CORDEX archiving: achievements, status and perspectives

Grigory Nikulin
Swedish Meteorological and Hydrological Institute (SMHI)

The International Conference on Regional Climate
ICRC-CORDEX 2016
17th - 20th May 2016 Stockholm, Sweden

CORDEX archiving: history

- in 2009-2010 the first CORDEX archiving specifications were built on experience gained from the PRUDENCE, ENSEMBLES and NARCCAP projects (RCMs)
- the first idea was about a central data portal + a number of regional databanks
- soon advantages of geographically-distributed, metadata-based searchable archives as the Earth System Grid Federation (ESGF) for CMIP5 become evident
- CORDEX archiving was adjusted towards CMIP5 as much as possible but still keeping necessary downscaling details (2011-2103)
- the main focus now is on ESGF (September 2013, first CORDEX data available), although other regional CORDEX data portals also exist

CORDEX-data team

development of the CORDEX archiving infrastructure was an joint effort of a number of European groups with coordination by 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)
- Grigory Nikulin and Michael Kolax (SMHI, Sweden)
- Prashanth Dwarakanath and Torgny Faxén (NSC-LIU, Sweden)
- and many more ….

CORDEX Archive Design

2009, 6 pages, including a variable list

CORDEX Archive Design

The following recommendations apply both to CORDEX runs driven by ERA-Interim and CMIP5-GCMs, in terms of : variables to be saved, output format and frequency, file naming and archiving.

(a) Three classes of data are defined:
Core, relevant to all communities: monthly and seasonal means;
Tier 1, relevant to most communities: daily surface/selected upper air data;

2014, present version, 26 pages (no variable list)

CORDEX Archive Design

Version 3.1, 3 March 2014

O.B. Christensen[1], W.J. Gutowski[2], G. Nikulin[3], and S. Legutke[4]

1 Introduction

This document specifies technical aspects of CORDEX archive file and data formats, as well as archive content[a]. It includes a common naming system, the Data Reference Syntax (DRS), which allows the identification of data sets wherever they

CORDEX Variable List

- the list includes 60 variables + static fields (land sea mask etc.)

2009

Tier 1: Daily average output.

1. 2-metre air temperature (K)
2. Maximum 2-metre air temperature (for the 24 hour period preceding the write point)
3. Minimum 2-metre air temperature (for the 24 hour period preceding the

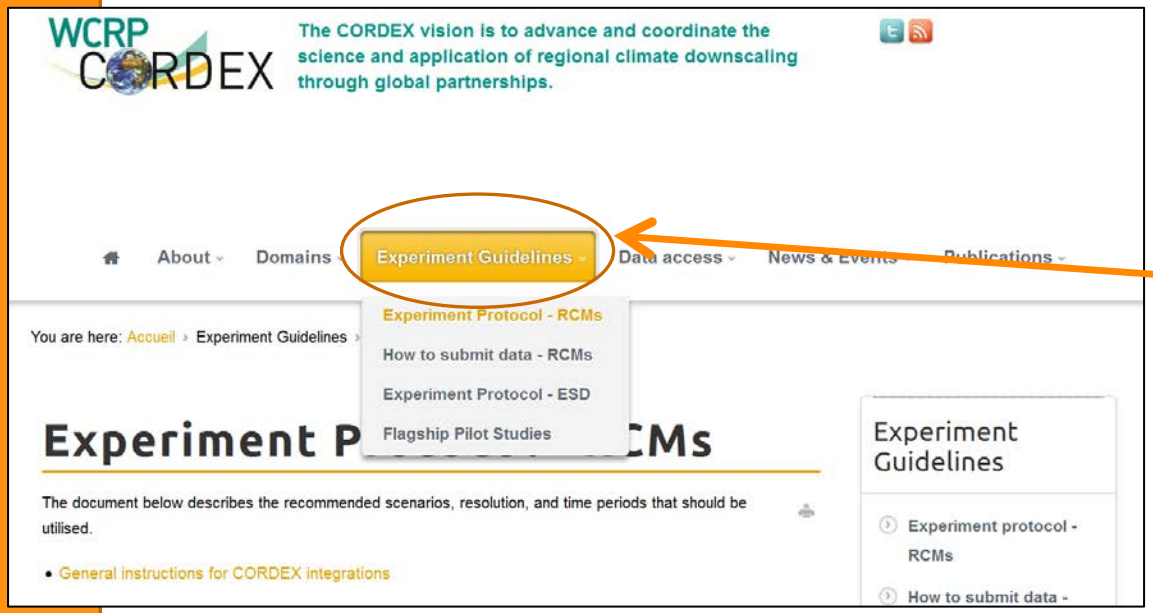
2014

CORDEX: requested variables

		Tier 2		Tier 1		Core					
output variable name	units	frq [1/day]	ag	frq [1/day]	ag	frq [1/mon]	ag	frq [1/sem]	ag	fx	long_name
tas	K	8	i	1	8	1	m*8	1	s*8		Near-Surface Air Temperatur
tasmax	K			1		1	m	1	s		Daily Maximum Near-Surface Air Ter
tasmin	K			1		1	m	1	s		Daily Minimum Near-Surface Air Tem
pr	kg m-2 s-1	8	a	1		1		1			Precipitation
ps	Pa	8	i	1	8						Surface Air Pressure
psl	Pa	8	i	1	8	1	m*8	1	s*8		Sea Level Pressure
huss	1	8	i	1	8	1	m*8	1	s*8		Near-Surface Specific Humidi
hurs	%	8	i	1	8	1	m*8	1	s*8		Near-Surface Relative Humidi

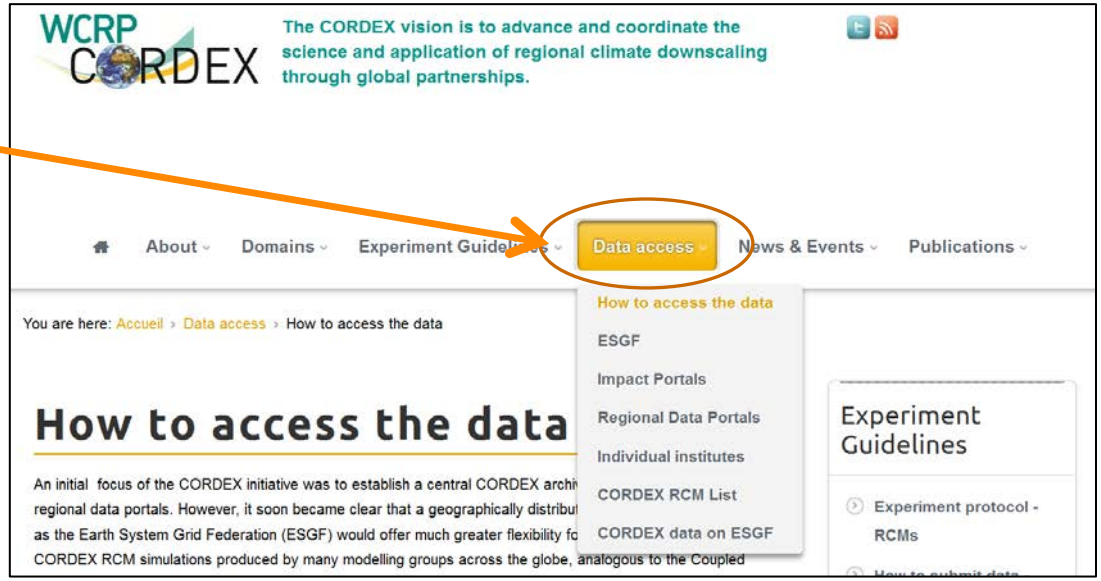
CORDEX web site

www.cordex.org



CORDEX documentation for the RCM and ESD communities

CORDEX Data for users



CORDEX data

CORDEX simulations can be accessed using:

