CMOR post-processing of NorESM output

–

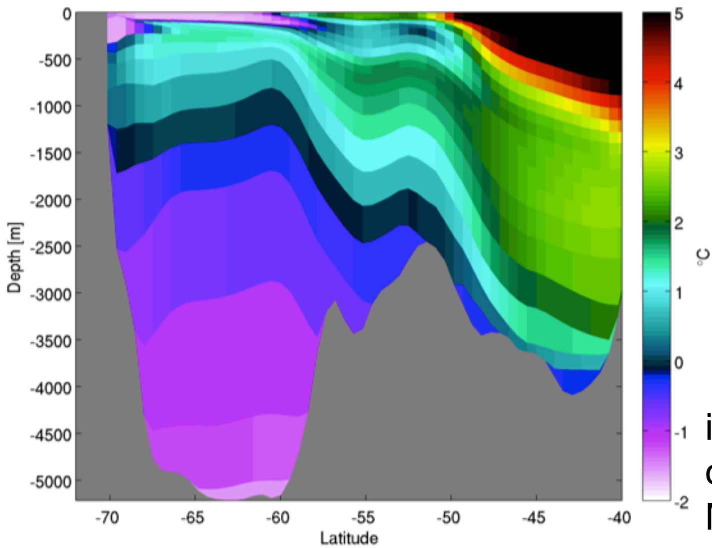
CMIP5 & CMIP6

Ingo Bethke (ingo.bethke@bjerknes.uib.no)

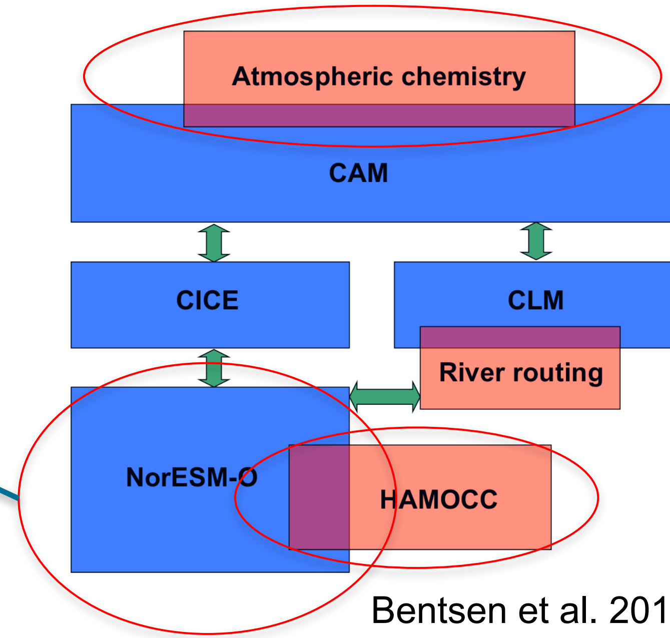
Norwegian Earth System Model (NorESM)

Flavor of NCARs Community Earth System Model version (CESM)

Temperature section through Southern Ocean



isopycnic coordinate ocean model based on MICOM (Bleck 2002)



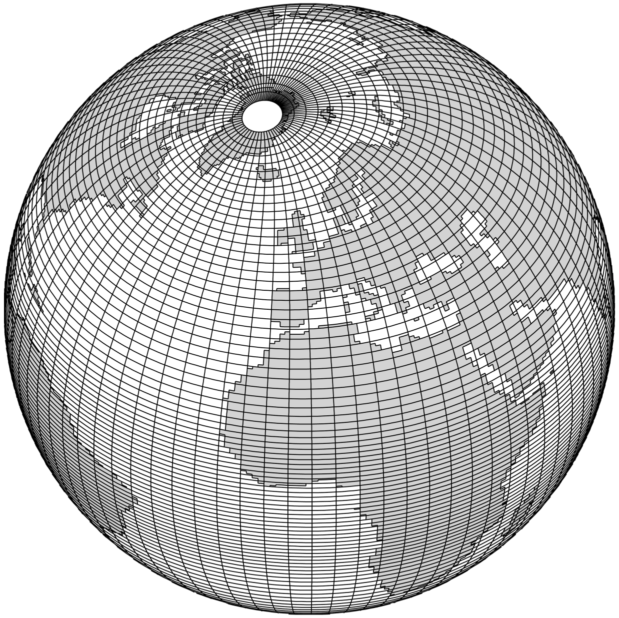
Bentsen et al. 2012

Challenges for post-processing:

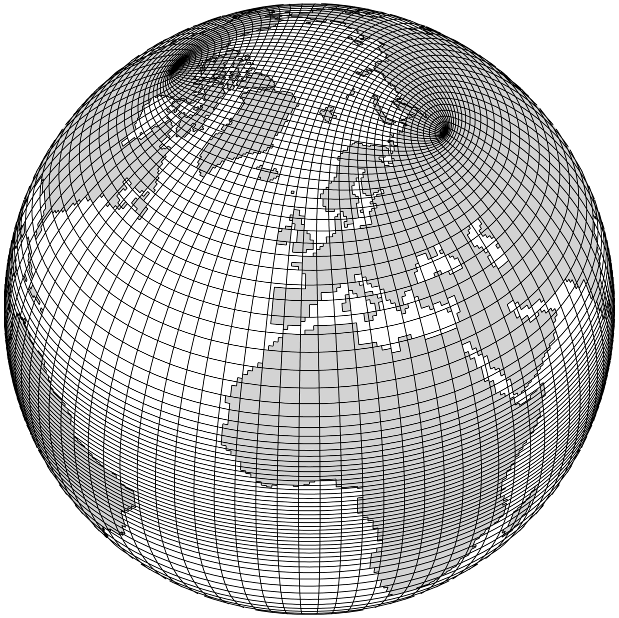
- time-varying vertical coordinate in ocean component
- exotic output fields from aerosol-chemistry component
- exotic output fields from ocean biogeochemistry component

Norwegian Earth System Model (NorESM)

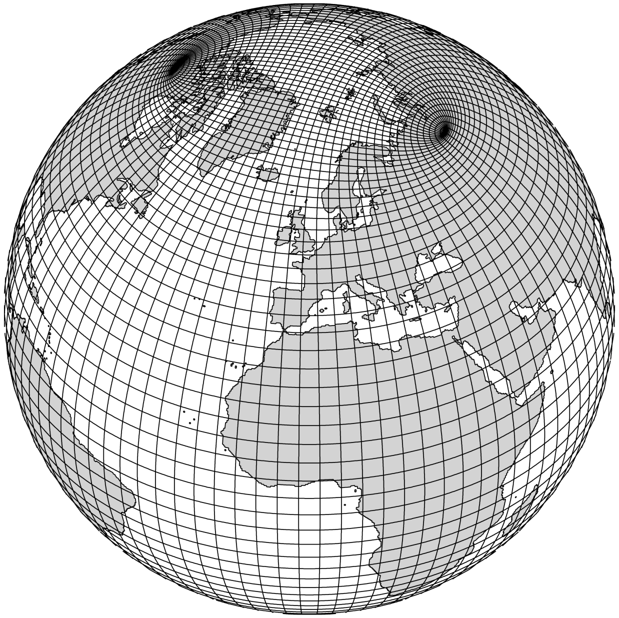
Horizontal ocean/sea-ice grids



1.125° bipolar grid (every 4th grid line shown):
- 320 × 384 grid cells.
- Used for the NorESM CMIP5 experiments.
- Enhanced meridional resolution near the equator (f_e = 1/4).



