

# Annual Statistics and Quality Assurance of the 2023 Ambient Monitoring Archive

December 2025

## Table of Contents

Purpose .....	1
Steps for Calculating Annual Statistics .....	2
Calculate derived local conditions .....	2
Pre-process the data .....	2
Calculate daily averages.....	2
Calculate annual averages from daily averages.....	3
Descriptions of R Code Used to Calculate Annual Statistics .....	4
Quality Assurance of the 2023 Archive.....	4
QA of the overlap years .....	5
QA of the newest year .....	9
Frequent Questions .....	10
What is the purpose of the Archive? Why does the Archive include data not in AQS? .....	10
Why are local conditions derived?.....	10
Why is there different criterion to calculate daily averages for remote and non-remote data?.....	10
Why are annual statistics separated by sampling duration?.....	10
How is the minimum number of samples of the different sampling durations determined when calculating a daily average?.....	10
Where is the information in the lookup table tabs obtained? .....	11
What is the minimum number of daily averages required to create a valid quarter? .....	11
How many valid quarters of data are needed to calculate an annual average? .....	11
What data are excluded from annual averages in the 2023 Archive?.....	11
Common Acronyms.....	12
References .....	12
Appendix .....	12

## Purpose

This document outlines the major steps in calculating annual statistics and Quality Assurance (QA) procedures for each pollutant, site, sampling duration, and year for the 2023 Ambient Monitoring Archive (AMA) for the Hazardous Air Pollutants (HAPs) (i.e., “the 2023 Archive”).

## Steps for Calculating Annual Statistics

### Calculate derived local conditions

Air quality samples are collected using either local conditions (LC) or standard conditions (STD), each of which have different applications (e.g., risk, model evaluation, defining detection limits, etc.). In instances where LC is not available, but STD is (i.e., when there is no collocated or nearby temperature or pressure data), the LC can be derived as outlined below. In short, when both conditions are available, the ratio between LC and STD is calculated and the best available ratio is applied elsewhere where needed. In the 2023 Archive, derived LC is calculated for approximately 1.17% of records.

### Calculate local conditions/standard conditions ratios

1. Subset Archive data to those records that have both STD and LC available.
2. Calculate the average STD and the average LC by pollutant/site/POC/day/sampling duration. (POC is the Parameter Occurrence Code.)
3. Average the ratios calculated above by site/day. Note: this is averaged across pollutants and sampling durations.
4. Average the ratios calculated above by site/quarter.
5. Average the ratios calculated above by site/year.
6. Average the ratios calculated above for the site across all available years.

### Pre-process the data

#### Subset data

1. Removed data that are not relevant or appropriate for the calculation of annual statistics (e.g., a sampling duration of 1 month and integrated 2-weeks samples, pollutants that are not HAPs, etc.).

#### Apply the LC/STD ratios to derive LC

1. Identify all the records which are only presented in STD (i.e., `SAMPLE_VALUE_STD_FINAL_TYPE=S`).
2. If the ratio exists for that pollutant/site/POC/day/sampling duration, derive LC by multiplying the sample collected using STD and the ratio.
3. If the above ratio does not exist but the ratio exists for that site/day, derive LC by multiplying the sample collected using STD and the ratio.
4. If the above ratio does not exist but the ratio exists for that site/quarter, derive LC by multiplying the sample collected using STD and the ratio.
5. If the above ratio does not exist but the ratio exists for that site/year, derive LC by multiplying the sample collected using STD and the ratio.
6. If the above ratio does not exist but the ratio exists for that site, derive LC by multiplying the sample collected using STD and the ratio.
7. If the ratio by site is missing for the sample collected using STD, assume  $ratio = 1$  (i.e., LC equal STD).

Sample are either originally collected in LC, converted to LC using collocated or nearby temperature and pressure data, or LC is derived using the methodology described above. These values are carried forth to calculate daily averages unless otherwise noted.

### Calculate daily averages

#### Separate data

1. Separate remote and non-remote data. Remote data include records from NOAA and MIT data sources. All other data collected from sources other than NOAA and MIT are considered “non-remote.”
2. Separate the non-remote data by minute sampling durations and hourly sampling durations.

### **Calculate daily averages from the minute sampling durations (non-remote data)**

1. Calculate the number of samples per pollutant/site/POC/day/hour/sampling duration. If the minimum number of samples is met (see Table 4), average up to the hour. This constitutes a valid hourly average. If the minimum number of samples is not met, all samples are removed.
2. From all valid hourly averages, calculate the number of samples per pollutant/site/POC/day/sampling duration. If there are at least 18 valid hourly averages in a day (see Table 4), average up to the day. This constitutes a valid daily average from the minute data. If the minimum number of hourly averages are not met, all samples are removed.

### **Calculate daily averages from the hourly sampling durations (non-remote data)**

1. Calculate the number of samples per pollutant/site/POC/day/sampling duration. If the minimum number of samples is met (see Table 4), average up to the day. This constitutes a valid daily average from the hourly data. If the minimum number of samples is not met, all samples are removed. Note: Sampling durations of 90 MINUTES and 150 MINUTES are considered to have an hourly sampling duration.

