



## Air Pollution Control Division

## Technical Services Program

### APPENDIX D3

### **Standard Operating Procedure for the Data Management Operations of Filter Based Particulate Data**

**Data Management**

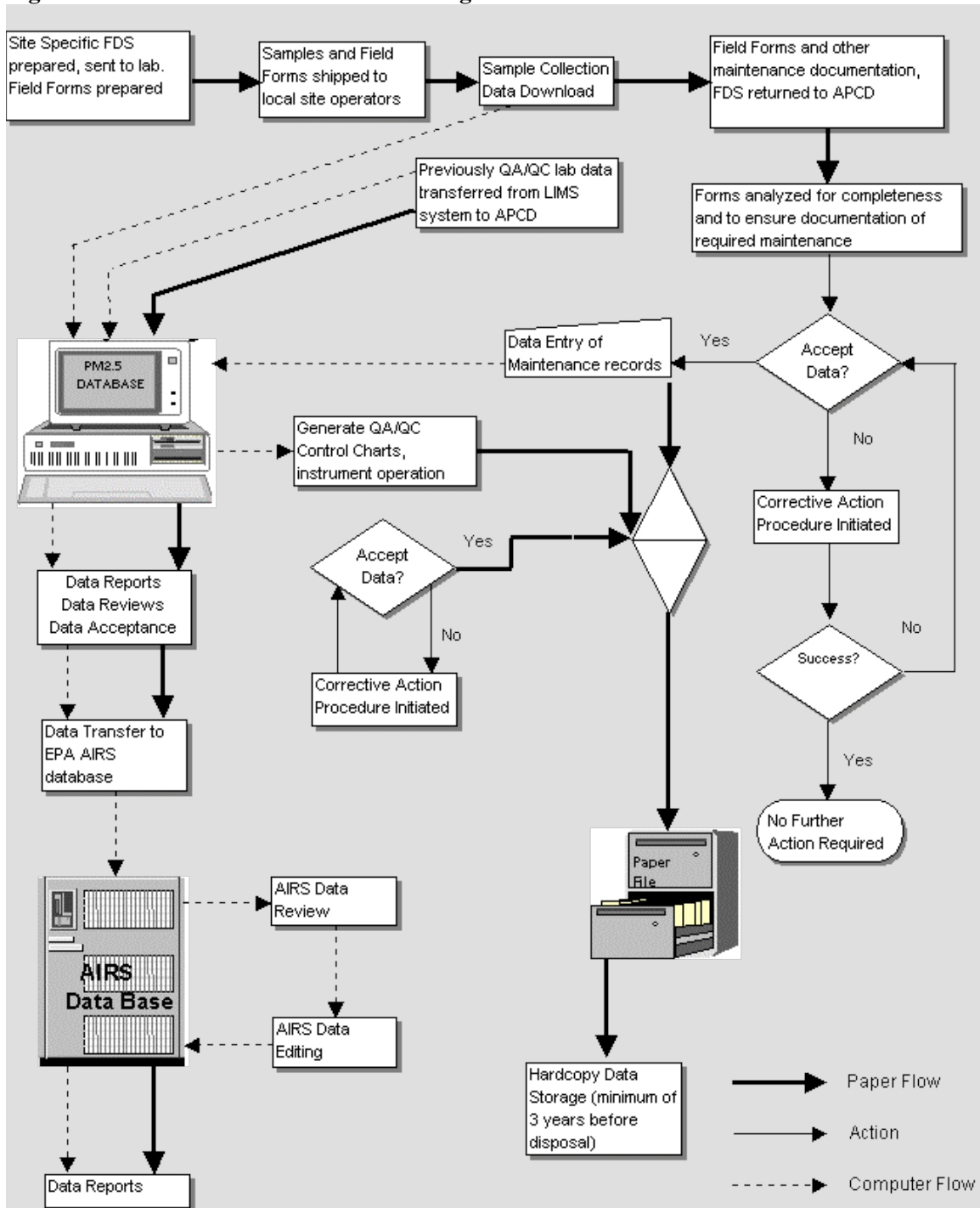
This section describes the data management operations pertaining to PM samples derived from low volume or high volume samplers for the SLAMS/NAMS and SPM stations operated by APCD. This includes an overview of the automated operations and analyses performed on raw (“as-collected”) PM data. These operations include data recording, validation, transmittal, transformation, reduction, analysis, management, storage, and retrieval.

Data processing steps for PM data are summarized in Figure 1. Originally, all PM data were processed in large batches using SPSS 9.0 software. Mid-year, 2000 the APCD acquired a database written by The Mid-Atlantic Regional Air Management Association (MARAMA). As written, this database tool has the ability to:

- Accept electronic files from both Andersen and R&P monitors as inputs;
- Accept filter data from laboratories in the form of an Excel spreadsheet;
- Review, edit, quality check and report their data; and
- Produce transaction files that can be sent to AQS.

