ICARDA
Science for Better Livelihoods in Dry Areas

IFAD
Enabling poor rural people to overcome poverty

Downloading prognostic meteorological information from Earth System Grid Federation (ESGF) Web-sites

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of the International Fund dor Agricultural Development (IFAD), International Center for Agricultural Research in the Dry Areas (ICARDA), Knowledge Management in Central Asian Countries Initiative for Lad Management (CACILM) phase II.

Feedback: ICARDA welcomes comment on and feedback on this publication. Please visit www.icarda.org to share your views and ideas..

International Center for Agricultural Research in the Dry Areas (ICARDA)
E-mail: ICARDA@cgiar.org
Website: www.icarda.org

Contents

INTRODUCTION

Future climate change has been recognized as one of the largest issues facing the world in the coming century (Evans, 2011). The Intergovernmental Panel on Climate Change (IPCC) has been tasked with compiling the state of knowledge in relation to climate change on a regular basis. To date they have produced five such assessments. These assessments are the basis of knowledge used by most governments to establish climate change related policy including the ongoing debates around the introduction of a price on greenhouse gas pollution. Global Climate Models (GCMs) are the main tools used to project the extent of this future climate change. The Coupled Model Intercomparison Project 3 (CMIP3, Meehl et al., 2007) was the international collaborative effort of GCM groups to produce projections that directly informed the IPCC fourth assessment report (IPCC, 2007). This construction of a many GCM ensemble is vital for dealing with the uncertainty associated with future projections. Every GCM that performs adequately for the recent past provides a plausible projection of future climate and it is difficult to know which of these plausible futures is more likely. Hence, the use of a many model ensemble is required to provide some measure of likelihood of the projected future climate. As the risks associated with large scale climate change have become better understood, more impact and adaptation studies have been performed. A significant spatial scale problem exists between the scale of the GCMs (200-400km) and the scales of interest for impacts and adaptation studies, which are often only tens of kilometers or less. In order to address this spatial scale problem various methods to downscale the GCM output have been developed. These downscaling methods can be generalized into two types: statistical and dynamical. Statistical downscaling involves deriving statistical relationships between some large-scale predictors and the local variable of interest. Dynamical downscaling uses mathematical representations of the physical processes that create the climate system, similar to GCMs, applied at a higher spatial resolution than the GCMs. In this way, they are able to capture climate phenomena not resolved by the GCMs including the influence of mountains and coastlines. Dynamical downscaling is done with a Regional Climate Model (RCM). When downscaling future climate projections RCMs assume that, the physical laws remain the same. Statistical downscaling techniques can also be applied to RCM output in order to provide information at point locations. One advantage of statistical techniques is that they are less computationally intensive and hence can be used to downscale many GCM (or RCM) climate projections. RCMs on the other hand, are quite computationally intensive and to date this has prevented them from being used to downscale many GCMs, hence they have not sampled the full range of plausible future climates.

The Coordinated Regional climate Downscaling Experiment (CORDEX) is a World Climate Research Program (WCRP) backed framework to produce ensembles of regional climate projections for all continents globally (Giorgi et al., 2009; http://wcrp.ipsl.jussieu.fr/SF_RCD_CORDEX.html). CORDEX brings together regional scale climate projections produced using both statistical and dynamical techniques. It aims to provide a framework to evaluate and benchmark model performance as well as producing projections for use in impact and adaptation studies. The first part of this framework is a set of common regional domains. These domains are shown in Figure 1. There is also an Arctic and an Antarctic domain. The model evaluation framework consists of RCM simulations performed using the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis (Uppala et al., 2008) as “perfect boundary conditions”. These

simulations are evaluated through a set of regional benchmark statistics, against regional datasets, that are designed and assembled by regional diagnostic teams. The climate projection framework within CORDEX is based on the set of new GCM simulations currently underway in support of the IPCC fifth Assessment Report, referred to as CMIP5. CORDEX focus on the GCM experiments using emission scenarios known as Representative Concentration Pathways (RCPs). CORDEX simulations span 1951 to 2100. Each RCM should perform these simulations for multiple GCMs. It is strongly recommended that all RCM groups should make their simulations available on ESGF (Earth System Grid Federation).

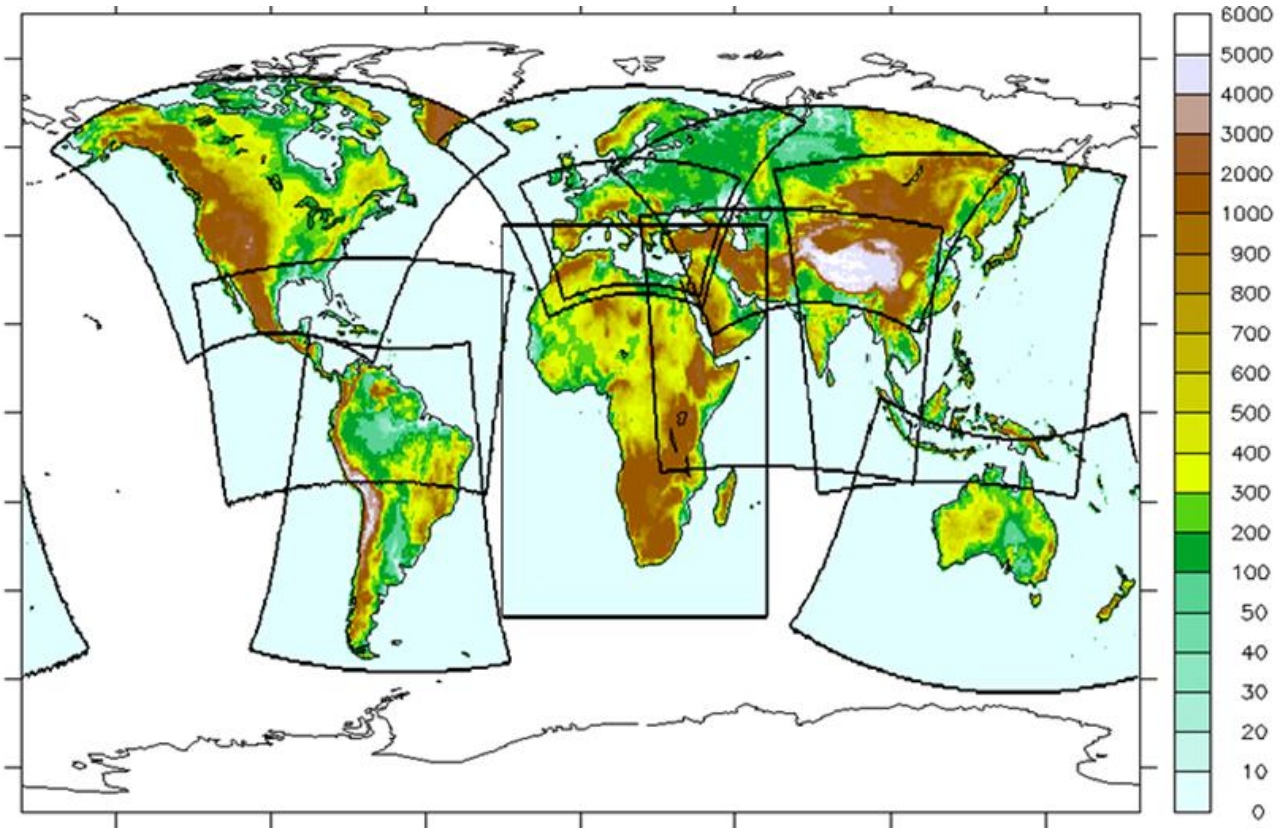


Figure 1. CORDEX domains and topography.

Despite the increasing use of regional climate model (RCM) simulations in climate-change impact studies, their application is challenging due to the risk of considerable biases (Teutschbein and Seibert, 2012). To deal with these biases, several bias correction methods have been developed recently, ranging from simple scaling to rather sophisticated approaches. The tool for bias correction was developed in Texas T&M University under support of ICARDA.

THE EARTH SYSTEM FEDERATION

The Earth System Grid Federation (ESGF) is a multi-agency, international collaboration of people and institutions working together to build an open source software infrastructure for the management and analysis of Earth Science data on a global scale.