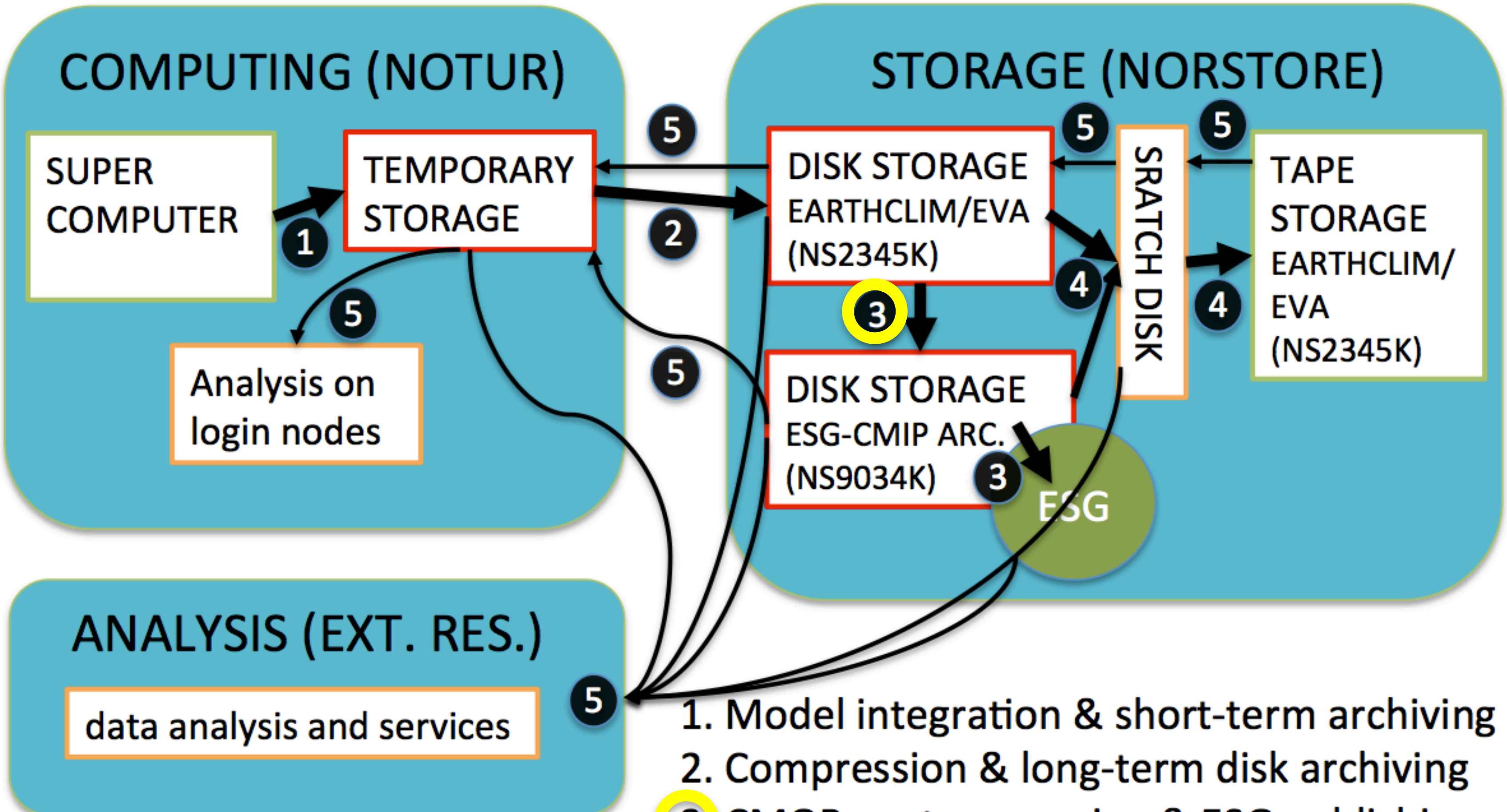
1° tripolar grid (every 4th grid line shown):
- 360 × 384 grid cells.
- Used for the NorESM-O CORE2 experiments.
- Enhanced meridional resolution near the equator (f_e = 1/4).



0.25° tripolar grid (every 16th grid line shown):
- 1440 × 1152 grid cells.
- Isotropic grid near equator.
- Target resolution for NorESM CMIP6 experiments.

Courtesy of Mats Bentsen

Data-flow in a typical NorESM production



1. Model integration & short-term archiving
2. Compression & long-term disk archiving
3. CMOR post-processing & ESG publishing
4. Tape archival (with staging=tar-bundling)
5. Data analysis

Why CMOR-izing?

What it does:

- post-processing step in which raw climate model output is rewritten in standardized format (“CMIP” standard)
- reorganization of output file structure (single record, multiple variables -> multiple records, single variables)
- basic annotation and quality control
- computation of derived products (linear combination, spatial integration)

Advantages:

- facilitates data-sharing of multi-model intercomparison studies through ESG
- facilitates multi-model output analysis and employment of community evaluation tools like ESMValTool across different projects (common output standard not only for models but also for all projects!)
- makes output more accessible to climate science community (over 500 publications have used NorESM1 output prepared for CMIP5)

Disadvantages:

- doubling of storage volume if raw output is kept
- cmor-izing takes time, system resources and human resources

Why CMOR-izing?

