Atmospheric Composition Variable Standard Name Convention

Status of this Memo

This RFC document describes the standard for atmospheric composition standard names, including guidelines on how to generate a standard name and lists of controlled variable names for use within the atmospheric composition standard name.

Distribution of this memo is unlimited.

Change Explanation

This RFC does not update or change a previous RFC.

Copyright Notice

Abstract

Development of the atmospheric composition standard names was initially motivated by the need to use the ICARTT V2.0 file format standard for data reporting from suborbital field studies. This version requires a standard name as part of the variable definition. Through test-implementation in several suborbital field studies, the standard name has demonstrated its role in supporting findable, accessible, interoperable, and reusable (FAIR) principles for data management and stewardship.

Over the last forty years, the number of measurements taken during suborbital field campaigns has increased more than tenfold. Typically, the only requirement for variable names has been that they be unique within a data provider's data file; with the rise in the number of variables and lack of constraints in how they are named, the complexity for data users to search and DAACs to archive and distribute the data has increased. Examination of the current naming systems has highlighted several limitations. First, only a quarter to a third of atmospheric composition measurements have standard names, the largest missing category being hydrocarbon measurements. Second, some of the current vocabulary and definitions are not suitable for atmospheric composition measurements. Third, optional qualifiers can lead to very long standard names with no consistent structure. Due to these limitations and the additional concerns, that variable names sometimes only provide a partial perspective of their intended physical phenomenon and that variable names do not always provide a clear message about the measurement, a new standard naming system was developed.

The atmospheric composition variable standard names are constructed using controlled vocabulary terms with four parts: measurement category (MeasurementCategory), core name (CoreName), acquisition method (AcquisitionMethod), and descriptive attributes

(DescriptiveAttributes), which are separated by an underscore. This structure is similar to that of the Climate and Forecast Metadata Conventions (CF) standard and the Geoscience Standard Names (GSN) ontology. The MeasurementCategory and CoreName can uniquely define the physical quantity and can be used to conduct a broad search to identify all measurements of the same physical quantity from different instruments and/or field studies. The AcquisitionMethod identifies the sampling technique that was used for the measurement, while the DescriptiveAttributes are intended to provide necessary description, which are needed to support research use or narrow down the search for data of interest.

This document provides a set of guidelines and controlled vocabulary for creating the standard name for different types of measurements conducted during atmospheric composition suborbital field studies. It is the intent that these lists will be used to support and improve the data reporting from NASA suborbital field studies. One practice is to directly support use of ICARTT 2.0 file format standards by fulfilling the variable standard name requirement, a tag for each data product variable. Not only can these standard names be used in the ICARTT 2.0 format, but they can be used as variable attributes in NetCDF/HDF formats. It is expected that the lists of controlled vocabulary will need to be updated in response to the advances in atmospheric composition measurements. This is evident in the over one hundred new core names that have been added since the 2019 FIREX-AQ field campaign.

Table of Contents

1 Introduction

A variable name readily identifies the physical phenomena that is the subject of a set of measurements, or some other data set, and serves as a link to its mathematical description. It is common for instrument scientists to use their intended measurable quantity as the data variable name. Datasets (i.e., collections of variables) gathered within the field of atmospheric science are used in a variety of ways, including computational modeling, interpretive analysis, measurement intercomparisons, and validation studies. To enable the use of datasets from different sources (i.e., data products from different measurement assets and/or models) in these various research activities, the data needs to be findable, accessible, interoperable, and reusable (FAIR) [1] across multiple computational resources.

Since the 1980's, over 30 major tropospheric airborne field campaigns have been conducted by NASA and partner agencies to investigate atmospheric composition. Since the beginning of these field campaigns, the number of variables measured during them has increased more than tenfold. In 1992, the Transport and Atmospheric Chemistry near the Equator-Atlantic (TRACE-A) campaign measured approximately 50 variables, in comparison to 2019 where the Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) campaign had over 600 variables reported. Hydrocarbon compounds and aerosol properties make up the largest component of these new variables. The large number of measured variables (especially when considering multiple field campaigns) increases the complexity for distributed active archive centers (DAACs) to distribute the data and for data users to search for and find their measurements of interest.

