

# **Publishing and Accessing CORDEX data via the Earth System Grid Federation (ESGF)**

**Grigory Nikulin**

Swedish Meteorological and Hydrological Institute

# CORDEX-ESGF team

**publication of CORDEX simulations on ESGF is an joint effort of several 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 ....



# Ensembles of regional climate simulations

## Large ensembles of regional climate simulations:

- **PRUDENCE (Europe) – about 11 RCMs**  
<http://prudence.dmi.dk/>
- **ENSEMBLES (Europe) - about 20 RCMs**  
<http://ensemblesrt3.dmi.dk/>
- **NARCCAP (North America) - 6 RCMs**  
<http://www.narccap.ucar.edu/>
- **CLARIS (South America) - 5 RCMs**  
<http://www.claris-eu.org/>
- **ENSEMBLES-AMMA (West Africa) - 10 RCMs**  
<http://ensemblesrt3.dmi.dk/>

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

# CORDEX initiative

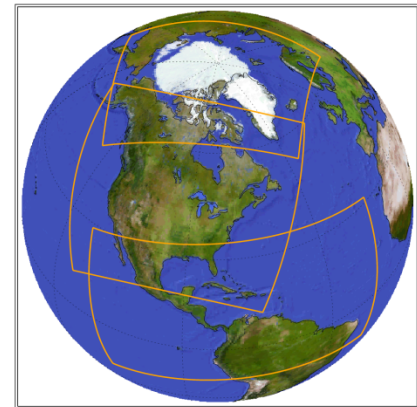
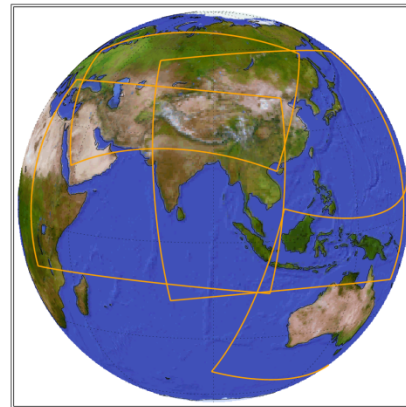
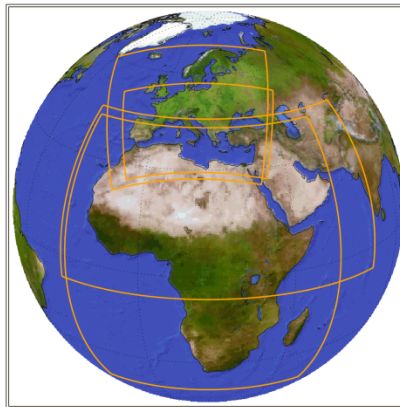
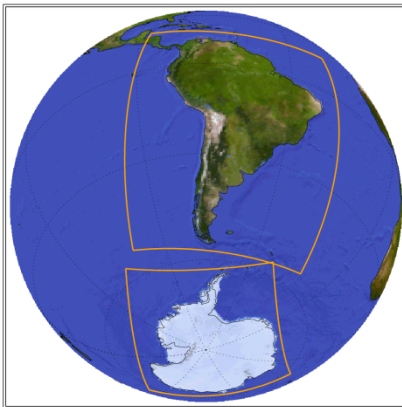
## Coordinated Regional Downscaling Experiment

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**

# CORDEX archives (not ESGF)

---

## Med-CORDEX (MED):

- <http://www.medcordex.eu/>  
users should be approved  
ToU – “non-commercial only” for all simulations (web)

## East Asia (EAS)

- <http://cordex-ea.climate.go.kr/>  
register and download  
ToU – “non-commercial only” for all simulations (pdf doc)

## South Asia (WAS)

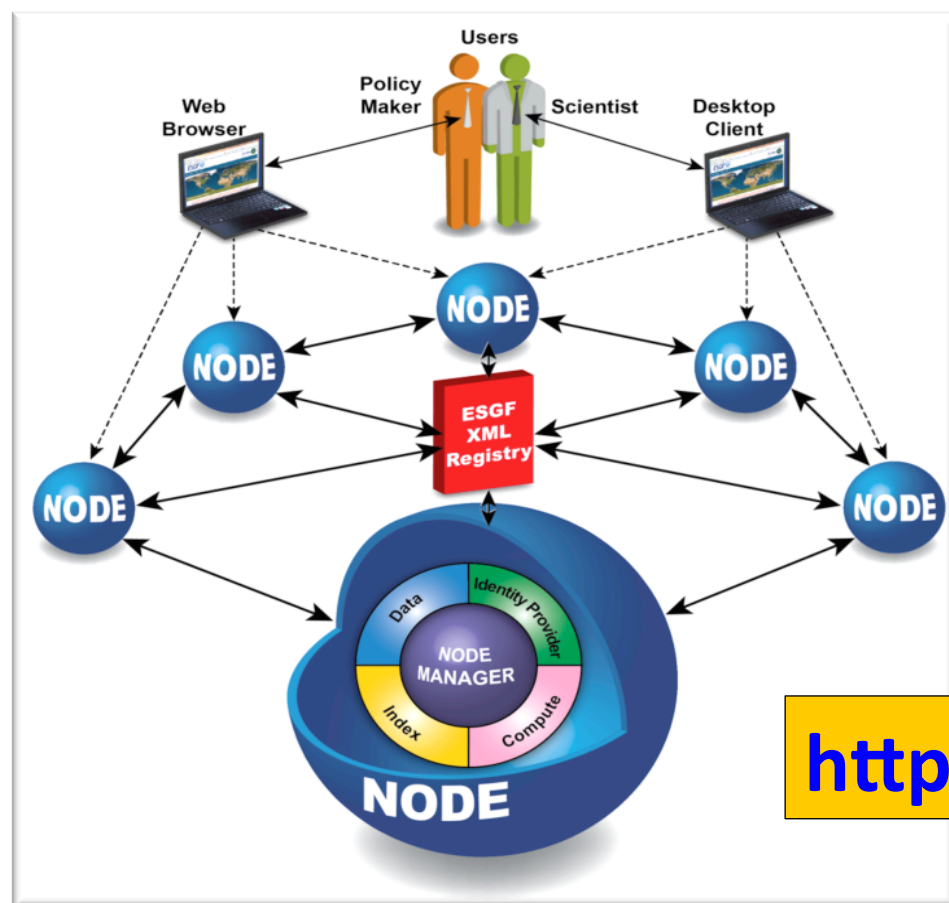
- <http://cccr.tropmet.res.in/cordex/>  
users should be approved  
ToU – “non-commercial only” for all simulations (web)

