



# **System for Analysis of Wind Collocations (SAWC):**

## **A Novel Archive and Collocation Software Application for the Intercomparison of Winds from Multiple Observing Platforms**

### **User Manual**

Version 1.2.0

September 2023

**Prepared for:**

U.S. Department of Commerce  
National Oceanic and Atmospheric Administration (NOAA)  
National Environmental Satellite, Data, and Information Service (NESDIS)

**Prepared by:**

Katherine E. Lukens, *U. Maryland/ESSIC/CISS, NOAA/NESDIS/STAR*  
Kevin Garrett, *NOAA/NWS/OSTI*  
Kayo Ide, *U. Maryland*  
David Santek, *U. Wisconsin-Madison/CIMSS*  
Brett Hoover, *NOAA/NWS/NCEP/EMC, Lynker Technologies*  
David Huber, *NOAA/NWS/NCEP/EMC, Redline Performance Solutions, LLC*  
Ross N. Hoffman, *U. Maryland/ESSIC/CISS, NOAA/NESDIS/STAR*  
Hui Liu, *U. Maryland/ESSIC/CISS, NOAA/NESDIS/STAR*

## Version Description Record

DOCUMENT TITLE: <b><i>System for Analysis of Wind Collocations (SAWC): A Novel Archive and Collocation Software Application for the Intercomparison of Winds from Multiple Observing Platforms User Manual</i></b>		
Revision	Date	Description of Change
1.0.0	April 15, 2023	Initial Release of User Manual
1.0.1	April 21, 2023	<b>Bug Fixes:</b> Corrected Plotting Tool's use of input path for archived index files. Corrected AMV type maps to show observations at all longitudes.
1.1.0	July 15, 2023	<b>Updates:</b> Updated application installation section. Converted HLOS_error to m/s in Aeolus Rayleigh-clear NetCDF files ( <i>original</i> (non-reprocessed) only: Sep 2018 - April 2020). Added missing observations to Aircraft NetCDF data files. Optimized Collocation and Plotting Tools to improve performance. Added a Height/Pressure vs Longitude plot type to the Plotting Tool.  <b>Bug Fixes:</b> Various bug fixes to the Plotting Tool.
1.2.0	September 22, 2023	<b>Update:</b> SAWC has officially moved to a new home! As such, the links to access SAWC have been updated, as have the supplemental information filenames and paths to various components throughout this document.  <b>Bug Fix:</b> Corrected density scatterplot function plotting issue for Plotting Tool.

## Preface

This document comprises the National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite, Data, and Information Service (NESDIS) publication of the System for Analysis of Wind Collocations (SAWC) *User Manual (UM)*.

This User Manual provides important information for wind observation users. It gives an overview of the System for Analysis of Wind Collocations (SAWC), including the wind observation archive, its datasets, and variables available. It also provides step-by-step instructions on the installation and operation of the collocation application developed for use in conjunction with the archived wind datasets.

The document will be updated as required.

# Table of Contents

<b>How to Use this Manual.....</b>	<b>1</b>
<b>1. Introduction.....</b>	<b>2</b>
<b>2. Overview of SAWC.....</b>	<b>3</b>
2.1 SAWC Contents.....	3
2.1.1 Wind Datasets.....	3
2.1.2 Collocation Application and Index Files.....	4
2.2 Directory Paths.....	4
2.2.1 Paths to Source Data.....	6
2.2.1.1 Aeolus Winds.....	6
2.2.1.2 All Other Winds.....	7
2.2.2 Paths to Collocation Application and Index Files.....	7
2.3 File Name Formats.....	8
2.3.1 Aeolus File Names.....	8
2.3.2 Non-Aeolus File Names.....	9
2.3.3 Collocation Index File Names.....	9
2.4 Supplemental Information.....	10
<b>3. Dataset Descriptions.....</b>	<b>11</b>
3.1 Aeolus.....	11
3.2 Aircraft.....	14
3.3 AMVs.....	15
3.4 Loon.....	16
3.5 Sondes.....	17
<b>4. Collocation Application.....</b>	<b>18</b>
4.1 Overview.....	18
4.1.1 Data Acquisition.....	19
4.1.2 Collocation of Winds.....	19
4.1.3 Analysis and Visualization.....	22
4.2 Installation.....	35
4.3 Running the Application.....	37
4.3.1 Collocation Tool.....	37
4.3.1.1 Order of Processing Steps.....	37
4.3.1.2 Steps for Running the Collocation Tool.....	38
4.3.1.3 Advanced Modifications: Step-by-step Examples.....	39
4.3.2 Plotting Tool.....	42
4.3.2.1 Order of Processing Steps.....	42
4.3.2.2 Steps for Running the Plotting Tool.....	43
4.3.2.3 Advanced Modifications: Step-by-step Examples.....	44
<b>Appendices.....</b>	<b>46</b>
A. Acronyms and Abbreviations.....	46
B. Tips, Tricks, and Other Useful Information.....	47
<b>Publications List.....</b>	<b>48</b>
<b>References.....</b>	<b>49</b>

---

## List of Figures

<b>Figure 1:</b> Spatial coverage of Aeolus observations for a 24-h period (2019-09-01). Colored dots denote the observation locations.....	<b>11</b>
<b>Figure 2:</b> Spatial coverage of aircraft observations for a 24-h period (2019-09-01). Colored dots denote the observation locations.....	<b>15</b>
<b>Figure 3:</b> AMV spatial coverage for a 24-h period (2019-09-01). Colored dots denote individual satellite-derived AMV observations with quality indicator (QI) > 80%. Each item in the legend lists the satellite name (top), the number of AMVs from the satellite (bottom left), and the percentage of all AMVs the satellite's count represents (bottom right in parentheses).....	<b>15</b>
<b>Figure 4:</b> Spatial coverage of Loon stratospheric balloon observations for Sept-Dec 2019. Colored dots denote the observation locations.....	<b>16</b>
<b>Figure 5:</b> Spatial coverage of sonde observations for a 24-h period (2019-09-01). Colored dots denote the observation locations.....	<b>17</b>
<b>Figure 6:</b> Flow chart illustrating SAWC's end-to-end process. Note that the user interfaces with the collocation and plotting tools in Steps (2) and (3), respectively.....	<b>19</b>
<b>Figure 7:</b> Flow chart illustrating the order of processing steps for the collocation tool.....	<b>37</b>
<b>Figure 8:</b> Flow chart illustrating the order of processing steps for the plotting tool.....	<b>42</b>

## List of Tables

<b>Table 1:</b> A list of all SAWC components, including source datasets, index files, and the collocation software application. Additionally, the temporal coverage and key variables available per dataset are listed.....	<b>5</b>
<b>Table 2:</b> General archive directory abbreviations, paths, and descriptions.....	<b>5</b>
<b>Table 3:</b> Paths to all SAWC components (source datasets, index files, and collocation application). Additionally, key variables that are available per dataset are listed.....	<b>6</b>
<b>Table 4:</b> Index file name abbreviations, available options per abbreviation, and corresponding descriptions.....	<b>9</b>
<b>Table 5:</b> List of L2B processor versions as indicated by Baseline number. The operational start date of each Baseline and relevant details are also provided.....	<b>12</b>
<b>Table 6:</b> List of input parameters the user can modify in the main bash script for the collocation tool (MAINSRIPT_for__Collocation.bash).....	<b>20</b>
<b>Table 7:</b> Default QC parameters currently available in SAWC as recommended by the data producers and broader wind community. QCs are applied to the listed datasets only if the user selects this option. Reject values are listed in the table. For Aeolus, p is pressure in hPa, is the L2B HLOS wind uncertainty in m/s, z is the height of the vertical range bins within which measurements are accumulated, and length is the length in km over which the measurements are accumulated. For AMVs, the quality indicator (QI) is the percent confidence value in % below which observations are rejected.....	<b>22</b>
<b>Table 8:</b> List of parameters the user can modify in the main bash script for the plotting tool (MAINSRIPT_for__Plotting.bash).....	<b>24</b>
<b>Table 9:</b> List of types of figures generated by the plotting tool.....	<b>25</b>
<b>Table 10:</b> List of file name abbreviations for figure files generated by the plotting tool.....	<b>27</b>
<b>Table 11:</b> Collocation application directory abbreviations, full paths, and descriptions.....	<b>35</b>

## Acknowledgments

The authors thank ESA, ECMWF, the NCEP/EMC Obsproc team, and Loon, LLC for providing the data; Peter Marinescu for his contributions towards converting Aeolus BUFR data into 6-hour NCEP prepBUFR for the purpose of data assimilation in NOAA NWP; Chia Moeller for her contribution towards the initial development of tools for analysis and visualization; Iliana Genkova, Jim Jung, Jaime Daniels, Cathy Thomas, and Emily Liu for their expertise on NCEP prepBUFR and winds in NOAA NWP; the Aeolus Cal/Val teams, Mike Hardesty, Erin Jones, Chris Barnet, and Sebastian Bley for their scientific and technical expertise; the SSEC at the University of Wisconsin (UW)-Madison for their support in creating and managing the archive; and Sid Boukabara for his leadership and support of this project.

The authors thank the administrators of the S4 supercomputer (Boukabara et al., [2016](#)) at UW-Madison's Space Science Engineering Center (SSEC) who provided the groundwork for launching SAWC's public release. The authors also thank the IT administrators at the NOAA/NESDIS/Center for Satellite Applications and Research (STAR) for their continued support in hosting and maintaining SAWC on the STAR public server.

The authors acknowledge support from the NOAA/NESDIS Office of Projects, Planning, and Acquisition (OPPA) Technology Maturation Program (TMP) through CICS and the Cooperative Institute for Satellite Earth System Studies (CISESS) at the University of Maryland (UMD)/Earth System Science Interdisciplinary Center (ESSIC) (grant nos. NA14NES4320003 and NA19NES4320002) and the Cooperative Institute for Meteorological Satellite Studies (CIMSS) at UW-Madison (grant no. NA20NES4320003).

# How to Use this Manual

## Inside Each Section

*Section 1* provides an introduction to the System for Analysis of Wind Collocations (SAWC).

*Section 2* describes the components of SAWC, including the structure of the wind observation archive, its contents, file name formats, and how to access it.

*Section 3* describes the wind datasets available in the archive.

*Section 4* provides an overview of the collocation application, including installation and how to run the collocation and plotting tools.

## How to Get Help

For help regarding the SAWC wind archive or collocation software application, please contact:

- Katherine E. Lukens ([katherine.lukens@noaa.gov](mailto:katherine.lukens@noaa.gov))

STAR Webmaster:

- Lori Brown ([lori.brown@noaa.gov](mailto:lori.brown@noaa.gov))



# 1. Introduction

The 2017-2027 decadal survey by the National Academies Press designated that highly accurate 3D atmospheric wind observations are a top priority in the science community (National Academies of Sciences, Engineering, and Medicine, [2018](#)). Numerous studies have been conducted to address this need (e.g., Velden and Holmlund, [1998](#); Velden and Bedka, [2009](#); Bormann et al., [2003](#); Genkova et al., [2008](#), [2010](#); Santek et al., [2014](#), [2019](#), [2021](#), [2022](#); Rani et al., [2022](#); Borde et al., [2016](#), [2019](#); Martin et al., [2021](#); Hoffman et al., [2021](#); Lukens et al., [2022](#)). Many involve AMV comparisons, with each study using widely varying collocation criteria. As there are few accepted collocation standards in the literature and broader wind community, researchers are often forced to invent their own criteria and algorithms. Further, researchers often need to acquire and reformat disparate data from numerous sources, and develop their own analysis tools, all of which can be time-consuming and delay progress.

In support of researchers' needs to acquire and analyze wind data from multiple sources, and to further address the 3D wind requirement, the System for Analysis of Wind Collocations (SAWC) was jointly developed through a collaboration between the NOAA/NESDIS/Center for Satellite Applications and Research ([STAR](#)), the University of Maryland/Cooperative Institute for Satellite Earth System Studies ([CISESS](#)), and the University of Wisconsin-Madison/Cooperative Institute for Meteorological Satellite Studies ([CIMSS](#)). SAWC provides—for the first time—the data and tools one needs for wind observation research in one place: a multi-year archive of global wind observations uniformly formatted in NetCDF—a universally accepted format—as well as select identified pairings of winds collocated in space and time, and a downloadable collocation software application developed for their intercomparison. The potential capability of SAWC is wide-ranging, from product validation and observation error characterization to advancements in data assimilation and the global Earth observing architecture.

This document presents an overview of SAWC and its components, as well as detailed instructions on how to install and operate the collocation application.

## 2. Overview of SAWC

SAWC consists of an archive of global wind observation datasets (hereafter, source datasets), collocation index files containing array locations of matched winds between select datasets, a collocation software application developed to match and statistically compare the wind observations, and a user manual. The archive is hosted and maintained on the NOAA/NESDIS/STAR public server. It is continuously updated and is publicly available online at the following links:

- **SAWC Home:** <https://www.star.nesdis.noaa.gov/data/sawc>
- **Wind Archive:** [https://www.star.nesdis.noaa.gov/data/sawc/wind\\_datasets](https://www.star.nesdis.noaa.gov/data/sawc/wind_datasets)
- **Collocation Application & Index Files:**  
<https://www.star.nesdis.noaa.gov/data/sawc/collocation>
- **User Manual:** [https://www.star.nesdis.noaa.gov/data/sawc/User\\_Manual](https://www.star.nesdis.noaa.gov/data/sawc/User_Manual)

### 2.1 SAWC Contents

#### 2.1.1 Wind Datasets

The archive is home to several source datasets each containing point wind observations. These datasets include:

- **Aeolus** Level-2B (L2B) winds (de Kloe et al., [2020](#))
- **Aircraft** observations (e.g., Moninger et al., [2003](#))
- **Atmospheric motion vectors (AMVs)** (e.g., Velden et al., [1997](#), [2005](#); Cotton et al., [2020](#)) derived from geostationary and polar-orbiting satellite imagery
- Observations from **Loon**, a stratospheric balloon network that is no longer operational (Rhodes and Candido, [2021](#))
- **Sonde** observations (e.g., Durre et al., [2006](#))

Aeolus data are provided by the European Space Agency (ESA). Aircraft, AMV, and sonde observations are provided by NCEP and are processed by the NCEP/EMC Obsproc team prior to the conversion to NetCDF. Note that all aircraft, AMV, and sonde datasets contain only unrestricted observations. Loon stratospheric balloon observations are provided by Loon, LLC.

All source datasets span the entire lifetime of Aeolus (September 2018 through April 2023), except for the acquisition of AMVs and aircraft and sonde winds that continues beyond Aeolus' lifetime to Present Day, and the Loon record that runs from 2011 to 2021 when the Loon project ended.

[Section 2.2](#) lists the paths to each dataset, and [Section 3](#) describes the datasets.

### **2.1.2 Collocation Application and Index Files**

Unique collocation techniques were developed to match wind observations based on user-specified criteria. The application consists of a collocation tool and a plotting tool.

The application's collocation tool is used to generate NetCDF index files that contain array indices or locations of the matched winds within each source data file. One index file is created per Driver-Dependent dataset pairing, where the "Driver" is the reference dataset and is defined as the dataset with which all other "Dependent" datasets are collocated. There are two index arrays per file, one pertaining to the Driver and the other the Dependent dataset; these arrays are the same length, although the lengths may vary between files. In addition, each index file contains the collocation distances and differences in time and height/pressure for each pair of collocated winds and are the same length as the corresponding index arrays. Finally, each index file contains the dataset names, paths to the source files in the archive, and the collocation criteria (maximum allowable difference in time, height/pressure, and distance) used for matching. Note that by saving indices of matched winds and not a copy of the collocated data itself, disk space is saved. Index files for select pairings and time periods are available in the archive.

The application's plotting tool utilizes the index files to extract the matched winds and visually and statistically compare them. See [Section 4.1.3](#) for details.

At present, the tools can only handle point observation wind datasets. Gridded datasets (e.g., NWP reanalyses) as well as other observational products such as cloud and aerosol are being considered for inclusion in future iterations of SAWC.

[Section 2.2](#) lists the locations of the index files and collocation application within the archive.

## **2.2 Directory Paths**

Table 1 lists all SAWC components, available data file formats, temporal coverage, and some key variables available in each. Table 2 lists the archive's main directories and directory abbreviations used in this document. Table 3 lists the paths to all available datasets within the archive.