### **Average across POCs (non-remote data)**

1. If there are multiple POCs per pollutant/site/day/sampling duration, remove all samples equal to zero if at least 50% of samples are NOT zero.
2. If there are multiple POCs per pollutant/site/day/sampling duration, remove all samples NOT equal to zero if more than 50% of the samples ARE zero.
3. After the appropriate valid collocated daily averages are removed (if necessary), average the remaining valid collocated daily averages by pollutant/site/day/sampling duration across POCs.

Note: Removing some collocated POCs ensures that collocated values that are mostly zero will result in a zero daily average and collocated values that are mostly NOT zero will result in a daily average that does not contain zeros.

### **Calculate daily averages for remote data**

1. Average by pollutant/site/day/sampling duration. Note: The remote data do not require a minimum number of minute or hourly samples. The remote data may have collocated monitors distinguished by POCs for a given pollutant/site/day/sampling duration.

### **Calculate annual averages from daily averages**

An annual average is calculated by averaging valid daily averages by pollutant/site/year/sampling duration meeting the criteria outlined below.

#### **Valid quarters**

Annual averages are calculated for a given pollutant/site/year/sampling duration only if there are at least *three* valid quarters for a given pollutant/site/year/sampling duration. A valid quarter is defined as a quarter having at least *six* daily averages per pollutant/site/quarter/sampling duration (McCarthy et al., 2007). If there are less than three valid quarters, all daily averages for a given pollutant/site/year/sampling duration are removed and an annual average is not calculated. Note: All valid daily averages are used in the calculation of an annual average if the three valid quarters threshold is met. Therefore, it is possible for an annual average to contain data from one “invalid” quarter of data if the criterion for the remaining three quarters is met. Note: a daily non-detect average assigned as zero can be included in the calculation of a valid quarter.

#### **Calculate annual averages**

Annual averages and corresponding statistics are calculated by pollutant/site/year/sampling duration (see Table 1 for all the annual statistics calculated). Some annual statistics are also calculated from valid daily averages using regression on order statistics (ROS) that allows for censored values (i.e., daily averages that are non-detect). To ensure a stable

statistic, the ROS annual average is only calculated if the percentage of valid daily averages below the method detection limit (MDL) does not exceed 80 (Helsel, 2005). The MDLs for the daily censored values are assigned:  $\min\{MDL, \min(\text{day in year})\}$ , where  $\min(\text{day in year})$  is the minimum daily non-zero values for a given pollutant/site/year/sampling duration. Note: The MDLs are in STD while the original samples may be collected using LC. If there were multiple MDLs used in the construction of the daily average (e.g., through multiple POCs or through a sub-daily sampling duration), the average MDL is taken.

**Table 1. Set of calculated annual statistics given the condition of the collected data and the treatment of non-detects.**

Condition of Data	Treatment of Non-Detects	Annual Statistics Calculated
Local, converted local, and derived local	zero	arithmetic mean, variance, maximum, 10th percentile, 25th percentile, median, 75th percentile, and 90th percentile
Local, converted local, and derived local	censored	arithmetic mean, 10th percentile, 25th percentile, median, 75th percentile, and 90th percentile
Standard only	zero	arithmetic mean, variance, maximum, 10th percentile, 25th percentile, median, 75th percentile, and 90th percentile
Standard only	censored	arithmetic mean, 10th percentile, 25th percentile, median, 75th percentile, and 90th percentile

## Descriptions of R Code Used to Calculate Annual Statistics

- `AMA_analysis.R`: Sources and calls all functions used to create annual statistics.
- `AMA_LC2STDRatio.R`: Creates an average ratio between LC to STD based on an average across days, quarters, years, and all years for a given site.
- `AMA_preprocessing.R`: Pre-processes the data including removing some select sites and parameters and applies the LC/STD ratio to data if applicable.
- `AMA_duration2daily.R`: Calculates daily averages by pollutant/site/day/sampling duration for sub-hourly and sub-daily sampling durations.
- `AMA_daily2annual.R`: Calculates annual statistics by pollutant/site/year/sampling duration based on local and standard conditions and treatments of non-detect data.
- `AMA_file.R`: Creates the final annual file.

## Quality Assurance of the 2023 Archive

QA of the 2023 Archive is primarily performed by comparing the ROS annual averages from the 2023 Archive to the 2022 Archive by pollutant/site/year/sampling duration using the criteria outlined below. Two sets of checks were performed: (1) analyzing “overlap years” and (2) analyzing the newest year of data. Currently, there is no corrective action for flagged records and/or annual averages.

Annual averages between the old and new versions of the Archive may be (1) the same, (2) different, (3) added, or (4) removed. Due to the comparison to the previous version of the Archive, these categories are only applied to the “overlap years” (i.e., 1990 – 2022, with the newest year, 2023, excluded). Individual records are investigated to identify the reasons why annual averages fall into these four categories. Most annual averages in the overlap years are the *same* between the 2022 Archive and the 2023 Archive (Table 2).

The *suspect categories* are categories used to identify the newest annual averages (i.e., annual averages from the year 2023) that are substantially different than the rest of the annual averages (i.e., averages from the years 1990 – 2022).