## CCCma (Canada)

- <http://www.cccma.ec.gc.ca/data/canrcm/CanRCM4/>  
register and download  
ToU – “unrestricted”  
only CanRCM4: AFR, ARC, EUR, NAM, both 0.22 and 0.44deg

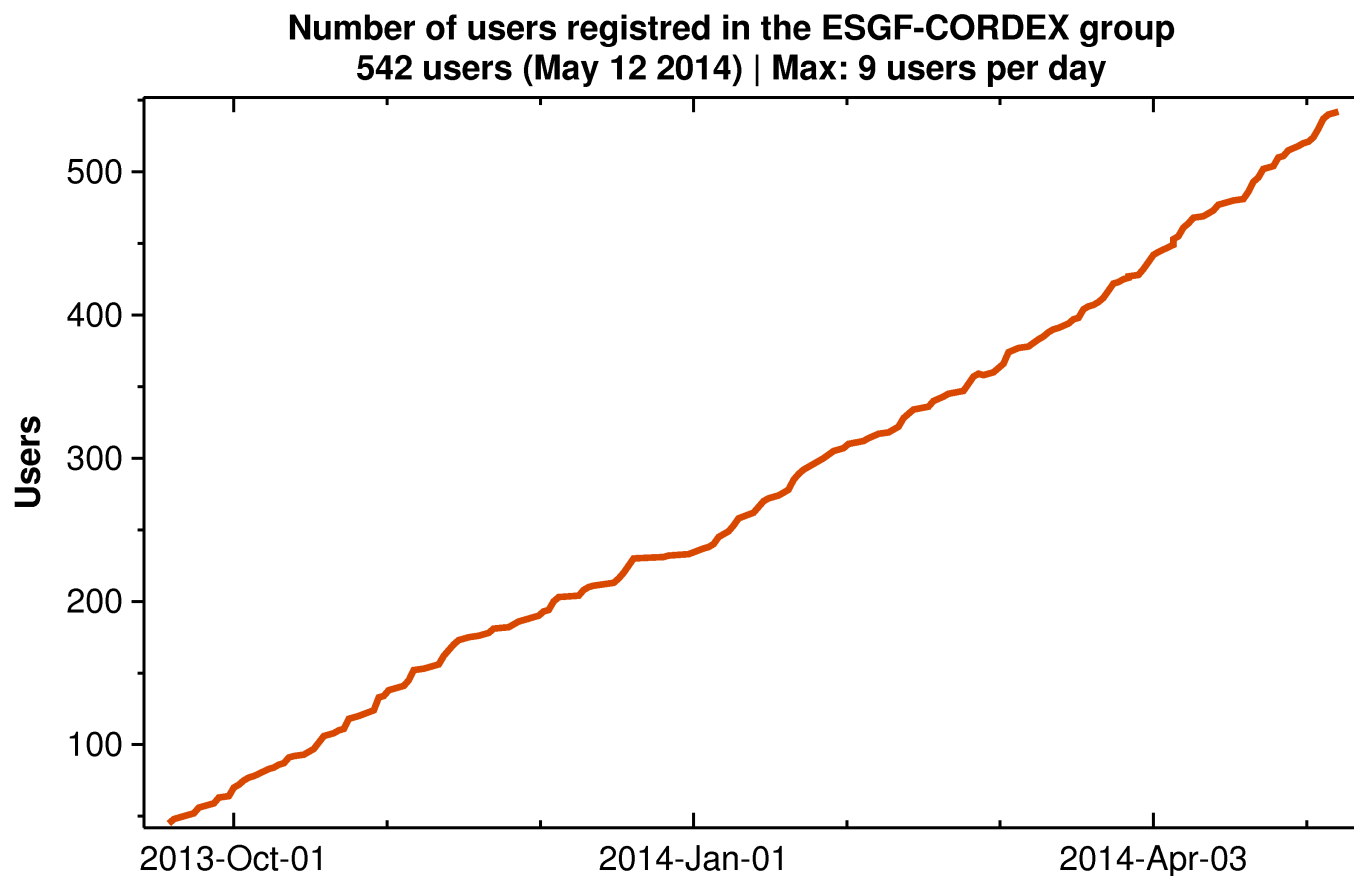
# ESGF

- 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.)



