

A simple note for HDF4/HDF-EOS2 to CF Mapping

1 Name conventions

In general, for any character not allowed by the CF name conventions, change that character to underscore ('_').

1.1 HDF-EOS2

In general, variable names of an HDF-EOS2 multi-grid/multi-swath/multi-zonal-average file should have the corresponding grid/swath/zonal-average names prefixed before the field names. Variable names of an HDF-EOS2 single grid/swath/zonal-average should just use the corresponding field names.

1.2 HDF4

1.2.1 SDS

SDS objects in HDF4 can attach to an HDF4 Vgroup, which is logically a container that can consist of many HDF4 objects including another Vgroup. Since individual SDS objects under different HDF4 Vgroups may share the same name, to avoid the obvious name clashing and to keep the group hierarchy of the SDS objects that attaches to a Vgroup, an SDS object name candidate mapped to CF include all the ancestor vgroup paths. Underscores ('_') are used to connect between the parent vgroup path and the child vgroup path or between the last vgroup path and the SDS name retrieved by calling the SDS APIs. Then an SDS object name candidate is mapped to CF by applying the CF name rule. For SDS objects that don't attach to any HDF4 Vgroups, the SDS object names retrieved by calling the SDS APIs are simply mapped to CF by applying the CF name rule.

1.2.2 Vdata

Like SDS, an HDF4 Vdata can also attach to an HDF4 Vgroup. Though a Vdata object will not map to a CF variable, but a Vdata field will be mapped to a CF variable. Therefore to avoid the ambiguity of Vdata field names, a vdata field name mapped to CF includes the corresponding vdata path. Moreover, to ensure that the CF users can understand the final mapped CF variable is originally an HDF4 Vdata field, string "vdata" is prefixed before the Vdata path and the string "vdf" (stands for Vdata Field) is used to connect the Vdata path and the Vdata field name.

1.3 HDF-EOS2 Hybrid

Since we find that some added SDS objects share the same names with the field names of some HDF-EOS2 grids or swaths, to avoid the massive name clashing and

to distinguish the non-HDFEOS2 objects from the HDFEOS2 objects, we add a string "NONEOS" at the end of the name of an added SDS object in an HDF-EOS2 file.

2 Dimensions

Dimensions in HDF-EOS2 will map to CF dimensions. Fake dimensions may be added to HDF4 SDS or Vdata fields.

3 Coordinate Variables

Latitude and longitude are properly retrieved from either HDF-EOS2 or HDF4 files. CF units "degrees_east" and "degrees_north" are added to latitude or longitude. CF requires that each dimension not associated with latitude or longitude has a corresponding coordinate variable. So the handler will try to first search such coordinate variable. If the handler cannot find such a variable, it will generate a proxy coordinate variable for that dimension. The value of this coordinate variable is index number 0,1,2,3..... Attribute units=level is added to this coordinate variable.

If latitude and longitude can be described as 1-D arrays, COARDS are followed.

If the file is an HDF-EOS2 swath or a curvilinear grid, we add a CF "coordinates" attribute to a variable to describe the coordinates of that variable.

4. CF attributes

CF attribute `_FillValue`, `valid_range`, `scale_factor`, `add_offset` are made to follow CF conventions for most NASA HDF-EOS2 products and some HDF4 TRMM, CERES and OBPG products.