



Collaborative Climate Community Data and Processing Grid (C3-Grid)

Education Workflow

Work Package:	cross-WP - WG Workflows
Author:	M. Stockhause
Version:	V0.1/2008
Publishing Date:	05/2008
Pages:	6
WP Coordination:	-
Partners:	MPI-M, IFM-GEOMAR, FUB, ZAIK, DLR, AWI, PIK, Robert-Bosch-Stiftung
Contact:	M. Stockhause
Email:	martina.stockhauseATzmaw.de



Content

1	Introduction.....	3
2	Functionality	3
2.1	Description and Restrictions	3
2.2	Usage	4
2.3	Input Data	5
2.4	Outlook	5
3	Installation at CP site.....	5
3.1	Scripts.....	5
3.2	Tools.....	5
4	References	5



1 Introduction

In discussion with the MPI-M's leader of the project "Klimawandel und seine Folgen" ("Climate Change and its Consequences"; <http://klimaprojekt.de>) the workflow 'Education' is developed. The project is a co-operation between scientific institutes (MPI-M, DKRZ, GKSS) and schools in the region of Hamburg, financed by the Robert-Bosch-Stiftung (Foundation). Pupils elaborate on self-defined climate topics; they are supported by scientists and students. At the moment an intensive attendance of the pupils is necessary as well as their presence at the scientific institutes during a considerable amount of time to work on the climate data. This is caused by the binary data formats and the unix tools for simple processing and selection operations on these raw data. However, the pupils' home computers are normally equipped with Windows operating systems. In addition, the elaborated topics cover a wide range of climate related questions.

The aim of the establishment of a C3Grid workflow for this project is to enable the pupils to work more self-dependently, i.e. less attended by the project leaders and more at home. Momentarily, the concept for such an 'Education' workflow is under discussion. Ease of use and the required high flexibility in possible application scenarios have to be balanced carefully.

2 Functionality

2.1 Description and Restrictions

The appropriate workflow for all possible applications cannot be set-up or at least not at once. The first realization step concentrates on the derivation of different statistical values: seasonal, multi-year seasonal, monthly, multi-year monthly, and temporal (over whole given time period) means, sums (for precipitation quantities), standard deviations, minimum, and maximum. These are performed using `cdos` [1] and are related to the `cdo` command families: ***seasX***, ***yseasX***, ***monX***, ***ymonX***, ***timX*** where X stands for '*mean*', '*sum*', '*std*', '*min*' or '*max*'.

There are specific quantities selected, for which these statistical values are provided: ***air_temperature-at2m***, ***precipitation_flux***, ***snowfall_flux***, ***water_evaporation_flux***, ***wind_speed***, ***sea_surface_temperature***, ***lwe_thickness_of_surface_snow_amount***. Since the statistical value 'mean' is not appropriate for precipitation quantities, for these the 'sum' is calculated instead. An overview over the functionality of the education workflow is given in **Table 1**.

The application is restricted to ECHAM5 output data at the moment.

Along with the selected data an accompanied metadata file in the C3Grid profile of ISO 19115 / 19139 format is prepared [1]. The script uses the agreed environment variables for the workspace [3].



CF output parameter	unit	input CFs	input data preparation	mean / sum
air_temperature-at2m	K	air_temperature-at2m		mean
precipitation_flux	kg m-2 s-1	convective_precipitation_flux, large_scale_precipitation_flux	RAW: after BOT: 142,143 for 260; cdo -setcode,260 -add - selcode,142 <in.grb> - selcode,143 <in.grb> <out.grb> CERA-E5: cdo -selcode,4 (260:TPREC)	sum
snowfall_flux	kg m-2 s-1	snowfall_flux		sum
lwe_thickness_of_ surface_snow_amount	m	lwe_thickness_of_ surface_snow_amount		mean
water_evaporation_flux	kg m-2 s-1	water_evaporation_flux		sum
sea_surface_temperature	K	sea_surface_temperature		mean
wind_speed	m s-1	eastward_wind, northward_wind, (RAW: after ATM: atmosphere_relative_vorticity, divergence_of_wind)	RAW: after ATM: 138,155 for 131,132,259; cdo -setcode,259 -sqrt -add -sqr -selcode,131 <in.grb> -sqr -selcode,132 <in.grb> <out.grb> CERA: cdo -setcode,259 - sqrt -add -sqr <in_u.grb> - sqr <in_v.grb> <out.grb>	mean

