Wikiprint Book

Title: Introduction and Motivation

Subject: TracMeteo - udg/ecoms

Version: 209

Date: 05/17/2022 03:50:56 AM

Table of Contents

The SPECS-EUPORIAS Data Portal

Different sector-specific impact activities to be undertaken in [?SPECS](#) and [?EUPORIAS](#) projects require a reduced number of variables (typically at surface) from different data sources (mainly seasonal forecasts, reanalysis, and observations). The [?SPECS-EUPORIAS Data Portal](#) has been established by the Santander Meteorology Group (UC-CSIC) as part of the data management activities in these projects to provide a unique access for these impact-relevant variables, gathered from existing datasets. The data portal is based on a THREDDS data server providing metadata and data access using OPeNDAP and other remote data access protocols. Moreover, a user-friendly [?R](#) package has also been developed for exploring and remotely accessing subsets of data, thus reducing the burden of data access in these activities. This package will be also a key component for other tasks of the projects based on R, including the validation and downscaling packages to be developed within SPECS and sector-specific calibration and modeling tools to be developed in EUPORIAS.

This trac/wiki page provides an up-to-date description of the SPECS-EUPORIAS Data Portal, including information of the available datasets and the documentation and code of the R package. This page is currently under construction, but both a first tutorial describing the basic functioning and a first version of the R package (a R function) are already available::

Code: [?loadSystem4.R](#)

Tutorial: [PDF file?](#)

Contents (under development):

Introduction and Motivation

The impact activities on seasonal timescales involved in SPECS ([?http://www.specs-fp7.eu](http://www.specs-fp7.eu)) and EUPORIAS ([?http://www.euporias.eu](http://www.euporias.eu)) projects require the use of different data sources (mainly seasonal forecasts, reanalysis, and observations). These activities include the calibration, downscaling, and modelling of sector-specific indices in agriculture, energy, health, etc., building on meteorological information. Typically, only a reduced subset of surface variables (precipitation, temperatures, mean sea level pressure, etc.) or in a reduced number of vertical levels (circulation and termodynamic drivers at, e.g., 850, 500, 200 hPa) is required for these activities. The *SPECS-EUPORIAS Data Portal* has been established by the **Santander Meteorology Group (UC-CSIC)** to gather the relevant information from existing datasets in order to provide a unique homogenized access to data for the SPECS and EUPORIAS partners (in particular for impact-users).

The *SPECS-EUPORIAS Data Portal* is based on a THREDDS data server providing metadata and data access using OPeNDAP and other remote data access protocols. Moreover, since the R language ([?http://www.r-project.org](http://www.r-project.org)) has been adopted for some key tasks in these projects (including the development of comprehensive validation and statistical-downscaling packages) a user-friendly R package has been developed to explore and access the data portal. This package can be used in R programs to remotely access subsets of data, thus reducing the burden of data access (versions for Python and Matlab are also available under request). This package will be continuously updated (keep informed at the documentation URL above) as part of the data management activities to build a data bridge for impact users and for the R developments to be done in these projects.

This document briefly describes the current state of the data portal, which has initially focused on data from the *ECMWF's System4 seasonal model*, as agreed in the downscaling parallel session of the kick-off meeting.

The THREDDS Data Server

The *SPECS-EUPORIAS Data Portal* is based on a password-protected THREDDS data server providing metadata and data access to a set of georeferenced atmospheric variables using OPeNDAP and other remote data access protocols. The variables names, units and additional metadata follow the [?CF convention](#). The variables are spatial grids based on multidimensional arrays of indexed values, following Unidata's _Coordinate convention_ ([?Coordinate Attributes](#) and [?Grid data types](#)).

Typically the data portal will include information at a daily resolution, but monthly-aggregated values could be also provided in some cases due to data limitations (in particular, *Mètèo-France* and *Met Office* have agreeed to provide monthly mean hindcasts for their use by the *SPECS* and *EUPORIAS* partners). In general, the data available will be typical surface variables (e.g. precipitation and near-surface temperature), although several variables (e.g. geopotential and temperature) on pressure levels will also be stored for the statistical downscaling activities.

The data gathering activities have initially focused on the *ECMWF System4 seasonal model*. The Meteorological Archival and Retrieval System (MARS) is the main repository of meteorological data at the *ECMWF* (European Centre for Medium-Range Weather Forecasts). It contains terabytes of operational and research data as well as data from special projects ([?MARS service](#)). The large amount of information stored and the inherent complexities of data access, download and post-processing is a first shortcoming for a flexible use of these datasets by a large number of partners. To overcome this issue, a reduced subset of surface variables ([?http://www.ecmwf.int/products/changes/system4/technical_description.html#description](http://www.ecmwf.int/products/changes/system4/technical_description.html#description)) (precipitation, temperatures and mean sea level pressure) have been downloaded from MARS (a colection of GRIB-1 files) at 0.75º spatial resolution and made available throught the *SPECS-EUPORIAS data portal*. The downloaded data has been exposed as three different virtual datasets using TDS:

- **System4 seasonal range (15 members)**: There are twelve initializations (hereafter called `runtimes`) per year (the first of January, February, ...) running for 7 months (hereafter called simply `times`). An ensemble of 15 members is available for the whole 1981-2010 period.
- **System4 seasonal range (51 members)**: There are only four `runtimes` per year (the first of February, May, August and November) and the forecasts run for 7 months. An ensemble of 51 members is available for the whole 1981-2010 period.
- **System4 annual range (15 members)**: As in the previous case, there are four `runtimes` per year, but the forecasts run for 13 months. An ensemble of 15 members is available for the whole 1981-2010 period.

Data gathering activities will next move to the CFS (?http://cfs.ncep.noaa.gov) version 2 hindcast, developed at the **Environmental Modeling Center at NCEP** and also to reanalysis and observational datasets.

Although the `TDS` provides a web interface to explore and access the datasets (shown in web access section), it is strongly recommented the use of `OPeNDAP` (a.k.a. `DODS`) client libraries to remotely access the data from scientific computing environments (`R`, `Matlab`, `Python`, etc.). For instance, the `R` function provided in this tutorial is based on the *NetCDF Java* OPeNDAP client (?http://www.unidata.ucar.edu/software/netcdf-java/documentation.htm), using the `rJava` `R` package (a similar approach is been also made for the `Matlab` implementation). Alternatively, the most recent *NetCDF library* versions provide access to `OPeNDAP` datasets (this is the solution for the `Python` implementation). In the following, we show a simple example of data access using the `R` package developed as part of the data portal. In particular the *System4* datasets can by directly accessed using the `loadSystem4` function, allowing the retrieval of slices for a particular variable in any of the dataset dimensions (`member`/space/`runtime`/`time`). Note that a more ellaborated worked example using `R` is shown in the R example section. Moreover, for a better understanding of the datasets structure, the use of the web interface for the `OPeNDAP` service is also illustrated web access section.