Breakout group on Cell methods

Climatological time axis and cell methods \textit{within} | \textit{between} \hspace{1em} \textit{days} | \textit{years}

Focus on

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

- What does the \textit{climatology} attribute mean?
- Its relation to cell methods \textit{within} and \textit{over}?
- Is \textit{climatology} necessary?
- Can \textit{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.
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*
    
    \[
    \text{time: mean within days time: mean over days} \implies \text{time: mean}
    \]
  - monthly mean tasmax does not have *climatology despite* the cell method constructs
    
    \[
    \text{within/between are used: time: maximum within days time: mean over days}
    \]

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*
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

<table>
<thead>
<tr>
<th>Standard name</th>
<th>Category</th>
<th>None</th>
<th>“Sea-level factors”</th>
<th>Area subset</th>
<th>Time subset</th>
<th>Time processing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description mentions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>climatology</td>
<td>*_anomaly</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>climatology (indirect)</td>
<td>various</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>cell_method</td>
<td>surface_*</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>flux_*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>mass_fraction_* (precip)</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>gust_*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>climatological time axis &amp; cell_method</td>
<td>derived statistics (climate indices)</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>12</td>
<td>26</td>
<td>4</td>
<td></td>
<td></td>
<td>52</td>
</tr>
</tbody>
</table>
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:

```python
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.