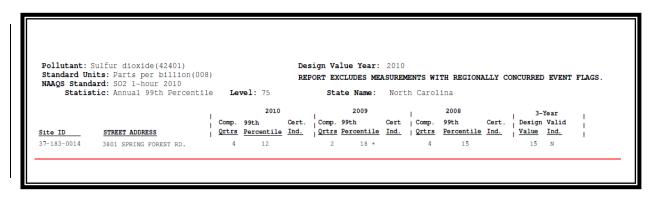# SO$_2$ Design Values in AQS Standard Retrievals AMP480/ Design Value Report

Introduction:
The AQS Design Value report for SO$_2$ displays a Design Value for a selected Design Value year. The 3-year Design Value is based on the average of a 3-year period, which includes the selected year plus the two prior years.

Also displayed is the following information for each year: the number of Complete Quarters for that year, the 99th Percentile (sample measurement of the year that falls in the 99th percentile) and the Certification Indicator, which indicates whether or not the data for that year have been certified by the reporting agency. For the 3-year Design Value, a Validity Indicator of "Y" indicates a valid Design Value.

```
Pollutant: Sulfur dioxide(42401)                    Design Value Year:  2010
Standard Units: Parts per billion(008)              REPORT EXCLUDES MEASUREMENTS WITH REGIONALLY CONCURRED EVENT FLAGS.
NAAQS Standard: SO2 1-hour 2010
     Statistic: Annual 99th Percentile    Level: 75           State Name:   North Carolina

                                    2010              2009              2008           3-Year
                            | Comp. 99th    Cert. | Comp. 99th    Cert | Comp. 99th    Cert. | Design Valid |
Site ID       STREET ADDRESS| Qrtrs Percentile Ind.| Qrtrs Percentile Ind.| Qrtrs Percentile Ind. | Value  Ind.  |
37-183-0014   3801 SPRING FOREST RD.     4     12           2     18 *         4     15           15    N
```

Background:
Design Values are the metrics that are compared to the NAAQS levels in order to determine compliance with the standard[1]. The design value for the primary 1-hour NAAQS is derived from the 3-year average of annual 99th percentile daily maximum 1-hour values for a monitoring site, also known as "the 1-hour primary standard design value." A list of the SO$_2$ Standards can be found in Appendix B.

SO$_2$ Design Values are calculated at the site-level. However, data from two or more monitors in the same year at the same site, reported to AQS under different POCs, may not be combined to meet data completeness requirements.

Calculating the 1-Hour Standard:
A series of calculations are performed to determine the 1-hour standard.

- Each **reported** sample value is converted to a **standard** sample value.
- A daily summary of the **standard** sample values is calculated.
- A quarterly summary of the **standard** sample values is calculated.
- An annual summary of the **standard** sample values is calculated.

---

[1] Refer to **40 CFR Part 50, Appendix T**

The calculations move in order from daily to annual, in the following order:

1. Standard Sample Value
   From each sample measurement reported, AQS computes a "standard value" with:
   Units: ppb
   Digits after decimal: 1
   Extra digits are truncated.

2. Compute Daily Summaries of Measurement Data
   AQS computes daily summaries for each monitor for each calendar day in which sample measurements are taken. The summaries include the maximum value measured for the day, and an indication as to whether or not measurements were taken for 75% of the hours for the day *(i.e.*, 18 hours).

   Multiple daily summaries are computed for each day where SO_2 sample measurements exist; each unique summary is identified by the combination of Exceptional Data Type (EDT_ID) and pollutant standard.

   Exceptional Data Type (EDT_ID):
       0: Summary for this day contains no flagged exceptional event data
       1: Day contained flagged exceptional event data; summary excludes all flagged data
       2: Day contained flagged exceptional event data; summary includes all flagged data
       5: Day contained flagged exceptional event data; summary excludes only flagged data
   that has been Regionally-concurred (with respect to the appropriate pollutant standard for the record)

   Pollutant Standard: For each Exceptional Data Type (EDT_ID) defined above, a summary record is created for each of the following SO_2 standards:
   - Annual primary standard
   - 1-hour primary standard (only for hourly data, *i.e.,* duration=1)

   The following information is also used in the calculations.

   Count of Observations: For each day, the count of values with the appropriate Exceptional Data Type (EDT_ID). Count of values does not include "missing" sample measurements that have a null data code.

