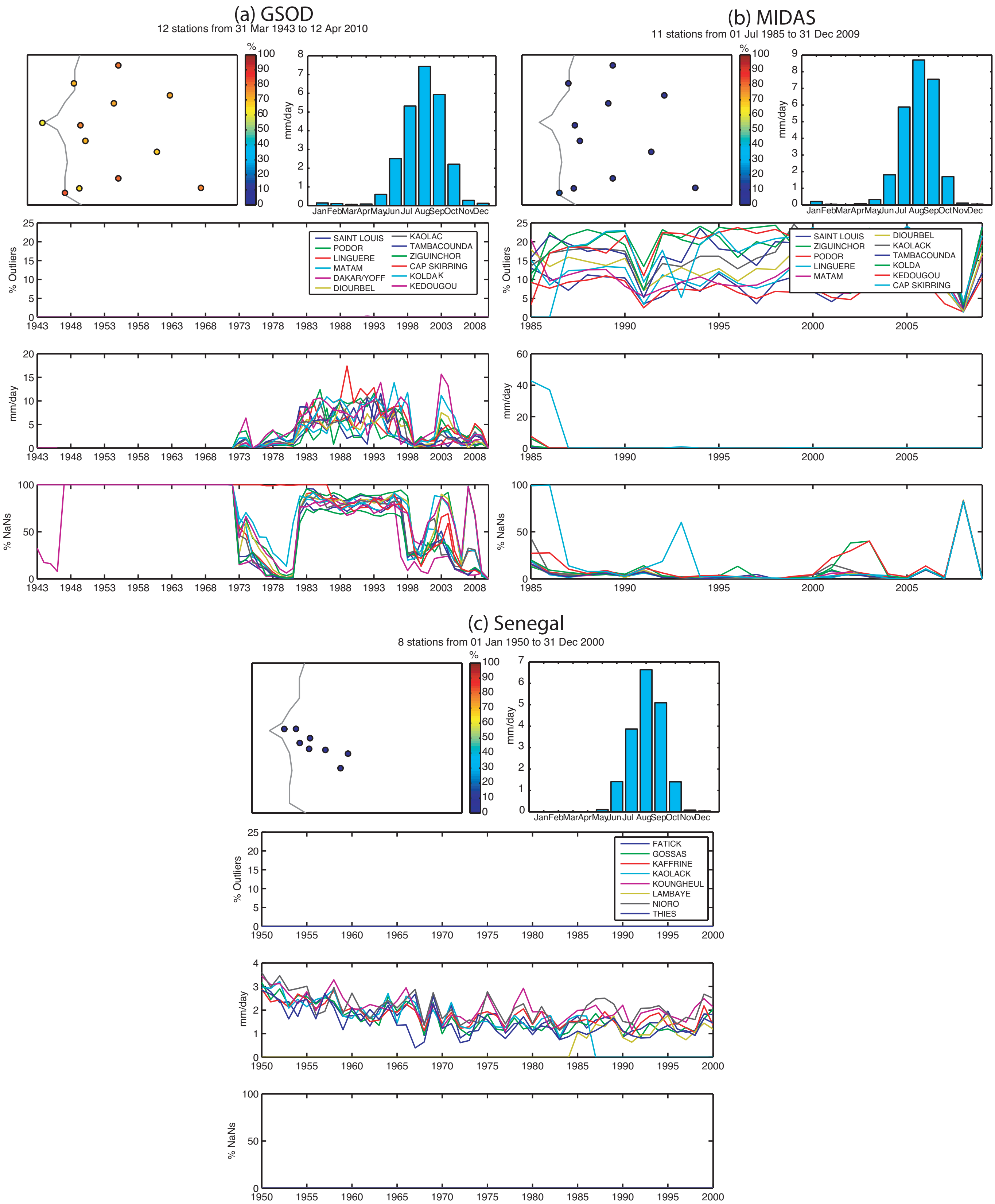

Figure 3: Intercomparison of datasets of precipitation in Senegal: (a) GSOD, (b) MIDAS and (c) Senegalese dataset. In rows from top to bottom: Mean percentage of missing data and monthly climatology (spatially averaged mean daily values) for the period above indicated, yearly percentage of outliers for each station, yearly climatology (mean daily values, in mm/day) for each station and yearly percentage of missing data for each station.