The suspect categories depend on the comparison sets for the newest annual averages. The suspect categories first establish if (1) there is a corresponding annual average in 2022 for that pollutant/site/sampling duration, (2) there is no corresponding annual average in 2022 but there is a time series for that pollutant/site/sampling duration, or (3) there is no annual average for 2022 and no time series for that pollutant/site/sampling duration.

**Table 2. Counts of annual averages in each QA category in the 2023 Archive.**

QA Category Name	Count
All annual averages	327,612
Annual averages (ROS*)	237,135
Overlap years^ (ROS)	229,162
Overlap years – Same (ROS)	214,108
Overlap years – Different (ROS)	10,870
Overlap years – Added (ROS)	4,184
Overlap years – Removed% (ROS)	428
Newest year& (ROS)	7,973

\*ROS = Regression on Order Statistics

^Overlap years = 1990 – 2022

%“Removed” is not included in the count of total ROS annual averages

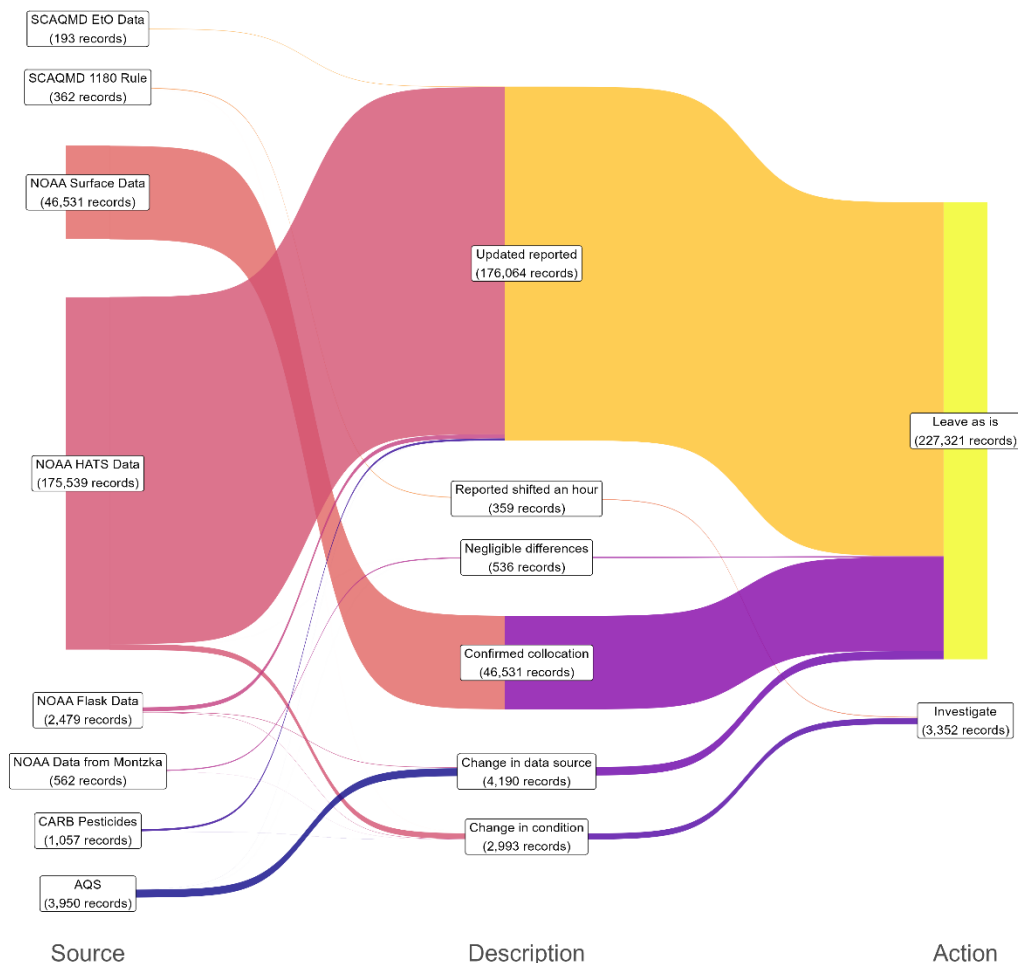
&Newest year = 2023

## QA of the overlap years

The QA of the overlap years (i.e., the years 1990 – 2022) explores the individual records that comprise the annual averages that are (1) different, (2) added, or (3) removed between the 2022 Archive and the 2023 Archive.

