

ReKliEs-De

Abschlussworkshop

Datenbearbeitung

Datenbearbeitung

- cdo Climate Data Operator
 - MPI-M Programmbibliothek
- nco netCDF Operators
 - Unidata Programmbibliothek
- Python netCDF-4 Library

Datenbearbeitung

Das Klimadaten Format ist für viele neuere Projekte: netCDF-CF

- NetCDF: **N**etwork **C**ommon **D**ata Format
 - <http://www.unidata.ucar.edu/software/netcdf/>
- CF: **C**limate and **F**orecast
 - <http://cfconventions.org/>

Datenbearbeitung Metadaten

- Zusatzinformationen zu den Daten
- Werden bei der Erzeugung hinzugefügt
- Müssen bestimmte Standards erfüllen
 - (Project Archiv Design / Daten Management Plan)

Datenbearbeitung Metadatenbereiche

Verzeichnisstruktur:

- Namen im Verzeichnisbaum
- Name der Datei

Datensuche

Data Reference Syntax (DRS)

Dateiattribute:

- Variablen Attribute
- Globale Attribute

Dateninhalt

NetCDF Attribute

Datenbearbeitung Metadatenbereiche

- **Projektinformationen**
 - Institut, Projekt
 - Personen
 - weitere Projektbeschreibungen
- **Experimentinformationen**
 - Name, Realisierung, Initialisierung, Physik
 - Kalender, Region, Gitter
- **Variableninformationen**
 - Name (long_name / standard_name)
 - Einheiten

Datenbearbeitung Konventionen

- Liste der im Projekt CMIP5 geforderten Modellergebnisse
http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.xls
- netCDF-CF Climate and Forecast (CF) Konvention
<http://cf-pcmdi.llnl.gov/>
- Data Reference Syntax (DRS)
Struktur von Dateiname und Verzeichnisbaum
- MIP Tabellen (CMIP5)
https://github.com/PCMDI/cmip5-cmor-tables/tree/master/Tables_csv
- MIP Tabellen (CORDEX [ReKliEs-De])
<https://github.com/PCMDI/cordex-cmor-tables/tree/master/Tables>
- CMOR (Climate Model Output Rewriter)
<https://pcmdi.github.io/cmor-site/>

Datenbearbeitung Besonderheiten

- Einheiten für dimensionslose Variablen: sind '1' und NICHT 'none'
- Flüsse haben keine Zeitangabe in den Einheiten
- Gibt es keinen CF standard_name für die Variable → keine Angabe
- Attribut 'Conventions' beginnt mit 'C'
- climatology Attribute ist eine Besonderheit
 - Beispiel: Variable sfcWindmax (monmean)
 - cell_methods="time: maximum within days time: mean over months"

<http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html#climatological-statistics>

<http://cfconventions.org/faq.html>

Datenbearbeitung DRS (Data Referenz Syntax)

- Jedes Projekt muss seine eigene DRS definieren
- Verzeichnisse und Dateinamen sind damit festgelegt
- Ausnahmen sind nicht zulässig
- Namen können gegen ein festes Vokabular (CV) getestet werden
 - CV: 'controlled vocabulary'
- Das CV kann ergänzt werden (neue Experimente, Institute,...)



Datenbearbeitung DRS Beispiel

Reklies	Syntax (wie CORDEX Archiv Design)
DRS (Pfad)	<activity>/<product>/<institute>/<domain/<driving_model>/<driving_experiment>>/<driving_ensemble member>/<rcm_name>/<frequency>/<variable name/<version>/
Beispiel	reklies/output/EUR-11/CLMcom/CCCma-CanESM2/rcp85/r1i1p1/CLMcom-CCLM4-8-17/v1/day/pr/v20170227/
Dateiname	pr_EUR-11_CCCma-CanESM2_rcp85_r1i1p1_CLMcom-CCLM4-8-17_v1_day_20060101-20101231.nc
Doku	http://cordex.dmi.dk/joomla/images/CORDEX/cordex_archive_specifications.pdf

Datenbearbeitung (NetCDF Dateiattribute)

NetCDF Attribute sind Metadaten in einer netcdf (.nc) Datei.