- ANL Node
- BADC Node
- BNU Node
- CMCC Node
- DKRZ Node
- DKRZ CMIP5 Node
- NOAA-GFDL Node
- IPSL Node
- NASA-GSFC Node
- NASA-JPL Node
- NCI Node
- NERSC Node
- ORNL Node
- PCMDI Node

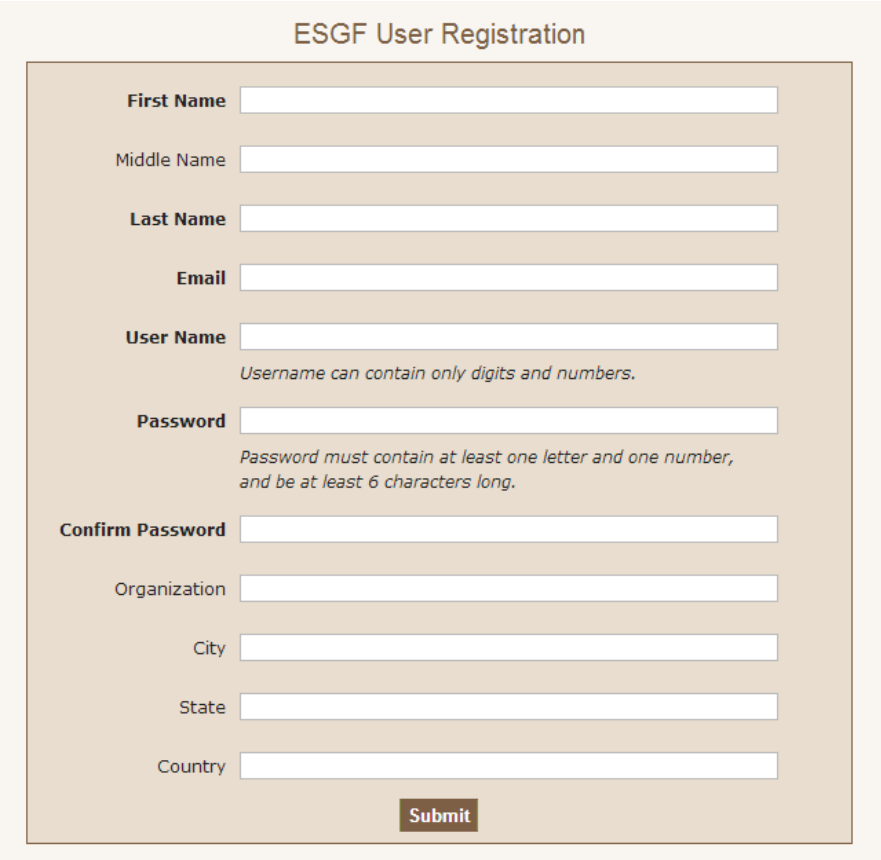
Data and metadata are published, stored and served from multiple centers "Nodes". There is no difference which of nodes you use to download data, however some nodes can fail in your region, but it does not mean that the system does not work at all, so if you have problems with one node, try others. For instance for Central Asia best choice is BADC node (http://esgf-index1.ceda.ac.uk) or CMCC node (http://adm07.cmcc.it/esgf-web-fe/)

! For successful work with ESGF sites (registration, searching and downloading the information)you should use one of the following browsers: Firefox 7+; Chrome 16+

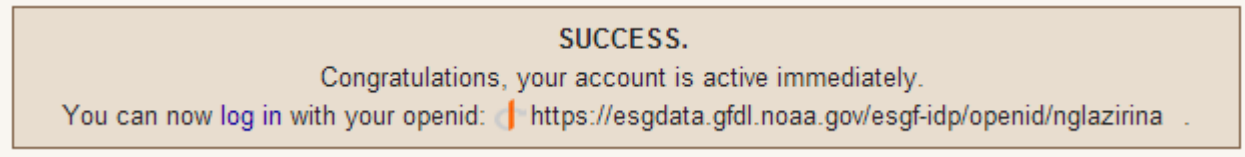
1. REGISTRATION

To download any information from the Web-site you need to create an account

When you click on "Create Account" security alert appears. Ignore it (click on "continue"). Fill in the registration form

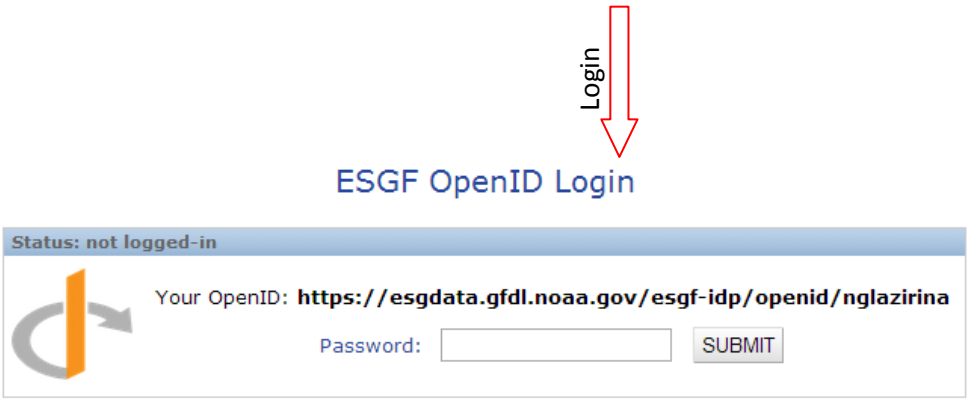


When you submit the form successfully the message appears



The link in the end is your username to login onto any of ESGF Nodes/sites. It will be sent also to the e-mail address that you provided during the registration.

Now you can login



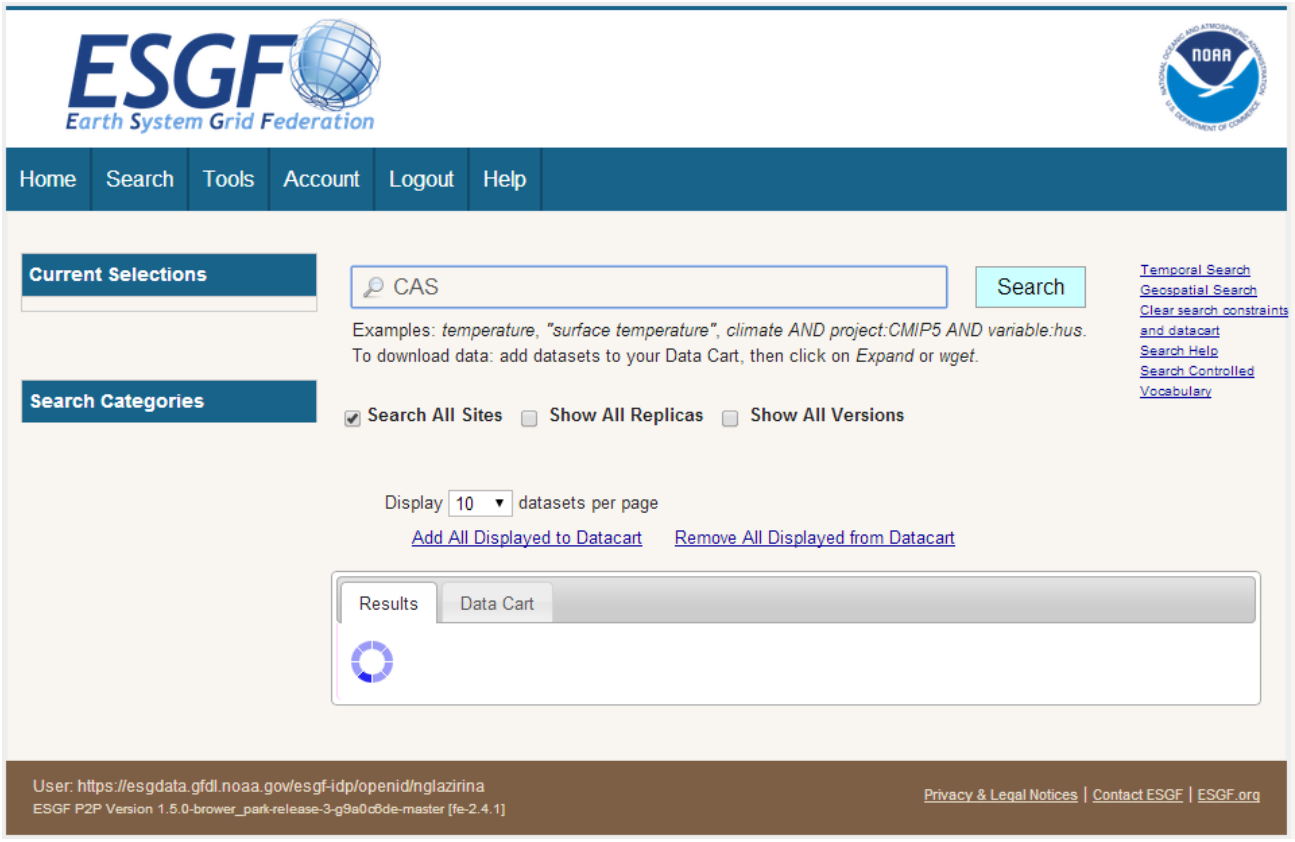
Enter the password you chose during registration and submit. If you logged in successfully, the page of you account opens

