CORDEX archive specifications for ocean variables

for CORDEX (currently based on Med-CORDEX needs)

Version 0: S. Somot, 8 march 2018, based on the documents produced for Med-CORDEX phase 1

somot_HyMeX-MedCORDEX_ocean_file_var_13aug2013.pdf

variable_names_sevault_lele_ali_9fev2015.pdf

and by CORDEX:

CORDEX_variables_requirement_table_21feb2014.pdf

CORDEX_archive_specifications_3march2014.pdf

Text in green underlines the new specificities adapted to the ocean component

Version 1: S. Somot, 15jan 2020, correction after the Toulouse Med-CORDEX meeting (Nov 2019), corrections are underlined in blue

Objectives of the document:

Following email exchanges with Gregory Nikulin from CORDEX-SAT, the goal of this document is not to develop a new archive specification document but to adapt the existing CORDEX document to new components of the regional climate system such as ocean, river, land-use, aerosols, …

***

1. Introduction

same as in CORDEX_archive_specifications_3march2014.pdf

that is to say:

• Core: monthly (and seasonal if usefull). For the ocean component, part of it is domain specific (e.g. sea ice variables for Baltic Sea)

• Tier 1: daily. For the ocean component, part of it is be domain specific (e.g. sea ice variables)

• Tier 2 : high frequency or very specific variables

Tier 2 data will be stored locally at modeling centers and made available on an informal basis upon request. Core and Tier 1 data will be published in central CORDEX archives. These may

be searchable archives relying on the metadata, like, for example, the CORDEX ESGF archives or Med-CORDEX database.

2. File format

same as in CORDEX_archive_specifications_3march2014.pdf

2.1 Netcdf attributes

same as in CORDEX_archive_specifications_3march2014.pdf

with new ocean-oriented examples in Appendix A (see below)

3. Grids

3.1 CORDEX domains

For the ocean component of coupled models, for example, running on the already-existing CORDEX domains (Mediterranean, Arctic), we will simply keep the CORDEX domain names defined for the atmospheric component (e.g. MED-44, MED-11 for Med-CORDEX). This for example apply for all the Med-CORDEX baseline runs. In particular, this will allow to keep the same file name whatever the published variable (atmosphere, ocean, river, …).

Examples from Med-CORDEX phase 2 :

CNRM-RCSM6 → atmosphere is at 12km (0.11°) and ocean is at 6-8km → MED-11 whatever the published variable

For non-coupled regional ocean model domains not defined in CORDEX, names have to be proposed, for example for the Baltic Sea region or sub-regions of the Mediterranean Sea.

The domain acronym has to be ‘domain’-‘resolution’, for example 'domain'-44, 'domain'-22, 'domain'-11, etc. corresponding to the chosen grid spacing. Names of the domains are provided in Table 1.

Example to define the ocean model resolution:
The model resolution is given in degrees. For example, MED-22 means Whole Mediterranean Sea at an horizontal resolution of 0.22° = 110 km x 0.22° = 25km (1/4°)
MED-09 = 0.09° = 10km
To choose your MED-XX, take your model mean resolution (ex: 12km) and make: 12 km / 110 km = 0.109 = 0.11 --> MED-11
some examples from Med-CORDEX phase 1 database:

INSTM-INSTMED06 --> 25km --> MED-22
NKUA-ALERMO30 for the Aegean Sea → 1/30° → AEG-03
ENEA-MITgcm1 → 1/xx → MED-12
CNRM-NEMOMED8 --> 9-12km --> 11km --> MED-10
CNRM-NEMOMED12 --> 6-8km --> 7 km --> MED-06
ENSTA-NEMOMED36 --> 2-2.5km --> MED-02
LA-SYMPHONIE111 --> 1km --> MED-01

For native coordinate systems data files must contain required grid information in the variable attribute - grid_mapping and in the coordinate variables (TO BE CLARIFIED FOR THE OCEAN MODELS) in accordance with the CF-1.4 or later convention [3]. Examples can be found in Appendix A. It is strongly recommended to provide the geographic latitudes and longitudes of the model grid cell positions as well, in order to facilitate the data analyses. Note that longitudes have to have absolute values as small as possible and monotonic. All variables have to be provided on their native computational grid.