Jede Datei hat einen Header mit folgenden Bereichen:

- dimensions (lat, lon, levels, time)
- variables
 - <variable> (time,lat,long)
 - lat (lat)
 - lon (lon)
 - time (time)
- global attributes

Datenbearbeitung NetCDF Header

```
ncdump -h pr_EUR-11_CCCma-CanESM2_rcp85_r1i1p1_CLMcom-CCLM4-8-17_v1_day_20060101-20101231.nc {
```

dimensions:

```
  rlon = 424 ;
```

```
  rlat = 412 ;
```

```
  time = UNLIMITED ; // (1825 currently)
```

```
  bnds = 2 ;
```

```
  vertices = 4 ;
```

Datenbearbeitung ncdump (1)

variables:

```
float lat(lat) ;
```

```
    lat:standard_name = "latitude" ;
```

```
    lat:long_name = "Latitude, south is negative" ;
```

```
    lat:units = "degrees_north" ;
```

```
float lon(lon) ;
```

```
    lon:standard_name = "longitude" ;
```

```
    lon:long_name = "Longitude, west is negative" ;
```

```
    lon:units = "degrees_east" ;
```

Datenbearbeitung ncdump (2)

```
float rlds(time, lat, lon) ;  
    rlds:units = "J m-2" ;  
    rlds:long_name = "Accumulated surface downwelling longwave radiation" ;  
int time(time) ;  
    time:standard_name = "time" ;  
    time:long_name = "Time" ;  
    time:units = "hours since 1970-01-01 00:00:00 +0000" ;  
    time:calendar = "gregorian" ;
```

Datenbearbeitung ncdump (3)

Globale Attribute (CORDEX EUR-11)

// global attributes:

```
:institution = "Climate Limited-area Modelling Community (CLM-Community)" ;  
:institute_id = "CLMcom" ;  
:experiment_id = "historical" ;  
:source = "CLMcom-CCLM4-8-17" ;  
:model_id = "CLMcom-CCLM4-8-17" ;  
:contact = "cordex-cclm@dkrz.de" ;  
:comment = "CORDEX Europe RCM CCLM 0.11 deg EUR-11" ;  
:references = "http://www.clm-community.eu/" ;  
:tracking_id = "90dcb0c9-3797-475e-9e6b-91da18194595" ;  
:title = "CLMcom-CCLM4-8-17 model output prepared for CORDEX historical" ;
```

Datenbearbeitung ncdump (4)

Globale Attribute (CORDEX EUR-11)

```
:CORDEX_domain = "EUR-11" ;  
:driving_experiment = "MOHC-HadGEM2-ES, historical, r1i1p1" ;  
:driving_model_id = "MOHC-HadGEM2-ES" ;  
:driving_model_ensemble_member = "r1i1p1" ;  
:driving_experiment_name = "historical" ;  
:rcm_version_id = "v1" ;  
:product = "output" ;  
:experiment = "historical" ;  
:frequency = "day" ;  
:creation_date = "2015-02-18T11:22:06Z" ;  
:Conventions = "CF-1.4" ;  
}
```


Globale Attribute

Attributname	DRS Name	Beispiel
project_id	activity	CORDEX
product	product	output
CORDEX_domain	Domain	EUR-11
institute_id	Institution	CLMcom
driving_model_id	GCMModelName	MOHC-HadGEM2-ES
experiment_id	CMIP5ExperimentName	historical , rcp45
driving_model_ensemble_member	CMIP5EnsembleMember	r1i1p1
model_id	RCMModelName	CLMcom-CCLM4-8-17
rcm_version_id	RCMVersionID	v1
frequency	Frequency	day
<in variable section>	VariableName	float rlds(time, lat, lon)

Variablen Attribute 1

Variable	Attributename	Wert
float tas		(time, rlat, rlon)
	coordinates	height lat lon
	missing_value	1.e+20f
	_FillValue	1.e+20f
	standard_name	air_temperature
	long_name	Near-Surface Air Temperature
	units	K
	frequency	day
	cell_methods	time: mean
	grid_mapping	rotated_latitude_longitude
	comment	daily-mean near-surface (usually, 2 meter) air temperature.