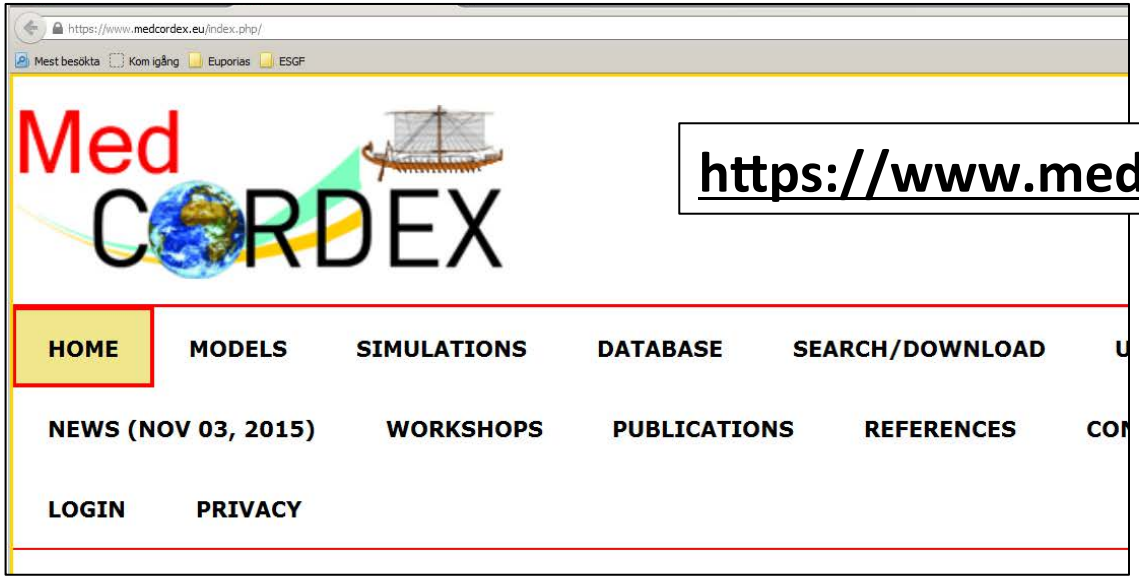
1. **Earth system Grid Federation (ESGF)**
2. **Regional Data Portals (Med-CORDEX, South and East Asia)**
3. **Impact portals (only the ESGF segment)**
4. **Servers at Individual modelling groups**

The CORDEX-ESGF segment was delayed and 3 regional data portals have been already established when first CORDEX simulations were published on ESGF:

- Med-CORDEX
- East Asia CORDEX
- South Asia CORDEX

During downtime of ESGF (June 2015 – January 2016) all portals were operational

Med-CORDEX data portal



https://www.medcordex.eu

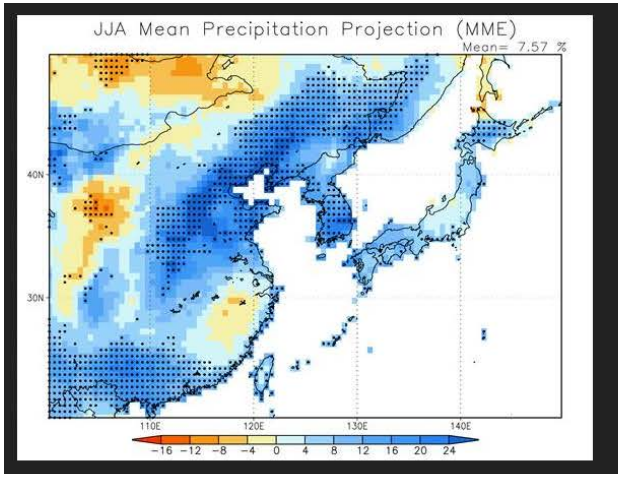
- Med-CORDEX simulations are available only at the portal
- users should be approved (quick response)
- Terms of Use (ToU): non-commercial for all datasets
- There is an advanced search, web-based post-processing (CDO etc.) and different statistics (simulations, variables, files etc.)

The last file was upload on 13th of May 2016. As of today:

- the repository contains **127692** data files for a total of **4928 GB**
- have been downloaded **414055** data files for a total of **13788 GB**
- **87** Med-CORDEX users (out of the **177** registered) and **43** HyMeX users downloaded data

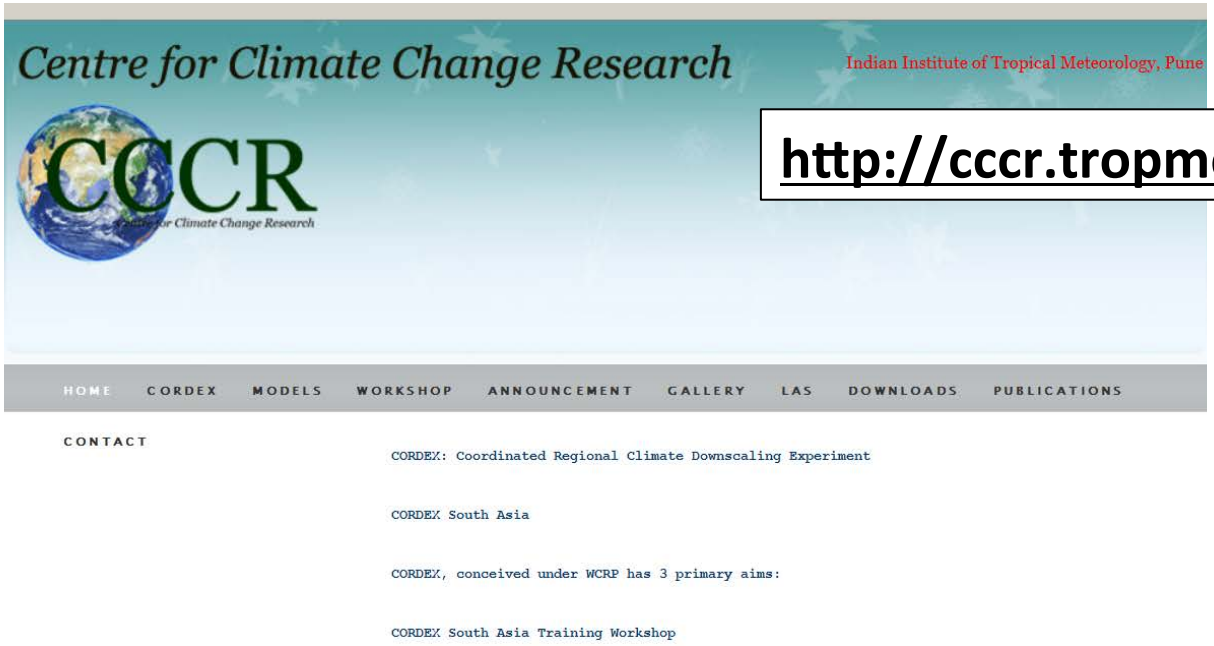
East Asia CORDEX data portal

cordex-ea.climate.go.kr



- register and download
- Terms of Use: non-commercial for all datasets at the portal
- Some EAS-CORDEX data is available on ESGF (unrestricted)
- there are plots with evaluation and projected changes (individual RCMs and MME)
- web-based plotting

South Asia-CORDEX data portal



http://cccr.tropmet.res.in/cordex/

- users should be approved
- Terms of Use not provided (a link to the CORDEX ToU)
- Some South Asia-CORDEX data is available on ESGF (unrestricted)

