



Evaluation reports per domain and user guide

Issued by: CSIC-IFCA

Date: 20/07/2021

Ref: C3S_D34d.3.3.1_validation_results_v3

Official reference number service contract: 2019/C3S_34d_CSIC-IFCA/SC1

This document has been produced in the context of the Copernicus Climate Change Service (C3S). The activities leading to these results have been contracted by the European Centre for Medium-Range Weather Forecasts, operator of C3S on behalf of the European Union (Delegation Agreement signed on 11/11/2014). All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission and the European Centre for Medium-Range Weather Forecasts has no liability in respect of this document, which is merely representing the authors view.



Contributors

CSIC-IFCA

Javier Díez
Sixto Herrera
Antonio S. Cofiño
José M. Gutiérrez

SMHI

Grigory Nikulin

Predictia

Daniel San Martín
Max Tuni



Table of Contents

1. Introduction	4
2. Diagnostic results per domain	5
3. Evaluation results per domain	6
4. User guidance	9
5. Summary	10
6. References	11



1. Introduction

The evaluation framework adopted in C3S_34d is described in D34d.3.1.1 and builds on the VALUE initiative (Maraun et al. 2015) for the evaluation of Regional Climate Models (RCMs). A reduced number of evaluation indices was selected (see Table 1), providing basic information on central tendency and variability (e.g. mean and variance), temporal (e.g. dry/wet spells) and extreme (e.g. extreme percentiles) aspects. The proposed selection is aligned with some recent studies (Maraun et al. 2019, Vautard et al. 2020, and Coppola et al. 2021) and also with the basic RCM evaluation results included in the IPCC-AR6 Atlas (which includes results only for biases in the mean values). The C3S worldwide CORDEX evaluation is designed as an extension of the evaluation results in the AR6 Atlas (which will be available by 9 August 2021).

Index Code	Measure	P	T	W	Description	C	T	E
Mean	Bias (% for P)	✓	✓	✓	Mean value	✓		
Variance	Bias (%)	✓	✓	✓	Quasi-variance	✓		
WetAnnual- MaxSpell	Bias (%)	✓			Median of the annual maximum wet (precip ≥ 1 mm) spells		✓	✓
DryAnnual- MaxSpell	bias (days)	✓			Median of the annual dry spell maxima		✓	✓
P98	bias (%)	✓	✓	✓	98th percentile (on wet days for precip)			✓
AnnualCycle- RelAmp	bias (%)	✓	✓	✓	Relative amplitude of the annual cycle (range of the monthly mean values over the annual mean)	✓	✓	
Interannual- Var	bias (%)	✓	✓	✓	Interannual variability (std of the annual time series)	✓	✓	
Rx1day	bias (%)	✓			Annual maximum 1-day precipitation			✓
TGx	bias		✓		Annual maximum of mean daily temperatures			✓
TGn	bias		✓		Annual minimum of mean daily temperatures			✓

Table 2. Validation indices and measures used in VALUE. Index codes refer to the package VALUE¹. Third, fourth and fifth columns indicate whether the indices are applied for validating precipitation (P), temperature (T) or wind (W). The last three columns indicate whether the indices analyze central tendency or variability (C), temporal (T) or extreme (E) aspects, respectively. Note that spells have been defined as at least two consecutive days fulfilling the particular condition.

¹ <https://github.com/SantanderMetGroup/VALUE>



The main goal of the proposed evaluation framework is providing homogeneous information for the whole ensembles available across the different domains – thus favoring comparability –. To this aim, two products were defined:

1. **Diagnostics:** statistics obtained directly from model data providing a visual outlook suitable to identify strange values and/or outliers in the data and/or models that deviate from the ensemble.
2. **Evaluation indices:** Indices (shown in Table 1) comparing model data and reference observations to provide objective measures of model performance informing on the merits and limitations of the simulations forming the ensemble.

Both diagnostics and evaluation indices can be computed from the available *evaluation* (ERA-Interim-driven) and *historical* (GCM-driven) simulations (see final inventory in D34d.1.1.3), providing complementary information for end users: 1) intrinsic performance of the RCM and 2) conditioned performance when driven by a particular CMIP5 GCM. Therefore, both experiments are used to calculate diagnostics and evaluation indices for both the RCM simulations and the driving models (ERA-Interim or the particular GCMs).

