

## Sub-orbital Order Tool (SOOT) Merge Functionality v1

This document describes SOOT's Merge functionality in terms of input, output, and algorithm.

### Introduction

The Sub-Orbital Order Tool (SOOT) is designed to provide merge capabilities through its [Power User Interface](#). This means that measurement data, from a common platform and date, with different sampling time stamps and/or integration times can be aligned to a common sampling time base. This enables easier comparison between datasets. The merge process is implemented using weighted time averages. For details on the weighting algorithm, see Annex 1.

For files to be merged into a single output file, input files must have the same campaign, platform, start date, and launch number (if applicable). If the input files do not share these attributes, separate output merge files are generated for each unique platform, campaign, start date, or launch number. A user can choose a target time base, either a continuous time with a constant interval, e.g., 10 s or 60 s, or can choose to specify a specific dataID to use as the source of data intervals for the merge file. All merge output files are normalized to have three time stamp variables: Time\_Start, Time\_Mid, and Time\_Stop.

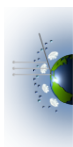
Once files are selected, variables associated with that dataID can be filtered to be included or not in the merge output file. Not all variables have to be merged from a single dataID. Currently, variable types, e.g., scalar, vector, or data flag, are obtained from the SOOT Metadata Database and are not required inputs by the user.

When doing a regular (i.e., non-Merge) download, the user downloads a Zip archive that contains the selected files. In contrast, when doing a Merge download, the user downloads a Zip archive that contains the following contents:

- A sub-directory that contains the merge file itself along with the input files that were merged together to create the merge file.
- A sub-directory that contains the non-mergeable files if applicable
- A README file that describes the contents of the Zip archive.

### Algorithmic Details

The variables in the input files are processed primarily as described in the methodology contained in the SSD-AC online merge tool documentation included as Annex I to this document. The only difference between the SOOT Merge Service and the SSD-AC online merge tool is how wind velocities are processed. When the destination data interval is larger than the source data interval, the source wind velocities must be aggregated. In the source files, wind speed is typically represented by two variables: "Wind\_Speed" (the magnitude of the wind velocity) and "Wind\_Direction" (the direction, in degrees, of the wind velocity). The two main ways of aggregating a set of wind velocities are:



## Atmospheric Science Data Center (ASDC) Sub-Orbital Order Tool's Merge Functionality

1. To compute the vector average of the directions (as unit vectors) and compute the arithmetic average of the speeds.
2. To compute the vector average of the directions (as unit vectors) and aggregate the speeds as the magnitude of the vector average of the input velocities.

The SSD-AC online merge tool uses the first technique to aggregate wind velocities. SOOT provides both aggregation techniques and allows a user to decide which best suits the analysis. The wind speed variable that corresponds to technique #1 has “\_arith” appended to the resulting variable name, and the one corresponding to technique #2 has “\_vec” appended. For example, if the wind velocity for a particular file is represented by two variables—“Wind\_Speed” and “Wind\_direction”—then the resulting merge file will have three variables: “Wind\_Speed\_arith,” “Wind\_Speed\_vec,” and “Wind\_direction.”

### Limitations

- When an input data file corresponds to more than one input housekeeping file (i.e., separate launch times), the SOOT merge service will generate multiple output merge files (i.e., one per launch).
- If an input data file has multiple launch times for a single day but the merge time base dataID only has a single date, the SOOT merge service will output a single merge file corresponding to the time base dataID selected.
- The SOOT merge service assumes that all input files are synchronized in sampling time. If this assumption is violated, then it is the user's responsibility to correctly interpret the merge outputs.
- Currently, the SOOT merge service only computes averages, it does not compute standard deviation.

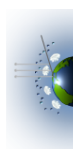
### Glossary

**DataID:** Short description of measured parameter/species, instrument, or model as specified by [ICARTT V2.0](#).

**Housekeeping file:** A file that contains meteorological and navigational parameters.

**Merge file:** An ICARTT file that is produced by merging a set of source files. It contains data from the input source files.