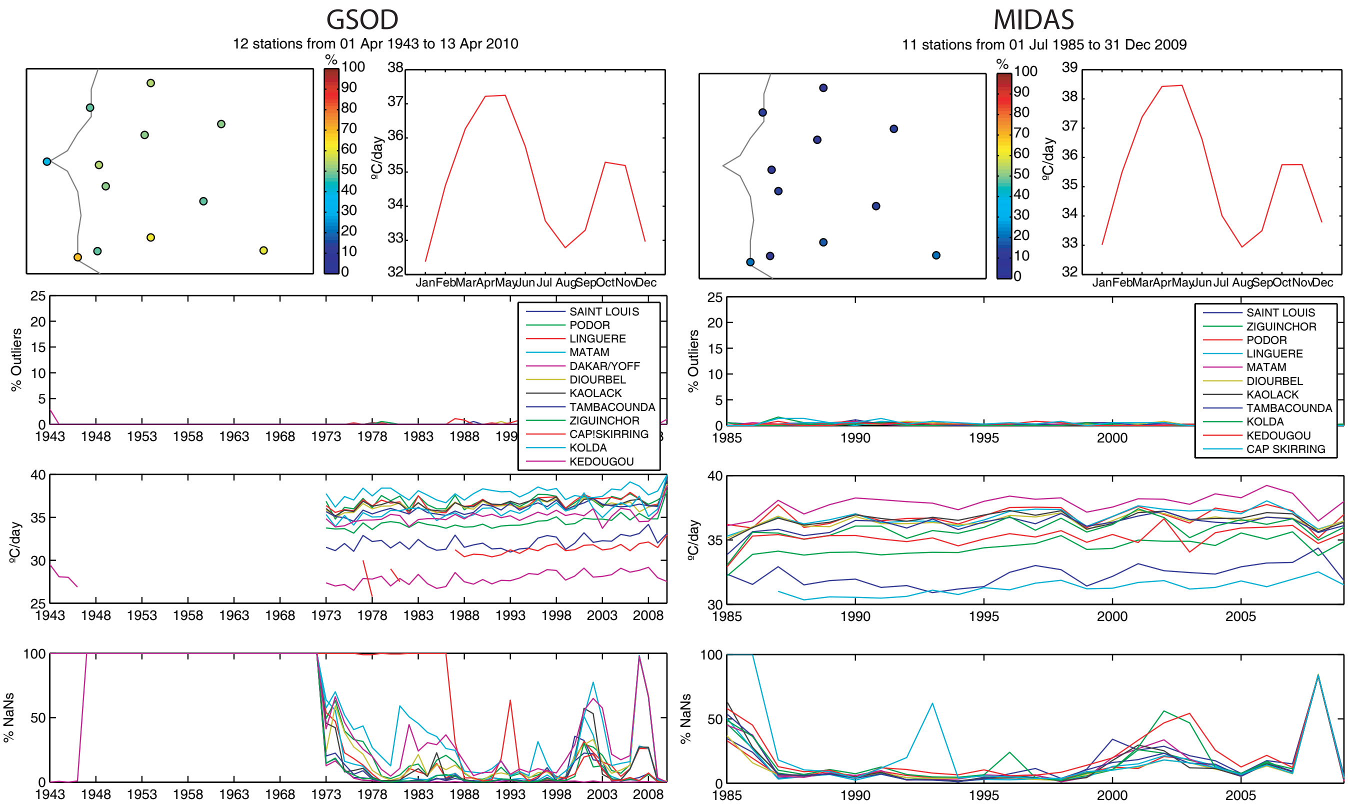

Figure 4: As Fig. 3 but for maximum temperature. In columns from left to right: GSOD and MIDAS.

The optimum observed dataset found for each of the variables after quality-checking was included in the portal as *predictands* with the labels *Precip* and *Tmax*, respectively. For instance, Fig. 5 shows the information for the maximum temperature dataset in Senegal, including the optimum six stations from the GSOD dataset in this country. Detailed information about the structure of the portal and the different options to visualize and create new predictands is given in the user documentation `http://www.meteo.unican.es/downscaling/doc/UserGuide.pdf`, which can be considered a companion of this deliverable.

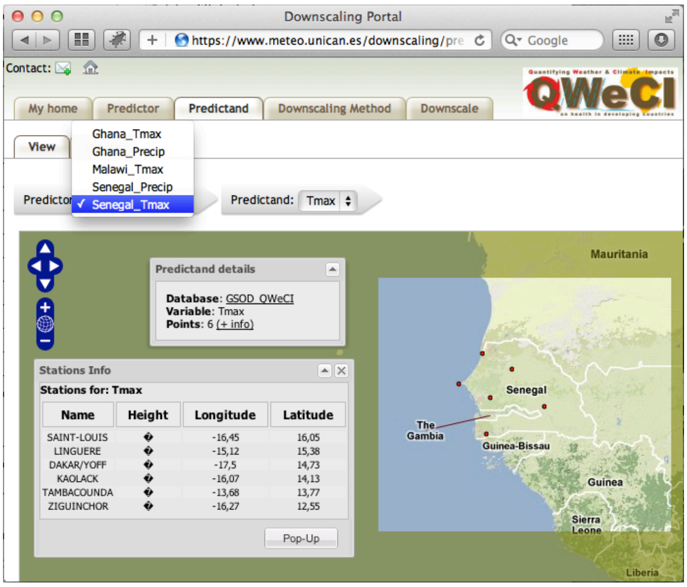

Figure 5: Panel of stations data (predictands) from the downscaling portal.

2.0.2 Dataset selected for Ghana

In Ghana, GSOD, MIDAS and the GMet dataset were intercompared. GSOD has 22 stations, distributed through the whole country, whereas MIDAS and GMet contain 21 and 19 of these stations, respectively. For both variables, the first GSOD records go back to 1942, but a continuous coverage starts in 1972. MIDAS covers the period 1985-2009 and GMet extends from 1960 to 2010.

For precipitation, the three datasets show a similar monthly climatology, although the rainiest month is June for GSOD and GMet and it is May for MIDAS (see Fig. 6). For GSOD, the rate of missing data is near the 100% in all stations for the whole period, whereas the rate is more irregular for MIDAS, but is still high in general, rounding the 50-60%. Furthermore, the latter rates in MIDAS are actually higher, since missing data are presumably filled with zeros (look at the interannual climatology). This explains the high amount of outliers within the dataset. For GMet, the rate of missing data is high in more than half of the stations, but it decreases notably in the others. Thus, we recommend for downscaling the seven stations with the smallest mean percentage of missing data within the GMet dataset: Navrongo, Tamale, Yendi, Kumasi, Akuse, Accra and Axim. The period selected in this case is 1960-2005.

For temperature (see Fig. 7), GSOD, MIDAS and GMet reproduce very similar monthly climatologies. For GSOD and MIDAS, the percentage of missing data is very high in all stations but one for the whole period (except Wenchi in the case of GSOD dataset). For GMet, the rate of missing data is high in the majority of the stations, but it quite low in three of them. Thus, we recommend using these stations for benchmarking the downscaling results: Tamale, Kumasi and Accra, considering the period 1977-2004.

2.0.3 Dataset selected for Malawi

GSOD and MIDAS were intercompared in Malawi. For precipitation/temperature, GSOD has 21/23 stations whilst MIDAS has 23/23, covering the whole country. GSOD/MIDAS records start in 1973/1985 and extend to 2010/2009. Both datasets are of very low quality for both variables with a high number of missing values. Concerning precipitation (Fig. 8), GSOD is mostly empty; only 3 of the stations have some data. Unfortunately, even these stations are mostly incomplete year by year. The shape of the monthly climogram is similar for both datasets, but observed rain in boreal winter months is higher in MIDAS, what could be due to the lack of data in GSOD. In three MIDAS stations, Lilongwe intl Airport, Chileka and Mzuzu, the mean percentage of missing data is rounding the 25% for the period 1992-2006, which is the optimum for this dataset. Unfortunately, missing data are presumably filled with zeros in those stations (as indicated by the zero-valued climatology starting in 1992, as compared with the previous period). This explains the elevate rate of outliers present in the dataset. Consequently, none of these datasets is recommended for downscaling. Those, precipitation in Malawi has not been considered in the portal at this stage.