2. Diagnostic results per domain

Diagnostics provide an outlook of the ensemble and allow for quick outlier or model departure detection. In D34d.3.1.1 the use of climate stripes² was proposed as a basic diagnostic for the CORDEX ensembles, showing visually the time series of annual spatially aggregated values over the whole domain (in columns) for the different available RCM simulations (in rows).

The results for 6 core variables of the CORDEX worldwide dataset (surface temperature, surface minimum temperature, surface maximum temperature, surface wind, precipitation, and specific humidity) have been computed for every domain (13) and scenario (5, evaluation, historical and scenarios RCP26, 45 and 85), as one pdf document per domain and are provided as a zip file together with this document; moreover, for tracking and further work, they have been stored (as csv and pdf files) in the “diagnostic” folder of the C3s_34d validation Git repository:

https://gitlab.com/c3s_34d_cordex4users/c3s_34d_validation

Figure 1 shows the results for a representative domain (CORDEX-AFR) stored as a single pdf file, which contains the 30 diagnostics resulting from the six variables analyzed and the five scenarios considered (evaluation and historical for the periods 1980-2005 and 1950-2005, respectively, as well as the future RCP26, 45 and 85, for the period 2006-2100).

² Climate stripes are a new graphical diagnostic consisting of a series of coloured stripes chronologically ordered to visually portray climatological values and variability.

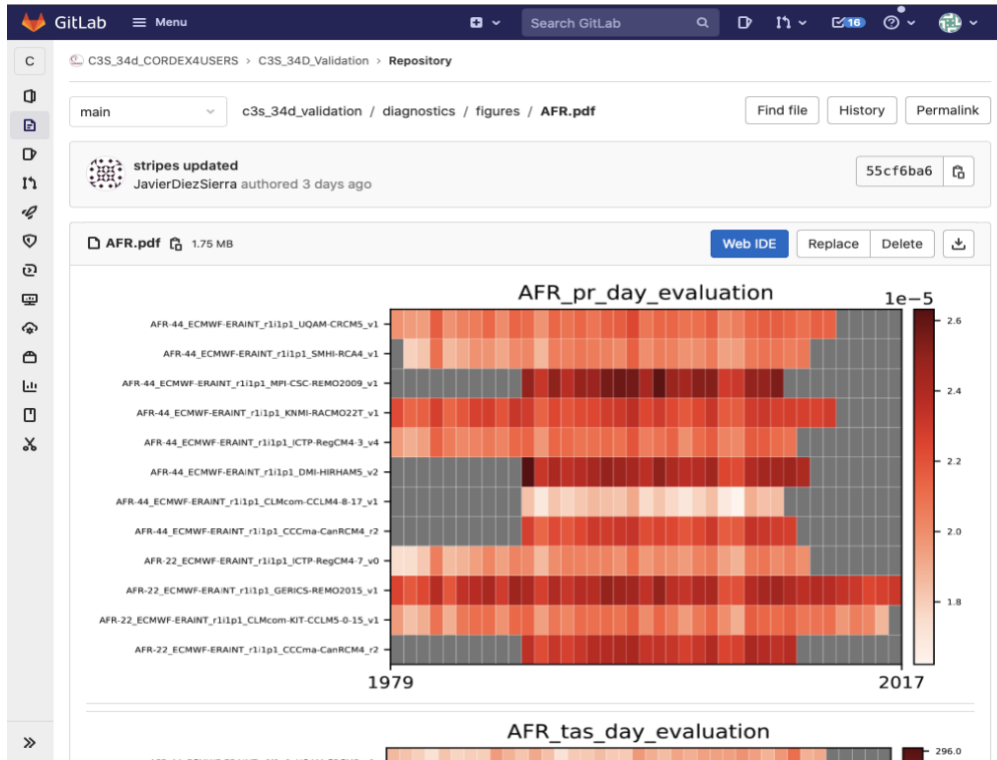


Figure 1. Screenshot of the “diagnostic” folder of C3s_34d validation Git repository showing an example of climate stripes diagnostics for CORDEX-AFR ensemble (the pdf file contains all the diagnostics for this domain organized by scenarios: evaluation, historical, RCP26, RCP45, RCP85).

3. Evaluation results per domain

The indices shown in Table 1 have been computed for the different CORDEX simulations (evaluation and historical scenarios) and for the observational reference used for model evaluation, the ERA5 bias adjusted version WFDE5 (Cucchi et al., 2020). All model results are regridded (conservative remapping) to the WFDE5 regular half-degree horizontal resolution grid before undertaking index calculation, so the indices for simulations and observations can be compared obtaining performance measures. The evaluation period considered is 1980-2005, which is a compromise between having a long reference period and using the common period of all datasets.