### Different annual averages – different reported values

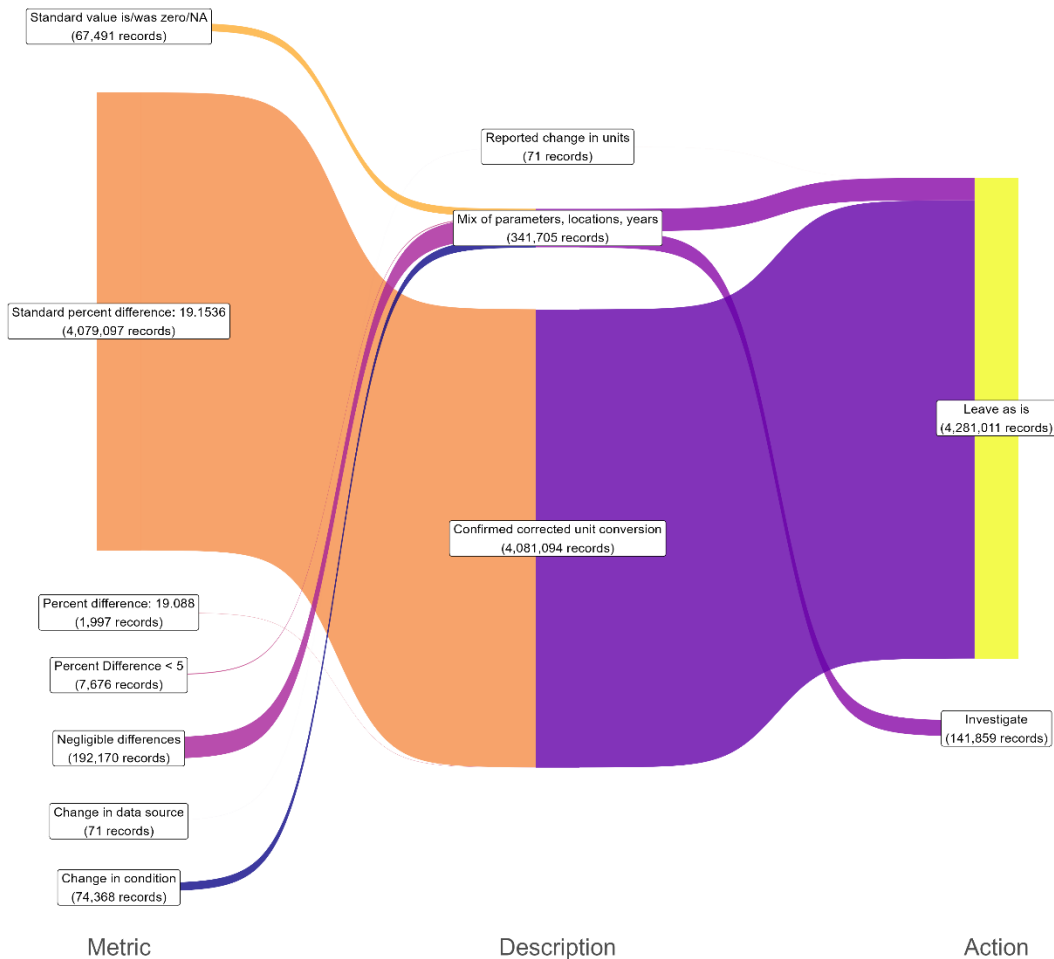
The individual records from the different annual averages are further divided into records that have (1) different reported values and (2) the same reported values but different final values. The individual records that are identical are not evaluated in the QA. The different reported values are first categorized by data source (e.g., AQS, TCEQ, NOAA, etc.) then grouped by observed dissimilarities in the data causing the different reported values. There were six dissimilarities identified among the different reported values. There were updated reported values as noted in the COMMENT field of the output data files observed for the NOAA HATS data, NOAA flask data, pesticides data from CARB, and ethylene oxide data from SCAQMD. Records from SCAQMD 1180 Rule were shifted by an hour (i.e., the corresponding start time for a given record was shifted by one hour for some records). NOAA data obtained from Stephen Montzka had small differences attributable to rounding. NOAA surface data had previously unaccounted for collocation that are now properly identified. Some AQS and NOAA flask data had a change from its previous data source (i.e., ODEQ and MNPCA for AQS data and NOAA HATS data for NOAA flask data) resulting in a change in significant figures of the reported value. There was a change in condition (i.e., STD and LC) for some records from NOAA HATS, NOAA flask, NOAA data from Stephen Montzka, and pesticides data from CARB. There was no further action for the records that had a noted updated reported value, negligible differences, previously unaccounted for collocation, or a change in data source. Records that were shifted an hour and had a change in condition were flagged for further investigation. Among the 230,673 records with different reported values originating from different annual averages, 227,321 records (98.5%) required no further action and 3,352 records (1.5%) were flagged for further investigation. See Figure 1 for record counts.



**Figure 1. Counts of individual records that had a change in reported values originating from different annual averages in the 2023 Archive.**

### Different annual averages – different final values

There was a total of 4,422,870 records originating from different annual averages that have the same reported values but different final values. Unlike the different reported values described above, these records were first categorized by metrics quantifying said differences. Seven metrics were defined. Several records had a 19.1536% difference when comparing the concentrations in STD. Other final values had small differences (i.e., < 5% difference) or negligible differences likely attributable to digit precision. Other records had a change in condition (i.e., STD and LC). Over 67,000 records either changed from zero/NA or changed to zero/NA. Lastly, a handful of records had a 19.088% difference or a change in data source. All records were explored to better characterize these metrics. Most records previously had erroneous unit conversions that were corrected. The records with a change in data sources had a corresponding change in units. The remaining records were a mix of parameter codes, locations, and years with no discernable patterns. The records that there previously zero/NA and the records with a change in condition were flagged for further investigation. These accounted for 141,859 records (3.2%). The remaining 4,281,011 records (96.8%) required no further action. See Figure 2 for records counts.