For temperature (Fig. 9), despite the extremely high number of missing data in both datasets, GSOD reproduces correctly the seasonal cycle throughout the year, whereas it is not realistic for MIDAS. In GSOD, only 3 of the stations (Lilongwe intl Airport, Chileka and Mzuzu) present acceptable rates of missing data for the period 1983-1994. The latter stations cover, from north to south, a wide range of the country. For MIDAS, all stations have a mean percentage of missing data near to the 100% (in fact, only 2 stations have data from 1995 onwards). Given the general bad quality of observations in Malawi, we can only indicate (not recommend) the three mentioned GSOD stations (Lilongwe intl Airport, Chileka and Mzuzu) as the best available for downscaling. Therefore, we request the user to be careful when interpreting results obtained for these stations since only 11 years of data (1983-1994) are considered in this case.

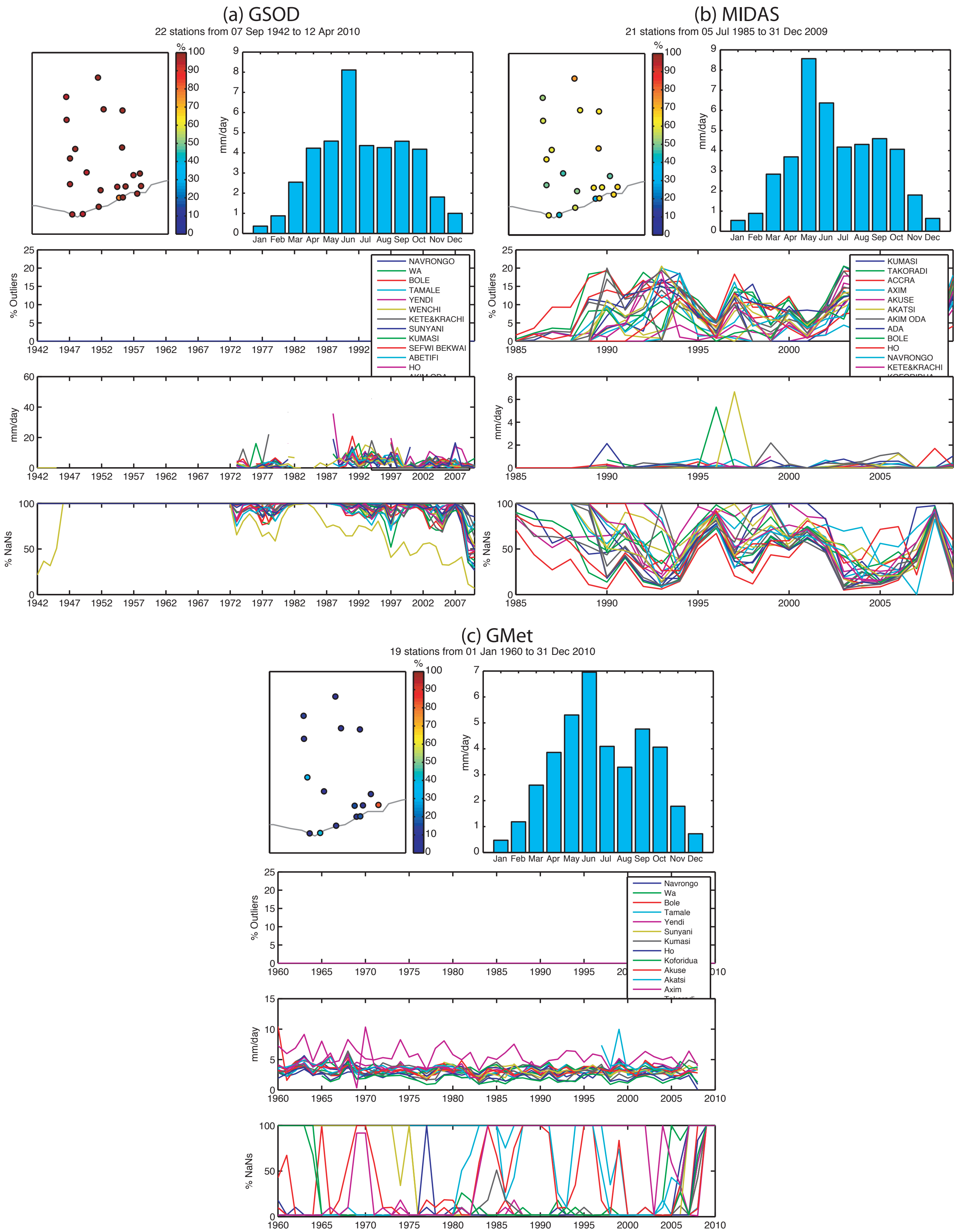

Figure 6: Intercomparison of datasets of precipitation in Ghana: (a) GSOD, (b) MIDAS and (c) GMet dataset. In rows from top to bottom: Mean percentage of missing data and monthly climatology (spatially averaged mean daily values) for the period above indicated, yearly percentage of outliers for each station, yearly climatology (mean daily values, in mm/day) for each station and yearly percentage of missing data for each station.