   Example: Total of 20 sample measurements. Six have exceptional event flags and 3 have EPA regional concurrence. For each Exceptional Data Type (EDT_ID), here is the Count of Observations:
   - 0: record will not exist
   - 1: Count of Observations = 14
   - 2: Count of Observations = 20
   - 5: Count of Observations = 17

Arithmetic Mean: Value = $\dfrac{\text{Sum of values}}{\text{Count of Observations}}$ where:

Sum of values: sums measurement values with the appropriate flagging and concurrence for the Exceptional Data Type (EDT_ID).

Count of Observations: From above

No rounding or truncation, and, in the event that all values for the day have been excluded because of the Exceptional Event Rule, the Arithmetic Mean is NULL.

Observation Percent: Value = $\dfrac{TotalNumberOfObservations}{MaximumNumberForDay}$ where:

Total Number Of Observations - the total number of values for the day, excluding none (i.e., no exclusions for flagged/concurred values.) For the above example, this value will be 20.

Maximum Number For Day - the maximum number of values for day (*i.e.*, 24)

Value is rounded to 0 digits after decimal.   For the above example, the Observation Percent is 20/24 = 83% for all Exceptional Data Types.

Maximum Sample Value: Maximum value for day with the appropriate Exceptional Data Type (EDT_ID) and concurrence.

Maximum Collection Hour: Hour that the Maximum Sample Value was reported. Where the same value was reported for multiple hours, it is the first hour that the value was reported.

Number of Values Greater than Primary Standard: Values, with the appropriate Exceptional Data Type (EDT_ID) and concurrence, greater than the primary standard. (For the Annual Standard, this is the count of values > 30 ppb, and for the 1-Hour standard, count of values > 75 ppb.)

Daily Criteria Indicator: Flag is set to 'Y' if the Observation Percentage is >= 75%.

3. Compute Quarterly Summaries of Measurement Data

AQS computes multiple summaries for each quarter in which SO_2 sample measurements exist; each unique summary is identified by the combination of Exceptional Data Type (EDT_ID) and the pollutant standard. (See item 2 above: Compute Daily Summaries of Measurement Data.)

Valid Day Count: Count of daily summaries with the corresponding pollutant standard and exceptional data type, where the summary criterion is met.

Arithmetic Mean: Value = $\dfrac{\text{Sum of values}}{\text{Count of Observations}}$ where:

Sum of Values – sums measurement values with the appropriate Exceptional Data Type (EDT_ID) and concurrence.

Count of Observations: From above.

Format: No rounding or truncation. In the event that all values for the day have been excluded because of the Exceptional Event Rule, the Arithmetic Mean is NULL.

Observation Percent: Value = $\dfrac{TotalNumberOfObservations}{MaximumNumberForQuarter}$ where:

Total Number Of Observations - the total number of values for the quarter, no exclusions for flagged/concurred values.

Maximum Number For Quarter –the total number of hours in the quarter

Percent Days: Value = $\dfrac{ValidDayCount}{NumberOfDaysInQuarter}$ where:

- *Valid Day Count*: From above.
- *Number Of Days In Quarter* - the number of calendar days in the quarter.

Quarterly Minimum: Minimum Daily Maximum value for valid days in the quarter with the appropriate flagging and concurrence.

Quarterly Maximum: Maximum Daily Maximum value over all days, not just valid days, in the quarter with the appropriate flagging and concurrence.

Daily Criteria Indicator: Flag is set to 'Y' if the Observation Percentage is >= 75%.

To meet completeness for the 1-hour primary standard, the summaries for each quarter include the percentage of valid days in the quarter. The 3-year Design Value requirements for the 1-hour primary standard require that the summaries for each quarter include the minimum and maximum daily maximum values for the quarter. However, these values are not displayed on the 480/Design Value report, but are stored in AQS and are available through a database query using Discoverer.

4. Compute Annual Summaries of Measurement Data

AQS computes multiple annual summaries for each year in which SO_2 sample measurements exist; each unique summary is identified by the combination of exceptional data type + pollutant standard. (See 2. Compute Daily Summaries of Measurement Data.)

Observation Percent: Value = $\dfrac{TotalNumberOfObservations}{MaximumNumberForYear}$ where:

Total Number Of Observations: Total number of values for the quarter. No exclusions for sample measurements affected by exceptional events, regardless of concurrence.

Maximum Number For Year: Total number of hours in the year.