The tool is an implementation of the Microsoft Access database. The data management system (DMS) allows the APCD to obtain reports on status of samples, location of specific filters, etc. The database has been modified substantially to accept lab files, RP2000 raw data, high volume PM<sub>10</sub> and TSP data, various QC data, QA Audits, and maintenance records, in addition to custom reports, forms, and queries. The following figure summarizes the data flow through the entire PM network.

**Figure 1 APCD Particulate Data Processing**



### **Low Volume Data:**

The Air Pollution Control Division's (APCD) low volume filter based PM network is entirely comprised of Rupprecht & Patashnick (RP) PM FRM or FEM (equipped with the VSCC) designated monitors. All the RP FRM instruments are equipped with internal data loggers. These data loggers provide data collection for sample and ambient data. For specifics regarding the RP instruments and their data storage, format, capability, etc. refer to the RP 2000 Partisol FRM and the RP 2025A Partisol Plus operating manuals. The data is retrieved from the instruments in the field by local site operators using Palm™ M105 PDAs or by the APCD as part of its regular maintenance routines. All the PDAs use the Palm operating system and RPCOPOS software. The software was provided by Rupprecht & Patashnick, Inc. as freeware upon purchase of a PM FRM sampler. Files downloaded from a sampler are named according to a convention that allows APCD personnel to identify what type of data was downloaded (see the RP2000FRM and RP2025 operating manuals for a complete description of data types), from which site the data originates, and when the data was downloaded. The raw data is a comma-delimited text (\*.csv) file.

Local personnel send the raw data to the APCD. In the cases where palmtops are used, the operators must transfer the data to their PCs using the Palm communications software provided by Palm. The files are stored electronically on the operator's PC in a specified directory (usually PM25\_data) on their local hard drive. The local operators compress all the data files into one file and send the compressed file to the APCD as an email attachment. The raw data files are combined and compressed using WinZip, PkZip, or Free Zip.

When the APCD receives the raw sampler data the data file (\*.zip or \*.csv) is copied to the network drive designated for storage of low volume PM data. If the data is compressed the entire zip file is saved to a secure directory on the APCD network server. Each of the two instrument types (RP2025 and RP2000) produces text files of slightly different format, requiring different processing. Each file is copied to the appropriate 'DUMP' directory on the network server. If the data originated from a RP2025 the data is copied to the secure 'Sequential' directory; if the data originated from a RP2000 the data is copied to the secure 'Single' directory. Each file is examined for format (transmission errors), type (mis-labeled Filter, Interval, or Input files), and completeness (ensuring continuity with prior transmission). After this initial examination the files reside in the appropriate directory, unchanged, until the next processing step.

A Field Data Sheet/Chain of Custody document (FDS) is created by APCD personnel and sent to the analytical lab prior to batch conditioning and tare weighing. The FDS is shipped to each site operator along with the unsampled filters. Post sampling, the FDS accompanies the filters back to the analytical lab where the filters are checked in upon receipt. Each filter has a unique record that was created when the filter was tare weighed and that resides on the analytical laboratory's DMS. The lab technicians may make appropriate comments on the FDS before sending the FDS to the APCD for review. Additionally, operator comments are entered into the DMS when the filters are checked-in at the analytical lab. When the FDSs are sent to the APCD from the analytical lab the transfer is followed by an electronic file containing the lab data (including the operator comments) for those filters included in that batch; the file is an export from the DMS. This exported file is a Microsoft Excel spreadsheet generated by exporting previously quality-checked data generated by the lab. The spreadsheet contains information on each filter in a batch including Filter ID, QA checks (including check weights, duplicate weights, and temperature and relative humidity controls), operator and lab comments, site id, and tare and gross weights. These

files are archived on the secure drive on the network server and immediately imported into the PM DMS.

When the FDSs are received at the APCD each sheet is checked to ensure that the filters in each batch from each site were stored and shipped cold (e.g. at or below 4°C), and that the data indicates all the sampling critical criteria were met. Any deviations from these critical criteria are noted on the FDS. The FDS accompanies a batch of filters from the analytical lab to the field. It allows the local operators to document site visits, filter sampler data, ‘condition’ anomalies (i.e. unusual weather conditions, nearby construction, sampler anomalies, etc.). Additionally, if there is an error in any of the electronic data (e.g. the operator generates an electronic record with an incorrect filter ID by programming the sampler incorrectly) the correct information can be noted on the FDS. The FDS allows the operator to make any corrections required and note anomalous conditions as well as serving as a data backup to the electronic record.

