**Protocols for CORDEX ESD Experiment 1**

*This protocol document defines the initial experimental framework for the ESD counterpart to the RCM activities within the CORDEX programme.*

*Please register for CORDEX ESD Experiment 1 at the following URL:*

*http://www.csag.uct.ac.za/cordex-esd/*

**General Concept**

The CORDEX-ESD Experiment 1 is a method evaluation experiment; that is, it focuses on the historical period[1] and does not use data from GCM climate change simulations[2]. The experiment seeks to engage as broadly as possible with the ESD community and provide opportunities to address dual objectives: characterizing optimal ESD skill under the constraints of intended transferability to using GCM predictors, and exploring the ESD limits in resolving the signal of the large-scale conditioning of local scale climate.

More specifically, the primary objectives are:

(1) to delineate the relative skill attributes of different statistical downscaling approaches when applied to a common source of predictors and predictands, and
(2) to assess the added-value[3] that ESD can provide from coarse scale predictor fields taken from GCMs.

With that in mind, the protocol permits a relatively straightforward application of an ESD method with the intent to include participation from as many of the ESD community as practical. The protocol gives flexibility in the choice of predictor variables within the limits of a common data source for the predictors, their frequency and spatial resolution, but is restricted to a common predictand dataset[4].

Participants in CORDEX-ESDM experiment 1 provide their best downscaling for the target locations and variables in order to establish a baseline for a method's skill and contribute to the first primary objective. These results will serve as a reference for subsequent analyses or future targeted experiments on a range of issues (e.g. to understand why different choices give different answers - e.g., some past sources of uncertainty).

Demonstration of added-value (the second primary objective) in a climate projection context requires downscaling results that use predictors that GCMs can produce (eg CMIP3/5/6). The experiment requires a complete description of the predictors and the method details for further processing and analysis. This metadata should be provided together with the downscaled results based on a set of templates for the experiment. Participants can submit results in netCDF and ASCII but files will be converted to netCDF, which will be the file of record (netCDF with attributes holding the metadata).

---

[1] 'perfect predictor experiment'
[2] while intending to do so in future work
[3] We define added value here in broad terms.
[4] The CLARIS LPB daily station or gridded WATCH data over the CLARIS LPB region (Argentina). The variables are: tasmax, tasmin, and precipitation (pr).

Within Experiment 1 participants may submit more than one set of results (subject to the protocols below) if, for example, they use more than one method or if they apply their method in more than one way (e.g. with a different mix of predictors).

The following protocol must be strictly adhered to in order to participate in the CORDEX experiment, in order to maximize the robustness of achieving the experiment goals. All the details in 1.1-1.5 must be enclosed in the metadata provided by the results according to the controlled vocabulary provided in the appendix.

**Experiment protocols:**

1. Predictors
   1.1 Primary source: ERA-Interim Reanalysis (ERA-I) 1979-2013.
   1.2 Variables: any variable(s) contained in the ERA-I, or which may be derived from ERA-I data without additional sources, and for which the complement is available from GCMs.
   1.3 Spatial resolution: Any native or post-processed resolution of ERA-I, at participants' choice based on their best judgment.
   1.4 Temporal resolution: Any native or post-processed resolution of ERA-I, at participants' choice based on their best judgment.
   1.5 Additional predictors source: Participants can additionally repeat the downscaling using other reanalysis data sets for their predictors, but only *after* completing the comparable experiment tier with ERA-I predictors. If additional runs are undertaken then MERRA should be the next reanalysis after ERA-I, after which other reanalyses may be accommodated.

2. Predictands
   2.1 Daily station data from the CLARIS LPB dataset for the CORDEX identified portion of South America (See map)
   2.2 WATCH gridded data for the CORDEX identified portion of South America (See map)
   2.3 The variables of Tmin (tasmin), Tmax (tasmax), and precipitation (pr).