Evaluation results are summarized for each of the CORDEX domains aggregating the gridbox evaluation results over the IPCC AR6 reference regions (see Figure 2). This allows obtaining user-friendly information in the form of tables or color matrices facilitating a quick overview of the results of a particular domain for the whole ensemble.

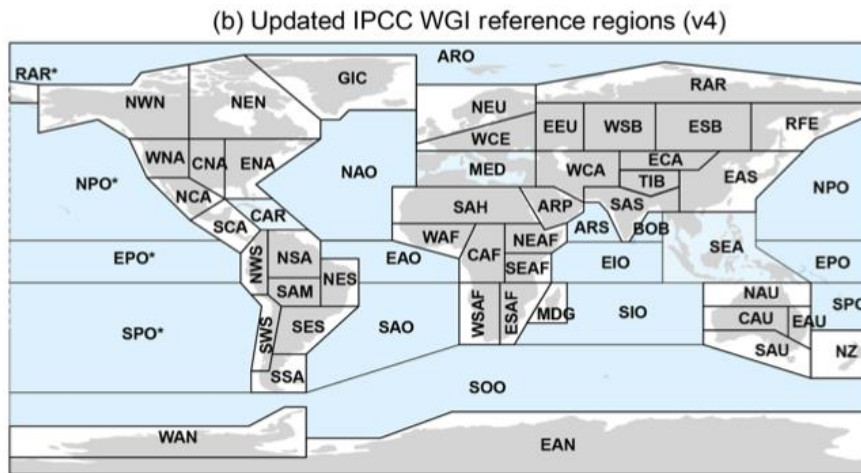


Figure 2. IPCC reference regions with 46 land (grey shading) and 15 ocean (blue) regions; stars indicate regions across the zero meridian which are split in two (Iturbide et al. 2020).

The results in form of pdf files (one pdf document per domain) are provided as a zip file together with this document; moreover, for tracking and further work, they are stored in the “indices/figures” folder of C3s_34d validation Git repository:

https://gitlab.com/c3s_34d_cordex4users/c3s_34d_validation

The results of all the indices for a single domain are stored in single pdf file (one index per page). The layout is the same for all indices and includes both the reanalysis-driven (left block) and the GCM-driven results (right block), as shown in Figure 3 for the first index in the Antarctica domain. The GCMs are represented in magenta, with all driven RCMs (in black) following the GCM. Each cell shows the biases corresponding to the annual (central), summer/JJA (upper-left corner), and winter/DJF (lower-right) values. The rows display results for different reference regions.

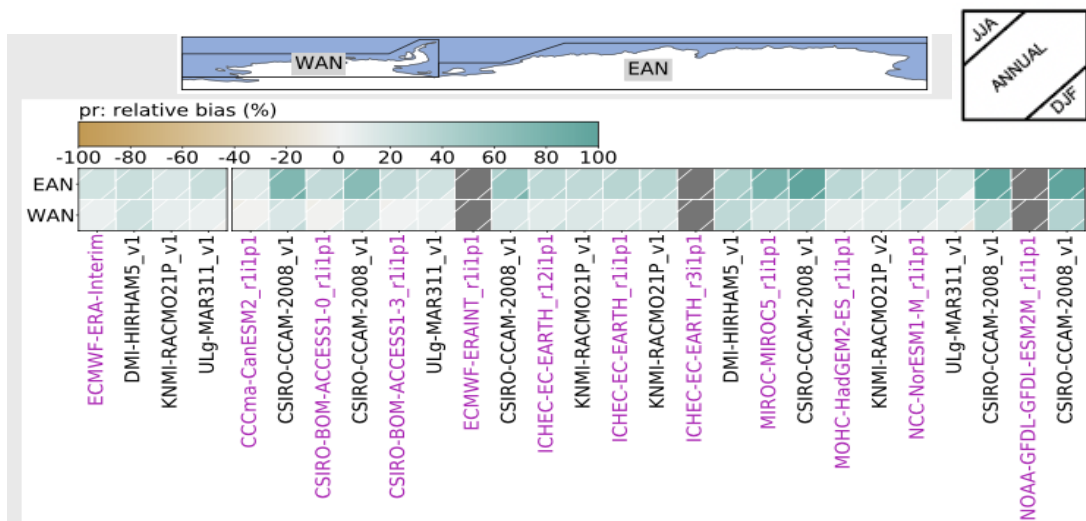


Figure 3. Screenshot of the first two pages (regions, top) and first index (bottom) for the Antarctica domain for the Western (WAN) and Eastern (EAN) Antarctica reference regions.