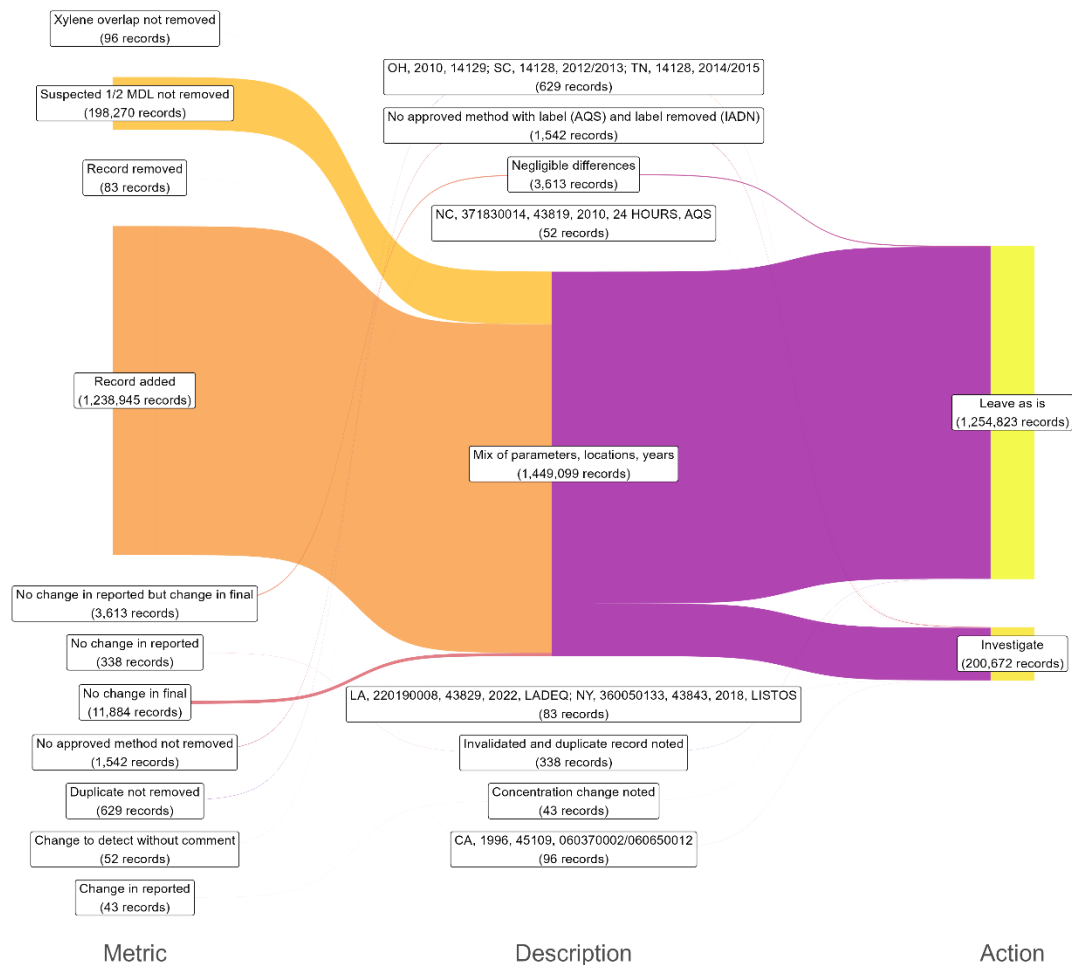
**Figure 2. Counts of individual records that had a change in final values originating from different annual averages in the 2023 Archive.**

### Added annual averages

Unlike the different annual averages, all records were considered in the added and removed annual averages, including the final values that were identical in the 2022 Archive. This is because the identical records contribute to the understanding as to why an annual average was added or removed. For example, records from a given pollutant/site/year/sampling duration may have been just below the completeness criterion for calculating an annual average in the 2022 Archive but met the criterion in the 2023 Archive. Knowing which records are identical gives additional context as to why an annual average may have been added.

All 1,455,495 records from the 4,184 added annual averages were investigated. Records were initially categorized into eleven metrics. Most records (over 1.2 million) were new to the 2023 Archive. Nearly 200,000 records were previously suspected to be non-detects. The remaining records were either identical to the 2022 Archive, had identical reported values but differences in the final, had no approved method but were not removed, were duplicate values but were not removed, had no change in the reported value, had overlapping xylene records but were not removed, were removed records, had a change in the reported value, or had a change from non-detect to detect without comment. These metrics were further analyzed by location, parameter code, and year into nine categories. Most records (over 1.4 million) were a mix of parameter codes, locations, and years with no discernable patterns. The remaining records mostly came from specific sites, parameter codes, and data sources. A total of 200,672 records (13.8%) were flagged for

further investigation while 1,254,823 records (86.2%) required no further action. The records labeled as not having a change in the final value, records added, negligible differences, no changes in reported values, and changes in reported required no further action. The remaining metrics were flagged as needing additional investigation, including those records with an overlap in xylene records that were not removed, duplicate value that were not removed, records with no approved methods that were not removed, records that changed from non-detect to detect without comment, records suspected of being non-detects, and records that were removed. See Figure 3 for records counts.