**Source file:** An ICARTT file that is an input file to Merge.



## ANNEX I. SSD-AC Online Merge Tool

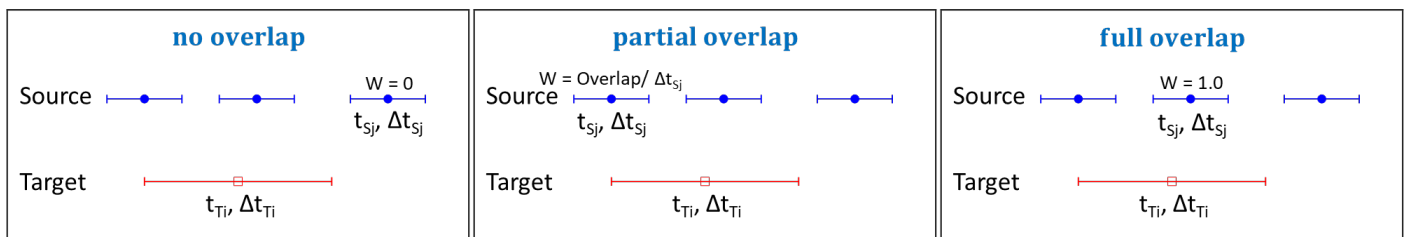
*“SSD-AC Online Merge Tool” was originally published on the Airborne Science Data for Atmospheric Composition website created by NASA Langley Research Center in cooperation with Earth Science Enterprise (ESE) and NASA headquarters.*

*For further details contact Dr. Gao Chen.*

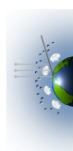
The measurement data from different instruments often have different sampling time stamps and/or integration times. The SSD-AC Online Merge Tool is designed to put various in-situ measurement data from a common platform to a common sampling time base. This is implemented using weighted time averages of different source data, i.e., individual measurement data, onto to the target (common) time base. The target time base can be a continuous time with a constant interval or the time base of an individual measurement. The weighting factor calculation is based on the overlap between a source sampling time interval and the target time interval or merge time interval (see figure below). The Online Merge Tool can also evaluate the weighted standard deviation of individual measurements within each target time interval as well as the number of source data points. This tool can also handle vector averages and propagate precision uncertainty using weighted quadrature average method. The online merge tool requires data to be reported in ICARTT 1.1 or 2.0 format as it interprets the missing/LOD data codes embedded in the headers. The users need to specify the variable type so the tool will handle the data properly. This document outlines how the weighting factor is assessed and how different types of the variables are merged.

### 1. Weighting Factor

The weighting factor is calculated for each target time interval and source data point involved in the merge. The value of the weighting factor for a source data point is determined by the overlap between the source interval and the target interval, see Figure below.



Here the Source sampling time base refers to the center ( $t_{sj}$ ) and period ( $\Delta t_{sj}$ ) of the sampling time for a given source data point. The target (merge) time base is defined by the center and interval, i.e.,  $t_{Ti}$  and  $\Delta t_{Ti}$  respectively.



The weighting factor for the  $j^{th}$  source data point and the  $i^{th}$  target time interval,  $w_{j,i}$ , is calculated as follows:

$$w_{j,i} = \begin{cases} 0 & \text{if } |t_{Ti} - t_{Sj}| \geq \frac{(\Delta t_{Ti} + \Delta t_{Sj})}{2} \\ \frac{(\Delta t_{Ti} + \Delta t_{Sj})}{2} - |t_{Ti} - t_{Sj}|}{\Delta t_{Sj}} & \left| \frac{\Delta t_{Ti} - \Delta t_{Sj}}{2} \right| < |t_{Ti} - t_{Sj}| < \frac{(\Delta t_{Ti} + \Delta t_{Sj})}{2} \\ 1 & \text{if } |t_{Ti} - t_{Sj}| \leq \left| \frac{\Delta t_{Ti} - \Delta t_{Sj}}{2} \right| \end{cases} \quad (1)$$

where:  $t_{Sj}$  = Source interval midpoint;  $\Delta t_{Sj}$  = Source interval duration;  $t_{Ti}$  = Target interval midpoint;  $\Delta t_{Ti}$  = Target interval duration;  $i$  = Target interval; and  $j$  = Source interval.

It is noted that equation (1) applies regardless if  $\Delta t_{Ti}$  is larger than  $\Delta t_{Sj}$  or not. This reflects that equation (2) is symmetric with respect to the target variable and source variables.

The total weighting factor for a target interval,  $w_i$ , is calculated as follows:

$$w_i = \sum_{j=1}^n w_{j,i} \quad (2)$$

where  $n$  is the number of source data points.