Figure 4 shows and screenshot of the “indices/figures” folder of C3s_34d validation Git repository including the results for all CORDEX domains.

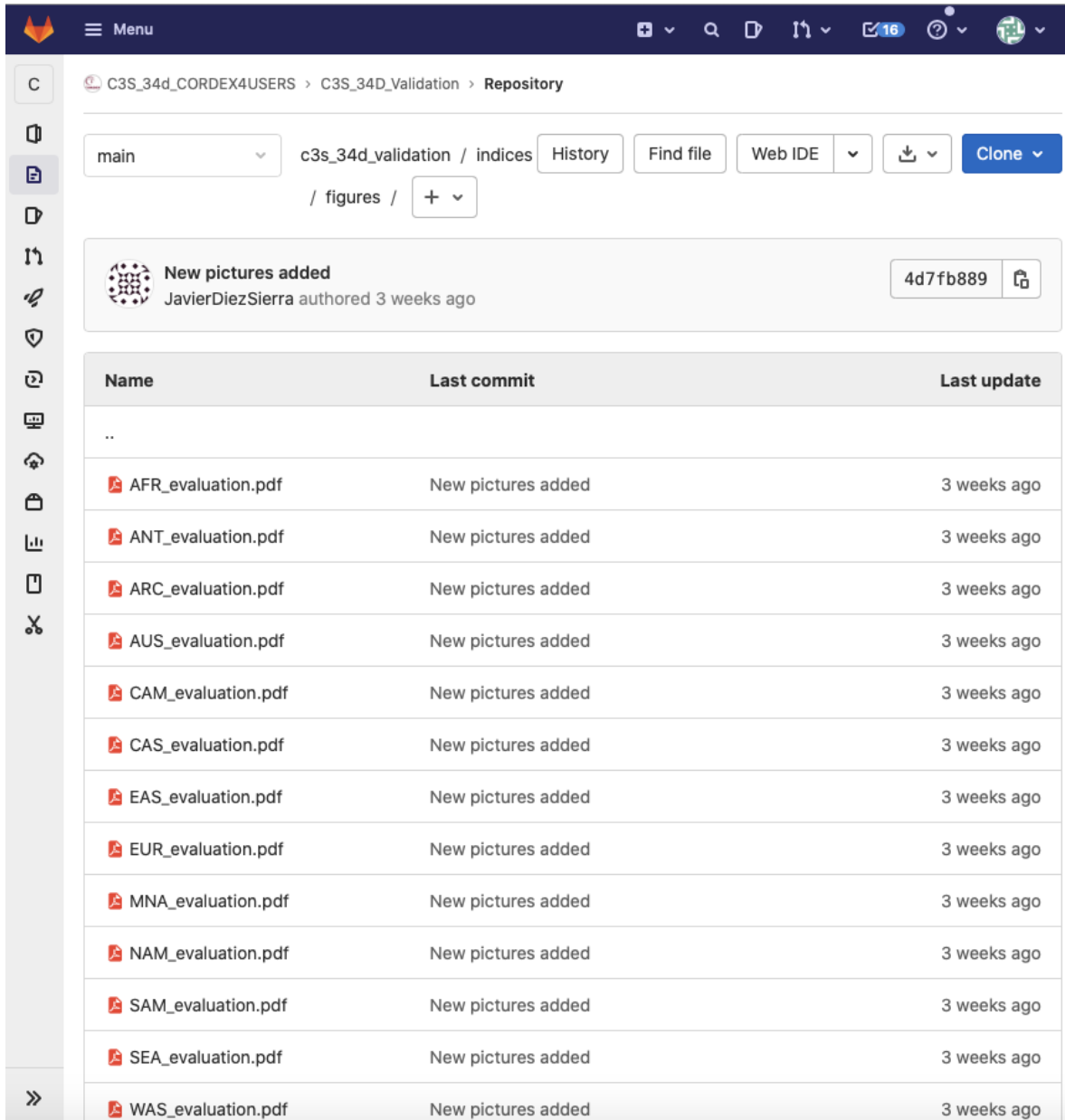


Figure 4. Screenshot of the “indices/figures” folder of C3S_34d validation Git repository showing the evaluation reports per regions available. Each of the pdf files contains the results of the twenty indices shown in Table 1 (one per page).



