GHOST QA Flag Definitions

Table of Contents

[Preface 1](#_Toc535249028)

[Definition Syntax 1](#_Toc535249029)

[QA FLAG DEFINITIONS 1](#_Toc535249030)

[Basic QA Flags 1](#_Toc535249031)

[Duplicate / Overlapping Time Flags 2](#_Toc535249032)

[Station Classifications 3](#_Toc535249033)

[Duplicate station 6](#_Toc535249034)

[Extreme Data Flags 6](#_Toc535249035)

[Invalid Data Provider Flags 7](#_Toc535249036)

[Recurring Value 7](#_Toc535249037)

[Hitting LOD 7](#_Toc535249038)

[Insufficient Measurement Resolution Flags 7](#_Toc535249039)

[Limit of Detection Flags 8](#_Toc535249040)

[Measurement Methodology Flags 8](#_Toc535249041)

[Temporal Period Flags 10](#_Toc535249042)

[Temporal Representation Flags 11](#_Toc535249043)

# Preface

# Definition Syntax

# QA FLAG DEFINITIONS

**QA\_flag\_name:**

standard\_name:

description:

## Basic QA Flags

**A1**

standard\_name: Missing Measurement

description: Measurement is missing (i.e. NaN)

**A2**

standard\_name: Infinite Value

description: Value is infinite -- happens when values are outside of the range that the float32 data type can handle (-3.4E+38 to +3.4E+38)

**A3**

standard\_name: Negative Measurement

description: Measurement is negative in absolute terms.

**A4**

standard\_name: Zero Measurement

description: Have measurement equal to zero

**A5**

standard\_name: No Key Metadata

description: One of the key metadata fields (latitude, longitude) is absent

**A6**

standard\_name: No Altitude Metadata

description: Altitude metadata field is absent, altitude is estimated using the ETOPO1 1 arc-minute global relief model

**A7**

standard\_name: Assumed Gas Volume

description: Have assumed gas volume when converting between mass density and volume mixing ratio. (Do not have either temperature or pressure, or both).

**A8**

standard\_name: Non-Integer Local Timezone (relative to UTC)

description: Local timezone has been determined to be non-integer for time of measurement, relative to UTC.

## Duplicate / Overlapping Time Flags

#Duplicate Time - First Value Kept

#Multiple measurements reported for the same temporal window - the first time is kept preferentially.

'Duplicate Time - First Value Kept':'B1',

#Overlapping Time - First Value Kept

#Measurements reported with overlapping temporal windows - these windows have the same temporal resolutions (e.g. all hourly), therefore the first time is preferentially kept.

'Overlapping Time - First Value Kept':'B2',

#Overlapping Time - Finest Temporal Resolution Kept

#Measurements reported with overlapping temporal windows - at least 2 of the overlapping windows have different temporal resolutions. Preferentially the finest temporal resolutions are kept. If have more than 1 windows of the same finest resolution, the first time is preferentially kept.

'Overlapping Time - Finest Temporal Resolution Kept':'B3',

## Station Classifications

#High Altitude - Metadata Altitude

#Station has an altitude >= 1500 metres relative to mean sea level, altitude taken from network provided metadata

'High Altitude - Metadata Altitude':'C1',

#High Altitude - Metadata Derived

#Station determined to be a 'mountain' station, derived from standardised 'terrain' network provided metadata

'High Altitude - Metadata Derived':'C2',

#High Altitude - ETOPO1

#Station determined to have an altitude >= 1500 metres relative to sea level datum, altitude taken from ETOPO1 digital elevation model

'High Altitude - ETOPO1':'C3',

#High Altitude - Iwahashi Global Landform Classification

#Stations determined to be high altitudes, derived from the European Soil Data Centre Iwahashi Global Landform Classification.

'High Altitude - Iwahashi Global Landform Classification':'C4',

#High Altitude - Meybeck Global Landform Classification

#Stations determined to be high altitudes, derived from the European Soil Data Centre Meybeck Global Landform Classification.