3.2 Regular geographic grid

not yet defined for CORDEX-Ocean domains

3.1 Vertical coordinates

Despite the CORDEX-atmosphere practice, 3D fields are allowed for the ocean components as in CMIP. Some standard levels could be defined later if required

4. Time coordinates

same as in CORDEX_archive_specifications_3march2014.pdf

5 Data Reference System

same as in CORDEX_archive_specifications_3march2014.pdf

5.1 DRS elements

The DRS element values have to consist of the characters a-z, A-Z, 0-9 and '-' (dash). No other character is allowed. The terms in brackets following the DRS element names in the list below indicate whether the values are prescribed ('single value'), have to be taken from a fixed list of values ('CV'), have to be registered with the CORDEX ('CV to register'), or can be chosen freely. Note that most elements must have the same value as a mandatory NetCDF attribute. The attribute name is included in the brackets after the element name in these cases.

The elements are (in alphabetical order):

activity (single value; project_id) has to have the value 'CORDEX'.

CMIP5EnsembleMember (CV; driving_model_ensemble_member) identifies the ensemble member of the CMIP5 experiment that produced the forcing data. It has to have the same value in CORDEX as in CMIP5. For evaluation runs it has to be r1i1p1. Invariant fields (frequency=fx) may have the value r0i0p0 or that of the corresponding GCMEnsembleMember attribute.

CMIP5ExperimentName (CV; driving_experiment_name) is either evaluation or the value of the CMIP5 experiment_id of the data used.

Domain (CV; CORDEX_domain) is the name assigned to each of the CORDEX regions and includes a flag for resolution as listed in Table 1or Table 2. Please refer also to section 3.1.

GCMModelName (CV; driving_model_id) 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). Note that dashes in either of the two parts are allowed. The CV list of GCMModelName can be found at http://cordex.dmi.dk/joomla/images/CORDEX/GCMModelName.txt.

Institution (CV; institute_id) is an identifier for the institution that is responsible for the scientific aspects of the CORDEX simulation (RCM configuration, experiments ...). The CV for Institution has to be coordinated in the worldwide CORDEX community. The actual state of the CV is found at

http://cordex.dmi.dk/joomla/images/CORDEX/RCMModelName.txt

New institute names should be added for participants who never published on the ESGF (ex : ITU, ENEA) → send an email to obc@dmi.dk

product (single value) has to have the value output.

RCMModelName (CV to register; model_id) is an identifier of the CORDEX RCM. It consists of the Institution identifier (see above) and a model acronym, connected by a dash (e.g. CNRM-RCSM6 or ICTP-RegCM-ES). The CV of the RCMModelName has to be coordinated in the worldwide CORDEX community. The actual state of the CV is found at http://cordex.dmi.dk/joomla/images/CORDEX/RCMModelName.txt
New names should be added for the new coupled configurations such as CNRM-RCSM6 or ICTP-RegCM-ES for example

RCMVersionID (free string; rcm_version_id) identifies reruns with perturbed parameters or smaller RCM release upgrades, i.e. equivalent simulations. Major upgrades and improvements should be reflected in the RCMModelName.

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

StartTime and **EndTime** (build rule) indicate the time span of the file content. The format is YYYY[MM[DD[HH[MM]]]], i.e. the year is represented by 4 digits, while the month, day, hour, and minutes are represented by exactly 2 digits, if they are present at all. In accordance with CMIP5, only those digits have to be included that are necessary to indicate the file content. The StartTime and EndTime of instantaneous data are based on the time values of the first and last record in the file, while averaged data use the left and right time_bnds values respectively. It is also allowed to use time values of the first and last records in NetCDF files for averaged data, however. The two dates are separated by a dash. All time stamps refer to UTC. Examples are found in Appendix C. See also section 5.4 for more details.

VariableName (CV) is the name of the target variable in the NetCDF files. The CV is found in column B of sheet 'all' of VR [1]p.

Most of the DRS elements correspond to a mandatory NetCDF global attribute which must have the same value. These pairs of DRS element and global NetCDF attribute are listed in section C of sheet 'Global attributes' of VR [1].

5.2 DRS file naming:

same as in CORDEX_archive_specifications_3march2014.pdf

+ see Appendix C

5.3 DRS Directory structure

same as in CORDEX_archive_specifications_3march2014.pdf

5.4 Time periods for each data file

same as in CORDEX_archive_specifications_3march2014.pdf

The time spans that have to be included into a single file depend on the aggregation, which is 3-hourly, 6-hourly, daily, monthly, seasonal, or invariant:

• 3-hourly or 6-hourly: one year,

• daily: 5 year or less,

• monthly : 10 years or less,

• invariant: single file.

