Storing Multi-Interval Samples in EBAS

1 Definition

Multi-Interval samples are defined as samples which have more then one (disjointed) sample interval.

This can occur in case of a measurement schedule where sampling media (e.g. filters) get exposed several times in independent intervals (e.g. only during daytime/nighttime). This results in an analysis sample that is related to multiple intervals over which the measurement value is averaged. See use case 1 (Puy de Dome EC/OC) and use case 2 (Whistler Mountain EC/OC).

Another reason for multi-interval values are multi annual monthly averages (see use case 3, Arosa ozone). Here the average for e.g. all Januarys in several years is known and should be stored.

2 Use cases

In recent years, three cases of multi-interval samples have been submitted to EBAS. The lack of a defined solution for storing them prevented the data from being ingested into the database so far.

2.1 Use Case 1: Puy de Dome EC/OC

Measurement schedule: 2 days in a week. Day (07:00-15:00) and night (22:00-05:00) samples are measured on two different filters, each filter exposed twice, then no measurements for the rest of the week. Weekly filter change.

2.2 Use Case 2: Whistler Mountain EC/OC

Measurement schedule: Usually filters are installed for a week (in exceptional cases longer, up to 12 days). Day and night samples are measured on two different filters, each filter is exposed several times.

The filter switch times are fixed and the switch is automated: 06:30 and 18:30 from Aug. 2008 to Jan. 2015 and 06:00 and 20:00 from Jan. 2015 to present.

The purpose for this schedule is to separate the possible signals of long range transport during nighttime (above free troposphere) from the local and regional influences during daytime (within the free troposphere).

Environment Canada mentioned that EC/OC observations at the Mauna Loa Observatory have been with this kind of schedule since 2014. This means we can expect data submission of another case of this data type.

2.3 Use Case 3: Arosa

Historical ozone data for Arosa are not fully preserved (at least not in digital form). In the given case, only monthly averages for several years (all Januarys 1954-1958, all Februaries 1954-1958, etc) are available.

These data are historically important and should be long term archived and published in EBAS.

3 Possible Solutions

3.1 Alternative 1: As one Dataset

The measurement data are stored as a single time series in EBAS. Each sub-interval (i.e. filter exposure for use case 1 and 2, or single month for use case 3) is one measurement interval in the database. This will result in repeating equal values for some some intervals. The filter analysis (use case 1 and 2) or monthly average (use case 3) represents an average over several intervals as stored in EBAS, thus each of those intervals will have the same measurement value.

Additional Flag 360

A new flag indicates that the actual measurement value belongs to more then this one time interval:

```
360  V  Sampling interval is one of several sampling intervals averaged
```

Flag 360 will be added to all samples.

Comment

Additionally, a comment should be added to all datasets stating the special type of measurements.

Example comment in the **VNAME** section of a variable:

```
total_carbon, ug C/m3, Comment=Daytime and nighttime samples each exposed for two intervals; flag 360 used
```

3.2 Alternative 2: As two Datasets

The measurements are stored as different datasets. For use cases 1 and 2, there will be one dataset for daytime and one for nighttime measurements, use case 3 will be implemented as 12 datasets (one for each month of the year).

The sample intervals in EBAS on the other hand will correspond to the full sampling or averaging interval of the data and not split up. An additional tag `time fraction` (see below) specifies that only a defined fraction of the full interval was actually measured or averaged.

For use cases 1 and 2 this means that the measurement intervals in EBAS represent the intervals between final filter changes (new filters weekly). Therefor the measurement interval equals the overall field time of one filter which is then analysed in the lab to yield one data point.

For use case 3 the measurement intervals in EBAS is the full averaging period (all years), the keyword `time fraction` (see below) will distinguish the datasets and specify which months the average belongs to.

Dataset Characteristic `Time fraction`

The time fraction measured needs to be set in a standardised way for all variables. A new characteristic type could be be defined for this purpose: `Time fraction`

This can be seen in analogy to the keyword `Fraction` used in the [OC/EC template](). There for example, the variable `organic_carbon, ug C/m3, Fraction=OCPyr` indicates that the component measured is still organic carbon, but only a fraction of it (the part that pyrolized during analysis). In analogy we use `time fraction` to define which defined fractions of the sample interval interval is measured.

VNAMES for all variables will then include a `Time fraction` tag specifying the fractions:

```
total_carbon, ug C/m3, Time fraction=Daytime (07:00-15:00)
total_carbon, ug C/m3, Time fraction=Nighttime (22:00-05:00)
```

Additional flag 385

For emphasising the reduced data coverage within the intervals, the definition of a new flag is useful:

```
385  V  Data completeness less then 100% because of time fractions, see variable m
etadata Time fraction
```

Flag 385 will be added to all samples.