**Table 1: A list of all SAWC components, including source datasets, index files, and the collocation software application. Additionally, the temporal coverage and key variables available per dataset are listed.**

Table 1 SAWC Component	File Formats	Temporal Coverage	Key Variables Available			
			Time	Horizontal Coords	Vertical Coords	Other
Aeolus	NetCDF, BUFR, ESA's Earth Explorer (EE) format	Sep 2018 - Apr 2023	Year, Month, Day, Hour	Latitude, Longitude	Height, Pressure	HLOS Wind Velocity, Azimuth Angle
Loon	NetCDF-4	2011 - Jan 2021				<i>u</i> -/ <i>v</i> -components, Wind Direction
Sonde		Sep 2018 - Present Day				Wind Speed, Wind Direction
Aircraft					Height	
AMV					Pressure	
Index Files		Select periods			N/A	
Collocation Application	One tarball per application version	N/A				

**Table 2: General archive directory abbreviations, paths, and descriptions.**

Table 2 Path Abbreviation	Full Path (if applicable)	Description
<b>\${HOME}</b>	<b>./sawc</b>	Home directory
<b>\${MAIN_WIND}</b>	<b>\${HOME}/wind-datasets/</b>	Main conventional dataset directory containing all wind datasets
<b>\${MAIN_COL}</b>	<b>\${HOME}/wind-datasets/collocation</b>	Main conventional dataset directory containing index files and the collocation application
<b>\${YYYY}</b>	N/A	Year, 4 digits (e.g., 2022)
<b>\${MM}</b>	N/A	Month, 2 digits (e.g., 01)
<b>\${DD}</b>	N/A	Day, 2 digits (e.g., 01)
<b>\${hh}</b>	N/A	Hour, 2 digits (possible range: 00-23)
<b>\${mm}</b>	N/A	Minute, 2 digits (possible range: 00-59)
<b>\${ss}</b>	N/A	Second, 2 digits (possible range: 00-59)

**Table 3: Paths to all SAWC components (source datasets, index files, and collocation application). Additionally, key variables that are available per dataset are listed.**

<b>Table 3 Dataset</b>	<b>Location (Path) in Archive</b>
<b>Aeolus</b>	<b><code>\${MAIN_WIND}/aeolus</code></b>
<b>Aircraft</b>	<b><code>\${MAIN_WIND}/aircraft</code></b>
<b>AMV</b>	<b><code>\${MAIN_WIND}/AMV</code></b>
<b>Loon</b>	<b><code>\${MAIN_WIND}/Loon</code></b>
<b>Sonde</b>	<b><code>\${MAIN_WIND}/sonde</code></b>
<b>Index Files</b>	<b><code>\${MAIN_COL}/index_files</code></b>
<b>Collocation Application</b>	<b><code>\${MAIN_COL}/application</code></b>

The source datasets and index files are stored by date within their respective directories. Their specific path structures are listed in Sections [2.2.1](#) and [2.2.2](#).

## 2.2.1 Paths to Source Data

### 2.2.1.1 Aeolus Winds

Aeolus winds are provided in daily files (spanning ~24 hours) and are stored by month. Their general location is:

**`${MAIN_WIND}/aeolus/${DataFormat}/${DataType}/${YYYY}/${MM}`**

- There are three options for **`${DataFormat}`**:
  1. **bufr**. This format is the Binary Universal Form for the Representation of meteorological data (BUFR).
  2. **EE-format**. This format is ESA's Earth Explorer (EE) binary format.
  3. **netcdf**. This format is NetCDF.
- There are two options for **`${DataType}`**:
  1. **original**. "Original" refers to the L2B winds processed using the original operational L2B processor version at the time of retrieval. Within this directory the data are stored by date, with the full path being:  
**`${MAIN_WIND}/aeolus/${DataFormat}/original/${YYYY}/${MM}/${DD}`**
  2. **reprocessed**. "Reprocessed" refers to the L2B winds that have been reprocessed using a different L2B processor version than what was originally used at the time of retrieval. Within this directory the data are first stored by the L2B processor version used to reprocess the data, and then by date, with the full path being:

**`${MAIN_WIND}/aeolus/${DataFormat}/reprocessed/${Processor}/${YYYY}/${MM}/${DD}`**

At the time of writing, there are three options for **`${Processor}`**: **2B10**, **2B11**, and **2B14**. “2B” refers to Aeolus Level-2B data. The numbers after “2B” indicate the version or “Baseline” of the L2B processor used to reprocess the data. More information on Baselines can be found in [Section 3.1](#).

- Within the **bufr** directory, there is a third option for **`${DataType}`**:
  3. **trim3hr**. This directory contains Aeolus BUFR data that has been reformatted to NCEP prepBUFR, with each file containing data centered around one of the four synoptic times (0000, 0600, 1200, 1800 UTC). There are two subsets of Aeolus data available in this format:
    - a. Original data for the year 2020, with the full path being:  
**`${MAIN_WIND}/aeolus/bufr/trim3hr/original/${YYYY}/${MM}/${DD}`**
    - b. Reprocessed data using Baseline 10 (2B10) spanning June-December 2019, with the full path being:  
**`${MAIN_WIND}/aeolus/bufr/trim3hr/reprocessed/2B10/${YYYY}/${MM}/${DD}`**

### 2.2.1.2 All Other Winds

Non-Aeolus winds are provided in 6-hourly NetCDF files (centered around the NWP analysis hours 00, 06, 12, and 18 UTC) and are stored by day. Their general location is:

**`${MAIN_WIND}/${Dataset}/${YYYY}/${MM}/${DD}`**

- There are several options for **`${Dataset}`**:
  - **aircraft**. This directory contains in situ atmospheric observations reported at aircraft flight level, with the full path being:  
**`${MAIN_WIND}/aircraft/${YYYY}/${MM}/${DD}`**
  - **AMV**. This directory contains atmospheric motion vectors derived from both geostationary and polar Earth-orbiting satellite imagery, with the full path being:  
**`${MAIN_WIND}/AMV/${YYYY}/${MM}/${DD}`**
  - **Loon**. This directory contains in situ atmospheric observations reported at balloon level in the upper troposphere/lower stratosphere, with the full path being:  
**`${MAIN_WIND}/Loon/${YYYY}/${MM}/${DD}`**
  - **sonde**. This directory contains in situ atmospheric observations reported in profile by sonde instrumentation, with the full path being:  
**`${MAIN_WIND}/sonde/${YYYY}/${MM}/${DD}`**

## 2.2.2 Paths to Collocation Application and Index Files

The directory **`${MAIN_COL}`** contains:

- *Collocation software application (all versions)*. The application was developed to match and statistically compare wind observations, with the full path being:  
**`${MAIN_COL}/application`**
- *Collocation index files*. These files are provided in 6-hourly NetCDF files (centered around the NWP analysis hours 00, 06, 12, and 18 UTC) and are stored by day. They

contain indices of identified pairings of winds between Driver and Dependent datasets, with the full path being:

**`${MAIN_COL}/index_files/${YYYY}/${MM}/${DD}`**

## 2.3 File Name Formats

This section provides structures and examples of file names available within the archive.

### 2.3.1 Aeolus Data File Names

- Within the **`${MAIN_WIND}/aeolus/bufr`** directory, the file names are listed as:  
**`AE_OPER_ALD_B_N_2B_${AEOLUS_DATETIME}_0001.BUFR`**  
 The path abbreviation **`${AEOLUS_DATETIME}`** indicates the date/time range of observations per BUFR file and refers to:  
**`${YYYY}${MM}${DD}T${hh}${mm}${ss}-${YYYY}${MM}${DD}T${hh}${mm}${ss}`**  
*Example file name:*  
**`AE_OPER_ALD_B_N_2B_20211231T233917_20220101T011004_0001.BUFR`**
- Within the **`${MAIN_WIND}/aeolus/bufr/trim3hr`** directory, the file name format is:  
**`Aeolus_ESA_${YYYY}${MM}${DD}${hh}.ncep`**  
*Example file name:*  
**`Aeolus_ESA_2020010100.ncep`**
- Within the **`${MAIN_WIND}/aeolus/EE-format`** directory, there are two files for each date/time range, a header file in xml (\*.HDR) and a binary data file (\*.DBL). The file names are listed as:  
**`AE_OPER_ALD_U_N_2B_${AEOLUS_DATETIME}_0001.DBL`**  
**`AE_OPER_ALD_U_N_2B_${AEOLUS_DATETIME}_0001.HDR`**  
*Example file names:*  
**`AE_OPER_ALD_U_N_2B_20220101T010941_20220101T032904_0001.DBL`**  
**`AE_OPER_ALD_U_N_2B_20220101T010941_20220101T032904_0001.HDR`**
- Within the **`${MAIN_WIND}/aeolus/netcdf`** directory, the file names are listed as:  
**`Aeolus.L2B.${AEOLUS_WIND_TYPE}.1day.${YYYY}${MM}${DD}.nc`**  
 There are two options for **`${AEOLUS_WIND_TYPE}`**:
  1. **MieCloud**. This refers to winds derived from Mie-cloudy (aerosol) backscatter retrievals.
  2. **RayClear**. This refers to winds derived from Rayleigh-clear (molecular) backscatter retrievals.*Example file names:*  
**`Aeolus.L2B.MieCloud.1day.20220101.nc`**  
**`Aeolus.L2B.RayClear.1day.20220101.nc`**

### 2.3.2 Non-Aeolus Data File Names

- Within the **`${MAIN_WIND}/aircraft`** directory, the file names are listed as:  
**`gdas.${YYYY}${MM}${DD}${hh}.aircft.tm00.bufr_d.nc4`**  
*Example file name:* **`gdas.2022010100.aircft.tm00.bufr_d.nc4`**
- Within the **`${MAIN_WIND}/AMV`** directory, the file names are listed as:  
**`gdas.${YYYY}${MM}${DD}${hh}.satwnd.tm00.bufr_d.nc4`**  
*Example file name:* **`gdas.2022010100.satwnd.tm00.bufr_d.nc4`**
- Within the **`${MAIN_WIND}/Loon`** directory, the file names are listed as:  
**`Loon_${YYYY}${MM}${DD}${hh}.nc4`**  
*Example file name:* **`Loon_2020010100.nc4`**
- Within the **`${MAIN_WIND}/sonde`** directory, the file names are listed as:  
**`gdas.${YYYY}${MM}${DD}${hh}.adpupa.tm00.bufr_d.nc4`**  
*Example file name:* **`gdas.2022010100.adpupa.tm00.bufr_d.nc4`**

### 2.3.3 Collocation Index File Names

Within the **`${MAIN_COL}/index_files`** directory, the file names are listed as:

**`index.${YYYY}${MM}${DD}${hh}.drv_${DRV_DSET}__${iQC}.dset1_  
 ${DEP_DSET}__${iQC}.nc4`**

*Example file name:*

**`index.2019090200.drv_Aeolus_MieCloud_ReprocB11__QC.dset1_  
 Aircraft__NoQC.nc4`**

This example index file contains array indices (i.e., array locations in the source data files) of collocated winds. Winds from the Dependent dataset (*Aircraft*) indicated by *dset1\_* are collocated with the Driver dataset (*Aeolus\_MieCloud\_ReprocB11*: Aeolus Mie-cloudy reprocessed using the Baseline 11 version of the L2B processor) indicated by *drv\_*. Note that:

- The Driver winds are QC'd prior to collocation (indicated by *\_\_QC*).
- The Dependent winds are not QC'd prior to collocation (indicated by *\_\_NoQC*).

Table 4 lists and defines the options for **`${DRV_DSET}`**, **`${DEP_DSET}`**, and **`${iQC}`** in the index file names.

**Table 4: Index file name abbreviations, available options per abbreviation, and corresponding descriptions.**

<b>Table 4</b> Abbreviation	Available Options	Description
<b><code>\${DRV_DSET}</code></b>	<b><code>Aeolus_\${AEOLUS_WIND_TYPE}_  \${COL_AEOLUS_TYPE}</code></b>	<i>Driver dataset:</i> Obs locations to which obs from the Dependent dataset are collocated
<b><code>\${DEP_DSET}</code></b>	<b>Aircraft  AMV_NCEP  Loon  Sonde</b>	<i>Dependent dataset:</i> Obs that are collocated to the Driver obs locations



<b>Table 4</b> Abbreviation	Available Options	Description
<b>\${COL_AEOLUS_TYPE}</b>	<b>Orig</b>	Aeolus winds processed with L2B processor version at time of retrieval
	<b>Reproc\${BASELINE}</b>	Aeolus winds reprocessed with different (typically updated) L2B processor (indicated by Baseline number <b>\${BASELINE}</b> ) than that at time of retrieval
<b>\${BASELINE}</b>	<b>B##</b>	L2B processor Baseline number, where <b>##</b> indicates the 2-digit Baseline number
<b>\${iQC}</b>	<b>NoQC</b>	Prior to collocation, no QC is applied to the dataset (name that precedes '___' in file name)
	<b>QC</b>	Prior to collocation, QC is applied to the dataset (name that precedes '___' in file name)

## 2.4 Supplemental Information

Information files in ASCII format that accompany the source data and collocation index files contain additional information about the datasets, including Aeolus data exclusion periods, AMV satellite identifier list, and other details about each dataset. Users are encouraged to familiarize themselves with these files, which can be found in the **\${MAIN\_WIND}/docs** directory.

All available variables for the NetCDF source datasets and index files are defined in a separate document called [SAWC Variable Definitions](#).

- For the datasets acquired from NCEP (aircraft, AMV, sonde), each NetCDF variable is associated with a BUFR Mnemonic Code, and the values of several of these variables correspond to integers with specific meanings listed in corresponding BUFR Code/Flag Tables. To look up this information for a NetCDF variable, simply find the associated BUFR mnemonic code (fifth column in the Variable Definitions document), look up that mnemonic code at the *BUFR Table B* website [https://www.nco.ncep.noaa.gov/sib/jeff/bufrtab\\_tableb.html](https://www.nco.ncep.noaa.gov/sib/jeff/bufrtab_tableb.html), and click on the *Code Table* or *Flag Table* link under the *Units* column.

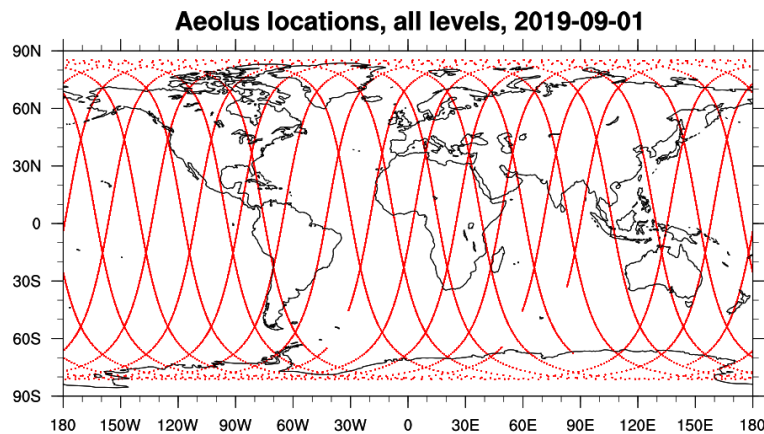
### 3. Dataset Descriptions

This section describes the source datasets currently available in the SAWC wind archive.

#### 3.1 *Aeolus*

Aeolus L2B winds are derived from backscattering signals retrieved by the Aladin Flight Module A (FM-A) (or the backup FM-B) Doppler wind lidar instrument on a dawn/dusk orbit. The satellite completes one orbit approximately every 92 min, with global coverage achieved every 7 days. Winds are observed in profile along the laser's Horizontal Line-of-Sight (HLOS). See the Aeolus Operations website for more information (ESA [2020](https://www.esa.int/ESA_Missions/Aeolus)).

Two main wind regimes are derived using ESA's L2B processor: (1) Rayleigh-clear (molecular backscattering) representing winds in clear scenes, and (2) Mie-cloudy (aerosol backscattering) representing winds in cloudy scenes. Aeolus spatial coverage for one day is shown in Fig. 1. Temporal coverage of the archived data spans the entire lifetime of Aeolus (September 2018 through April 2023).



**Figure 1: Spatial coverage of Aeolus observations for a 24-h period (2019-09-01). Colored dots denote the observation locations.**

The L2B processor has been updated several times over the course of Aeolus's lifetime; these updates are designated as Baselines and are implemented operationally for certain periods of time. Table 5 lists the different Baselines available in the archived Aeolus dataset and the dates on which they became operational. Note that because different processors have been implemented over the course of Aeolus' lifetime, the data quality and metadata available may change over time.

