Observation Variable Specifications

Variable naming conventions

Regardless of the file type that IODA will eventually use (currently netcdf), we want to use the netcdf-CF naming convention for variables inside the file. Unfortunately, the netcdf-CF convention doesn't cover the entire space of observations types meaning that observation variables that exist which have no corresponding netcdf-CF name. For these cases we should create names that are similar to those in the netcdf-CF set. For example, there is no convention for aerosol optical depth (AOD). netcdf-CF names tend to be full words separated by underscores (air_temperature for sensible atmospheric temperature) so AOD could be named aerosol_optical_depth to make it appear netcdf-CF-like.

Here are some links to the netcdf-CF website:

- netcdf-CF main website
- search form for looking up netcdf-CF names

Following are tables that show the names we are using for variables in the observation files. We can update this as we add new obs types and new netcdf-CF names appear. The purpose of this table is to have a single reference for the names that we use for observation variables. Also included in the table are the names and types of units that go with the variable data.

These tables represent what we should be moving toward in the files. Currently (as of 3/8/19), we have a couple exceptions in our observation files that need to be corrected.

- Observation date and time are currently represented using a reference date with a time offset. The reference date is a netcdf attribute (integer) called "date_time" of the form YYYYMMDDHH representing the analysis date (e.g., 2018041500 for April 15, 2018 00Z). The time is a netcdf variable (float) called "time" that is an offset from the reference time in hours.
- air_pressure is currently in the units of hectopascals (mb).

Data organization in the IODA file

For each observation data variable, it is required to provide an observation value, error estimate, and QC mark. These three quantities are marked with the "group" names ObsValue, ObsError and PreQC respectively. Observation variables are stored as vectors with a length equal to the number of locations (nlocs). A location is a unique combination of meta data quantities that specify the spatial and temporal location of the corresponding observation data. For integrated quantities (e.g., radiance, brightness_temperature, aerosol_optical_depth) each location is specified with a value for latitude, longitude and date_time. For point quantities (e.g., air_temperature, eastward_wind) each location is specified with a value for latitude, longitude, date_time and some form of vertical coordinate such as air_pressure.

In the IODA netcdf file the group names are designated by adding a suffix to the variable name where the suffix is composed of an "@" symbol followed by the group name. For example with the air_temperature variable, the netcdf file would contain three vectors named:

- air_temperature@ObsValue
  - Observed value, float
- air_temperature@ObsError
  - Estimate of the error in the observed value, typically a variance, float
- air_temperature@PreQC
  - QC marks for observation values, integer

For satellite instrument observation data, there are typically multiple channels (representing different frequency bands) that each produce an observed value at each location. We are currently treating each channel as a separate variable so that these data can fit into the scheme of multiple vectors of nlocs length. By convention, the channel number is tagged onto the end of the name of the quantity being measured so that each channel has a unique name. Let's say we have an instrument yielding brightness temperature measurements that has 11 channels. The variable names we would use for the observed values would be:

- brightness_temperature_1@ObsValue
- brightness_temperature_2@ObsValue
- ...
- brightness_temperature_11@ObsValue

> (i) JEDI has a channel selection mechanism that allows for the reading in of a subset of the channels available from a given instrument. This means that it is not necessary to store all channels in the netcdf file which will become useful when hyperspectral obs types come on line.

Meta data that corresponds to locations, are stored in vectors (nlocs long) that use a "@MetaData" suffix. For example, the scan angle of an instrument would be stored in a netcdf variable named "scan_angle@MetaData".

Meta data can also correspond to the variables in the file. Using our satellite instrument example, there would be a total of 11 variables (nvars = 11) in the file. Meta data such and channel number and channel frequency would then be store in vectors that are nvars (11) in length. MetaData related to the variables use the variable name suffix "@VarMetaData". For example the channel frequencies would be stored in a netcdf variable named "channel_frequency@VarMetaData".

File Variable Conventions

The variables inside a IODA observation data file need to adhere to the following conventions:

NETCDF files

- Variable names must contain a group name
    - The group name is the "@<group_name>" suffix tagged onto the netcdf variable name.
    - Each variable in the file is specified as "<variable_name>@<group_name>"
- Variables must use the netcdf fill value for missing data marks
    - The netcdf variable attribute "_FillValue" contains the fill value
- Variables must use the correct data types
    - Allowed types are: int, float, char (no double precision)
    - All numeric data are float except for QC marks which are int
        - The checker script (see below) only verifies that variables with group "PreQC" are integer type
- Variable data cannot contain invalid numerical values (nan, inf, -inf)

The plan is to issue warnings now about variables that violate these conventions so that these violations can be located and fixed. Once all files are fixed, then the convention violation warnings will be changed to errors.

A checker script (check_ioda_nc.py) has been written for netcdf files, and will be made available in a pull request that is due to be merged around November 1, 2019. Once all files pass the checker, a ctest will be added to the ioda repository that runs the checker script on all of the netcdf files in the ioda repository so new files that come in will be immediately verified.

Variable names by observation types

In the following tables, the Data Group column denotes if the variable is observation data, meta data related to locations, or meta data related to variables. The entries in this column are the variable group name suffixes without the "@". An entry of ObsValue implies that you also have the ObsError and PreQC variables in the file.

All observation types have the location meta data as shown in this table:

Variable Name	Units Name	Description	Group Type
latitude	degrees	latitude, zero at equator, positive toward the north	MetaData
longitude	degrees	longitude, zero at prime meridian, positive toward the east	MetaData
date_time	iso_8601_format	absolute date and time of observation, in ISO 8601 date, time format	MetaData

Radiosonde, Aircraft

Variable Name	Units Name	Description	Group Type
air_temperature	kelvin	atmospheric sensible temperature	ObsValue
specific_humidity	kilogram_per_kilogram	specific humidity	ObsValue
eastward_wind	meters_per_second	zonal wind component	ObsValue
northward_wind	meters_per_second	meridional wind component	ObsValue
air_pressure	pascal	atmospheric pressure	MetaData (vertical location coordinate)

AMSU-A

Variable Name	Units Name	Description	Group Type
brightness_temperature		brightness temperature	ObsValue
Scan_Angle	degrees	instrument scan angle in degrees	MetaData
Scan_Position	dimensionless	instrument scan position	MetaData
Sat_Zenith_Angle	degrees	satellite zenith angle	MetaData
Sat_Azimuth_Angle	degrees	satellite azimuth angle	MetaData
Sol_Zenith_Angle	degrees	sun zenith angle	MetaData
Sol_Azimuth_Angle	degrees	sun azimuth angle	MetaData

chaninfoindex			VarMetaData
frequency	hertz	instrument channel frequency	VarMetaData
polarization			VarMetaData
wavenumber			VarMetaData
error_variance		statistical variance of channel measurement	VarMetaData
mean_lapse_rate			VarMetaData
use_flag			VarMetaData
sensor_chan			VarMetaData
satinfo_chan			VarMetaData

AOD

Variable Name	Units Name	Description	Group Type
aerosol_optical_depth		aerosol optical depth	ObsValue
sol_zenith_angle	degrees	sun zenith angle	MetaData
sol_azimuth_angle	degrees	sun azimuth angle	MetaData
surface_type			MetaData
modis_deep_blue_flag			MetaData
frequency	hertz	instrument channel frequency	VarMetaData
polarization			VarMetaData
wavenumber			VarMetaData
gsi_use_flag			VarMetaData
sensor_channel			VarMetaData