Valid Day Count: Count of daily summaries with the corresponding Exceptional Data Type (EDT_ID) and pollutant standard, where the summary criterion is met.

Required Day Count: Number of calendar days in a year.
Count of Complete Quarters: Number of quarterly summaries, with the corresponding Exceptional Data Type (EDT_ID) and pollutant standard,, where the summary criterion is met.
Criteria Indicator: Flag is set to "Y" if the Count of Complete Quarters = 4

Arithmetic Mean for the annual standard: Value = $\dfrac{\text{Sum of values}}{\text{Count of Observations}}$ where:

Sum of Values: sum of **hourly** measurement values with the appropriate flagging and concurrence for the exceptional data type.
Count of Observations: From above.
No rounding or truncation, and, in the event that all values for the day have been excluded because of the Exceptional Event Rule, the Arithmetic Mean is NULL.

Arithmetic Mean for the 1-hour standard: = $\dfrac{\text{Sum of valid daily maxima}}{\text{Count of Valid Days}}$ where:

Sum of valid daily maxima: **sum of valid daily maxima**, from the daily summaries, with the appropriate Exceptional Data Type (EDT_ID) and pollutant standard.
Count of Valid Days: From above.
Format: No rounding or truncation. In the event that all values for the day have been excluded because of the Exceptional Event Rule, the Arithmetic Mean is NULL.

Arithmetic Standard Deviation: Value = $\sqrt{\dfrac{\left(n*\sum_{j=1}^{n} s_j^2\right) - \left(\sum_{j=1}^{n} s_j\right)^2}{n*\left(n-1\right)}}$ where the definition of *s* and

n depend on the pollutant standard for the record:

- Annual Standard:
  s = Value with appropriate Exceptional Data Type (EDT_ID) and concurrence
  n = Count of Observations
  and *n* > 1.

- 1-Hour Standard
  s = Daily Maximum with appropriate Exceptional Data Type (EDT_ID) and concurrence
  n = Count of Valid Days
  and *n* > 1.

Maximum Values 1 – 10 and *Date and Time of Maximum Values 1 - 10:*

- Annual Standard: Ten highest measurement values for year.
- 1-Hour Standard: Ten highest Daily Maximum Values where the summary criteria is met.

Percentile Values:

- Annual Standard: Percentiles are computed directly from the measurement values as per the algorithm below. (Note: Percentiles other than the 99th are computed but only the 99th is displayed on the report; other percentiles are only available through a database query using Discoverer or another SQL tool.)
- 1-Hour Standard: Percentiles are computed for two distinct sets of data (see the algorithm below) and the <u>larger</u> of these values is taken as the percentile value for the year;   Set 1 has the valid daily maxima (*i.e.,* maximum for days with 18 or more total hours), and set 2 has all daily maxima (*i.e.*, includes all days with at least 1 non-excluded value.) In both cases the daily maxima must correspond to the exceptional data type and pollutant standard of the annual summary being computed.
  Algorithm:

$$k = CEIL\left(n - \frac{nP}{100}\right) \text{ where:}$$

  - k is the kth maximum (e.g. 1st maximum for k = 1)
  - *CEIL(x)* is the function that produces the smallest integer greater than or equal to x; e.g. *CEIL(1.0)* = 1 and *CEIL(1.1)* = 2.
  - *n* is the total number of values (days) with at least 1 non-excluded value
  - *P* is the percentile level, e.g. 99 for the 99th percentile

Format: No rounding or truncation.

5. <u>Compute 3-year Summaries of Measurement Data</u> (1-hour annual summaries)
AQS computes the 3-year summaries "on-the-fly" (i.e. when they are retrieved), rather than calculated and stored as permanent records in the database.

The 3-year Design Values are computed at the site-level rather than at the monitor-level. When there are multiple monitors at a site, these rules apply:
- Within any given year, data from two or more monitors will NOT be combined.
- For each year of the 3-year period, data from the monitor with the highest average quarterly completeness will be used.

Multiple 3-year summaries are computed for each 3-year period in which SO_2 sample measurements exist; each unique summary is identified by the combination of Exceptional Data Type (EDT_ID) and pollutant standard. (See 2. Compute Daily Summaries of Measurement Data.)

For each 3-year period, each year would be defined as follows:
Design Value Year: the <u>ending year</u> of the 3-year summary period.
Design Value Year - 1: the year <u>prior to</u> the DV Year of the 3-year summary period.
Design Value Year - 2: two years <u>prior to</u> the DV Year of the 3-year summary period.