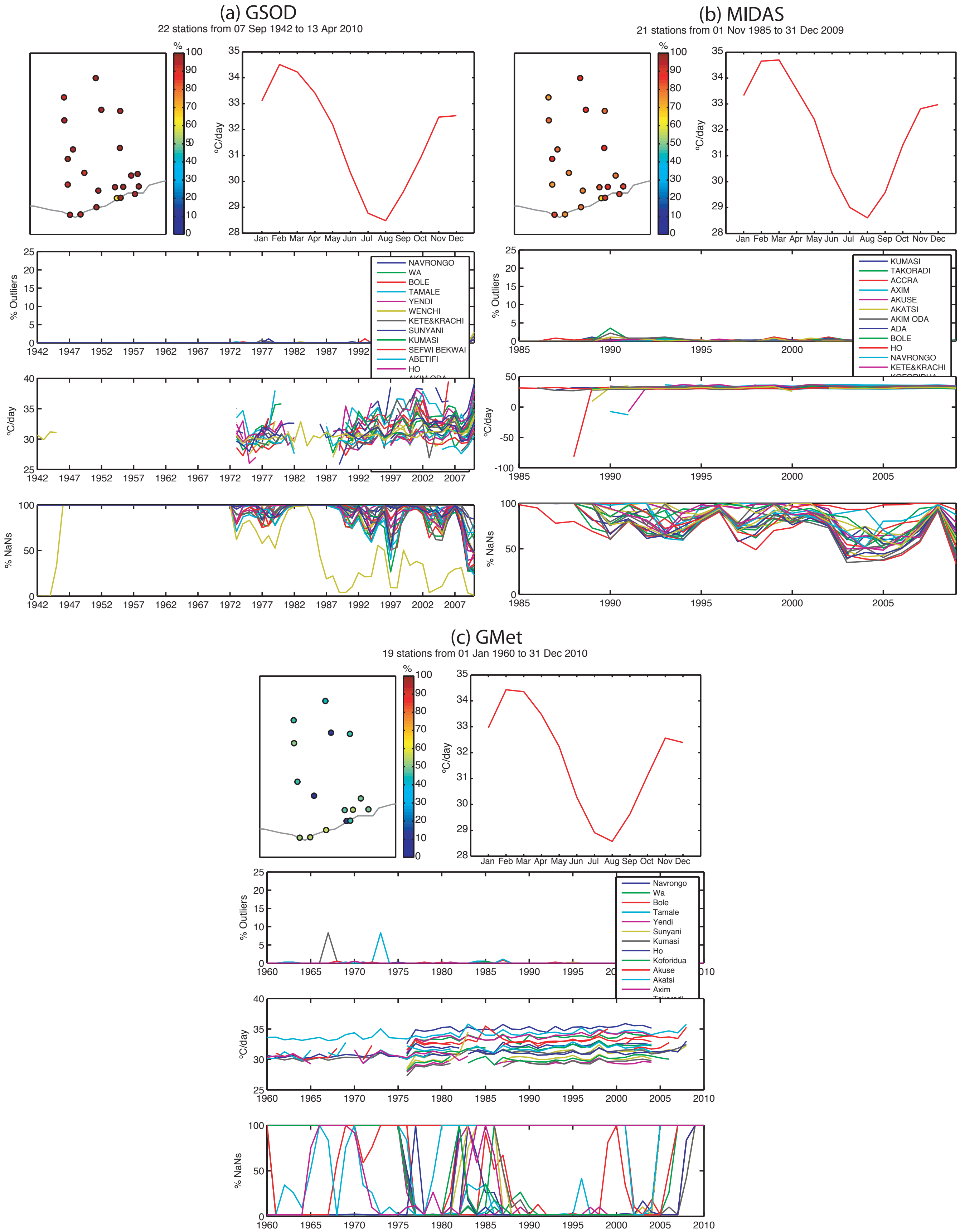

Figure 7: As Fig. 6 but for maximum temperature.

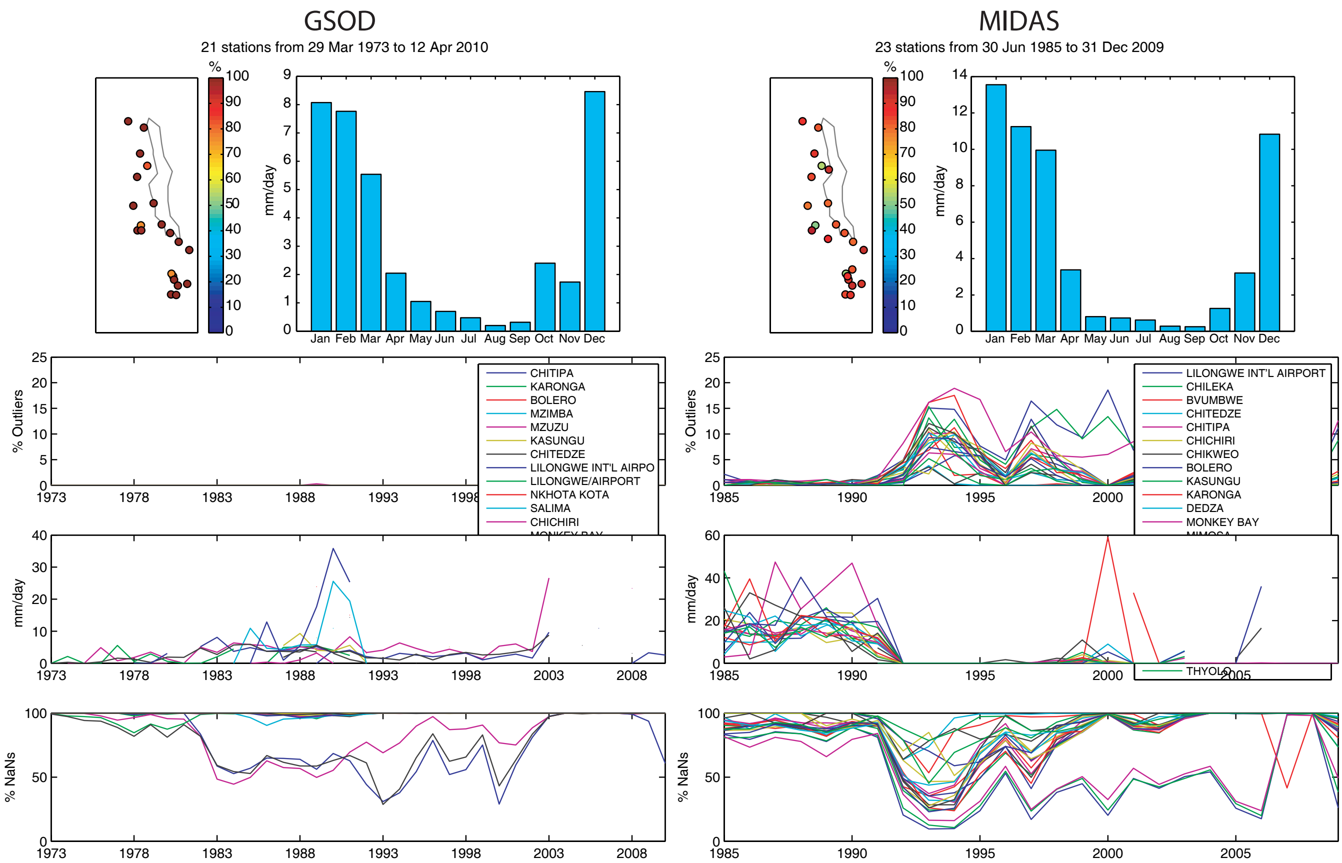

Figure 8: Intercomparison of datasets of precipitation in Malawi: GSOD and MIDAS. In rows from top to bottom: Mean percentage of missing data and monthly climatology (spatially averaged mean daily values) for the period above indicated, yearly percentage of outliers for each station, yearly climatology (mean daily values, in mm/day) for each station and yearly percentage of missing data for each station.