ESGF

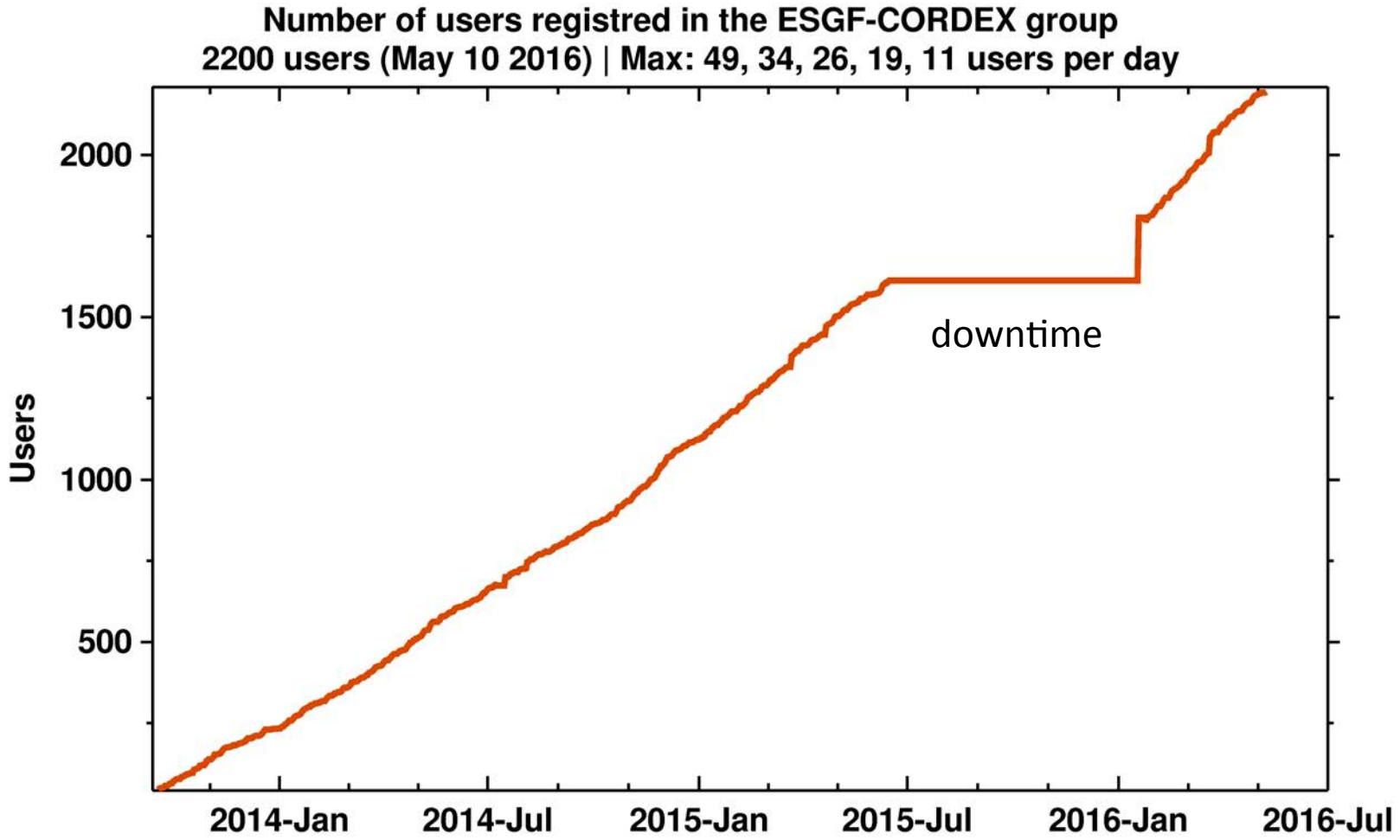
- an international collaboration that develops, deploys and maintains software infrastructure for the management, dissemination, and analysis of model output and observational data
- Peer-To-Peer architecture(p2p): geographically-distributed, decentralized database (many data nodes, no central archive)
- ESGF's primary goal is to facilitate advancements in Earth System Science
- mid June 2015 - January 2016 was downtime due to a security issue
- back to operation since February 2016 (not all nodes)

http://esgf.llnl.gov/

ESGF
Earth System Grid Federation

CORDEX-ESGF

- **CORDEX-ESGF** is in operation since mid September 2013
- CORDEX data can be accessed in the same way as CMIP5, Obs4MIPs, Ana4MIPs
- 2200 users in the CORDEX-ESGF group (10 May 2016)



CORDEX ESGF nodes

- Tier 1 nodes (graphical search interface) , May 2016:
  - esg-dn1.nsc.liu.se (SMHI-NSC-LIU, Sweden)
  - esgf-data.dkrz.de (DKRZ, Germany)
  - esgf-index1.ceda.ac.uk (CEDA, UK)
  - esgf-node.ipsl.fr (IPSL, France)
- Tier 2 nodes (only data, no graphical search interface):
  - cordexesg.dmi.dk (DMI, Denmark)
  - data.meteo.unican.es (Univ. of Cantabria, Spain)
  - esg-cccr.tropmet.res.in (IITM, India, not in operation)

It doesn't matter from which of the index nodes CORDEX simulations can be accessed, a federative system, no central archive, links lead to the same files

How to access the CORDEX data:
"Data access" section on www.cordex.org

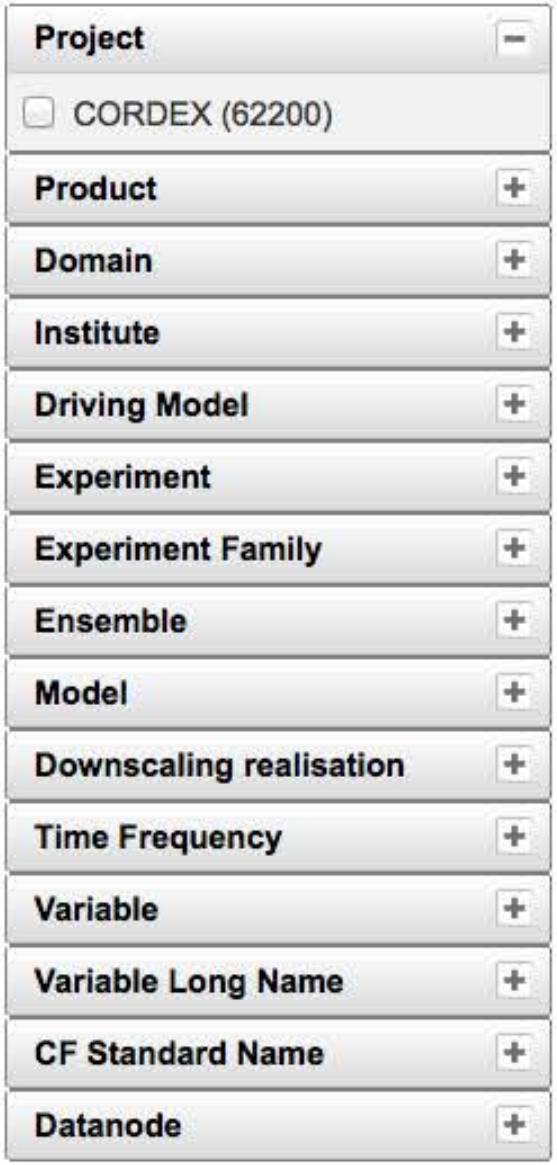
ESGF training workshop for CORDEX Asia

- 4-5 December 2014, WMO Regional Training Centre, Nanjing University of Information Science and Technology, China (funding: WMO, WCRP, MARIS, NUITS)
- 23-25 February 2016, Jeju International Convention Center, Jeju, Korea, hosted by National Institute of Meteorological Sciences at Korea Meteorological Administration (funding: WCRP, NIMS, IS-ENES2)
- Technical workshops: setting up ESGF node, quality control and publication of CORDEX simulations on ESGF

