

7.3. Cell Methods

To describe the characteristic of a field that is represented by cell values, we define the `cell_methods` attribute of the variable. This is a string attribute comprising a list of blank-separated words of the form "*name: method*". Each "*name: method*" pair indicates that for an axis identified by *name*, the cell values representing the field have been determined or derived by the specified *method*. For example, if data values have been generated by computing time means, then this could be indicated with `cell_methods="t: mean"`, assuming here that the name of the time dimension variable is "t".

In the specification of this attribute, *name* can be a dimension of the variable, a scalar coordinate variable, a valid standard name, or the word "area". (See Section 7.3.4, *Cell methods when there are no coordinates, concerning the use of standard names in cell_methods*.) The values of *method* should be selected from the list in [Appendix E, Cell Methods](#), which includes `point`, `sum`, `mean`, `maximum`, `minimum`, `mid_range`, `standard_deviation`, `variance`, `mode`, and `median`. Case is not significant in the method name. Some methods (e.g., `variance`) imply a change of units of the variable, as is indicated in [Appendix E, Cell Methods](#).

It must be remembered that the method applies only to the axis designated in `cell_methods` by *name*, and different methods may apply to other axes. If, for instance, a precipitation value in a longitude-latitude cell is given the method `maximum` for these axes, it means that it is the maximum within these spatial cells, and does not imply that it is also the maximum in time. Furthermore, it should be noted that if any *method* other than "point" is specified for a given axis, then `cell_bounds` should also be provided for that axis (except for the relatively rare exceptions described in Section 7.3.4, "Cell methods when there are no coordinates").

The default interpretation for variables that do not have the `cell_methods` attribute specified depends on whether the quantity is extensive (which depends on the size of the cell) or intensive (which does not). Suppose, for example, the quantities "accumulated precipitation" and "precipitation rate" each have a time axis. A variable representing accumulated precipitation is extensive in time because it depends on the length of the time interval over which it is accumulated. For correct interpretation, it therefore requires a time interval to be completely specified via a boundary variable (i.e., via a `cell_bounds` attribute for the time axis). In this case the default interpretation is that the cell method is a sum over the specified time interval. This can be (optionally) indicated explicitly by setting the cell method to `sum`. A precipitation rate on the other hand is intensive in time and could equally well represent either an instantaneous value or a mean value over the time interval specified by the cell. In this case, the default interpretation for the quantity would be "instantaneous" (which, optionally, can be indicated explicitly, by setting the cell method to `point`). More often, however, cell values for intensive

Deleted: the

Deleted: The token name can be a dimension of the variable, a scalar coordinate variable, or a valid standard name.

Deleted: and this also is specified by [Appendix D, Dimensionless Vertical Coordinates](#).

Formatted: Pattern: Clear

Deleted: indicated

Deleted: for instance,

Deleted: have cells associated with their grid points, but

Deleted: ,

Deleted: n't

Deleted: So

Deleted: , suppose

Deleted: and that time intervals are associated with each point on the time axis via a boundary variable

Deleted: and

Deleted: . Hence its

Deleted: should be

Deleted: associated with the grid point represents the time

Deleted: over which the precipitation was accumulated.

Deleted: is

Deleted: However, if the mean method is not specified then

Deleted: . The default method is indicated explicitly

Formatted: Pattern: Clear

quantities are means, and this should be indicated explicitly by setting the cell method to mean and specifying the cell bounds.

Because the default interpretation for an intensive quantity differs from that of an extensive quantity and because this distinction may not be understood by some users of the data, it is recommended that every data variable include for each of its dimensions and each of its scalar coordinate variables the `cell_methods` information of interest (unless this information would not be meaningful). It is especially recommended that `cell_methods` be explicitly specified for each spatio-temporal dimension and each spatio-temporal scalar coordinate variable.