'High Altitude - Meybeck Global Landform Classification':'C5',

#Near Coast - Metadata Derived

#Station is located near the coast - derived from standardised 'terrain' network provided metadata

'Near Coast - Metadata Derived':'C6',

#Near Coast - GSFC

#Stations located within 50km of the coast - using GSFC nearest to coastline dataset (0.01 degree grid)

'Near Coast - GSFC':'C7',

#Rural Station - Lenient Metadata Derived

#Station has been determined to be rural, using standardised network provided metadata, following lenient classifications.

'Rural Station - Lenient Metadata Derived':'C8',

#Urban Station - Lenient Metadata Derived

#Station has been determined to be urban, using standardised network provided metadata, following lenient classifications.

'Urban Station - Lenient Metadata Derived':'C9',

#Unclassified Station - Lenient Metadata Derived

#Station has been determined to be unclassified, using standardised network provided metadata, following lenient classifications.

'Unclassified Station - Lenient Metadata Derived':'C10',

#Rural Station - Strict Metadata Derived

#Station has been determined to be rural, using standardised network provided metadata, following strict classifications.

'Rural Station - Strict Metadata Derived':'C11',

#Urban Station - Strict Metadata Derived

#Station has been determined to be urban, using standardised network provided metadata, following strict classifications.

'Urban Station - Strict Metadata Derived':'C12',

#Unclassified Station - Strict Metadata Derived

#Station has been determined to be unclassified, using standardised network provided metadata, following strict classifications.

'Unclassified Station - Strict Metadata Derived':'C13',

#Rural Station - Anthrome

#Rural station as defined by using the UMBC Anthrome gridded classification dataset

'Rural Station - Anthrome':'C14',

#Urban Station - Anthrome

#Urban station as defined by using the UMBC Anthrome gridded classification dataset

'Urban Station - Anthrome':'C15',

#Rural Station - TOAR

#Rural station as defined by using a TOAR approach to classification (Tropospheric Ozone Assessment Report).

'Rural Station - TOAR':'C16',

#Urban Station - TOAR

#Urban station as defined by using a TOAR approach to classification (Tropospheric Ozone Assessment Report).

'Urban Station - TOAR':'C17',

#Unclassified Station - TOAR

#Unclassified station as defined using a TOAR approach to classification (Tropospheric Ozone Assessment Report).

'Unclassified Station - TOAR':'C18',

#Rural Station - Joly-Peuch

#Rural station as defined using a Joly-Peuch approach to classification

'Rural Station - Joly-Peuch':'C19',

#Unclassified Station - Joly-Peuch

#Unclassified station as defined using a Joly-Peuch approach to classification

'Unclassified Station - Joly-Peuch':'C20',

## Duplicate station

#Station has been decreed to be a duplicate (i.e. reporting the same data as from another network, but the data from another network has been preferred)

'Duplicate Station':'D',

## Extreme Data Flags

#Extreme Data - Scientifically Non-Feasible

#Data is greater than a scientifically feasible limit (variable by parameter).

'Extreme Data - Scientifically Non-Feasible':'E1',

#Extreme Data - Distributional Outlier

#Data is screened through adjusted boxplot to determine distributional outliers

'Extreme Data - Distributional Outlier':'E2',

#Extreme Data - Manually Decreed

#Data has been found and decreed manually to be extreme, a select section of data is flagged

'Extreme Data - Manually Decreed':'E3',

## Invalid Data Provider Flags

#Invalid Data Provider Flags - Scientifically Decreed

#Measurements are associated with data quality flags given by the data provider which have been scientifically decreed to suggest the measurements are associated with substantial uncertainty/bias

'Invalid Data Provider Flags - Scientifically Decreed':'F1',

#Invalid Data Provider Flags - Network Decreed

#Measurements are associated with data quality flags given by the data provider which have been decreed by the reporting network to suggest the measurements are associated with substantial uncertainty/bias

'Invalid Data Provider Flags - Network Decreed':'F2',

## Recurring Value

#Do check for persistently recurring values. check is done by using a moving window of 12 measurements. If 11/12 of values in the window are valid.

'Recurring Value':'G',

## Hitting LOD

#More than 25% of station data (per month) is below/above exactly on the LOD, therefore significantly biasing the remaining distribution of data

'Hitting LOD':'H',

## Insufficient Measurement Resolution Flags

#Insufficient Measurement Resolution - Reported

#The reported resolution of measurement is coarser than a set limit (variable by measured parameter)

'Insufficient Measurement Resolution - Reported':'I1',

#Insufficient Measurement Resolution - Empirical

#The resolution of measurement is analysed month by month. If the minimum difference between observations is coarser than a set limit (variable by measured parameter), measurements are flagged.