During NASA suborbital atmospheric composition field studies, variable names are given by the instrument scientists. Since there have not been many guidelines related to variable names, and the only requirement has been that they are unique within a data provider's data file, variable names and structures can vary significantly, even for the same type of variable. For example, some instrument scientists add additional information to their variable names to indicate units and/or instruments (e.g., O3_ppbv for ozone mixing ratio reported in parts per billion by volume and DLH_H2O_ppmv for diode laser hygrometer (DLH) measurement of water vapor mixing ratio reported in parts per million by volume). This practice can satisfy variable name uniqueness, at least within one data file, which is necessary especially when there is more than one instrument to measure the same physical quantity. It is not uncommon to have multiple instruments measure mission critical variables on a single measurement platform and in a single field study. Even the same instrument scientist may choose a different variable name for the same measurement during a different field campaign (Figure 1).

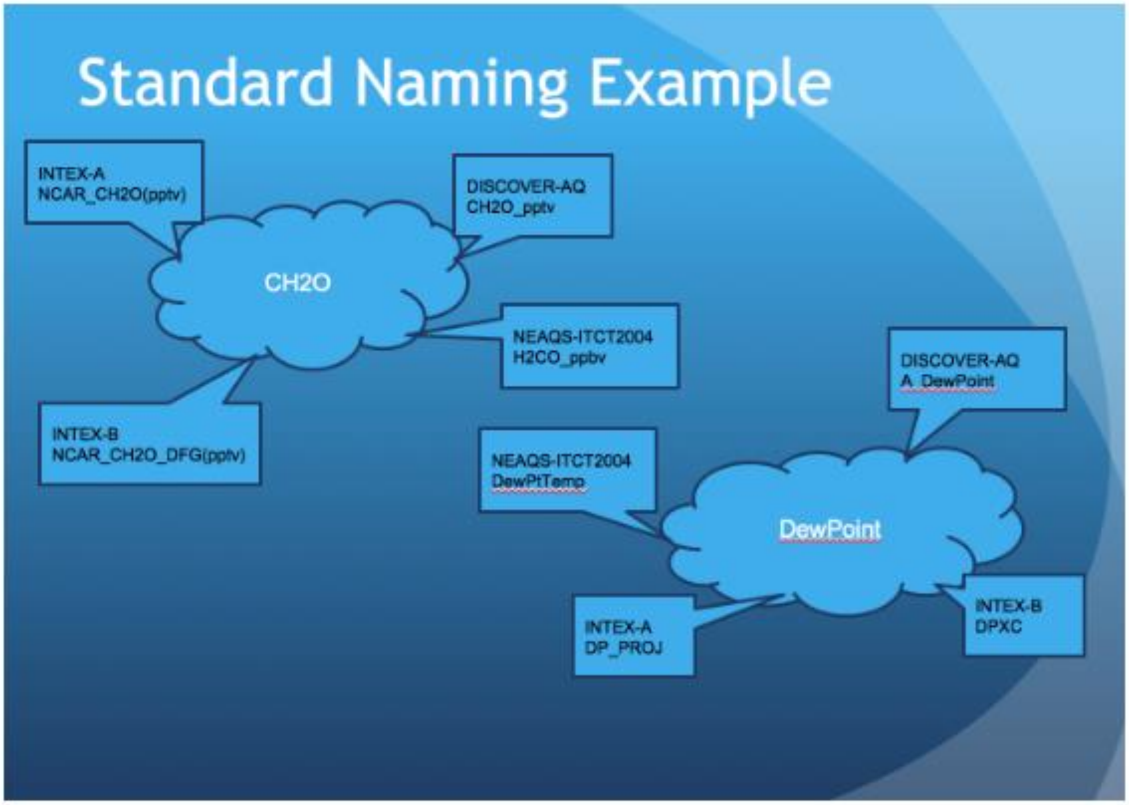

Figure 1: Example of formaldehyde (CH_2O) and dew point variable names across multiple campaigns.