4 Advantages and Disadvantages of the Proposed Alternatives

4.1 Correctness according to data model

Alternative 1 violates the data model by relating the mean value over several time intervals to each one of the time interval. Taking this idea to the extremes it is comparable storing an annual mean value as 12 monthly averages in the database. Flag 360 is somehow an attempt to fix the data model, but there is no formal way to find the real resolution of the data in the metadata. The metadata about the real nature of the measurement intervals can only be implemented as comment (especially in use case 2, where the number of filter exposures is not constant).

Alternative 2 also violates the data model as we know it, this time relates the mean value to a bigger interval. But the `time fraction` mechanism is a formal and strict way to fix the data model. The `fraction` mechanism is already used and `time fraction` is a consistent extension.

4.2 Resolution code

With alternative 1, the resolution code will appear finer then the real resolution. The degree of distortion varies with the number of intervals per sample and possible uncovered intervals (like in use case 1).

In the three uses cases we see a wrong resolution code by a factor of between 11 and 56:

- Use case 1 has an apparent resolution code `15h` vs. real `1w`.
- Use case 2 has an apparent resolution code `12h` vs. real `1w`.
- Use case 3 has an apparent resolution code `1mo` vs. real `56mo`.

This means, data users might use the data with the assumption that they have a much finer time resolution.

4.3 Further Averaging

Calculating higher averages works easier for alternative 1. For example the day/night samples in use cases 1 and 2 would be naturally averaged to a correct overall monthly mean by existing ebas statistic routines. There are no existing routines that would work for alternative 2.

On the other hand, averaging along the same time fractions can be done with standard EBAS routines, while there are no existing routines for this for alternative 1. E.g. calculating the night time monthly mean in use cases 1 and 2 easy for alternative 2.

4.4 Intuitiveness and Misunderstandings

In order to avoid misunderstandings (and thus wrong use of data), the implementation should be as simple and intuitive as possible.

Unfortunately, neither of the two alternatives seem to be the one that everybody thinks to be the natural solution. Not only in internal discussions, also the data submitters seem to be split in two fractions.

Use case 1 and 3 have been submitted in a form equivalent to alternative 1, while use case 2 have been submitted in a form equivalent to alternative 2. There is at least also on the data reporter side not clear concept how this kind of data should be expressed.

4.5 Measurement schedule changes

In case of changes of measurement schedule (e.g. use case 2 changed day/night switch times in 2015: 06:30 / 18:30 changed to 06:00 / 20:00)

With Alternative 1, changes in the measurement schedule are visible in the data and do not need to be part of the metadata.

With

Use case 2: changed 6:30am – 6:30 pm from Aug. 2008 to Jan 2015

6:00am - 8:00pm from Jan 2015 to present

4.6 Summary:

Alternative 1

- The variable appears to be of a much higher time resolution. Impossible to see the real averaging interval
- One variable contains all measurements, `time fraction` needed
- Easy to average "overall", difficult to average along time fractions (e.g. night time monthly mean)

Alternative 2

- The variable has the right resolution code
- The variable has an additional time restriction. E.g. its not only `TC` anymore, but `TC measured during night times`
- Easy to average along time fractions (e.g. night time monthly mean), difficult to average "overall"

5 Conclusions

needed...

Appendix: Examples for the Known Use Cases for Each of the Two Alternatives

Below one can find example Nasa Ames snippets to illustrate the effects of alternatives 1 and 2 for the three use cases.

- Puy de Dome (EC/OC day/night filters multiply exposed)
- Whistler Mountain (EC/OC day/night filters multiply exposed)

- Arosa ozone multi year monthly averages (all Januarys 1954-1958, all Februaries 1954-1958, etc)

Use Case 1: Puy de Dome

Example for Alternative 1:

This is just a small part of the Nasa Ames file, the full file is attached as `Puy_Alt1.nas`.

VNAMES:

```
total_carbon, ug C/m3, Comment=Daytime and nighttime samples each exposed for two
intervals; flag 360 used
numflag, no unit
```

Resolution code:

```
Resolution code: 15h
```

(not really correct, but appears to be 15h)

DATA:

```
start_time end_time  TC TC_numflag
  0.292000 0.625000 2.24 0.360
  0.917000 1.208000 2.86 0.360
  1.292000 1.625000 2.24 0.360
  1.917000 2.208000 2.86 0.360
  7.292000 7.625000 0.24 0.360
  7.917000 8.208000 0.61 0.360
  8.292000 8.625000 0.24 0.360
  8.917000 9.208000 0.61 0.360
```

Example for Alternative 2:

This is just a small part of the Nasa Ames file, the full file is attached as `Puy_Alt2.nas`.