Example 7.4. Methods applied to a time_series

Consider 12-hourly time_series of pressure, temperature and precipitation from a number of stations, where pressure is measured instantaneously, maximum temperature for the preceding 12 hours is recorded, and precipitation is accumulated in a rain gauge. For a period of 48 hours from 6 a.m. on 19 April 1998, the data is structured as follows:

```
dimensions:
  time = UNLIMITED; // (5 currently)
  station = 10;
  nv = 2;
variables:
  float pressure(station,time);
  pressure:long_name = "pressure";
  pressure:units = "kPa";
  pressure:cell_methods = "time: point";
  float maxtemp(station,time);
  maxtemp:long_name = "temperature";
  maxtemp:units = "K";
  maxtemp:cell_methods = "time: maximum";
  float ppn(station,time);
  ppn:long_name = "depth of water-equivalent precipitation";
  ppn:units = "mm";
  ppn:cell_methods = "time: sum";
  double time(time);
  time:long_name = "time";
  time:units = "h since 1998-4-19 6:0:0";
  time:bounds = "time_bnds";
  double time_bnds(time,nv);
data:
  time = 0., 12., 24., 36., 48.;
  time_bnds = -12.,0., 0.,12., 12.,24., 24.,36., 36.,48.;
```

Note that in this example the time axis values coincide with the end of each interval. It is sometimes desirable, however, to use the midpoint of intervals as coordinate values for variables that are representative of an interval. An application may simply obtain the midpoint values by making use of the boundary data in `time_bnds`.

7.3.1 Statistics for more than one axis

If more than one cell method is to be indicated, they should be arranged in the order they were applied. The left-most operation is assumed to have been applied first. Suppose, for example, that within each grid cell a quantity varies in both longitude and time and that these dimensions are named "lon" and "time", respectively. Then values representing the time-average of the zonal maximum are labeled `cell_methods="lon: maximum time: mean"` (i.e. find the largest value at each instant of time over all longitudes, then average these maxima over time); values of the zonal maximum of time-averages are labeled `cell_methods="time: mean lon: maximum"`. If the methods could have been applied in any order without affecting the outcome, they may be put in any order in the `cell_methods` attribute.

If a data value is representative of variation over a combination of axes, a single method should be prefixed by the names of all the dimensions involved (listed in any order, since in this case the order must be immaterial). Dimensions should be grouped in this way only if there is an essential difference from treating the dimensions individually. For instance, the standard deviation of topographic height within a longitude-latitude gridbox could have `cell_methods="lat: lon: standard_deviation"`. (Note also, that in accordance with the recommendation of the following paragraph, this could be equivalently and preferably indicated by `cell_methods="area: standard_deviation"`.) This is not the same as `cell_methods="lon: standard_deviation lat: standard_deviation"`, which would mean finding the standard deviation along each parallel of latitude within the zonal extent of the gridbox, and then the standard deviation of these values over latitude.

To indicate variation over horizontal area, it is recommended that instead of specifying the combination of horizontal dimensions, the special string "area" be used. The common case of an area-mean can thus be indicated by `cell_methods="area: mean"` (rather than, for example, "lon: lat: mean"). The horizontal coordinate variables to which "area" refers are in this case not explicitly indicated in `cell_methods` but can be identified, if necessary, from attributes attached to the coordinate variables, scalar coordinate variables, or auxiliary coordinate variables, as described in Chapter 4, Coordinate Types.

7.3.2 Recording the spacing of the original data and other information

To indicate more precisely how the cell method was applied, extra information may be included in parentheses () after the identification of the method. This information includes standardized and non-standardized parts. Currently the only standardized information is to provide the typical interval between the original data values to which the method was applied, in the situation where the present data values are statistically representative of original data values which had a finer spacing. The syntax is (interval: value unit), where *value* is a numerical value and *unit* is a string that can be recognized by UNIDATA's Uunits package [UDUNITS]. The *unit* will usually be dimensionally equivalent to the unit of the corresponding dimension, but this is not

~~Deleted:~~ (

~~Deleted:~~) within each gridbox. Values that represent

~~Deleted:~~ |

~~Deleted:~~ ",

~~Deleted:~~ labelled

~~Deleted:~~ , whose

~~Deleted:~~ is

~~Deleted:~~ them

~~Deleted:~~ would

~~Deleted:~~ standardized

~~Deleted:~~ standarddized

~~Deleted:~~ does not have to

~~Deleted:~~ name

required (which allows, for example, the interval for a standard deviation calculated from points evenly spaced in distance along a parallel to be reported in units of length even if the zonal coordinate of the cells is given in degrees). Recording the original interval is particularly important for standard deviations. For example, the standard deviation of daily values could be indicated by `cell_methods="time: standard_deviation (interval: 1 day)"` and of annual values by `cell_methods="time: standard_deviation (interval: 1 year)"`.

Deleted: although it often will be.