Samuel's comment: I think that the 5-year organisation is a mess … I would prefer to go to one year aggregation for daily and also perhaps for the monthly. It would really simplify the work for everyone I guess

6. Variables to output:

same as in CORDEX_archive_specifications_3march2014.pdf

7. Storage in the Med-CORDEX database:

Med-CORDEX phase 1:
When files are ready and names checked with the Med-CORDEX coordinators (S. Somot) and the database manager (E. Lombardi), you can upload your files on the Med-CORDEX/HyMeX database (www.medcordex.eu) with your data producer login and password.
Ask Emanuele Lombardi <emanuele.lombardi@enea.it> if you need help.

Then your ocean dataset will be searchable activating the ocean realm in the searching tools of the database: https://www.medcordex.eu/search/index.php

Med-CORDEX phase 2 :
When files are ready, model declared to ESGF and names checked with the Med-CORDEX coordinators (S. Somot), you can upload your files on the ESGF under the project_id CORDEX

8. Metadata for the ocean model runs:

Med-CORDEX phase 1:
Metadata are key elements of a reliable dataset. It allows the users to know what are the main characteristics of the model and simulations and also, when needed, the known issues of the dataset. In HyMeX/Med-CORDEX, we decided to write one metadata sheet per simulation. Those sheets remains simple and can be full-filled on-line in the HyMeX database:
http://mistrals.sedoo.fr/HyMeX/Model-Data/
Ask Sophie Cloché <Sophie.Bouffies-Cloche@ipsl.jussieu.fr> or Karim Ramage <Karim.Ramage@ipsl.polytechnique.fr> if you need help.
You can find example of already completed metadata sheets in:
http://mistrals.sedoo.fr/HyMeX/Plateform-search/?datsType=2
and an example for a coupled model here:
http://mistrals.sedoo.fr/?editDatsId=1312&datsId=1312&project_name=HyMeX

The metadata files are clickable links in the list of runs: for example for the coupled runs of Med-CORDEX
https://www.medcordex.eu/Tabelle_RUNs/Listofruns_Med-CORDEX_phase1_RCSM.pdf

Med-CORDEX phase 2:
The metadate files can be found here https://www.medcordex.eu/baseline-runs.php
Contact S. Somot to have access to the editable tables and to complete them for your runs.

Acknowledgments:

same as in CORDEX_archive_specifications_3march2014.pdf

+ some additions: currently the Med-CORDEX contributors to the file naming and variable lists are Gianmaria Sannino, Ali Harzallah, Samuel Somot, Florence Sevault and Emanuele Lombardi

References:

same as in CORDEX_archive_specifications_3march2014.pdf

Tables

Table 1 CORDEX Domains

Area	Short name	Resolution	limits	
Mediterranean Sea	MED-06, MED-10, MED-12, MED-22	Model dependant	- Black Sea not included - Gibraltar Strait at the West	
Aegean Sea	AEG-03			
Arctic Sea				
Baltic Sea				
Carabian Sea				
China Sea				

Table 2 Regular CORDEX grid

nothing defined yet but could be a good idea as the biodiversity community in the Med Sea do share a common and regular grid with a 10km grid mesh

Appendix A Examples of Netcdf headers

Examples of global attributes :

:Conventions = "CF" ;

:production = "CNRM-NEMOMED8 ocean model" ;

:institution = "CNRM (Centre National de Recherches Meteorologiques), Meteo-France, Toulouse" ;

:contact = "florence.sevault@meteo.fr" ;

:experiment_id = "NM8-24" ;

:frequency = "mon" ;

:creation_date = "2014-11-18 15:29:44" ;

and for 1D : :basin = "Mediterranean Sea East of 5.6W" ;

Examples of coordinates attributes:

float latitude(y, x) ;

latitude:units = "degrees_north" ;

latitude:standard_name = "latitude" ;

float longitude(y, x) ;

longitude:units = "degrees_east" ;

longitude:standard_name = "longitude" ;

Examples of vertical coordinates attributes:

double gdept(z) ;

gdept:units = "m" ;

gdept:standard_name = "mediterranean_vertical_grid" ;

Examples of time attributes:

float time(time) ;

time:standard_name = "time" ;

time:units = "seconds since 1961-01-01 00:00:00" ;