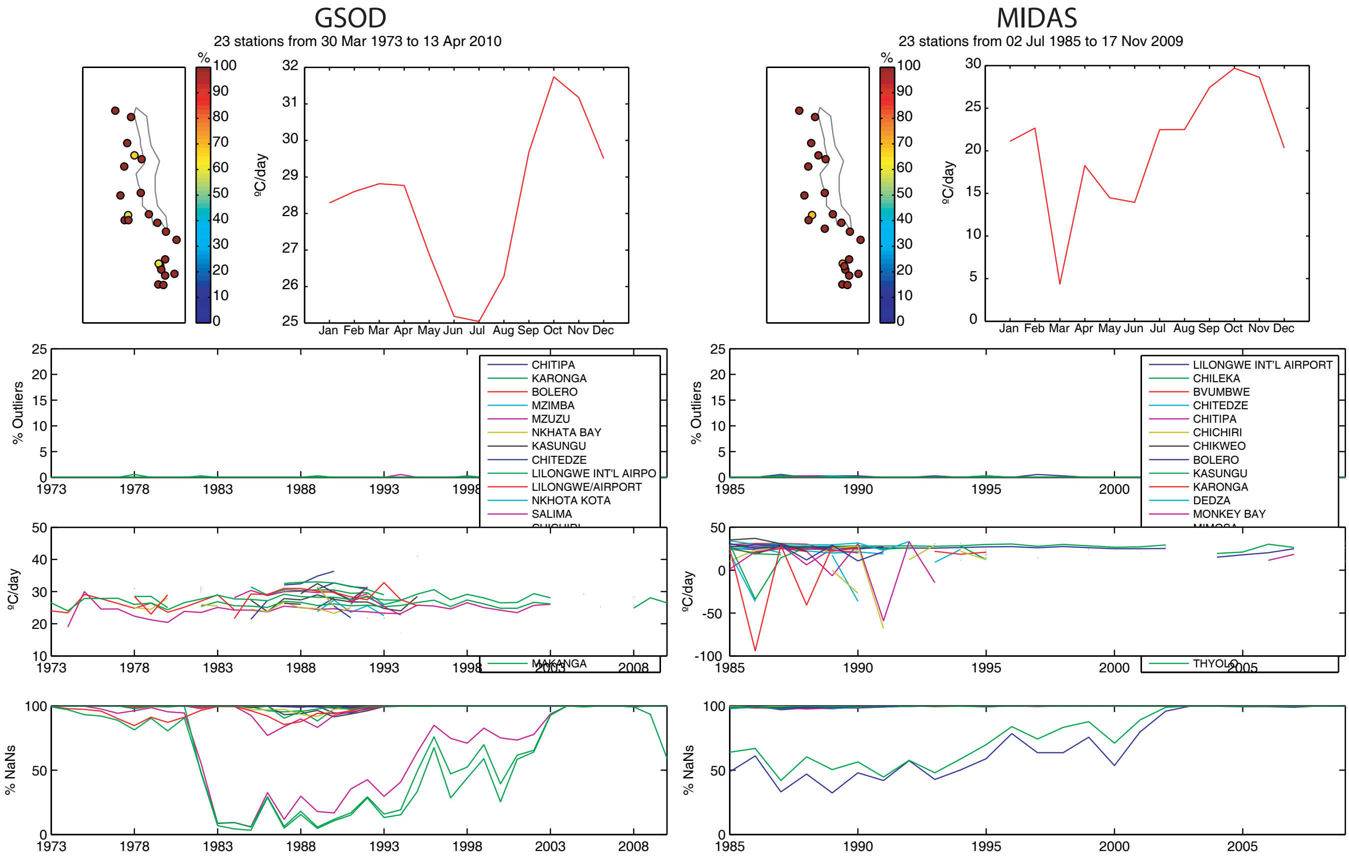

Figure 9: As Fig. 8 but for maximum temperature.

3 Predictors: Reanalysis Data

The statistical downscaling methods included in the portal work on Perfect Prognosis (PP) basis. These methods usually work in two steps: 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 historical period (the intersection of the reanalysis time-window and the observations availability period, with a minimum desirable of 20 years). Then, the resulting statistical model is applied to data from different GCM seasonal predictions to obtain the projected local forecast (in this case the predictor data is build considering the predictor variables from the GCM outputs and bias correction is applied as a post-processing step to GCM outputs in order to correct deviations with the reanalysis data used to train the statistical method [1]). This process is schematically illustrated in Fig. 2; more information about this process and its implementation in the portal is provided in the user guide.

In order to manage a homogeneous basic set of parameters for the different reanalysis and GCMs included in the portal, a common dataset of frequently-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
Maximum temp. (Tmax)	surface	daily
Minimum temp. (Tmin)	surface	daily
Precipitation (Precip)	surface	daily

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.

In the present version of the portal, reanalysis data from the ERA40 ECMWF reanalysis has been included in the portal, for the above mentioned variables; ERA-Interim and NCEP/NCAR Reanalysis1 will be included in a future version. Figure 10 shows the predictor panel of the portal, allowing to select the desired reanalysis, time-window, geographical domain and variables, in order to create a predictor dataset. The selection of an appropriate predictor, with a physical connection to the desired predictand is the most important task in the downscaling process and expert knowledge about the climate over the region of interest is necessary at this step (e.g. provided by the local meteorology services, or research institutes).