# CORDEX-ESGF in operation

- **CORDEX-ESGF** is in operation since mid September 2013
- now ESGF provides the same interface for both global (CMIP5) and regional (CORDEX) climate simulations
- 542 users in the CORDEX-ESGF group (12 May 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

[http://cordex.dmi.dk/joomla/images/CORDEX/cordex\\_terms\\_of\\_use.pdf](http://cordex.dmi.dk/joomla/images/CORDEX/cordex_terms_of_use.pdf)

# 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 ESGF archive

- CORDEX simulations are visible on all ESGF index nodes
- Not all ESGF index nodes support the full CORDEX search: the CORDEX specific search facets “Domain”, “Driving model”, “Downscaling realisation”

At moment (May 2014) full CORDEX support only on the European ESGF index nodes:

- [esg-dn1.nsc.liu.se](http://esg-dn1.nsc.liu.se) (SMHI-NSC, Sweden)
  - [cordexesg.dmi.dk](http://cordexesg.dmi.dk) (DMI, Denmark)
  - [esgf-data.dkrz.de](http://esgf-data.dkrz.de) (DKRZ, Germany)
  - [esgf-index1.ceda.ac.uk](http://esgf-index1.ceda.ac.uk) (BADC, UK)
  - [esgf-node.ipsl.fr](http://esgf-node.ipsl.fr) (IPSL, France)
- 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

# CORDEX ESGF datanodes (12 May 2014)

Datanode	Number of Datasets	Number of CORDEX domains
<b>esg-dn1.nsc.liu.se</b> SMHI, NCS-LIU, Sweden	27804	AFR-44, ARC-44, EUR-11, EUR-44, MNA-44, MNA-22, NAM-44, SAM-44, WAS-44
<b>cordexesg.dmi.dk</b> DMI, Denmark	5394	AFR-44, ANT-44, EUR-11, EUR-44, NAM-44
<b>carbon.dkrz.de</b> DKRZ, Germany	3324	AFR-44, EUR-11, EUR-44, WAS-44
<b>esgf-data1.ceda.ac.uk</b> BADC, UK	1004	AFR-44, AUS-44, EUR-44, EAS-44
<b>data.meteo.unican.es</b> Univ. Cantabria, Spain	228	EUR-22, EUR-44
<b>esgf.extra cea.fr</b> CEA, France	196	EUR-11, EUR-44
<b>noresg.norstore.uio.no</b> NorStore, Norway	-	published EUR-44 and EAS-44 earlier, not available now

- one dataset: one variable, experiment, domain, frequency
- full CORDEX output for one simulation (Core+Tier1): 235 datasets

# CORDEX Domains on ESGF (12 May 2014)

CORDEX Domain	Number of Datasets	Number of groups
Europe (44, 22, 11)	6091/113/3894	8/1/5
Africa (44)	6234	4
North America (44)	1752	2
Arctic (44)	2718	1
MENA (44, 22)	1551/700	1/1
South Asia (44)	830	2
South America (44)	564	1
Antarctica (44)	604	1
Australia (44)	171	1
East Asia (44)	171	1

- 10 CORDEX domains are available (different number of simulations)
- a RCM group publishes simulations on its own datanode (if available)
- where to publish is decided within the CORDEX domains:
  - ESGF non-ESGF
  - EUR-11 to DKRZ and EUR-44 to DMI (Euro-CORDEX)

# ESGF: the first steps

All ESGF nodes have the same user interface

**ESGF** Earth System Grid Federation

is-enes  
INFRASTRUCTURE FOR THE EARTH SYSTEM NETWORK  
FOR DATA SYSTEM MODELLING

**NSC**  
National Supercomputer Centre

Linköpings universitet

**SMHI**

Home **Search** Tools Login Help

Welcome to this ESGF P2P Node

**Peer Nodes**

- [ANL Node](#)
- [BADC Node](#)
- [BNU Node](#)
- [CMCC Node](#)
- [DKRZ Node](#)
- [DKRZ CMIP5 Node](#)
- [NOAA-ESRL Node](#)
- [NOAA-GFDL Node](#)

**About the datanode/NSC**

The SMHI-LIU-NSC ESGF datanode at the National Supercomputer Centre, Linköping, is Sweden's first datanode in the ESGF framework. It hosts climate data produced by the Swedish Meteorological and Hydrological Institute (SMHI). NSC is an independent organization within Linköping University (LIU), and is funded by the Swedish Research Council via SNIC (Swedish National Infrastructure for Computing).