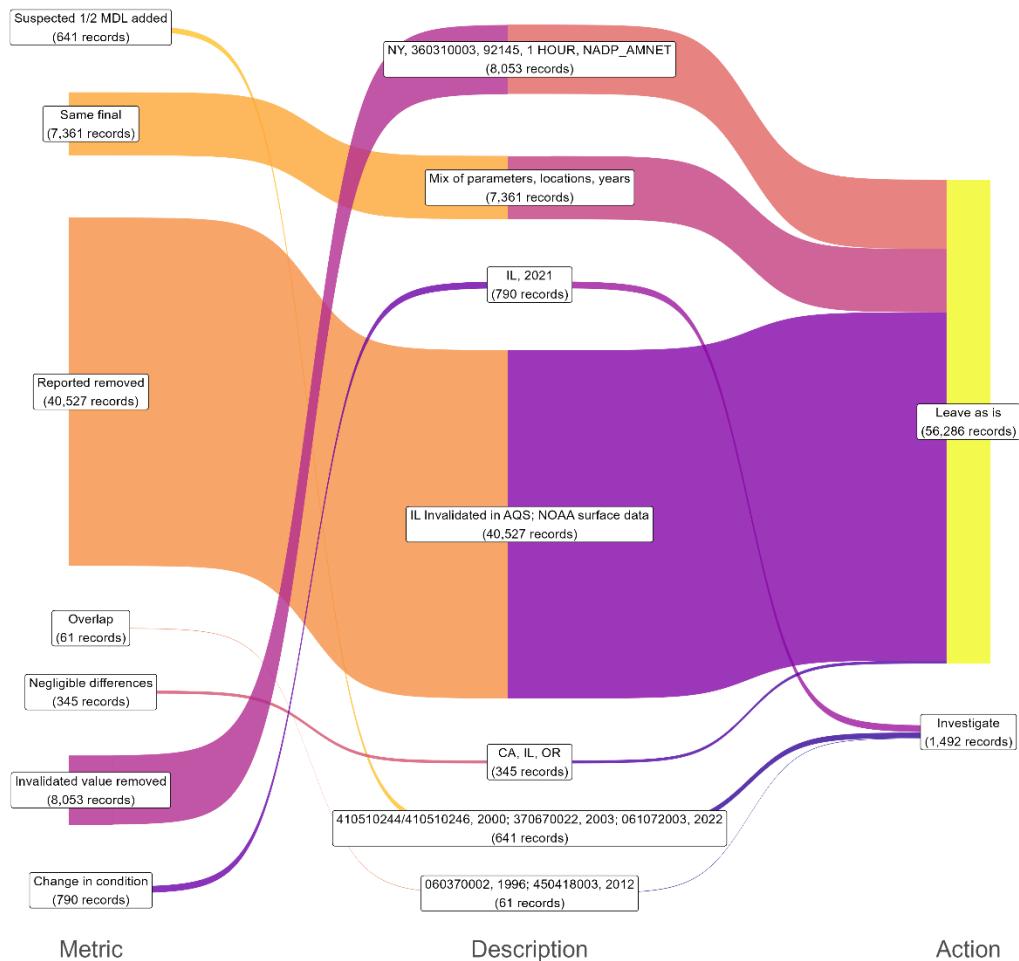


**Figure 3. Counts of individual records originating from added annual averages in the 2023 Archive.**

### Removed annual averages

All 57,778 records from the 428 removed annual averages were investigated. There was a total of seven metrics explored for the individual records. Most records were removed from the 2023 Archive or had the same final value as the 2022 Archive. Other records were previously suspected of being non-detect but were not properly flagged, overlapped, had negligible differences, were invalidated and removed, or had a change in condition (i.e., STD and LC). These metrics were further analyzed by location, parameter code, sampling duration, and data source into seven categories. The removed records originated from NOAA surface data that previously did not properly account for collocation (described in the Different annual averages – different reported values section above) and from Illinois data that were invalidated in AQS. One site in New York invalidated data from one site from the NADP data source. There was a mix of parameters, locations, and years among the records with the same final value with no discernable pattern.

Records from Illinois in 2021 had a change in condition (i.e., STD and LC). A handful of sites had records added that were suspected of being non-detect, had negligible differences, or overlapped. A total of 1,492 records (2.6%) were flagged for further investigation while the remaining 56,286 records (97.4%) required no further action. See Figure 4 for record counts. All numbers in Figures 1-4 are presented in the Appendix.



**Figure 4. Counts of individual records originating from removed annual averages in the 2023 Archive.**

### QA of the newest year

There were 7,973 annual averages from the newest year of data (i.e., 2023) in the 2023 Archive. The QA suspect categories in the 2023 Archive explored whether or not the newest year of data (1) had a large difference compared to the previous year (i.e., 2022), (2) had a large difference within a time series for a given pollutant/site/sampling duration (if the previous year did not exist), or (3) was greater than the 95<sup>th</sup> percentile for a given pollutant/year/sampling duration if the a time series or previous year did not exist.

Unlike in the overlap years, the suspect category thresholds are determined using records only from the 2023 Archive (i.e., the 2022 Archive is not used as a comparison). The count of annual averages falling into each suspect category is presented in Table 3. Of the 7,973 annual averages, 45 (0.6%) fell into suspect categories.

