"CLIMATE DATA OPERATORS" AS A USER-FRIENDLY PROCESSING TOOL FOR CMSAF'S SATELLITE-DERIVED CLIMATE MONITORING PRODUCTS

Frank Kaspar[1], Uwe Schulzweida[2], Ralf Müller[2]

[1]Satellite Application Facility on Climate Monitoring (CM SAF), Deutscher Wetterdienst, Frankfurter Str. 135, 63067 Offenbach, frank.kaspar@dwd.de
[2]Max Planck Institute for Meteorology, Bundesstr. 53, 20146 Hamburg

Abstract

EUMETSAT's Satellite Application Facility on Climate Monitoring (CM SAF) provides satellite-derived geophysical parameter data sets suitable for climate monitoring. CM SAF provides data on several cloud parameters, surface albedo, radiation fluxes at the top of the atmosphere and at the surface as well as atmospheric water vapour, temperature and humidity profiles. Potential applications are for instance input for climate models and validation of climate models. Although there is a growing interest in the climate modelling community to use satellite data for such validation studies, the different data formats and structures of both communities are still a drawback for such applications.

NetCDF with metadata according to the "climate and forecast (CF) metadata conventions" is a frequently used data format of the climate modelling community and is also an accepted standard of other applications. Providing satellite products according to this standard will therefore significantly increase the acceptance of such data in this community. The 'climate data operators (CDO)' are a collection of operators that were developed for processing and analysis of such data and are widely used by climate modellers. In order to allow easy access to CM SAF datasets for this community, the possibility to import CM SAF data has been integrated into the 'climate data operators'. Here we describe the concept and provide some examples for the usage of this tool.

PRODUCTS AND DATA FORMAT OF CM SAF

CM SAF exploits data from polar-orbiting and geostationary meteorological satellites in order to provide climate monitoring products of the following parameters (Schulz et al., 2009):

- Cloud parameters: cloud fraction (CFC), cloud type (CTY), cloud phase (CPH), cloud top height, pressure and temperature (CTH, CTP, CTT), cloud optical thickness (COT), cloud water path (CWP).
- Surface radiation components: Surface albedo (SAL); surface incoming (SIS) and net (SNS) shortwave radiation; surface downward (SDL) and outgoing (SOL) longwave radiation, surface net longwave radiation (SNL) and surface radiation budget (SRB).
- Top-of-atmosphere radiation components: Incoming (TIS) and reflected (TRS) solar radiative flux at top-of-atmosphere. Emitted thermal radiative flux at top-of-atmosphere (TET).
- Water vapour: Vertically integrated water vapour (HTW), layered vertically integrated water vapour and layer mean temperature and relative humidity for 5 layers (HLW), temperature and mixing ratio at 6 pressure levels

Daily and monthly mean products can be ordered via the CM SAF web page (http://www.cmsaf.eu) and are provided free-of-charge. More information on the data is available on the CM SAF homepage (http://www.cmsaf.eu). Products with higher spatial and temporal resolution, i.e. instantaneous swath-based products, can be provided on request (contact.cmsaf@dwd.de). One category of CM SAF products are compiled as first-guess products in near-real time mode, i.e., shortly after availability of the satellite observations. Their main use is for getting a quick overview of conditions for the most recent time (i.e., by localizing large anomaly patterns). CM SAF also works on the reprocessing of

long time series based on inter-calibrated radiances. These time series will be processed with constant retrieval algorithms and are therefore more suitable for climate analysis and esp. climate trend studies.

CM SAF's near-real time climate monitoring products are provided as HDF5 (Hierarchical Data Format, release 5). Reasons for selection of this data format were its high compression efficiency and the features to include several data models and self-describing datasets. However, it has become obvious that the data format as well as the projection type of the data is difficult to handle for several users. In order to facilitate the access to CM SAF data for these users, the possibility to convert and reproject the data with a user-friendly software tool is now offered. In order to minimize the effort for users (i.e. for installing and understanding additional software), this functionality was implemented into a software package that is all ready used by a large community (CDO; climate data operators; see Schulzweida et al., 2010). The concept of the tool is therefore already known to these users and the software is eventually already available at their computing centre or on their local platforms.