While it is important to have uniqueness within a dataset from one study, particularly when data are reported in ASCII format (e.g., International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) format), the variable name differences make it difficult for users to locate and interact with a particular variable across multiple data sets. This is especially true for data users who are not familiar with the field missions and who would therefore not have the knowledge they need to search for A_DewPoint in one field campaign and DPXC in another to find data on dew point, for example. One effective solution to this problem, identified by the Earth Science Data System (ESDS) ICARTT Refresh Working Group, is to introduce variable standard names that can be used as tags for each data variable [2]. This will allow similar measurements (e.g., dew point) to be categorized and located across field campaigns, regardless of what the instrument scientists gave as the variable name. Based on the successful implementation of using standard names in the NASA Toolsets for Airborne Data (TAD) web application, later Subsetting Tools for Advanced Analysis of Airborne Chemistry Data (STA3CD), the use of standard variable names was adopted in the NASA suborbital data format standard ICARTT V2.0 (https://cdn.earthdata.nasa.gov/conduit/upload/6158/ESDS-RFC-029v2.pdf) [2]. These standard names are intended to enable data discovery (i.e., findable and accessible) and support data ingest across different studies (i.e., interoperable and reusable). By providing a common terminology, all users (instrument scientists, science team members, students, researchers, and DAACs) will be able to find, distribute and use atmospheric composition data in an efficient manner.

To date, the current list of atmospheric composition standard names has been successfully implemented in and accepted by science team members (from instrument scientists to modelers) in FIREX-AQ, the Cloud, Aerosol and Monsoon Processes – Philippines Experiment (CAMP^{2}EX), the Aerosol Cloud Meteorology Interactions over the Western Atlantic Experiment (ACTIVATE), and the Dynamics and Chemistry of the Summer Stratosphere (DCOTSS) field campaigns. With the help of field study science team members, over one hundred new measurement core names have been added to the list of controlled vocabulary, since prior to the aforementioned field campaign data submissions.

The use of standard names has also helped the Atmospheric Science Data Center (ASDC) streamline their data ingest and archival process by saving time assigning data to its respective data product grouping, or collection. The standard names allow for data product groupings, based on its physical and/or chemical properties, to be readily defined and for the data files to accurately be assigned to their respective grouping. This allows a DAAC to ingest and archive the data collected more efficiently, along with distributing it to the public.

A few different variable standard naming systems exist, but they do not adequately cover the needs of atmospheric composition measurements, and specifically the needs of the NASA suborbital atmospheric science field studies (i.e., field campaigns involving aircraft, ground sites, ships, and mobile labs). One of the approved standards recommended for use in NASA Earth Science Data Systems is the Network Common Data Form (NetCDF) Climate and Forecast (CF) Metadata Conventions. The CF "standard is intended for use with climate and forecast data, for atmosphere, surface and ocean, and designed with model-generated data particularly in mind" [3]. While many principles of the CF convention can be extended beyond its originally intended uses, it has been increasingly recognized by various communities that the CF standard name and its structure have shortcomings when handling in-situ atmospheric composition measurements involved in suborbital field studies.

Currently, CF standard names only cover a quarter to a third of the atmospheric composition relevant measurements taken in suborbital field campaigns. The largest group of missing measurements are hydrocarbon measurements. While variables for aerosol and cloud measurements are generally well represented, many attributes (qualifiers) are missing or do not adequately or accurately describe measurements. These attributes are needed to properly define the measurement for research use. For example, the CF standard name "atmosphere_optical_thickness_due_to_particulate_organic_matter_ambient_aerosol" does not clearly define the size range of the aerosol particles, which is critical for comparing different in-situ measurements of aerosol particles. There are also inconsistencies within the CF standard names. For example, some aerosol properties have both dry and ambient relative humidity versions of their standard name, but others do not. With CF explicitly defining standard names, the addition of tens of thousands of extra CF names would be required to include all the additional species and attributes needed to account for various combinations.

Much of the specific vocabulary and terms currently used in CF are not suitable for atmospheric composition measurements since they are designed for model and forecast data. For example, “mass_concentration_of_pm1_ambient_aerosol_particles_in_air” specifically defines the measurement of PM1 aerosol as "…particulate compounds with an aerodynamic diameter of less than or equal to 1 micrometer”. This definition is misleading and incorrect if used for a mass measurement of ambient aerosol particles that is based on optical sizing rather than aerodynamic sizing technique.

Lastly, the construction of CF names is governed by a set of guidelines that allows for the addition of qualifiers, such as surface, component, medium, process, and condition, to a base standard name using underscores. The difficulty in this, particularly from an interoperability standpoint, is that many of these qualifiers are optional, making it hard for a system or user to know the ‘base standard name’, since there is no set number of underscores or qualifiers. The standard names may also be derived from other standard names following a set of specified rules. In these cases, standard names can begin with words such as “tendency_of_X” or “product_of_X_and Y” where X and Y are the other standard names. The range of possibilities for CF standard names poses a problem for discoverability and interoperability.

