Breakout group on **Cell methods**

Climatological time axis and cell methods *within* | *between* *days* | *years*

Focus on

- cf-conventions issue #197
- Chapter 7.4 Climatological statistics
- related to Trac ticket 82

- What does the `climatology` attribute mean?
- Its relation to cell methods `within` and `over`?
- Is `climatology` necessary?
- Can `climatology` be disconnected from the cell methods?
- What to do for CMIP7?
- What is the status quo, and what minimal changes may we want to make [to section 7.4]?
-
- A new or alternative mechanism that allows for a more flexible description of more complex and/or multi-step temporal processing of data.

Status quo

Chapter 7.4 is not quite as clear as one could wish for (as evidenced by the discussion in #197)

- It seems that *climatology* is required whenever cell methods *within* / *between* are used.
- It seems that *climatology* should be used to describe the “special time axis” required for describing the climatological annual/seasonal or diurnal cycle, i.e. calculations over a set of disconnected time intervals.
- Thus we have CMIP6 files where
  - monthly mean tas does not have *climatology* because

    time: mean within days time: mean over days ==> time: mean
  - monthly mean tasmax does not have *climatology* despite the cell method constructs *within* / *between* are used: *time: maximum within days time: mean over days*

  Although they are very similar from a climatological (general sense) point of view

- Currently allowed formats are
  - *time: method1 within years time: method2 over years*
  - *time: method1 within days time: method2 over days*
  - *time: method1 within days time: method2 over days time: method3 over years*

Different time intervals

There are four types of time intervals

- A continuous sequence of non-overlapping periods, such as a time series of hourly, daily, or annual data
- A continuous sequence of overlapping periods, such as a hourly (period: 6 hours), daily (period: 3 days), decadal data (period: 30 years). That is, some kind of running statistic.
- A discontinuous sequence of non-overlapping periods, such as what is needed to calculate a 30-year climatology of the annual cycle at daily resolution

```
time_bounds = 1971-01-01 00:00, 2000-01-02 00:00,
              1971-01-02 00:00, 2000-01-03 00:00,
              …
              1971-12-31 00:00, 2001-01-01 00:00
```

- A discontinuous sequence of overlapping periods, e.g. to calculate a 30-year climatology of the 5-day smoothed annual cycle at daily resolution

```
time_bounds = 1970-12-29 00:00, 2000-01-03 00:00,
              1970-12-30 00:00, 2000-01-04 00:00,
              …
              1970-12-29 00:00, 2001-01-03 00:00
```

Relation to existing standard names

Standard name		Cell method "processing"					
Description mentions	**Category**	**None**	**"Sea-level factors"**	**Area type subset**	**Time subset**	**Time processing**	**Total**
climatology	*_anomaly	8					8
climatology (indirect)	various	4					4
cell_method	surface_*		1	6			7
	flux_*			17			17
	mass_fraction_* (precip)			2			2
	other			1			1
	gust_*				4		4
climatological time axis & cell_method	derived statistics (climate indices)					9	9
Total		12	1	26	4	9	52

Ideas/thoughts/questions that has come up (1)

- Is there a [slight] conceptual difference between "a climatology" and "a climatological time-series"?
  -- Personally I would say yes, and this is not clear in Chapter 7.4.

- What is the function of the *climatology* attribute more precisely?
  - Is it necessary? -- **No** because all information is in the time bounds in combination with the cell methods
  - Well, **Yes** it is useful to clearly signal whether it is a “proper” time-series” calculated over a sequential series of time periods, or if it is calculated over a set of discontinuous time periods so as to describe the typical conditions (i.e. “a climatology”)
- If we keep *climatology* could it be disconnected from the cell methods constructs *within / over*?

Ideas/thoughts/questions that has come up (2)

- An extension based on a forecasting use-case:

```
cell_methods = "leadtime: mean within days
                forecast_reference_time: mean over days within months”
```

Ideas/thoughts/questions that has come up (3)

- The list of permissible combinations of within and over contain 3 alternatives. This needs to be extended (Trac ticket 82): to more than 3 steps, more flexible time specs., incl. runnings stats.
- Martin Juckes suggested, partly building on Trac ticket 82, an extension that would introduce substantially more flexibility.