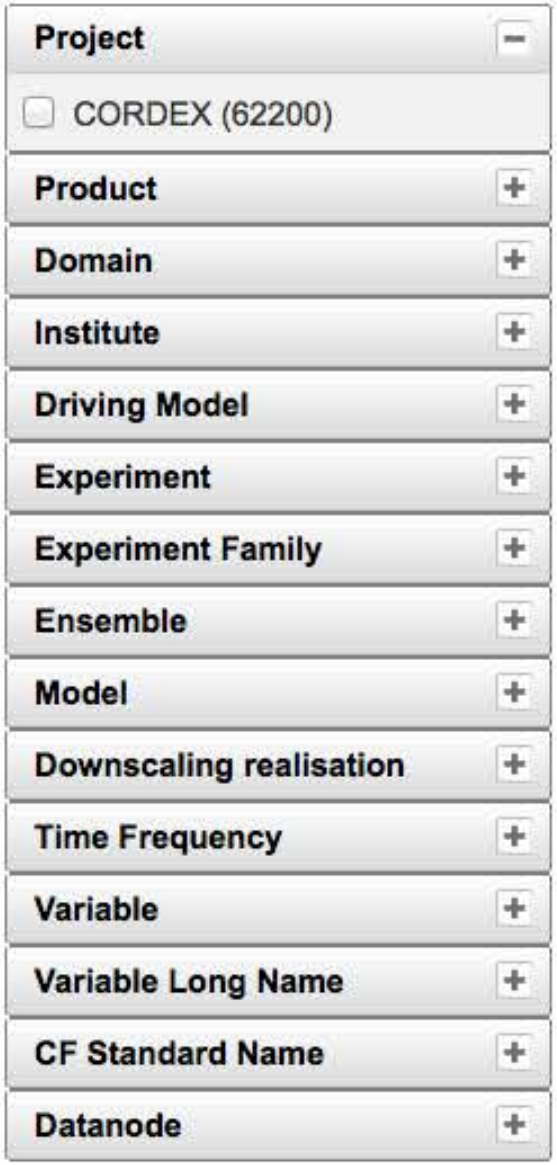
Nanjing, December 2014

Jeju, February 2016

CORDEX ESGF search interface

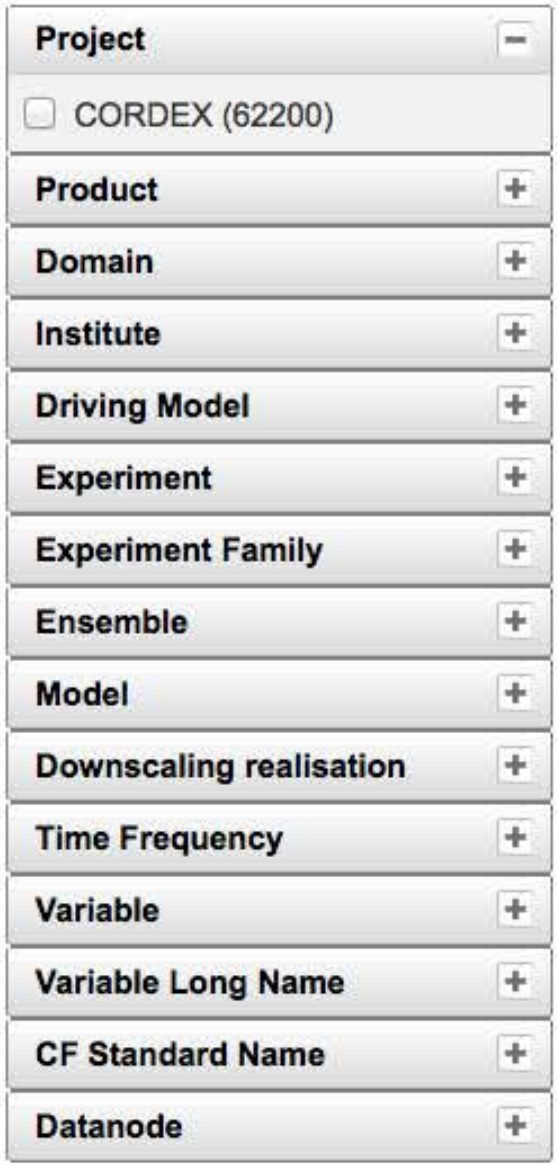


CORDEX ESGF search interface



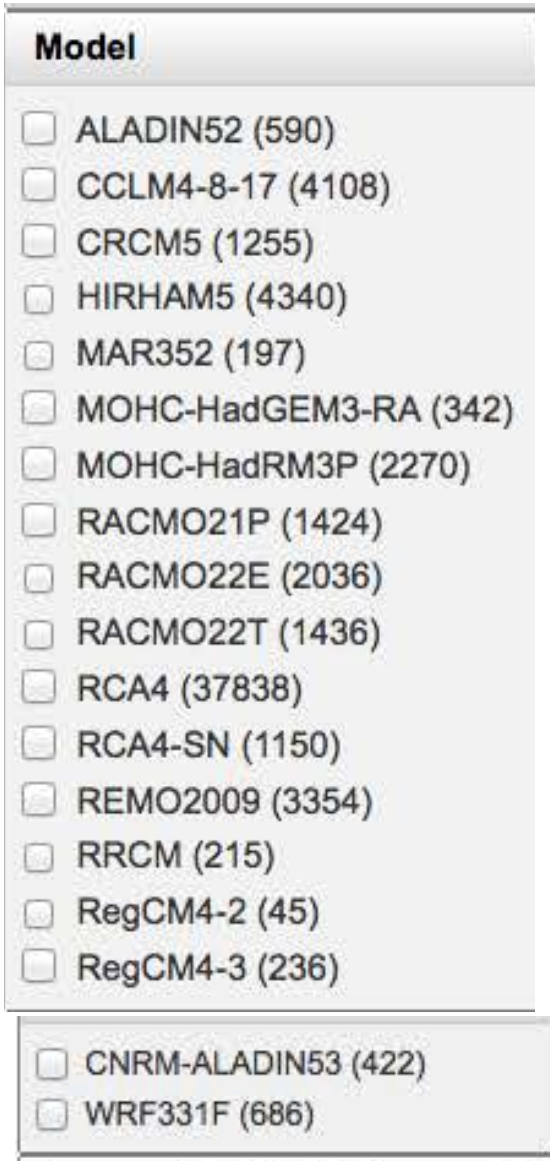
12 of the 14 CORDEX domains

CORDEX ESGF search interface



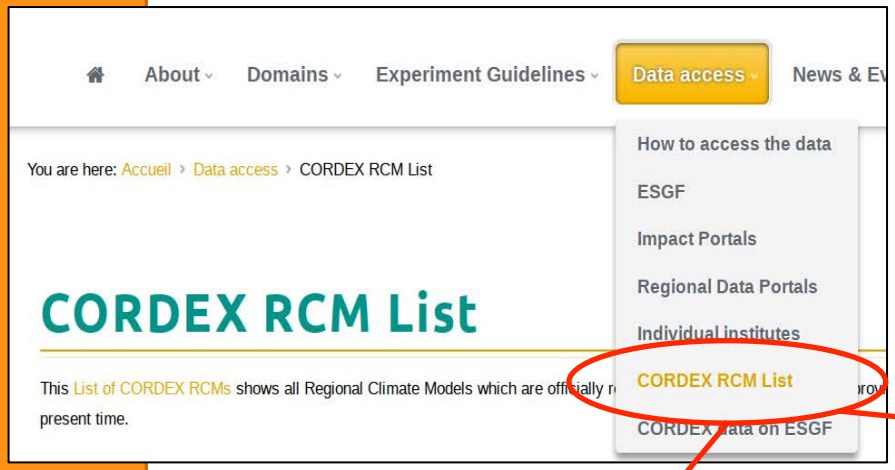
12 of the 14 CORDEX domains

RCMs



CORDEX RCM list

All CORDEX RCM groups have to provide info to **cordex-registration@cordex.org**



1. Full institution name
2. Short institution name (acronym)
3. Contact person and e-mail
4. RCM name (acronym)
5. **Terms of Use**: unrestricted or non-commercial only
6. CORDEX domains you are interested in.