**Resources**

**Quick Links**

- [Create Account](#)
- [MyProxyLogon](#)
- [Expert Search \(XML\)](#)
- [Wget Script Generator](#)
- [ESGF aggregated RSS feed](#)
- [Contact ESGF](#)

# ESGF: the first steps

## Select project CORDEX

[Home](#)
[Search](#)
[Tools](#)
[Login](#)
[Help](#)

**Current Selections**  
 No search criteria selected

[Temporal Search](#)  
[Geospatial Search](#)  
[Clear search constraints and datacart](#)  
[Search Help](#)  
[Search Controlled Vocabulary](#)

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.  
 To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

☒ Search All Sites
 ☐ Show All Replicas
 ☐ Show All Versions

Display  datasets per page  
[Add All Displayed to Datacart](#)
[Remove All Displayed from Datacart](#)

Results

Data Cart

**Search Categories**  

Project
AmeriFlux (3)
BioClim (2)
C20C (1)
CMIP5 (60654)
<b>CORDEX (37950)</b>
COUND (4)
CSSEF (614)
EUCLIPSE (29)
GASS-YoTC-MIP (2667)

# 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
Instrument
Experiment Family
Experiment
Time Frequency
Product
Realm

ALADIN52 (198)
CCLM4-8-17 (1781)
HIRHAM5 (2820)
RACMO21P (968)
RACMO22E (1588)
RACMO22T (980)
RCA4 (26654)
RCA4-SN (1150)
REMO2009 (390)
WRF331F (60)
WRF331G (170)
Instrument
Experiment Family
Experiment
Time Frequency
Product
Realm

## Driving Model (GCM)

Variable
Variable Long Name
CMIP Table
CF Standard Name
Ensemble
Domain
Driving Model
CCCma-CanESM2 (2744)
CNRM-CERFACS-CNRM-CM5 (3141)
ECMWF-ERAINT (4372)
ICHEC-EC-EARTH (13157)
IPSL-IPSL-CM5A-MR (1210)
MIROC-MIROC5 (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

## Domain

Ensemble
Domain
AFR-44 (6234)
AFR-44i (2628)
ANT-44 (604)
ANT-44i (364)
ARC-44 (2718)
ARC-44i (1621)
AUS-44 (171)
AUS-44i (80)
EAS-44 (171)
EAS-44i (80)
EUR-11 (3894)
EUR-11i (1600)
EUR-22 (113)
EUR-44 (6091)