4. User guidance

The evaluation framework described in this document has been designed to provide users with homogeneous information for the whole ensembles available across the different domains using two products. First, a simple diagnostic (climate stripes of the annual time series of the whole ensemble) provides users with an outlook of full ensemble for all available scenarios, suitable to identify strange values (for models and/or years) and/or outliers in the data and/or models that deviate from the ensemble. Second, a number of validation indices characterizing different performance dimensions (central, temporal and extreme aspects) are computed for the evaluation and the historical RCM simulations. The former informs about the intrinsic performance of the RCM, driven by “perfect” boundary conditions; the latter informs about the conditional performance when driven by a particular CMIP5 GCM. Overall, the results for the evaluation and historical scenarios allow analyzing the biases of the different RCMs and how they change when driven by different GCMs. This information may help to detect problems in particular RCM-GCM couplings in particular domains.

These results provide some basic evaluation of the CORDEX worldwide dataset, harmonized across the different CORDEX domains. However, the available pdf documents and figures hampers a user-friendly analysis of the information. Therefore, we propose the development of a simple web page which allows selecting the particular domain and additional relevant elements (scenario, variable, index) and presenting the selection thus facilitating a user-friendly access to the data. Figures 5 and 6 show a mockup of the web page with the selectors for the diagnostics and the evaluation components. In each case, every choice corresponds to a single figure which is presented to the user. The integration of similar results from C3S_34b Lot1 will be considered (see the GCM-Evaluation tab in the mockup figures) and designed in case this proposal is accepted by ECMWF.

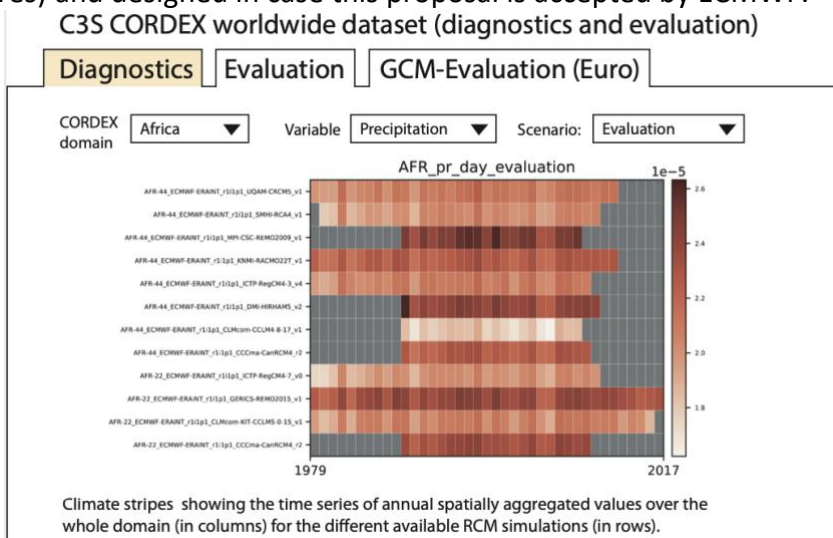


Figure 5. Mockup of the diagnostics and evaluation web page, showing the options for the diagnostics component.

