

## Data Set Documentation Guidelines

The documentation (i.e., the "Readme" file) that accompanies each project data set is as important as the data itself. This information permits collaborators and other analysts to understand any limitations or special characteristics of the data that may impact its use. Data set documentation should accompany all data set submissions, including both preliminary and final. The following outline and content is recommended and should be adhered to as closely as possible to make the documentation consistent across all data sets.

### Data Set Documentation/Readme Outline:

**Title:** This should match the data set name

**Author(s):**

- Name(s) of the Lead author and any co-authors
- All authors are requested to provide an [ORCID](#), email address, and institution/organization name
- Complete mailing address, telephone/facsimile numbers, title or position, and website address (if applicable)
- Similar contact information for data questions (if different than above)

Minghui Diao, Assistant Professor

[Minghui.diao@sjsu.edu](mailto:Minghui.diao@sjsu.edu)

ORCID: 0000-0003-0324-0897

Affiliation: San Jose State University

Address: One Washington Square, San Jose State University, San Jose, CA, 95192-0104

### 1.0 Data Set Overview:

- Introduction or abstract
- Time period covered by the data
- Physical location (including lat/lon/elev) of the measurement or platform
- Data source if applicable (e.g., for operational data include agency)
- Any web address references (i.e., additional documentation such as Project web site)

The VCSEL hygrometer provided measurements of water vapor volume mixing ratio (unit: ppmv) during the NSF SOCRATES campaign.

The VCSEL water vapor 1-Hz and 25-Hz data from the NSF SOCRATES campaign have been processed based on laboratory calibrations conducted by Minghui Diao in the summer of 2018. After the calibration, statistical distributions of in-cloud relative humidity have improved 3% - 5% compared with the original data hosted at EOL directory.

The main change in the version 2 water vapor data compared with version 1 is that a stricter QAQC procedure has been applied to filter out more problematic samples. For example, data with very low

laser intensity (< 100) are deleted. +/- 2 seconds surrounding the error codes or mode changes are deleted. The calibration parameters remain the same as the version 1 data processed by PI Diao.

Because of the time and effort spent in QA/QC this water vapor dataset, please consider offering co-authorship, if you decide to use this PI-based calibrated data. If you have questions regarding the data, please contact the PI (Minghui Diao) at [minghui.diao@sjsu.edu](mailto:minghui.diao@sjsu.edu).

More information on the laboratory calibration procedure, parameters used for the calibration, and improvements of the relative humidity distributions in liquid, mixed and ice phase clouds can be found on: <https://www.cloud-research.org/water-vapor>

## **2.0 Instrument Description:**

- Brief text (i.e., 1-2 paragraphs) describing the instrument with references
- Figures (or links), if applicable
- Table of specifications (i.e., accuracy, precision, frequency, resolution, etc.)

## **3.0 Data Collection and Processing:**

- Description of data collection
- Description of derived parameters and processing techniques used
- Description of quality assurance and control procedures
- Data intercomparisons, if applicable

For calibration parameters and experimental setup, please refer to the PowerPoint slides uploaded in this webpage.

## **4.0 Data Format:**

- Data file structure and file naming conventions (e.g., column delimited ASCII, NetCDF, GIF, JPEG, etc.)
- Data format and layout (i.e., description of header/data records, sample records)
- List of parameters with units, sampling intervals, frequency, range
- Data version number and date
- Description of flags, codes used in the data, and definitions (i.e., good, questionable, missing, estimated, etc.)

VCSEL version 2 data for SOCRATES

Updated date: September 3, 2021

Data version for this version 2 dataset is named as `socrfxx_H2O_XXHz_v2_Diao.txt`.

Text files (.txt) are included for the QA/QC water vapor data, including five columns: timeGV (UTC time in seconds), H2O\_conc18qc (water vapor molecule number concentrations with a unit of # of molecule / cm<sup>3</sup>), H2O\_VXL18qc (water vapor mixing ratio by VCSEL hygrometer, with a unit of ppmv, calculated as water vapor partial pressure divided by total air pressure then multiplying 1e6), RHliq18qc (relative humidity with respect to liquid in %), RHice18qc (relative humidity with respect to ice in %).

timeGV	H2O_conc18qc	H2O_VXL18qc	RHliq18qc	RHice18qc
82200	1.92e+17	7.47e+03	40.7	34.7
82201	1.91e+17	7.45e+03	40.4	34.4
82202	1.91e+17	7.46e+03	40.1	34.1
82203	1.92e+17	7.49e+03	40.0	34.0

Error codes for the reported variables are listed below:

H2O_conc18qc	H2O_VXL18qc	RHliq18qc	RHice18qc
9.99e+19	9.99e+04	999.0	999.0

The calculations of RHliq and RHice use water vapor mixing ratio measured by the VCSEL hygrometer, temperature measured by the Rosemount temperature probe, and total pressure. Equations of saturation vapor pressure with respect to ice and liquid are based on Murphy and Koop (2005).

### 5.0 Data Remarks:

- PI's assessment of the data (i.e., disclaimers, instrument problems, quality issues, etc.)
- Missing data periods
- Software compatibility (i.e., list of existing software to view/manipulate the data)

In the version 2 data, the reported RHliq has 2019 seconds above 105%, and there is no RHliq value above 110%. Note that the uncertainties of RHliq and RHice reflect the uncertainties from both water vapor and temperature measurements.

### 6.0 References:

- List of documents cited in this data set description. Please provide links for any publications, if available.