NetCDF seasonal-to-decadal data handling strategy in ENSEMBLES: Attributes for physical variables

F. J. Doblas-Reyes, ECMWF

Version 22/10/2007

1. Introduction

The CF conventions are designed to promote the processing and sharing of files created using the NetCDF format. The conventions define metadata that provide a definitive description of what the data in each file represents, and of the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, regridding, and display capabilities. The ENSEMBLES seasonal-to-decadal (s2d) ocean data should be written in CF-compliant (http://www-pcmdi.llnl.gov/cf) NetCDF format. The atmospheric data are also written in NetCDF using a GRIB to NetCDF convertor. The data is publicly disseminated via an OPeNDAP (also known as DODS) interface, which can work with NetCDF files.

NetCDF files describe their content using, among other information, variables, which include coordinates (spatial and temporal) and physical magnitudes. The CF convention provides a syntax that can be used to clearly identify the physical variables. Three main attributes are used to provide a description of the physical variable:

- Standard name: The “standard_name” attribute provides a definite description of the physical quantity, which allows users of the data from different sources to determine whether quantities are in fact comparable. Modifications to the magnitude specified in the “standard_name” attribute due to common statistical operations (such as a time average or accumulation) can be expressed via the “cell_methods” attribute.

- Long name: The “long_name” attribute contains a long descriptive name of the variable, which may be used, for example, for labelling plots. It is

highly recommended that this attribute along with the “standard_name” be provided to make the file self-describing.

- Units: The units attribute is required for all variables that represent dimensional quantities. The value of the “units” attribute is a string. Units are not required for dimensionless quantities. A variable with no units attribute is assumed to be dimensionless. However, a “units” attribute specifying a dimensionless unit may optionally be included. The conforming unit for quantities that represent fractions, or parts of a whole, is "1”. Note that case is significant in the “units” strings.

The specification of these attributes is important for data exchange to avoid descriptions arbitrarily assigned by the institution that provides the data. This report provides a list of the attributes to be used to encode the ENSEMBLES s2d variables defined in the common list:

http://www.ecmwf.int/research/EU_projects/ENSEMBLES/news/common_variables.html

2. Names of the variables

In the following, it is assumed that data to be exchanged are raw model output. A list of the standard names, along with the long names and units to describe the physical variables to be archived as part of the common dataset in the s2d ENSEMBLES integrations follows. These attributes can be checked with the list maintained by the CF group in http://cf-pcmdi.llnl.gov/documents/cf-standard-names, where thorough descriptions of the variables are available. Another site with information on short names for the variables is http://www-pcmdi.llnl.gov/ipcc/standard_output.html.

Short names

Short names, prepared following a convention from the IPCC data archive, are used as tags in the NetCDF file to simplify the handling of the physical variables. This feature is not a CF convention. The name appears in front of all the attributes. An example is “thetao” for the ocean temperature. This is expected to facilitate the convergence with the rules used to archive and disseminate data of experiments for different time scales.

Ocean, 3D fields

- GRIB code: 129
    - thetao:standard_name = "sea_water_potential_temperature" ;
    - thetao:long_name = "potential temperature wrt surface" ;
    - thetao:units = "K" ;
- GRIB code: 130
    - so:standard_name = "sea_water_salinity" ;
    - so:long_name = "sea water salinity" ;
    - so:units = "PSU" ;
- GRIB code: 131
    - uo:standard_name = "sea_water_x_velocity" ;
    - uo:long_name = "zonal current positive to the east" ;
    - uo:units = "m s-1" ;
- GRIB code: 132
    - vo:standard_name = "sea_water_y_velocity" ;
    - vo:long_name = "meridional current positive to the north" ;
    - vo:units = "m s-1" ;
- GRIB code: 133
    - wo:standard_name = "upward_sea_water_velocity" ;
    - wo:long_name = "vertical current positive to the surface" ;
    - wo:units = "m s-1" ;

Ocean, 2D fields

- GRIB code 145

- zoh:standard_name = "sea_surface_height_above_geoid" ;
- zoh:long_name = "sea surface height above the geoid" ;
- zoh:units = "m" ;