Another current list of standard names is the Geoscience Standard Names Ontology (GSN), which is based on and extended from the Community Surface Dynamics Modeling System (CSDMS) standard names. The GSN standard names are a set of variable names using a series of rules and controlled vocabulary designed to avoid ambiguous variable names and domain-specific terminology, use generic or already-standardized object names, are human and machine readable, and standardized. The standard names are constructed through pairings of object names and quantity names (identification of measurement concept used to quantify the object in some way) and uses CF names for atmospheric chemistry (a key part of atmospheric composition studies). The Scientific Variables Ontology (SVO) is a successor to GSN [4]. However, due to the reliance on CF names, which as mentioned previously does not fully capture the range of measurements currently taken in atmospheric field campaigns nor the attributes necessary to describe measurement sampling and data reporting, these standard naming schemes are not suitable for suborbital atmospheric composition field studies.

Due to the limitations in the current set of standards available for use as standard names, and to address other goals or concerns (described below), a new system of standard names has been formed.

First, standard names generated by this system attempt to strike a good working compromise between completeness (or explicitness) and generality. For example, neither particle_concentration, nor concentration_of_particles_in_size_ranges_10_to_55_nm_measured_by_Joe_Doe_particle_counter_on_the_P3B_while_flying_over_the_South_Pacific serve the purpose of machine interoperability, while preserving readability for human users.

Additionally, the standard names are intended to address two important aspects of atmospheric composition measurements. First, atmospheric composition measurements sometimes can only provide a partial perspective to the intended physical phenomena or quantity. For example, the number size distribution of aerosol particles is always used in the assessment of aerosol particle microphysical and optical properties. In theoretical description, aerosol particle size is defined as the geometric size, but this cannot be readily measured by instruments commonly used in the field studies. What is typically reported in data sets are optical size (based on particle scattering properties), mobility size, or aerodynamic size. These measurements are related to, but different from the geometric size. Thus, standard names containing attributes provide information relative to the measurements and enhance data usability.

Second, it is also the case that some variable names, commonly used in literature to describe rather complicated physical quantities or processes, are unlikely to be understood by users outside of the subject area by simply reading the variable name. For example, SSA, CCN, and NOy:

- SSA: single scattering albedo of aerosol particles within an observed size range and at a particular wavelength
- CCN: concentration of cloud condensation nuclei measured at set instrument supersaturation levels
- NOy: total mixing ratio, which is a lumped measurement of reactive nitrogen species comprising NO, NO2, HNO3, PAN, HONO, NO3, N2O5, organic nitrates, and possible particulate nitrates

These variable names do not convey a clear message to those who do not have background knowledge in the subject area. However, these terms can be used by users to get a precise definition and to learn about the subject area, since they are widely used by the research community in journal publications. A goal of the atmospheric composition variable standard names, then, is to be accessible to and provide context for all users, while still remaining connected to those in the subject area.

This document provides a set of guidelines for creating standard names for different types of measurements conducted during suborbital field studies on atmospheric composition. It is the intent that these lists will be used to support and improve the data reporting from NASA suborbital field studies by fulfilling the variable standard name requirement in the ICARTT V2.0 file format standards, in which each data product variable is tagged with a standard name. It is noted that the application of the atmospheric composition variable standard names is not limited to the ICARTT V2.0 data format. For example, they can also be used as variable attributes in NetCDF/HDF files.

2 Standard Name Recommendations

The proposed standard names are constructed using controlled vocabulary terms with four parts: measurement category (MeasurementCategory), core name (CoreName), acquisition method

(AcquisitionMethod), and descriptive attributes (DescriptiveAttributes), which are separated by an underscore:

Standard Name = MeasurementCategory_CoreName_AcquisitionMethod_DescriptiveAttributes

It is designed to support data discovery, distribution, interoperability, and use, by accurately describing all variables from different measurements/instruments while using a consistent format for interoperability. For data discovery, the MeasurementCategory and CoreName (i.e., corename themselves and respective descriptions) can be used to conduct a broad search to identify all measurements of the same physical quantity from different instruments and/or field studies. The DescriptiveAttributes can then be used to narrow down the search for data of interest.

Measurement category broadly groups all measurement standard names into one of thirteen categories (Gas, AerComp, AerMP, AerOpt, CldComp, CldMicro, CldMacro, CldOpt, Met, GasJValue, AquJValue, Platform, and Rad). This provides uniqueness as using only CoreNames could be ambiguous (e.g., a particle number concentration could be describing cloud or aerosol particles). Within each MeasurementCategory, the format of each standard name is consistent (i.e., they have the same number and type of descriptive attributes).