2. SELECTION OF THE DATA TO DOWNLOAD

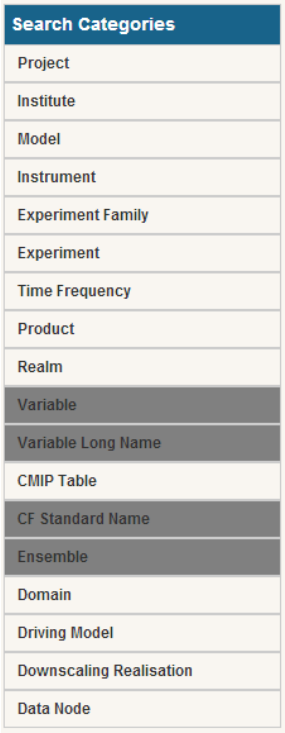
Click Search and set your domain (area of interest) in the Search field. For example domain for Central Asia is CAS, for Middle East and North Africa - MNA, etc. (the complete listing of domains you can find on page 10-11 of http://cordex.dmi.dk/joomla/images/CORDEX/cordex_archive_specifications.pdf).

You should keep in mind that the CORDEX domains are defined as rectangles in rotated-pole coordinates, but core variables are provided on a regular geographic latitude/longitude grid also (this grids are marked with i, for instance CAS-44 is in rotated-pole coordinates, CAS-44i is in regular geographic coordinates). However, SWAT-programs (CORDEX2SWAT or CORDEX2SWAT-bias) for data extraction can be used only for datasets in rotated-pole coordinates.

Select ***Search All Sites*** and click on Search



Then the panel with other search categories activates, however you should keep in mind that only options available for your previous selections are visible.



Main selections you can make are following:

- domain

- time frequency (daily, monthly, seasonal)

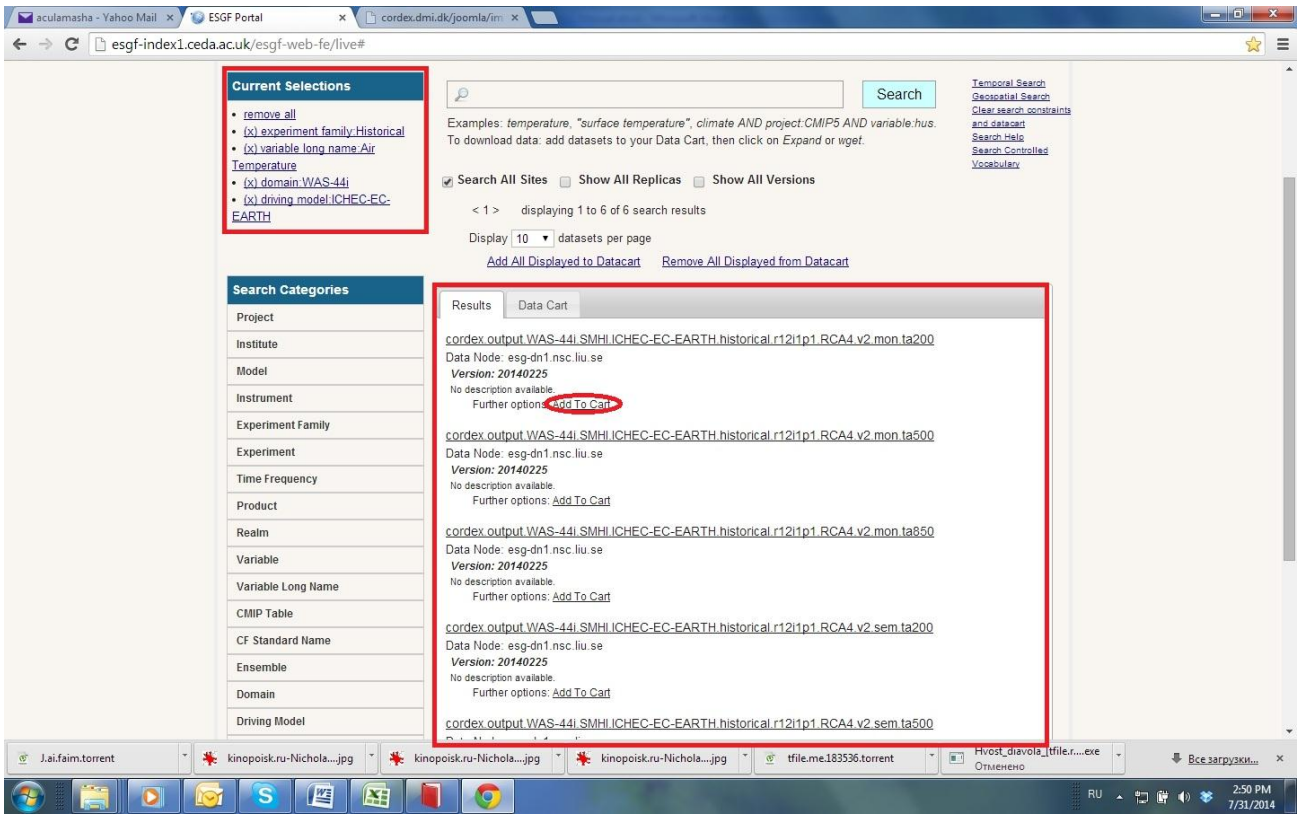
- experiment or experiment family (historical, evaluation and rcp, the information concerning rcp can be found on http://www.pik-potsdam.de/~mmalte/rcps/)

- variable or variable long name (see page 16-18 of http://cordex.dmi.dk/joomla/images/CORDEX/cordex_archive_specifications.pdf)

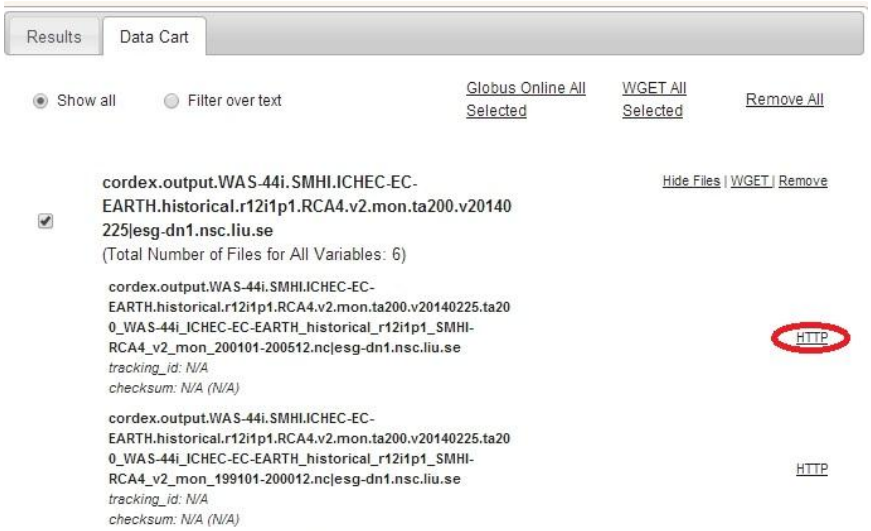
- driving model.