- GRIB code: 148
  - zmlo:standard_name = "ocean_mixed_layer_thickness" ;
  - zmlo:long_name = "depth of the ocean mixed layer from the surface" ;
  - zmlo:units = "m" ;
- GRIB code: 163
  - t20d:standard_name = " depth" ;
  - t20d:long_name = "depth " ;
  - t20d:units = "m" ;
  - t20d:coordinates = “d20“ ;

  with a coordinate variable to specify the temperature
  - float sc
  - sc:data_type = float ;
  - sc:standard_name = "sea_water_potential_temperature" ;
  - sc:long_name = "potential temperature wrt surface " ;
  - sc:units = "K" ;
  - In the data section: sc=293.15 ;
- GRIB code: 164
  - thetaot:standard_name = "sea_water_potential_temperature" ;
  - thetaot:long_name = "vertically averaged ocean temperature" ;
  - thetaot:units = "K" ;

- thetaot:cell_methods = "depth: mean" ;

with a coordinate variable to specify the range of depths used in the vertical averaging (with dimensions depth = 1 and depth_bnd = 2):

- float depth(depth) ;
- depth:data_type = "float" ;
- depth:standard_name = "depth" ;
- depth:units = "m" ;
- depth:bounds="depth_bnd" ;
- float depth_bnd(depth, depth_bnd) ;
- depth_bnd:units = "m" ;
- In the data section depth = 150 and depth_bnd = 0, 300 ;

Atmosphere, 3D fields

- GRIB code: 129
  - g:standard_name = "geopotential" ;
  - g:long_name = "geopotential" ;
  - g:units = "m2 s-2" ;
- GRIB code: 130
  - ta:standard_name = "air_temperature" ;
  - ta:long_name = "air temperature" ;
  - ta:units = "K" ;
- GRIB code: 131
  - ua:standard_name = "eastward_wind" ;
  - ua:long_name = "zonal wind positive to the east" ;
  - ua:units = "m s-1" ;

- GRIB code: 132
    - ua:standard_name = "northward_wind" ;
    - ua:long_name = "meridional wind positive to the north" ;
    - ua:units = "m s-1" ;
- GRIB code: 133
    - hus:standard_name = "specific_humidity" ;
    - hus:long_name = "mass fraction of water vapour in moist air" ;
    - hus:units = "kg kg-1" ;

Atmosphere, near-surface fields

- GRIB code: 139
    - tas:standard_name = "surface_temperature" ;
    - tas:long_name = "SST over sea, soil temperature over land and ice temperature over sea ice" ;
    - tas:units = "K" ;
- GRIB code: 141
    - snld:standard_name = "lwe_thickness_of_surface_snow_amount" ;
    - snld:long_name = "snow amount in metres of water equivalent" ;
    - snld:units = "m" ;
- GRIB code: 146
    - hfssd:standard_name = "surface_downward_sensible_heat_flux" ;
    - hfssd:long_name = "heat exchange between the surface and the air by motion of air" ;
    - hfssd:units = "W m-2" ;
    - hfssd:cell_methods = "leadtime: sum" ;

where leadtime is the short name for the coordinate variable with standard_name “forecast_period”.

- GRIB code: 147
    - hflsd:standard_name = "surface_upward_latent_heat_flux" ;
    - hflsd:long_name = "heat exchange between the surface and the air by water phase changes" ;
    - hflsd:units = "W m-2" ;
    - hflsd:cell_methods = "leadtime: sum" ;
- GRIB code: 151
    - psl:standard_name = "air_pressure_at_sea_level" ;
    - psl:long_name = "mean sea level pressure" ;
    - psl:units = "Pa" ;
- GRIB code: 164
    - clt:standard_name = "cloud_area_fraction" ;
    - clt:long_name = "cloud area fraction for the whole atmosphere column" ;
    - clt:units = "1" ;
- GRIB code: 165
    - uas:standard_name = "eastward_wind" ;
    - uas:long_name = "near-surface zonal wind positive to the east" ;
    - uas:units = "m s-1" ;
    - uas:coordinates = “zns“ ;

  with a coordinate variable to specify the height:

    - double zns
    - zns:standard_name = "height" ;