**Table 5: List of L2B processor versions as indicated by Baseline number. The operational start date of each Baseline and relevant details are also provided.**

<b>Table 5 Baseline</b>	<b>Operational Start Date</b>	<b>Details</b>
<b>B01</b>	2018-09-03	<ul style="list-style-type: none"> <li>FM-A instrument used for retrieval</li> <li>Valid for 2018-09-03 through 2018-10-16</li> </ul>
<b>B02</b>	2018-09-03	<ul style="list-style-type: none"> <li>FM-A instrument used for retrieval</li> <li>Valid for 2018-09-03 through 2019-05-16</li> </ul>
<b>B03</b>	2019-05-16	<ul style="list-style-type: none"> <li>FM-A instrument used for retrieval</li> <li>Valid for 2019-05-16 through 2019-06-14</li> </ul>
<b>B04</b>	2019-06-14	<ul style="list-style-type: none"> <li>FM-A instrument used for retrieval</li> <li>Valid for 2019-06-14 through 2019-06-16</li> </ul>
<b>B05</b>	2019-06-28	<ul style="list-style-type: none"> <li>Switch from FM-A to FM-B initiated on 2019-06-16 due to FM-A power loss</li> <li>FM-B instrument used for retrieval</li> <li>Valid for 2019-06-28 through 2019-09-05</li> </ul>
<b>B06</b>	2019-06-28	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2019-06-28 through 2019-10-31</li> </ul>
<b>B07</b>	2019-10-31	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2019-10-31 through 2020-04-02</li> </ul>
<b>B08</b>	2020-04-02	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2020-04-02 through 2020-04-20</li> </ul>
<b>B09</b>	2020-04-20	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2020-04-20 through 2020-07-09</li> </ul>
<b>B10</b>	2020-04-20	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2020-04-20 through 2020-10-08</li> <li>Usage of M1 mirror temperature correction in L2B processor enabled</li> <li>Regeneration of B09 into B10 completed (<i>B09 is equivalent to B10</i>)</li> <li>Used for first reprocessing campaign: reprocess winds from 2019-06-28 through 2019-12-31</li> </ul>
<b>B11</b>	2020-10-08	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2020-10-08 through 2021-05-26</li> <li>Used for second reprocessing campaign: reprocess winds from 2019-06-28 through 2020-10-06</li> </ul>
<b>B12</b>	2021-05-26	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2021-05-26 through 2021-12-06</li> </ul>
<b>B13</b>	2021-12-06	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2021-12-06 through 2022-03-29</li> </ul>
<b>B14</b>	2022-03-29	<ul style="list-style-type: none"> <li>FM-B instrument used for retrieval</li> <li>Valid for 2022-03-29 through 2022-09-13</li> <li>Used for third reprocessing campaign: reprocess winds from 2018-08-31 through 2019-06-16, when FM-A instrument was operational</li> </ul>

<b>Table 5 Baseline</b>	<b>Operational Start Date</b>	<b>Details</b>
<b>B15</b>	2022-09-13	<ul style="list-style-type: none"> <li>Completed switch from FM-B back to FM-A on 2022-11-29</li> <li>FM-B instrument used for retrieval for 2022-09-13 through 2022-10-04</li> <li>FM-A instrument used for retrieval for 2022-11-27 onward</li> </ul>
<b>B16</b>	2023-04-18	<ul style="list-style-type: none"> <li>FM-A instrument used for retrieval</li> <li>Valid for 2023-04-18 through 2023-04-30</li> <li>Mie-cloudy winds have been significantly improved with this baseline</li> </ul>

*Additional notes on Aeolus:*

- EE-format data are from ESA's public Aeolus Online Dissemination System: <https://aeolus-ds.eo.esa.int/oads/access/> (ESA, 2020)
  - The EE files can be converted into ASCII using ESA's [CODA](#) tool.
- *NetCDF files:*
  - NetCDF files are converted from EE format with metadata retained. Because only EE data flagged as "valid" are saved as NetCDF, there may be some dates without NetCDF files but with EE files.
    - Aeolus Rayleigh wind observations come with accompanying reference pressures and temperatures derived from ECMWF model forecasts (e.g., Dabas et al., 2008; Šavli et al., 2021), whereas Mie winds do not.
    - During the NetCDF conversion, each Mie wind observation is assigned a reference pressure and temperature from the nearest Rayleigh observation. (First, the closest Rayleigh profile to the Mie observation in space and time is identified. Then, the Rayleigh observation within that profile with the smallest vertical distance from the Mie observation is found and the pressure and temperature from that Rayleigh observation is assigned to the Mie observation; note that this only happens if the minimum vertical distance between the observations is  $\leq 1.0$  km. If there is more than one Rayleigh observation with the same minimum vertical distance to the Mie observation in question, then the pressure/temperature of the Rayleigh observation with the most similar wind speed to the Mie observation is chosen.)
  - Off-orbit profiles are removed prior to conversion to NetCDF. See the document file `${MAIN_WIND}/docs/aeolus/Aeolus_L2B_Offorbit_Profiles.txt` for more information.
  - Duplicate profiles are removed prior to conversion to NetCDF. See the document file `${MAIN_WIND}/docs/aeolus/Aeolus_L2B_Duplicate_Profiles.txt` for more information.
  - NetCDF files are daily files covering ~24 hour periods (00:00-23:59 UTC) and are stratified by Aeolus wind regime listed in each file name (e.g., `Aeolus.L2B.${AEOLUS_WIND_TYPE}.1day.20190801.nc`)
  - Aeolus QC flags are added to the NetCDF files based on ESA- and ECMWF-recommended QC decisions (Rennie and Isaksen, 2020). See the

document file [\\${MAIN\\_WIND}/docs/aeolus/Aeolus\\_L2B\\_Quality\\_Controls.txt](#) for more information.

- Please be aware that starting with Baseline B10, the units for the estimated HLOS wind error (name key in CODA: "hlos\_error\_estimate") are now  $\text{cm s}^{-1}$  for both L2B Rayleigh and Mie winds in the EE-format files; prior to Baseline B10, the units were  $\text{m s}^{-1}$ .
  - Note that during the NetCDF conversion process, the estimated HLOS wind error is converted back to  $\text{m s}^{-1}$ . See the document file [\\${MAIN\\_WIND}/docs/aeolus/Aeolus\\_L2B\\_Obs\\_Errors.txt](#) for more information.
- *Data Exclusion Dates:*
  - Certain time periods are known to have degraded data content and are NOT recommended for use in research or operations. See the document file [\\${MAIN\\_WIND}/docs/aeolus/Aeolus\\_L2B\\_Data\\_Exclusion\\_List.txt](#) for more information.
- *Hot Pixels:*
  - It should be noted that a patch was implemented in the operational processor which corrects pixel intensities for elevated dark signal levels (defined as "hot pixels") on Friday June 14, 2019 at 08:30 UTC. The L2B processing flags winds as invalid if a strong bias from uncorrected hot pixels occurs (ESA, [2021](#)). For more information on hot pixels and their correction, see Weiler et al. ([2021a](#)).
- *Mirror Temperature Corrections:*
  - Small fluctuations in the temperatures across the telescope's primary mirror onboard Aeolus was linked to wind biases as high as  $8 \text{ m s}^{-1}$ . This systematic error has since been corrected, reducing much of the wind bias. The mirror temperature correction has been implemented in operations since April 2020. For more information, see Weiler et al. ([2021b](#)).
- For additional information on Aeolus, visit:
 

[https://www.esa.int/Applications/Observing\\_the\\_Earth/FutureEO/Aeolus](https://www.esa.int/Applications/Observing_the_Earth/FutureEO/Aeolus)

and

<https://space.oscar.wmo.int/satellites/view/aeolus>

## 3.2 Aircraft

Aircraft observations are reported at flight level in the upper troposphere/lower stratosphere (cruise altitude) and in the ascending and descending legs of each flight. The spatial coverage is regional, mostly in the NH as seen in Fig. 2. Temporal coverage of the archived data spans from September 2018 to Present Day.

The data are converted to NetCDF from NCEP prepBUFR files ingested by NOAA operations for assimilation. The data are stratified by aircraft type (which are listed in the corresponding prepBUFR table containing additional metadata information: [\\${MAIN\\_WIND}/docs/bufr\\_tables/prepbufr\\_table.aircraft](#)) and include winds and other observations.

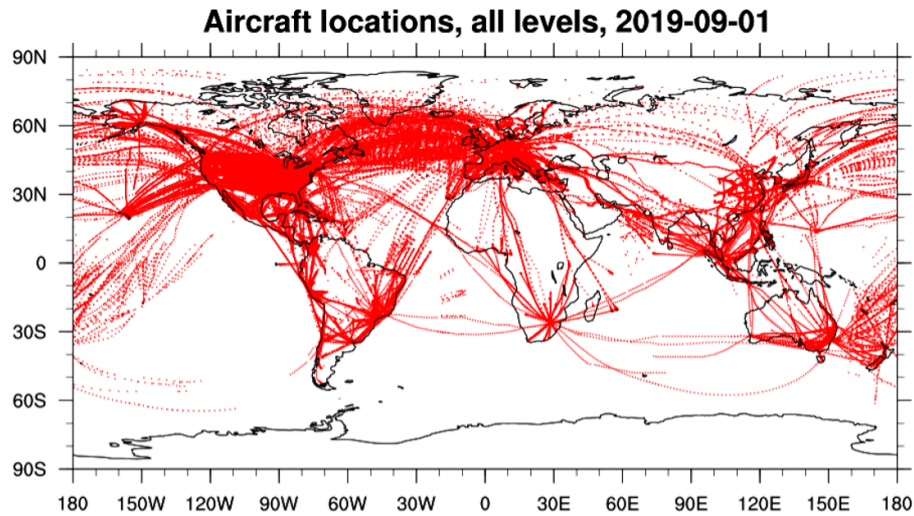


Figure 2: Spatial coverage of aircraft observations for a 24-h period (2019-09-01). Colored dots denote the observation locations.

### 3.3 Atmospheric Motion Vectors (AMVs)

Atmospheric motion vectors are derived from tracking clouds and water vapor features in satellite imagery through time. The archived AMVs include derived winds from both geostationary (GEO) and polar Earth-orbiting (LEO) satellites. Spatial coverage is near-global at various vertical levels, as shown in Fig. 3. Temporal coverage of the archived data spans from September 2018 to Present Day.

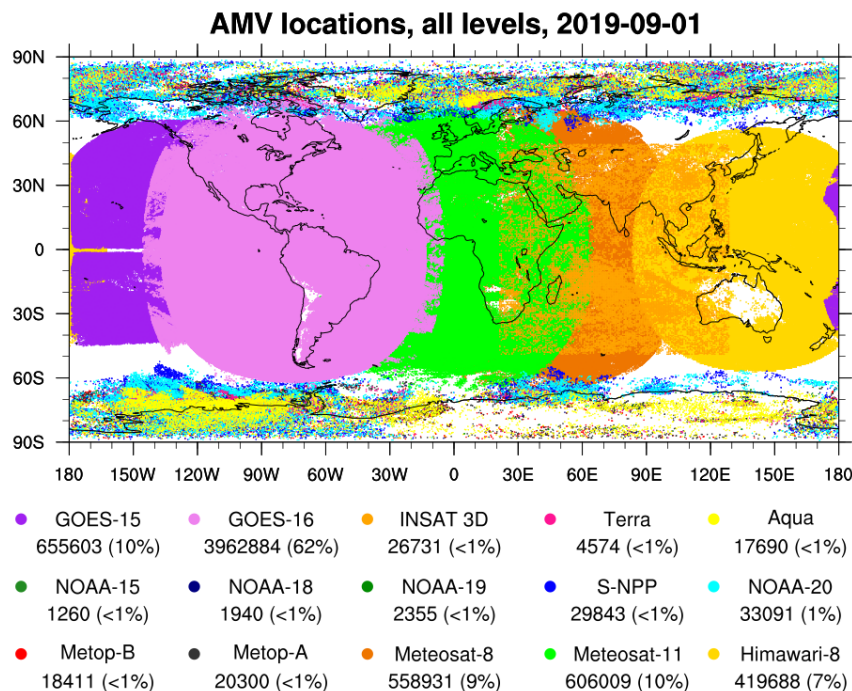


Figure 3: AMV spatial coverage for a 24-h period (2019-09-01). Colored dots denote individual satellite-derived AMV observations with quality indicator (QI) > 80%. Each item in the legend lists the satellite name (top), the number of AMVs from the satellite (bottom left), and the percentage of all AMVs the satellite's count represents (bottom right in parentheses).

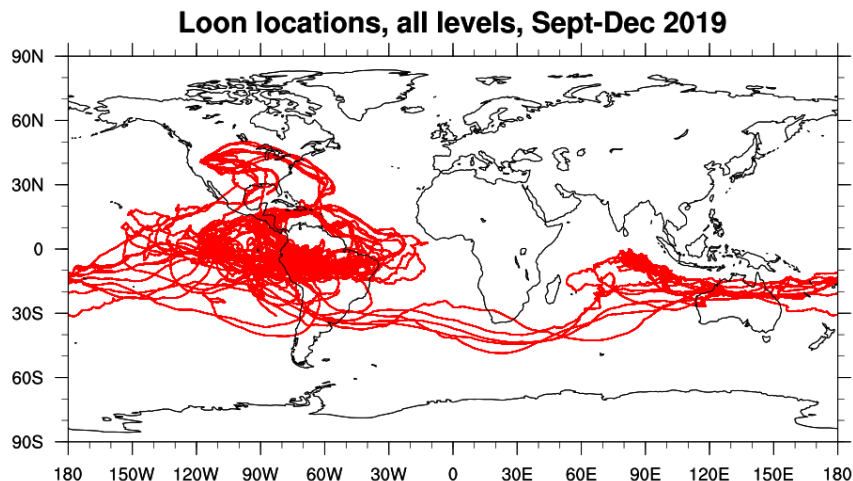


AMVs are converted to NetCDF from NCEP prepBUFR files that are ingested by NOAA operations for assimilation. The AMV data can be stratified by the satellite from which they are derived; the full list of possible satellite sources (**AMV\_Satellite\_Identifer\_List.txt**) as well as corresponding prepBUFR tables (**prepbufr\_table.AMV\***) that define certain metadata items can be found in the **`${MAIN_WIND}/docs/bufr_tables`** directory.

### 3.4 Loon

Loon was a network of stratospheric superpressure balloons deployed by Loon, LLC, a subsidiary of Google's parent company Alphabet, to provide internet connectivity to regions with limited access from 2011-2021 ([Rhodes and Candido, 2021](#)). The balloon configurations consisted of an instrument payload platform tethered to a tennis court-sized balloon specifically designed to withstand harsh stratospheric conditions. The balloons remained aloft for months at a time, and a ground team was able to control their movements by remotely adjusting the balloon pressure in order to move the configurations into different airstreams. The spatial coverage is regional (Fig. 4) and depended upon the Loon mission at the time of deployment. In situ atmospheric observations were reported at balloon level (~50-100 hPa or 18-20 km) at a frequency of ~1-20 minutes. Temporal coverage of the archived data spans from 2011 to 2021 when the Loon project ended.

A full description of the Loon dataset is available here: <https://doi.org/10.5281/zenodo.3755988>



**Figure 4: Spatial coverage of Loon stratospheric balloon observations for Sept-Dec 2019. Colored dots denote the observation locations.**

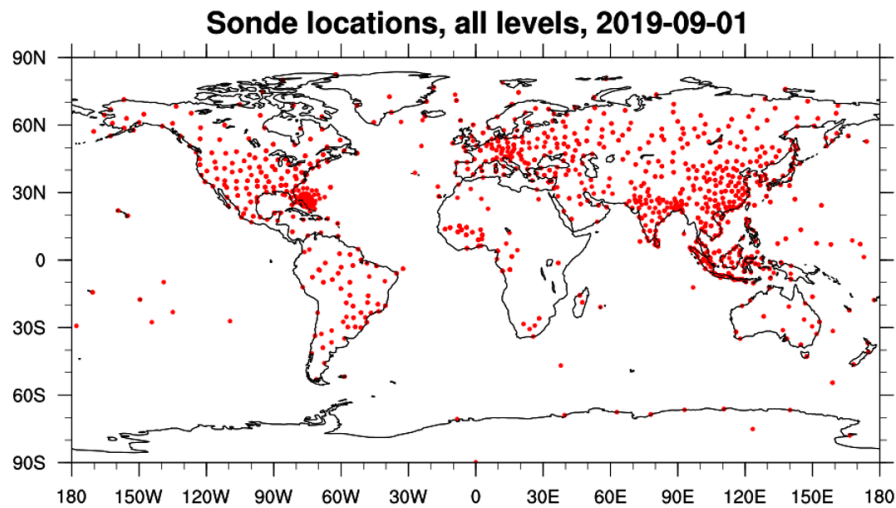
*A note on Loon:*

- Before downloading or using the Loon data, please read the following files located in **`${MAIN_WIND}/docs`**:
  - **Loon\_Flight\_Data.pdf**
  - **Loon\_License\_Agreement–Noncommercial\_Use.pdf**

### 3.5 Sondes

Sondes (radio-, drop-, or rawinsondes) report in situ observations in profile at balloon level throughout the troposphere and into the lower stratosphere. The spatial coverage is regional, mostly over land (Fig. 5). Temporal coverage of the archived data spans from September 2018 to Present Day.

The data are converted to NetCDF from NCEP prepBUFR files ingested by NOAA operations for assimilation and are stratified by sonde type (which are listed in the prepBUFR table `${MAIN_WIND}/docs/bufr_tables/prepbufr_table.sonde`). They include winds and other observations.



**Figure 5: Spatial coverage of sonde observations for a 24-h period (2019-09-01). Colored dots denote the observation locations.**



## 4. Collocation Software Application

### 4.1 Overview

The collocation software application was developed as part of a collaboration between NOAA/NESDIS/STAR, CISESS at UMD, and CIMSS at UW-Madison to support researchers' needs to acquire and analyze wind data from multiple sources, and to further address the National Academies observation requirement for highly accurate 3D winds (National Academies of Sciences, Engineering, and Medicine, [2018](#)). The application is Python-based and is controlled via Bash scripting. A basic understanding of each language is recommended, although not required.