text list

```
# List of CORDEX RCMs (short, machine readable version)
# Do not change this file - it is auto-generated based on the contents
# of the CORDEX RCM registration sheet
# To register CORDEX simulations please contact cordex-registration /at/
# Timestamp:  2016-05-04 14:43:11
# ----------------------------------------------------------------
# model id           institute id            ToU
# ----------------------------------------------------------------
AUTH-LHTEE-WRF321B       AUTH-LHTEE      unknown
AUTH-MC-WRF371           AUTH-MC         unrestricted
AWI-HIRHAM5              AWI             unrestricted
BCCR-WRF331              BCCR            unrestricted
BCCR-WRF331C             BCCR            unrestricted
BCCR-WRF361              BCCR            unrestricted
CCCma-CanRCM4            CCCma           unrestricted
CHMI-ALADIN53            CHMI            non-commercial
CLMcom-CCLM4-8-17        CLMcom          non-commercial
CNRM-ALADIN52            CNRM            non-commercial
CNRM-ALADIN53            CNRM            non-commercial
CNRM-ARPEGE52            CNRM            unrestricted
CNRM-RCSM4               CNRM            non-commercial
CRP-GL-WRF331A           CRP-GL          unrestricted
CSIRO-CCAM               CSIRO           unrestricted
CUNI-RegCM4-2            CUNI            unknown
DHMZ-RegCM4-2            DHMZ            non-commercial
DMI-HIRHAM5              DMI             unrestricted
ENEA-RegCM4-3            ENEA            non-commercial
```

html list

CORDEX Models overview

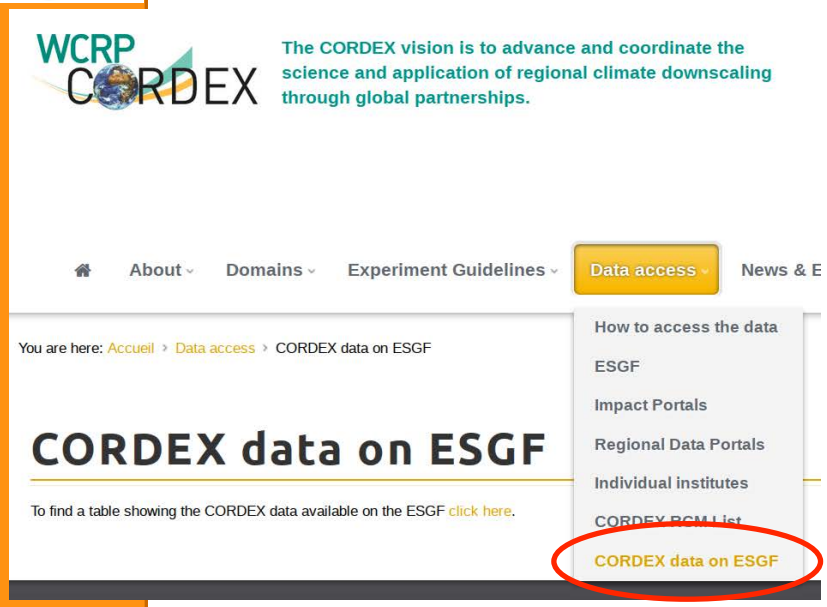
Information automatically generated based on ESGF CORDEX registration sheet.

To register CORDEX simulations please contact cordex-registration /at/ smhi.se

Timestamp: # 2016-05-04 14:43:11 #

Model Name	Institute	Institution Name	
AUTH-LHTEE-WRF321B	AUTH-LHTEE	Aristotle University of Thessaloniki, Laboratory of Heat Transfer and Environment	
AUTH-MC-WRF371	AUTH-MC	Aristotle University of Thessaloniki, Department of Meteorology & Climatology	
AWI-HIRHAM5	AWI	Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research	
BCCR-WRF331	BCCR	Uni Research and the Bjerknes Centre for Climate Research	unrestricted
BCCR-WRF331C	BCCR	Uni Research and the Bjerknes Centre for Climate Research	unrestricted
BCCR-WRF361	BCCR	Uni Research and the Bjerknes Centre for Climate Research	unrestricted

What's available ?



- automatically generated once a week or when a larger number of new datasets published
- many thanks to DKRZ for providing this service

