

How to download CORDEX data from the ESGF

The way the ESGF works is to generate a wget script based on input you give. So you select which domain, which variable, which scenario, which regional model, what time step etc then ESGF generates a wget script that you then run to do the actual data retrieval.

You need to have the software wget and this instruction is for the Linux environment so if you are in a windows environment you will need to install wget for windows.

Here are the steps to download data in Linux:

1. ESGF NODE

There are a few to choose from for example: Sweden, Germany, UK or France linked below

<https://esg-dn1.nsc.liu.se/search/esgf-liu/> or <https://esgf-data.dkrz.de/projects/esgf-dkrz/> or <https://esgf-index1.ceda.ac.uk/search/esgf-ceda/> or <https://esgf-node.ipsl.upmc.fr/projects/esgf-ipsl/>

You will need to **create an account**. This is done by clicking the Create Account link in the top right. This will provide you with an OpenId (like a username) and a password you need to execute the wget script at the end of the process.

Once you have an account, **Login**, and you can continue.

2. Which CORDEX do you want?

On ESGF is data for CORDEX “Phase 1” (0.44 degree data, which has by far the largest GCM-RCM downscaled data available), CORDEX-Adjust if you want bias-corrected data (smaller GCM-RCM dataset). In the Search block enter your choice and then click the “Search” button. For example “CORDEX”

Now you have narrowed the search to CORDEX data, you will see this in the **Search Constraints** just under the Search button:

For example: Search Constraints: CORDEX

3. Select the domain...

Click on “Domain” in the special CORDEX box on the left and tick the domain you want.

If you want the **CORDEX-CORE data** (the new 0.22 degree downscaling, but a very limited GCM-RCM matrix – 3 RCMs downscaled 4 GCMs) select for example AFR-22.

Then click the search button.

Now you will see your selections under the Search button:

For example: Search Constraints: CORDEX|AFR-44

4. Now select the variable... If you want multiple variables download one variable at a time because you can only download 1000 files at a time in a wget script and sometimes there are more than this number of files. So you can think you are downloading all the files but then some will be missing if there are over 1000.

You can select variables in one of 2 ways – I will use precipitation as an example:

1 – If you are familiar with the abbreviations used for variables on ESGF you can select the “**Variable**” tab and select “**pr**” from the list

2 – If you are not familiar with the variable conventions on ESGF you can select the “**Variable Long Name**” tab and select “**Precipitation**” from the list

Using (1) as an example, tick “pr” and click the “Search” button again and now you will see your selection:

For example: Search Constraints: CORDEX|AFR-44|pr

Some other popular variables:

tas = average surface temperature, tasmin = average surface minimum temperature,
tasmax = average surface maximum temperature, pr = precipitation

5. Now choose your experiment...

Click “Experiment” on the left menu and you will see:

Evaluation – CORDEX RCMS downscaled ERA-Interim data

historical – CORDEX RCMS downscaled GCM 1950-2005

rcp26, rcp 45, rcp85 – CORDEX RCMS downscaled GCM 2006-2100 for each RCP

The evaluation period is needed to evaluate the bias in each RCM. It is recommended to always download the evaluation run.

Tick for example the rcp85 and click the “Search” button again and now you will see your selection.

For example: Search Constraints: CORDEX | AFR-44 | pr | rcp85

6. Now you can select the time frequency you are interested in, daily or monthly data

Click “Time Frequency” on the left menu and select daily or monthly, Monthly is suggested until you are confident in using ESGF because the data files are smaller. Remember you can only download 1000 files at a time in a wget script and sometimes there are more than this number of files, the excess of which will not be downloaded.

CORDEX-CORE also has 3- and 6-hourly data, there are lots of files if you select this option so be sure of what you are trying to do before you download these.

Using monthly data – tick “mon” and Search again to give

For example: Search Constraints: CORDEX | AFR-44 | pr | rcp85 | mon

7. Next choose your Ensemble

We generally use r1i1p1 so click “Ensemble” and tick “r1i1p1” then Search to give

For example: Search Constraints: [CORDEX](#) | [AFR-44](#) | [pr](#) | [rcp85](#) | [mon](#) | [r1i1p1](#)

8. Then in the CORDEX box choose your “**Downscaling Realization**” on the left menu and select “v1” then Search to give

For example: Search Constraints: [CORDEX](#) | [AFR-44](#) | [pr](#) | [rcp85](#) | [mon](#) | [r1i1p1](#) | [v1](#)

9. Now in the CORDEX box click “**RCM Model**” to select the RCMs you want the data from. Do them one at a time because you can only download 1000 files at a time in a wget script and sometimes there are more than this number of files. This is also why downloading 1 variable at a time is suggested.

For e.g. the RCA4 model, click on “RCM Model” in the CORDEX box and select a model then click Search to give

For example: Search Constraints: [CORDEX](#) | [AFR-44](#) | [pr](#) | [rcp85](#) | [mon](#) | [r1i1p1](#) | [v1](#) | [RCA4](#)

FYI:

To find out which GCMs each RCM has downloaded select a RCM from the “RCM Model”, then “Search”. Once this is done click “Driving Model” just above the CORDEX box and you will see which GCMs are downscaled. E.g. Select “RCA4” from the “RCM Model” tab and “Search”, then select “Driving Model” and you will see 9 GCMs here.

To look at other RCMS, e.g. CCLM, go up to the “Search Constraints” filter and delete RCA4 by clicking on the red “X”, wait for the results and then select “CCLM” from the “RCM Model” tab and repeat.

RCA4 has downscaled 9 GCMs, CCLM4 downscaled 3 GCMs, REMO 4 and RACMO 1

If you want to have RCMs that downscaled the same GCM you can select the GCM in the “Driving Model” part of the left menu before you do step 9.

10. Making the wget script

There are two options again (1) make a series of wget script for each of the results you see in the results window or (2) make a wget script that will download all the files in your search window:

(1) Simply click on the link [**WGET Script**] for each line of your search return page and you will be prompted to save a wget script for each search return result. This downloads the wget script that you will need to execute (see 11 below). The “Show Data” and “Metadata” links are useful to see the files you will be downloading and also the metadata associated with the files. If you have followed my example there will be 9 lines so 9 links.

(2) Above the list of results you will see a link “**Add all displayed results to Data Cart**“. Clicking this will add all your search results to your “My Data cart”. Once this is done you will see in your “Data Cart” on the top right the results have been added. Click on the “My Data Cart” link to be taken to your Data Cart page.

Click the “**Select All Databases**” tab and then the [**WGET Script**] link – this then generates a box that says “For better performance, WGET scripts are generated for each Data Center separately. Click on each link below to retrieve the script for each Data Center.” and a link to download the wget script e.g. “WGET Script for esg-dn1.nsc.liu.se” click the link and download the wget script.

Save the wget scripts to where you will be downloading the data.

11. Execute the wget script (See here for lots more info... <https://www.earthsystemcog.org/projects/cog/doc/wget>)

Then you need to make the wget script executable, so on the command line do

```
chmod 744 wget-xxxxxxxx.sh
```

Then execute the script:

```
./wget-xxxxxxxx.sh -H
```

You will be asked to enter your ESGF OpenId and Password

Instructions originally written by Chris Lennard