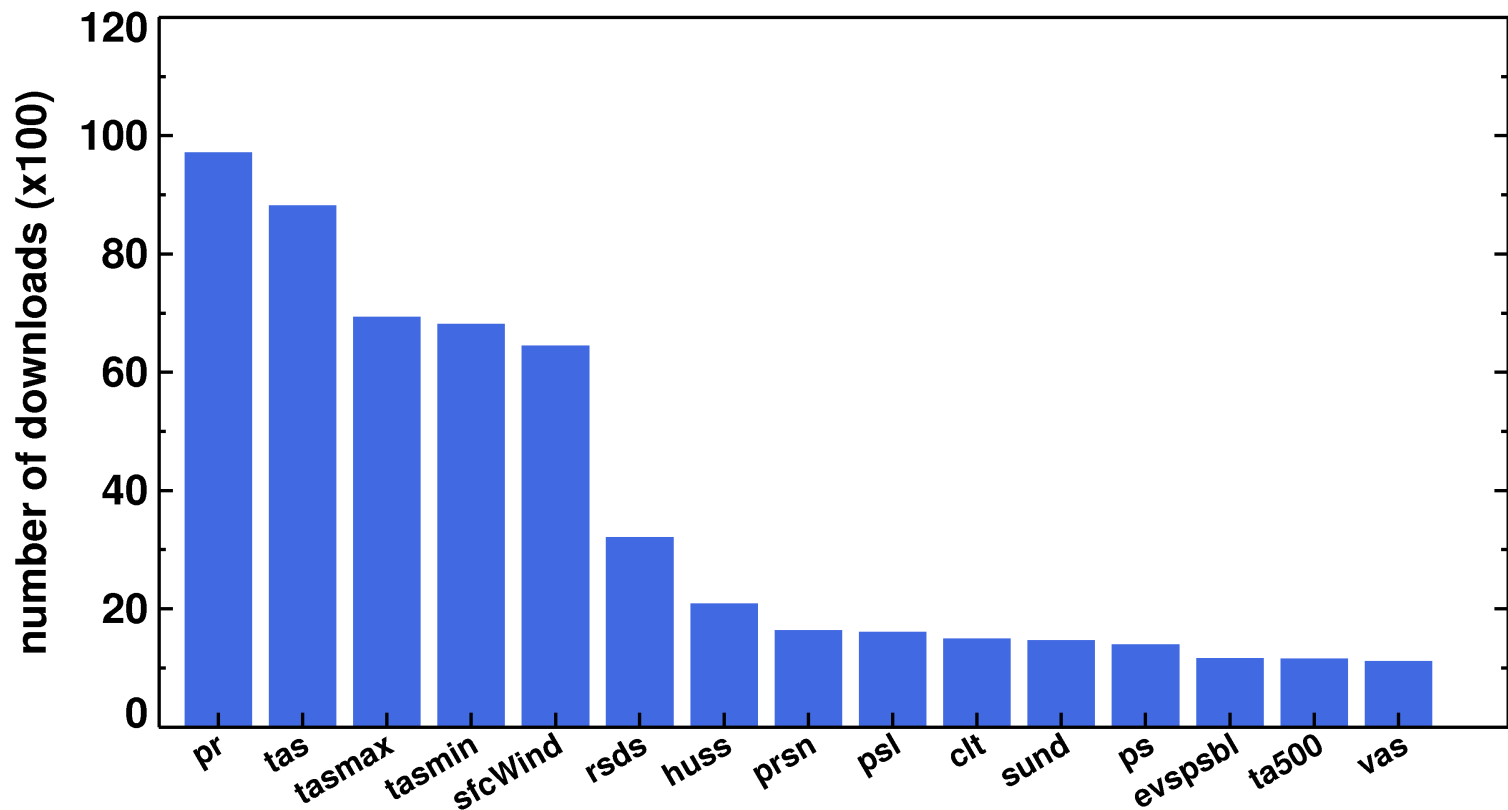
## Domain name convention

- 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

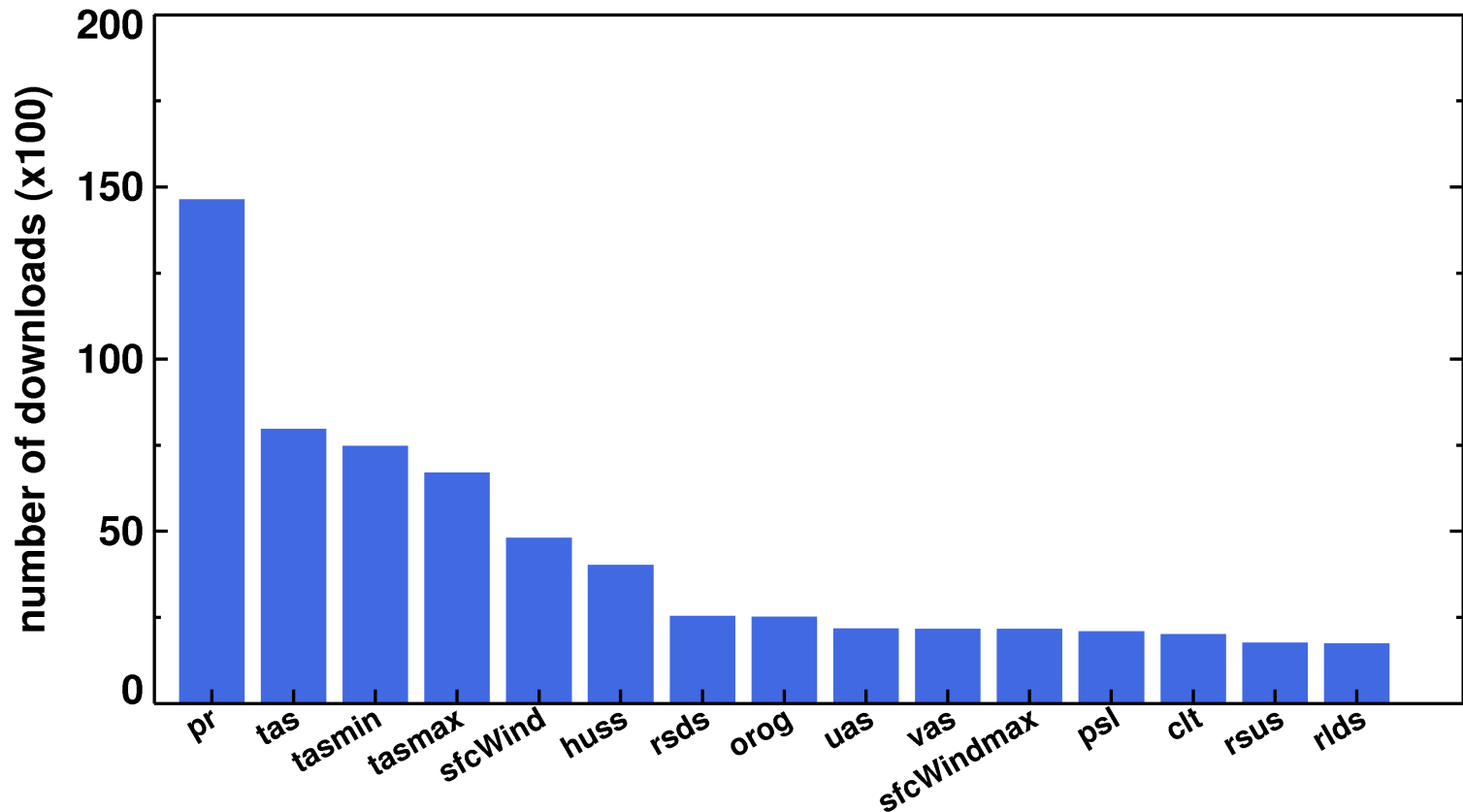
Total number of downloads per variable (SMHI-RCA4)  
CORDEX EUR-44 | <http://esg-dn1.nsc.liu.se> | 20140512