VNAMES:

```
total_carbon, ug C/m3, Time fraction=Daytime (07:00-15:00)
total_carbon, ug C/m3, Time fraction=Nighttime (22:00-05:00)
numflag, no unit
```

Resolution code:

```
Resolution code: 1w
```

DATA:

```
start_time end_time   TCday   TCnight   TC_numflag
  0.292000 2.208000   2.24      2.86    0.385
  7.292000 9.208000   0.24      0.61    0.385
```

Use Case 2: Whistler Mountain EC/OC

The schedule is similar to the Puy de Dome case, but due to the fact that filters are exposed more often, the differences between alternative 1 and 2 are more extreme (Up to 24 samples intervals in ebas for one filter change).

Example for Alternative 1:

This is just a small part of the Nasa Ames file, the full file is attached as `WSL_Alt1.nas`.

VNAMES:

```
total_carbon, ug C/m3, Comment=Daytime and nighttime samples each exposed for seve
ral intervals; flag 360 used
numflag, no unit
```

Resolution code:

```
Resolution code: 12h
```

(not really correct (1w in reality), but appears to be 12h)

DATA:

```
start_time  end_time      TC   flag
 0.625000   0.750000  11.100  0.360
 0.750000   1.250000   8.380  0.360
 1.250000   1.750000  11.100  0.360
 1.750000   2.250000   8.380  0.360
 2.250000   2.750000  11.100  0.360
 2.750000   3.250000   8.380  0.360
 3.250000   3.750000  11.100  0.360
 3.750000   4.250000   8.380  0.360
 4.250000   4.750000  11.100  0.360
 4.750000   5.250000   8.380  0.360
 5.250000   5.750000  11.100  0.360
 5.750000   6.250000   8.380  0.360
 6.250000   6.750000  11.100  0.360
 6.750000   7.250000   8.380  0.360
```

```
 7.250000   7.750000  11.100  0.360
 7.750000   8.250000   8.380  0.360
 8.250000   8.750000  11.100  0.360
 8.750000   9.250000   8.380  0.360
 9.250000   9.750000  11.100  0.360
 9.750000  10.250000   8.380  0.360
10.250000  10.750000  11.100  0.360
10.750000  11.250000   8.380  0.360
11.250000  11.750000  11.100  0.360
11.750000  12.250000   8.380  0.360
12.250000  12.750000  11.100  0.360
12.750000  12.827778   8.380  0.360
12.827778  13.250000   1.540  0.360
13.250000  13.750000   1.960  0.360
13.750000  14.250000   1.540  0.360
14.250000  14.750000   1.960  0.360
14.750000  15.250000   1.540  0.360
15.250000  15.750000   1.960  0.360
15.750000  16.250000   1.540  0.360
16.250000  16.750000   1.960  0.360
16.750000  17.250000   1.540  0.360
17.250000  17.750000   1.960  0.360
17.750000  18.250000   1.540  0.360
18.250000  18.750000   1.960  0.360
18.750000  19.250000   1.540  0.360
19.250000  19.750000   1.960  0.360
19.750000  20.250000   1.540  0.360
20.250000  20.750000   1.960  0.360
20.750000  20.878472   1.540  0.360
20.878472  21.250000   2.880  0.360
21.250000  21.750000   4.310  0.360
21.750000  22.250000   2.880  0.360
22.250000  22.750000   4.310  0.360
22.750000  23.250000   2.880  0.360
23.250000  23.750000   4.310  0.360
23.750000  24.250000   2.880  0.360
24.250000  24.750000   4.310  0.360
24.750000  25.250000   2.880  0.360
25.250000  25.750000   4.310  0.360
25.750000  26.250000   2.880  0.360
26.250000  26.750000   4.310  0.360
26.750000  27.250000   2.880  0.360
27.250000  27.729167   4.310  0.360
```

Example for Alternative 2:

This is just a small part of the Nasa Ames file, the full file is attached as `WSL_Alt2.nas`.

