[THE FOLLOWING TEXT TO REPLACE THAT OF SECTION 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). The portion concerned is constant in time in some cases, but it could be time-varying. Grid cell "portions" that can be considered are only those permitted to be associated with the standard_name "area_type". There are two options for indicating when a quantity represents a portion of a cell.

The first method can be used for the common case that the cell_method applies to 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 is "area: mean where sea_ice", then the data would represent a mean over only the sea ice portion of the grid cell. When this first option is adopted, none of the variables in the netCDF file should be given a name identical to the string that names the area_type. This restriction is imposed so that it will be clear that the metadata should not be interpreted following the second option (described in the next paragraph), which takes precedence.

The second method for indicating that a statistic applies to only a portion of a cell is more general because it can reference multiple area-types. This may be needed when a variable has a dimension that ranges across various area types. 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 method 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 7.6 SHOULD APPEAR HERE]

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 above Example). 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 "`area: mean where` *type1* over *type2*" indicates the mean is calculated by integrating over the *type1* portion of the cell and dividing by the area of the *type2* portion. When "`over` *type2*" is omitted, it is assumed to be the same as *type1*.

When the "where" construct is used, and when "area" is not the only "dimension" to which it applies, the interpretation more generally is that a "weighted" mean is reported. Specifically, the quantity of interest is integrated over the additional dimension(s) with weights proportional to the fraction of "type1" area_type that exists, and then this is divided by the integral over the same dimension(s) of the fraction of "type2" area_type that exists. [Note that certain variables might be undefined if the fraction of the area_type considered is 0; for example the temperature of sea ice is not defined if there is no sea ice. In this case, a time-mean value can still be computed for cells containing some sea ice during at least a portion of the averaging interval because no matter what the value assumed for temperature when sea ice is missing, those values are given zero weight in computing the time-mean.]

Note that "`all_area_types" is one of the valid strings permitted for a variable with the standard_name area_type, so a cell_methods string of the form "area: mean over *type1* where all_area_types"

---

indicates the mean is calculated by integrating over the *type1* portion of the grid cell and dividing by the entire area of the grid cell.

The following three examples illustrate cases when one might want to use "where" or "where … over" in defining the cell_methods:

1. Suppose that in a grid cell the fractional sea ice varies over time, but there is interest in the time-mean surface temperature of the sea ice. The time-samples, each representing a spatially-averaged sea ice temperature can be summed and then divided by the number of samples to obtain an unweighted mean where sea ice exists. This would be indicated with:

    cell_methods = "area: mean where sea_ice time: mean"

2. Suppose there is interest in recording the mean fractional area covered by sea ice and the mean sea ice thickness in such a way that their product would equal the time-mean volume of sea ice in each grid cell. In this case the sea ice area would be reported as an unweighted time-mean, while the mean sea ice thickness would be calculated with time samples weighted by the fractional area of sea ice. Thus, for sea ice thickness:

    cell_methods = "area: time: mean where sea_ice"

3. Suppose the time-mean contributions to total heat flux from different portions of a grid cell (e.g., ice-free and ice-covered) are of interest, and there are reasons to report these in such a way that the total heat flux is the sum of the individual contributions. Then the cell_methods attribute would be defined:

    cell_methods="area: mean where sea_ice over all_area_types time: mean"

In some cases a variable referencing a specific area_type will actually be defined even in the absence of that area_type (i.e., over the entire grid cell). Consider the surface_snow_thickness, which could sensibly be considered to be 0 in the absence of snow. In this case one might in some instances want to report "area: time: mean where snow" (giving a measure of the typical snow depth when snow exists) and in other instances "area: time: mean where snow over all_area_types" (which in this case would be identical to "area: time: mean") or "area: time: mean where snow over land".

[I'M NOT SURE WHETHER EXAMPLE 7.7 SHOULD BE RETAINED, OMITTED OR MODIFIED]

Page 1: [1] Deleted	Karl Taylor	10/26/16 1:44 PM

conventions may be used.

Page 1: [2] Deleted	Karl Taylor	10/26/16 3:07 PM

convention is a method that can be

Page 1: [3] Deleted	Karl Taylor	10/26/16 3:10 PM

were **mean** and the **area_type** were **sea_ice**

Page 1: [4] Deleted	Karl Taylor	10/26/16 3:14 PM

then

| Page 1: [4] Deleted | Karl Taylor | 10/26/16 3:14 PM |

then

| Page 1: [5] Deleted | Karl Taylor | 10/26/16 1:44 PM |

the

| Page 1: [6] Comment [4] | Karl Taylor | 10/26/16 3:40 PM |

I've included "area" here because the succeeding statement only applies when the method is applied to an area.

| Page 1: [7] Comment [5] | Karl Taylor | 10/26/16 3:41 PM |

The original says "summing" which is not right because its really a discrete evaluation of an integral over space, with each values weighted by a (finite) area.

| Page 1: [8] Comment [6] | Karl Taylor | 10/26/16 3:42 PM |

I've moved the sentence that originally appeared here (dealing with "all_area_types") to the beginning of the paragraph after the next.

| Page 1: [9] Comment [7] | Karl Taylor | 10/26/16 3:28 PM |

I've inserted this paragraph because I think we should explicitly indicate what is meant when "where" is used to describe something that varies in time, as well as space.

| Page 1: [10] Deleted | Karl Taylor | 10/26/16 1:44 PM |

within

within

I've addressed here Martin's concern about making it clear how to treat time-means when some samples are missing. Because samples are weighted by fraction of cell occupied by the area_type, it doesn't matter what value is assumed for a variable when the area_type is missing; it won't contribute to the integral.

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. (

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. (

I've modified the example in this sentence because I think the original would rarely apply.