# Euro-CORDEX downloads (EUR-11)

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

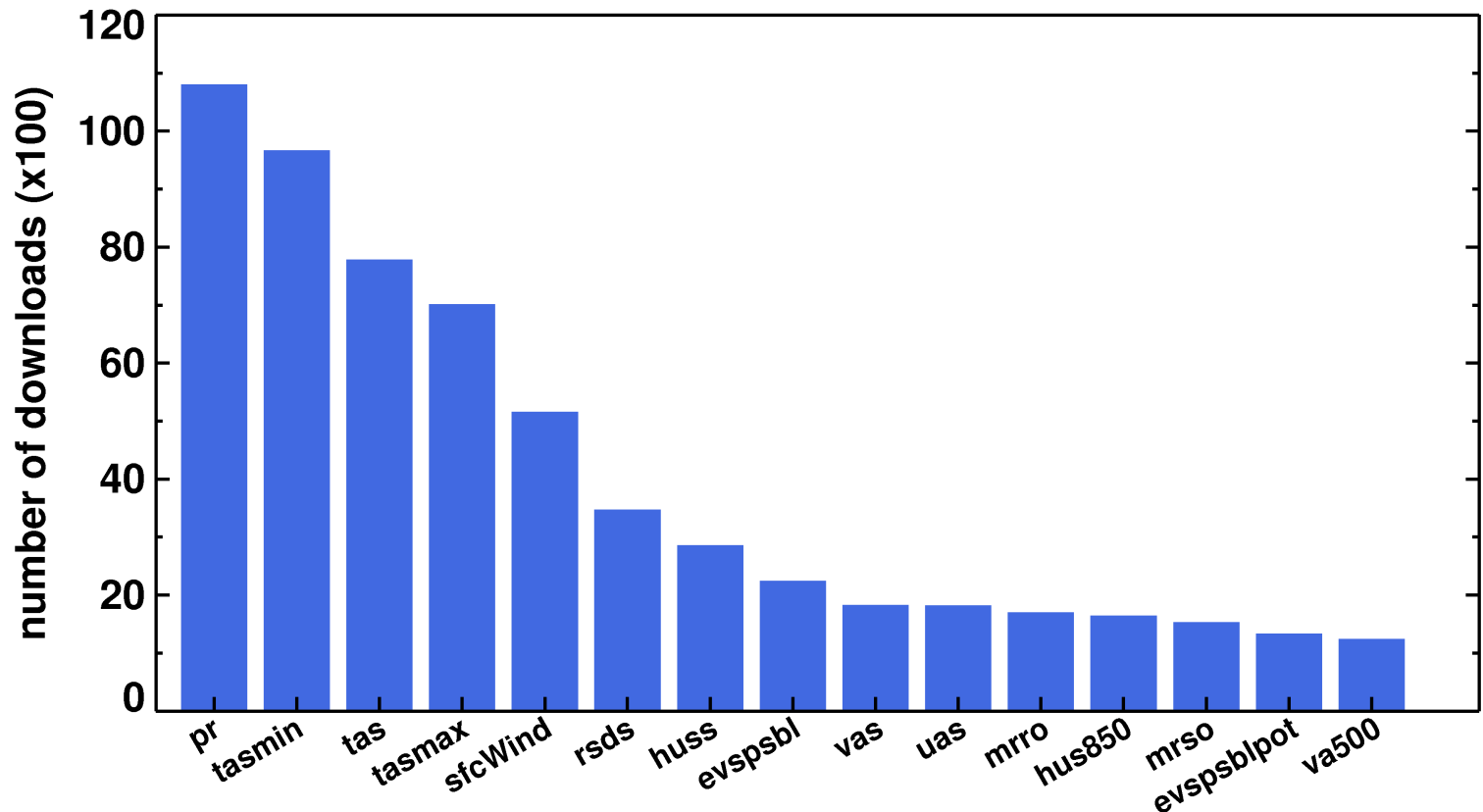
Total number of downloads per variable (SMHI-RCA4)  
CORDEX EUR-11 | <http://esg-dn1.nsc.liu.se> | 20140512