Table 1: CF standard names for which the statistical values of the education workflow are provided, their units, input CF standard names and derivation, and if 'mean' or 'sum' is calculated for 'mean' option (cf. section 0).

2.2 Usage

usage: python education_wf.py -i <inputfiles> -l <logfile> -m <metainfile> -p <processings> -r <temporal basis> -d <dir infiles> -o <dir outfiles> -f <outfile>

- p processings: 'min','max','mean', **and** 'std' possible with performance of mean or sum respectively for 'mean' option (cf. **Table 1**)
- r temporal basis: 'M' (monthly), 'YM' (multi-year monthly), 'S' (seasonal), 'YS' (multi-year seasonal), **or** 'T' (temporal: whole time period)
- i name of input files as ','-separated list
- l name of logfile in <dir outfiles>
- m name of input metadata files as ','-separated list (in same order as inputfiles)
- d path to directory of input files
- o path to directory of output files
- f name of output file

Example: multi-year monthly mean, standard deviation, minimum, and maximum for the CF standard names in input files edu_in*.grb:

```
python education_wf.py -i edu_in1.grb,edu_in2.grb -l edu1.log -m edu_in1.xml
-p min,max,mean,std -r YM -d /path/in -o /path/out -f edu_out1.grb
```



The script `education_wf.py` analyses the command line and reads an initialization file to look up the needed processing (information of performing mean or sum and commands to be run to derive output CF standard name; cf. **Table 1**).

2.3 Input Data

The education workflow can handle only ECHAM5, direct model output (raw data) and post-processed data available from CERA database, which was selected using the C3Grid data request interface (interface D; [4]).

2.4 Outlook

- Flexibility improvement / enhancement
- Additional quantities, e.g. `beaufort_wind_force`, `sea_surface_height_above_sea_level`, `global_average_sea_level_change`, etc.
- Unit conversions
- Additional statistical values

3 Installation at CP site

3.1 Scripts

There are two scripts needed that have to be placed in the same directory:

- `education_wf.py` : execution script
- `meta_readISO.xml`: read necessary metadata information out of metadata file
- `meta_writeISO.xml`: metadata update script called by `education_wf.py`
- `education_proc_echam5.lst`: initialization file for parameter specific processing (information of **Table 1**).

The python version should be at 2.4 or newer. (It should work with 2.1 as well but this is not tested.)

3.2 Tools

The following tools have to be installed for application:

- `cdo` (climate data operators - <http://www.mpimet.mpg.de/~cdo>)
- `xmlstarlet`

These tools can be either installed in a directory in the search path or loaded by modules during execution from `os.environ['MODULESHOME']+'/init/python'`.

4 References

- [1] U. Schulzweida:
CDO user's Guide
<http://www.mpimet.mpg.de/~cdo/cdo.pdf>



- [2] S. Kindermann:
C3Grid ISO Metadata Profile
http://www.c3grid.de/documents/c3grid_metadata.pdf
http://www.c3grid.de/documents/C3Datenkatalog-V_1_0.pdf
http://www.c3grid.de/documents/xml-schema_G0-C3GRID.xml
- [3] T. Langhammer, V. Achter, and V. Winkelmann:
Richtlinien für die Konfiguration des Workspaces [German]
http://www.c3grid.de/documents/Konfiguration_Workspace_deliverable.pdf
- [4] T. Langhammer, F. Schnintke:
T5.1 (Supplement A): Interface Specification for C3-Grid Archives
http://www.c3grid.de/documents/interface_archive.pdf