Upon receipt of the data from the analytical lab along with the Field Data Sheets the sampler records can now be imported into the DMS. Every sample data file generated is imported into the DMS individually. The records are automatically processed into a screening file. While there, they are manually checked for data completeness and verified against the information provided on the FDS. Spurious records are deleted (e.g. when the RP2025 is placed in ‘STOP’ mode by the operator the instrument creates a false record that must be deleted) and any necessary changes to filter IDs are made (note: changes are made only if operator comments on the FDS indicate that a specific filter sampled under an incorrect Filter Id). The records are checked for continuity with preceding records for the site and any missing or incomplete records are recreated from information provided on the FDS (e.g. the RP2000FRM doesn’t create a record when a Field Blank is taken, the record must be created by hand).

After these checks are made the records are transferred from the screening table into an intermediate storage table. Once there, null codes are entered, where necessary, summary reports are generated for review. Other summary reports such as anomalous or missing data as well as field blank reports are available. Management reviews these reports for completeness and accuracy. After these reviews, necessary changes are made, the data is flagged as having been QA checked, and flagged for AQS submittal.

### **High Volume Data:**

The APCD’s high volume network is predominantly comprised of GMW 1200 samplers. All local operators in APCD’s high volume network are shipped a supply of tare weighed high volume filters in Particulate Sample Field Envelopes, which are used as chain of custody forms and to record field data. Operator’s collect a high volume sample as described in the HIGH VOLUME SOP and return the filters to APCD. APCD enters the field data associated with each sample from the PSFE into the DMS and forwards the filters to the lab. The lab reenters the field data into a separate database and conducts the remaining portion of the gravimetric analysis according to the HIGH VOLUME LAB SOP. Lab technicians then compare the field data in their system to the data independently entered by APCD and makes changes to their data or advises APCD to alter their data. All discrepancies are verified by the lab with the information on the PSFE. Gravimetric data and verified field data are then available for APCD to pull from the laboratory’s database and import into the DMS. The following table summarizes the validation checks performed on the PM data.

**Table 1: Validation Check Summaries**

Type of Data Check	Electronic Transmission and Storage	Manual Checks	Automated Checks
Data Parity and Transmission Protocol Checks	X		
Manual Key Entry		X	
Date and Time Consistency		X	X
Completeness of Required Fields		X	X
Range Checking		X	X
Statistical Outlier Checking			X
Manual Inspection of Charts and Reports		X	

Once the data have been reviewed an AQS transaction file is created and the file is transferred to the APCD Ambient Data Specialist for upload to AQS.

### Data Validation

Data validation is a combination of checking that data processing operations have been carried out correctly and of monitoring the quality of the field and laboratory operations. Data validation can identify problems in any of these areas. Once problems are identified, the data can be corrected or invalidated, and corrective actions can be taken for field or laboratory operations. Numerical data stored in the PM DMS are never overwritten by flags or null codes. Flags or null codes denoting error conditions or QA status are saved as separate fields in the database. It is always possible to revert to the original data.

The following validation functions are incorporated into the PM DMS to ensure quality of data entry and data processing operations:

**Key Entry Error Checks** – When data are manually entered into the database either at the lab or the APCD, the entry operator compares the values with what is written on the entry forms. Changes are made to correct entry errors.

**Completeness Checks** - When the data are processed certain data completeness criteria must be met. For example, each filter must have a start time, an end time, an average flow rate, dates weighed, and operator and technician names. The data entry operator will be notified if an incomplete record has been entered before the record can be closed. Where possible the data from an incomplete record is filled in using information already available (e.g. the start time from the next sample from the same instrument can be used as the stop time if unavailable).

**Internal Consistency and Other Reasonableness Checks** - Several other internal consistency checks are built into the PM DMS. For example, the end time of a filter must be greater than the start time. Computed filter volume (integrated flow) must be equal to the exposure time multiplied by the average flow rate. Net weights must be positive and should they exceed a specified value the entries are re-verified with the lab logbook. Ambient temperature and pressure

values must be within a reasonable range, depending on location and time of year.

**Data Retention** - Raw data sheets, including copies of the FDS-COC forms, PSFEs, and 4-Week Verification forms, are retained on file in the APCD office for a minimum of five years, and are readily available for audits and data verification activities. Physical samples such as filters are archived for a minimum of three years and shall be discarded with appropriate attention to proper disposal of potentially hazardous materials.

**Statistical Data Checks** - Errors found during statistical screening will be traced back to original data entry files and to the raw data sheets, if necessary. These checks shall be run prior to any data submission to AQS. Data validation is the process by which raw data are screened and assessed before it can be included in the main database (i.e., the PM DMS).