For each year (DV, DV-1, and DV-2), the following five fields are calculated:

Count of Complete Quarters: number of quarters with > 75% completeness (comes from the Annual Summary.

99th Percentile: comes from the Annual Summary.

99th Percentile Completeness Indicator: comes from the Annual Summary.

High Substituted 99th Percentile: follows these rules:

- If all quarters for the year have *Completeness Percent* > 75%, then *High Substituted 99th Percentile = 99th Percentile*
- If any quarter for the year has *Completeness Percent* < 50%, then *High Substituted 99th Percentile* = NULL
- If any quarter has *Completeness Percent* < 75%, the annual 99th percentile (*Percentile Value for Annual Standard*) is calculated using all hourly measurements, and, for each quarter where the Percent Days < 75%, the 3-year quarterly **maximum^2** is substituted for **all missing values**.

Low Substituted 99th Percentile: follows these rules:

- If all quarters for the year have *Completeness Percent* > 75%, then *Low Substituted 99th Percentile = 99th Percentile*
- If any quarter has *Completeness Percent* < 75%, the annual 99th percentile (*Percentile Value for Annual Standard*) is calculated using all hourly measurements, and, for each quarter with Percent Days < 75%, the 3-year quarterly **minimum** is substituted for **all missing values**.

For each 3-year period, the following are calculated:

3-Year 1-Hour Design Value: Average of annual 99th percentile values.   Format: Rounded to 0 digits after decimal.

3-Year 1-Hour Design Value Validity Indicator: Set to 'Y' if **any** of the following are true, else set to 'N'

- Count of Complete Quarters = 4 for all 3 years
- The 3-Year 1-Hour Design Value <= the hourly standard and **all** of the following are true:
  - For each quarter, there are a total of 200 valid days for that quarter in the 3-year period
  - The Percent Days for each quarter in each year >= 50% (i.e. High-Substituted 99th Percentile values are not null)
  - The average of the High-Substituted 99th Percentile values <= the hourly standard
- The 3-Year 1-Hour Design Value > the hourly standard AND at least 75% of the days in each quarter in the three year period have at least one reported hourly value.
- The 3-Year 1-Hour Design Value > the hourly standard and **all** of the following are true:
  - For each quarter, there are a total of 200 valid days for that quarter in the 3-year period
  - The average of the Low-Substituted 99th Percentile values > the hourly standard

---

^{2} The 99th percentile is the 4th maximum for a complete year (i.e. 366 days). Unless the substituted value (min or max) is in the range of maximum 1 – 4, it will not become the new 99th percentile; the only effect will be on the number of observations that determine which maximum is the 99th percentile. (You do not have to look at all data for the year, but only the top 4 maxima.) Applies to both the high-substituted and low-substituted values.

Appendix A: Definitions

Term	Definition
Design Value	The metric (i.e., statistic) that is compared to the NAAQS level to determine compliance.
Exceptional Event	An event that affects air quality, is not reasonably controllable or preventable, is an event caused by human activity that is unlikely to recur at a particular location or a natural event, and is determined by the EPA in accordance with 40 CFR 50.14 to be an exceptional event.
Exceptional Data Type	An AQS designation as to how a summary is affected by exceptional events. Possible values include: 0 – No Events: None of the measurement data in the time period for the summary has been flagged for exceptional event exclusion. 1 – All Events Excluded: The summary excludes any measurements that have been flagged for exceptional event exclusion independent of whether or not the EPA has concurred with the flagging. 2 – Events Included: Measurements included in the summary have been flagged for exceptional event exclusion. (Note: They are included in the summary independent of whether or not the EPA has concurred with the flagging.) 5 – Concurred Events Excluded: The summary excludes any measurements that have been flagged for exceptional event exclusion and the flagging has been concurred by the EPA Regional Office. Note: For any site/monitor and summary time period, either a EDT 0 summary will exist (meaning that no data was flagged) OR all three EDT 1, EDT 2 and EDT 5 summaries will exist (meaning that data was flagged).
Daily Maximum	The highest 1-hour value within a given calendar day.
National Ambient Air Quality Standards	Sets the levels of air quality for the United States, in the Code of Federal Regulations (40 CFR Part 50), to protect the population's health and the environment.
Parameter	An AQS term for a physical quantity that is measured. This may be a pollutant concentration (such as SO2) or a meteorological quantity such as temperature.
Year	A calendar year.