Examples of 3D variable attributes : should at least contain :

thetao:units = "K" ;

thetao:_FillValue = 1.e+20f ;

thetao:standard_name = "sea_water_potential_temperature" ;

thetao:valid_min = -1.e+20f ;

thetao:valid_max = 1.e+20f ;

thetao:cell_method = "time:mean" ;

thetao:coordinates = "latitude longitude gdept time" ;

Examples of 2D variable attributes : should at least contain :

zos:units = "m" ;

zos:_FillValue = 1.e+20f ;

zos:standard_name = "sea_surface_height_above_geoid" ;

zos:valid_min = -1.e+20f ;

zos:valid_max = 1.e+20f ;

zos:cell_method = "time:mean" ;

zos:coordinates = "latitude longitude time" ;

Examples of 1D variables attributes : should at least contain :

sosa:units = "PSU" ;

sosa:standard_name = "sea_surface_salinity" ;

sosa:cell_method = "basin:mean" ;

Appendix B Lists of CORDEX variables

see the ocean variable requirement file for a full description:
https://www.medcordex.eu/CORDEX-ocean_variables_requirement_table_15jan2020.pdf

Description, standard_name	Short name
sea_water_potential_temperature	thetao
sea_water_salinity	so
sea_water_x_velocity	uo
sea_water_y_velocity	vo
sea_water_z_velocity	wo
sea_surface_height_above_geoid	zos
sea_surface_temperature	tos
sea_surface_salinity	sos
ocean_mixed_layer_thickness_defined_by_sigma_t	mlotst
surface_downward_heat_flux_in_sea_water	hfds
heat_flux_correction	hfcorr
net_downward_shortwave_flux_at_sea_water_surface	rsntds
surface_downward_x_stress	tauuo
surface_downward_y_stress	tauvo
water_flux_into_sea_water	wfo
water_flux_into_sea_water_without_runoff_without_flux_correction	wfonorunoffnocorr
water_flux_into_sea_water_without_flux_correction	wfonocorr
surface_temperature_from_atmospheric_model	tsa

surface_average_water_flux_into_sea_water	wfoa
surface_average_surface_heat_flux t	hfloa
volume_average_sea_water_potential_temperature	thetaoa
volume_average_sea_water_salinity	soa
surface_average_sea_surface_temperature	tosa
surface_average_sea_surface_salinity	sosa
surface_average_sea_surface_height	zosa
surface_average_sea_surface_height_at_the_western_boundary	zosw
gibraltar_net_water_flux	gibfx
gibraltar_net_water_flux_in	gibfxin
gibraltar_heat_flux	gibhf
gibraltar_heat_flux_in	gibhfin

Files will be stored as monthly means for the 3D variables (thetao, so, uo, vo), as daily means for the 2D variables (tos, sos, zo, mlotst) and 1D variables. 1D variables are surface or volume averages of the more classical variables plus strait transports. Sub-daily (hourly) is requested for tos only to evalute the diurnal cycle.

Appendix C Examples of file names

Real examples of files found in the Med-CORDEX database currently:

example for the hindcast/evaluation runs with ocean stand-alone:
tos_MED-22_CNRM-ARPERA2_evaluation_r1i1p1_INSTM-INSTMED06_v2_mon_198801-199712.nc
hfds_MED-06_CNRM-ALDERA1_evaluation_r1i1p1_CNRM-NEMOMED12_v1_3hr_19981201-19981231.nc
so_AEG-03_CNRM-ARPERA_evaluation_r1i1p1_NKUA-ALERMO30_v1_mon_196101-197012.nc

example for the hindcast runs for coupled regional climate models:
tos_MED-10_ECMWF-ERAINT_evaluation_r1i1p1_CNRM-RCSM4_v1_day_200901-200912.nc
zos_MED-06_CNRM-ALDERA1_evaluation_r1i1p1_ENEA-MITgcm12_v1_mon_200001-200012.nc

Real examples of files found on the ESGF for Med-CORDEX phase 2:

example for the evaluation runs with coupled models:
tos_MED-11_ECMWF-ERAINT_evaluation_r1i1p1_CNRM-RCSM6_v1_mon_198001-198912.nc

example for the hindcast runs for coupled regional climate models:
to be completed

Endnotes

same as in CORDEX_archive_specifications_3march2014.pdf