NETCDF AS A STANDARD DATA FORMAT IN THE CLIMATE MODELLING COMMUNITY

A widely established data format in the climate modelling community is netCDF. With the so-called "climate and forecast (CF) metadata conventions" a metadata standard for netCDF-files has been established in order to promote the processing and sharing of files in this community (see http://cf-pcmdi.llnl.gov). The CF conventions are increasingly gaining acceptance and have been adopted by a number of projects and groups as a primary standard (e.g. some large model intercomparison projects (so-called "MIPs"), like the Coupled Model Intercomparison Project CMIP5, see http://cmip-pcmdi.llnl.gov/cmip5).

Based on this standard a number of applications have been developed that allow easy handling and visualisation of such data. Many of them are free software products and thereby easily accessible to users (e.g. GrADS for visualisation of gridded netCDF data; see http://www.iges.org/grads). Providing satellite products according to this standard will therefore significantly increase the acceptance of such data in this community.

CLIMATE DATA OPERATORS (CDO)

The 'climate data operators (CDO)' are a collection of command line operators that were originally developed for processing and analysis of data produced by a variety of climate and numerical weather prediction models (e.g. for file operations, simple statistics, arithmetics, interpolation or the calculation of climate indices). Supported file formats are therefore the frequently used output formats of such models as GRIB, netCDF and several binary formats. With installations in several hundred groups world-wide, the package is widely established in the climate modelling community.

The main CDO features are:
- More than 400 operators available
- Modular design and easily extendable with new operators
- Very simple UNIX command line interface
- A dataset can be processed by several operators, without storing the interim results in files
- Most operators handle datasets with missing values
- Fast processing of large datasets
- Support of many different grid types
- Tested on many UNIX/Linux systems, Cygwin, and MacOS-X

AN IMPORT FILTER FOR CM SAF PRODUCTS

In order to allow easy access to CM SAF datasets for the climate modelling community, the possibility to import CM SAF data has been integrated into the 'climate data operators'. Besides the pure conversion of CM SAF-HDF5-files to netCDF and GRIB, this offers additional possibilities for prepossessing the data for validation studies, especially interpolation to other grid types and selection of regions. The implementation considers special features, e.g. methods for interpolation of non-continuous datasets as cloud types. Daily and monthly mean products of CM SAF are provided in equal-area projections that are described in the metadata entries of the HDF5 files. CDO reads the projection parameters from the metadata in the HDF5-headers in order to allow spatial operations like remapping. For spatial operations with CM SAF's intermediate products on original satellite projection, additional files with arrays of latitudes and longitudes are needed. These can be obtained from CM SAF together with the data.

CM SAF products are partly generated on the basis of NWC SAF's MSG- and PPS-packages (see http://www.nwcsaf.eu). Output of both packages is used to calculate the daily and monthly means of CM SAF. Processing of these intermediate products on original pixel-resolution for polar-orbiting satellites as well as geostationary satellites is also facilitated by CDO when pixel-related geolocation is available. CDO can therefore also be of interest for users of both packages.

EXAMPLE 1: CONVERSION AND REMAPPING OF DATA

Many users are interested to receive the data in different formats or projections. Here we give a few hints, how such conversions can be performed.

The pure conversion of CM SAF's HDF5 files to netCDF can be done as follows:

> *cdo –f nc import_cmsaf ifile.h5 ofile.nc*

(With "*-f nc*" the output format is set to "netCDF". "*import_cmsaf*" is the operator to import the CM SAF's HDF5 files (here: "*ifile.h5*"); "*ofile.nc*" is the name of the output file. Please note that the orginal product names from the HDF5-metadata are retained in this operation, in order to guarantee consistency with the product documentation. Product names are therefore not compliant to the CF convention).