Appendix B: SO$_2$ Standards

1. 1-Hour Primary Standard

Level: 75 ppb
Form: 99th percentile of daily maxima for year
Completeness: Annual: 4 quarters with 75% valid days in quarter (days valid if 75% hours present)

2. Annual Primary Standard

Level: 30 parts per billion (ppb)
Form: Annual arithmetic mean of hourly measurements.
Completeness: 4 quarters with 75% of hours in quarter reported.

3. Daily (24-Hour) Primary Standard

Level: 140 parts per billion (ppb)
Form: Annual second highest maximum of daily (24-hour) averages
Completeness: 4 quarters with 75% valid days in quarter (days valid if 75% hours present)

4. Secondary Standard

Level: 500 ppb
Form: Annual 2nd highest maximum of 3-Hour block averages
Completeness: 4 quarters with 75% hours in quarter.

5. 3-Year 1-Hour Design Value

Level: 75 ppb
Form: Average of annual 99th percentile
Completeness: 3 complete years

6. 5-Minute Maximum by Hour (AQS parameter code 42406)

There are no regulatory requirements for this data. It is processed as if it were non-criteria data.

7. 5-Minute Block Average Data (AQS parameter code 42401, duration H)

There are no regulatory requirements for this data. It is processed as if it were non-criteria data.

Appendix C: SO2_Design_Values folder in Discoverer

Queries against this folder are resource-intensive; <u>we recommend limiting your query as much as possible before running it.</u> You can use the Workbook Wizard to select the state, county and site, as well as the Design Value Year, before running the query. Fields available:

- State code
- County code
- Tribal code
- Site ID
- Parameter code
- Pollutant name
- DV Year *(the year for which you are calculating the Design Value)*
- Design value
- Street address
- Local site ID
- Local site name
- Latitude
- Longitude
- State name
- County name
- Tribal name
- City name
- Urbanized area
- CBSA name
- Year DV POC *(The POC used to supply the data for the year this design value was calculated.)*
- Year DV Complete Quarters
- Year DV 99th percentile
- Year DV Complete Ind *(Completeness Indicator – Y or N)*
- Year DV Cert IND *(Certification Indicator – Y or N)*
- Year DV-1 POC *(The POC used in this year. "DV-1" is the year prior to the year for which you are calculating the Design Value.)*
- Year DV-1 Complete Quarters
- Year DV-1 99th percentile
- Year DV-1 Complete Ind *(Completeness Indicator – Y or N)*
- Year DV-1 Cert IND *(Certification Indicator – Y or N)*
- Year DV-2 POC *(The POC used in this year. "DV-2" is two years prior to the year for which you are calculating the Design Value.)*
- Year DV-2 Complete Quarters
- Year DV-2 99th percentile
- Year DV-2 Complete Ind *(Completeness Indicator – Y or N)*
- Year DV-2 Cert IND *(Certification Indicator – Y or N)*
- Standard ID
- Standard description
- Statistic name
- Primary standard level
- Standard units code
- Standard units name
- Three year design validity

Example of one row (site metadata + 3-year summary with annual and quarterly statistics):

State code	County code	Tribal code	Site id	Parameter code	Pollutant name	Dv year	Design value	Street address		Local site id	Local site name
37	183	NULL	0014	42401	Sulfur dioxide	2010	15	3801 SPRING FOREST RD.		NULL	E Millbrook Middle School

Latitude	Longitude	State name	County name	Tribal name	City name	Urbanized area	CBSA name
+35.856111	-78.574167	North Carolina	Wake	NULL	Raleigh	Raleigh, NC	Raleigh-Cary, NC

Year DV POC	Year DV Complete Quarters	Year DV 99th Percentile	Year DV Complete IND	Year DV Cert IND
2	4	12	Y	NULL

Year DV -1 POC	Year DV -1 Complete Quarters	Year DV -1 99th Percentile	Year DV -1 Complete IND	Year DV -1 Cert IND
2	2	18	N	NULL

Year DV -2 POC	Year DV -2 Complete Quarters	Year DV -2 99th Percentile	Year DV -2 Complete IND	Year DV -2 Cert IND
1	4	15	Y	NULL

Standard id	Standard description	Statistic name	Primary standard level	Standard units code	Standard units name	Three year design validity	Edt id
19	SO2 1-hour 2010	Annual 99th Percentile	75	008	Parts per billion	N	0