3. Time range of data to be used: 1979-2011 (see item 5 for subset periods)

4. Results submission
   Submission of results will be faciliated by a web based submission service currently under development. The submission service will facilitate the accurate submission of data and meta-data for each results sets. Results will need to be submitted in one of two formats:
   4.1 Text files in the same format as the text version of the predictand data (CLARIS) described in Appendix 1.
   4.2 NetCDF files following the structure describe in Appendix 1

5. Experiment tiers

5.1 Tier 1 - Baseline
   5.1.1 Calibration period: 1979-1995 (17 years)
   5.1.2 Validation period: 1996-2006 (12 years)

5.2 Tier 2 – K-Folding

5.2.1 Five sub-periods within the 25-year time span of 1979-2003
The 5 sub-periods are:

1979-1983 1984-1988 1989-1993 1994-1998 1999-2003

5.2.2 Use 4 of 5 sub-periods for calibration, validate on the fifth. Reserve each of the sub-periods in succession to produce 5 sets of calibration/validation.
5.2.3 Tier 2.1 - Weak stationarity test: For each calibration/validation set in 5.2.2, additionally validate on 2004-2012, thus applying 5 different calibrations to the same validation period.

6. Analysis and diagnostics.

6.1 File of daily data submitted as per the requirements of the data protocol document.
6.2 All submitted downscaled data will be passed through the CORDEX standardized quality control and diagnostics code, and the results made available to all registered CORDEX participants.
6.3 Participants are welcome undertake their own diagnostics that are specific to evaluation their particular method, with summary results submitted to the CORDEX results archive

7. Data policy

This is defined in the data policy document and is as for the CORDEX-RCM activities. Note that for downscaled results to be within the CORDEX framework, they must be submitted to the CORDEX archive as per the above protocol, and the data will become public to registered CORDEX participants once the quality-control checks are passed.

**APPENDIX 1: Controlled vocabulary and data structure for downscaled results**

***Controlled Vocabulary***

A controlled vocabulary (CV) is used to ensure that names of models, methods, experiments, datasets, and variables are used consistently across a project. This allows for comprehensive collation, analysis and comparison of diverse submissions with minimized risk of confusion or misrepresentation of results.

The experiments will therefore follow the following CV for:

Variable names:
*tasmin, tasmax, pr* (as per the CMIP5 specification)

GCM/reanalysis:
*ECMWF-ERAINT* (as per CORDEX and CMIP5)

Frequency:
*mon, day, 3hr, 6r* (only *day* used at this point)

Development of a full CV for statistical downscaling is ongoing, coordinated currently by NCPP but including participation from CORDEX and COST-VALUE. This CV may be deployed at the submission stage in order to describe each ESD method in a structured way. Details to follow.

***Data formats***

In order to submit data to the experiments described above, certain standards need to be met. CSAG will provide the data submission service and as such will support submission of data in two formats; netCDF or the plain text (ASCII) format that is used for the Claris LPB station data. If submitting plain text, the submission form provided by CSAG will enable you to enter extra meta-data (see below) not supported by the plain text format.

**Only the validation period data need be submitted.** If data is submitted for the complete calibration and validation period only the calibration periods will be extracted for analysis.

***Data Reference Syntax and formatting***

For reference and for those who wish to submit netCDF format data files, the following DRS (Data Reference Syntax - filenaming scheme) and netCDF file structure will be used.

For CORDEX ESD the following variation of the CORDEX RCM DRS will be used:

*VariableName_PredictandData_CalibrationData_ESDExperimentName_GCMModelName_CMIP5ExperimentName_CMIP5EnsembleMember_ESDModelName_ESDVersionID_Frequency_StartTime_EndTime.nc*

The DRS elements are described below. DRS elements that are required as NetCDF global attributes are shown in square brackets after the element name:

VariableName:
Target variable name[5] (eg. tasmax, tasmin, pr)

[5] Complying with CMIP5 standard output table:
http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.pdf

PredictandData [predictand_data]:
Predictand dataset descriptor (eg. clarislpb-stations, WFDEI)

CalibrationData [calibration_data]:
Model/analysis used for model calibration (eg. ECMWF-ERAINT)

ESDExperimentName [esd_experiment_id]:
ESD experiment name (eg. cordex-esd1-t1, cordex-esd1-t2, cordex-esd1-t2.1)

GCMModelName [driving_model_id]:
Model/analysis used to drive the model (eg. ECMWF-ERAINT)