What it does:

- post-processing step in which raw climate model output is rewritten in standardized format (“CMIP” standard)
- reorganization of output file structure (single record, multiple variables -> multiple records, single variables)
- basic annotation and quality control
- computation of derived products (linear combination, spatial integration)

Advantages:

- facilitates data-sharing of multi-model intercomparison studies through ESG
- facilitates multi-model output analysis and employment of community evaluation tools like ESMValTool across different projects (common output standard not only for models but also for all projects!)
- makes output more accessible to climate science community (over 500 publications have used NorESM1 output prepared for CMIP5)

Disadvantages:

- doubling of storage volume if raw output is kept
- cmor-izing takes time, system resources and human resources

Why CMOR-izing?

What it does:

- post-processing step in which raw climate model output is rewritten in standardized format (“CMIP” standard)
- reorganization of output file structure (single record, multiple variables -> multiple records, single variables)
- basic annotation and quality control
- computation of derived products (linear combination, spatial integration)

Advantages:

- facilitates data-sharing of multi-model intercomparison studies through ESG
- facilitates multi-model output analysis and employment of community evaluation tools like ESMValTool across different projects (common output standard not only for models but also for all projects!)
- makes output more accessible to climate science community (over 500 publications have used NorESM1 output prepared for CMIP5)

Disadvantages:

- doubling of storage volume if raw output is kept
- cmor-izing takes time, system resources and human resources

Past CMOR-ization of NorESM output

Project	Extended title	Model	Volume	When	NorESM contact
CMIP5	fifth Coupled Model Intercomparison Project	NorESM1-M NorESM1-ME	~60 TB	2011	Ingo Bethke (UNI)
GeoMIP	Geoengineering Model Intercomparison Project	NorESM1-M	~10 TB	2011	Helene Muri (UiO), Kari Altersjær (UiO)
PlioMIP	Pliocene Model Intercomparison Project	NorESM-L	~1 TB	2012	Zhongshi Zhang (UNI)

- people with NorESM cmor-izing experience: Ingo Bethke (UNI), Helene Muri (UIO), Kari Altersjær (UiO), Alf Grini (MET), Zhongshi Zhang (UNI), Jerry Tjiputra (UNI), Anna Lewinschal (MISU), Bjørn Samset (CICERO)
- NorESM participation in intercomparison projects that did not require cmorization: CMIP2, CMIP3, CORE2, PEGASOS...

Upcoming CMOR-ization of NorESM output

Project	Extended title	Volume	When	NorESM contact
		NON-CMIP		
GeoMIP	Geoengineering MIP	?? TB	2016-	Helene Muri (UiO)
CDR-MIP	Carbon Dioxide Removal MIP	?? TB	2016-	Helene Muri (UiO)
EXPECT	Exploring the potential and side effects of climate engineering	?? TB	2016	Alf Grini (MET) Helene Muri (UiO)
		CMIP6		
DECK	Diagnosis, Evaluation, and Charact. of Klima	200 TB	2016	Seland (MET), Bethke (UNI)
AerChemMIP	Aerosols Chemistry MIP	32 TB	2017	Dirk Olivie
C4MIP	Coupled Climate Carbon Cycle MIP	5 TB	2017	Christoph Heinze (UiB)
CFMIP	Cloud Feedback MIP	5 TB	2017	Alf Grini (MET)
DAMIP	Detection and Attribution MIP	38 TB	2016/7	Seland (MET), Bethke (UNI)
DCPP	Decadal Climate Prediction Project	100 TB	2017	Noel Keenlyside (UiB)
OMIP	Ocean MIP	21 TB	2016	Mats Bentsen (UNI)
PDRMIP	Precipitation Driver Response MIP	37 TB	2017	Alf Kirkevåg (MET)
PMIP	Paleoclimate MIP	30 TB	2017	Zhongshi Zhang (UNI)
RFMIP	Radiative Forcing MIP	15 TB	2017	Alf Kirkevåg (MET)
ScenarioMIP	Scenario MIP	16 TB	2016/7	Iversen (MET), Heinze (UiB)
VolMIP	MIP on the Response to Volcanic Forcing	5 TB	2017	Odd Helge Otterå
GeoMIP CMIP6	Geoengineering MIP	21 TB	2017	Muri (UiO), Kristjansson (UiO)
HighResMIP	High Resolution MIP	??? TB	2017/8	Alok Gupta (UNI)