Variablen Attribute 2

Variable	Attributname	Wert
double time	standard_name	time
	long_name	Time
	units	Days since 1970-01-01 00:00:00
	calendar	gregorian
	bounds	(time_bnds)
double time_bnds	(time, bnds)	(see dimensions)
double lat	standard_name	latitude
	units	degrees north
double lon	Standard_name	longitude
	units	degrees east

Variablen Attribute 3

Variable	Attributname	Wert
double height		2 (tas, hus etc.) 10 (sfcWind etc.)
	units	m
	axis	Z
	positive	up
	long_name	height
	standard_name	height

Ungenauigkeiten beim CF-Standard

- Variableneinheiten ohne Dimension sind '1' und nicht 'none'
- Variablen die Flüsse repräsentieren haben keine Zeiteinheiten
- Variable die keinen CF Standardnamen haben sollen leer sein
- Eintrag 'Conventions' beginnt mit 'C'
- Attribute climatologies sind Ausnahmen z.B. sfcWindmax:

sfcWindmax:cell_methods="time: maximum within days time: mean over months"

<http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html#climatological-statistics>

<http://cfconventions.org/faq.html>

Datenbearbeitung (nc0 Beispiele I)

VAR1=2m_temp

newVAR=tas

1. Variablennamen ändern

```
ncrename -O -v ${VAR1},${newVAR} <file>
```

2. Variable mit Faktor multiplizieren

```
ncap2 -O -s '${VAR1}=${VAR1}*${FACT}' $fil
```

3. Variablen lat + lon in zwei Dateien extrahieren

```
ncap2 -s 'lat=double(lat); lon=double(lon)' latlon.nc dlatlon.nc
```

4. Umwandeln einer Variablen von Typ ,double' in Typ ,float'

```
ncap2 -s 'MyVar=float(MyVar)' input_file output_file
```

5. Multiplikation der Werte einer Variablen mit einem Faktor:

```
ncap2 -O -s 'MyVar=MyVar*0.001' input_file output_file
```

6. Druck von Pa nach hPa umrechnen :

langsam: **ncap2** -s 'pressure=pressure/100.' wa12clmN_eraint-ctrl_cclmprs_2011-12.nc prs_1.nc

schnell: **ncatted** -O -a units,pressure,o,c,'hPa' prs_1.nc

Datenbearbeitung (nco Beispiele II)

7. Variable umbenennen:

ncrename -h -O -v Altername,Neuename Datei.nc

9. Lat / Lon Variablen aus einer Datei in eine Zwischendatei speichern:

ncks -v lat,lon pr_mon_19790101-19800101.nc latlon.nc

ncap2 -s 'rlat=double(rlat); rlon=double(rlon)' latlon.nc dlatlon.nc

10. Einbau in eine vorhandene Datei:

ncks -A dlatlon.nc MeineDatei.nc

Datenbearbeitung (nco Beispiele III)

8.a Setzen eines Attributes (hier: tracking_id):

```
# uuid erstellen (unter Unix): uuid=`echo $(python -c 'import uuid; print uuid.uuid1()')`  
    ncatted -h -O -a tracking_id,global,o,c,"$uuid"Datei.nc
```

8.b Setzen mehrerer Attribute in einem Schritt:

```
# setzen (unter Unix) einiger Variablen  
# export VAR= Variablenname  
# export UNIT= Einheit  
# export SNAME=standard_name  
# export LNAME=long_name  
    ncatted -h -O -a _FillValue,${VAR},o,f,1.e20 \  
    -a missing_value,${VAR},o,f,1.e20 \  
    -a units,${VAR},o,c,$UNIT \  
    -a standard_name,${VAR},o,c,$SNAME \  
    -a tracking_id,global,o,c,"$uuid" \  
    -a long_name,${VAR},o,c,"$LNAME" Datei.nc
```


Datenbearbeitung (cdo Beispiele I)

Umgebungsvariablen (unter Linux) für CDO:

```
export IGNORE_ATT_COORDINATES=1
```

```
export CDO_HISTORY_INFO=0
```