The application has two main components:

1. *Collocation Tool*. Developed to match observations from any number of wind datasets using user-specified collocation criteria. Output index files are generated and contain the locations of matched winds in the source data files.
2. *Plotting Tool*. Uses the index files to extract and statistically and visually compare the matched winds by calling a suite of plotting functions.

It is important to note that the collocation and plotting tools are flexible and designed to handle additional wind datasets not yet available in the archive. Later in this section, details are provided on the installation and utility of the application, including more advanced modifications the user can perform such as adjusting certain parameters and adding new datasets for comparison.

Figure 6 illustrates SAWC's end-to-end process that is composed of three main steps:

1. Data Acquisition
2. Collocation of Winds
3. Analysis and Visualization

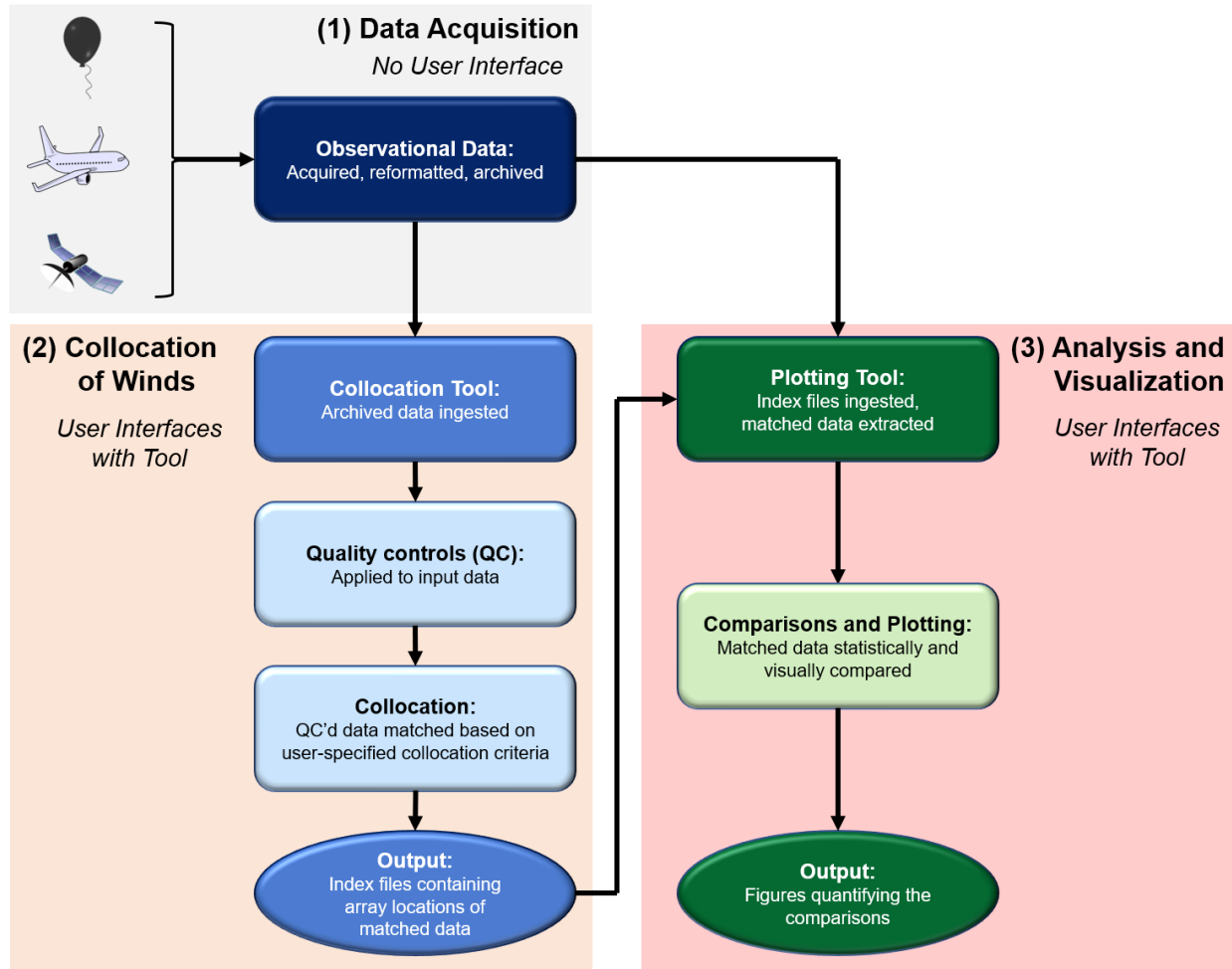


Figure 6: Flow chart illustrating SAWC's end-to-end process. Note that the user interfaces with the collocation and plotting tools in Steps (2) and (3), respectively.

#### 4.1.1 Data Acquisition

Data is acquired from various sources (including ESA, ECMWF, NOAA/NCEP, and Loon) and converted to a common format (NetCDF) for archival. All NetCDF data files have corresponding metadata embedded to provide additional information about the observations and data sources. See [Appendix B](#) for tips on how to quickly view a NetCDF file.

#### 4.1.2 Collocation of Winds

The collocation tool matches wind observations between different datasets. It reads NetCDF files from a minimum of two datasets: one is the Driver (or reference) dataset, and the other is the Dependent dataset. The Driver is the dataset with which all Dependent datasets are collocated. The tool allows for one or more Dependent datasets to be collocated against a single Driver dataset at one time.

Users can modify numerous input parameters for the collocation tool in its main bash script:

**MAINSRIPT\_for\_\_Collocation.bash**

This initiates the collocation tool as a SLURM job that runs in the background on your machine.

A full list of input parameters the user can change in the main bash script is provided in Table 6.

**Table 6: List of input parameters the user can modify in the main bash script for the collocation tool (MAINSRIPT\_for\_\_Collocation.bash).**

<b>Table 6 Parameter</b>	<b>Type</b>	<b>Description</b>
<b>yyyy</b>	Integer	Year. Must be one 4-digit number.
<b>mm</b>	Integer	Month array. Must contain at least one array item.
<b>ddarr</b>	Integer	Day array. Must contain at least one array item.
<b>dir_home</b>	String	Full path to location of <b>MAINSRIPT_for__Collocation.bash</b>
<b>dir_out</b>	String	Full path to location where collocation index files are to be saved.
<b>archive_parent</b>	String	Full path to the home directory (parent directory of <b>/wind_datasets</b> and <b>/collocation</b> directories) on your local machine.
<b>dataset_names</b>	String	Names of datasets to be collocated. The first dataset is set as the Driver; all others are set as Dependents. Must contain at least 2 dataset names.
<b>dst_max</b>	Float	Maximum collocation distance. Must have one array item for every Dependent dataset listed in <b>dataset_names</b> . Units = km
<b>prs_max</b>	Float	Maximum log10(pressure) difference. Must have one array item for every Dependent dataset listed in <b>dataset_names</b> . Units = log10(hPa)
<b>tim_max</b>	Float	Maximum time difference. Must have one array item for every Dependent dataset listed in <b>dataset_names</b> . Units = minutes
<b>hgt_max</b>	Float	Maximum height difference. Must have one array item for every Dependent dataset listed in <b>dataset_names</b> . Units = km
<b>qc_flags</b>	Integer	Quality control (QC) flags for each dataset listed in <b>dataset_names</b> . <i>Options available: 0</i> (no QC applied), or <i>1</i> (QC applied)
<b>amv_qi_flags</b>	Integer	Quality indicator (QI) value for AMV observations. This parameter must be set even if AMVs are not listed in <b>dataset_names</b> . Units = % (percent)
<b>qi_choices</b>	String	QI choice for AMVs from NCEP ("AMV_NCEP"). <i>Options available: NO_FC</i> (default; use QI variable without forecast), or <i>YES_FC</i> (use QI variable with forecast)
<b>n_max</b>	Integer	Number of Dependent observations allowed to match each Driver observation.. Default = 50

Table 6 Parameter	Type	Description
<b>nproc</b>	Integer	Number of processors to use during parallelization. Default = 50
<b>account</b>	String	HPC account. This is the account you the user are allowed to use.
<b>qos</b>	String	HPC QoS. This is for setting SLURM jobs.
<b>partition</b>	String	HPC partition. This is for setting SLURM jobs.
<b>timelimit</b>	String	HPC runtime limit for SLURM jobs.
<b>bline</b>	String	Abbreviation for Aeolus L2B dataset type. <i>Options available:</i> <b>orig</b> (choose dataset processed with original L2B processor at time of retrieval), or <b>B##</b> (choose dataset re-processed with different L2B processor than that used at time of retrieval. <b>##</b> is a 2-digit number indicating the Baseline number, e.g., B10 = Baseline 10)
<b>dtype</b>	String	Subdirectory based on <b>bline</b> choice. <i>Options available:</i> <b>original</b> (if <b>bline</b> = <b>orig</b> ), or <b>reprocessed/2BB</b> (if <b>bline</b> is <u>not</u> <b>orig</b> . <b>BB</b> is the parameter <b>bline</b> )
{load modules}	–	A series of module load commands needed to run the collocation tool. Please check that the following modules are loaded properly as per your HPC system's protocols: Hdf, hdf5, netcdf4, udunits2, ncview, nco

Note that **n\_max** (Table 6) indicates the maximum number of Dependent observations allowed to be collocated with a single Driver observation; the default is 50 and can be changed by the user, if desired. This means that multiple Dependent observations may be collocated to one Driver observation, and one Dependent observation may be collocated with multiple Driver observations. The option to “super-ob” these multiple collocations is available within the plotting tool and is described in [Section 4.1.3](#).

The collocation approach is based on that developed at CIMSS (e.g., Santek et al., [2021](#)): Observations from one or more Dependent datasets are collocated to the 4D locations (latitude, longitude, height/pressure, and time) of each Driver observation. The default collocation criteria are:

- 100 km maximum great circle distance (150 km for sondes)
- 60 minute maximum absolute time difference (90 minutes for sondes)
- 0.04 maximum absolute log<sub>10</sub>(hPa) pressure difference
- 1 km maximum absolute height difference

Both pressure and height criteria are required as input. This is because during the collocation, a choice is made to use either the pressure or height criterion. Pressure is used if pressure is available in both the Driver and Dependent datasets. If neither the Driver nor Dependent dataset has a pressure variable, or if only one of them has pressure, then the height criterion is used. If

height is to be used, and only pressure but not height is available for one dataset, that dataset's pressure variable is converted to height following the NWS formulation:

$$h_{alt} = \frac{0.3048}{1000} * (145366.45 * (1 - (\frac{p}{1013.25})^{0.190284}))$$

where  $p$  is the input pressure in hPa, and  $h_{alt}$  is the resulting height in km.

Once the collocation tool reads in the input data, QCs are applied if the user chooses this option. At present, QCs are readily available for two datasets, Aeolus and AMVs, based on recommendations from the data producers and broader wind community (Table 7). It is important to note that the user is able to change the QC parameters and/or add QCs for the other datasets if desired. After QC, the collocation tool compares the QC'd data from the Driver and Dependent datasets and finds the array locations of the matched winds within those files. All Dependent observations are compared to each Driver observation, and if a Dependent observation meets all selected collocation criteria, its array location (i.e., index) is saved to the corresponding output index file.

**Table 7: Default QC parameters currently available in SAWC as recommended by the data producers and broader wind community. QCs are applied to the listed datasets only if the user selects this option. Reject values are listed in the table. For Aeolus,  $p$  is pressure in hPa,  $\sigma$  is the L2B HLOS wind uncertainty in m/s,  $z$  is the height of the vertical range bins within which measurements are accumulated, and  $length$  is the length in km over which the measurements are accumulated. For AMVs, the quality indicator (QI) is the percent confidence value in % below which observations are rejected.**

<b>Table 7 Dataset</b>	<b>QC Parameters</b>
<b>Aeolus Mie-cloudy</b>	$p > 800$ hPa $\sigma > 5$ m/s
<b>Aeolus Rayleigh-clear</b>	$p > 800$ hPa $\sigma > 8.5$ m/s for $800 \geq p > 200$ hPa $\sigma > 12$ m/s for $p \leq 200$ hPa $z < 0.3$ km $length < 60$ km
<b>AMV</b>	QI < 80

Note that if Aeolus is chosen as either a Driver or Dependent dataset, the collocation tool does not need to be run multiple times to match winds from each wind regime (Rayleigh-clear and Mie-cloudy); rather, the tool initiates jobs for both Aeolus wind regimes simultaneously.

### 4.1.3 Analysis and Visualization

The plotting tool visually and statistically compares matched winds. For a user-specified range of dates, the tool reads in the array locations of collocated winds from the index files and extracts the matched wind observations from the source data files.

Once extracted but before the intercomparisons are performed, the winds undergo a mathematical conversion if Aeolus is listed as either a Driver or Dependent dataset. Specifically, non-Aeolus winds that are matched with Aeolus are projected onto the Aeolus HLOS wind direction prior to any statistical analysis or plotting taking place. If Aeolus is the Driver, all Dependent winds are projected onto the HLOS direction, and all statistical analyses are computed using the HLOS-projected winds. If Aeolus is listed as a Dependent dataset, the matched Driver winds are projected onto the HLOS direction, and statistical analyses only concerning Aeolus are computed using the HLOS-projected winds; otherwise, the Driver-Dependent matches are compared using the wind data as is. Eqs. (1-4) are used to project non-Aeolus winds  $y$  onto the Aeolus HLOS direction  $x_{dir}^{HLOS}$ :

$$u_y = -y_{spd} \sin(y_{dir}) \quad (1)$$

$$v_y = -y_{spd} \cos(y_{dir}) \quad (2)$$

$$y_{spd}^{HLOS} = (-u_y \sin(x_{dir}^{HLOS})) + (-v_y \cos(x_{dir}^{HLOS})) \quad (3)$$

$$y_{dir}^{HLOS} = x_{dir}^{HLOS} \quad (4)$$

where  $y_{spd}$  and  $y_{dir}$  are the non-Aeolus wind speed and direction, respectively,  $u_y$  and  $v_y$  are the u- and v-components of the non-Aeolus wind,  $y_{spd}^{HLOS}$  is the non-Aeolus wind velocity projected onto the HLOS direction, and  $y_{dir}^{HLOS}$  is the new direction of the non-Aeolus wind and is equal to the HLOS direction  $x_{dir}^{HLOS}$ . Due to the unique geometry of Aeolus, HLOS winds can have both positive and negative values (unlike non-HLOS wind speeds that can only be positive).

After the HLOS projection (if applicable), a second set of QCs is applied to all winds, i.e., a general wind gross check is applied where absolute wind velocity differences between the Driver and Dependent observations are rejected if they exceed 25 m/s. This threshold was chosen to omit extraneous outliers and follows other established collocation practices (e.g., Daniels et al., 2012). The gross check is applied to all dataset pairings selected for comparison, after which the QC'd Dependent winds are compared with the QC'd Driver winds through a series of statistical calculations and plotting functions. Numerous figures are generated that quantify the wind comparisons.

Global and regional statistics are computed simultaneously, where regional refers to the Northern Hemisphere extratropics (NH), Tropics (TR), and Southern Hemisphere extratropics (SH) all bordered by  $|30^\circ|$  latitude. Seasonal statistics may be computed by rerunning the plotting tool for different date ranges. In addition, AMV wind type comparisons between infrared (IR), visible, water vapor cloudy sky (WVcloud), and water vapor clear sky (WVclear) AMVs are automatically performed only if the user selects an AMV dataset for analysis. Further, the

plotting tool must be run once for each Aeolus wind regime, if statistics for more than one Aeolus wind regime are desired.

Users can modify numerous parameters in the main bash script that initiates the plotting tool:

**MAINSRIPT\_for\_\_Plotting.bash**

A full list of parameters the user can change in this script is provided in Table 8.

