

Data Reference Syntax (DRS) for bias-adjusted CMIP5 simulations

Levavasseur, G.¹ and Noël, T.²

¹ Institute Pierre Simon Laplace (IPSL, France) | ² The Climate Data Factory (TCDF, France)

June 06th, 2018

This document specifies the Data Reference Syntax (DRS) elements for managing bias-adjusted CMIP5 simulation data. The document includes file naming conventions and metadata as NetCDF attributes. The DRS elements are allowed to either assume values defined by Controlled Vocabularies (CV), or free text, or free text with build rules.

1. Bias-adjustment DRS sub-elements

It is proposed that the DRS for bias-adjusted CMIP5 simulation data should be as close as possible to the [CMIP5 archiving specifications](#). If needed, the bias-adjusted CMIP5 DRS could also include Bias-correction information and grid-label information are added to the CMIP5 DRS following some the guidelines of [CORDEX_Adjust](#) and [CMIP6](#) archives design.

Three bias-correction DRS sub-elements are introduced:

- ***bc_name*** is an identifier for the applied bias-correction method that includes a dash-separated combination of acronyms for the institute and the bias-correction method (e.g. SMHI-DBS43, LSCE-IPSL-CDFt, UCAN-EQM etc.).
- ***obs_name*** is an acronym for the observation/reanalysis datasets used as reference data for bias adjustment. Presently, there is no unique CV for regional observational datasets, and acronyms for observations have to be defined in consultation with institutions responsible for the observational products.
- ***ref_period*** is the reference or calibration period in YYYY-YYYY format (e.g. 1971-2000 or 1981-2010).

These 3 sub-elements are attached using dashes ("-") to the CMIP5 DRS creating a new element *called* ***bias_adjustment***. The new ***bias_adjustment*** element is a bit long but provides all necessary information about the bias adjustment methodology.

One grid label DRS element is introduced:

- *grid_label* is the information of the regridded data used for bias-adjustment.

Example:

A CMIP5 simulation is bias-adjusted by IPSL-CDFt method using WFDEI as a reference observational dataset for the 1981-2005 period, the *bias_adjustment* becomes IPSL-CDFt-WFDEI-1981-2005 (*i.e.*, *bc_name-obs_name-ref_period*). Note that dashes in sub-elements can be dropped for consistency and easy automatic parsing.

2. File names, variable names, and NetCDF attributes

The names of the files in the CMIP5-Adjust project are made up of the CMIP5 DRS elements, CMIP6 DRS and CORDEX-Adjust DRS with the changes described above. The elements are separated by underscores (“_”) and must appear in the following order:

```
VariableName_Frequency_GCModelName_CMIP5ExperimentName_CMIP5EnsembleMember_GridLabel_BiasAdjustment_Frequency[_StartTime-EndTime].nc
```

In order to avoid any confusion and clearly distinguish original and bias-adjusted CMIP5 simulation data, it has been decided to follow an approach used in CMIP5 ([CMOR Table Amon: 2-D bias-corrected fields on atmospheric grid](#)) for the adjustment of decadal experiment results by appending ‘Adjust’ to the variable name DRS elements in file names and in NetCDF files, for instance *pr* variable becomes *prAdjust* (used also in [ISI-MIP](#)).

The long variable names (*long_name* NetCDF attribute) have to be also modified by pre-pending “Bias-Adjusted”, for instance Near-Surface Air Temperature becomes Bias-Adjusted Near-Surface Air Temperature.

One issue which has to be taken into account is a situation when the reference/calibration period includes years from both historical and scenario experiments. In this case a different bias-adjusted historical simulation is created for each scenario experiment instead of the same input one for all scenarios. It is proposed to use only the scenario acronyms (rcp26, rcp45 and rcp85) in file names for the entire bias correction period even for the historical experiment (until 2005 in the CMIP5 and CORDEX). No changes are needed for the ERA-Interim driven CORDEX evaluation experiment.

Example:

input files containing original uncorrected model results:

```
tas_day_IPSL-CM5A-MR_rcp45_r1i1p1_20060101-20551231.nc
```

bias-adjusted file (new/modified information in blue)

tasAdjust_day_IPSL-CM5A-MR_rcp45_r1i1p1_gr1_IPSL-CDFt-WFDEI-1989-2005_day_20060101-20551231.nc

gr1 is the metadata for the regridding method.