If the cell method applies to a combination of axes, they may have a common original interval e.g. `cell_methods="lat: lon: standard_deviation (interval: 10 km)"`. Alternatively, they may have separate intervals, which are matched to the names of axes by position e.g. `cell_methods="lat: lon: standard_deviation (interval: 0.1 degree_N interval: 0.2 degree_E)"`, in which 0.1 degree applies to latitude and 0.2 degree to longitude.

If there is both standardized and non-standardized information, the non-standardized follows the standardized information and the keyword `comment:`. If there is no standardized information, the keyword `comment:` should be omitted. For instance, an area-weighted mean over latitude could be indicated as `lat: mean (area-weighted)` or `lat: mean (interval: 1 degree_north comment: area-weighted)`.

A dimension of size one may be the result of "collapsing" an axis by some statistical operation, for instance by calculating a variance from time series data. We strongly recommend that dimensions of size one be retained (or scalar coordinate variables be defined) to enable documentation of the method (through the `cell_methods` attribute) and its domain (through the `cell_bounds` attribute).

Deleted: and used

Deleted: document

Example 7.5. Surface air temperature variance

The variance of the diurnal cycle on 1 January 1990 has been calculated from hourly instantaneous surface air temperature measurements. The time dimension of size one has been retained.

Deleted: measurements.

```
dimensions:
  lat=90;
  lon=180;
  time=1;
  nv=2;
variables:
  float TS_var(time,lat,lon);
  TS_var:long_name="surface air temperature variance"
  TS_var:units="K2";
  TS_var:cell_methods="time: variance (interval: 1 hr comment:
  sampled instantaneously)";
  float time(time);
  time:units="days since 1990-01-01 00:00:00";
  time:bounds="time_bnds";
  float time_bnds(time,nv);
```

Deleted: of hourly instantaneous);

```
data:
  time=.5;
  time_bnds=0.,1.;
```

Notice that a parenthesized comment in the `cell_methods` attribute provides the nature of the samples used to calculate the variance.

7.3.3 Statistics applying to portions of cells

By default, the statistical method indicated by `cell_methods` is assumed to have been evaluated over the entire horizontal area of the cell. Sometimes, however, it is useful to limit consideration to only a portion of a cell (e.g. a mean over the sea-ice area). To indicate this, one of two conventions may be used.

The first convention is a method that can be used for the common case of a single area-type. In this case, the `cell_methods` attribute may include a string of the form "*name: method where type*". Here *name* could, for example, be *area* and *type* may be any of the strings permitted for a variable with a `standard_name` of *area_type*. As an example, if the method were *mean* and the *area_type* were *sea_ice*, then the data would represent a mean over only the sea ice portion of the grid cell. If the data writer expects *type* to be interpreted as one of the standard *area_type* strings, then none of the variables in the netCDF file should be given a name identical to that of the string (because the second convention, described in the next paragraph, takes precedence).

The second convention is the more general. In this case, the `cell_methods` entry is of the form "*name: method where typevar*". Here *typevar* is a string-valued auxiliary coordinate variable or string-valued scalar coordinate variable (see Section 6.1, "Labels") with a `standard_name` of *area_type*. The variable *typevar* contains the name(s) of the selected portion(s) of the grid cell to which the *method* is applied. This convention can accommodate cases in which a method is applied to more than one area type and the result is stored in a single data variable (with a dimension which ranges across the various area types). It provides a convenient way to store output from land surface models, for example, since they deal with many area types within each surface gridbox (e.g., *vegetation*, *bare_ground*, *snow*, etc.).

Example. Mean surface temperature over land and sensible heat flux averaged separately over land and sea.

```
dimensions:
  lat=73;
  lon=96;
  maxlen=20;
  ls=2;
variables:
  float surface_temperature(lat,lon);
    surface_temperature:cell_methods="area: mean where land";
  float surface_upward_sensible_heat_flux(ls,lat,lon);
```

Comment [KET1]: We need a number here ... I think Example 7.6, and this will require incrementing all subsequent numbers in this chapter.

```

    surface_upward_sensible_heat_flux:coordinates="land_sea";
    surface_upward_sensible_heat_flux:cell_methods="area: mean where
land_sea";
    char land_sea(1s,maxlen);
    land_sea:standard_name="area_type";
data:
    land_sea="land", "sea";