'Insufficient Measurement Resolution - Empirical':'I2',

## Limit of Detection Flags

#Below Documented Lower Limit of Detection

#Measurement is below or equal to the instrumental documented lower limit of detection

'Below Documented Lower Limit of Detection':'L1',

#Below Reported Lower Limit of Detection

#Measurement is below or equal to the network reported lower limit of detection

'Below Reported Lower Limit of Detection':'L2',

#Above Documented Upper Limit of Detection

#Measurement is above or equal to the instrumental documented upper limit of detection

'Above Documented Upper Limit of Detection':'L3',

#Above Reported Upper Limit of Detection

#Measurement is above or equal to the network reported upper limit of detection

'Above Reported Upper Limit of Detection':'L4',

## Measurement Methodology Flags

#Methodology Not Mapped

#The measurement methodology used has not yet been mapped to standardised dictionaries of measurement methodologies

'Methodology Not Mapped':'M1',

#Assumed Primary Sampling

#A level of assumption has been made in determining the primary sampling type

'Assumed Primary Sampling':'M2',

#Assumed Sample Preparation

#A level of assumption has been made in determining the sample preparation

'Assumed Sample Preparation':'M3',

#Assumed Measurement Methodology

#A level of assumption has been made in determining the measurement methodology

'Assumed Measurement Methodology':'M4',

#Unknown Primary Sampling Instrument

#The specific name of the primary sampling instrument is unknown

'Unknown Primary Sampling Instrument':'M5',

#Unknown Measuring Instrument

#The specific name of measuring instrument is unknown

'Unknown Measuring Instrument':'M6',

#Erroneous Primary Sampling

#The primary sampling is not appropriate to prepare the specific parameter for subsequent measurement.

'Erroneous Primary Sampling':'M7',

#Erroneous Sample Preparation

#The sample preparation is not appropriate to prepare the specific parameter for subsequent measurement.

'Erroneous Sample Preparation':'M8',

#Erroneous Measurement Method

#The measurement methodology used is not known to be able to measure the specific parameter. Only do check when known (or have assumed method).

'Erroneous Measurement Methodology':'M9',

#Invalid QA Measurement Method

#The specific measurement methodology has been decreed not to conform to QA standards as the method is not sufficiently proven/ subject to substantial biases/uncertainty. Only do check when known (or have assumed method).

'Invalid QA Measurement Methodology':'M10',

#No EU Reference/Equivalent Measurement Method

#The specific measurement methodology does not have European standard reference or equivalence status for the specific parameter being measured, as decreed by the European Parliament.

'No EU Reference/Equivalent Measurement Method':'M11',

#No US Reference/Equivalent Measurement Method

#The specific measurement methodology does not have US standard reference or equivalence status for the specific parameter being measured, as decreed by the United States government.

'No US Reference/Equivalent Measurement Method':'M12',

## Temporal Period Flags

for classifying periods of time measurements are made within

#Daytime

#Time of measurement is daytime. Done by calculating the solar elevation angle for a latitude/longitude/measurement height at a certain timestamp.

'Daytime':'P1',

#Nightime

#Time of measurement is nighttime. Done by calculating the solar elevation angle for a latitude/longitude/measurement height at a certain timestamp.

'Nighttime':'P2',

#Weekday

#Time of measurement is weekday.

'Weekday':'P3',

#Weekend

#Time of measurement is weekend.

'Weekend':'P4',

## Temporal Representation Flags

Periods are decreed not representative if <50% of period is not represented with valid measurements (after screening by key QA flags), or a data gap exists which is >15% of total length of period

#Not Hourly Representative

'Not Hourly Representative':'R1',

#Not Diurnally Representative

'Not Diurnally Representative':'R2',

#Not Weekly Representative

'Not Weekly Representative':'R3',

#Not Monthly Representative

'Not Monthly Representative':'R4',

#Not Annually Representative

'Not Annually Representative':'R5',