VNAMES:

```
total_carbon, ug C/m3, Time fraction=Daytime (06:30-18:30 UTC-8)
total_carbon, ug C/m3, Time fraction=Nighttime (18:30-06:30 UTC-8)
numflag, no unit
```

Resolution code:

```
Resolution code: 1w
```

DATA:

```
start_time   end_time    TCday  TCnight   flag
 0.625000   12.827778   11.100    8.380   0.654385
12.827778   20.878472    1.540    1.960   0.654385
20.878472   27.729167    4.310    2.880   0.385000
```

Use Case 3: Arosa

Example for Alternative 1:

This is just a small part of the Nasa Ames file, the full file is attached as `ARO_Alt1.nas`.

VNAMES:

```
ozone, ug/m3, Comment=Multi-year monthly averages; flag 360 used
numflag, no unit
```

Resolution code:

```
Resolution code: 1mo
```

(not really correct (56mo in reality), but appears to be 1mo)

DATA (May 1954-Dec 1958):

```
 start_time     end_time   O3  flag
 120.000000   151.000000   30  0.360
 151.000000   181.000000   28  0.360
 181.000000   212.000000   22  0.360
 212.000000   243.000000   26  0.360
 243.000000   273.000000   20  0.360
 273.000000   304.000000   21  0.360
 304.000000   334.000000   15  0.360
 334.000000   365.000000   12  0.360
```

```
  365.000000   396.000000   13  0.360
  396.000000   424.000000   19  0.360
  424.000000   455.000000   21  0.360
  455.000000   485.000000   26  0.360
  485.000000   516.000000   30  0.360
  516.000000   546.000000   28  0.360
  546.000000   577.000000   22  0.360
  577.000000   608.000000   26  0.360
  608.000000   638.000000   20  0.360
  638.000000   669.000000   21  0.360
  669.000000   699.000000   15  0.360
  699.000000   730.000000   12  0.360
  730.000000   761.000000   13  0.360
  761.000000   790.000000   19  0.360
  790.000000   821.000000   21  0.360
  821.000000   851.000000   26  0.360
  851.000000   882.000000   30  0.360
  882.000000   912.000000   28  0.360
  912.000000   943.000000   22  0.360
  943.000000   974.000000   26  0.360
  974.000000  1004.000000   20  0.360
 1004.000000  1035.000000   21  0.360
 1035.000000  1065.000000   15  0.360
 1065.000000  1096.000000   12  0.360
 1096.000000  1127.000000   13  0.360
 1127.000000  1155.000000   19  0.360
 1155.000000  1186.000000   21  0.360
 1186.000000  1216.000000   26  0.360
 1216.000000  1247.000000   30  0.360
 1247.000000  1277.000000   28  0.360
 1277.000000  1308.000000   22  0.360
 1308.000000  1339.000000   26  0.360
 1339.000000  1369.000000   20  0.360
 1369.000000  1400.000000   21  0.360
 1400.000000  1430.000000   15  0.360
 1430.000000  1461.000000   12  0.360
 1461.000000  1492.000000   13  0.360
 1492.000000  1520.000000   19  0.360
 1520.000000  1551.000000   21  0.360
 1551.000000  1581.000000   26  0.360
 1581.000000  1612.000000   30  0.360
 1612.000000  1642.000000   28  0.360
 1642.000000  1673.000000   22  0.360
 1673.000000  1704.000000   26  0.360
 1704.000000  1734.000000   20  0.360
 1734.000000  1765.000000   21  0.360
 1765.000000  1795.000000   15  0.360
 1795.000000  1826.000000   12  0.360
```

Example for Alternative 2:

This is just a small part of the Nasa Ames file, the full file is attached as `ARO_Alt2.nas`.

VNAMES:

```
ozone, ug/m3, Time fraction=only months January
ozone, ug/m3, Time fraction=only months February
ozone, ug/m3, Time fraction=only months March
ozone, ug/m3, Time fraction=only months April
ozone, ug/m3, Time fraction=only months May
ozone, ug/m3, Time fraction=only months June
ozone, ug/m3, Time fraction=only months July
ozone, ug/m3, Time fraction=only months August
ozone, ug/m3, Time fraction=only months September
ozone, ug/m3, Time fraction=only months October
ozone, ug/m3, Time fraction=only months November
ozone, ug/m3, Time fraction=only months December
numflag, no unit
```

Resolution code:

```
Resolution code: 56mo
```

DATA (May 1954-Dec 1958):

```
 start_time     end_time  O3Jan O3Feb O3Mar O3Apr O3May O3Jun O3Jul O3Aug O3Sep O3
Oct O3Nov O3Dez flag
 120.000000  1826.000000     13    19    21    26    30    28    22    26    20
 21    15  12   0.385
```