The core name is the basic identification of the physical quantity being reported. The CoreNames chosen are those that have been commonly used in literature. As described previously, the terms commonly used in literature are sometimes unlikely to be understood by users outside of the subject area. To remedy this, a brief description is provided along with each core name. This description provides additional information for when the core name chosen isn’t explicit in what it means.

The acquisition method refers to the sampling technique of the measurement. The modes chosen are similar to the European Space Agency (ESA) Atmospheric Validation Data Centre (EVDC) acquisition method metadata attributes, which are InSitu, Numerical Simulation, Remote Sensing, and Sample [5].

The descriptive attributes provide measurement and/or data reporting information relevant for data use and faceted data search. The number and types of descriptive attributes are dependent on the MeasurementCategory, but consistent within each MeasurementCategory. For example, all trace gas standard names have two descriptive attributes, MeasurementSpecificity and Reporting, whereas all aerosol optical property standard names have four descriptive attributes, wavelength (WL), MeasurementRH, SizeRange, and Reporting.

The atmospheric composition standard names document containing the controlled vocabulary lists for each term is intended to be a living document. To stay relevant to the measurements and user community, CoreNames will be updated and/or modified as part of each major field campaign. Descriptive attributes may also be updated, but far less often as the current lists capture all possibilities that are known to exist in the atmospheric composition field study data holdings. Due to the ever-growing nature of the controlled lists, it is located at the Suborbital Science Data for Atmospheric Composition data repository, www-air.larc.nasa.gov [6]. This document will be maintained by the ASDC DAAC in conjunction with field campaign data management teams.

3 References

Normative References

[6] Atmospheric Composition Variable Standard Name Recommendations, October 2021, https://www-air.larc.nasa.gov/missions/etc/AtmosphericCompositionVariableStandardNames.pdf

Informative References

[1] M. D. Wilkinson, et al, "The FAIR Guiding Principles for scientific data management and stewardship", Scientific Data, 3, (2016), https://www.go-fair.org/fair-principles/

[2] B. Eaton et al, NetCDF Climate and Forecast (CF) Metadata Conventions, September 2021, https://cfconventions.org/Data/cf-conventions/cf-conventions-1.9/cf-conventions.html#_introduction

[3] Guidelines for Construction of CF Standard Names, December 2008, http://cfconventions.org/Data/cf-standard-names/docs/guidelines.html

[4] Scientific Variables Ontology, 2019, scientificvariablesontology.org/documentation/

[5] EVDC Metadata Guidelines, December 2019, https://evdc.esa.int/documentation/browse-meta-data/data_discipline/

4 Authors' Addresses

Primary Authors

Morgan Silverman, SSAI/NASA LARC, MS 401B, Hampton, VA, USA Tel: 757-864-3219
email: morgan.l.silverman@nasa.gov
Michael Shook, NASA LARC, michael.a.shook@nasa.gov
Rebecca Hornbrook, NCAR/UCAR, rsh@ucar.edu
Luke Ziemba, NASA LARC, luke.ziemba@nasa.gov
Sam Hall, NCAR/UCAR, halls@ucar.edu
Kirk Ullmann, NCAR/UCAR, ullmannk@ucar.edu
John Crounse, CalTech, crounjd@caltech.edu
Ryan Bennett, NSRC, r.bennett@baeri.org
Megan Buzanowicz, SSAI/NASA LARC, megan.e.buzanowicz@nasa.gov
Gao Chen, NASA LARC, gao.chen@nasa.gov

