Accessing CORDEX data via the Earth System Grid Federation ESGF

Grigory Nikulin
Swedish Meteorological and Hydrological Institute
CORDEX-ESGF team

Publication of CORDEX simulations via ESGF is an joint effort of several European groups with a lot support from the IS-ENES2 FP7 project

- Martin Juckes and Stephen Pascoe (BADC, UK)
- Ole B. Christensen (DMI, Denmark)
- Stephanie Legutke and Stephan Kindermann (DKRZ, Germany)
- Sebastien Denvil (IPSL, France)
- Michael Kolax (SMHI, Sweden)
- Prashanth Dwarakanath (NSC-LIU, Sweden)
- and many more ....
Funding of the European and S. Africa nodes
Ensembles of regional climate simulations

Large ensembles of regional climate simulations:

- **PRUDENCE (Europe)** – about 11 RCMs
  [http://prudence.dmi.dk/](http://prudence.dmi.dk/)

- **ENSEMBLES (Europe)** - about 20 RCMs
  [http://ensemblesrt3.dmi.dk/](http://ensemblesrt3.dmi.dk/)

- **NARCCAP (North America)** - 6 RCMs
  [http://www.narccap.ucar.edu/](http://www.narccap.ucar.edu/)

- **CLARIS (South America)** - 5 RCMs

- **ENSEMBLES-AMMA (West Africa)** - 10 RCMs
  [http://ensemblesrt3.dmi.dk/](http://ensemblesrt3.dmi.dk/)

All these projects have substantially contributed to regional climate studies but: different protocols, output lists, standards etc.
Coordinated Regional Downscaling Experiment

sponsored by the World Climate Research Programme
http://wcrp-cordex.ipsl.jussieu.fr/

• Coordinate high-resolution Regional Climate Modeling (unified protocols, experiments, format, variables etc.)
• Provide a set of high-resolution Regional Climate Scenarios for the majority of land-regions of the globe
• Make these data sets available and useable to the impact and adaptation communities

13 CORDEX Domains defined
Establishing CORDEX archive

How to make all CORDEX simulations accessible?

- The first idea was about standard data portals like in ENSEMBLES (http://ensemblesrt3.dmi.dk/)
- Work begun in 2010
  - CORDEX archive specifications (file names, directories etc.)
  - CORDEX variable requirement table (variable names, units etc.)
  - Controlled vocabularies (RCM names, experiments etc.)
- CORDEX specs were built on CMIP5 as much as possible, as well as on experience gained from the NARCCAP and ENSEMBLES projects
- Pretty soon benefits of geographically-distributed, metadata-based searchable archives as the Earth System Grid Federation (ESGF) for CMIP5 become evident
- The main focus is on ESGF now, although other CORDEX data portals also exist
ESGF

- an international collaboration for handling climate science data
- Peer-To-Peer architecture (p2p): geographically-distributed, decentralized database (many data nodes, no central archive)
- metadata-based search (model names, experiments, frequency etc.)

http://esgf.org
CORDEX-ESGF in operation

- CORDEX-ESGF is in operation since mid September 2013 (with a lot of support from the IS-INES2 FP7 project)
- now ESGF provides the same interface for both global (CMIP5) and regional (CORDEX) climate simulations
- 437 users in the CORDEX-ESGF group (3 Apr 2014)
CORDEX Terms of Use (ToU)

- the CORDEX ToU are mainly based on the CMIP5 ToU
- both define two groups: “non-commercial research and educational purposes” and “commercial purposes” (unrestricted)
- all users have to select one of two groups
- ToU are defined by individual CORDEX RCM groups but the situation is more complex compared to CMIP5 (a combination of two ToU)
- a unrestricted RCM downscales a non-commercial AOGCM: RCM output is non-commercial by default but there are exemptions
- MIROC5 is non-commercial but dynamically downscaled data driven by MIROC5 does not inherit the restrictions (a derived product)
- still not so clear with Empirical-Statistical Downscaling (ESD): many different methodologies of different complexity

Quality control

**data quality is the well known problem for all analysing large ensembles of climate simulations**

- typical problems: missing time steps, wrong units, names etc.
- a simulation with a mistake has been downloaded by 100 users: each user is spending time on the same problem
- takes a lot of resources
- an advanced Quality Checker was developed for CORDEX by DKRZ (Hamburg): missing time steps, units, file and variable name, suspicious max/min values and many other aspects
- for the European ESGF data nodes (IS-ENES) all CORDEX simulations have to be quality checked before publishing
- the DKRZ QC is openly available and it’s strongly recommended to all RCM CORDEX groups running QC before distributing their CORDEX simulations by any means (ESGF, direct contacts)
CORDEX results: how to access