Table 1 summarizes the validation checks applicable to the PM data.

### Data Transformation

Calculations the DMS computes while running reports and queries for transforming raw data from measured units to final concentrations are relatively straightforward, and many are carried out in the sampler data processing unit before being recorded. The following relations in Table 2 pertain to PM monitoring:

**Table 2: Raw Data Calculations**

Parameter	Units	Type of Conversion	Equation
Filter Volume (V <sub>a</sub> ) *	m <sup>3</sup>	Calculated from average Flow Rate (Q <sub>ave</sub> ) in L/min, and total elapsed time (t) in min. multiplied by the unit conversion (m <sup>3</sup> /L)	$V_a = Q_{ave} \times t \div 10^3$
Mass on Filter (M)	µg	Calculated from filter post-weight (M <sub>f</sub> ) in µg and filter pre-weight (M <sub>i</sub> ) in µg, multiplied by the unit conversion (µg/mg)	$M = M_f - M_i$
PM Concentration (MC)	µg/m <sup>3</sup>	Calculated from laboratory data and sampler volume	$MC = \frac{M}{V_a}$

\* - R&P PM FRM instruments provide this value from the internal data logger, the DMS generates flow rates and sample volumes for high volume samples when the data is imported and stores those values permanently.

### Data Transmittal

Data transmittal occurs when data are transferred from one person or location to another or when data are copied from one form to another. Some examples of data transmittal are copying raw data from a data entry form for keying into a computer file and electronic transfer of data over a telephone or computer network. Table 3 summarizes data transfer operations.

**Table 3: Data Transfer Operations**

<b>Description of Data Transfer</b>	<b>Originator</b>	<b>Recipient</b>	<b>QA Measures Applied</b>
Keying Weighing Data into the lab’s DMS system	Laboratory Technician (hand-written data form)	Data Processing Personnel	Double Key Entry
Electronic data transfer	Between computers or over network	--	Parity Checking; transmission protocols
Filter Receiving and Chain-of-Custody	Shipping and Receiving Clerk	The lab’s DMS system Computer (accessioning /receiving clerk enters data at a local terminal)	Filter numbers are verified automatically; inspections indicate missing filters and/or incorrect data entries
Calibration, FRM, and Audit Data	Auditor or field supervisor	PM DMS	Entries are checked by QA Officer
AQS data summaries	PM Data Manager	AQS (U.S. EPA)	Automatic statistical checks in AQS

The APCD will report all PM ambient air quality data and information specified by the AQS Users Guide (Volume II, Air Quality Data Coding, and Volume III, Air Quality Data Storage), coded in the AQS format. Such air quality data and information will be fully screened and validated and will be submitted directly to the AQS via electronic transmission, in the format of AQS, and in accordance with the quarterly schedule. The specific quarterly reporting periods and due dates are shown in Table 4.

**Table 4: Data Reporting Schedule**

<b>Reporting Period</b>	<b>Due Date</b>
January 1-March 31	June 30
April 1-June 30	September 30
July 1-September 30	December 31
October 1-December 31	March 31

**Data Analysis**

The APCD is currently implementing the data summary and analysis requirements contained in 40CFR Part 58, Appendix A. It is anticipated that as the PM Program develops, additional data analysis procedures will be developed. Two key operational criteria for PM sampling are bias and precision. As defined in 40CFR Part 58, Appendix A, these are based on differences between collocated sampler results and FRM performance evaluations. The APCD will inspect the results of collocated sampling during each batch of validation activity. This data will be evaluated as early in the process as possible, so that potential operational problems can be addressed. The objective of the APCD will be to optimize the performance of its PM monitoring equipment. Multiple collocation results must be accumulated to assess data quality with confidence.



However, even limited data can be used for system maintenance and corrective action. The following specific summary statistics will be tracked for the APCD's PM network:

- Single sampler bias or accuracy (based on collocated FRM data, flow rate performance audits, and FRM performance evaluations)
- Single sampler precision (based on collocated data)
- Network-wide bias and precision (based on collocated FRM data, flow rate performance audits, and FRM performance evaluations)
- Data completeness

Equations used for these reports are given in Table 5.