Datasets corresponding to your selection are listed in the tab Results (see below)

Those you want to download you should ***Add to Cart***

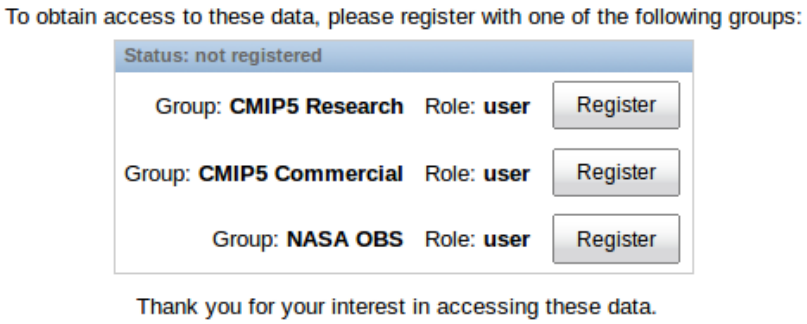


To brows the files you are interested in and download them you should open Data Cart tab and click on Show files (file naming is described below)



and you click on HTTP to download the file.

When you download data for the first time, you are asked to register yourself with one of groups. You should chose the CMIP5 Research group.



If you use Windows operating system, you have to download files one by one. For Linux the software is available to get a set of files.

To simplify your further work create a well structured database of your climate change information, where at least all the files for a domain are located in one folder, as Software processing this files cannot work with a number of folders at a time.

FILE NAMING

The names of the files in CORDEX archives are made up of the elements described below. The elements are separated by underscores ('_') and appear in the following order

VariableName_Domain_GCMModelName_CMIP5ExperimentName_CMIP5EnsembleMember_
RCMModelName_RCMVersionID_Frequency[_StartTime-EndTime].nc

VariableName is the name of the target variable:

tas - near-surface air temperature (mean at a height between 1.5 to 10.0 m);
tasmax - daily-maximum near-surface air temperature (mean at a height between 1.5 to 10.0 m);
tasmin - daily-minimum near-surface air temperature (mean at a height between 1.5 to 10.0 m);
pr - precipitadtion;
ps - surface pressure;
psl - mean sea level pressure;
hurs - near-surface relative humidity (mean at a height between 1.5 to 10.0 m);
huss - near-surface specific humidity (mean at a height between 1.5 to 10.0 m);
sfcWind - near-surface wind speed (mean at a height between 1.5 to 10.0 m);
sfcWindmax - daily-maximum near-surface wind speed (mean at a height between 1.5 to 10.0 m);
clt - total cloud cover;
sund - sunshine hours (duration when surface solar radiation flux exceeds 120 W/m2);
rsds - surface downwelling shortwave radiation;
rlds - surface downwelling longwave radiation;
hfls - surface latent heat flux;
hfss - surface sensible heat flux;
rsus - upwelling surface shortwave radiation;

rlus - upwelling longwave radiation;
evspsbl - surface evaporation;
evspsblpot - potential evapotranspiration;
mrfso - soil frozen water content;
mrros - surface runoff;
mrro - total runoff;
mrso - total soil moisture content;
snw - surface snow amount;
snm - snow melt;
prhmax - maximum 1-hour precipitation rate within 24 hour period;
prc - convective precipitation;
rlut - top of atmosphere (TOA) outgoing longwave radiation;
rsdt - TOA incident shortwave radiation;
rsut - TOA outgoing shortwave radiation;
uas - eastward near-surface wind (mean at a height between 1.5 to 10.0 m);
vas - northward near-surface wind (mean at a height between 1.5 to 10.0 m);
wsgsmax - maximum near-surface gust wind speed (mean at a height between 1.5 to 10.0 m);
tauu - surface downward eastward wind stress;
tauv - surface downward northward wind stress;
ts - surface (skin) temperature;
zmla - atmospheric boundary layer thickness;
prw - column water vapour;
clwvi - column condensed (liquid+ice) water content;
clivi - column ice water content;
ua850 - zonal (eastward) wind at 850 hPa;
va850 - meridional (northward) wind at 850 hPa;
ta850 - temperature at 850 hPa;
hus850 - specific humidity at 850 hPa;
ua500 - zonal (eastward) wind at 500 hPa;
va500 - meridional (northward) wind at 500 hPa;
zg500 - geopotential height at 500 hPa;
ta500 - temperature at 500 hPa;
ua200- zonal (eastward) wind at 200 hPa;
va200- meridional (northward) wind at 200 hPa;
ta200 - temperature at 200 hPa;
zg200 - geopotential height at 200 hPa;
clh - high clouds (p<440hPa);
clm - medium clouds (680 hPa > p > 440 hPa);
cll - low clouds (p>680hPa); the lower plev_bnds value of cll has to be 100000Pa; actually the cloud fraction between 68000Pa and the surface is requested;
snc - snow area fraction;
snd - snow depth;
sic - sea ice area fraction;
prsn - snowfall flux;
areacella - atmosphere grid-cell area;
orog - Surface Altitude;
sftlf - land area fraction;
sftgif - fraction of grid cell covered with glacier;
mrsofc - capacity of soil to store water;
rootd - maximum root depth.

Domain is the name assigned to each of the CORDEX regions (http://wcrp-cordex.ipsl.jussieu.fr/images/pdf/cordex_regions.pdf)

GCMModelName is an identifier of the driving data. The name consists of an institute identifier and a model identifier. For reanalysis driven runs these are ECMWF and a name for the reanalysis data (ERAINT). For runs driven by CMIP5 model data these are the associated CMIP5 institute_id and the CMIP5 model_id. The two parts of the name are separated by a dash. The list of GCMModelName can be found at http://cordex.dmi.dk/joomla/images/CORDEX/GCMModelName.txt

CMIP5ExperimentName is historical, evaluation or the value of the CMIP5 experiment_id (rcp2.6, rcp4.5, rcp8.5) of the data used.

CMIP5EnsembleMember identifies the ensemble member of the CMIP5 experiment that produced the forcing data. It has same value in CORDEX as in CMIP5

RCMModelName is an identifier of the CORDEX RCM (regional circulation model). It consists of the Institution identifier (an identifier for the institution that is responsible for the scientific aspects of the CORDEX simulation) and a model acronym, connected by a dash (e.g. DMI-HIRHAM5 or SMHI-RCA4). The CV of the RCMModelName has to be coordinated in the worldwide CORDEX community. Full list of the models can be found at http://cordex.dmi.dk/joomla/images/CORDEX/RCMModelName.txtm

RCMVersionID identifies reruns with perturbed parameters or smaller RCM release upgrades, i.e. equivalent simulations. Major upgrades and improvements are reflected in the RCMModelNamen.

Frequency is the output frequency indicator: 3hr=3 hourly, 6hr=6 hourly, day=daily, mon=monthly, sem=seasonal, and fx=invariant fields.

RCM DATA PROCESSING

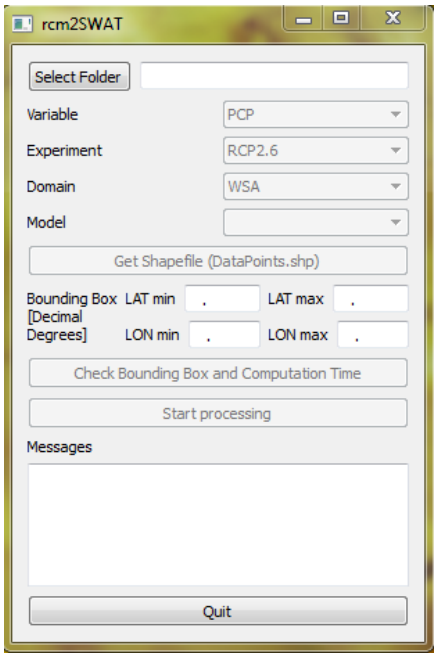
The software RCMEXTRACT is for extracting the data from data-files downloaded from ESGF portal in the form of daily time series for historical or future period for each grid cell of the area of interest. During one run it processes files for selected variable, experiment, domain and model.