CORDEX - ESGF data availability overview

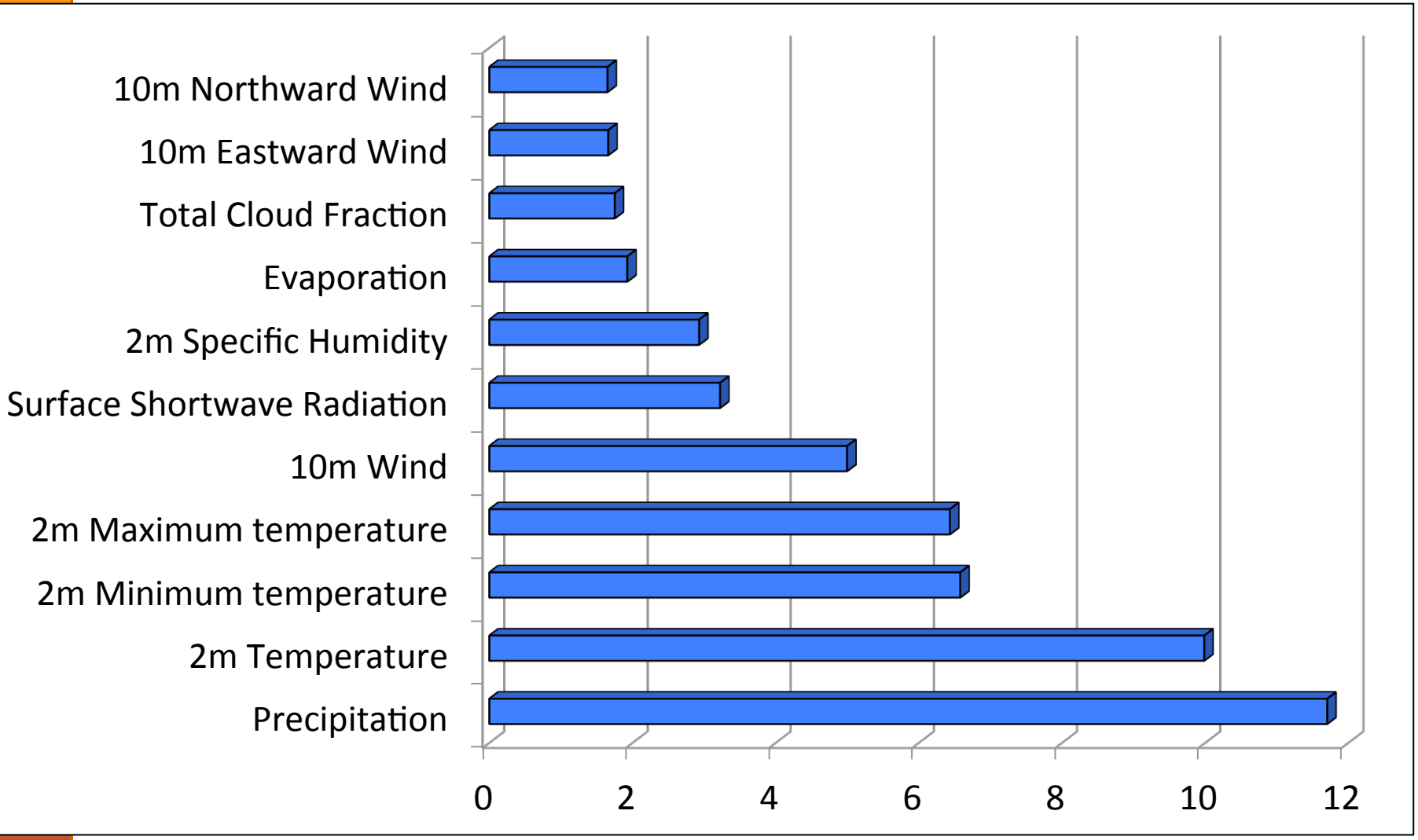
- sorted by domains -

Information retrieved from ESGF index nodes at # 2016-05-04 14:43:21 #

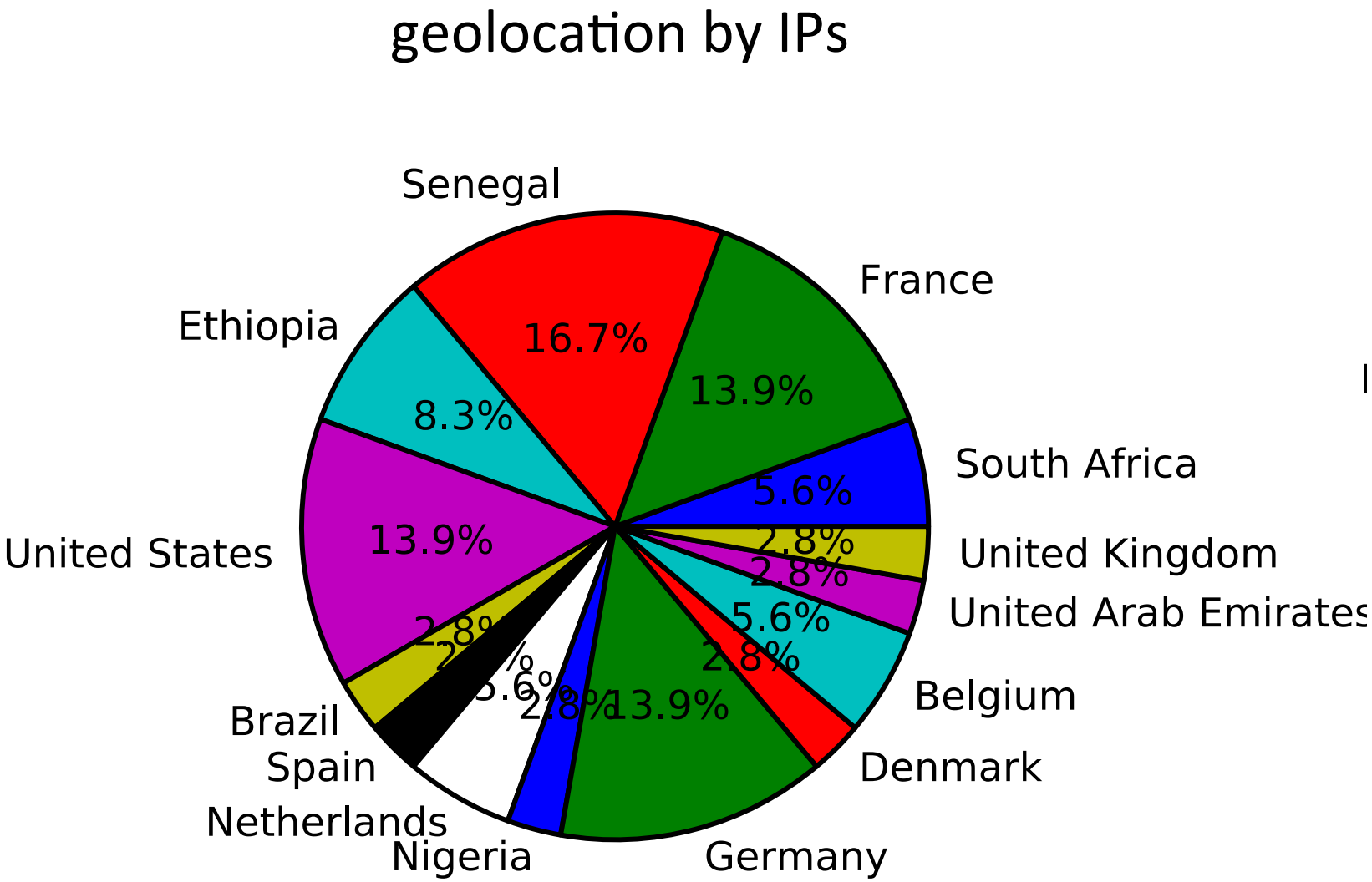
Domain	Model	Driving Model
AFR-44	CCLM4-8-17	CNRM-CERFACS-CNRM-CM5
AFR-44	CCLM4-8-17	ECMWF-ERAINT
AFR-44	CCLM4-8-17	ICHEC-EC-EARTH
AFR-44	CCLM4-8-17	MOHC-HadGEM2-ES
AFR-44	CCLM4-8-17	MPI-M-MPI-ESM-LR
AFR-44	CRCM5	CCCma-CanESM2
AFR-44	CRCM5	ECMWF-ERAINT
AFR-44	CRCM5	MPI-M-MPI-ESM-LR
AFR-44	HIRHAM5	ECMWF-ERAINT
AFR-44	HIRHAM5	ICHEC-EC-EARTH
AFR-44	HIRHAM5	NCC-NorESM1-M
AFR-44	MOHC-HadGEM3-RA	ECMWF-ERAINT
EUR-11	RACMO22E	ICHEC-EC-EARTH
EUR-11	RACMO22E	MOHC-HadGEM2-ES
EUR-11	RCA4	CNRM-CERFACS-CNRM-CM5
EUR-11	RCA4	ECMWF-ERAINT
WAS-44	REMO2009	MPI-M-MPI-ESM-LR
WAS-44i	MOHC-HadRM3P	ECMWF-ERAINT
ARC-44	RRCM	ECMWF-ERAINT
ARC-44i	HIRHAM5	ECMWF-ERAINT
ARC-44i	MAR352	ECMWF-ERAINT

Most popular CORDEX variables

11 variables are about 50% of total CORDEX downloads
(based on 7 CORDEX domains, ~130 simulations, SMHI-RCA4)

