Climate Services xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Climate Services

journal homepage: www.elsevier.com/locate/cliser

The ECOMS User Data Gateway: Towards seasonal forecast data provision and research reproducibility in the era of Climate Services

A.S. Cofiño [a], J. Bedia [b], M. Iturbide [b], M. Vega [a], S. Herrera [a], J. Fernández [a], M.D. Frías [a], R. Manzanas [b], J.M. Gutiérrez [b,*]

[a] Meteorology Group, Dpto. de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander 39005, Spain
[b] Meteorology Group, Instituto de Física de Cantabria (CSIC – Univ. de Cantabria), Santander 39005, Spain

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Seasonal forecasting
Open science
THREDDS
Reproducibility
R software
Bias adjustment
Downscaling
Validation

ABSTRACT

Sectorial applications of seasonal forecasting require data for a reduced number of variables from different datasets, mainly (gridded) observations, reanalysis, and predictions from state-of-the-art seasonal forecast systems (such as NCEP/CFSv2, ECMWF/System4 or UKMO/GloSea5). Whilst this information can be obtained directly from the data providers, the resulting formats, temporal aggregations, and vocabularies may not be homogeneous across datasets. Moreover, different data policies hold for the different databases, being only some of them publicly available. Therefore, obtaining and harmonizing multi-model seasonal forecast data for sector-specific applications is an error-prone, time consuming task.

In order to facilitate this, the ECOMS User Data Gateway (ECOMS-UDG) was developed in the framework of the ECOMS initiative as a one-stop-service for climate data. To this aim, the variables required by end users were identified, downloaded from the data providers and locally stored as virtual datasets in a THREDDS Data Server (TDS), implementing fine-grained user management and authorization via the THREDDS Access Portal (TAP). As a result, users can retrieve the subsets best suited to their particular research needs in a user-friendly manner using the standard TDS data services. Moreover, an open source, R-based interface for data access and postprocessing was developed in the form of a bundle of packages implementing harmonized data access (one single vocabulary), data collocation, bias adjustment and downscaling, and forecast visualization and validation. This provides a unique comprehensive framework for end-to-end applications of seasonal predictions, hence favoring the reproducibility of the ECOMS scientific outcomes, extensible to the whole scientific community.

© 2017 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Practical Implications

The integration of seasonal predictions in different impact sectors such as agriculture, energy, hydrology and health require data from different sources, including observations, reanalysis and seasonal predictions/hindcasts from state-of-the-art forecasting systems. Typically, only a reduced number of surface variables is needed, which can be directly obtained from the different data providers. However, the resulting formats, temporal scales/aggregations and vocabularies (variable naming and units) may not be homogeneous across datasets. Thus, obtaining and harmonizing the datasets (particularly seasonal predictions) is typically an error-prone, time consuming task. Moreover, different data policies hold for the various datasets (which are freely available only in some cases) and therefore data access may not be straightforward.

The ECOMS User Data Gateway (ECOMS-UDG) was developed in order to mitigate the above mentioned problems, facilitating data provision to end users and favouring science transparency, openness and reproducibility. To this aim, ECOMS-UDG was built upon different open-source software components publicly available: The UNIDATA THREDDS data server, the THREDDS Access Portal implementing fine-grained user management and authorization, and the *climate4R bundle* providing data access and post-processing tools (including bias adjustment and downscaling) based on the R language and computing environment. As a result, ECOMS-UDG provides a unique framework to explore seasonal predictability allowing for the development of end-to-end seasonal

* Corresponding author.
E-mail address: gutierjm@unican.es (J.M. Gutiérrez).

http://dx.doi.org/10.1016/j.cliser.2017.07.001
2405-8807/© 2017 Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: Cofiño, A.S., et al. The ECOMS User Data Gateway: Towards seasonal forecast data provision and research reproducibility in the era of Climate Services. clim. Ser. (2017), http://dx.doi.org/10.1016/j.cliser.2017.07.001

forecast applications using state-of-the-art seasonal forecasting systems (such as NCEP/CFSv2, ECMWF/System4 or UKMO/GloSea5).

The functionalities of ECOMS-UDG are illustrated with a case study application over Europe, analyzing seasonal predictability of winter (DJF) temperatures and precipitation, in connection to North Atlantic Oscillation (NAO) predictability at seasonal time scales. Although some of the datasets used in this work are restricted to ECOMS partners due to data access constraints imposed by the data providers, there is a minimum amount of information (e.g. WFDEI observations, NCEP/NCAR reanalysis and CFSv2 seasonal forecasts) openly accessible, which allows reproducing the results here presented and undertaking further research activities.

Currently, ECOMS-UDG does not provide operational forecasts, but only retrospective forecasts (hindcasts) and reference data (observations and reanalysis). Therefore, operational applications would require accessing (downloading) the operational predictions directly from the data provider; however, the tools provided in ECOMS-UDG can be used to transparently access the downloaded local dataset (e.g. using the corresponding hindcast dictionary), thus facilitating this task.