*NorESM2 versions: MH, HH, MM, LM, LME, LMEC

Upcoming CMOR-ization of NorESM output

Project	Extended title	Volume	When	NorESM contact
	NON-CMIP			
GeoMIP	Geoengineering MIP	?? TB	2016-	Helene Muri (UiO)
CDR-MIP	Carbon Dioxide Removal MIP	?? TB	2016-	Helene Muri (UiO)
EXPECT	Exploring the potential and side effects of climate engineering	?? TB	2016	Alf Grini (MET) Helene Muri (UiO)
	CMIP6			
DECK				Bethke (UNI)
AerChemMIP				
C4MIP				ze (UiB)
CFMIP				
DAMIP				Bethke (UNI)
DCPP				e (UiB)
OMIP				UNI)
PDRMIP				ET)
PMIP	Paleoclimate MIP	30 TB	2017	Zhongshi Zhang (UNI)
RFMIP	Radiative Forcing MIP	15 TB	2017	Alf Kirkevåg (MET)
ScenarioMIP	Scenario MIP	16 TB	2016/7	Iversen (MET), Heinze (UiB)
VolMIP	MIP on the Response to Volcanic Forcing	5 TB	2017	Odd Helge Otterå
GeoMIP CMIP6	Geoengineering MIP	21 TB	2017	Muri (UiO), Kristjansson (UiO)
HighResMIP	High Resolution MIP	??? TB	2017/8	Alok Gupta (UNI)