**Table 5: Report Equations**

Criterion	Equation	Reference
Accuracy of Single Sampler Flow - Single Check ( $d_i$ ) $X_i$ is reference flow; $Y_i$ is measured flow	$d_i = \frac{Y_i - X_i}{X_i} \times 100$	40 CFR 58 Appendix A, Section 5.5.1.1
Bias of a Single Sampler - Annual Basis ( $D_j$ )- average of individual percent differences between sampler and reference value; $n_j$ is the number of measurements over the period	$D_j = \frac{1}{n_j} \times \sum_{i=1}^{n_j} d_i$	40 CFR 58 Appendix A, Section 5.5.1.2
Percent Difference for a Single Check ( $d_i$ ) - $X_i$ and $Y_i$ are concentrations from the primary and duplicate samplers, respectively.	$d_i = \frac{Y_i - X_i}{(Y_i + X_i) / 2} \times 100$ $CV_i = \frac{ d_i }{\sqrt{2}}$	40 CFR 58 Appendix A, Section 5.5.2.1
Pooled Coefficient of Variation, Quarterly Basis ( $CV_{j,q}$ ). The $CV_i$ will only be used when the two measurements are both greater than 6 $\mu\text{g}/\text{m}^3$ .	$CV_{j,q} = \sqrt{\frac{\sum_{i=1}^{n_j} CV_i^2}{n_{j,q}}}$	40 CFR 58 Appendix A, Section 5.5.2.3 (a)
Completeness	$\text{Completeness} = \frac{N_{\text{valid}}}{N_{\text{theoretical}}} * 100\%$	--

### Validation and Verification Methods

Many of the processes for verifying and validating the measurement phases of the PM data collection operation have been discussed above. If these processes, as written in the QAPP, are

followed, and the sites are representative of the conditions for which they were selected, one would expect to achieve the PM DQOs. However, exceptional field events may occur, instrument malfunction and field and laboratory activities may negatively affect the integrity of samples. In addition, it is expected that some of the QC checks will fail to meet the acceptance criteria. The review of this routine data and their associated QC data will be verified and validated on a sample batch basis. The sample batch is the most efficient entity for verification/validation activities. It is assumed that if measurement uncertainty can be controlled within acceptance criteria, at a batch level, then the overall measurement uncertainty will be maintained within the precision and bias DQOs.

### **Process for Validating and Verifying Data Verification of Sample Batches**

The APCD ships and receives low volume filters on an alternating two-week schedule and high volume filters on a monthly schedule. After a low volume sample batch is completed, a thorough review of the data will be conducted for completeness and data entry accuracy. All raw data that is hand entered into the DMS from data sheets will be proofed as discussed previously. Raw sampler data received electronically at the APCD is checked for form and completeness. The data values are compared to reduce the possibility of entry and transcription errors. Once the data is entered into the DMS, a report is generated that identifies outliers, and inconsistent data (like data checks), and data outside of acceptance criteria. These data are tagged and reverified to ensure completeness and accuracy. When APCD receives a high volume filter from an operator the data from the PSFE is entered into the DMS as soon as practicable. Those filters are then forwarded to the contract lab. As this occurs on an ad hoc basis there is no ‘batch’ equivalent in the high volume system. Consequently data verification occurs when APCD receives notice from the contract lab that high volume data is finalized and available.

### **Validation**

Validation of measurement data will require two stages, one at the measurement value level, and the second at the batch level. Comments on all invalid samples will be noted in the DMS. Information will include a brief summary of why the sample was invalidated including the value out of range. This record will be available on the DMS since all samples that were pre-weighed or otherwise used in the field will be recorded. At least one AQS null code will be associated with an invalid sample, indicating which critical criteria were violated. Validity flags may be used to note special circumstances or otherwise to note suspect samples. No other flags will be associated with the record. If a record has a null code associated with it the system automatically produces the appropriate AQS transaction for AQS upload.

### **Validation of Measurement Values**

Certain criteria based upon CFR rules and/or APCD judgments have been developed that will be used to invalidate a sample or measurement. Field operator notes on the FDS or PSFE forms will be used to determine if specific samples, samples from a particular instrument, or samples from a shipping batch will be invalidated. In all cases the sample(s) will be returned to the laboratory for further examination and analysis. When the laboratory technician reviews the FDS forms he/she will look for operator notes. Filters that have notes indicating obvious contamination, filter damage, or field accidents will be immediately examined. Typically, these samples will be invalidated. A null record will be generated and substituted for the sample record for AQS upload.

Violations of the critical criteria may be used alone or in combination to invalidate samples. Since the possible combinations are overwhelming and cannot be anticipated, each sample will be reviewed to determine if individual samples or multiple samples from a site for a particular time period will be invalidated. The Department will keep a record of the combination of conditions that resulted in invalidating a sample or set of samples. As mentioned above, all data invalidation will be documented. The sample record will be maintained in the PM DMS. Table 6 and Table 7 contain criteria that can be used to invalidate samples based on a single violation of the critical criteria or a combination of violations.