CDO allows combining this conversion with additional operations. For example a specific field can be selected from the HDF5-file. Here, we select "cloud liquid water path":

> *cdo –f nc –selname,"clwp" –import_cmsaf ifile.h5 ofile.nc*

The data can be projected to other grids. Different methods are available. Here, we use bilinear remapping (operator: "*remapbil*"). For this operator, information on geolocation of the pixels is required. In case of the intermediate data, these have to be provided by additional files containing the latitude/longitude information (should be provided with operator "*setgrid*"):

> *cdo –f nc*
> *–remapbil,"grid_description.txt"*
> *–setgrid,cmsaf_lat_lon.hdf*
> *–selname,"clwp"*
> *–import_cmsaf ifile.h5 ofile.nc*

The description of the destination grid has to be provided in a separate text file (here: "*grid_description.txt*"). Please refer to the CDO manual for details on grid definitions. The following example defines a regular lat-lon-grid with a resolution of 0.025° for a section of the Atlantic (as u sed for Figure 1):

```
gridtype = lonlat
xsize = 320
ysize = 280
xfirst = -53
xinc = 0.025
yfirst = 20.0
yinc = 0.025
```

For data stored as scaled integer values (in the original HDF5 files), it could be desirable (or necessary) to increase the accuracy of the converted products[1]. This can be done with the parameter "*-b*". Here, accuracy is set to type 'float' with 32 bit accuracy:

```
cdo –f nc –b f32
 –fldmean
 –remapbil,"grid_description.txt"
 –setgrid,cmsaf_lat_lon.hdf
 –selname,"clwp"
 –import_cmsaf ifile.h5 ofile.nc
```

In this example, "*fldmean*" is used to calculate the mean of the data in the remapped domain. Further statistical operators can be applied. Please refer to the manual for a full list of operators.
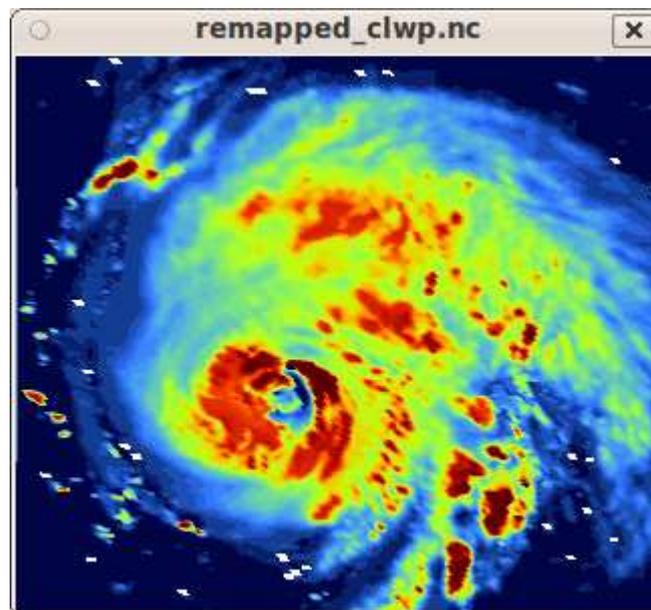

Figure 1: **An example of a plot generated with *ncview* based on a netCDF-file converted with CDO from a CM SAF product. The plot shows Hurricane Helene as visible in CM SAF cloud liquid water path data derived from SEVIRI for September, 18th 2006, 10:45 UTC. At that day, Hurricane Helene reached its peak intensity.**
CDO was used to remap the data from the original satellite projection to a 0.025° grid (with the grid description as shown in the text). The converted file can then be open with *ncview* and will be directly visualized as shown here. Footnote 2 shows the CDO call.

EXAMPLE 2: MONTHLY NUMBER OF CLEAR AND CLOUDY DAYS

CDO can also be used to perform arithmetical operations and simple statistical analysis of data. Here we show an example how the tool can be used to aggregate daily values of cloud fraction to a typical monthly product as used by meteorological services for climate monitoring. Daily values of CM SAF cloud fraction are used to derive monthly values of clear and cloudy days. Such products are typically derived from interpolated station data. A typical definition of such products could be, e.g. the number of days with cloud fraction less than 2 octas (clear days) or greater than 7 octas (cloudy days).