# Africa-CORDEX downloads (AFR-44)

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

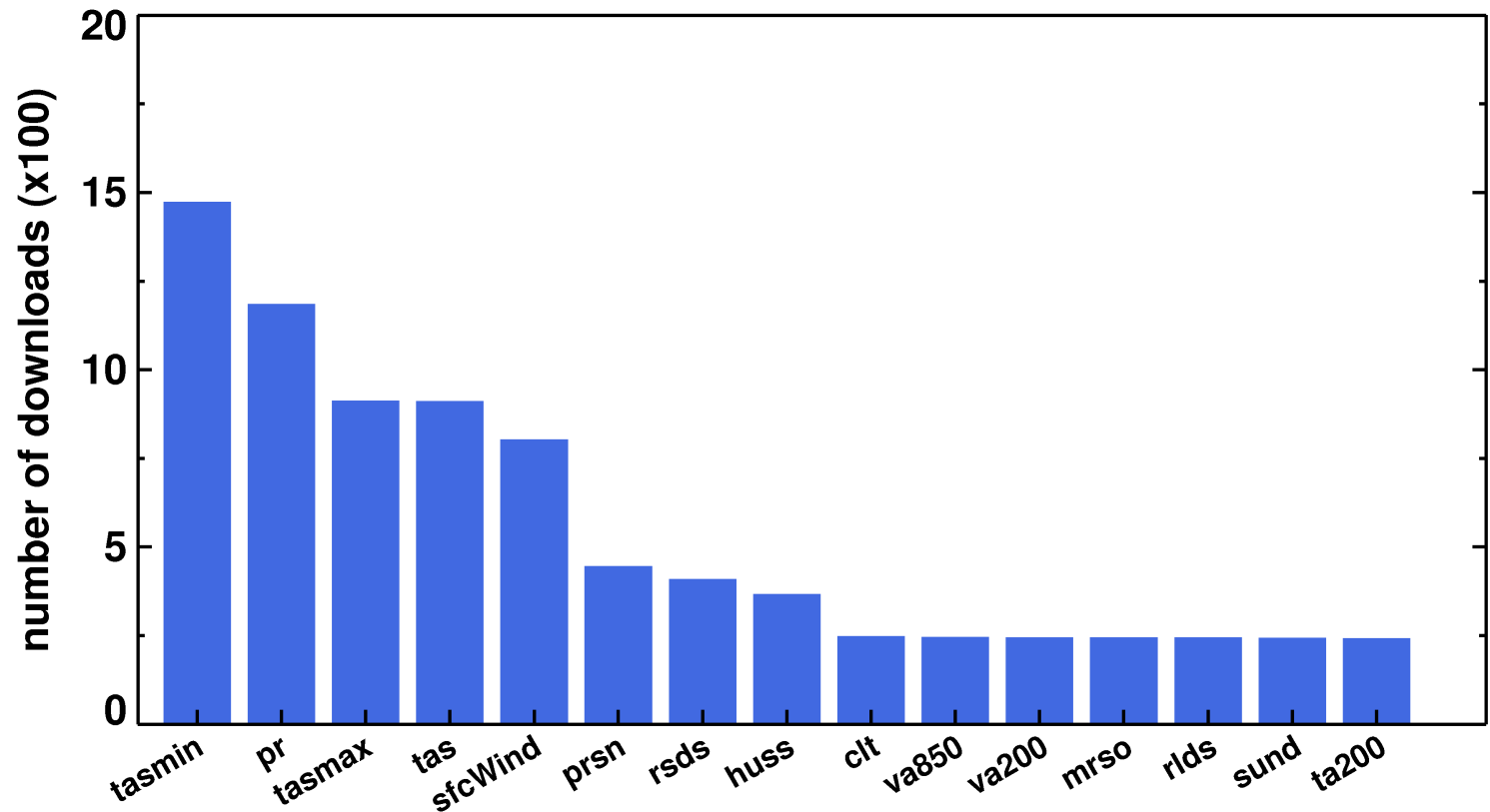
Total number of downloads per variable (SMHI-RCA4)  
CORDEX AFR-44 | <http://esg-dn1.nsc.liu.se> | 20140512



# MENA-CORDEX downloads (MNA-44)

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

Total number of downloads per variable (SMHI-RCA4)  
CORDEX MNA-44 | <http://esg-dn1.nsc.liu.se> | 20140512