**Table 6: Single Flag Invalidation Criteria for Single Samples**

Criteria	AQS Null Code	Comment
Contamination	AM or AR	Concurrence with lab technician and PM Coordinator
Filter Damage	AK	Concurrence with lab technician and PM Coordinator
Event	Various	Exceptional, known field event expected to have affected sample. Concurrence with lab technician and PM Coordinator
Laboratory Accident	AR	Concurrence with lab technician and PM Coordinator
Field Accident	AL	Concurrence with lab technician and PM Coordinator
Flow Rate Cutoff	AN	Termination of sample collection due to flow rate > 10% design flow rate for 60 seconds.

**Table 7: Single Sample Validation Template**

Requirement	Acceptance criteria	Major <sup>1</sup>	Minor <sup>2</sup>	Null Code
Flow Rate	$\leq \pm 5\%$ of 16.67 L/min. for < 5 min	>10%	>5%	AH
Flow Rate Verification	$\leq 4\%$ of transfer standard	> 6%	> 4%	AM
Filter Temp	> 5° C for < 30 min	> 10°C	> 5° C	AM
Elapsed Sample Time	1380 < Sample Time < 1500 Minutes	> 1530	>1500	AG
<b>Holding Times</b>				
Pre-sampling	$\leq 30$ days	>32 days	>30 days	AM
Sample Recovery	$\leq 96$ hours	>100 hours	>96 hours	AB
Post-sampling				
25°C	$\leq 10$ days	>12 days	>10 days	AM
4°C	$\leq 30$ days	>32 days	>30 days	AM

### Validation of Sample Batches

Due to the nature and holding times of the routine samples, it is critical that the Department minimize the amount of data that is invalidated. Based on the types of QC samples that are

included in the batch and the field and laboratory conditions that are reported along with the batch (field/lab flags), the Department has developed a validation template that will be used to determine when routine data will be invalidated and when major corrective actions need to be instituted. Table 8 represents the lab validation template.

**Table 8: Validation Template**

Requirement	Frequency	Acceptance Criteria	Major <sup>1</sup>	Minor <sup>2</sup>	Null Code <sup>3</sup>
<b>Blanks</b> Field Blanks Lab Blanks	10% 10%	$\leq \pm 30 \mu\text{g}$ $\leq \pm 15 \mu\text{g}$	All blanks $> \pm 30 \mu\text{g}$ All blanks $> \pm 15 \mu\text{g}$	One blank $> \pm 30 \mu\text{g}$ One blank $> \pm 15 \mu\text{g}$	
<b>Precision Checks</b> Collocated pairs Duplicate weight	3 10%	PD $\leq 10\%$ $\leq \pm 15 \mu\text{g}$	Both samples $> 15\%$ Duplicate $> \pm 20 \mu\text{g}$	One sample $> 10\%$ Duplicate $> \pm 15 \mu\text{g}$	
<b>Accuracy</b> Balance Checks	10%	$\leq \pm 3 \mu\text{g}$	10% checks $> \pm 3 \mu\text{g}$	20% checks $> \pm 3 \mu\text{g}$	
<b>Lab Conditions</b> Temperature Humidity	Each weighing Each weighing	Mean 20-23°C $\leq \pm 2^\circ\text{C}$ 30-40% $\leq \pm 5\%$	Mean $> 25^\circ$ or $< 18^\circ$ $> \pm 4^\circ$ Mean $> 45\%$ or $< 20\%$ $> \pm 7\%$	Mean 23-25° or 18-20° $> \pm 2^\circ < \pm 4^\circ$ Mean $> 45\%$ or $< 20\%$ $> \pm 5\% < \pm 7\%$	

1-if 2 majors occur data invalidated

2-if 4 minors occur data invalidated. 2 minors equal 1 major

3-AIRS AQS

Based upon the number of major and minor flags associated with the batch, the batch may be invalidated or flagged as suspect. The DMS system will evaluate the batch and generate a report based upon the results described in the validation template. If the report describes invalidating the batch of data, the batch will be reanalyzed. Prior to reanalysis, all efforts will be made to take corrective actions, depending on the type of QC checks that were outside of acceptance criteria, to correct the problem. If the batch remains outside the criteria, the routine samples will be invalidated and the appropriate null code substituted for the value in AQS.

### Data Validation, Manual Methods

To fully complete the validation process all flagged and unflagged data must be examined for validation criteria based on information sources available within the APCD. The validation tables in Section 22.0 of this document should be consulted for specific validation criteria. Some validation criteria that are not easily automated are listed below:

- *Manual Data Inspection*—The purpose of manual data inspection is to spot unusually high (or low) values that might indicate a gross error in the data collection system. It is often helpful to plot data in a time series.
- *Systems Audit Report Results*—Audits occasionally turn up serious deficiencies that could affect the validity of the data. For example, if it were found that a field operator is mounting sample modules incorrectly, it would be necessary to flag or invalidate all the data corresponding to that sample module when that operator was working.