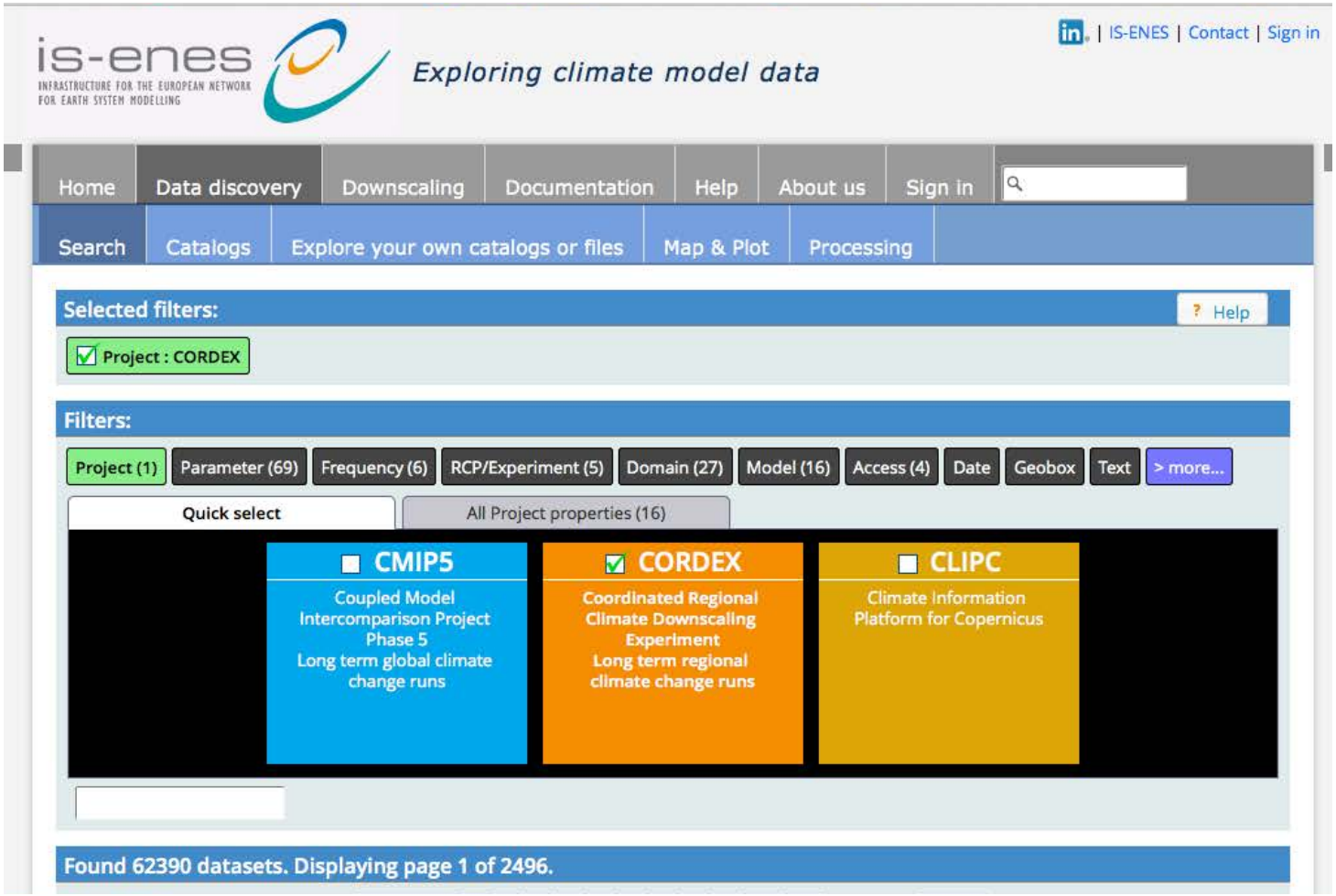


Download statistics for CORDEX-Africa



Impact Portals

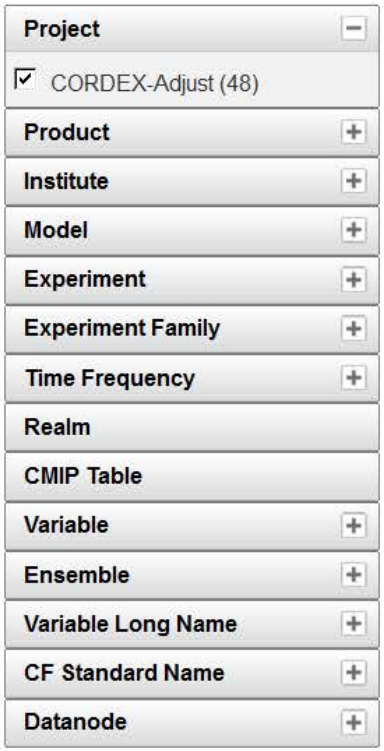
- CORDEX simulations available on ESGF can also be accessed via **climate4impact.eu** with an user friendly interface (registration on ESGF)
- web-based post-processing
- a tool for statistical downscaling is promised

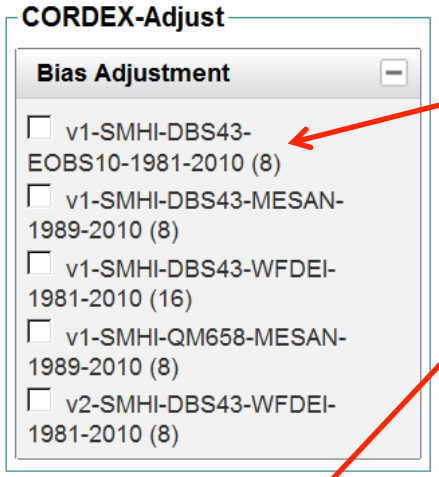


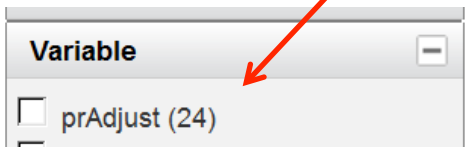
Bias-adjusted (corrected) CORDEX simulations

- a common question: “Where can I get bias-adjusted CORDEX data?”
- many methods and many groups run BC but usually BC simulations are not easily available to a wider community
- 2nd CORDEX SAT meeting (Feb 2015): “**CORDEX-Adjust**” project on ESGF
- work is ongoing and the first BC CORDEX simulations should be available soon on ESGF (summer 2016)

Just a test configuration







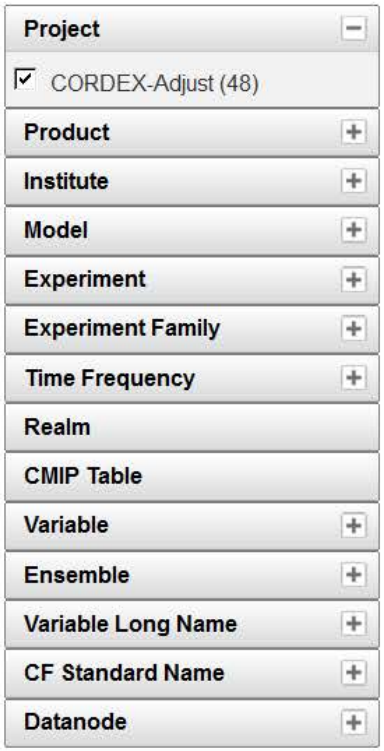
BC info is included to one element “Bias Adjustment”: “**BC_name**”-”**OBS**”-”**Period**”

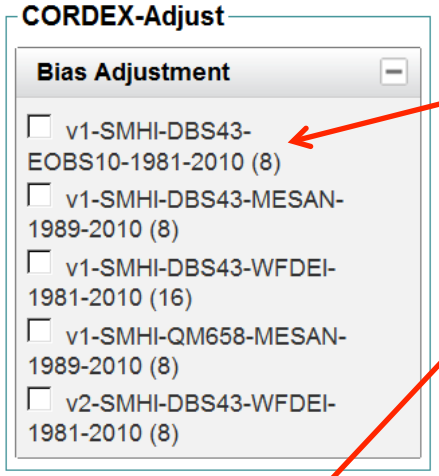
variable names: **pr** --> **prAdjust**

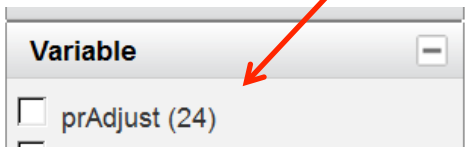
Bias-adjusted (corrected) CORDEX simulations

- a common question: “Where can I get bias-adjusted CORDEX data?”
- many methods and many groups run BC but usually BC simulations are not easily available to a wider community
- 2nd CORDEX SAT meeting (Feb 2015): “**CORDEX-Adjust**” project on ESGF
- work is ongoing and the first BC CORDEX simulations should be available soon on ESGF (summer 2016)

Just a test configuration







BC info is included to one element “Bias Adjustment”: “**BC_name**”-”**OBS**”-”**Period**”

variable names: **pr** --> **prAdjust**

- **BC is a controversial approach**
- **still debated**
- **no best BC method ?**
- **BC-related uncertainties unclear**