```

If the *method* is mean, various ways of calculating the mean can be distinguished in the `cell_methods` attribute with a string of the form "mean where *type1* [over *type2*]"'. Here, *type1* can be any of the possibilities allowed for *typevar* or *type* (as specified in the two paragraphs preceding Example 7.6). The same options apply to *type2*, except it is not allowed to be the name of an auxiliary coordinate variable with a dimension greater than one (ignoring the dimension accommodating the maximum string length). A `cell_methods` attribute with a string of the form "mean where *type1* over *type2*" indicates the mean is calculated by summing over the *type1* portion of the cell and dividing by the area of the *type2* portion. In particular, a `cell_methods` string of the form "mean where all_area_types over *type2*" indicates the mean is calculated by summing over all types of area within the cell and dividing by the area of the *type2* portion. (Note that "all_area_types" is one of the valid strings permitted for a variable with the standard_name `area_type`.) If "over *type2*" is omitted, the mean is calculated by summing over the *type1* portion of the cell and dividing by the area of this portion.

Example. Thickness of sea-ice and snow on sea-ice averaged over sea area.

```

variables:
    float sea_ice_thickness(lat,lon);
    sea_ice_thickness:cell_methods="area: mean where sea_ice over sea";
    sea_ice_thickness:standard_name="sea_ice_thickness";
    sea_ice_thickness:units="m";
    float snow_thickness(lat,lon);
    snow_thickness:cell_methods="area: mean where sea_ice over sea";
    snow_thickness:standard_name="lwe_thickness_of_surface_snow_amount";
    snow_thickness:units="m";

```

In the case of sea-ice thickness, the phrase "where `sea_ice`" could be replaced by "where `all_area_types`" without changing the meaning since the integral of sea-ice thickness over all area types is obviously the same as the integral over the sea-ice area only. In the case of snow thickness, "where `sea_ice`" differs from "where `all_area_types`" because "where `sea_ice`" excludes snow on land from the average.

7.3.4 Cell methods when there are no coordinates

To provide an indication that a particular cell method is relevant to the data without having to provide a precise description of the corresponding cell, the "name" that appears in a "name: method" pair may be an appropriate standard_name (which identifies the dimension) or the string, "area" (rather than the name of a scalar coordinate variable or a dimension with a coordinate variable). This convention cannot be used, however, if the

Comment [KET2]: We need a number here ... I think Example 7.7, and this will require incrementing all subsequent numbers in this chapter.

name of a dimension or scalar coordinate variable is identical to *name*. There are two situations where this convention is useful.

First, it allows one to provide some indication of the method when the cell coordinate range cannot be precisely defined. For example, a climatological mean might be based on any data that exists, and, in general, the data might not be available over the same time periods everywhere. In this case, the time range would not be well defined (because it would vary, depending on location), and it could not be precisely specified through a time dimension's bounds. Nevertheless, useful information can be conveyed by a `cell_methods` entry of "time: mean" (where time, it should be noted, is a valid `standard_name`). (As required by this convention, it is assumed here that for the data referred to by this `cell_methods` attribute, "time" is not a dimension or coordinate variable.)

Second, for a few special dimensions, this convention allows one to indicate (without explicitly defining the coordinates) that the method applies to the domain covering the entire permitted range of those dimensions. This is allowed only for longitude, latitude, and area (indicating a combination of horizontal coordinates). For longitude, the domain is indicated according to this provision by the string "longitude" (rather than the name of a longitude coordinate variable), and this implies that the method applies to all possible longitudes (i.e., from 0E to 360E). For latitude, the string "latitude" is used and implies the method applies to all possible latitudes (i.e., from 90S to 90N). For area, the string "area" is used and implies the method applies to the whole world.

In the second case if, in addition, the data variable has a dimension with a corresponding labeled axis that specifies a geographic region (Section 6.1.1, "Geographic Regions"), the implied range of longitude and latitude is the valid range for each specified region, or in the case of area the domain is the geographic region. For example, there could be a `cell_methods` entry of "longitude: mean", where longitude is *not* the name of a dimension or coordinate variable (but is one of the special cases given above). That would indicate a mean over all longitudes. Note, however, that if in addition the data variable had a scalar coordinate variable with a `standard_name` of `region` and a value of `atlantic_ocean`, it would indicate a mean over longitudes that lie within the Atlantic Ocean, not all longitudes.

We recommend that whenever possible, cell bounds should be supplied by giving the variable a dimension of size one and attaching bounds to the associated coordinate variable.