- *Performance Audit Results*—Large audit bias found in critical measurements such as sample volume accuracy could result in data invalidation or flagging back to the last previous acceptable audit or control check result.
- *Collocated Sample Results*—Performance on collocated duplicates should be evaluated if the appropriate information is available. Poor performance audit results should be investigated further, and if a specific, identifiable problem is uncovered that affects reportable data, that data should be flagged or invalidated. In general, reportable data should not be invalidated unless a specific, identifiable cause for the discrepancy can be found.
- *Operator's Notes and Site-Specific Information*—Operator's notes can contain information that would call for data invalidation due to lack of sampling representativeness. Examples include meteorological events such as sand storms, temporary violations of site criteria such as nearby construction, or operational difficulties with the sampling equipment. The APCD should use its best judgment about the impact of site conditions on the acceptability of the data and may consult with EPA via the DOPO or the Regional Office if there are questions.

### Automatic Checks

Because the data will be delivered from the analytical laboratory to the APCD in electronic format, automated checking could be implemented. Some automated screening methods that might be considered include the following:

- *Time Series Analysis*—This analysis is typically the examination of a set of data for a single observable acquired over a period of time. Time series data are often best examined graphically, and it is often helpful to chart related variables together on the same graph.
- *Outlier Checks*—are another means of identifying potential problems. **An environmental observation should never be invalidated simply because it is identified as a possible outlier by statistical techniques.** Observed environmental concentration distributions tend to be somewhat skewed, so that a small number of concentrations significantly higher than the long-term average should be expected. Selecting the top 1 or 2 percent of values in a data set for investigation however is often a good rule of thumb for data assessment because high data values are sometimes the result of analytical, procedural, or calculation errors.

## Hi volume data import from LSD

Process to extract correct data set from LSD's MTL database (MTL), verify independent data entry, and upload records into APCD's particulate data management system (PMT)

Identify the correct data set

LSD does not report data to APCD, rather APCD pulls data from the MTL system. The Lab Manager notifies APCD that hi volume data is ready to be pulled from the MTL database and characterizes the data that's available to be pulled, usually by an indication that all filters with a gross weight date on or before 'mm/dd/yy' are available. To extract data from MTL follow these steps:

Open the MTL database, open the table, 'Filter'.

Filter all the available records so that only hi vol records are available (Note: every hi vol filter ID begins with a 'Q').

Filter the resulting dataset by selecting only those records that satisfy the criteria 'with a gross weight data on or before 'mm/dd/yy'.

Sort the resulting dataset by 'BatchID', records where 'BatchId' are blank are available to be pulled.

For these records apply a batchID, usually the date specified by the lab manager in the following format, 'yymmdd'.

Open a sql (design) view of the query, qryExpHiVol.

In the sql view edit the 'Where' statement of the query so that it specifies the batchID used in the previous step as the condition of the where statement (i.e. WHERE (((dbo\_Filter.batchID)=yymmdd)).

Save the query and execute it.

Copy the resulting dataset into a new Msexcel spreadsheet.

Close the query, close the MTL system.

Verifying independent data entry.

Rename that sheet 'exp' (for export). Sort the entire spreadsheet on the field, 'comment'.

Replace blanks comments with a value of 'nc' (no comment).

Sort the entire sheet on 'rundate' (note the first and last sample dates of the dataset)

Sort the entire sheet on 'AQS\_ID' and 'rundate'. Copy the entire spreadsheet onto a new sheet, rename that sheet, 'checks'.

Navigate to the PMT system, open the table, 'hivolcheckin'.

Filter all the available records by sample date according to the first and last sample dates of the export data, noted in a previous step (i.e. Filter – between).

Use the mouse to select the column, 'FilterId', by clicking on the column header. While holding down the left key grab the column and move it so that 'FilterId' is the first column of the table.

Select all the records, copy and paste them into a new sheet of the operative excel spreadsheet, rename this sheet, 'apcd checkin'.

Sort the sheet just created by 'Monitor\_ID' and 'SampDate'.

Data Entry Comparisons

The two independently entered datasets need to be compared for data entry agreement. APCD verifies the accuracy of the field data of every hi vol record that is imported into the PMT system. Where differences exist APCD determines which data set is correct and makes the appropriate changes.