CMIP5ExperimentName [driving_experiment_name]:
CMIP5 experiment of driving GCM or 'evaluation' for re-analysis

CMIP5EnsembleMember [driving_model_ensemble_member]:
Ensemble member of the driving GCM (eg. r1i1p1)

ESDModelName [model_id]:
A unique name giving the submitter's institution, method and configuration version number)
ESDVersionID [esd_version_id]:
v[1,2,3,4…] to keep track of updates/replacements
Frequency [frequency]:
day (for now)
StartTime:
Start time of the data in the file yyyymmdd[hh[mm]]
EndTime:
End time of the data in the file yyyymmdd[hh[mm]]

So for example:

*tasmax_claris-stations_ECMWF-ERAINT_cordex-esd1-t1_ECMWF-ERAINT_evaluation_r1i1p1_SOMD_v1_day_19960101-20051231.nc*

***NetCDF data structure***

The netcdf data structure should comply with the CF1.6 conventions and follow the following structure.

dimensions:
time = UNLIMITED;
station = 81 ;
str_length = 50 ;
variables:
long time(time) ;
time:units = "days since 1800-01-01" ;
time:calendar = "standard" ;
float pr(time, station) ;
pr:standard_name = "precipitation_flux";

pr:long_name = “Precipitation”;
pr:coordinates = "latitude longitude" ;
pr:units = “mm/day”;
float latitude(station) ;
latitude:standard_name = “latitude”;
latitude:long_name = “station latitude”;
latitude:units = "degrees_north" ;
short longitude(station) ;
longitude:standard_name = “longitude”;
longitude:long_name = “station longitude”;
longitude:units = "degrees_east" ;
float elevation(station) ;
elevation:standard_name = “height”;
elevation:long_name “station altitude above sea level”
elevation:units = "meters" ;
elevation:positive = "up" ;
char id(station, str_length) ;
id:cf_role = ″timeseries_id″;
char name(station, str_length) ;

***Global attributes:***

Whether submitting as netCDF files or text files, all DRS elements For text file submissions this will be facilitate by the data upload service. NOTE: This is a preliminary set of attributes and will be finalized at a later date to align with the ongoing development of the statistical downscaling controlled vocabulary (see above)

| | |
|---|---|
| ‘project_id’: | Must be ‘CORDEX-ESD’ |
| ‘institution’: | Unique name of the source institution |
| ‘predictand_data’ | Shold be ‘claris-lpb’ or ‘watch-wfdei’ |
| ‘calibration_data’ | Should be ‘ERAINT’ |

‘esd_experiment_id’
‘driving_model_id’
‘driving_experiment’
‘driving_model_ensemble_member’
‘predictor_source’ (ERAINT)
‘predictor_names’ (z500, u10, v10, etc..)
‘predictor_spatial-resolution’ (0.5deg, 1.0deg, 2.5deg, etc.)
‘predictor_typical_domain_size’ (eg. 1000km, 5x5 grid boxes, 90W-50W/40S-20S, etc.)
‘predictor_lon-range’ (eg. 95degW-50degW? ‘NA’ for not applicable)
‘predictor_lat-range’ (eg 40degS-20degS? ‘NA’ for not applicable)
‘predictor_number_eofs’ (e.g 15; ‘NA’ for not applicable) ?
‘predictor_temporal-resolution’ (6 hourly, daily, monthly, etc..)
‘predictor_history’ ? (regridded to 3x3 degrees; derived saturation pressure from T(2m), original, etc)
‘predictand_source’ (CLARIS LPB)
‘predictand_history’ ?
‘method_name’
‘method_downscaling_strategy’ (e.g MOS, PP, hybrid)

‘method_model-type’ (analog, regression, weather-generator, neural-nets, CCA, SVD, bias-correction, or a combination of these)
‘method_output’ (deterministic, probabilistic)

‘calibration_period’
‘validation_period’
‘author’
‘date_stamp’

optional:
‘calibration_results’
‘URL’ (the URL of the webpage describing the experiment or our workshops - esentially this document [on the CORDEX web?])
‘reference’
other information.