3. Time periods for each data file

Bias-corrected daily CMIP5 data sets have to include the same years (time records) as requested for the input CMIP5 files (see 5.4 “Time periods for each data file” in [CORDEX archiving specifications](#)).

4. Global attributes

A number of global attributes have to be copied from input CMIP5 files and some of them have to be modified. Also, a number of new global NetCDF attributes have to be added to bias-adjusted CMIP5 data sets. See attached table [CMIP5-Adjust DRS attributes](#).

5. DRS directory structure

The data have to be managed with the following directory structure:

```
<activity>/
  <product>/
    <institute>/
      <model>/
        <experiment>/
          <frequency>/
            <modeling_realm>/
              <cmor_table>/
                <ensemble>/
                  <grid_label>/
                    <bias_adjustment>/
                      <version>/
                        <variable>/
```

Note that the upper 2 levels <activity>/<product> are fixed to CMIP5-Adjust/bias-adjusted-output.

Examples of bias-adjusted CMIP5 netcdf files (new information in blue)

IPSL-CM5A-MR simulation interpolated at 0.5° and bias-adjusted by IPSL using CDFt v2.2 and the WFDEI daily gridded observational dataset, 1979-2005 period as reference.

```

tasAdjust_day_IPSL-CM5A-MR_rcp85_rlilpl_gr1_IPSL-CDFT22s-WFDEI-1979-2005_1951
0101-19651231 {
dimensions:
    time = UNLIMITED ; // (5475 currently)
    lat = 360 ;
    lon = 720 ;
    bnds = 2 ;
variables:
    double time(time) ;
        time:bounds = "time_bnds" ;
        time:units = "days since 1850-01-01 00:00:00" ;
        time:calendar = "365_day" ;
        time:axis = "T" ;
        time:long_name = "time" ;
        time:standard_name = "time" ;
    double time_bnds(time, bnds) ;
    double lat(lat) ;
        lat:bounds = "lat_bnds" ;
        lat:units = "degrees_north" ;
        lat:axis = "Y" ;
        lat:long_name = "latitude" ;
        lat:standard_name = "latitude" ;
    double lat_bnds(lat, bnds) ;
    double lon(lon) ;
        lon:bounds = "lon_bnds" ;
        lon:units = "degrees_east" ;
        lon:axis = "X" ;
        lon:long_name = "longitude" ;
        lon:standard_name = "longitude" ;
    double lon_bnds(lon, bnds) ;
    double height ;
        height:units = "m" ;
        height:axis = "Z" ;
        height:positive = "up" ;
        height:long_name = "height" ;
        height:standard_name = "height" ;
    float tasAdjust(time, lat, lon) ;
        tasAdjust:standard_name = "air_temperature" ;
        tasAdjust:long_name = "Bias-Adjusted Near-Surface Air
Temperature" ;
        tasAdjust:units = "K" ;
        tasAdjust:original_name = "tasAdjust" ;
        tasAdjust:cell_methods = "time: mean" ;
        tasAdjust:cell_measures = "area: areacella" ;
        tasAdjust:history = "2018-04-27T23:30:15Z altered by CMOR:
Treated scalar dimension: \'height\'. 2018-04-27T23:30:23Z altered by CMOR:
Converted type from \'d\' to \'f\'. " ;
        tasAdjust:coordinates = "height" ;
        tasAdjust:missing_value = 1.e+20f ;
        tasAdjust:_FillValue = 1.e+20f ;