In the following we apply a limit of 25% (i.e. 2 octas) and 75% (i.e. 6 octas) to the CM SAF cloud fraction product to define clear and cloudy days. Two CDO calls are used to generate a product with monthly number of clear and cloudy days for Europe. In the first step, CDO is applied in a simple UNIX-Shell-Loop to remap the daily data to a European domain[3] and select the clear days (i.e. based on the 'lower equal'-operator ("-lec") a map is created where all clear days are set to 1). In a final call, all intermediate files are added up:

```
day=20100801
while [ $day -le 20100831 ]
do
  # next line: remapping and selection of clear days
  cdo -f nc -lec,25 -remapbil,grid.txt -import_cmsaf CFCdm${day}*MA.hdf CFCdm${day}_l25.nc
  day=`expr $day + 1`
done
# next line: adding up all days of the month
cdo enssum CFCdm201008??_l25.nc clear_days_201008.nc
```

Depending on the domain configuration, remapping can be computational intensive. In order to avoid repeating this time-consuming step each time within such loops, CDO allows separating the interpolation into two steps. The first step is the generation of the interpolation weights (operator "*genbil*"), which is the most time-consuming part. These interpolation weights can than be reused for every remapping process with the operator "*remap*". This method should be used only if all input fields are on the same grid and the location of missing values does not change.

The same example as before (but now for cloudy days) using a pre-calculated weighting file:

```
day=20100801
# next line: generation of interpolation weights. weights are stored in "we.nc"
cdo -f nc -genbil,grid.txt -import_cmsaf CFCdm${day}0000320070023201MA.hdf we.nc

while [ $day -le 20100831 ]
do
  # next line: remapping with weighting file and selection of cloudy days.
  cdo -f nc -gec,75 -remap,grid.txt,we.nc -import_cmsaf CFCdm${day}*.hdf CFCdm${day}_g75.nc
  day=`expr $day + 1`
done
cdo enssum CFCdm201008??_g75.nc cl_201008.nc
```

Results of both scripts are shown in Figure 2. Here we used GrADS (Grid Analysis and Display System; see http://www.iges.org/grads/) to visualize the results. GrADS is also a tool that allows visualisation of CF-compliant netCDF-files with very simple commands. It is freely available and widely used in meteorology.
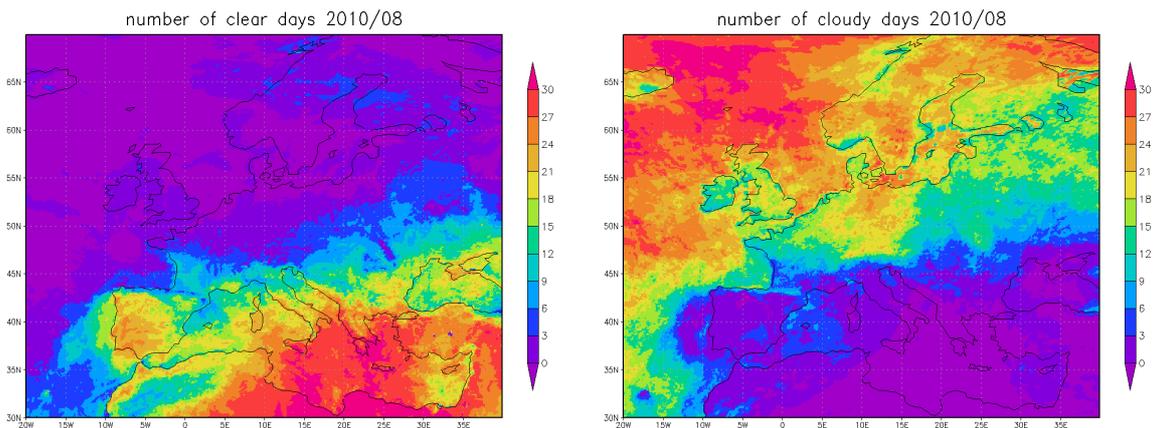
Figure 2: Number of clear and cloudy days for August 2010. The product is derived from CM SAF daily means of SEVIRI-based cloud fraction. CDO was used to create the data file (see text) which was then visualized with *GrADS*. The few simple commands for generating the plots are shown in Footnote 4.