In order to create a first set of downscaling configurations in each of the pilot countries, using the quality-checked local observations and the ERA40 reanalysis data, we have considered a simple predictor configuration, which will serve as a benchmark for further improvements of the methods considering appropriate large-scale predictors for each variable in each country. This will be done in collaboration with the partners from the different local countries. The simple configuration used in the pre-defined experiments consider a geographical region restricted to the particular country (it has been shown in different studies that a country-scale geographical domain is appropriate to define the predictors when working at daily timescale). Moreover, in order to get a parsimonious configuration, a unique variable has been considered as predictand: the maximum temperature model output for *Tmax* and the model output precipitation for *Precip*. Note that these variables, in particular the later, are not large-scale circulation variables, well represented in the GCMs and, thus, suitable for downscaling. However, they provide a simple benchmark on the statistical downscaling error in perfect model conditions.

[1]Thus, 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 with promising results. These methods will be included in a future version of the downscaling portal.

4 Downscaling Methods

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. Usually, the different statistical downscaling methodologies are broadly categorized into three classes:

- *Weather typing (analogs)*, 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, thus allowing to test and compare the performance of several approaches (note that the skill of statistical

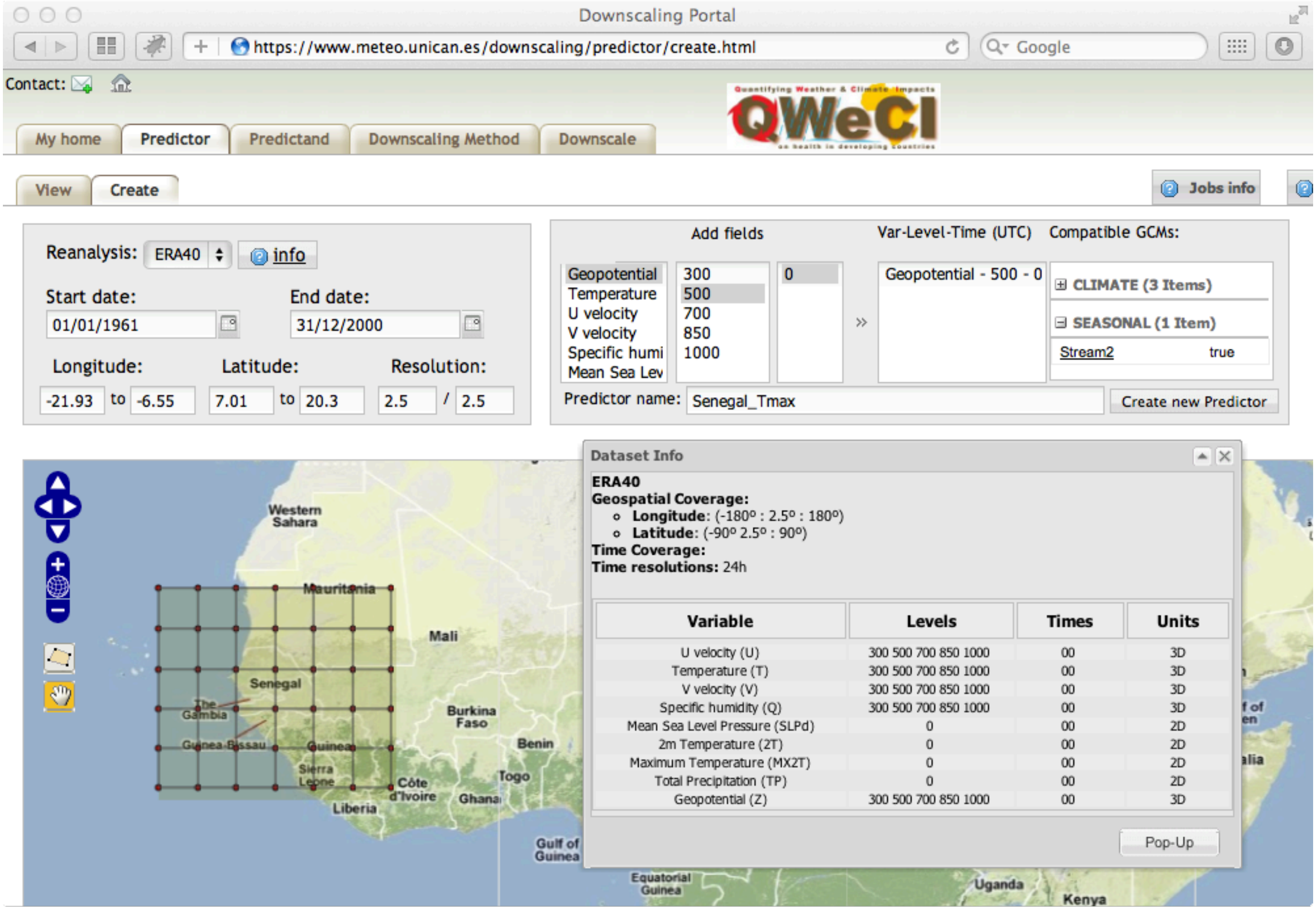

Figure 10: Panel to create predictors (from reanalysis data) in the downscaling portal.

downscaling methods varies from variable to variable and from region to region). For a particular predictor and predict and, a number of methods can be selected, configured from the "Downscaling Method" window, and automatically validation using cross-validation with 75% of the data used for training and 25% reserved for testing the model (see the user documentation). This automatic validation feature is an important help for users for the iterative process of creating the optimum predictor configuration (variables, geographical domain, etc.).

In order to create a first set of downscaling configurations in each of the pilot countries, we have considered the *analog* methodology, since it can be applied to both precipitation and temperature. For instance, Fig. 11 shows the validation report generated automatically (station by station) for the maximum temperature downscaling in Senegal.

5 Downscaling S2D Forecasts

In the present version of the portal, we considered the seasonal hindcast from the multi-model STREAM2 experiment of the EU ENSEMBLES project, comprising five state-of-the-art coupled atmosphere-ocean models from the following centers: The UK Met Office (UKMO), Météo France (MF), the European Centre for Medium-Range Weather Forecasts (ECMWF), the Leibniz Institute of Marine Sciences (IFM-GEOMAR) and the Euro-Mediterranean Centre for Climate Change (CMCC-INGV). All models included the main radiative forcings. None had flux adjustments. The atmosphere and the ocean were initialized using realistic estimates of their observed states and each model was run from an ensemble of nine initial conditions (nine members). Seven months-long hindcasts were issued four times a year within the period 1960-2005, starting the first of February, May, August and November.

