CF meeting
9-11 June 2020
Zoom session link

Discussion topic notes

Cell methods: "within"|"over" "days"|"months" and time axis (Section 7.4) (Lars Bärring)

https://github.com/cf-convention/cf-conventions/issues/197

There are two components:

- One is to clarify, and agree on a solution/interpretation of what the "climatology" attribute means and its relation to the cell methods "within" and "over": How are they connected? What does it mean for different typical cases? Can the "climatology" attribute be disconnected from the cell methods? Is in fact "climatology" necessary at all? What to do vs. existing CMIP6 data and what to do for CMIP7? What is the status quo, and what minimal changes may we want to make? Once these matters have been resolved this may come down to rather small changes to Section 7.4.

- The other one, is the more far-reaching need for a new or alternative mechanism that allows for a more flexible description of more complex and/or multi-step temporal processing of data.

Lars Barring, Carlee L, Charlotte Pascoe, Sebastien Villaume, Alice Bui, Anna Lo Piccolo, Antonio Cofino, Hélène Cote, Karl Taylor, Matt Donovan, Pierre-Antoinne Bretonniere, Sadie Bartholomew, Seth McGinnis, Thomas Hearty, mamc, James Johnson, Martin Juckes., Daniel Heydebreck, Sylvain Herledan, Lu Guan

---

Notes (*Please correct me if I have misinterpreted your comment!*)

Jonathan: Are there examples of climatological cell methods that is not a climatology, and vice versa?

- Lars: “Climatology” is composed of a sequence of non-contiguous intervals. So “monthly mean of daily maximum” is *not* a climatology

- Seth: Climatology only applies to the ”discontinuous” timeseries. Thus, the climatology attribute has been useful to distinguish between the continuous and discontinuous
- Jonathan, what if the data variable had a size 12 time axis (one for each month) - is that continuous?
- Seth & Martin: No - it’s discontinuous, as each value of the data variable is composed of discontinuous months
    - Martin Juckes: In the monthly mean tasmax example described by Lars, a consecutive set of values within one month is processed to produce one data value; in the 30 year average of monthly means introduced by Seth, each data value is an average of values in 30 separate intervals.
- Martin: It is somewhat of a defect that you need to examine the bounds to fully understand what the climatology means. This is difficult! This is touched in in Trac #82 https://cf-trac.llnl.gov/trac/ticket/82
- Jonathan: Summaries Seth as saying climatology meaning discontinuous”, and Lars as saying “climatology means there is within/over” . Is that summary right?
- David: The defect Martin mentions is even more difficult to deal with if there’s no climatology attribute
- Sebastien: At ECMWF, we use regularly the cell_method construct "leadtime: mean within days (interval: 1 hour) forecast_reference_time: mean over days within months" ;
    This is because we usually take the daily forecast and compute the daily average of a number of variables, then at the end of each month we compute the average of the daily average. We call this “monthly mean of daily means” but it is not a climatology so we do not want to use “climatology”, it is something completely different. We would welcome it if the cell_method construct “within/over” is NOT constrained to climatology because we have several use cases for the construct that are not climatology. In NWP, a climatology is typically a sampling of 30 years. In a sense it is the statistical asymptote that a variable would converge to. If you take the mean temperature of January 2020, it does not represent well any January. But if you take the mean of the last 30 January month, that mean would represent the “typical” temperatures in January, and that is what we call “climatology”. There are also many types of climatology, one could compute a centennial climatology by taking the last 100 years, a millennial climatology with 1000 years, etc. To me climatology is not the same thing as doing 2 consecutive post processing operations wrt time of a variable..
- Seth: it was not right that CF chose “climatology” to apply to shorter length time periods (e.g. one month). The CF usage needs to accord with the common-sense meaning of the word.
- Jonathan: there is an analogy between the annual cycle and the diurnal cycle.
- Jonathan: Perhaps a different word is needed - deprecate “climatology” and choose a new name for the future.
- Martin: Idea was to put more info into cell methods by extending the “interval” mechanism (see GitHub #197). Idea being that you can get essential information without

having to resort to inspecting the bounds. Idea: “window” syntax (first mooted in Trac #82)