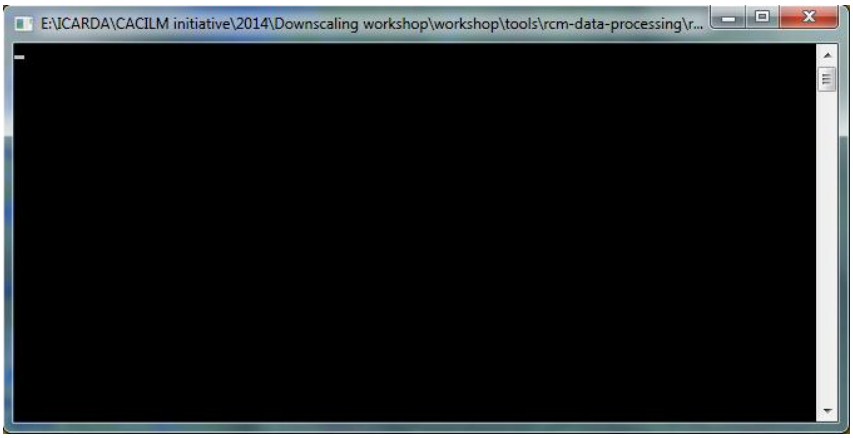
Download software from (link to Web-site).

Software does not require installation!

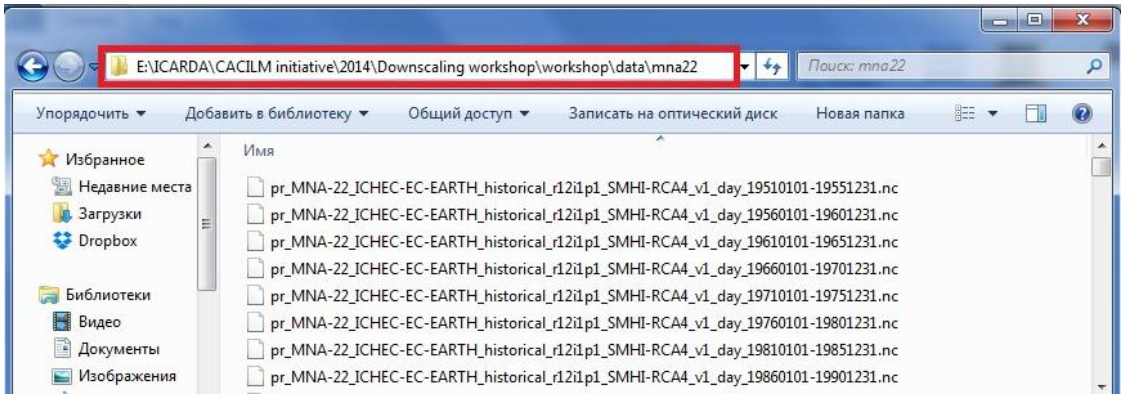
Open folder rcmextract and launch **swat_01.exe**.

Two windows opens. First one is for settings, and in the second one the program progress messages are displayed



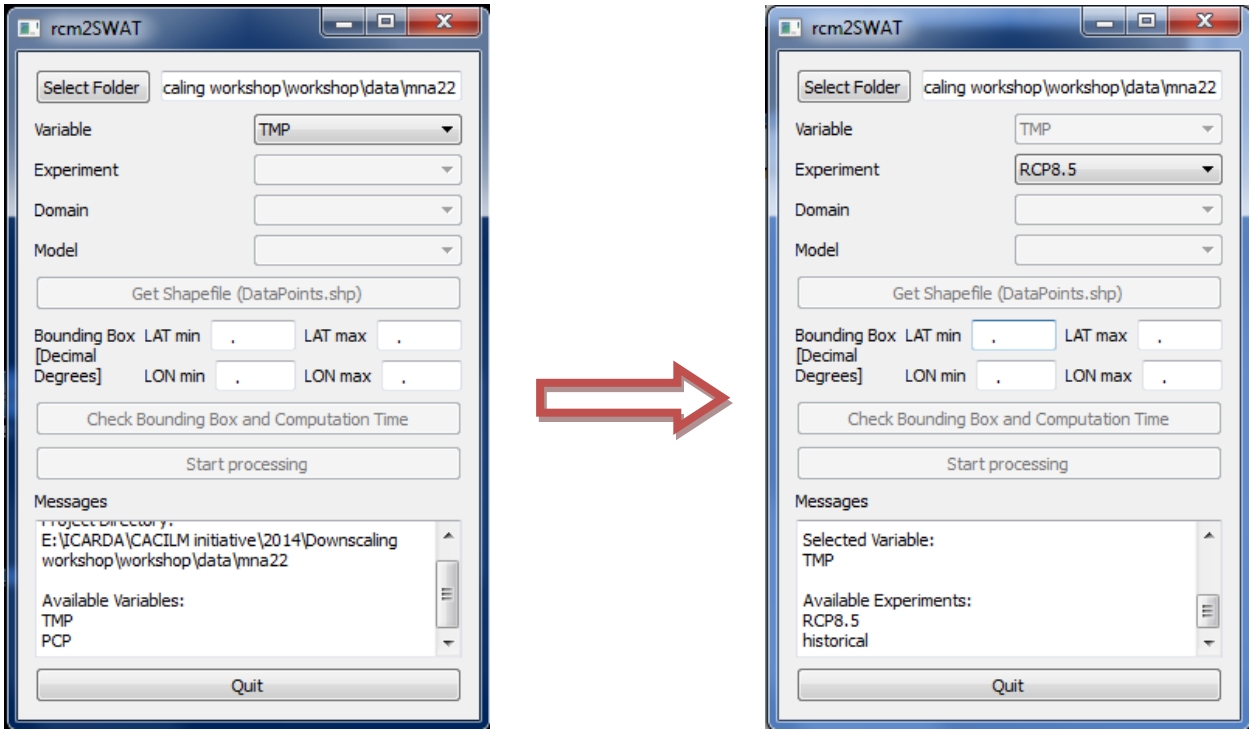


To process data you should set the path to the folder containing data downloaded from ESGF portal. Keep in mind that the Software does not recognize sub-folders, so all the data you want to process have to be located in one main folder

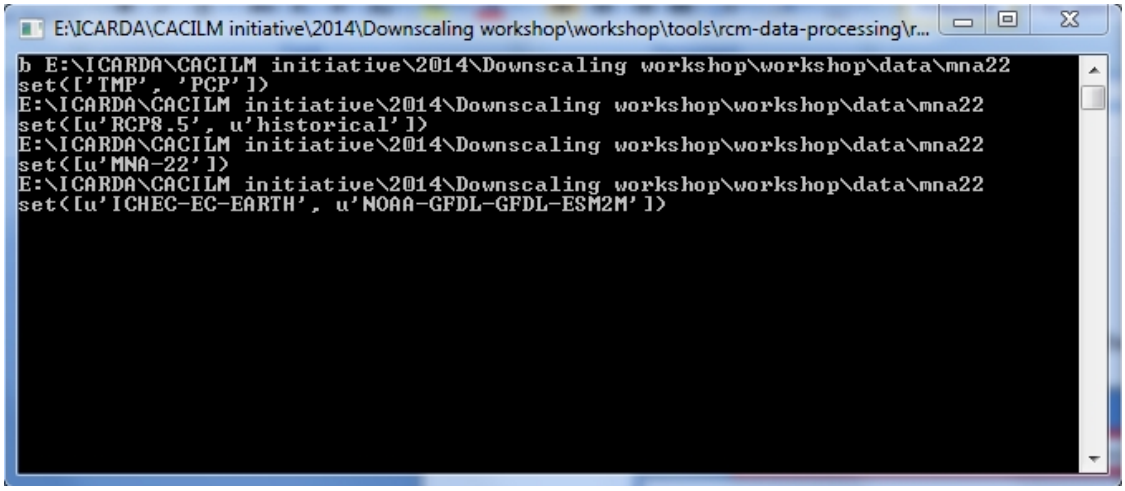


When you select folder the program analyses the folder content and display variables contained, i.e. for which data can be extracted

To make selection you should click on the selected item even if it is indicated by default. Then program processes you selection and activates next option