## 2. Scalars

$$x_{Ti} = \frac{\sum_{j=1}^n w_{j,i} x_{Sj}}{w_i} \quad (3)$$

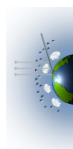
Equation 3: Weighted average of scalar variable data for the  $i^{th}$  target interval,  $x_{Ti}$ , where  $x_{Sj}$  is the Source data value for the  $j^{th}$  point.

$$S_{Ti} = \sqrt{\frac{\left( \left( \sum_{j=1}^n w_{j,i} x_{Sj}^2 \right) - w_i x_{Ti}^2 \right) w_i}{w_i \times w_i - \sum_{j=1}^n w_{j,i} \times w_{j,i}}} \quad (4)$$

Equation 4: Weighted standard deviation (variation/dispersion from the average) for scalar variables for the  $i^{th}$  target interval.

$$SigPre_{Ti} = \frac{\sqrt{\sum_{j=1}^n w_{j,i} \times SigPre_{Sj}^2}}{w_i} \quad (5)$$

Equation 5: Propagated precision using weighted quadrature average, where  $SigPre_{Sj}$  is the  $j^{th}$  point of a source data precision variable.



### 3. Vectors

Vector variable,  $v$ , averaging is done using  $x$  and  $y$  components.

$$v_{x,sj} = \cos\left(v_{sj} \times \frac{\pi}{180}\right) \quad (6a)$$

$$v_{y,sj} = \sin\left(v_{sj} \times \frac{\pi}{180}\right) \quad (6b)$$

Equation 6a and 6b:  $x$  and  $y$  components of the  $j^{\text{th}}$  vector data point. Vector variables are assumed to be in degrees from 0 to 360, hence converting to radians.

$$v_{Ti} = \begin{cases} \text{atan2}(v_{y,i}, v_{x,i}) \frac{180}{\pi} & \text{if } \text{atan2}(v_{y,i}, v_{x,i}) \geq 0 \\ \text{atan2}(v_{y,i}, v_{x,i}) \frac{180}{\pi} + 360 & \text{if } \text{atan2}(v_{y,i}, v_{x,i}) < 0 \end{cases} \quad (7)$$

$$\text{Where: } v_{y,i} = \frac{\sum_{j=1}^n w_{j,i} \times v_{y,sj}}{w_i}, \quad v_{x,i} = \frac{\sum_{j=1}^n w_{j,i} \times v_{x,sj}}{w_i}$$

Equation 7: The merged vector variable,  $v_{Ti}$ , converted back to degrees between 0 and 360. It is noted the atan2 function may behave differently in different programming languages. For example,  $v_{y,i}$  and  $v_{x,i}$  should be switched, i.e.,  $\text{atan2}(v_{x,i}, v_{y,i})$ , if equation (7) were applied in an Excel spreadsheet.

$$S_{x,i} = \sqrt{\frac{(\sum_{j=1}^n w_{j,i} \times v_{x,sj}^2 - w_i v_{x,i}^2) w_i}{w_i^2 - \sum_{j=1}^n w_{j,i}^2}} \quad (8a)$$

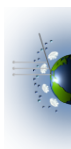
$$S_{y,i} = \sqrt{\frac{(\sum_{j=1}^n w_{j,i} \times v_{y,sj}^2 - w_i v_{y,i}^2) w_i}{w_i^2 - \sum_{j=1}^n w_{j,i}^2}} \quad (8b)$$

Equation 8:  $x$  and  $y$  components of the standard deviation for the  $i^{\text{th}}$  target interval.

$$\delta f = \frac{x \delta y + y \delta x}{x^2 + y^2} \quad (9)$$

$$S_{Ti} = \frac{\left| \frac{v_{y,i}}{v_{x,i}} \right| \times \sqrt{\left( \frac{S_{x,i}}{v_{x,i}} \right)^2 + \left( \frac{S_{y,i}}{v_{y,i}} \right)^2}}{1 + \left( \frac{v_{y,i}}{v_{x,i}} \right)^2} \times \frac{180}{\pi} \quad (10)$$

Equations 9: Error propagation of the arctangent used to recombine the  $x$  and  $y$  components of the standard deviation



Equation 10: The merged standard deviation

#### 4. Wind

$$U_{Sj} = -1 \times spd_{Sj} \times \sin\left(\theta_{Sj} \times \frac{\pi}{180}\right) \quad (11a)$$

$$V_{Sj} = -1 \times spd_{Sj} \times \cos\left(\theta_{Sj} \times \frac{\pi}{180}\right) \quad (11b)$$

Equation 11a and 11b: Converting the  $j^{th}$  point of wind speed ( $spd_{Sj}$ ) and direction ( $\theta_{Sj}$ ) to the east-west component,  $u_{Sj}$ , and north-south component  $v_{Sj}$ . It is assumed that wind direction is in degrees. The weighted average for vector components ( $U_i$  and  $V_i$ ) are calculated via equation 3. The vector average of the wind speed can be constructed through quadrature sum of the u and v components.