Fig. 12 shows the downscaling panel, which allows selecting the desired experiment (only the multi-model ENSEMBLES Stream2 in this version of the portal), the initialization month (e.g. February) and the forecast months to be downscaled. Note that according to the selected initialization month, different forecast months will be available (e.g. March-August in this cast), since the predictions run for seven months. The different GCMs available are shown in columns and a check-box is associated with each combination of GCM and forecast month, so the desired combi-

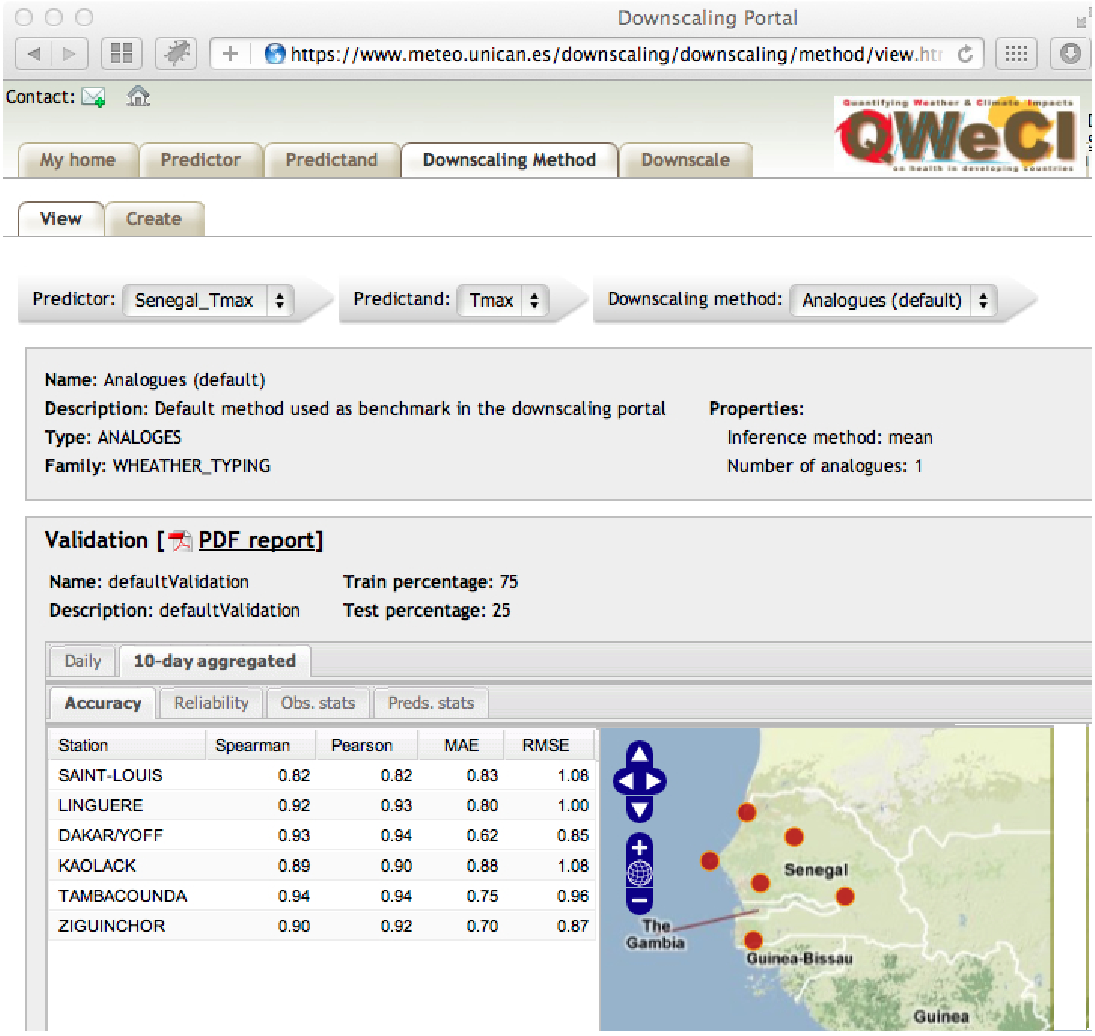

Figure 11: Validation panel to test a particular downscaling configuration for a predictor and predict and dat sets. Standard verification scores are provided station by station both a daily and 10-daily aggregated basis.

nation can be selected and downscaled (see the user guide for more information about the resulting downscaling files, format, etc.).

Finally, all the available experiments, the status of the jobs, etc., can be consulted from the main window (the home tab) of the portal. Fig. 13 shows this panel for the pre-defined user *qweci*, which contains the benchmarking experiments performed in this deliverable. These experiments will be also visible (read-only mode) for all other users defined in the portal.

6 Acknowledgements

We acknowledge the QWeCI project, funded by the European Comission's 7th Framework Programme through contract 243964, for the financial support of this work; as well as the ENSEMBLES project, funded by the European Commission's 6th Framework Programme through contract GOCE-CT-2003-505539, for the data provided.