**Table 8: List of parameters the user can modify in the main bash script for the plotting tool (MAINSRIPT\_for\_\_Plotting.bash).**

<b>Table 8</b> Parameter	Type	Description
<b>yyyyS</b>	String	Start date year. Must be 4 digits.
<b>mmS</b>	String	Start date month. Must be 2 digits.
<b>ddS</b>	String	Start date day. Must be 2 digits.
<b>hhS</b>	String	Start date hour. Must be 2 digits.
<b>yyyyE</b>	String	End date year. Must be 4 digits.
<b>mmE</b>	String	End date month. Must be 2 digits.
<b>ddE</b>	String	End date day. Must be 2 digits.
<b>hhE</b>	String	End date hour. Must be 2 digits.
<b>archive_parent</b>	String	Full path to the home directory (parent directory of <b>/wind_datasets</b> and <b>/collocation</b> directories) on your local machine.
<b>dir_home</b>	String	Full path to location of <b>MAINSRIPT_for__Plotting.bash</b>
<b>dir_in</b>	String	Full path to location where collocation index files are located.
<b>dep_dset</b>	String	Array of strings listing Dependent dataset names
<b>dep_dset_qc</b>	String	Array of strings listing whether QC was applied to Dependent datasets prior to collocation. Must be the same length as <b>dep_dset</b> . <i>Options available: <b>NoQC</b> (QC was not applied to dataset prior to collocation), and <b>QC</b> (QC was applied).</i>
<b>drv_dset</b>	String	String listing Driver dataset.
<b>drv_dset_qc</b>	String	String indicating whether QC was applied to Driver dataset prior to collocation. Must be the same length as <b>drv_dset</b> . <i>Options available: <b>NoQC</b> (QC was not applied to dataset prior to collocation), and <b>QC</b> (QC was applied).</i>
<b>filename_segment</b>	String	Segment of the collocation index filename listing the required Driver dataset information. <b>USER SHOULD NOT MODIFY THIS PARAMETER.</b>
<b>dir_out</b>	String	Full path to location where all figures are to be saved.

<b>Table 8</b> Parameter	Type	Description
<b>superob_choice</b>	Integer	Choice to super-ob (average) multiple collocations per Driver observation, or use all collocations for statistical analysis. <i>Options available: -1 (use all collocations), or 0 (super-ob)</i>
<b>account</b>	String	HPC account. This is the account you the user are allowed to use.
<b>qos</b>	String	HPC QoS. This is for setting SLURM jobs.
<b>partition</b>	String	HPC partition. This is for setting SLURM jobs.
<b>timelimit</b>	String	HPC runtime limit for SLURM jobs.
{load modules}	–	A series of module load commands needed to run the collocation tool. Please check that the following modules are loaded properly as per your HPC system's protocols: Hdf, hdf5, netcdf4, udunits2, ncview, nco, ncl

*Notes on super-obbing (superob\_choice = 0):*

- The term “super-obbing” with respect to SAWC refers to the averaging of all matched winds from a Dependent dataset with the same Driver observation, resulting in one Dependent observation (representing the mean of the multiple collocations) for every Driver observation. For example, say you are collocating Aeolus winds (Driver) and AMVs (Dependent), and 25 AMVs pass all collocation criteria and are subsequently collocated with the same Driver observation: If super-obbing is chosen, the 25 AMVs are all averaged together prior to any analysis or visualization taking place.
- Super-obbing the observations reduces the sizes of the matched Driver and Dependent arrays, and may in turn decrease the processing time for analysis and plotting.

All figures generated by the plotting tool can be found in the **dir\_out** directory (Table 8). Descriptions of the types of figures generated by the plotting tool are listed in Table 9.

**Table 9: List of types of figures generated by the plotting tool.**

<b>Table 9</b> Start of Figure Name	Description
<b>DENSITY_SCATTER</b>	Density scatterplots comparing one collocated variable from one Dependent dataset to the Driver. <i>Note: If the differences are statistically significant (at 95% level), a line of text will appear on the figure stating as much.</i>
<b>HIST</b>	Histograms displaying the collocation differences (time differences, pressure/height difference, or collocation distance) for all Dependent datasets vs the Driver.
<b>MAP_2Dcontour</b>	2D cylindrical equidistant map displaying a variable summed (observation number density) or averaged (wind speed, difference) per cell on a lat/lon grid within a specified vertical layer, for one Dependent dataset compared to the Driver.



<b>Table 9</b> <b>Start of Figure Name</b>	<b>Description</b>
<b>MAP_2Dpoints</b>	2D cylindrical equidistant map displaying a variable at all observed lat/lon locations (not transformed to a lat/lon grid) within a specified vertical layer, for one Dependent dataset compared to the Driver.
<b>MAP_3D</b>	3D projection of a variable at all observed 3D locations, for one Dependent dataset compared to the Driver.
<b>MAP_CE.Match_Locations</b>	2D cylindrical equidistant map of observation locations, for all datasets.
<b>MAP_Ortho.Match_Locations</b>	Same as <b>MAP_CE</b> but on a polar projection.
<b>MAP_Ortho_NorthPole</b>	Same as <b>MAP_2Dcontour</b> but on a polar projection over the North Pole.
<b>MAP_Ortho_SouthPole</b>	Same as <b>MAP_2Dcontour</b> but on a polar projection over the South Pole.
<b>MAP_Rotate</b>	Rotating globe displaying some variable (e.g., number density) for one dataset.
<b>MAP_Rotate.Match_Locations</b>	Rotating globe displaying observation locations, for all datasets.
<b>SCATTER</b>	Scatterplot comparing one collocated variable from all Dependent datasets to the Driver.
<b>STATSvsWSPD</b>	Wind speed difference, standard deviation (SD) of the difference, and SD of each dataset vs Driver wind speed, with corresponding observation count, per Driver wind speed bin for all Dependent datasets. <i>Note:</i> If differences between the Driver and a Dependent dataset are statistically significant (at 95% level) for a certain Driver wind speed bin, a solid circle will appear at that speed along the Dependent difference line.
<b>TIME_SERIES</b>	Time series displaying the mean wind differences (Dependent - Driver), correlation coefficients, SDs of the differences, root mean square differences (RMSDs), and observation counts per day (24-h period) for all Dependent datasets.
<b>VERT_DISTR</b>	Mean vertical distribution of wind speed differences (Dependent - Driver), SDs of the differences, and the corresponding observation counts, per vertical bin for all Dependent datasets. <i>Note:</i> If the differences between the Driver and a Dependent dataset are statistically significant (at 95% level) at a certain vertical level, a solid circle will appear at that level along the corresponding Dependent dataset profile line.
<b>VERTvsLAT_LonRange</b>	Vertical cross section (pressure/height vs latitude) displaying a variable summed (e.g., observation number density) or averaged (e.g., wind speed, difference) per grid cell within a specified longitude range, for one Dependent dataset compared to the Driver.
<b>VERTvsLAT_ZonalMean</b>	Same as <b>VERTvsLAT_LonRange</b> but averaged for all longitudes (i.e., zonal mean).
<b>VERTvsLON_MeridMean</b>	Vertical cross section (pressure/height vs latitude) displaying a variable summed (e.g., observation number density) or averaged (e.g., wind speed, difference) per grid cell averaged over all latitudes (i.e., meridional mean), for one Dependent dataset compared to the Driver.

<b>Table 9</b> Start of Figure Name	Description
<b>VERTvsTIME</b>	Pressure/height vs time figure displaying a variable summed (e.g., observation number density) or averaged (e.g., wind speed, difference) per grid cell for each day (24-h period), for one Dependent dataset compared to the Driver

Table 10 lists all possible abbreviations that are used to make up a figure's file name.

**Table 10: List of file name abbreviations for figure files generated by the plotting tool.**

<b>Table 10</b> Abbreviation	Description
<b>#{All_DEPs}</b>	A list of all Dependent datasets delimited by an underscore “_” <i>Example: Aircraft_Loon_Sonde</i>
<b>#{DateRange}</b>	Period of time the figure represents. There is one option: <b>#{StartDate}-#{EndDate}</b>
<b>#{DSETabbrev}</b>	Indicates the dataset that is plotted. There are several options: <ul style="list-style-type: none"> <li>○ <b>DRV</b>: The Driver listed as <b>#{DRV_DSET}</b> in the file name is plotted</li> <li>○ <b>DEP</b>: The Dependent listed as <b>#{DEP_DSET}</b> in the file name is plotted</li> <li>○ <b>DEP-DRV</b>: The difference (<b>#{DEP_DSET}</b> minus <b>#{DRV_DSET}</b>) is plotted</li> </ul>
<b>#{LatPOV}</b>	Latitude on which a rotating globe's point-of-view is centered.
<b>#{Level}</b>	Represents the mid-level of the vertical layer in question. <i>Examples: 7.0km, 70.0hPa</i>
<b>#{LonRange}</b>	Range of longitudes over which variables are summed or averaged in vertical cross section figures. <i>Example: 30.0-60.0</i>
<b>#{Pole}</b>	Refers to the polar perspective of the projection. There are two options: <b>NorthPole</b> , <b>SouthPole</b>
<b>#{Region}</b>	Indicates region. There are four options: <b>Global</b> , <b>NH</b> , <b>TR</b> , <b>SH</b> , where NH stands for Northern Hemisphere, TR for Tropics, and SH for Southern Hemisphere.
<b>#{ScatterPlots}</b>	The plotting tool generates two types of scatter plots: <b>DENSITY_SCATTER</b> , <b>SCATTER</b>
<b>#{StartDate}, #{EndDate}</b>	<b>#{StartDate}</b> and <b>#{EndDate}</b> have the same form <b>#{YYYY}#{MM}#{DD}#{hh}</b> and correspond to the start date and end date, respectively, that the user specifies in the main plotting bash script. <i>Example: 2019090100</i>
<b>#{SuperobOption}</b>	<b>#{SuperobOption}</b> indicates whether the observations were super-obbed prior to statistical analysis and plotting. There are two options: <ul style="list-style-type: none"> <li>○ <b>SuperOb</b>: The observations were super-obbed</li> <li>○ <b>AllMatches</b>: The observations were not super-obbed; rather, all matches were included in the analysis</li> </ul>

Table 10 Abbreviation	Description
<b><code>\${VertLatType}</code></b>	Refers to the type of pressure/height vs latitude plot. There are two options: <b><code>LonRange\${LonRange}</code></b> , <b><code>ZonalMean</code></b>

File name formats for each type of figure are defined below. Note that the abbreviation **`${Variable}`** is unique for each figure type.

- There are two file name formats for the figures **DENSITY\_SCATTER** or **SCATTER**:
  1. **`${ScatterPlots}.${Variable}.${Region}.${DateRange}.x_${DRV_DSET}.y_${DEP_DSET}.${SuperobOption}.png`**
  2. **`${ScatterPlots}.${Variable}.AMVtype.${Region}.${DateRange}.x_${DRV_DSET}.y_${DEP_DSET}.${SuperobOption}.png`**
    - The first displays the comparison of **`${Variable}`** between the Driver and one Dependent dataset.
    - The second displays the comparison of **`${Variable}`** corresponding to an AMV type between the Driver and one Dependent dataset if an AMV dataset is selected as either the Driver or a Dependent dataset.
    - There are several options for **`${Variable}`**: **HLOS\_Velocity**, **WindSpeed**, **Height**, **Pressure**.
    - All other abbreviations have been previously defined.
- There are two file name formats for **HIST** figures:
  1. **`HIST.Match_${Variable}.${Region}.${DateRange}.x_${DRV_DSET}.y_${DEP_DSET}.png`**
  2. **`HIST.Match_${Variable}.AMVtype.${Region}.${DateRange}.x_${DRV_DSET}.y_${DEP_DSET}.png`**
    - The first displays **`${Variable}`** for all Dependent datasets specified in the main plotting bash script.
    - The second displays **`${Variable}`** for all AMV types, and is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
    - Here, **`${Variable}`** corresponds to the collocation difference variables. There are several options:
      - **Distance**: Collocation distance between Dependent and Driver
      - **HeightDiff**: Height difference (Dependent - Driver), if the vertical variable is *height*
      - **PressureDiff**: Pressure difference (Dependent - Driver), if the vertical variable is *pressure*
      - **TimeDiff**: Time difference (Dependent - Driver)
    - All other abbreviations have been previously defined.

- There are four file name formats for **MAP\_2Dcontour** figures:
  1. **MAP\_2Dcontour.Nobs\_Density.\${DSETabbrev}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  2. **MAP\_2Dcontour.\${Variable}.\${DSETabbrev}.\${Level}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  3. **MAP\_2Dcontour.Nobs\_Density.\${DSETabbrev}.AMVtype.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  4. **MAP\_2Dcontour.\${Variable}.\${DSETabbrev}.AMVtype.\${Level}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
    - The first displays the number of density of collocations for either the Driver or a Dependent dataset per cell on a lat/lon grid.
    - The second displays a collocated variable (e.g., wind speed, wind difference) averaged per cell on a lat/lon grid within a specified vertical layer. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
    - The third displays the number density of collocations for an AMV type per cell on a lat/lon grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
    - The fourth displays the wind speed or wind difference for an AMV type averaged per cell on a lat/lon grid within a specified vertical layer. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
    - There are several options for **\${Variable}**: **Wind** (i.e., speed or velocity), **Wind\_Diff**, **Wind\_SDdiff**
    - All other abbreviations have been previously defined.
- There are two file name formats for **MAP\_2Dpoints** figures:
  1. **MAP\_2Dpoints.\${Variable}.\${DSETabbrev}.\${Level}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  2. **MAP\_2Dpoints.\${Variable}.\${DSETabbrev}.AMVtype.\${Level}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
    - The first displays a collocated variable (e.g., wind speed, wind difference) corresponding to all observations within a specified vertical layer.
    - The second displays all collocated AMVs of one type within a specified vertical layer.
    - There are two options for **\${Variable}**: **Wind** (i.e., speed or velocity), **Wind\_Diff**
    - All other abbreviations have been previously defined.
- There are two file name formats for **MAP\_3D** figures:
  1. **MAP\_3D.\${Variable}.\${DSETabbrev}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  2. **MAP\_3D.AMVtype.\${Variable}.\${DSETabbrev}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
    - The first displays a collocated variable (e.g., wind speed, wind difference) corresponding to all observations at their 3-dimensional locations.

- The second displays all collocated AMVs of one type at their 3-dimensional locations.
  - There are several options for **Variable**: **Wind** (i.e., speed or velocity), **Wind\_Diff**
  - All other abbreviations have been previously defined.
- There are two file name formats for **MAP\_CE.Match\_Locations** figures:
  1. **MAP\_CE.Match\_Locations.DateRange.x\_Drv\_DSET.y\_All\_DEPs.SuperobOption.png**
  2. **MAP\_CE.Match\_Locations.AMVtype.DateRange.x\_Drv\_DSET.y\_All\_DEPs.SuperobOption.png**
  - The first displays the lat/lon locations of collocated observations from all Dependent datasets.
  - The second displays the lat/lon locations of all available types of collocated AMVs.
  - All other abbreviations have been previously defined.
- There are two file name formats for **MAP\_Ortho.Match\_Locations** figures:
  1. **MAP\_Ortho.Match\_Locations.Pole.DateRange.x\_Drv\_DSET.y\_All\_DEPs.SuperobOption.png**
  2. **MAP\_Ortho.Match\_Locations.Pole.AMVtype.DateRange.x\_Drv\_DSET.y\_All\_DEPs.SuperobOption.png**
  - The first displays the lat/lon locations of collocated observations from all Dependent datasets on the corresponding polar projection **Pole**.
  - The second displays the lat/lon locations of all available types of collocated AMVs on the corresponding polar projection **Pole**.
  - All other abbreviations have been previously defined.
- There are four file name formats for **MAP\_Ortho.Pole**:
  1. **MAP\_Ortho.Pole.Nobs\_Density.DSETabbrev.DateRange.x\_Drv\_DSET.y\_Dep\_DSET.SuperobOption.png**
  2. **MAP\_Ortho.Pole.Variable.DSETabbrev.Level.DateRange.x\_Drv\_DSET.y\_Dep\_DSET.SuperobOption.png**
  3. **MAP\_Ortho.Pole.Nobs\_Density.DSETabbrev.AMVtype.DateRange.x\_Drv\_DSET.y\_Dep\_DSET.SuperobOption.png**
  4. **MAP\_Ortho.Pole.Variable.DSETabbrev.AMVtype.Level.DateRange.x\_Drv\_DSET.y\_Dep\_DSET.SuperobOption.png**
  - The first displays the number of density of collocations for either the Driver or a Dependent dataset per cell on a lat/lon grid.
  - The second displays a collocated variable (e.g., wind speed, wind difference) averaged per cell on a lat/lon grid within a specified vertical layer. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.