EXAMPLE 3: MERIDIONAL PROFILE OF LAYERED WATER VAPOUR

Most CM SAF products contain two-dimensional data (as those used in the previous examples), but some also contain an additional temporal or spatial dimension. For example, cloud and radiation products are also provided as mean-monthly diurnal cycle (with hourly resolution, i.e. containing 24 monthly averages for each hour of a day). ATOVS-derived water vapour is available as a vertically layered product. Such file structures have also been considered in the 'import_cmsaf' operator. When converting such data, a netCDF-file with appropriate three-dimensional structure is generated and additional operators can be applied to process the data.

The following call selects level 5 (which is the layer from the surface to 850 hPa) of the layered product and calculates a zonal mean from the global ATOVS water vapour product:

    cdo -f nc -zonmean -sellevel,5
            -import_cmsaf HLWdm20100401000003201100171101GL.hdf HLWdm20100401_ZM_L5.nc
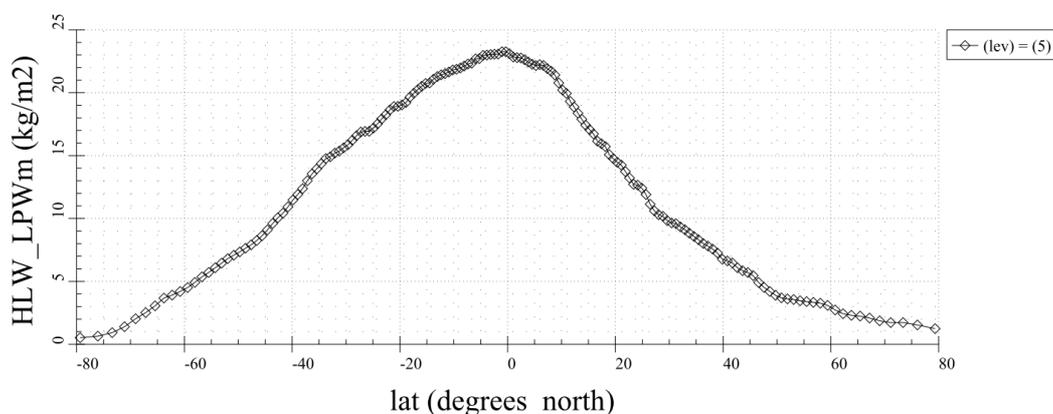
The result is shown in Figure 3.

Figure 3: Meridional profile of zonally averaged atmospheric water vapour between surface and 850 hPa for April, 1st 2010 (based on CM SAF ATOVS product). Data of the plot were generated with the CDO call as shown in the text and visualized with *ncview*.

OPTIONS FOR DATA CONVERSION IN CM SAF'S WEB USER INTERFACE

Users typically order CM SAF products via the CM SAF Web User Interface (accessible via http://www.cmsaf.eu or directly at http://wui.cmsaf.eu). After submitting the order, the requested products are retrieved from the data base and are then typically provided via an ftp-server. Until recently, files were provided in the original format as stored in the data base (i.e. HDF5 for the first-guess products). For users that do not want to convert the data locally on their own platform, a few simple options for conversion based on CDO are now also integrated into this ordering process. Users can convert the data to netCDF, remap them to regular latitude/longitude grids and cut out regions from the products. Before providing the data on the ftp-server, the files are converted to the requested format.

LICENSE, AVAILABILITY AND COMPONENTS

CDO was developed at the Max-Planck-Institute for Meteorology. It is open source and released under the terms of the GNU General Public License v2 (GPL). The source code is available at https://code.zmaw.de/projects/cdo.

CDO uses the GNU configure and build system for compilation. The only requirement is a working ANSI C99 compiler.

CDO should easily compile on every Posix compatible operating system like IBM's AIX, HP-UX, Sun's Solaris as well as on most Linux distributions, BSD variants and cygwin. Thus it is possible to use CDO similarly on general purpose PCs and unix-based high performance clusters.
Precompiled packages are available for some LINUX distributions (e.g. debian, ubuntu; see comments on https://code.zmaw.de/projects/cdo for details).