```

```

tasAdjust:associated_files = "baseUrl:
http://cmip-pcmdi.llnl.gov/CMIP5/dataLocation gridspecFile:
gridspec_atmos_fx_IPSL-CM5A-MR_rcp85_r0i0p0.nc areacella:
areacella_fx_IPSL-CM5A-MR_rcp85_r0i0p0.nc" ;

// global attributes:
:grid_label = "gr1" ;
:grid_resolution = "0.5°" ;
:grid_interpolation_method = "remapbic" ;
:grid_info = "WATCH" ;
:bc_method = "Cumulative Distribution Function Transform (CDFT)
method - Vrac, M., T. Noël, and R. Vautard (2016), Bias correction of
precipitation through Singularity Stochastic Removal: Because occurrences
matter, J. Geophys. Res. Atmos., 121, 5237-5258, doi:10.1002/2015JD024511." ;
:bc_method_id = "IPSL-CDFT22s" ;
:bc_observation = "Watch Forcing Data methodology applied to
ERA-Interim data (WFDEI) Weedon et al., 2011, J. Hydromet., 12, 823-848,
doi:10.1175/2011JHM1369.1" ;
:bc_observation_id = "WFDEI" ;
:bc_period = "1979-2005" ;
:bc_info = "IPSL-CDFT22s-WFDEI-1979-2005" ;
:input_tracking_id = "02e64e77-8b3d-410b-af8c-0a0e8bbf4a81" ;
:input_institution = "IPSL (Institut Pierre Simon Laplace,
Paris, France)" ;
:input_institute_id = "IPSL" ;
:institution = "IPSL (Institut Pierre Simon Laplace)" ;
:institute_id = "IPSL" ;
:experiment_id = "rcp85" ;
:source = "IPSL-CM5A-MR" ;
:model_id = "IPSL-CM5A-MR" ;
:forcing = "Nat,Ant,GHG,SA,Oz,LU,SS,Ds,BC,MD,OC,AA" ;
:parent_experiment_id = "historical" ;
:parent_experiment_rip = "r1i1p1" ;
:branch_time = 2005. ;
:contact = "robert.vautard@lsce.ipsl.fr Data manager : Robert
VAUTARD (LSCE) - Sebastien DENVIL (IPSL) - Thomas NOEL (TCDF)" ;
:comment = "Collaboration between IPSL and TCDF" ;
:references = "P.-A. Michelangeli, M. Vrac, H. Loukos.
\'Probabilistic downscaling approaches: Application to wind cumulative
distribution functions\'. Geophysical Research Letters, 36, L11708,
doi:10.1029/2009GL038401, 2009" ;
:initialization_method = 1 ;
:physics_version = 1 ;
:product = "bias-adjusted-output" ;
:experiment = "RCP8.5" ;
:frequency = "day" ;
:creation_date = "2018-04-27T23:30:23Z" ;
:history = "2018-04-27T23:30:23Z CMOR rewrote data to comply
with CF standards and CMIP5-Adjust requirements." ;
:Conventions = "CF-1.4" ;
:project_id = "CMIP5-Adjust" ;

```

```
        :table_id      =      "Table      day      (August      2016)
2522b7852917861b4bd16b79eb21d790" ;
        :title = "IPSL-CM5A-MR model output prepared for CMIP5-Adjust
RCP8.5" ;
        :parent_experiment = "historical" ;
        :modeling_realm = "atmos" ;
        :realization = 1 ;
        :cmor_version = "2.9.3" ;
        :tracking_id = "34da2f94-04a6-4f40-bec9-ff4a5442d4c5" ;
}
```

NetCDF Attribute	Status	Value	Example
experiment_id	Unchanged		
experiment	Unchanged		
model_id	Unchanged		
foucing	Unchanged		
parent_experiment_id	Unchanged		
parent_experiment_rip	Unchanged		
branch_time	Unchanged		
initialisation_method	Unchanged		
physics_version	Unchanged		
realisation	Unchanged		
parent_experiment	Unchanged		
product	Modified	Fixed	bias-adjusted
project_id	Modified	Fixed	CMIP5-Adjust
contact	Modified	Contact information of institution that is responsible for bias-adjusted datasets	
institution	Modified	Full name of institution that is responsible for bias-adjusted datasets	
institute_id	Modified	Short acronym for the institution responsible for bias-adjusted data sets	
creation_date	Modified	Creation date of the dataset	
tracking_id	Modified	New UUID to generate	
bc_method	New	Full name of the bias correction methods applied and its references	
bc_method_id	New	Acronym of the bias correction methods (i.e., bc_name DRS sub-element)	
bc_observation	New	Full name of the observation data used as a reference for bias correction and its references	
bc_observation_id	New	Acronym for the observation data used as a reference for bias correction (i.e., obs_name DRS sub-element)	
bc_period	New	Reference period used for bias correction (i.e., ref_period DRS sub-element)	
bc_info	New	Combination of bc_method_id, bc_observation_id and bc_period separated by dashes (i.e., bias_adjustement DRS element)	
input_institution	New	Full name of institution that is responsible for input CMIP5 datasets	
input_institute_id	New	Short acronym for the institution responsible for input CMIP5 datasets	
input_tracking_id	New	UUID from input CMIP5 files	
grid_label	Optional	Acronym for the grid name (i.e., grid_label DRS element)	
grid_resolution	Optional	Grid resolution	
grid_interpolation_method	Optional	Interpolation method	
grid_info	Optional	Additional information on the grid	