- The third displays the number density of collocations for an AMV type per cell on a lat/lon grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - The fourth displays the wind speed or wind difference for an AMV type averaged per cell on a lat/lon grid within a specified vertical layer. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - There are several options for **Variable**: **Wind** (i.e., speed or velocity), **Wind\_Diff**, **Wind\_SDdiff**
  - All other abbreviations have been previously defined.
- There are two file name formats for **MAP\_Rotate**:
    1. **MAP\_Rotate.Nobs\_Density.DSETabbrev.centerlat.LatPOV.DateRange.x\_DRV\_DSET.y\_DEP\_DSET.SuperobOption.gif**
    2. **MAP\_Rotate.Nobs\_Density.DSETabbrev.AMVtype.centerlat.LatPOV.DateRange.x\_DRV\_DSET.y\_DEP\_DSET.SuperobOption.gif**
    - The first displays the number of density of collocations for either the Driver or a Dependent dataset per cell on a lat/lon grid on a rotating globe centered on the latitude **LatPOV**.
    - The second displays the number density of collocations for an AMV type per cell on a lat/lon grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset on a rotating globe centered on the latitude **LatPOV**.
    - All other abbreviations have been previously defined.
- There are two file name formats for **MAP\_Rotate.Match\_Locations**:
    1. **MAP\_Rotate.Match\_Locations.centerlat.LatPOV.DateRange.x\_DRV\_DSET.y\_All\_DEPs.SuperobOption.gif**
    2. **MAP\_Rotate.Match\_Locations.AMVtype.centerlat.LatPOV.DateRange.x\_DRV\_DSET.y\_All\_DEPs.SuperobOption.gif**
    - The first displays the lat/lon locations of collocated observations from all Dependent datasets on a rotating globe centered on the latitude **LatPOV**.
    - The second displays the lat/lon locations of all available types of collocated AMVs on a rotating globe centered on the latitude **LatPOV**.
    - All other abbreviations have been previously defined.
- There are two file name formats for the figures **STATSvsWSPD**:
    1. **STATSvsWSPD.Variable.DEP-DRV.Region.DateRange.x\_DRV\_DSET.y\_DEP\_DSET.SuperobOption.png**
    2. **STATSvsWSPD.Variable.DEP-DRV.AMVtype.Region.DateRange.x\_DRV\_DSET.y\_DEP\_DSET.SuperobOption.png**
    - The first displays the difference of **Variable** between the Driver and one Dependent dataset, the SD of that difference, and collocated observation count per wind speed/velocity bin.

- The second displays the difference of **`\${Variable}`** corresponding to an AMV type between the Driver and one Dependent dataset, the SD of that difference, and collocated observation count per wind speed/velocity bin. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - There are two options for **`\${Variable}`**:
    - **HLOS\_Velocity\_Diff**: This is only created if the Aeolus dataset is selected as either the Driver or a Dependent dataset.
    - **WindSpeed\_Diff**: This is only created if the Aeolus dataset is neither the Driver nor a Dependent dataset.
  - All other abbreviations have been previously defined.
- There are two file name formats for the figures **TIME\_SERIES**:
    1. **TIME\_SERIES`\${Variable}`.DEP-DRV`\${Region}``\${DateRange}`.x\_`\${DRV\_DSET}`.y\_`\${DEP\_DSET}``\${SuperrobOption}`.png**
    2. **TIME\_SERIES`\${Variable}`.DEP-DRV.AMVtype`\${Region}``\${DateRange}`.x\_`\${DRV\_DSET}`.y\_`\${DEP\_DSET}``\${SuperrobOption}`.png**
    - The first displays the difference of **`\${Variable}`** between the Driver and all Dependent datasets, the SD of that difference, correlation coefficients between the Driver and Dependent datasets, RMSD, and collocated observation count per wind speed/velocity bin.
    - The second displays the difference of **`\${Variable}`** corresponding to all AMV types between the Driver and one Dependent dataset, the SD of that difference, correlation coefficients between the Driver and Dependent datasets, RMSD, and collocated observation count per wind speed/velocity bin. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
    - There are two options for **`\${Variable}`**:
      - **HLOS\_Velocity\_Diff**: This is only created if the Aeolus dataset is selected as either the Driver or a Dependent dataset.
      - **WindSpeed\_Diff**: This is only created if the Aeolus dataset is neither the Driver nor a Dependent dataset.
      - **Pressure\_Diff**: This is created using all Dependent datasets with *pressure* as the vertical variable.
      - **Height\_Diff**: This is created using all Dependent datasets with *height* as the vertical variable.
    - All other abbreviations have been previously defined.
  - There are four file name formats for **VERT\_DISTR**:
    1. **VERT\_DISTR.HLOS\_Velocity\_Diff.DEP-DRV`\${Region}``\${DateRange}`.x\_`\${DRV\_DSET}`.y\_`\${All\_DEPs}``\${SuperrobOption}`.png**
    2. **VERT\_DISTR.WindSpeed\_Diff.DEP-DRV`\${Region}``\${DateRange}`.x\_`\${DRV\_DSET}`.y\_`\${All\_DEPs}``\${SuperrobOption}`.png**
    3. **VERT\_DISTR.HLOS\_Velocity\_Diff.DEP-DRV.AMVtype`\${Region}``\${DateRange}`.x\_`\${DRV\_DSET}`.y\_`\${All\_DEPs}``\${SuperrobOption}`.png**



4. **VERT\_DISTR.WindSpeed\_Diff.DEP-DRV.AMVtype.\${Region}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${All\_DEPs}.\${SuperobOption}.png**
  - The first displays the average HLOS wind velocity difference, SD of the difference, and observation number density per pressure/height bin (prescribed in **MAIN.plotting\_code.py** and **MAIN.plotting\_code.AMVtypes.py**). This is only created if the Aeolus dataset is selected as either the Driver or a Dependent dataset.
  - The second displays the average wind speed difference, SD of the difference, and observation number density per pressure/height bin. This is only created if the Aeolus dataset is neither the Driver nor a Dependent dataset.
  - The third displays the average HLOS wind velocity difference, SD of the difference, and observation number density per pressure/height bin corresponding to each AMV type. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset, and the Aeolus dataset is selected as either the Driver or a Dependent dataset.
  - The fourth displays the average HLOS wind velocity difference, SD of the difference, and observation number density per pressure/height bin corresponding to each AMV type. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset, and the Aeolus dataset is neither the Driver nor a Dependent dataset.
  - All other abbreviations have been previously defined.
- There are four file name formats for **VERTvsLAT\_LonRange** and **VERTvsLAT\_ZonalMean**:
  1. **VERTvsLAT\_\${VertLatType}.Nobs\_Density.\${DSETabbrev}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  2. **VERTvsLAT\_\${VertLatType}.\${DSETabbrev}.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  3. **VERTvsLAT\_\${VertLatType}.Nobs\_Density.\${DSETabbrev}.AMVtype.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  4. **VERTvsLAT\_\${VertLatType}.\${DSETabbrev}.AMVtype.\${DateRange}.x\_\${DRV\_DSET}.y\_\${DEP\_DSET}.\${SuperobOption}.png**
  - The first displays the number of density of collocations for either the Driver or a Dependent dataset per cell on a pressure/height vs latitude grid.
  - The second displays a collocated variable (e.g., wind speed, wind difference) averaged per cell on a pressure/height vs latitude grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - The third displays the number density of collocations for one AMV type per cell on a pressure/height vs latitude grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - The fourth displays the wind speed or wind difference for one AMV type averaged per cell on a pressure/height vs latitude grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.



- There are two options for **VertLatType**:
  - **LonRangeLonRange**: Average over the longitudes in range **LonRange**
  - **ZonalMean**: Average over all longitudes
- There are several options for **Variable**: **Wind** (i.e., speed or velocity), **Wind\_Diff**
- All other abbreviations have been previously defined.
- There are four file name formats for **VERTvsLON\_MeridMean**:
  5. **VERTvsLON\_MeridMean.Nobs\_Density.DSETabbrev.DateRange.x\_**  
**DRV\_DSET.y\_DEP\_DSET.SuperobOption.png**
  6. **VERTvsLON\_MeridMean.DSETabbrev.DateRange.x\_**  
**DRV\_DSET.y\_DEP\_DSET.SuperobOption.png**
  7. **VERTvsLON\_MeridMean.Nobs\_Density.DSETabbrev.AMVtype.**  
**DateRange.x\_DRV\_DSET.y\_DEP\_DSET.SuperobOption.png**
  8. **VERTvsLON\_MeridMean.DSETabbrev.AMVtype.DateRange.x\_**  
**DRV\_DSET.y\_DEP\_DSET.SuperobOption.png**
  - The first displays the number of density of collocations for either the Driver or a Dependent dataset per cell on a pressure/height vs longitude grid.
  - The second displays a collocated variable (e.g., wind speed, wind difference) averaged per cell on a pressure/height vs longitude grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - The third displays the number density of collocations for one AMV type per cell on a pressure/height vs longitude grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - The fourth displays the wind speed or wind difference for one AMV type averaged per cell on a pressure/height vs longitude grid. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - There are several options for **Variable**: **Wind** (i.e., speed or velocity), **Wind\_Diff**
  - All other abbreviations have been previously defined.
- There are two file name formats for **VERTvsTIME**:
  1. **VERTvsTIME.Nobs\_Density.DSETabbrev.Region.DateRange.x\_**  
**DRV\_DSET.y\_DEP\_DSET.SuperobOption.png**
  2. **VERTvsTIME.Nobs\_Density.DSETabbrev.AMVtype.Region.**  
**DateRange.x\_DRV\_DSET.y\_DEP\_DSET.SuperobOption.png**
  - The first displays the number of density of collocations for either the Driver or a Dependent dataset per cell on a pressure/height vs time grid.
  - The second displays the number of density of collocations for either the Driver or a Dependent dataset per cell on a pressure/height vs time grid, for one AMV type. This is only created if an AMV dataset is selected as either the Driver or a Dependent dataset.
  - Here, there are only two options for **DSETabbrev**: **DRV, DEP**

## 4.2 Installation

The collocation application should be installed on an HPC compute node and not a gateway node.

It is recommended that you download a copy of the data you will need to your local machine prior to running the collocation application. All the data should be located in the same **archive\_parent** (i.e., home) directory (see Table 6) and should have the same subdirectory structure as the online archive (**\${MAIN\_WIND}**).

Follow these steps to install the collocation application:

1. Download the **\*.tar.gz** file from the archive in **\${MAIN\_COL}/application** and place it in the directory where you plan to install and use the collocation application.
2. Unzip the file with the following command:  

```
gunzip *.tar.gz
```
3. Untar the file with the following command:  

```
tar -xf *.tar
```
4. You should now be able to see and access the **SAWC\_Collocation\_Application\_v###** directory (where **###** refers to the version number; it is recommended that the latest version be used). The main directories within the collocation application are listed in Table 11.

Table 11: Collocation application directory abbreviations, full paths, and descriptions.

Table 11 Path Abbreviation	Full Path (if applicable)	Description
<b>\${COL_APP}</b>	<b>/SAWC_Collocation_Application_v###</b>	Collocation application main directory. Contains main bash scripts and SLURM job scripts. Note that <b>v###</b> indicates the version number of the collocation application
<b>\${SRC}</b>	<b>\${COL_APP}/src</b>	Directory containing all source code

5. Now it is time to set up your conda environment. You should only have to do this once on a single HPC system; even if you download additional versions of SAWC in the future onto the same system, you should not need to do this again.
  - a. In your **.bashrc** file in your home directory (or **.cshrc**, etc.), initialize your conda environment (*Below is an example in bash syntax. You will need to change the **/opt** paths to point to what is available on your machine*):

---

```
__conda_setup="$('/opt/miniconda/3.8-s4/bin/conda' 'shell.bash' 'hook' 2> /dev/null)"
if [ $? -eq 0 ]; then
    eval "$__conda_setup"
else
    if [ -f "/opt/miniconda/3.8-s4/etc/profile.d/conda.sh" ]; then
        . "/opt/miniconda/3.8-s4/etc/profile.d/conda.sh"
    else
        export PATH="/opt/miniconda/3.8-s4/bin:$PATH"
    fi
fi
unset __conda_setup
```

- b. Save your **.bashrc** file and source it to enact your new changes:

```
source .bashrc
```

- c. Load the miniconda 2.7 module, e.g.:

```
module load miniconda/2.7-base
```

- d. Build upon your initialized conda environment using the file **SAWC\_conda\_env.yml** located in **\${SRC}**:

```
conda env create -f ${SRC}/SAWC_conda_env.yml
```

- e. Activate this environment:

```
conda activate SAWC_conda_env
```

*Note:* If the above command doesn't work, try:

```
source activate SAWC_conda_env
```

- f. Check that the **SAWC\_conda\_env** environment is active (note the double dash -- in the command below):

```
conda info --envs
```

The environment **SAWC\_conda\_env** should be listed on your screen.

You are now ready to run the application! (See [Section 4.3](#))

## 4.3 Running the Application

The collocation application should be run on an HPC compute node and not a gateway node.

### 4.3.1 Collocation Tool

#### 4.3.1.1 Order of Processing Steps

Figure 7 illustrates the collocation tool's order of processing steps.

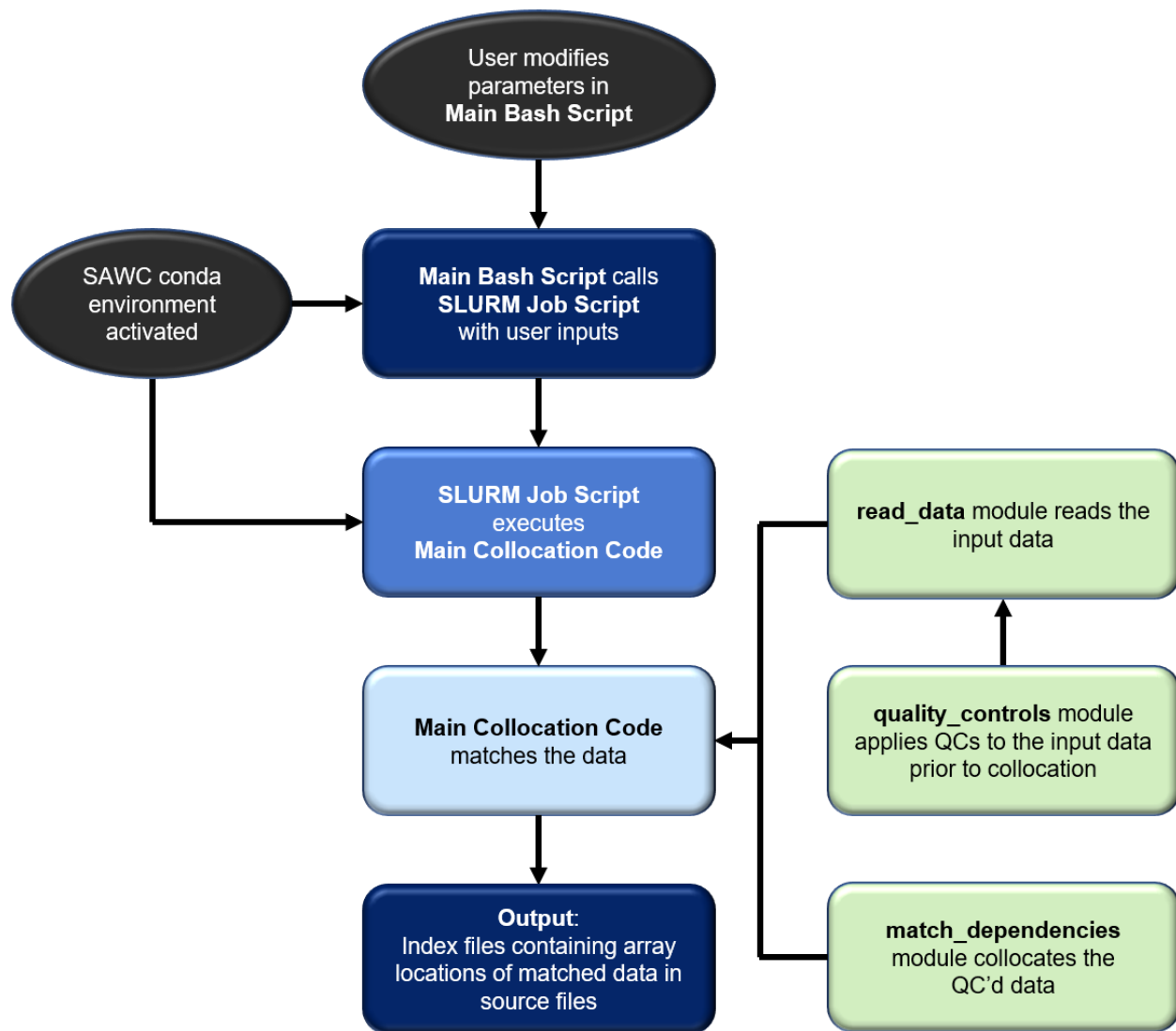


Figure 7: Flow chart illustrating the order of processing steps for the collocation tool.

The script **MAINSRIPT\_for\_\_Collocation.bash** initiates the collocation between wind datasets, and is located in **\${COL\_APP}**. Specifically, this script activates the conda environment (**SAWC\_conda\_env**, as per the \*.yaml file) used to run the collocation tool in the

background, and calls the SLURM script **run\_collocation\_code.job** that initiates all jobs for the tool, with each job name beginning with **COL\***. Each job calls the Python script **MAIN.match\_driver\_dependents.py** located in the **\${COL\_APP}/src** directory. This script reads in the source data to be collocated, matches each Dependent dataset with the Driver, and writes the matched indices to one collocation index file in NetCDF per Driver-Dependent combination for every date **\${YYYY}\${MM}\${DD}\${hh}** in the user-specified time period. One job is initiated for each NWP analysis time (00, 06, 12, or 18 UTC), so users will have at minimum 4 jobs in their queue for a single day (8 jobs for a single day if Aeolus is to be matched: 4 for Rayleigh-clear comparisons, and 4 for Mie-cloudy comparisons). Each job has a corresponding log file, with file names beginning with **LOG\_COL\_\***.