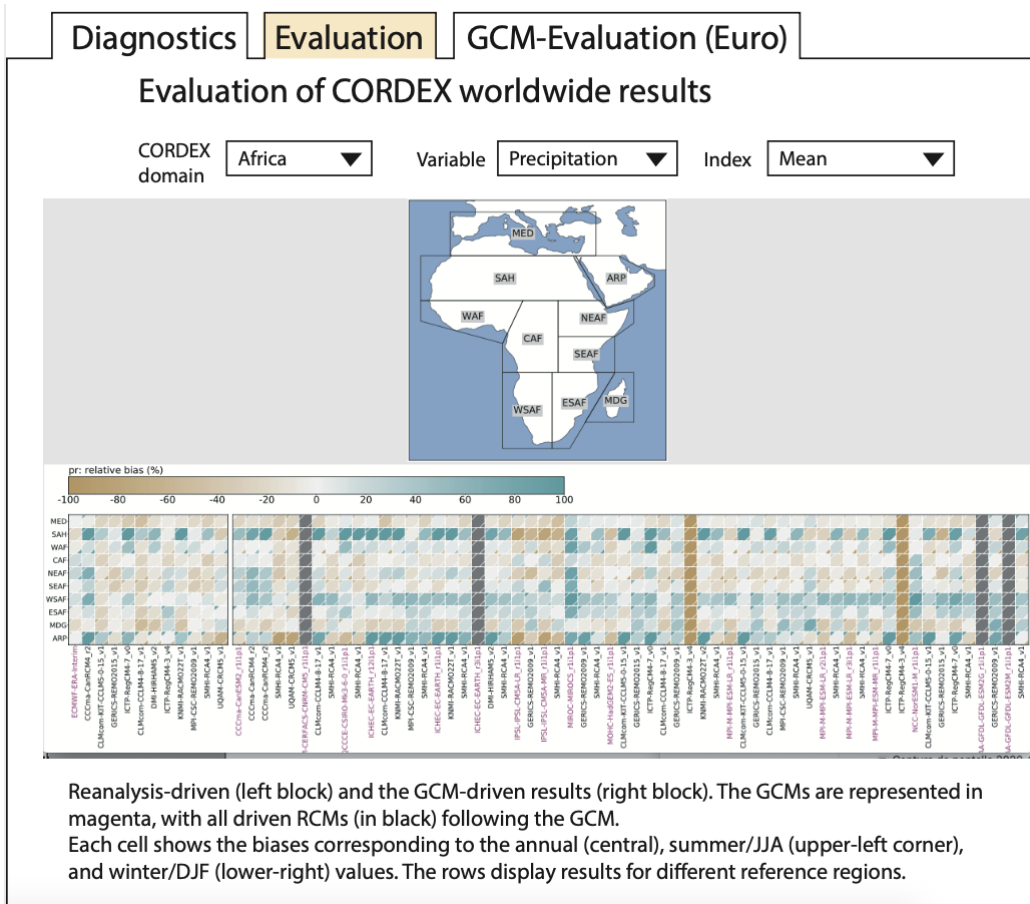


Figure 6. Mockup of the diagnostics and evaluation web page, showing the options for the evaluation component.

5. Summary

Diagnostic and evaluation results have been computed for each CORDEX worldwide domain for six (precipitation, mean/minimum/maximum daily temperature, surface wind, specific humidity) and three (precipitation, temperature and surface wind) variables, respectively. The simple evaluation framework is designed to provide key user-friendly evaluation information uniformly across regions for the users of the C3S worldwide CORDEX data. Summary information of the full ensembles available in the different domains have been produced and described in this document, so the users can quickly analyze the homogeneity of the dataset for the different variables and aspects of interest and get some preliminary information for sub-ensemble selection. User guidelines and a proposal to show the information produced to the users using a simple web page (where users can select the domain, variable and index of interest) are also presented.



The original gridded results computed in this contract and used to calculate the regional summaries described in this report are available for the sake of reproducibility and reusability, as described in M34d.3.3.1 (evaluation datasets).

A description of the software used to produce the result is presented in Deliverable D34d.3.2.1 (Evaluation package and unitary tests).

6. References

Coppola, E. et al., 2021: Climate hazard indices projections based on CORDEX-CORE, CMIP5 and CMIP6 ensemble. *Climate Dynamics*, doi:10.1007/s00382-021-05640-z.

Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., and Buontempo, C.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies, *Earth Syst. Sci. Data*, 12, 2097–2120, <https://doi.org/10.5194/essd-12-2097-2020>, 2020.

Iturbide, M. et al. (2020) An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. *Earth Syst. Sci. Data*, 12, 2959–2970, <https://doi.org/10.5194/essd-12-2959-2020>

Maraun, D., Widmann, M., Gutiérrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R. and Wilcke, R.A. (2015), VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3: 1-14. <https://doi.org/10.1002/2014EF000259>

Maraun, D, Widmann, M, Gutiérrez, JM. (2019) Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment. *Int J Climatol*, 39: 3692– 3703. <https://doi.org/10.1002/joc.5877>

Vautard, R. et al. (2020). Evaluation of the large EURO-CORDEX regional climate model ensemble. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD032344. Accepted Author Manuscript. <https://doi.org/10.1029/2019JD032344>