Navigate to the 'checks' sheet. Beginning in column 'p' create new column headers, 'aq', 'instno' (column Q), 'smpdate' (column R), and 'runtime' (column S). Using the 'vlookup' function of Excel APCD needs to verify that the records collected from the MTL system match with what APCD has previously entered. If the previous steps have been followed then the sheets of the spreadsheet have a known format. The following Excel phrases can be used:

For AQS ID: =IF(MID(VLOOKUP(\$F2,'apcd checkin'!\$A\$2:\$K\$1012,2,FALSE),1,9) = C2,0,1),

For Instrument #: =IF(VLOOKUP(\$F2,'apcd checkin'!\$A\$2:\$K\$1012,3,FALSE) = D2,0,1),

For sample date: =IF(VLOOKUP(\$F2,'apcd checkin'!\$A\$2:\$K\$1012,4,FALSE) = E2,0,1),

For sample time: =IF(VLOOKUP(\$F2,'apcd checkin'!\$A\$2:\$K\$1012,7,FALSE) = M2,0,1).

The user should make sure that the number of records in sheet 'apcd checkin' is less than the number of records identified in the phrases (in this instance the number of records is 1012), otherwise that portion of the phrase can be edited to accommodate the number of records on the 'apcd checkin' sheet. This

comparison needs to be made for every record in ‘checks’, the values in p2 – s2 should be copied to the rest of the records in the sheet ‘checks’.

#### Evaluation

The phrases, above will return a value of ‘0’ if the value provided by the lab corresponds with what has been entered by APCD; a value of ‘1’ will be returned where there is a discrepancy. APCD identifies every record with a value of ‘1’s in any of the columns (p:s) and explore why there is a difference between the data sets. If the data entry error is on the part of APCD then the data should be corrected in ‘hivolcheckin’ and a comment made in column u. If the data entry error was made by the lab then the corresponding record in ‘exp’ should be changed and a comment made in column u. After all the entry errors have been resolved and ‘exp’ has been made the first sheet in the spreadsheet save the excel spreadsheet in the ‘Field\_data\DUMP\FRM\_Hivol’ directory as ‘LSD Hi Vol YYMMDD.xls (or xlsx) where YYMMDD is the date of the operation (i.e. today’s date).

Note: If the use/intent of the vlookup phrases isn’t clear just open a previously verified data set and copy the cells from that ‘checks’ sheet into the current ‘checks’ sheet. The user will have to verify that the references are correct.

#### Uploading to PMT System

From the Main Menu press the button ‘HiVol’.

Enter today’s date in the ‘Upload Date’ box (Note: this date can be used to filter sample records to easily identify the records loaded ‘today’)

Press button ‘1’, Select the MS Excel spreadsheet created earlier.

Press button ‘3’, Append the data to the Filter table in the PMT system.

### **Initial Evaluation of HiVol Data**

#### Applying Initial Flags

If not already open then open the PMT system. Open the table, ‘Filter’, scroll right until you see the field, ‘RetrievedDate’. Filter this field, you’ll want to only see the records just loaded into the system, this data should equal the same date input in a previous step while uploading data.

User can sort the table to identify values that don’t meet critical criteria for sample time and sample flow rate and otherwise to identify sample criteria that are beyond the typical range of data. The correct null codes for these samples can be applied in ‘AIRSReEngNullCode’. Audit results can be incorporated in this step. The table, ‘Audit\_QA’ should be inspected, filtered based on the sample dates included in this set of filters and sorted by ‘AIRS\_POC’. Should there be any notable audit results in this time frame that can be associated with the samples under inspection the appropriate flags or null codes can be applied. The table, ‘AIRSNULValueCodes’ identifies the null codes and their descriptions. The table ‘AQSValidityCode’ identifies various sample flags and their descriptions.

Once the correct records have been identified and null codes or sample flags have been applied, the data set can be further evaluated. Navigate to the Main Menu. From the Main Menu press the button, ‘Data Report Menu’, in the first box next to the field ‘Param Code’ enter ‘85101’. Press ‘Create Dataset, when available press ‘Datasheet View (with Flags)’. These operations will run a series of queries that report the data set recently imported into PMT and organize it in a table organized by AQS ID and sample date. Additionally, these steps result in a dataset that includes the concentration, any flags applied in the previous step, and null codes applied in the previous step in the following format: xx.x (()), where ‘xx.x’ is the mass concentration of the sample, the first () would contain any sample flags (if appropriate), the second () would contain any null codes (if appropriate). Each column (instrument) should be sorted from high to low and inspected for negative values, extremely high values (in excess of 140  $\mu\text{g}/\text{m}^3$ ) and values that make sense in terms of spatially related sites (especially collocated samples). Each issue identified in the previous step should be investigated – for example if a negative value the weighing process should be investigated for that specific filter. For high values look for similarly high values at spatially related sites. For poorly collocated results look at the order the sample was taken, investigate if samples could have been switched, etc. When necessary LSD can reweigh the sample, special care must be taken so that this sample isn’t loaded to AQS until after the reweight.