The **read\_data.py** module contains all functions that read in each of the wind datasets. The **match\_dependencies.py** module contains all dependent functions for the collocation tool. The **quality\_controls.py** module contains all functions that perform quality controls on each of the wind datasets prior to collocation.

#### 4.3.1.2 Steps for Running the Collocation Tool

Follow these steps to run the collocation tool for matching  $n$  Dependent datasets against one Driver dataset:

1. Open **MAINSRIPT\_for\_\_Collocation.bash** and set all parameters listed in Table 6.
2. Run the tool by entering on the command line:

**./MAINSRIPT\_for\_\_Collocation.bash**

All jobs run in the background. See Appendix B for tips on how to check the status of the job.

3. Once a job completes, you will find the corresponding index file in the output directory **dir\_out** that you previously prescribed in **MAINSRIPT\_for\_\_Collocation.bash**.

### 4.3.1.3 Advanced Modifications: Step-by-step Examples

Below are some step-by-step examples of offline modifications the user can make to the collocation tool's source code:

1. *Enabling the collocation tool to include and match a wind dataset not available in the archive*
  - a. Make sure the new dataset files are all in NetCDF, ideally in 6-hr format centered around the NWP analysis times 00, 06, 12, and 18 UTC.
  - b. In the collocation tool's main bash script **MAINSCRIPT\_for\_\_Collocation.bash**:
    - i. Add the new dataset name **`${NewDataset}`** to **dataset\_names** as either the Driver (the first name in the list) or a Dependent dataset (the 2nd, 3rd, etc. name in the list), where **`${NewDataset}`** is the name the user chooses to represent the new dataset.
    - ii. If **`${NewDataset}`** is a Dependent dataset, add an element corresponding to the new dataset's collocation criteria to each **\*\_max** array.
    - iii. Add an element corresponding to the new dataset's QC flag to the **qc\_flags** array.
    - iv. If **`${NewDataset}`** is an AMV dataset, add an element corresponding to the new dataset's AMV QI flag to the **amv\_qi\_flags** array, as well as an element corresponding to the new dataset's QI choice to the **qi\_choices** array.
  - c. In **read\_data.py**, copy and paste one of the functions to the end of the script.  
*Suggestion:* If the new data in each NetCDF file are organized into groups (as is common for HDF or NetCDF-4), copy the **read\_aircraft** function. If not, copy the **read\_loon** function.
    - i. Rename the new function as **read\_`${NewDataset}`**.
    - ii. Change the **\*\_path**, **\*\_path\_B4**, and **\*\_path\_A** variables to the paths to your new dataset for the current date, date before, and date after, respectively. Follow the syntax of these paths in the original **read\_\*** function.  
*Note:* If all dates are located in the same directory, then **\*\_path**, **\*\_path\_B4**, and **\*\_path\_A** are all the same.
    - iii. Change/add the variable names **\*\_var** as they are listed in the new dataset's NetCDF files.
    - iv. Below the line **# Load dataset** are lines of code corresponding to each variable name, and these lines are grouped. Add/remove lines of code from each of these groups following the variable names listed in *Step iii*.  
*Note:* Check that all new temporary variables are deleted (**del**) where appropriate.
  - d. In **MAIN.match\_driver\_dependents.py**:
    - i. Add the following line to the *Import python modules* section at the top of the script:

**from read\_data import read\_`${NewDataset}`**

- ii. In the if-else block below the line **# Define and load DRIVER dataset**, add another **elif** condition corresponding to the new dataset just above the **else** condition. Follow the syntax of the other **elif** conditions.

*Notes:*

1. Check that the output variables within the new **elif** condition (left side of the equal sign = ) match those listed in the last line of the new function **read\_\${NewDataset}** (on **return** line) in **read\_data.py**.
  2. Check that the input variables within the new **elif** condition (within the parentheses after the function name on the right side of the equal sign) match those listed in parentheses in the first line of the new function **read\_\${NewDataset}** in **read\_data.py** (to the right of the function name).
- iii. Repeat *Step ii* but this time in the if-else block below the line **# Define dependent datasets...** Follow the syntax of the other **elif** conditions in this if-else block.

## 2. Adding/modifying QC functions

- a. In **quality\_controls.py**, copy and paste the function **qc\_amv** to the end of the script.
  - i. Rename the new function as **qc\_\${NewDataset}**, where **\${NewDataset}** is the name the user chooses to represent the new dataset.
  - ii. The **idx** variable contains indices of the data where the input conditions within **np.where()** are **True**; these input conditions are the QC criteria. Change the QC criteria here to correspond to the new dataset.

*Notes:*

1. Write the QC criteria so that you KEEP the data that pass QC. For example, if your max time difference criterion is 60 minutes, the criterion syntax should be: **np.where(time\_max < 60)**
2. Each condition should be separated by a set of parentheses within the **np.where()** function
  - a. An asterisk **\*** between each condition is equivalent to “and”
  - b. A plus sign **+** between each condition is equivalent to “or”

*Example:*

**idx = np.where( ((prs < pmax)\*(prs > pmin))+(wspd>wspdmin) )**

does the following:

It finds and keeps all indices **idx** where ((pressure **prs** is less than maximum pressure **pmax**) AND (**prs** is greater than minimum pressure **pmin**)) OR (wind speed **wspd** is greater than minimum wind speed **wspdmin**)).

- iii. **qc\_list** is a string that lists in words the QC criteria you added in *Step ii*. Change this string to match the QC criteria you added in *Step ii*. This is important as **qc\_list** is included as metadata in the output index files.

- b. In the function `read_${Dataset}` in `read_data.py` that corresponds to the QC function you just made:
- Add the following line to the *Import python modules* section at the top of the script:  

```
from quality_controls import qc_${Dataset}
```
  - In the `if runtime=="match"` block toward the end of the function, add under `if bool_qc` a call to the new function `qc_${Dataset}`, following the syntax below:

```
tindexes,qc_list = qc_${Dataset}(...input variables for new QC
function...)
sindexes = np.asarray(tindexes)
indexes = sindeces.flatten()
d_lat = tdset_lat[indexes]
```

Be sure to add a line like `d_lat=...` for each variable in the dataset.

*Notes:*

- If you are adding a QC function for a new dataset that is not currently available in the archive, make sure you first complete Advanced Modification #1 in this section ([Section 4.3.1.3](#)), then go to Advanced Modification #2b.
- The output variables for the new QC function are **tindexes** and **qc\_list**.
- Check that the input variables (within the parentheses after the function name on the right side of the equal sign in the above example) match those listed in parentheses in the first line of the new function `qc_${NewDataset}` in `quality_controls.py` (to the right of the function name).



### 4.3.2 Plotting Tool

#### 4.3.2.1 Order of Processing Steps

Figure 8 illustrates the plotting tool's order of processing steps.

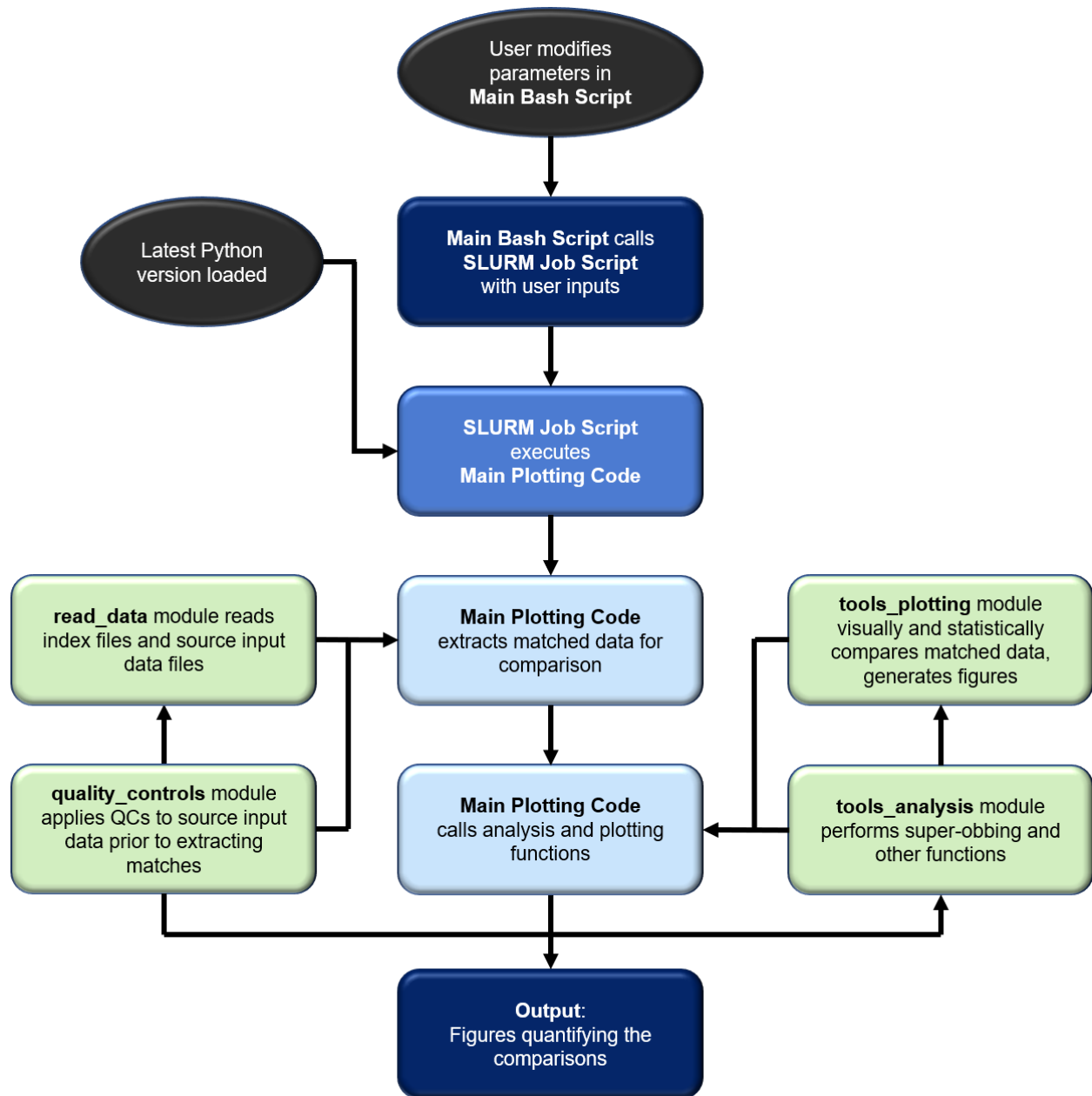


Figure 8: Flow chart illustrating the order of processing steps for the plotting tool.

The script **MAINSRIPT\_for\_Plotting.bash** initiates the comparison of matched observations between the Driver and Dependent datasets tool, and is located in **\${COL\_APP}**. Specifically, this script calls the SLURM script **run\_plotting\_code.job** that initiates one job, with the job

name beginning with **PLOT\***. The job calls the plotting source code **MAIN.plotting\_code.py** (or **MAIN.plotting\_code.AMVtypes.py** which is identical to **MAIN.plotting\_code.py** except that it compares the different AMV types instead of comparing each dataset) located in the **{COL\_APP}/src** directory. The plotting source code reads in the matched indices from the index files within the user-specified time period, and uses them to extract the matched winds from the source data files. Then, the collocated Driver and Dependent winds are compared through visualization and statistical analyses, and numerous figures that quantify the comparisons are generated. The corresponding log file name begins with **LOG\_PLOT\_\***.

The **read\_data.py** module contains the function that reads the index files. The **quality\_controls.py** module contains all functions that perform quality controls on each of the wind datasets prior to any analysis. The **tools\_analysis.py** module contains functions that perform statistical, mathematical, and other analyses to compare the wind datasets. The **tools\_plotting.py** module contains functions that plot the comparisons as output figures.

#### *4.3.2.2 Steps for Running the Plotting Tool*

Follow these steps to run the plotting tool for comparing  $n$  Dependent datasets against one Driver dataset:

1. Open **MAINSRIPT\_for\_\_Plotting.bash** and set all parameters listed in Table 8 (exception: **filename\_segment**).
2. Run the tool by entering on the command line

**./MAINSRIPT\_for\_\_Plotting.bash**

The job runs in the background. See Appendix B for tips on how to check the status of the job.

3. The figures will be saved in the output directory **dir\_out** that you prescribed in **MAINSRIPT\_for\_\_Plotting.bash**.

### 4.3.2.3 Advanced Modifications: Step-by-step Examples

Below are a few examples of more advanced offline modifications the user can make to the plotting tool.

1. *Enabling the plotting tool to include in the intercomparisons a new wind dataset not available in the archive*
  - a. If you haven't already done so, make sure you have already added this new wind dataset to the **read\_data.py** module. See [Section 4.3.1.3](#) Advanced Modifications #1 for details.
  - b. In the plotting tool's main bash script **MAINSRIPT\_for\_Plotting.bash**:
    - i. Add the new dataset name **`${NewDataset}`** to **dep\_dset** if it is a Dependent dataset, or set **drv\_dset=`${NewDataset}`** if it is the Driver, where **`${NewDataset}`** is the name the user chooses to represent the new dataset.
    - ii. If **`${NewDataset}`** is a Dependent dataset, add an element corresponding to the new dataset's QC flag in the **dep\_dset\_qc** array. Your choices are "QC" or "NoQC", and they depend on what is listed in the output index filename next to **`${NewDataset}`**.
  - c. In **MAIN.plotting\_code.py** (and in **MAIN.plotting\_code.AMVtypes.py**):
    - i. Add the following line to the *Import python modules* section at the top of the script:
 

```
from read_data import read_${NewDataset}
```
    - ii. In the if-else block below the line **# Define and load DRIVER dataset**, add another **elif** condition corresponding to the new dataset just above the **else** condition. Follow the syntax of the other **elif** conditions.
 

*Notes:*

      1. Check that the output variables within the new **elif** condition (left side of the equal sign = ) match those listed in the last line of the new function **read\_`${NewDataset}`** (on **return** line) in **read\_data.py**.
      2. Check that the input variables within the new **elif** condition (within the parentheses after the function name on the right side of the equal sign) match those listed in parentheses in the first line of the new function **read\_`${NewDataset}`** in **read\_data.py** (to the right of the function name).
    - iii. Repeat *Step ii* but this time in the if-else block within the section **# Define dependent datasets** below the line **# read dependent dataset**. Follow the syntax of the other **elif** conditions in this if-else block.
2. *Change cross section longitude bins for **VERTvsLAT\_LonRange** plots*
  - a. The arrays **vertcrossLONSmin** and **vertcrossLONSmax** govern the cross section longitude boundary minimums and maximums, respectively. They can be found in **MAIN.plotting\_code.py** (and in **MAIN.plotting\_code.AMVtypes.py**) under the section **# More plotting prep**. Change their values if desired.

Note: **vertcrossLONSmin** and **vertcrossLONSmax** must be the same size, and the value of each min must be less than the corresponding value of each max.

### 3. Change grid resolutions for 2D maps

- a. In **MAIN.plotting\_code.py** (and in **MAIN.plotting\_code.AMVtypes.py**) under the section **# Set global parameters**:
  - i. Modify **latlonstride** (i.e., the step for the horizontal resolution arrays **flats** (latitudes) and **flons** (longitudes)). This can be found below the line **# set horizontal resolution**.
  - ii. Modify **plevstride** (i.e., the step for the vertical *pressure* resolution array **fpres**). Also modify **hlevstride** (i.e., the step for the vertical *height* resolution array **fhgts**). These can be found below the line **# set vertical resolution**.

### 4. Change horizontal extent of 3D maps

- a. In **tools\_plotting.py** within the function **map\_3d\_profile** under **# Define horizontal extent**, modify the values of the array **extent**. **extent** must always have 4 values, which together represent the lower left and upper right corners of the horizontal space to be plotted.  
*Example:* If **extent = [-180, 180, -90, 90]**, the lower left corner is (-180,-90) and the upper right corner is (180, 90), where corner values are defined as (longitude, latitude).

### 5. Change 3D map perspective

- a. In **tools\_plotting.py** within the function **map\_3d\_profile** under **# 3D viewpoint**:
  - i. Modify **azim\_pt**
    1. **azim\_pt** moves the plot's center point left or right. When **extent = [-180, 180, -90, 90]**, **azim\_pt=230** indicates the lower-left corner (-180,-90) is in the center of the plot. If **azim\_pt<230**, the center of the plot moves to the right of this point; if **azim\_pt>230**, the center of the plot moves to the left of this point.
  - ii. Modify **elev\_pt**
    1. **elev\_pt** moves the plot's center point up or down.

### 6. Change plotted vertical levels for 2D maps

- a. In **MAIN.plotting\_code.py** (and in **MAIN.plotting\_code.AMVtypes.py**) under the section **# Set global parameters** below the line **# Vertical levels for plotting**:
  - i. Modify **levsP**. Make sure there is at least one pressure level (units = hPa). This variable is type float.
  - ii. Modify **levsZ**. Make sure there is at least one height level (units = km). This variable is type float.