**Table 3. Suspect category descriptions and counts for the newest year of data in the 2023 Archive.**

Suspect Category	Description	Count
Overall	Annual averages added in the 2023 Archive from newest year	7,973
Previous year exists	2022 exists and 2023 annual average >10x 2022 annual average	16
Time series exists	2022 does not exist but time series exists in overlap years for a given pollutant/site/year/sampling duration and 2023 annual average >10x time series max in overlap years	2
Other	2022 and time series in overlap years do not exist and 2023 annual average >95 <sup>th</sup> %ile for pollutant/year	27
Total suspect	Flagged suspect annual averages for the newest year	45

## Frequent Questions

### What is the purpose of the Archive? Why does the Archive include data not in AQS?

The Archive brings together multiple data sources of ambient monitoring HAPs data across the US in a unified format that allows for user-ready data analysis. The Archive pulls in AQS data and other high quality data not housed in AQS. Due to data reporting requirements of HAPs, some state air agencies collect data that are appropriate for AQS but are not uploaded to AQS. Other federal agencies and established national programs outside of a given EPA regulatory network also collect relevant data. See Technical Report for additional information.

### Why are local conditions derived?

Sample values are either collected in LC or STD. Data collected in STD can be converted to LC from collocated or nearby temperature and pressure data or they can be derived from calculated LC/STD ratios. Converting or deriving to LC allows the data to be utilized for air toxics applications where LC is required. The LC/STD ratios derive LC where the final is in STD due to a lack of collocated or nearby temperature and pressure data.

### Why is there different criterion to calculate daily averages for remote and non-remote data?

Non-remote sub-daily data, depending on the sampling duration, require a minimum number of samples to calculate a daily average. This ensures that a daily average is representative. However, remote data do not have this requirement. For example, a daily average for remote data may be represented by one 5-minute sample. This criterion is loosened because the pollutants and monitoring locations of remote data typically represent background concentrations that do not meaningfully fluctuate at sub-daily time scales.

### Why are annual statistics separated by sampling duration?

Annual statistics are separated by pollutant/site/year/sampling duration to account for broad method differences among different sampling durations.

### How is the minimum number of samples of the different sampling durations determined when calculating a daily average?

A minimum number of samples is needed to calculate a daily average when the sampling duration is sub-daily. The minimum number of samples varies depending on the scale of the sampling duration (i.e., hourly or sub-hourly). These are found under the “Durations” tab in AMA2023\_lookups.xlsx. If the duration description is under 60 minutes, the minimum count is defined as  $\left\lceil 0.75 \times \frac{60}{DUR\_DESC} \right\rceil$ , where *DUR\_DESC* is the sub-hourly duration description in minutes. If

the duration description is 60 minutes or above, the minimum count is defined as  $\left\lceil 0.75 \times \frac{24}{DUR\_DESC} \right\rceil$ , where the *DUR\_DESC* is the hourly duration description in hours. In other words, the minimum count is rounding up the minimum number of samples needed to have 75% completeness for an hour (or day).

**Table 4. Minimum sampling count needed to construct an hourly or daily average presented by sampling duration.**

Duration Description	Average Up To	Minimum Count
5 MINUTES	HOURLY	9
10 MINUTES	HOURLY	5
15 MINUTES	HOURLY	3
30 MINUTES	HOURLY	2
150 MINUTES	DAILY	8
90 MINUTES	DAILY	12
1 HOUR	DAILY	18
2 HOUR	DAILY	9
3 HOURS	DAILY	6
4 HOUR	DAILY	5
5 HOUR	DAILY	4
6 HOUR	DAILY	3
8 HOUR	DAILY	3
12 HOUR	DAILY	2
24 HOURS	DAILY	1

Where is the information in the lookup table tabs obtained?

- AQS\_CAS: <https://aqs.epa.gov/aqsweb/documents/codetables/parameters.csv> (accessed September 2025)
- NATTS: <https://www.epa.gov/system/files/documents/2023-03/NATTS-site-list-2023.pdf> (accessed September 2025)
- NEI: PollutantCode.xlsx from [https://www.epa.gov/system/files/other-files/2022-05/EISCodeList\\_1.zip](https://www.epa.gov/system/files/other-files/2022-05/EISCodeList_1.zip) (accessed September 2025)
- AirToxScreen: 2020AirToxScreenPollutants.xlsx from <https://www.epa.gov/system/files/other-files/2024-06/2020-airtoxscreen-supplemental-data-files.zip> (accessed September 2025)

What is the minimum number of daily averages required to create a valid quarter?

A minimum of *six* daily averages is required per pollutant/site/sampling duration to create a valid quarter.

How many valid quarters of data are needed to calculate an annual average?

A minimum of *three* valid quarters in a year is needed per pollutant/site/sampling duration to create an annual average. This is to ensure that the annual average is representative of the year. All daily averages are used in the calculation of an annual average, assuming the threshold of valid quarters is met. Therefore, an annual average may include data from three “valid” quarters and one “invalid” quarter.

What data are excluded from annual averages in the 2023 Archive?