Contributing Authors

Barbara Barletta, University of California – Irvine, bbarlett@uci.edu
Helen Conover, UAH ITSC/NASA MSFC/GHRC DAAC, helen.conover@uah.edu
Josh DiGangi, NASA LARC, joshua.p.digangi@nasa.gov
Keith Evans, JCET/UMBC, evans@umbc.edu
Arlene Fiore, Columbia University, af2544@columbia.edu
Melanie Follette-Cook, Morgan State University/NASA GSFC, melanie.cook@nasa.gov
Peter Hall, NASA GSFC, peter.f.hall@nasa.gov
Peter Leonard, ADNET Systems/NASA GSFC, peter.j.leonard@nasa.gov
Qing Liang, NASA GSFC, qing.liang@nasa.gov
John Nowak, NASA LARC, john.b.nowak@nasa.gov
Scott Peckham, University of Colorado – Boulder, scott.peckham@colorado.edu
Hampapuram Ramapriyan, SSAI/NASA GSFC, hampapuram.ramapriya@ssaihq.com
Scott Ritz, NASA GSFC, scott.a.ritz@nasa.gov
Martin Schultz, Forschungszentrum Jülich, m.schultz@fz-juelich.de
Maria Stoica, University of Colorado – Boulder, maria.stoica@colorado.edu
Qian Tan, NASA ARC, qian.tan@nasa.gov
Andrew Thorpe, JPL, Andrew.K.Thorpe@jpl.nasa.gov
Jian Zeng, ADNET Systems/NASA GSFC, jian.zeng@nasa.gov

Working Group Members

Aubrey Beach, NASA LaRC/ASDC currently at Booz Allen Hamilton
Stephane Beland, University of Colorado – Boulder, sbeland@colorado.edu
Stephen Berrick, NASA ESDIS, Stephen.W.Berrick@nasa.gov
Allan Doyle, International Interfaces, Inc., adoyle@intl-interfaces.com
Amanda Benson Early, NASA LaRC/ASDC currently at ActiveCampaign
Scott Gluck, JPL, Scott.Gluck@jpl.nasa.gov
Nathan James, NASA ESDIS, nathan.l.james@nasa.gov
James Johnson, NASA GSFC, james.johnson@nasa.gov
Jeanne Laurencelle, University of Alaska – Fairbanks, jclaurencelle@alaska.edu
Dave Meyer, NASA GSFC, david.j.meyer@nasa.gov
George Milly, Columbia University, gpm2109@columbia.edu
David Moroni, JPL PO.DAAC, david.f.moroni@jpl.nasa.gov
Emily Northup, NASA LaRC/ASDC currently at Trader Interactive
Deborah Smith, UAH – ITSC/NASA MSFC/GHRC DAAC, deborah.smith@uah.edu
Chris Stoner, University of Alaska Satellite Facility, cstoner5@alaska.edu
Richard Strub, SSAI/NASA GSFC, richard.f.strub@nasa.gov
Jeff Walter, NASA LaRC, jeff.walter@nasa.gov

Appendix A - Glossary of Acronyms

Acronym	Description
ACTIVATE	Aerosol Cloud Meteorology Interactions over the Western Atlantic Experiment
ARC	Ames Research Center
ASCII	American Standard Code for Information Interchange
ASDC	Atmospheric Science Data Center
CAMP2EX	Cloud, Aerosol and Monsoon Processes – Philippines Experiment
CCN	Cloud Condensation Nuclei
CF	Climate and Forecast (CF) Metadata Conventions
CSDMS	Community Surface Dynamics Modeling System
DAAC	Distributed Active Archive Center
DCOTSS	Dynamics and Chemistry of the Summer Stratosphere
DLH	Diode Laser Hygrometer
ESA	European Space Agency
ESDIS	Earth Science Data Information Systems
ESDS	Earth Science Data Systems
EVDC	Atmospheric Validation Data Centre
FAIR	Findable, Accessible, Interoperable, and Reusable
FIREX-AQ	Fire Influence on Regional to Global Environments and Air Quality
GHRC	Global Hydrometeorology Resource Center
GSFC	Goddard Space Flight Center
GSN	Geoscience Standard Names Ontology
HDF	Hierarchical Data Format
ICARTT	International Consortium for Atmospheric Research on Transport and Transformation
ITSC	Information Technology and Systems Center
JCET	Joint Center for Earth Systems Technology
JPL	Jet Propulsion Laboratory
LARC	Langley Research Center
MSFC	Marshall Space Flight Center
NASA	National Aeronatuics and Space Administration
NetCDF	Network Common Data Form
NSRC	National Suborbital Research Center
PO	Physical Oceanography
SSA	Single Scattering Albedo
SSAI	Science Systems & Applications, Inc
STA3CD	Subsetting Tools for Advanced Analysis of Airborne Chemistry Data

SVO	Scientific Variables Ontology
TAD	Toolsets for Airborne Data
TRACE-A	Transport and Atmospheric Chemistry near the Equator-Atlantic
UAH	University of Alabama - Huntsville
UCAR	University Corporation for Atmospheric Research
UMBC	University of Maryland, Baltimore County
WL	Wavelength