# Appendices

## A. Acronyms and Abbreviations

3D	3-dimensional
2D	2-dimensional
AMV	Atmospheric motion vector
ASCII	American Standard Code for Information Interchange
BUFR, bufr	Binary Universal Form for the Representation of meteorological data
CIMSS	UW-Madison Cooperative Institute for Meteorological Satellite Studies
CISESS	UMD/ESSIC Cooperative Institute for Satellite Earth System Studies
CODA	ESA Common Data Access
ECMWF	European Centre for Medium-Range Weather Forecasts
EE, EE-format	Earth Explorer format
EMC	Environmental Modeling Center
ESA	European Space Agency
ESSIC	Earth System Science Interdisciplinary Center
FM-A	Aeolus laser nominal flight model A
FM-B	Aeolus laser nominal flight model B (backup)
FTP	File Transfer Protocol
GEO	Geostationary satellite
HLOS	Horizontal Line-of-Sight
IR	Infrared
L2B	Level-2B
LEO	Low Earth-orbiting satellite
LOS	Line-of-Sight
NCEP	National Centers for Environmental Protection
NESDIS	NOAA National Environmental Satellite, Data, and Information Service
NetCDF, netcdf	Network Common Data Form
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical weather prediction
NWS	NOAA National Weather Service
OSTI	NOAA/NWS Office of Science and Technology Integration
prepBUFR	Preprocessed BUFR
QC	Quality control
QI	Quality indicator
SH	Southern Hemisphere
SSEC	Space Science and Engineering Center
STAR	NOAA/NESDIS Center for Satellite Applications and Research
TR	Tropics
UMD	University of Maryland
URL	Uniform Resource Locator (online website)
UW-Madison	University of Wisconsin-Madison
WMO	World Meteorological Organization
WVclear	Water vapor clear sky AMVs
WVcloud	Water vapor cloudy sky AMVs
xml	Extensible Markup Language

## B. Tips, Tricks, and Other Useful Information

1. To view a NetCDF file **`\${FILENAME}`**, open a terminal window and enter either of the following commands on the command line:

```
ncl_filedump `${FILENAME}` ... to view metadata information
ncview `${FILENAME}` ... to view variables graphically
```

Note that the file extension (either **.nc** or **.nc4**, whichever corresponds to the file in question) must be included to execute the above commands.

2. Use the following **squeue** command to check the status of your SLURM jobs.

```
squeue -u `${username}` -o "%.18i %.9P %.30j %.8u %.8T %.10M %.9I %.6D %R"
```

where **`\${username}`** is your username on the HPC account you are using.

*Tip:* Create an alias for the **squeue** command in your **.rc** file in your home directory so you don't have to type it out every time! For example, in a **.bashrc** file, add the following:

```
alias myq='squeue -u YourUserName -o "%.18i %.9P %.30j %.8u %.8T %.10M %.9I %.6D %R"
```

3. If you see the following warning in a **LOG\*** file, do not be alarmed.

**WARNING: missing\_value not used since it cannot be safely cast to variable data type**

It is a standard Python warning letting the user know that the **missing\_value** attribute in a NetCDF file is outside the "valid range" and hence it ignores the "valid range." This has no effect on the application, data, or output.

4. To use the collocation application on your local machine, make sure you modify the bash variables **archive\_parent**, **account**, **qos**, **partition**, and **timelimit**, as well as all **{loaded modules}** according to your machine's protocols, in both main bash scripts:

```
MAINSCRIPT_for__Collocation.bash
MAINSCRIPT_for__Plotting.bash
```

## Publications List

*List of publications that make use of SAWC:*

Hoffman, R. N., Lukens, K. E., Ide, K., and Garrett, K., 2021: A collocation study of atmospheric motion vectors (AMVs) compared to Aeolus wind profiles with a feature track correction (FTC) observation operator. Q. J. R. Meteorol. Soc., 148(742):321–337, Jan. (Part A), <https://doi.org/10.1002/qj.4207>.

Lukens, K. E., Garrett, K., Ide, K., Santek, D., Hoover, B., Huber, D., Hoffman, R. N., and Liu, H., 2023: System for Analysis of Wind Collocations (SAWC): A Novel Archive and Collocation Software Application for the Intercomparison of Winds from Multiple Observing Platforms. In preparation.

Lukens, K. E., Ide, K., Garrett, K., Liu, H., Santek, D., Hoover, B., and Hoffman, R. N., 2022: Exploiting Aeolus level-2b winds to better characterize atmospheric motion vector bias and uncertainty, Atmos. Meas. Tech., 15, 2719–2743, <https://doi.org/10.5194/amt-15-2719-2022>.

Santek, D., Dworak, R., Wanzong, S., Rink, T., Lukens, K., Reiner, S., and García-Pereda, J., 2022: NWC SAF Winds Intercomparison Study Report: 2021. International Winds Working Group, available at [http://cimss.ssec.wisc.edu/iwvg/Docs/CIMSS\\_AMV\\_Comparison\\_2021\\_Report\\_02Nov2022.pdf](http://cimss.ssec.wisc.edu/iwvg/Docs/CIMSS_AMV_Comparison_2021_Report_02Nov2022.pdf).

## References

- Borde, R., Carranza, M., Hautecoeur, O. and Barbieux, K., 2019: Winds of change for future operational AMV at EUMETSAT. *Remote Sensing*, 11(18), 2111. <https://doi.org/10.3390/rs11182111>.
- Borde, R., Hautecoeur, O. and Carranza, M., 2016: EUMETSAT global AVHRR wind product. *Journal of Atmospheric and Oceanic Technology*, 33, 429–438. <https://doi.org/10.1175/JTECH-D-15-0155.1>.
- Bormann, N., Saarinen, S., Kelly, G., and Thepaut, J.-N., 2003: The Spatial Structure of Observation Errors in Atmospheric Motion Vectors from Geostationary Satellite Data, *Mon. Weather Rev.*, 131, 706–718, [https://doi.org/10.1175/1520-0493\(2003\)131<0706:TSSOOE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0706:TSSOOE>2.0.CO;2).
- Boukabara, S. A., Zhu, T., Tolman, H. L., Lord, S., Goodman, S., Atlas, R., et al., 2016: S4: An O2R/R2O infrastructure for optimizing satellite data utilization in NOAA numerical modeling systems. a step toward bridging the gap between research and operations. *Bulletin of the American Meteorological Society*, 97, 2359–2378. <https://doi.org/10.1175/bams-d-14-00188.1>
- Cotton, J., Doherty, A., Lean, K., Forsythe, M., and Cress, A., 2020: NWP SAF AMV monitoring: the 9th Analysis Report (AR9), Tech. rep., NWP SAF, Version 1.0, REF: NWPSAFMO-TR-039, <https://nwp-saf.eumetsat.int/site/monitoring/winds-quality-evaluation/amv/amv-analysis-reports/> (last access: 9 May 2021).
- Dabas, A., Denneulin, M., Flamant, P., Loth, C., Garnier, A. and Dolfi-Bouteyre, A., 2008: Correcting winds measured with a rayleigh doppler lidar from pressure and temperature effects. *Tellus A: Dynamic Meteorology and Oceanography*, 60, 206–215. <https://doi.org/10.1111/j.1600-0870.2007.00284.x>.
- Daniels, J.; Bresky, W.; Wanzong, S.; Velden, C.; and Berger, H., 2012: NOAA/NESDIS/STAR GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document For Derived Motion Winds. Available online: <https://www.star.nesdis.noaa.gov/goesr/docs/ATBD/DMW.pdf> (accessed on 19 July 2023).
- de Kloe, J., Stoffelen, A., Tan, D., Andersson, E., Rennie, M., Dabas, A., Poli, P., and Huber, D., 2020: Aeolus Data Innovation Science Cluster DISC ADM-Aeolus Level-2B/2C Processor Input/Output Data Definitions Interface Control Document. Tech. rep., KMNI, Aeolus, DISC, REF: AED-SDECMWF-L2B-037, <https://earth.esa.int/eogateway/documents/20142/37627/Aeolus-L2B-2C-Input-Output-D-ICD.pdf> (last access: 17 April 2021).
- Durre, I., Vose, R. S., and Wuertz, D. B., 2006: Overview of the Integrated Global Radiosonde Archive. *J. Climate*, 19, 53-68, doi: <https://doi.org/10.1175/JCLI3594.1>.
- European Space Agency (ESA), 2020: Aeolus L2B Earth Explorer dataset, ESA [data set], <https://aeolus-ds.eo.esa.int/oads/access/> (last access: 25 February 2021).



- European Space Agency (ESA), 2021: Summary of Quality of Aeolus Data Products from 2nd Reprocessing Campaign covering June 2019 to October 2020. <https://earth.esa.int/eogateway/documents/20142/0/Aeolus-Summary-Reprocessing-2-DI-SC.pdf> (last access: 13 April 2023).
- Genkova, I., Borde, R., Schmetz, J., Daniels, J., Velden, C., Holmlund, K.: Global atmospheric motion vector intercomparison study. In Proceedings of the 9th International Winds Workshop, Annapolis, MD, USA, 14–18 April 2008. [https://www.researchgate.net/profile/Johannes-Schmetz/publication/237834561\\_GLOBAL\\_ATMOSPHERIC\\_MOTION\\_VECTOR\\_INTERCOMPARISON\\_STUDY/links/0c96052825b3d80e7f000000/GLOBAL-ATMOSPHERIC-MOTION-VECTOR-INTERCOMPARISON-STUDY.pdf](https://www.researchgate.net/profile/Johannes-Schmetz/publication/237834561_GLOBAL_ATMOSPHERIC_MOTION_VECTOR_INTERCOMPARISON_STUDY/links/0c96052825b3d80e7f000000/GLOBAL-ATMOSPHERIC-MOTION-VECTOR-INTERCOMPARISON-STUDY.pdf) (last access: 13 April 2023).
- Genkova, I.; Borde, R.; Schmetz, J.; Velden, C.; Holmlund, K.; Bormann, N.; Bauer, P. Global atmospheric motion vector intercomparison study. In Proceedings of the 10th International Winds Workshop, Tokyo, Japan, 22–26 February 2010. <http://cimss.ssec.wisc.edu/iwgg/iww10/talks/genkova2.pdf> (last access: 13 April 2023).
- Hoffman, R. N., Lukens, K. E., Ide, K., and Garrett, K., 2021: A collocation study of atmospheric motion vectors (AMVs) compared to Aeolus wind profiles with a feature track correction (FTC) observation operator. Q. J. R. Meteorol. Soc., 148(742):321–337, Jan. (Part A), <https://doi.org/10.1002/qj.4207>.
- Lukens, K. E., Ide, K., Garrett, K., Liu, H., Santek, D., Hoover, B., and Hoffman, R. N., 2022: Exploiting Aeolus level-2b winds to better characterize atmospheric motion vector bias and uncertainty, Atmos. Meas. Tech., 15, 2719–2743, <https://doi.org/10.5194/amt-15-2719-2022>.
- Martin, A., Weissmann, M., Reitebuch, O., Rennie, M., Geiß, A., and Cress, A., 2021: Validation of Aeolus winds using radiosonde observations and numerical weather prediction model equivalents, Atmos. Meas. Tech., 14, 2167–2183, <https://doi.org/10.5194/amt-14-2167-2021>.
- Moninger, W. R., Mamrosh, R. D., and Pauley, P. M., 2003: Automated Meteorological Reports from Commercial Aircraft. BAMS, 84(2), 203–216, <https://doi.org/10.1175/BAMS-84-2-203>.
- National Academies of Sciences, Engineering, and Medicine: Thriving on Our Changing Planet, 2018: A Decadal Strategy for Earth Observation from Space. National Academies Press, Washington, DC, USA, <https://doi.org/10.17226/24938>.
- Rani, S.I., Jangid, B.P., Kumar, S., Bushair, M.T., Sharma, P., George, J.P., et al., 2022: Assessing the quality of novel Aeolus winds for NWP applications at NCMRWF. Quarterly Journal of the Royal Meteorological Society, 148(744), 1344–1367. Available from: <https://doi.org/10.1002/qj.4264>.
- Rennie, M. and Isaksen, L., 2020: The NWP impact of Aeolus Level-2B winds at ECMWF. Tech. rep., Aeolus Data Innovation Science Cluster DISC, Ref: AED-TN-ECMWF-NWP-025, <https://doi.org/10.21957/alift7mhr>.

- Rhodes, B., and Candido, S., 2021: Loon Stratospheric Sensor Data [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.3755988>.
- Santek, D., García-Pereda, J., Velden, C., Genkova, I., Wanzong, S., Stettner, D., and Mindock, M., 2014: A new atmospheric motion vector intercomparison study, in: 12th International Winds Workshop, 16–20 June 2014, Copenhagen, Denmark, CIMSS, Technical Report, <http://www.nwcsaf.org/aemetRest/downloadAttachment/225> (last access: 18 December 2020).
- Santek, D., Dworak, R., Nebuda, S., Wanzong, S., Borde, R., Genkova, I., García-Pereda, J., Negri, R. G., Carranza, M., Nonaka, K., Shimoji, K., Oh, S. M., Lee, B.-I., Chung, S.-R., Daniels, J., and Bresky, W., 2019: 2018 Atmospheric Motion Vector (AMV) Intercomparison Study, *Remote Sens.*, 11, 2240, <https://doi.org/10.3390/rs11192240>.
- Santek, D., Dworak, R., Wanzong, S., Rink, T., Lukens, K., Reiner, S., and García-Pereda, J., 2022: NWC SAF Winds Intercomparison Study Report: 2021. International Winds Working Group, available at [http://cimss.ssec.wisc.edu/iwwg/Docs/CIMSS\\_AMV\\_Comparison\\_2021\\_Report\\_02Nov2022.pdf](http://cimss.ssec.wisc.edu/iwwg/Docs/CIMSS_AMV_Comparison_2021_Report_02Nov2022.pdf).
- Santek, D., Hoover, B., Zhang, H., and Moeller, C., 2021: Evaluation of Aeolus Winds by Comparing to AIRS 3D Winds, Rawinsondes, and Reanalysis Grids, in: 15th International Winds Workshop, 12–16 April 2021, Virtual, CIMSS, <https://www.ssec.wisc.edu/meetings/iwwg/2021-meeting/presentations/oral-santek/>.
- Šavli, M., Pourret, V., Payan, C. and Mahfouf, J.-F., 2021: Sensitivity of Aeolus HLOS winds to temperature and pressure specification in the L2B processor. *Atmospheric Measurement Techniques*, 14, 4721–4736. <https://doi.org/10.5194/amt-14-4721-2021>.
- Velden, C. S. and Bedka, K. M., 2009: Identifying the Uncertainty in Determining Satellite-Derived Atmospheric Motion Vector Height Attribution, *J. Meteorol. Clim.*, 48, 450–463, <https://doi.org/10.1175/2008JAMC1957.1>.
- Velden, C. S. and Holmlund, K., 1998: Report from the working group on verification and quality indices (WG II), in: 4th International Winds Workshop, 20–23 October 1998, Saanenmöser, Switzerland, EUMETSAT, [http://cimss.ssec.wisc.edu/iwwg/iww4/p19-20\\_WGReport3.pdf](http://cimss.ssec.wisc.edu/iwwg/iww4/p19-20_WGReport3.pdf).
- Velden, C., Daniels, J., Stettner, D., Santek, D., Key, J., Dunion, J., Holmlund, K., Dengel, G., Bresky, W., and Menzel, P., 2005: Recent innovations in deriving tropospheric winds from meteorological satellites, *B. Am. Meteorol. Soc.*, 86, 205–223, <https://doi.org/10.1175/BAMS-86-2-205>.
- Velden, C. S., Hayden, C. M., Nieman, S. J., Menzel, W. P., Wanzong, S., and Goerss, J. S., 1997: Upper-tropospheric winds derived from geostationary satellite water vapor observations, *B. Am. Meteorol. Soc.*, 78, 173–195, [https://doi.org/10.1175/1520-0477\(1997\)078<0173:UTWDFG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<0173:UTWDFG>2.0.CO;2).
- Weiler, F., Kanitz, T., Wernham, D., Rennie, M., Huber, D., Schillinger, M., Saint-Pe, O., Bell, R., Parrinello, T., and Reitebuch, O., 2021a: Characterization of dark current signal

measurements of the ACCDs used on-board the Aeolus satellite. *Atmos. Meas. Tech.*, 14, 5153–5177, <https://doi.org/10.5194/amt-14-5153-2021>.

Weiler, F., Rennie, M., Kanitz, T., Isaken, L., Checa, E., de Kloe, J., Okunde, N., and Reitebuch, O., 2021b: Correction of wind bias for the lidar on board Aeolus using telescope temperatures. *Atmos. Meas. Tech.*, 14, 7167–7185, <https://doi.org/10.5194/amt-14-7167-2021>.