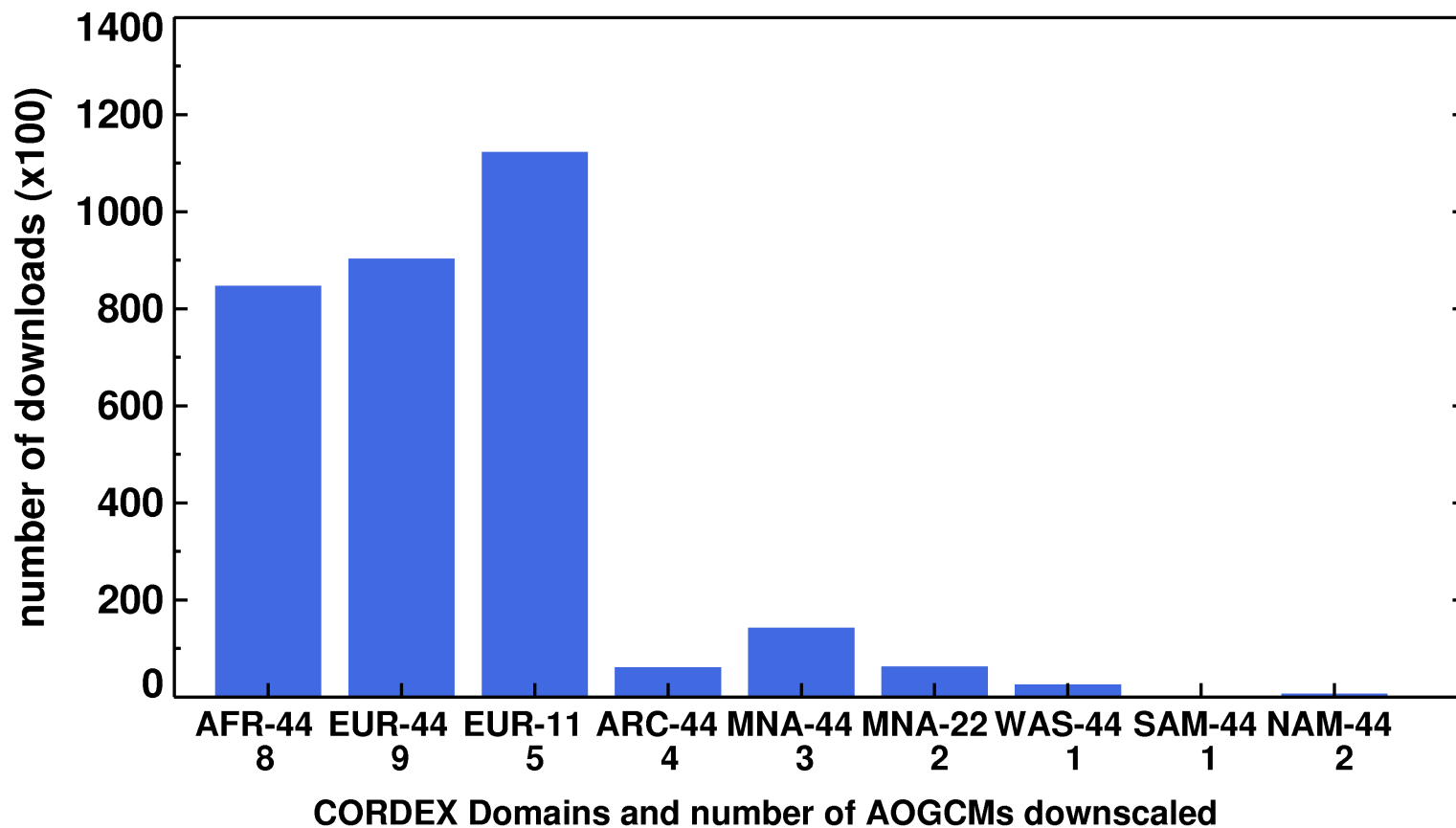


- User needs: minimum temperature is important for agriculture

# Downloads per CORDEX domain

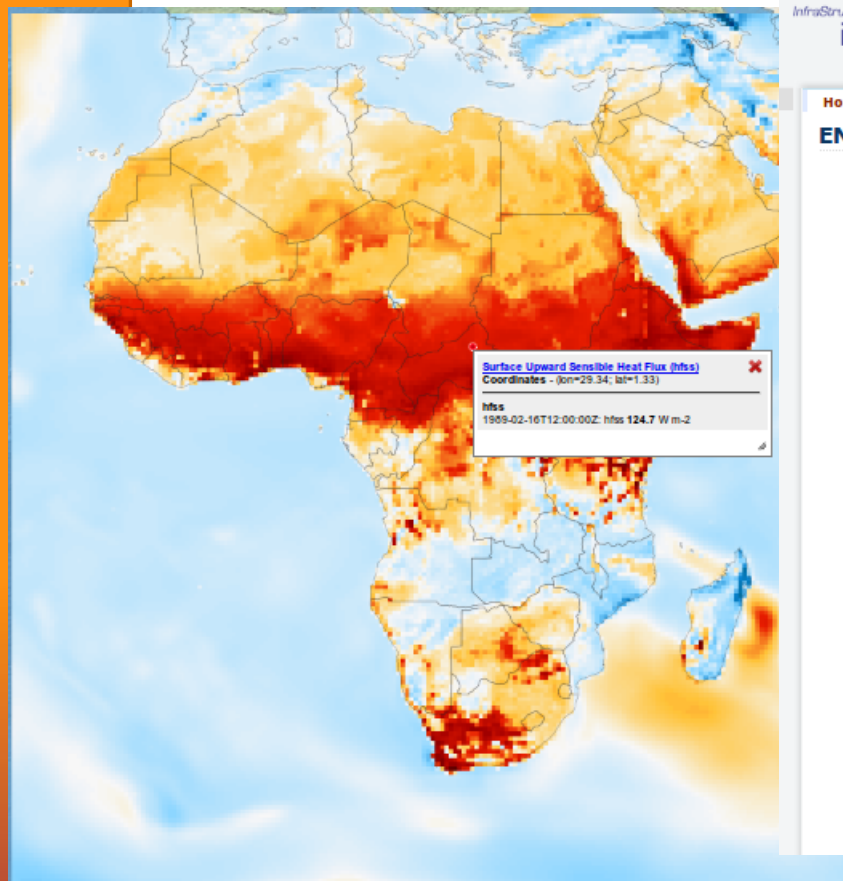
EUR-11 – hi-res Euro-CORDEX is a clear favourite

Total number of downloads per CORDEX domain  
 SMHI-RCA4 | <http://esg-dn1.nsc.liu.se> | 20140512



# Climate4impact.eu (operational)

- 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.



Infrastructure for the European Earth System Modelling

ENES Portal Interface for the Climate Impact Communities (Prototype)

15-ENES Website

Home Data discovery Map & Plot Documentation Help About us Log in

## ENES Portal Interface for the Climate Impact Communities

Welcome to the **ENES Portal Interface for the Climate Impact Communities (EPICIC)**, oriented towards climate change impact modellers, impact and adaptation consultants, as well as other experts using climate change data. The portal is in its evaluation phase, [contact us](#) if you have comments or would like to participate.

Here you will find **access to data** and **quick looks** of global climate models (GCM) scenarios, as well as some regional climate model (RCM) and downscaled higher resolution climate data. The portal provides data transformation tooling for **tailoring data** to your needs and **mapping & plotting** capabilities. All using standardised interfaces and common data processing tools to access and process data, properly described with standardised metadata.

**Guidance** on how to use climate scenarios, **documentation** on the climate system, frequently asked questions (FAQ) and **examples** in several impact and adaptation themes (Use Cases) are presented and described, along with the steps required to go from the GCM data to the impact model input data (workflow).

Agriculture/Forestry
 Energy
 Health
 Infrastructure/Urban
 Marine/Coastal
 Nature/Biodiversity
 Tourism
 Water Management