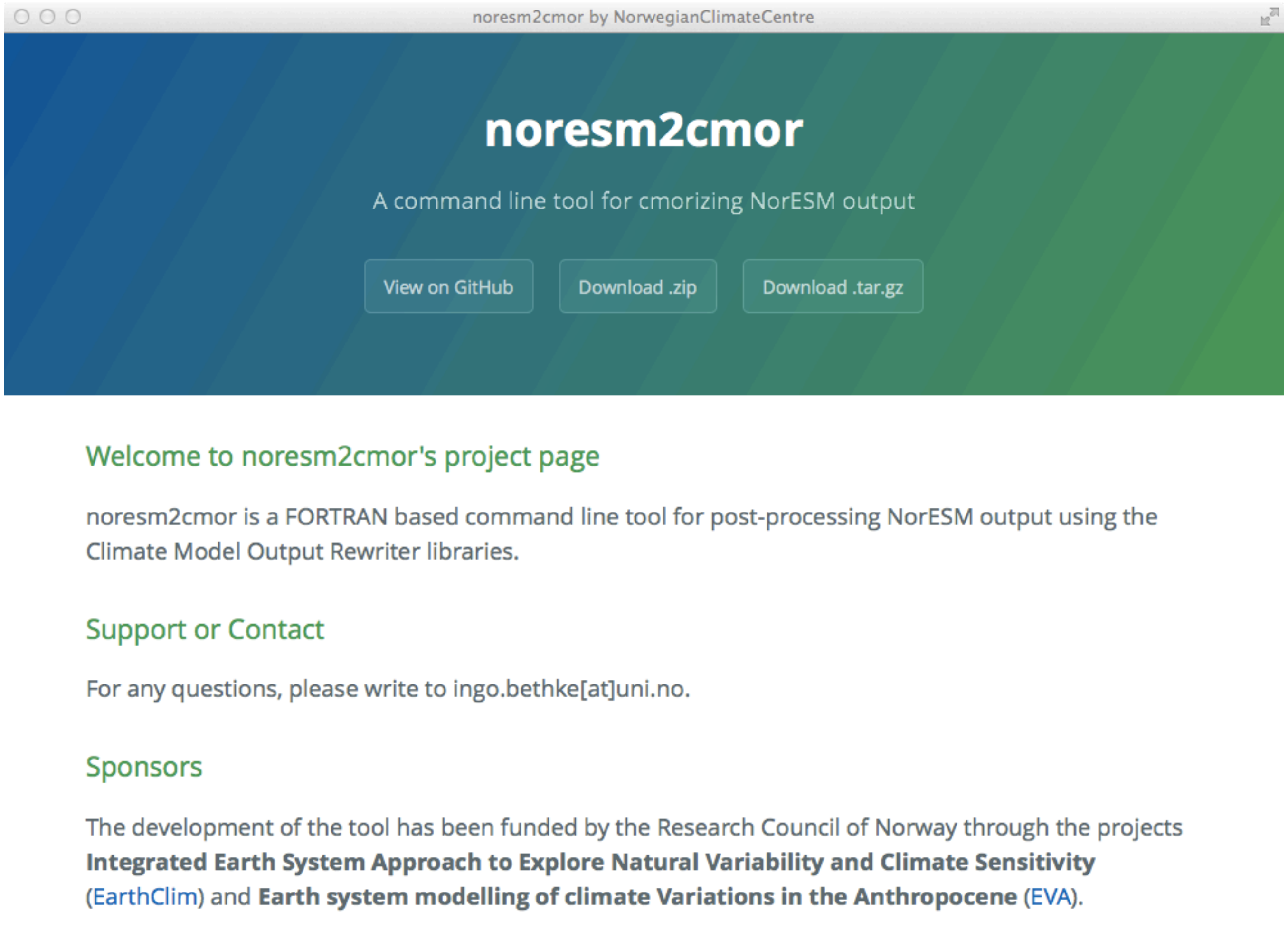
Post-processing challenges

- many MIPS=many experiments to process
- new output diagnostics
- more model configurations (CMIP5=3, CMIP6=6)
- big data volume (CMIP6 ~ 10 X CMIP5)

➢ **high demands on post-processing tools**

*NorESM2 versions: MH, HH, MM, LM, LME, LMEC

The CMOR-ization tool for NorESM output



noresm2cmor by NorwegianClimateCentre

noresm2cmor

A command line tool for cmorizing NorESM output

View on GitHub | Download .zip | Download .tar.gz

Welcome to noresm2cmor's project page

noresm2cmor is a FORTRAN based command line tool for post-processing NorESM output using the Climate Model Output Rewriter libraries.

Support or Contact

For any questions, please write to ingo.bethke[at]uni.no.

Sponsors

The development of the tool has been funded by the Research Council of Norway through the projects **Integrated Earth System Approach to Explore Natural Variability and Climate Sensitivity** (EarthClim) and **Earth system modelling of climate Variations in the Anthropocene** (EVA).

https://github.com/NorwegianClimateCentre/noresm2cmor

The CMOR-ization tool for NorESM output

Basic features

- FORTRAN based command line tool
- uses namelist input file: info about system, model, experiment and output
- performs various tasks:
  - renaming
  - unit conversion
  - vector rotation
  - vertical interpolation (e.g., from hybrid to pressure level)
  - global integration (e.g., co2mass)
  - computation of derived variables (e.g., barotropic streamfunction)
  - annotation

Advanced features

- advanced file scanning and handling of time information with minimal requirements on input format -> allows post-processing in a single step
- MPI-parallelization along variable dimension

Content of noresm2cmor/

bin		
	noresm2cmor	executable
build		
	makefile	makefile for noresm2cmor
data		
	cmorout	sym link to folder where output is stored
	griddata	sym link to folder with grid information
	sampledata	sym link to folder where sample data is installed
namelists		
	noresm2cmor_NorESM_GENERIC_template.nml	
	noresm2cmor_CMIP5_NorESM1-M_historical_r1i1p1.nml	
	…	
scripts		
	cmorize_generic.sh	wrapper script for cmor-ization of generic experiments
	cmorize_sampledata.sh	script for cmor-ization of sample output
	install_cmorout.sh	script for installing setting up output folder
	install_griddata.sh	script for downloading/installing NorESM grid data
	install_sampledata.sh	script for downloading/installing NorESM sample data
source		noresm2cmor source files
tables		CMOR table files
README		

Installation (https://github.com/NorwegianClimateCentre/noresm2cmor/blob/master/README)

Download code

```
git clone https://github.com/NorwegianClimateCentre/noresm2cmor
```

Build code

```
cd noresm2cmor/build/
make -f Makefile_cmor2.nird_intel Makefile     # need to customize
make
```

creates executable in noresm2cmor/bin (requires cmor2 and NetCDF library)

Install grid data (1.3 GB) and sample model output (15 GB)

```
noresm2cmor/scripts/install_griddata.sh <local path for storing grid data>
noresm2cmor/scripts/install_sampledata.sh <local path for storing sample output>
```

Configure path where CMOR-ized output is to be stored

```
noresm2cmor/scripts/install_cmorout.sh <local path where cmor-output to be stored>
```

Test

```
noresm2cmor/scripts/cmorize_sampledata.sh
```

Usage

Direct use

Syntax

./noresm2cmor <path to namelist file>

Example

./noresm2cmor CMIP5_NorESM1-M_historical_r1i1p1.nml

Wrapper

Syntax

./cmorize_generic.sh <absolute path to NorESM output folder> <start year> <last year>