$$\theta_i = 180 + \frac{180}{\pi} \times \text{atan2}(U_i, V_i) \quad (12a)$$

$$\text{Where: } U_i = \frac{\sum_{j=1}^n w_{j,i} U_{Sj}}{w_i}, \quad V_i = \frac{\sum_{j=1}^n w_{j,i} V_{Sj}}{w_i} \quad (12b)$$

Equation 12: Weighted average wind direction value for the  $i^{th}$  target interval. Equation (12a) is based on C++ and Python programming languages. One may switch  $U_i$  and  $V_i$  in the equation for different programming languages.

$$\theta_{v,i} = \frac{\sum_{j=1}^n w_{i,j} \cos\left(\theta_{Sj} \times \frac{\pi}{180}\right)}{w_i} \quad (13a)$$

$$\theta_{u,i} = \frac{\sum_{j=1}^n w_{i,j} \sin\left(\theta_{Sj} \times \frac{\pi}{180}\right)}{w_i} \quad (13b)$$

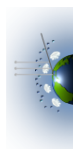
Equation 13: Weighted merge of the angle components.

$$S_{Ti} = \sin^{-1} \left[ 1 + \left( \frac{2}{\sqrt{3}} - 1 \right) \epsilon^3 \right] \times \frac{180}{\pi} \quad (14)$$

$$\text{where } \epsilon = \sqrt{1 - (\theta_{u,i}^2 + \theta_{v,i}^2)}$$

Equation 14: Yamartino method<sup>1</sup> for wind direction standard deviation.

<sup>1</sup> Yamartino, R.J. (1984). "A Comparison of Several "Single-Pass" Estimators of the Standard Deviation of Wind Direction". Journal of Climate and Applied Meteorology 23 (9): 1362-1366.



## 5. Data flag handling

This feature is designed to handle data flags which designate either sampling locations, sampling events (e.g., fire plume), or sampling conditions (e.g., cloud conditions). These flags should not be averaged as it may yield misleading results. The online merge tool will give a binary output to indicate if there is a change in the data flag values within the target (merge) interval. The goal is to alert users to further investigate the data flag values in this case. The value of flag variables are restricted to be from 0 to 1E8.

$$xFlag_{Ti} = \begin{cases} xFlag_{Sj} & \text{if } xFlag_{Sj} = \text{contant when } w_{j,i} > 0 \\ -6666 & \text{if } xFlag_{Sj} \neq \text{contant when } w_{j,i} > 0 \end{cases} \quad (15)$$

Equation 15: Data flag handling for flag variable, xFlag, where xflag<sub>s<sub>j</sub></sub> represent the *j*<sup>th</sup> point of the source data value and xFlag<sub>T<sub>i</sub></sub> denotes the output for the *i*<sup>th</sup> target interval. “-6666” is the indicator code for flag value changes.

## 6. Data code handling

The online merge tool handles ICARTT format missing data codes and limit of detection (LOD) codes.

### 6.1 Missing data code handling

The summation skips over missing data entries when it encounters one within a target interval. For example, equation (2) and (3), in practice, are modified as below:

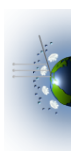
$$w_i = \sum_{j=1}^n w_{j,i} \quad \text{if } x_j \neq -9999 \quad (2')$$

$$x_i = \frac{\sum_{j=1}^n w_{j,i} x_j}{w_i} \quad \text{if } x_j \neq -9999 \quad (3')$$

where “-9999” is the missing data code. If no non-missing source data is encountered within a target interval, the output for that target interval is the missing data code.

### 6.2 Limit of detection code handling

The online merge tool outputs the same LOD code if one LOD code is encountered within a target interval.



$$x_i = \begin{cases} -8888 & \text{if } x_j = -8888 \text{ when } w_{j,i} > 0 \\ -7777 & \text{if } x_j = -7777 \text{ when } w_{j,i} > 0 \\ \frac{\sum_{j=1}^n w_{j,i} x_j}{w_i} & \text{if } x_j \neq -8888 \text{ and } x_j \neq -7777 \text{ when } w_{j,i} > 0 \end{cases} \quad (16)$$

where “-8888” represents the lower limit of detection and -7777 denotes the upper limit of detection. Cases where both upper and lower limit of detection codes are encountered in the same target interval are considered highly unlikely for commonly used merge time intervals and therefore treated in the same way as encountering just the lower limit of detection code (-8888) within the target interval.