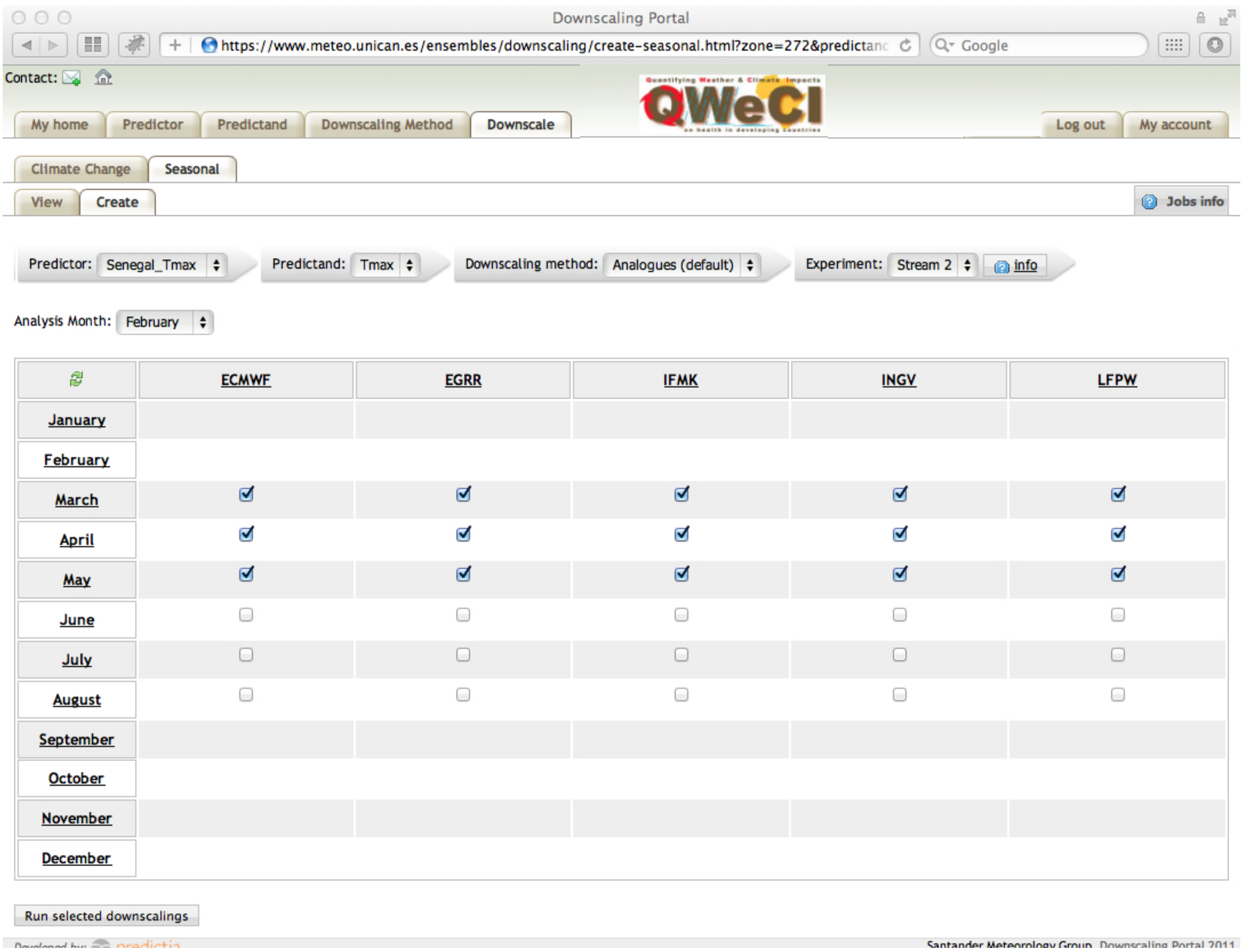

Figure 12: Downscaling panel, allowing to select the desired experiment (e.g. multi-model ENSEMBLES Stream2), the initialization month (e.g. February) and the forecast months to be downscaled. The different models forming the experiment are shown in columns.

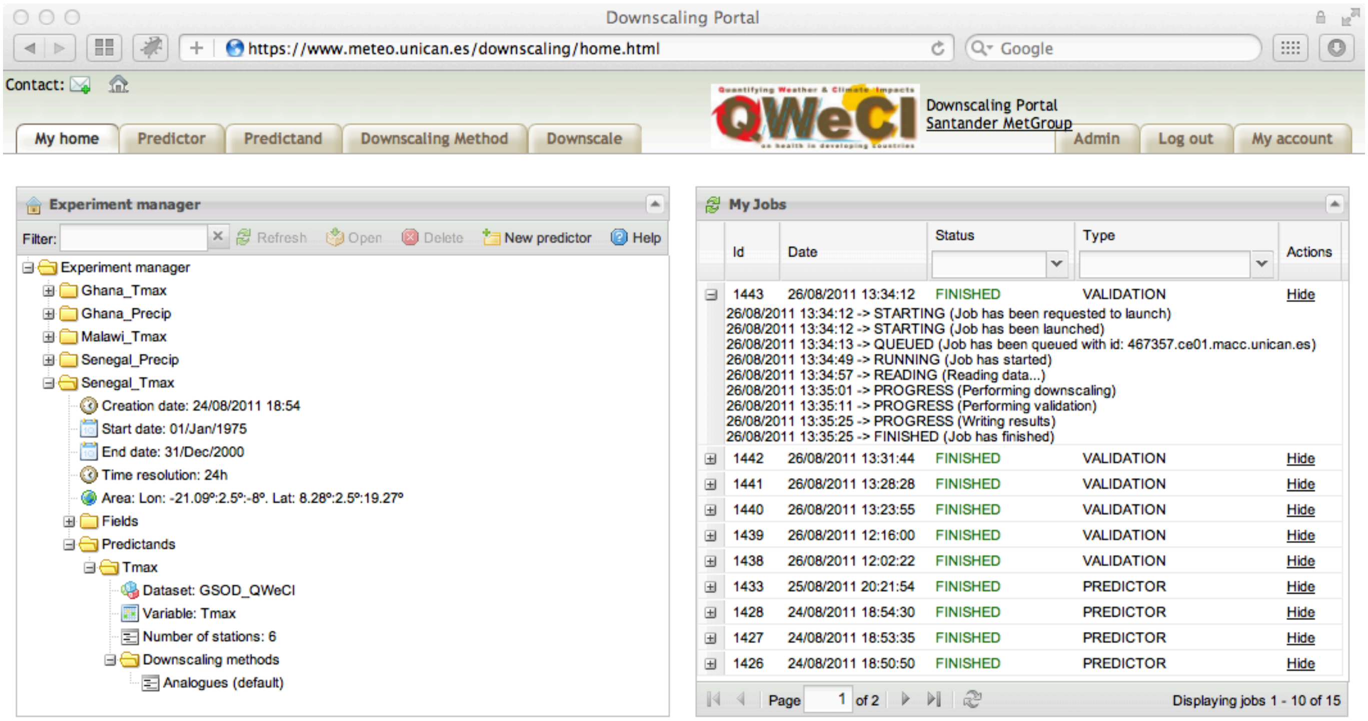

Figure 13: Home panel with the different experiments, jobs and the account information. The details of a particular experiment (left) and a particular job (right) are displayed in this window.