Example

./cmorize_generic.sh /work/ingo/sampledata/N20TRAERCN_f19_g16_01 2000 2000

Namelist file configuration

System namelist with path settings etc

```
&system
 ibasedir      = '../data/sampledata',
 obasedir      = '../data/cmorout/N20TRAERCN_f19_g16_01',
 griddata      = '../data/griddata',
 tabledir      = '../tables'
 createsubdirs = .false.,
 verbose       = .true.,
/
```

Namelist file configuration

Model namelist with model specific information

```
&model
 model_id      = 'NorESM1-M',
 source        = 'NorESM1-M 2011',
                 ' atmosphere: CAM-Oslo (CAM4-Oslo-noresm-ver1_cmip5-r112, f19L26);',
                 ' ocean: MICOM (MICOM-noresm-ver1_cmip5-r112, gx1v6L53);',
                 ' sea ice: CICE (CICE4-noresm-ver1_cmip5-r112);',
                 ' land: CLM (CLM4-noresm-ver1_cmip5-r112)',
 institution   = 'Norwegian Climate Centre',
 institute_id  = 'NCC',
 references    = ' ',
 contact       = 'Please send any requests or bug reports to noresm-ncc@met.no.',
 tagoyr        = 'micom.hy',
 tagoyrbgc     = 'micom.hbgcy',
 tagomon       = 'micom.hm',
 tagomonbgc    = 'micom.hbgcm',
 tagoday       = 'micom.hd',
 tagimon       = 'cice.h',
 tagiday       = 'cice.h1',
 tagamon       = 'cam2.h0',
 tagaday       = 'cam2.h1',
 taga6hr       = 'cam2.h2',
 taga3hr       = 'cam2.h3',
 taga3hri      = 'cam2.h4',
 taglmon       = 'clm2.h0',
 taglday       = 'clm2.h1',
 tagl3hr       = 'clm2.h2',
 tagl3hri      = 'clm2.h3',
 rhotablesuff  = 'OnRho'
 atmgridfile   = 'grid_atm_1.9x2.5_gx1v6.nc',
 ocngridfile   = 'grid_gx1v6.nc',
 ocninitfile   = 'inicon_gx1v6.nc',
 ocnmertfile   = 'mertraoceans_gx1v6.dat',
 secindexfile  = 'secindex_gx1v6.dat'
/
```

Namelist file configuration

Experiment namelist with experiment specific information

```
&experiment
 casename      = 'N20TRAERCN_f19_g16_01',
 experiment_id = 'historical',
 parent_experiment_id  = 'pre-industrial control',
 parent_experiment_rip = 'r1i1p1',
 history       = ' ',
 comment       = ' ',
 forcing       = 'GHG, SA, Oz, Sl, Vl, BC, OC',
 realization   = 1,
 branch_time   = 255135.,
 year1         = 1850,
 yearn         = 2005,
 month1        = 1,
 monthn        = 12,
 exprefyear    = 1850
/
```

Namelist file configuration

Variable namelists (one namelist per CMIP table) containing information on

- how to map model variables on cmor variables
- linear combination of variables
- special operations (streamfunction computation, integration, etc)
- unit conversion

records per file

cmor variables

special operations

```
&table_omon
 domon          = .true.,
 romon          = 100000,
 tomon          = 'CMIP5_Omon',
 vomon          =
   'fresh_ai      ','fsitherm  ','cmFW day-1 -> kg m-2 s-1',
   'fsalt_ai      ','sfdsi     ','kg m-2 day-1 -> kg m-2 s-1',
   'fhocn_ai      ','hfsithermds','          ',
   'uflx,vflx     ','msftbarot ','strmf     ',
   'lip+sop+eva+rnf+rfi+fmltfz ','wfo','     ',
...
/
```

model variables

CMIP6: Impractical to hard-code 43 tables in source code. Should explore alternatives.

Design considerations

Choice of programming language FORTRAN vs C/python

- doesn't require special proficiency from scientists
- easy to debug
- high-degree of memory control
- good performance, particularly when using intel compiler
- easy to MPI parallelize

Intention to use python for CMIP7

Memory usage

- memory: ¼° X 60L X 8 byte precision ~ 500 GB per field
- accurate memory planning required when designing tools
- testing with high-resolution data important
- avoid creating unnecessary memory copies

"Optimal" file size/chunking strategy

- keep file sizes close but below 2 GB?
- same chunking for fields in same table?
- align time chunks with decades (e.g. 1950-1959, 1960-1969)?
- same number of records in each chunk?