    - zns:long_name = "height" ;
    - zns:positive = "up" ;
    - zns:units = "m" ;
    - zns=10 ;

- GRIB code: 166
    - vas:standard_name = "northward_wind" ;
    - vas:long_name = "near-surface meridional wind positive to the north" ;
    - vas:units = "m s-1" ;
    - vas:coordinates = “zns“ ;

  with a coordinate variable to specify the height:
    - double zns
    - zns:standard_name = "height" ;
    - zns:long_name = "height" ;
    - zns:positive = "up" ;
    - zns:units = "m" ;
    - zns=10 ;

- GRIB code: 167
    - tas:standard_name = "air_temperature" ;
    - tas:long_name = "near-surface temperature" ;
    - tas:units = "K" ;
    - tas:coordinates = “zns“ ;

  with a coordinate variable to specify the height:
    - double zns

    - zns:standard_name = "height" ;
    - zns:long_name = "height" ;
    - zns:positive = "up" ;
    - zns:units = "m" ;
    - zns=1.5 or 2 (depending on the model) ;
- GRIB code: 168
    - tdps:standard_name = "dew_point_temperature" ;
    - tdps:long_name = "near-surface dewpoint temperature" ;
    - tdps:units = "K" ;
    - tdps:coordinates = “zns“ ;

  with a coordinate variable to specify the height:

    - double zns
    - zns:standard_name = "height" ;
    - zns:long_name = "height" ;
    - zns:positive = "up" ;
    - zns:units = "m" ;
    - zns=1.5 or 2 (depending on the model) ;
- GRIB code: 169
    - rsds:standard_name = "surface_downwelling_shortwave_flux_in_air" ;
    - rsds:long_name = "solar radiation incident at the surface" ;
    - rsds:units = "W m-2" ;
    - rsds:cell_methods = "leadtime: sum" ;
- GRIB code: 175

    - rlds:standard_name = "surface_downwelling_longwave_flux_in_air" ;
    - rlds:long_name = "longwave radiation incident at the surface" ;
    - rlds:units = "W m-2" ;
    - rlds:cell_methods = "leadtime: sum" ;
- GRIB code: 176
    - rss:standard_name = "surface_net_downward_shortwave_flux" ;
    - rss:long_name = "difference of solar radiation from above and from below at the surface" ;
    - rss:units = "W m-2" ;
    - rss:cell_methods = "leadtime: sum" ;
- GRIB code: 177
    - rls:standard_name = "surface_net_downward_longwave_flux" ;
    - rls:long_name = "difference of longwave radiation from above and from below at the surface" ;
    - rls:units = "W m-2" ;
    - rls:cell_methods = "leadtime: sum" ;
- GRIB code: 178
    - rst:standard_name = "toa_net_downward_shortwave_flux" ;
    - rst:long_name = "difference of solar radiation from above and from below at the top of the atmosphere" ;
    - rst:units = "W m-2" ;
    - rst:cell_methods = "leadtime: sum" ;
- GRIB code: 179
    - rlut:standard_name = "toa_net_downward_longwave_flux" ;

- rlut:long_name = "difference of longwave radiation from above and from below at the top of the atmosphere" ;
- rlut:units = "W m-2" ;
- rlut:cell_methods = "leadtime: sum" ;

- GRIB code: 182
  - evlwr:standard_name = "lwe_water_evaporation_rate" ;
  - evlwr:long_name = " moisture flux from the surface into the atmosphere " ;
  - evlwr:units = "m s-1" ;
  - evlwr:cell_methods = "leadtime: sum" ;
- GRIB code: 201
  - tasmax:standard_name = "air_temperature" ;
  - tasmax:long_name = "near-surface maximum temperature in the previous 24 hours computed from 6-hourly data" ;
  - tasmax:units = "K" ;
  - tasmax:cell_methods = "leadtime: maximum (interval: 6h)" ;
  - tasmax:coordinates = “zns“ ;

  with a coordinate variable to specify the height:
  - double zns
  - zns:standard_name = "height" ;
  - zns:long_name = "height" ;
  - zns:positive = "up" ;
  - zns:units = "m" ;
  - zns=1.5 or 2 (depending on the model) ;

