

CORDEX experiment design and archiving specifications for statistical downscaling of CMIP6

 **Draft version**

This document is still in draft version with no official release

Version History

First version, Feb 2023. Open to internal CORDEX-SAT review.

Second version, Apr 2024. Open to comments from the CORDEX community.

The **CO**ordinated **R**egional **c**limate **D**ownscaling **E**xperiment ([CORDEX](#)) was implemented under the auspices of the World Climate Research Program ([WCRP](#)) in order to improve downscaling techniques and their use in the provision of robust regional climate information for application in vulnerability, impacts and adaptation (VIA) studies. Prior CORDEX activities included a validation stream aimed at assessing and improving regional climate downscaling models and techniques, along with a regional projection stream based on downscaling of global projections from the fifth phase of the Climate Model Intercomparison Project ([CMIP5](#)). The experimental protocol for the regional projections was limited to dynamical downscaling (RCM), while Empirical Statistical Downscaling (ESD) projection activities were carried out in a more uncoordinated manner.

This document presents an experimental framework for empirical-statistical downscaling aligned with the CORDEX-CMIP6 experimental protocol for RCMs¹ when possible. The general aim is to downscale a subset of climate scenarios from the CMIP6 ensemble of projections generated within the [ScenarioMIP](#) activity and to make these downscaled regional scenarios publicly available ([Gutowski et al. 2016](#)). In order to facilitate comparability with the RCM projections, the experimental framework focuses on the CORDEX standard domains (and leaves flexibility to domains to define additional experiments depending on resources and domain-specific scientific questions).

For the sake of simplicity, the framework presented here focuses on GCM-driven ESD methods and will be extended to more sophisticated hybrid approaches in future versions (in particular, to hybrid

RCM-ESD approaches, including emulators). The framework consists of the following components.

1. CORDEX domains

14 CORDEX standard domains have been approved and include: *South America, Central America, North America, Europe, Mediterranean, Middle East North Africa, Africa, Central Asia, South Asia, East Asia, Southeast Asia, Australasia, Arctic and Antarctica*. The simulations have to follow the size specifications for the CORDEX domains on regular longitude-latitude coordinates and cover the minimum domain sizes (<https://cordex.org/domains>). However, a flexible choice of additional sub-domains (and experimental configuration) is allowed to accommodate needs and capacities of different communities and groups, preferably in connection with FPS activities or in coordination with ESD topical domain activities. These additional experiments could be defined by the particular regional communities in the different domains and should be coordinated with **the International Project Office for CORDEX (IPOC)** and **CORDEX POCs** for their respective domains.

2. Evaluation experiment: Cross-validation

Groups must first carry out an evaluation (cross-validation) downscaling covering at least the entire 1979-2020 period. A two-fold (split) cross-validation will be performed in the evaluation experiment using two 21-year periods: 1979-1999, 2000-2020. Training will be performed separately for each of the periods (1979-1999 and 2000-2020) providing predictions for the complementary period (2000-2020 and 1979-1999, respectively). The two predicted time-slices will be pooled together into a single cross-validated series covering 1979-2020.

Predictors will be defined using the **ERA5** reanalysis. The choice of the predictors will be decided by each modeling center based on the key large- and regional-scale drivers affecting the domain of interest (or just considering the target variables of interest for MOS-based methods, including bias adjustment and super-resolution methods).

Predictands will be specified for each domain using a reference interpolated observational gridded dataset defined on a regular latitude-longitude grid with 0.25° resolution (or higher, aligned with the RCM experiments) depending on data availability; this must be coordinated with the POCs and other groups within each CORDEX domain to choose a common reference dataset and a common output grid. For instance, the reference dataset for the European domain could be E-OBS² (v25.0e ensemble mean, 0.1° resolution version). As a fallback option the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3) observational input dataset provides gridded observations with global coverage at 1/12° resolution (W5E5-CHELSA³ dataset, 1979-2016; note that in this case the second training period should be adapted to 2000-2016).

The regional CORDEX communities can also carry out additional evaluation experiments (e.g. using other reanalyses and/or observational datasets and/or cross-validation strategies) to complement the common experiment. It is up to these communities to set up their own experiment design for such additional empirical-statistical downscaling depending on resources and domain-specific scientific questions (e.g. observational uncertainty). This should be coordinated with the [CORDEX POCs](#) for their respective domains and [the International Project Office for CORDEX](#) (IPOC).

3. Historical and scenario experiment

The CORDEX **historical** empirical-statistical downscaling must cover the entire **1950-2014** period. **SSP3-7.0** and **SSP1-2.6** scenarios are prioritized and downscaled results must be provided for the entire **2015-2100** period. It is recommended to downscale **additional members** of these two scenarios (to account for internal variability) and the **SSP2-4.5** and **SSP5-8.5** scenarios, if resources allow.