To take full advantage of CDO features the following additional libraries should be installed:
- Unidata netCDF library (http://www.unidata.ucar.edu/packages/netcdf) version 3 or higher. This is needed to process netCDF files with CDO.
- HDF5 szip library (http://www.hdfgroup.org/doc_resource/SZIP) version 2.1 or higher. This is needed to process szip compressed GRIB files with CDO.
- HDF5 library (http://www.hdfgroup.org/HDF5) version 1.6 or higher. This is needed to import CM-SAF HDF5 files with the CDO operator import_cmsaf.
- PROJ.4 library (http://trac.osgeo.org/proj) version 4.6 or higher. This is needed to convert Sinusoidal and Lambert Azimuthal Equal Area coordinates of CM SAF data to geographic coordinates, for e.g. remapping.

A precompiled 32-bit Windows version of CDO is also available at the CDO homepage (including the libraries). With that version CDO can be used in the Windows Command Prompt and in Windows batch files.

CONCLUSIONS

EUMETSAT's Satellite Application Facility on Climate Monitoring provides satellite-based monitoring products of various parameters of the atmospheric energy and water cycle. Not all users are familiar to the data formats used in the satellite community (HDF5) and were therefore frequently asking for conversion tools. In order to make the use of such tools as easy as possible for users, CM SAF decided to incorporate their data formats into a tool that is already well-known to a large number of users. A new import filter for CM SAF data has been implemented into the software package 'climate data operators'. This solution now offers the possibility to convert and reproject CM SAF data to the

formats requested by users. Due to wide acceptance of this tool (e.g. by climate modellers and in numerical weather forecast), people can use CM SAF products without any additional training and will also profit from the large number of operators that are already available in that package. As the package is provided as freely available open source software, CM SAF and the users will profit from further developments. Conversion of CM SAF products to CF compliant netCDF-files also allows very easy handling and visualisation of the products with other software tools. Some examples are shown in this document. Based on this experience CM SAF decided to provide all upcoming new datasets (long reprocessed time series) as CF-compliant netCDF files.

REFERENCES

Schulz, J., et al., (2009) Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF). Atmos. Chem. Phys., 9, pp 1687-1709

Schulzweida, U., Kornblueh, L., Quast, R., (2010) CDO user's guide – Climate Data Operators, Version 1.4.6. pp 1-173

1. The examples shown in this section were originally used to convert cloud data from the hurricane season in 2006 (Hurricane Helene; as also examined in the framework of the African Monsoon Multidisciplinary Analysis (AMMA) project). During that period some metadata were set incorrectly in the intermediate (instantaneous) products. Such problems can also be corrected with the help of some CDO operators (in that specific case: missing value should be set to -5, and values are divided by 25 (in order to correct the incorrectly set 'gain factor' of the original data, which is 5 but should be 0.2).
   The specific examples would then look like this:
   cdo –f nc –b f32 -divc,25 -setmissval,-5 –selname,"clwp"
       –setgrid,cmsaf_lat_lon.hdf –import_cmsaf ifile.h5 ofile.nc
   In case of such operations, it is important to adapt the data type in the netCDF-output, so that the range of the output data still fits into the range that can be covered by the selected data type.
   The data from the lifetime of Hurricane Helene is provided as *"CM SAF cloud product demonstration dataset"*.

2. The following command and input data were used for generating the plot shown in Figure 1 (see also Footnote 1 for details):
   cdo –f nc –b f32 –remapbil,grid.txt –setgrid,CM_SAF_03636_03636_LAT_LON.hdf
       –divc,25 -setmissval,-5 –selname,clwp
       –import_cmsaf CPPin20060918104530000500018299MD.hdf remapped_clwp.nc

3. In this example, the European domain was defined as follows (including the Mediterranean):
   gridtype = lonlat
   xsize = 600
   ysize = 400
   xfirst = -20
   xinc = 0.1
   yfirst = 30
   yinc = 0.1

4. The following GrADS commands were used to generate the plot in Figure 2 (right plot):
   sdfopen cl_201008.nc   ** open the netCDF data file*
   set grads off
   set gxout grfill   ** switch plotting mode to "filled grid cells"*
   d data1   ** plot the data*
   draw title number of cloudy days 2010/08   ** draw the title*
   cbarn   ** plot the legend*
   printim cl_201008.png png x1024 y800 white   ** create a graphics file (png)*