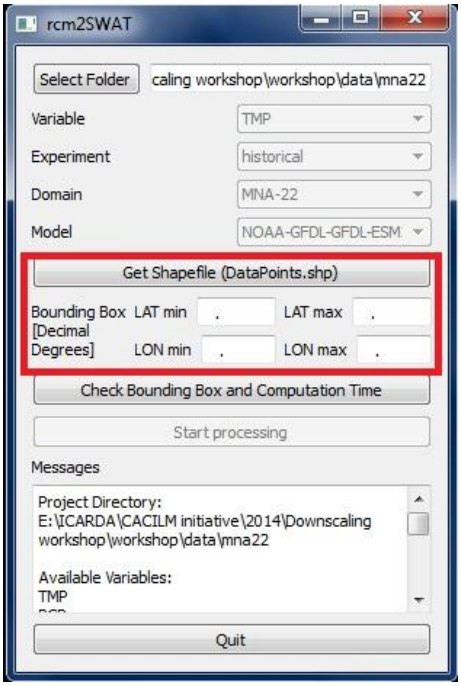
Each step is reflected on the program panel

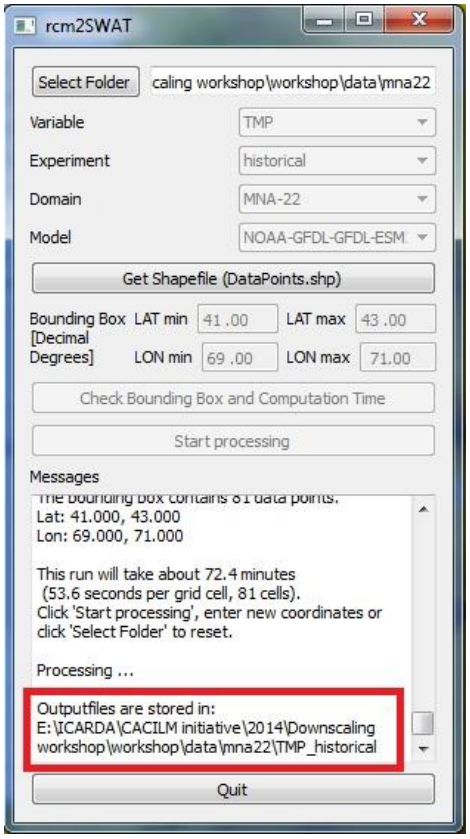


After selection of variable, experiment, domain and model, you should set the area of interest for which you want to extract data. You can do it two ways:

- using shape-file with points, for which you want to extract data (experimental sites, meteorological stations, etc.);

- bounding area using geographical coordinates (decimal degrees).





After specifying the area of interest the Software checks how many grid cells will be extracted and estimates time necessary for this.

You can run the extraction, make new settings or specify other area.

The extracted data are stored in the newly created sub-folder of the folder you set as containing initial data. It has got name according to your selection of variable and experiment (in our sample is TMP_historical).

The sub-folder contains location-file (TMP_historical_loc.txt) where you can find coordinates and altitude of all the selected points) and data-files created for each selected point and starting from date of the beginning of time series.

Sample of location file:

```
ID,NAME,LAT,LONG,ELEVATION
1,TMP_0001,41.03000000000001,69.19000000000000,100.0000000000000
2,TMP_0002,41.03000000000001,69.41000000000000,100.0000000000000
3,TMP_0003,41.03000000000001,69.63000000000000,100.0000000000000
4,TMP_0004,41.03000000000001,69.84999999999999,100.0000000000000
5,TMP_0005,41.03000000000001,70.06999999999999,100.0000000000000
6,TMP_0006,41.03000000000001,70.28999999999999,100.0000000000000
7,TMP_0007,41.03000000000001,70.50999999999999,100.0000000000000
8,TMP_0008,41.03000000000001,70.72999999999999,100.0000000000000
9,TMP_0009,41.03000000000001,70.94999999999999,100.0000000000000
10,TMP_0010,41.25000000000000,69.19000000000000,100.0000000000000
11,TMP_0011,41.25000000000000,69.41000000000000,100.0000000000000
12,TMP_0012,41.25000000000000,69.63000000000000,100.0000000000000
….
```

Sample of data file for temperature:

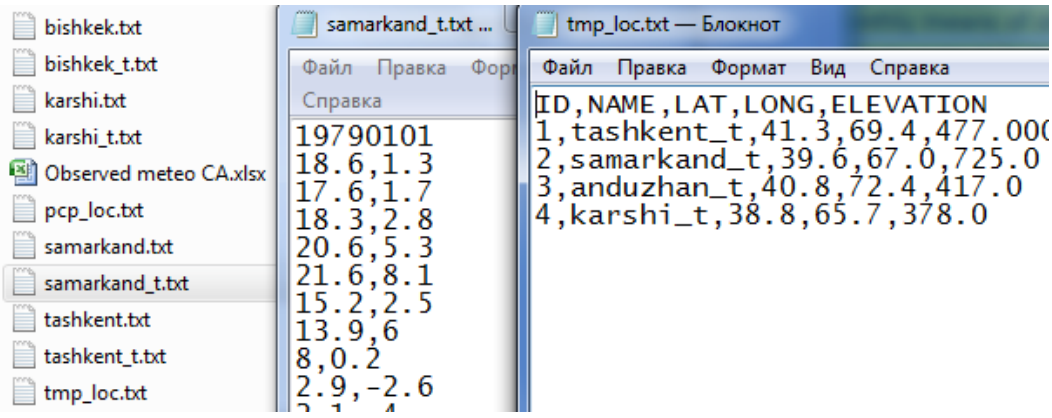
```
19510101
0.00,0.00
0.00,0.00
0.00,0.00
0.00,0.00
0.74,0.00
3.16,0.00
5.04,0.00
5.20,0.96
4.50,1.32
6.08,0.40
3.59,0.00
2.80,0.00
7.72,0.39
7.75,0.00
6.80,0.00
4.91,2.81
7.86,2.02
8.59,0.99
8.51,0.00
0.00,0.00
0.00,0.00
0.00,0.00
0.00,0.00
```

```
1.04,0.00
4.63,0.20
1.40,0.00
0.03,0.00
0.00,0.00
0.00,0.00
0.00,0.00
1.24,0.00
0.00,0.00
0.00,0.00
0.00,0.00
```

BIAS CORRECTION

As values of the meteorological parameters projected by different GCMs can have biases even after downscaling, it is reasonable to make bias correction based on observed values for historical period. For this purpose observed and simulated (downloaded from ESGF portal) daily data for historical period averaged over months using CORDEX2SWAT_bias software and compared for each available meteorological station. The software identifies the nearest climate model grid point for each available meteorological station and derives correction factors (or differences) for each month by comparing monthly means of observed and RCM/GCM data. Afterwards it corrects the RCM/GCM data and saves the climate output in the form of time series for separate grid cells. Besides it saves the correction factors (or differences).

For bias correction one should have data of observation on meteorological parameters to be corrected, which were obtained from meteorological stations located in territory of interest. Long-term data on each parameter for each station should be saved as a separate text-files with meaningful names, for instance station name_parameter name.txt. File content starts from date of first observation, following lines contain daily measured values (missing values are signed as -99.9). Be aware that temperature requires two values per line (maximum and minimum temperature) separated by a coma. Names of data files for each parameter along with station ID, latitude, longitude, and elevation of each station must be listed in correspondent loc-file (pcp_loc.txt, tmp_loc.txt, etc.). The station location file and all data-files of a variable must be stored in one folder.