Only ambient HAPs monitoring with a sampling duration of less than or equal to 24 HOURS are considered for annual averages. Therefore, records with a 2-week sampling duration are excluded from the annual average calculation. This

sampling duration may be included in future versions of the Archive after conducting a more thorough analysis. Fenceline monitoring data are currently excluded from annual averages.

## Common Acronyms

AMA	Ambient Monitoring Archive
AQS	Air Quality System
CARB	California Air Resource Board
EPA	Environmental Protection Agency
EtO	Ethylene oxide
HAP	Hazardous Air Pollutant
LC	Location Conditions
MDL	Method Detection Limit
MIT	Massachusetts Institute of Technology
MNPCA	Minnesota Pollution Control Agency
NA	Not Available (missing value)
NADP	National Atmospheric Deposition Program
NOAA	National Oceanic and Atmospheric Administration
NOAA HATS	NOAA Halocarbons and other Atmospheric Trace Species
ODEQ	Oregon Department of Environmental Quality
POC	Parameter Occurrence Code
QA	Quality Assurance
ROS	Regression on Order Statistics
SCAQMD	South Coast Air Quality Management District
STD	STanDard conditions
TCEQ	Texas Commission on Environmental Quality
US	United States

## References

- Helsel, D.R., 2005. *Nondetects and data analysis. Statistics for censored environmental data* (pp. xv+-250).
- McCarthy, M.C., Hafner, H.R., Chinkin, L.R. and Charrier, J.G., 2007. Temporal variability of selected air toxics in the United States. *Atmospheric Environment*, 41(34), pp.7180-7194.

## Appendix

**Table 5. Sankey figure counts from Sources/Metrics presented in Figures 1-4.**

Overlap Years Category	Source/Metric	Count
Different - Different Reported Value	AQS	3,950
Different - Different Reported Value	CARB Pesticides	1057
Different - Different Reported Value	NOAA Data from Montzka	562
Different - Different Reported Value	NOAA Flask Data	2,479
Different - Different Reported Value	NOAA HATS Data	175,539
Different - Different Reported Value	NOAA Surface Data	46,531
Different - Different Reported Value	SCAQMD 1180 Rule	362
Different - Different Reported Value	SCAQMD EtO Data	193
Different - Different Final Value	Change in condition	74,368

Different - Different Final Value	Change in data source	71
Different - Different Final Value	Negligible differences	192,170
Different - Different Final Value	Percent Difference < 5	7,676
Different - Different Final Value	Percent difference: 19.088	1,997
Different - Different Final Value	Standard percent difference: 19.1536	4,079,097
Different - Different Final Value	Standard value is/was zero/NA	67,491
Added	Change in reported	43
Added	Change to detect without comment	52
Added	Duplicate not removed	629
Added	No approved method not removed	1,542
Added	No change in final	11,884
Added	No change in reported	338
Added	No change in reported but change in final	3,613
Added	Record added	1,238,945
Added	Record removed	83
Added	Suspected 1/2 MDL not removed	198,270
Added	Xylene overlap not removed	96
Removed	Change in condition	790
Removed	Invalidated value removed	8,053
Removed	Negligible differences	345
Removed	Overlap	61
Removed	Reported removed	40,527
Removed	Same final	7,361
Removed	Suspected 1/2 MDL added	641

**Table 6. Sankey figure counts from Descriptions presented in Figures 1-4.**

<b>Overlap Years Category</b>	<b>Description</b>	<b>Count</b>
Different Reported Value	Change in condition	2,993
Different Reported Value	Change in data source	4,190
Different Reported Value	Confirmed collocation	46,531
Different Reported Value	Negligible differences	536
Different Reported Value	Reported shifted an hour	359
Different Reported Value	Updated reported	176,064
Different Final Value	Confirmed corrected unit conversion	4,081,094
Different Final Value	Mix of parameters, locations, years	341,705
Different Final Value	Reported change in units	71
Added	CA, 1996, 45109, 060370002/060650012	96
Added	Concentration change noted	43
Added	Invalidated and duplicate record noted	338
Added	LA, 220190008, 43829, 2022, LADEQ; NY, 360050133, 43843, 2018, LISTOS	83
Added	Mix of parameters, locations, years	1,449,099
Added	NC, 371830014, 43819, 2010, 24 HOURS, AQS	52
Added	Negligible differences	3,613
Added	No approved method with label (AQS) and label removed (IADN)	1,542

Added	OH, 2010, 14129; SC, 14128, 2012/2013; TN, 14128, 2014/2015	629
Removed	060370002, 1996; 450418003, 2012	61
Removed	410510244/410510246, 2000; 370670022, 2003; 061072003, 2022	641
Removed	CA, IL, OR	345
Removed	IL Invalidated in AQS; NOAA surface data	40,527
Removed	IL, 2021	790
Removed	Mix of parameters, locations, years	7,361
Removed	NY, 360310003, 92145, 1 HOUR, NADP_AMNET	8,053

**Table 7. Sankey figure counts from Actions presented in Figures 1-4.**

<b>Overlap Years Category</b>	<b>Investigate</b>	<b>Leave as is</b>
Different Reported Value	3,352	227,321
Different Final Value	141,859	4,281,011
Added	200,672	1,254,823
Removed	1,492	56,286