- Seth: Having multiple cycles (annual/diurnal) is confusing a little bit. Makes us lose sight of the fact that we're doing the same thing: sampling from with a period - this is the *intent* of the word “climatology”
    - Notionally: window: 1 day cycle: 1 hour
- Jonathan: Extending double time processing to things that are not the diurnal/annual cycle makes it more complicated because you have to define the period (whereas it was implicit before (daily or yearly)
- Klaus: Could have a climatology type attribute on the (climatology) bounds variable (David’s interpretation, at least)
- Jonathan: Could only allow “climatology” attribute when *annual* cycles are involved, and choose another name when diurnal or any other period is involved (David: Have I got that right?)
- Klaus: A negative first reaction to the word “climatology” is not always correct.
- Karl & Seth: “Max” or ”min” is not tied to a particular point in the cycle, unlike “mean” for a particular time of day.
- Martin: e.g. in the cell methods: “time: mean (window: day) time: mean (window: month) time: mean (cycle: years)” to mean (1) calculate daily means, (2) average consecutive days over each month, (3) average months over consecutive annual cycles.
- Jonathan: Could call them “cyclical time bounds”
- Klaus: Important to distinguish between annual and daily cycle
    - Jonathan: The cell methods “over” and “within” bits gives you that
- Jonathan: Might want to calculate maximum over lunar months. CF-1.8 can not do that
- Seth: Make sure that for any resolution we come to, remember that these things are not always calculated all at once. E.g. we sometimes make some monthly mean data and then later someone else turns into annual climatology, so it must be easy to chain on further processing to the cell methods string.
- Seth: keep in mind chained processing with whatever solutions we consider, so that it’s amenable to documentation via appending to existing attributes rather than completely rewriting them
- Lars: What are the next steps?
- Jonathan: More thinking!
- Klaus: some well-defined use cases would be very useful to help us understand what we need to be encoding
- Seth: Where should we enumerate these use cases? In GitHub? Yes
    - Klaus: Also update the initial proposal
- Lars: Collect use cases in this google doc (linked from the #197)
- David: Action: check with than that this google doc is not going anywhere in the near future.
    - David (2020-06-11): It is confirmed that this doc in this folder will be around for "a good long while"
- Seth: Uses cases for daily and annual cycle *together* would be useful

    - Jonathan: There should be some examples of this already in CF-1,8
    - Likely other cycles might include tides / lunar month; forecasting cycles?
- Seth: weekly cycles could be useful when human behaviour is a factor (e.g. air pollution)

- Sebastien: After yesterday's discussion I realised that Jonathan was right, we use "climatology" for 2 completely different things and part of the misunderstanding comes from this. Then I tried to understand the need of having a "climatology" when doing 2 consecutive processing over a time range (and climatology bounds). I think it is connected to the fact that CF does not allow 2 time coordinates (2 coordinates, not 2 axes, the distinction is extremely important because many people in CF mix the 2 concepts) . So when you do a daily mean and then a mean of these means, you can't represent it properly in CF because you are forced to have 1 time coordinate but you need 2 time bounds. And this is connected to another long discussed issue of "having multiple time coordinates", see https://cf-trac.llnl.gov/trac/ticket/117 . So the way to "solve" the problem of having 1 coordinate variable and 2 bounds was to create the "climatology" and associate the second bounds to it.
  At ECMWF we usually use 3 time coordinates (actually 2, the 3rd one can be derived for the first 2 but we include it for convenience: time = forecast_ref_time + lead_time).
We still have only one of these coordinates bearing the attribute "axis: T" like this:

```
dimensions:
        t = 1;
        bnd2 = 2 ;
variables:
 double forecast_reference_time(t) ;
        forecast_reference_time:units = "hours since 1900-01-01 00:00:00" ;
        forecast_reference_time:standard_name = "forecast_reference_time" ;
        forecast_reference_time:calendar = "gregorian" ;
        forecast_reference_time:long_name = "Forecast reference time" ;
        forecast_reference_time:bounds = "forecast_reference_time_bnds" ;

 double forecast_reference_time_bnds(t,bnd2);

 double leadtime(t) ;
        leadtime:units="hours" ;
        leadtime:long_name="hours elapsed since the start of the forecast" ;
        leadtime:standard_name="forecast_period" ;
        leadtime:bounds = "leadtime_bnds" ;

 double leadtime_bnds(t,bnd2);

 double time(t) ;
        time:units = "hours since 1900-01-01 00:00:00" ;
```

time:standard_name = "time" ;
time:calendar = "gregorian" ;
time:axis = "T" ;
time:bounds = "time_bnds" ;
time:long_name = "time" ;

...lat/lon defined here ….

float iicethic(t, lat, lon) ;
iicethic:units = "m" ;
iicethic:standard_name = "sea_ice_thickness" ;
iicethic:long_name = "Ice thickness" ;
iicethic:coordinates = "time latitude longitude" ;
iicethic:cell_methods = "leadtime: mean within days (interval: 1 hour)
forecast_reference_time: mean over days within months" ;

data:

forecast_reference_time = [00]

forecast_reference_time_bnds = [00, 720]

leadtime = [00]

leadtime_bnds = [00, 23]

time = [00]

Iicethic = […]