Climate indices

- there are many climate indices defined by different communities
- different software packages to calculate different indices
- Can we provide a set of climate indices for the CORDEX simulations in a coordinated way?
- the main problem is that there is no standard definition of metadata for the climate indices (standard and long names etc.)
- work on development of common metadata standards for the climate indices is ongoing in the CLIPC project (Climate Information Platform for Copernicus; http://www.clipc.eu/)

Metadata description for climate indices[1] according to the Climate & Forecasting (CF) convention – variable names and attributes

Draft version 5 May 2016

Variable name	Expert Team Index name	Expert Team Definition	Index units (user-friendly)	calculation frequency	long_name (standarised)	CF units	CF standard_name	CF cell-method
fd	Frost days	Count when TN < 0°C	days	ann/sea/mon	Number of Frost Days (Tmin < 0C)	1	number_of_days_with_air_temperature_below_threshold	time: minimum within days time: sum over days
fd0	Frost days	Count when TN < 0°C	days	ann/sea/mon	Number of Frost Days (Tmin < 0C)	1	number_of_days_with_air_temperature_below_threshold	time: minimum within days time: sum over days
fd2	Frost days 2	Count when TN < 2°C	days	ann/sea/mon	Number of Frost Days (Tmin < +2C)	1	number_of_days_with_air_temperature_below_threshold	time: minimum within days time: sum over days
fdm2	Hard freeze	Count when TN < -2°C	days	ann/sea/mon	Number of Hard Freeze Days (Tmin < -2C)	1	number_of_days_with_air_temperature_below_threshold	time: minimum within days time: sum over days

Climate indices

- there are many climate indices defined by different communities
- different software packages to calculate different indices
- Can we provide a set of climate indices for the CORDEX simulations in a coordinated way?
- the main problem is that there is no standard definition of metadata for the climate indices (standard and long names etc.)
- work on development of common metadata standards for the climate indices is ongoing in the CLIPC project (Climate Information Platform for Copernicus; http://www.clipc.eu/)

Metadata description for climate indices[1] according to the Climate & Forecasting (CF) convention – variable names and attributes

Draft version 5 May 2016

Variable name	Expert Team Index name	Expert Team Definition	Index units (user-friendly)	calculation frequency	long_name (standarised)	CF units	CF standard_name	CF cell-method
fd	Frost days	Count when TN < 0°C	days	ann/sea/mon	Number of Frost Days (Tmin < 0C)	1	number_of_days_with_air_temperature_below_threshold	time: minimum within days time: sum over days
fd0	Frost days	Count when TN < 0°C	days	ann/sea/mon	Number of Frost Days (Tmin < 0C)	1	number_of_days_with_air_temperature_below_threshold	time: minimum within days time: sum over days
fd2	Frost days 2	Count when TN < 2°C	days	ann/sea/mon	Number of Frost Days (Tmin < +2C)	1	number_of_days_with_air_temperature_below_threshold	time: minimum within days time: sum over days
fdm2	Hard freeze	Count when TN < -2°C	days	ann/sea/mon	Number of Hard Freeze Days (Tmin < -2C)	1	number_of_days_with_air_temperature_below_threshold	time: minimum within days time: sum over days

a “CORDEX-Index” project on ESGF ?

Empirical Statistical Downscaling (ESD)

- ESD is an integral part of CORDEX, not only RCMs
- however, up to now only RCM simulations are available (ESGF+portals)
- ESD community is very diverse: many different methods, more “local” downscaling compared to size of the CORDEX domains
- common metadata for ESD methods, Data Reference Syntax (DRS) are in development

Suggestion for the extension of the *CORDEX Archive Design* document

Frank Kreienkamp, Deutscher Wetterdienst, frank.kreienkamp@dwd.de
José Manuel Gutiérrez, IFCA-CSIC/UC, gutierjm@unican.es
Chris Jack, University if Cape Town, cjack@csag.uct.ac.za
Sixto Herrera, Universidad de Cantabria, herreras@unican.es
Version 1.03; January 29, 2016

The following text suggests several extensions for the current[1] CORDEX Archive Design to include needed aspects for ESD data. The section number corresponds to the defined in the CORDEX Archive Design document.

- still a lot of work
- many technical aspects are different between RCM and ESD

CMIP6

- **CORDEX** is a CMIP6-Endorsed Model Intercomparison Project (MIP)
- a diagnostic MIP (no new experiments but specific output)

- 30 years of the pre-industrial experiment
- 1950-2014, the historical simulations
- 2015-2100, the scenario runs (RCP26, 45 and 85)

we may expect some changes to the CORDEX archive design to be close to the CMIP6 specifications as much as possible (simplifies CORDEX data management using tools developed in CMIP6)

- CMIP6 GCM data is expected in 2018

CORDEX output variables

CORE: 43 variables (monthly and seasonal, ESGF/portals)
TIER1: 60 variables (daily, ESGF/portals)
TIER2: 16 variables (3-hr) and 36 variables (6-hr): available by request from the CORDEX RCM groups (some available on ESGF)

CORDEX output variables

CORE: 43 variables (monthly and seasonal, ESGF/portals)
TIER1: 60 variables (daily, ESGF/portals)
TIER2: 16 variables (3-hr) and 36 variables (6-hr): available by request from the CORDEX RCM groups (some available on ESGF)

- near-surface relative humidity (**hurs**) was not included but specific humidity (**huss**), added in December 2013 (not all RCM modelling groups have saved relative humidity)

CORDEX output variables

CORE: 43 variables (monthly and seasonal, ESGF/portals)
TIER1: 60 variables (daily, ESGF/portals)
TIER2: 16 variables (3-hr) and 36 variables (6-hr): available by request from the CORDEX RCM groups (some available on ESGF)

- near-surface relative humidity (**hurs**) was not included but specific humidity (**huss**), added in December 2013 (not all RCM modelling groups have saved relative humidity)
- only total soil moisture content (**mrso**) has been defined but satellite-based observational data sets present a few upper cm

CORDEX output variables

CORE: 43 variables (monthly and seasonal, ESGF/portals)
TIER1: 60 variables (daily, ESGF/portals)
TIER2: 16 variables (3-hr) and 36 variables (6-hr): available by request from the CORDEX RCM groups (some available on ESGF)

- near-surface relative humidity (**hurs**) was not included but specific humidity (**huss**), added in December 2013 (not all RCM modelling groups have saved relative humidity)
- only total soil moisture content (**mrso**) has been defined but satellite-based observational data sets present a few upper cm
- only 3 pressure levels (850, 500 and 200mb), in many regions not enough (monsoon circulation, jet streams, vertical cross sections etc.): more pressure levels (synchronised to CMIP6), the CORDEX domains should decide how many pressure levels to save

CORDEX output variables

CORE: 43 variables (monthly and seasonal, ESGF/portals)
TIER1: 60 variables (daily, ESGF/portals)
TIER2: 16 variables (3-hr) and 36 variables (6-hr): available by request from the CORDEX RCM groups (some available on ESGF)

- near-surface relative humidity (**hurs**) was not included but specific humidity (**huss**), added in December 2013 (not all RCM modelling groups have saved relative humidity)
- only total soil moisture content (**mrso**) has been defined but satellite-based observational data sets present a few upper cm
- only 3 pressure levels (850, 500 and 200mb), in many regions not enough (monsoon circulation, jet streams, vertical cross sections etc.): more pressure levels (synchronised to CMIP6), the CORDEX domains should decide how many pressure levels to save
- ocean variables were not defined from the beginning, only atmospheric ones, should be included (ocean variables are available from the Med-CORDEX portal only)

Thank you for your attention
Questions ?