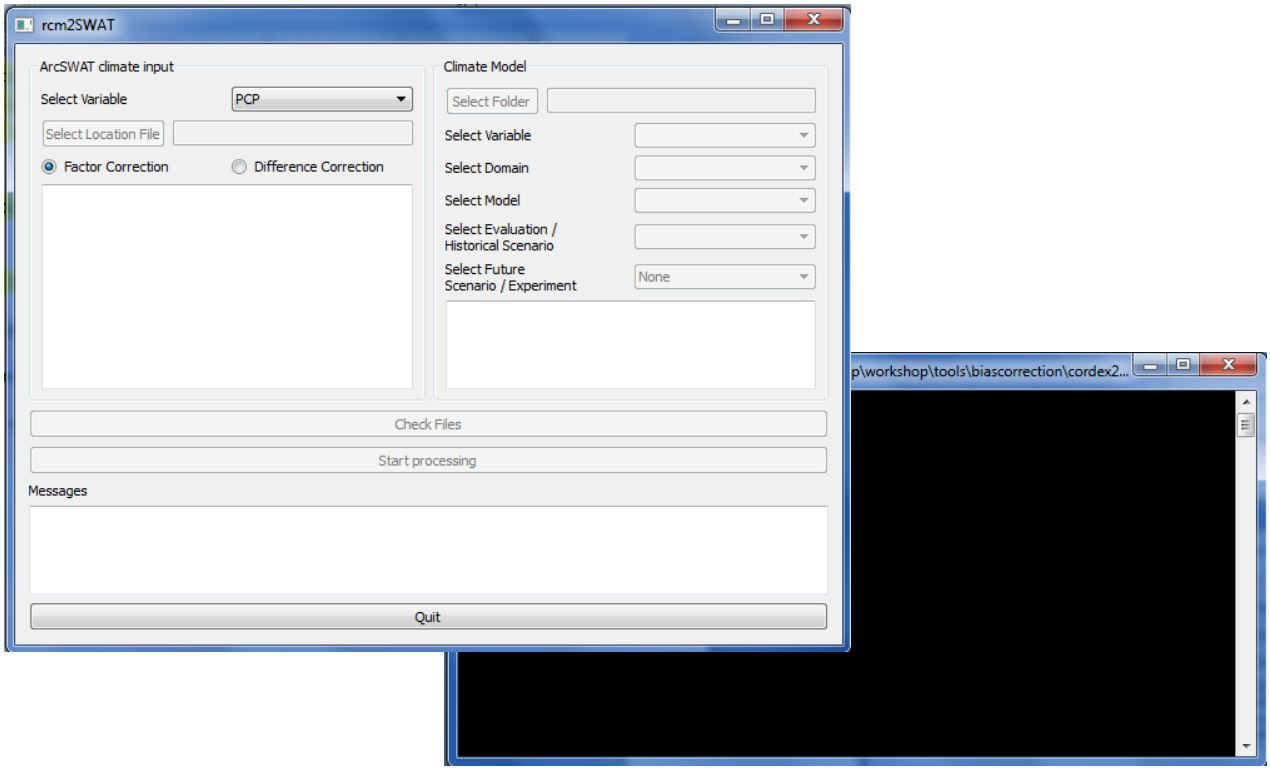


Download software from (link to Web-site).

Software does not require installation.

Open folder cordex2swat_bias and launch **swat_c_01.exe**.

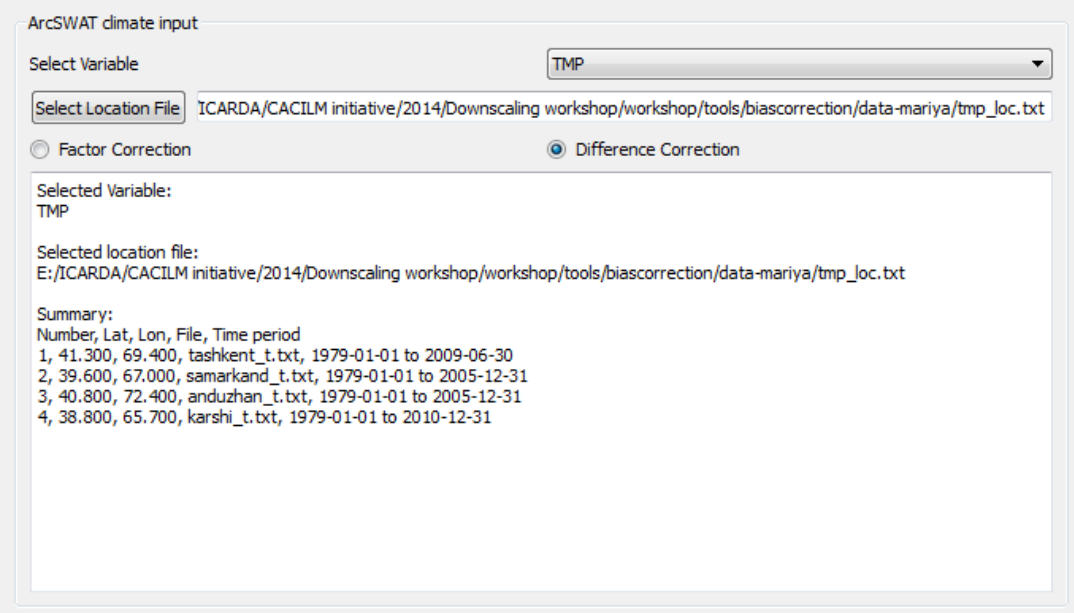
Two windows opens. First one is for settings, and in the second one the program progress messages are displayed



First you should make settings for observed data in the left side of the window (to make selection you should click on the selected item even if it is indicated by default, then program processes you selection and activates next option):

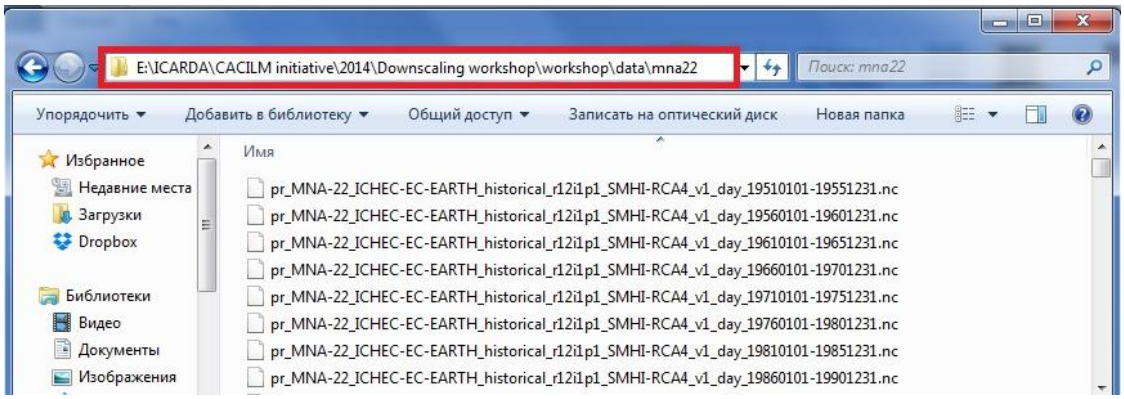
- select variable;
- select location file;
- type of correction (depending on variable: factor for precipitation and difference for temperature).

After selection you can look through summary of observed data available for comparison and bias-correction



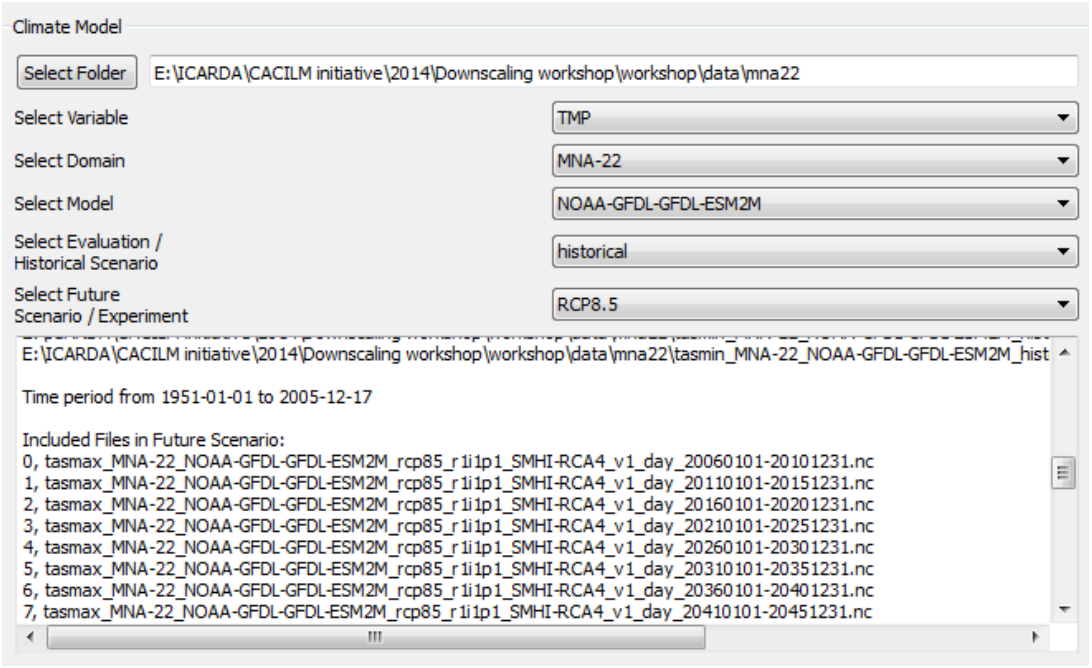
Next step is to make settings for the data downloaded from ESGF portal:

- set the path to the folder containing the data downloaded from ESGF portal (keep in mind that the Software does not recognize sub-folders, so all the data you want to process have to be located in one main folder, besides it can take long time, as the software analyses the folder content)

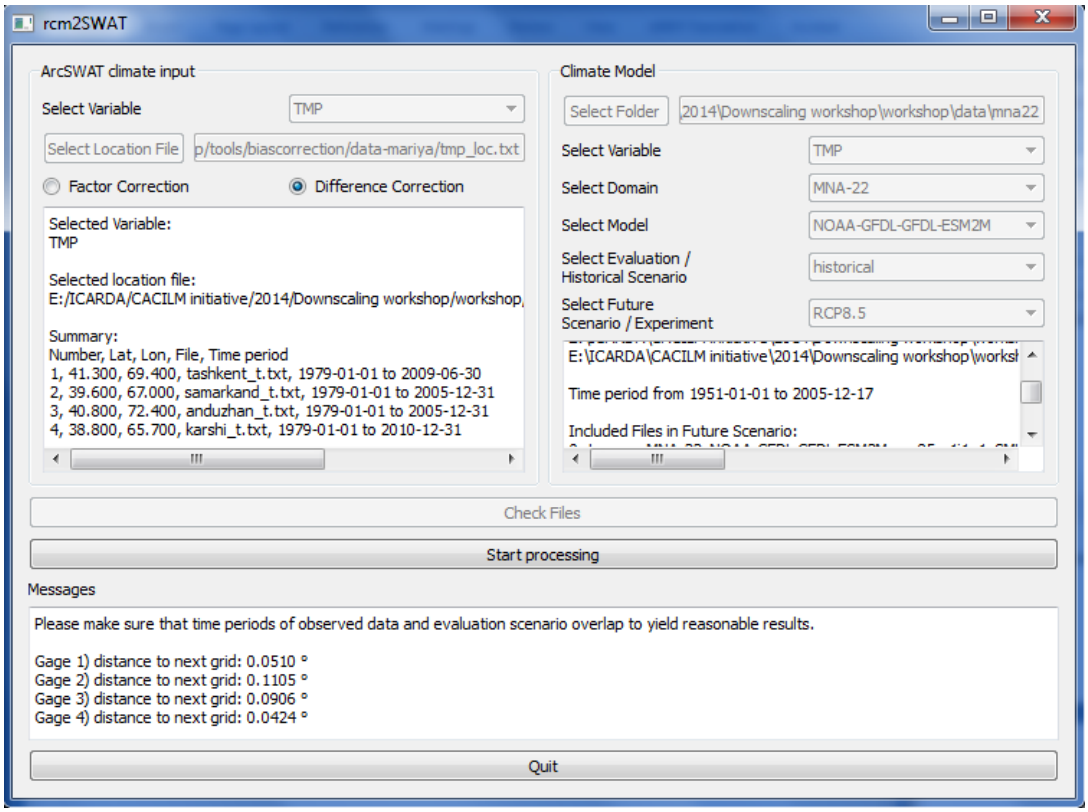


from the options available in the folder you set
- select variable (it must agree with variable chosen in the left part)
- select domain
- select model (GCM)
- select historical scenario (although there are other options, you should chose "historical" for bias correction)
- select future scenario to be corrected.

After selection you can look through summary of the selected modeled data

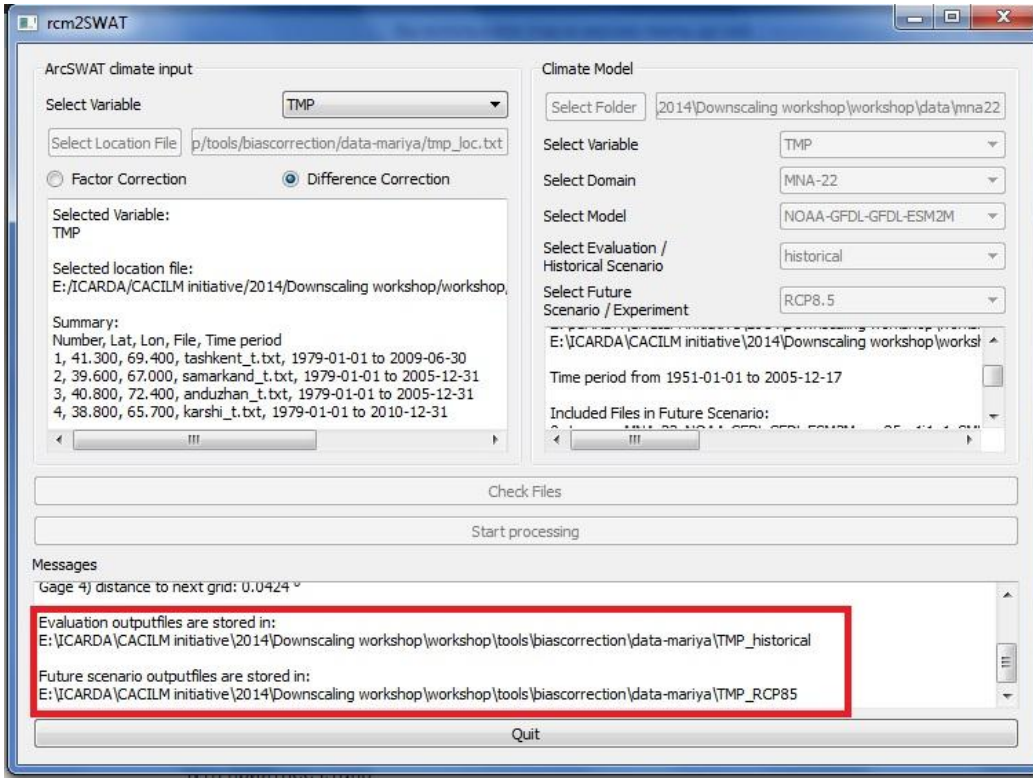


Now you need to check files (the software identifies rid cells closest to the available meteorological stations)

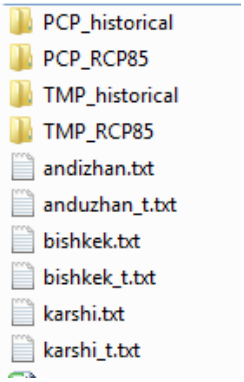


and Start processing

Processing takes rather long time. After it information about correction results appears in the window "messages"



The tool saves the output files in sub-folders of the folder containing observed data. They are named after the variable and experiment. The program generates one data file for each meteorological station. It also provides the correction factors (or differences) (factor_<gagename>.txt) and the station location summary file (*_loc.txt) that can be used in ArcSWAT.



The tool provides bias corrected and original RCM/GCM data.

Be aware that data from the meteorological stations may be influenced by local features and are not necessarily representative for a grid cell of the climate model. Compare RCM/GCM and observed data before using the data for hydrologic simulations.

REFERENCES

1. Evans, J.P., 2011. *CORDEX – An international climate downscaling initiative. 19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011*. http://mssanz.org.au/modsim2011

2. Giorgi, F., C. Jones, and G. R. Asrar (2009), *Addressing climate information needs at the regional level: the CORDEX framework, WMO Bulletin*, 58(3), 175-183.

3. IPCC (2007), *Climate Change 2007: The Physical Science Basis*, edited by S. Solomon, edited by , D. Qin and edited by , M. Manning, edited by , Z. Chen, edited by , M. Marquis, edited by , K.B. Averyt, edited by , M. Tignor, edited by , H.L. Miller, Cambridge University Press, Cambridge, United Kingdom.

4. Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007. *The WCRP CMIP3 multimodel dataset - A new era in climate change research, Bull. Amer. Meteor. Soc.*, 88(9), 1383-1394.

5. Teutschbein, C., J. Seibert, 2012. *Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology*, V.456–457, 12–29

6. Uppala S., Dee D., Kobayashi S., Berrisford P., Simmons A. (2008): *Towards a climate data assimilation system: status update of ERA-Interim*. ECMWF Newsletter No. 115, 12-18.