1. Introduction

The European Climate Observations, Modelling and Services (ECOMS) initiative coordinates the activities of three European projects (EUPORIAS, SPECS and NACLIM) focusing on seasonal to decadal prediction. Different studies carried out in these projects have tested the integration of seasonal prediction in several impact sectors such as agriculture, energy, hydrology and health (Lowe et al., 2016; Ogutu et al., 2016; Bedia et al., 2017a). These sectorial studies require data from different sources (observations, reanalysis and seasonal predictions/hindcasts) for verification and downscaling purposes. Typically, only a reduced number of surface (and upper-air for downscaling) variables is needed, which can be directly obtained from the different data providers. However, the resulting formats, temporal scales/aggregations and vocabularies (variable naming and units) may not be homogeneous across datasets. Thus, obtaining and harmonizing the datasets (particularly seasonal predictions) is typically an error prone, time consuming task. Moreover, different data policies hold for the various datasets (which are freely available only in some cases) and therefore data access may not be straightforward. These difficulties and challenges for moving towards end-to-end climate data services have been reported in several studies (Coelho and Costa, 2010) and constitute a major bottleneck for real-world applications.

The ECOMS User Data Gateway (ECOMS-UDG) was developed as part of the ECOMS data management activities in order to mitigate the above mentioned problems, facilitating data provision to end users (see Fig. 1 for a schematic illustration of the main components). To this aim, the variables required in the sectorial applications were identified, downloaded from data providers, and stored locally in a THREDDS (THematic Real-time Environmental Distributed Data Services) data server implementing fine-grained user authorization via the TAP (THREDDS Access Portal). This provides a one-stop-service for climate data access where users can efficiently retrieve the subsets best suited to their particular research aims (for particular regions, periods and/or ensemble members). Besides the standard data access services (such as OPeNDAP or NetCDF Subset Service), an additional interface (the `loadeR` package) is provided for R users (R Core Team, 2017), providing also appropriate R data structures for data manipulation. Thus, ECOMS-UDG data subsets can be efficiently accessed directly from R using a single line of code.

Furthermore, some common transformation/calibration postprocessing steps are typically applied to raw model data before their use in sectorial models, including data collocation (e.g. regridding, temporal aggregation, subsetting), bias adjustment (e.g. local scaling or quantile mapping) and forecast validation. In some cases, these steps are very technical and they are not always appropriately documented in practical applications, thus making the reproducibility of the results difficult. In order to facilitate these tasks, an R bundle for data postprocessing was developed as an extra layer of ECOMS-UDG, implementing a generic package for data

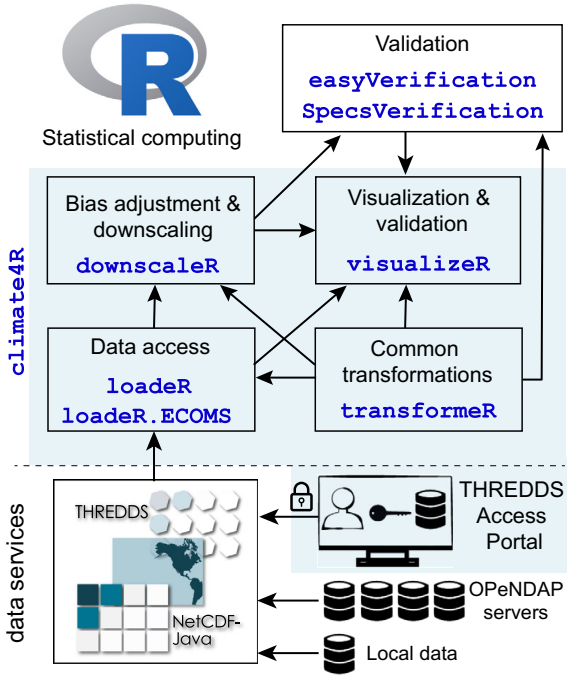

Fig. 1. Schematic illustration of the ECOMS-UDG components, including the THREDDS Data Server (TDS), the THREDDS Access Portal (TAP) and the `climate4R` interface for data access and postprocessing, formed by several packages for data access, transformation, bias adjustment and visualization and validation. Compatibility with some external packages has been achieved by appropriate two-way bridging functions (for the corresponding data structures). Arrows indicate data flow and blue shading indicates in–house developments. All componentes are distributed under GNU General Public License. Some of the images are courtesy of UCAR/Unidata. The R logo is ©2016 The R Foundation.

transformation (`transformeR` package, Bedia and Iturbide, 2017) and bridging some existing packages developed in the framework of ECOMS for bias adjustment and downscaling (`downscaleR` package, Bedia et al., 2017b), and forecast visualization and validation (`visualizeR` package, Frías et al., 2017). Moreover, transparent connection to other external packages (e.g. `easyVerification` package, MeteoSwiss, 2017) was also developed by implementing two-way bridging functions for the corresponding data structures (documented in the corresponding packages). The resulting R bundle (referred to as `climate4R`) provides a unique framework with which data access, postprocessing and validation can be performed using a few lines of code. This allows end-to-end experimental reproducibility and facilitates the description (metadata) and documentation of the whole data flow. An up-to-date description of ECOMS-UDG, including information on the available datasets, variables and tools is provided in the wiki page: http://www.meteo.unican.es/ecoms-udg.

Please cite this article in press as: Cofiño, A.S., et al. The ECOMS User Data Gateway: Towards seasonal forecast data provision and research reproducibility in the era of Climate Services. clim. Ser. (2017), http://dx.doi.org/10.1016/j.cliser.2017.07.001

Download English Version:

https://daneshyari.com/en/article/8867283

Download Persian Version:

https://daneshyari.com/article/8867283

Daneshyari.com