Design considerations

Choice of programming language FORTRAN vs C/python

- doesn't require special proficiency from scientists
- easy to debug
- high-degree of memory control
- good performance, particularly when using intel compiler
- easy to MPI parallelize

Intention to use python for CMIP7

Memory usage

- memory: ¼º X 60L X 8 byte precision ~ 500 GB per field
- accurate memory planning required when designing tools
- testing with high-resolution data important
- avoid creating unnecessary memory copies

"Optimal" file size/chunking strategy

- keep file sizes close but below 2 GB?
- same chunking for fields in same table?
- align time chunks with decades (e.g. 1950-1959, 1960-1969)?
- same number of records in each chunk?

Design considerations

Choice of programming language FORTRAN vs C/python

- doesn't require special proficiency from scientists
- easy to debug
- high-degree of memory control
- good performance, particularly when using intel compiler
- easy to MPI parallelize

Intention to use python for CMIP7

Memory usage

- memory: ¼º X 60L X 8 byte precision ~ 500 GB per field
- accurate memory planning required when designing tools
- testing with high-resolution data important
- avoid creating unnecessary memory copies

"Optimal" file size/chunking strategy

- keep file sizes close but below 2 GB?
- same chunking for fields in same table?
- align time chunks with decades (e.g. 1950-1959, 1960-1969)?
- same number of records in each chunk?

Moving to CMIP6 cmor-ization using cmor3

- Successful first test of cmor-izing NorESM output with cmor3 library
- Full implementation of CMIP6 cmor-ization planned for beginning of August
- Main challenges relative to CMIP5
  - CMOR3 functionality and calls have been changed and extended as result of new table format
  - MANY new tables (43 in total) and variables (~1200 in total)
  - Statistical reductions (zonal means, climatologies of e.g. diurnal cycle)
  - Requirement for horizontal regridding?

Moving to CMIP6 cmor-ization using cmor3

- Implementation of CMIP6 tables and variables is work in progress
- Use google-sheet to organize & coordinate collaborative effort

docs.google.com

	A	B	C	D	E	F	G	H	I
1	Priority	Component	CMIP name	NorESM implementation	CMOR implementation	Frequencies	Long name	Description	Units
2	1	aerosol	zg500	Z500	no	6hrPt	Geopotential Height at 500 hPa	geopotential heig	m
3	1	atmos	areacella	area	yes	fx	Grid-Cell Area for Atmospheric Grid Variables	For atmospheres	m2
4	1	atmos	ccb	PCONVB	yes	mon	Air Pressure at Convective Cloud Base	Where convectiv	Pa
5	1	atmos	cct	PCONVT	yes	mon	Air Pressure at Convective Cloud Top	Where convectiv	Pa
6	1	atmos	ch4	CH4	yes	mon,monC	CH4 volume mixing ratio	Mole fraction is u	mol mol-1
7	1	atmos	ch4global	ch4vmr	yes	mon,monC	Global Mean Mole Fraction of CH4	Global Mean Mol	1e-09
8	1	atmos	ci	FREQZM	yes	mon	Fraction of Time Convection Occurs in Cell	Fraction of time t	1
9	1	atmos	cl	CLOUD	yes	mon	Percentage Cloud Cover	Percentage clouc	%
10	1	atmos	cli	CLDICE	yes	mon	Mass Fraction of Cloud Ice	Includes both lar	kg kg-1
11	1	atmos	clivi	TGCLDIWP	yes	mon	Ice Water Path	mass of ice wate	kg m-2
12	1	atmos	clt	CLDTOT	yes	3hr,day,mon	Total Cloud Cover Percentage	Total cloud area	%
13	1	atmos	clw	CLDLIQ	yes	mon	Mass Fraction of Cloud Liquid Water	Includes both lar	kg kg-1
14	1	atmos	clwvi	TGCLDLWP+TG	yes	mon	Condensed Water Path	Mass of condens	kg m-2

Notes | All | CMIP=DECK+historical | AerChemMIP | C4MIP | CFMIP | CORDEX | DAMIP | Explore

https://docs.google.com/spreadsheets/d/154dL8auLawUQybsKWxLNfiqqDwAhAgzTQVhw0jXfIl/edit?usp=sharing