For the selection of GCMs and ensemble members, groups must prioritize those aligned with the CORDEX-CMIP6 experimental protocol for RCMs in the different domains and those used in ISIMIP3⁴:

- GFDL-ESM4 r1i1p1f1
- UKESM1-0-LL r1i1p1f2
- MPI-ESM1-2-HR r1i1p1f1
- IPSL-CM6A-LR r1i1p1f1
- MRI-ESM2-0 r1i1p1f1

The latter are common for all domains and will facilitate intercomparison across domains and with ISIMIP3 climate input products.

4. Output variables

Most ESD models have been designed and tested using daily mean/maximum/minimum temperatures (*tas*, *tasmin*, *tasmax*) and precipitation (*pr*). Therefore, these are the four variables targeted in this experiment at daily resolution (other variables are also welcome). Methods with a stochastic component and those predicting distributions (i.e. downscaling parameters of the daily distributions) could provide the downscaled parameters and/or multiple realizations (following the archiving protocol described in the next section).

5. Documentation, archiving and publishing specifications

CORDEX output for coordinated experiments (standard or FPS domains⁵) shall be published on the Earth System Grid Federation (ESGF) in order to be consistent with the CMIP6 archive and to make the output available to as many users as possible by adhering to the FAIR-principles⁶. All CORDEX-CMIP6 results have to be formatted (cmorised) according to the CORDEX-CMIP6 archive specification document⁷ that provide technical aspects of CORDEX-CMIP6 data format and ESGF directory structure..

In particular, ESD results must include the following DRS elements and corresponding values⁸:

```
project_id: CORDEX
activity_id: ESD
source_type: ESD-PP, ESD-MOS, ESD-Combined
```

Stochastic methods should include the new global attribute `stochastic_output: yes` and use the `version_realization: v1-rM` attribute to indicate the particular realization "M" (e.g. `v1-r1`, ..., `v1-r100` for a sample of 100 realizations).

Moreover, statistical downscaling methods should be fully documented facilitating reproducibility. This information will be requested upon registration before archiving downscaled results in ESGF, following the details provided in the CORDEX GitHub⁹. In particular, contributing institutions will register both `institution_id` and `source_id` (an URL is also requested during registration for full reproducibility, `further_info_url`). The `source_id` is a full identifier (acronym) of the CORDEX ESD method (e.g. DeepESD-5E) which should be generated indicating a code for the ESD method (`esd_method_id`) and another code for the particular configuration used for training (`training_id`), separated by a dash "-" (e.g. DeepESD refers to the particular deep learning method used¹⁰ and the suffix 5E indicates the particular configuration used to train the model). These sub-elements give rise to new mandatory global attributes for CORDEX-CMIP6 ESD linking these IDs to free text details as described in the table below.

CORDEX-CMIP6 ESD global attribute	description	Example (for DeepESD-5Ea)
<code>esd_method</code>	Free text describing the ESD method used (in particular if different from	

	that described in the reference).	Deep convolutional neural network including 5 convolution and one dense layer, with ReLU activation functions.
esd_method_id	A short, valid identifier of the ESD method	DeepESD
training	Free text fully describing the training methodology and data used for this training configuration, including predictor / predictand datasets, variables, predictor domain, etc.	Trained on ERA5 and E-OBS data. Predictor area 15W-30N to 50E-70N, using Z, T, U, V and Q at 850, 700 and 500mb. Input data is standardized at a gridbox level using the mean/std of the reanalysis in the training period. In the downscaling phase the GCM predictors are bias adjusted (month by month) against the reanalysis variables using a simple linear scaling (adjusting the mean).
training_id	A short, valid identifier of the training configuration. Avoid long IDs trying to summarize all training details. No dashes are allowed in this field.	5E

Downscaling plans must be reported to the CORDEX simulation status repository¹¹.

1. <https://cordex.org/experiment-guidelines/cordex-cmip6/experiment-protocol-rcms/> ↩
2. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-gridded-observations-europe> ↩
3. <https://data.isimip.org/10.48364/ISIMIP.836809.3> (see *.300arcsec_global_daily, corresponding to 1/12 deg). ↩
4. <https://protocol.isimip.org/protocol/ISIMIP3b/#climate-forcing> ↩
5. Note that statistical downscaling results from uncoordinated ESD efforts should not be published under the project_id 'CORDEX'. ↩
6. <https://www.go-fair.org/fair-principles> ↩
7. <https://doi.org/10.5281/zenodo.10961068> ↩

8. Note that this is an initial core definition which will evolve to reflect new science. The controlled vocabularies can be discussed and updated in the GitHub CORDEX repository:

<https://github.com/WCRP-CORDEX/cordex-cmip6-cv> ↩

9. <https://github.com/WCRP-CORDEX/cordex-cmip6-cv> ↩

10. <https://doi.org/10.5194/gmd-15-6747-2022> ↩

11. <https://cordex.org/2022/09/05/information-on-cordex-cmip6-simulation-plans-and-status-is-available-2> ↩