At moment (April 2014) CORDEX simulations can be accessible from a few European ESGF data nodes:

- esg-dn1.nsc.liu.se (SMHI-NSC, Sweden)
- esgf-data.dkrz.de (DKRZ, Germany)
- esgf-index1.ceda.ac.uk (BADC, UK)
- cordexesg.dmi.dk (DMI, Denmark)
- esgf-node.ipsl.fr (IPSL, France)
- noresg.norstore.uio.no (Norway)

- It doesn’t matter from which of the data nodes CORDEX simulations can be accessed, links lead to the same files

Other CORDEX data portals (not ESGF):
- Med-CORDEX http://www.medcordex.eu/
- East Asia http://cordex-ea.climate.go.kr/
- South Asia http://cccr.tropmet.res.in/cordex/
ESGF: the first steps

All ESGF nodes have the same user interface
ESGF: the first steps

Select project CORDEX
ESGF: the first steps

**Search facets**

Current Selections

(x) project: CORDEX

Search Categories

- Project
- Institute
- Model
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm
- Variable
- Variable Long Name
- CMIP Table
- CF Standard Name
- Ensemble
- Domain
- Driving Model
- Downscaling realisation
- Data Node

**Model (RCM)**

Current Selections

(x) project: CORDEX

Search Categories

- Project
- Institute
- Model
- ALADIN52 (198)
- CCLIM-4-17 (1781)
- HIRHAM5 (2820)
- RACMO21P (968)
- RACMO22E (1588)
- RACMO22T (980)
- RCA4 (26654)
- RCA4-SN (1150)
- REMO2009 (390)
- WRF331F (66)
- WRF331G (170)
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm

**Driving Model (GCM)**

Variable

Variable Long Name

CMIP Table

CF Standard Name

Domain

Driving Model

- CCCma-CanESM2 (2744)
- CNRM-CERFACS-CNRM-CM5 (3141)
- ECMWF-ERAINT (4372)
- ICHEC-EC-EARTH (13157)
- IPSL-IPSL-CM5A-MR (1210)
- MIROC-MIROCS (1365)
- MOHC-HadGEM2-ES (2466)
- MPI-M-MPI-ESM-LR (3761)
- NCC-NorESM1-M (2055)
- NOAA-GFDL-GFDL-ESM2M (2488)

Downscaling realisation

Data Node
ESGF: the first steps

CORDEX Domains available via ESGF (3 Apr 2014)

- Europe (EUR-11, EUR-22 and EUR-44)
- Africa (EUR-44)
- Middle East and North Africa (MNA-22 and MNA-44)
- South America (SAM-44)
- North America (NAM-44)
- South Asia (WAS-44)
- Arctic (ARC-44)
- Antarctic (ANT-44)

Domain name convention:

- AFR-44, EUR-11 – the native RCM grids (daily, monthly, seasonal)
- AFR-44i, EUR-11i – the regular grids
- (only monthly and seasonal)
Euro-CORDEX downloads (EUR-44)

Most popular variables: precipitation, mean, maximum, minimum temperature and wind

Total number of downloads per variable (SMHI-RCA4)
CORDEX EUR-44 | http://esg-dn1.nsc.liu.se | 20140403
Euro-CORDEX downloads (EUR-11)

Most popular variables: precipitation (favourite), mean, minimum, maximum temperature and specific humidity

Total number of downloads per variable (SMHI-RCA4)
CORDEX EUR-11 | http://esg-dn1.nsc.liu.se | 20140403
Africa-CORDEX downloads (AFR-44)

Most popular variables: precipitation, minimum, mean, maximum temperature and wind

Total number of downloads per variable (SMHI-RCA4)
CORDEX AFR-44 | http://esg-dn1.nsc.liu.se | 20140403
MENA-CORDEX downloads (MNA-44)

Most popular variables: minimum temperature, precipitation, mean, maximum temperature and wind
Downloads per CORDEX domain

EUR-11 – hi-res Euro-CORDEX is a clear favourite

Total number of downloads per CORDEX domain
SMHI-RCA4 | http://esg-dn1.nsc.liu.se | 20140403
The climate4impact portal is able to analyse and display data held at ESGF nodes.

For example all 27 climate extreme indices (ETCCDI) will be implemented and can be calculated for all CORDEX simulations available via ESGF.

There is no need to download CORDEX simulations, install software etc.