1. Selektion der Monate 3...11 aus einer Datei:

```
"cdo -f nc4c -seasmean -selmonth,3/11 input_file output_file"
```

2. Selektion des Monats 12 aus einer Datei:

```
"cdo -f nc4c -seasmean -selmonth,12 input_file output_file"
```

3. Selektion von Zeitschritten aus einer Datei:

```
"cdo -f nc4c -s tinselmean,3,2 input_file output_file"
```

4. Zusammenbinden zweier Dateien (z.B. Winde u + v Variablen):

```
"cdo merge Input1_file Input2_file Output_file_u_v"
```

5. Entrotieren von Variablen u + v (Winde):

```
"cdo rotuvb,u,v Output_file_u_v Output_file_entrotiert"
```

6. Selektion einer Variablen aus einer Datei:

```
Für u: "cdo selvar,u Output_file_entrotiert Output_file_entrotiert_Var_u"
```

```
Für v: "cdo selvar,v Output_file_entrotiert Output_file_entrotiert_Var_v"
```

Datenbearbeitung (cdo Beispiele II)

7. Selektion eines Jahres aus einer Datei:

```
"cdo -f nc4c -s selyear,1981 input_file output_file"
```

8. Verschiebung der Zeit in einer Datei um 30 Tage:

```
"cdo -f nc4c -s -shifttime,30days input_file output_file"
```

9. Setzen einer neuen Zeitmarke:

```
"cdo -f nc4c -s setreftime,'days since 1950-01-01 00:00:00' input_file output_file"
```

10. Operatoren sind kombinierbar: Abarbeitung von rechts nach links!

```
"cdo -f nc4c -s selyear,1997 -shifttime,27days -setreftime, 1950-01-01, 00:00:00 input_file output_file"
```

11. Selektion eines Levels 3 (z.B. aus einer Datei mit verschiedenen Schichten einer Bodenvariablen):

```
"cdo -f nc4c sellevidx,3 input_file output_file "
```

12. Multiplikation der Variablen mit einem Faktor:

```
"cdo mulc,0,02745 input_file output_file"
```

13. Selektion einer bestimmten Stunde (oder mehrerer Stunden) aus einer Datei:

```
"cdo -f %s -s selhour,2,3,4 input_file output_file"
```

Datenbearbeitung (cdo Beispiele III)

14. Erstellung von Monatsdaten aus täglichen Daten:

```
"cdo -f nc4c -s -monmean -daymean .mulc,27,3%s input_file output_file"
```

15. Die Zeitachse neu definieren (incl. time_bnds):

```
"cdo --cmor setreftime,1980-11-10,10:00:00,seconds infile outfile"
```

Datenbearbeitung I

cdo: [Installation-und-Unterstützte-Plattformen](#)

nco: [Installation und Programme](#)

python netCDF4 z.B.: `pip install netcdf4`

cygwin: [cygwin](#) | [cygwin Paketliste \(installierbare Programme\)](#)

spack (aus Projekt ESIWACE): [Software stack deployment for Earth System Modelling using Spack](#)

docker: [Installation \(alle Systeme\)](#) | [Windows Installation](#)

Oracle VirtualBox (VM Interpreter): [Benutzerhandbuch inkl. Installation \(nur Englisch\)](#)

Datenbearbeitung II

cdo: <https://code.mpimet.mpg.de/projects/cdo/wiki#Installation-and-Supported-Platforms>

nco: <http://nco.sourceforge.net/#Executables>

python netCDF4 Library, Installation mit (Linux): [pip install netcdf4](#)

Cygwin (Windows):

https://www.ibm.com/support/knowledgecenter/de/SS2GNX_7.2.1/com.ibm.tivoli.tpm.ins.doc/install_72ga/tins_cygwin_72ga.html

cygwin Paketliste (installierbare Programme): | https://cygwin.com/packages/package_list.html

spack (Softwarestack aus dem Projekt ESIWACE): https://www.esiwace.eu/results/conference-contributions/esiwace-usability-prace-days-2017/at_download/file

Docker Installation: <https://docs.docker.com/engine/installation/>

Docker Installation (Windows): https://docs.docker.com/toolbox/toolbox_install_windows/

Oracle VirtualBox (Benutzerhandbuch): <https://www.virtualbox.org/manual/UserManual.html>