- GRIB code: 202
    - tasmin:standard_name = "air_temperature" ;
    - tasmin:long_name = "near-surface minimum temperature in the previous 24 hours computed from 6-hourly data" ;
    - tasmin:units = "K" ;
    - tasmin:cell_methods = "leadtime: minimum (interval: 6h)" ;
    - tasmin:coordinates = “zns“ ;

  with a coordinate variable to specify the height:

    - double zns
    - zns:standard_name = "height" ;
    - zns:long_name = "height" ;
    - zns:positive = "up" ;
    - zns:units = "m" ;
    - zns=1.5 or 2 (depending on the model) ;
- GRIB code: 228
    - prlr:standard_name = "lwe_precipitation_rate" ;
    - prlr:long_name = "total precipitation" ;
    - prlr:units = "m s-1" ;
    - prlr:cell_methods = "leadtime: sum" ;
- GRIB code: 229
    - mrsov:standard_name = "volume_fraction_of_water_in_soil " ;
    - mrsov:long_name = "total volumetric soil moisture (liquid and solid) content" ;
    - mrsov:units = "1" ;

- mrsov:cell_methods = "soil_layers: sum" ;

Fixed fields

- GRIB code: 172
  - standard_name = "land_area_fraction" ;
  - long_name = "land-sea mask, 1 all land, 0 all sea" ;
  - units = "1" ;
- GRIB code: equivalent to 129 at the surface, but with different units
  - standard_name = "surface_altitude" ;
  - long_name = "orography" ;
  - units = "m" ;

3. Accumulated variables and monthly means: cell methods

The “cell_methods” attribute is used to describe operations carried out with the field. This is the case of accumulated variables, such as the daily fluxes (which are accumulated over 24 hour periods), variables obtained after specific operations, such as the maximum and minimum near-surface temperature, and all monthly means. This is a string attribute comprising a list of blank-separated words of the form "name: method (interval: time or levels)". Each "name: method" pair indicates that for the axis identified by name, the values representing the field have been determined or derived by the specified method. The token name can be a dimension of the variable, a scalar coordinate variable, or a valid standard name. The values of the method can be point, sum, mean, maximum, minimum, mid_range, standard_deviation, variance, mode, and median. Case is not significant in the method name. It must be remembered that the method applies only to the axis indicated, and different methods may apply to other axes. The interval between the cells used in the operation is included in brackets after the key word “interval”.

As an example, suppose the quantities "accumulated precipitation" and "precipitation rate", which apply, respectively, to the daily and monthly mean values defined in ENSEMBLES. Each physical variable has a time axis, with time intervals

associated with each point on the time axis via a boundary variable. A variable representing accumulated precipitation is extensive in time and requires a time interval to be completely specified. Hence its default interpretation should be that the cell represents the time interval over which the precipitation was accumulated. This is indicated explicitly by setting the cell method to sum. A precipitation rate on the other hand is intensive in time and could equally well represent an instantaneous value or a mean value over the time interval specified by the cell. If the mean method (necessary to indicate a monthly mean) is not specified, the default interpretation for the quantity would be as an instantaneous value, which is not the case in ENSEMBLES.

If more than one cell method is to be indicated, the methods should be arranged in the order they were applied. The left-most operation is assumed to have been applied first. This rule should be applied to the 300-metre averaged temperature in the ocean:

thetaot:cell_methods = "depth: mean time: mean (interval 1 hour)" ;

where the “interval 1 hour” additional information for the “time” axis corresponds to the case of the HOPE model, which accumulates the values every time step (which is of one hour). If the methods could have been applied in any order without affecting the outcome, they may be put in any order in the cell_methods attribute.

The description of all physical variables should include the following line

physical_variable:coordinates = "experiment_id source realization institution";

which specifies that those coordinates provide essential information to interpret the physical variable. That line should include also all coordinate variables that determine the depth, temperature or any other additional feature of the physical variable. For instance, for the 300-metre averaged temperature:

thetaot:coordinates = "experiment_id source realization institution depth" ;

Besides, in the case of the forecasts it should also include the leadtime and reftime coordinates. When this information is missing, any retrieval with THREDDS or NCO will fail to include this basic information among the retrieved variables.