18.0 Reconciliation with Data Quality Objectives

Section 3 described the data quality objective (DQO) process, which is an important planning tool to determine the objectives of an environmental data operation, to understand and agree upon the allowable uncertainty in the data and, with that, to optimize the sampling design. This information, along with sampling and analytical methods and appropriate QA/QC, should be documented in an organization's QAPP. The QAPP is then implemented by the monitoring organizations under the premise that if it is followed, the DQOs should be met. Reconciliation with the DQO involves reviewing both routine and QA/QC data to determine whether the DQOs have been attained and that the data are adequate for their intended use. This process of evaluating the data against the DQOs has been termed data quality assessment (DQA).

The DQA process has been developed for cases where formal DQOs have been established. However, these procedures can also be used for data that do not formally have DQOs. Guidance on the DQA process can be found in the documents titled *Data Quality Assessment: A Reviewer's Guide (EPA QA/G-9R)*[1] and its companion document *Data Quality Assessment: Statistical Tools for Practitioners (EPA QA/G-9S)*[2]. This document focuses on evaluating data for fitness in decision-making and also provides many graphical and statistical tools.

As stated in EPA QA/G-9R "*Data quality, as a concept, is meaningful only when it relates to the intended use of the data*". By using the DQA Process, one can answer four fundamental questions:

1. Can the decision (or estimate) be made with the desired level of certainty, given the quality of the data set?
2. How well did the sampling design perform?
3. If the same sampling design strategy is used again for a similar study, would the data be expected to support the same intended use with the desired level of uncertainty?
4. Is it likely that sufficient samples were taken to enable the reviewer to see an effect if it was really present?

DQA is a key part of the assessment phase of the data life cycle (Figure 18.1), which is very similar to the ambient air QA life cycle described in Section 1. As the part of the assessment phase that follows data validation and verification, DQA determines how well the validated data can support their intended use.

18.1 Five Steps of the DQA Process

As described in *EPA QA/G-9R*[1] and *EPA QA/G-9S*[2], the DQA process is comprised of five steps. The steps are detailed below. Since DQOs are available for the $PM_{2.5}$ program, they will be used as an example for the type of information that might be considered in each step. The $PM_{2.5}$ information is italicized and comes from a model $PM_{2.5}$ QAPP[3] for a fictitious reporting organization called Palookaville. The model QAPP was developed to help monitoring organizations develop QAPPs based upon the new R-5 QAPP requirements.

Comment [ZJW1]: Make sure all references are up to date and the current version of the document.

[1] http://www.epa.gov/quality1/qs-docs/g9r-final.pdf
[2] http://www.epa.gov/quality1/qs-docs/g9s-final.pdf
[3] http://www.epa.gov/ttn/amtic/pmqainf.html

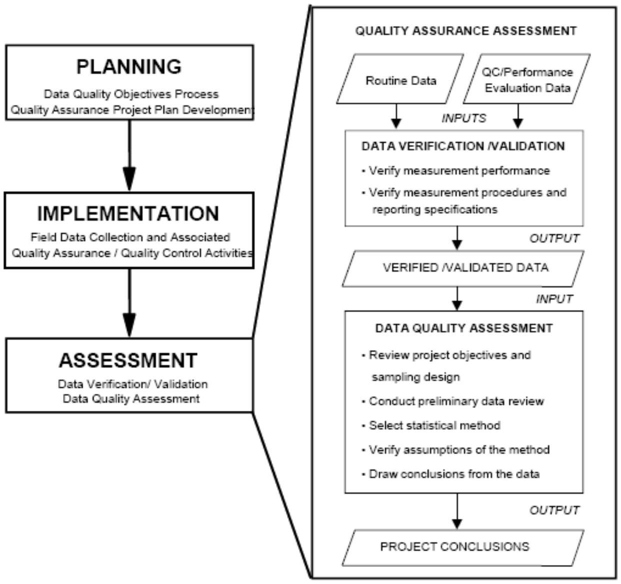

Figure 18.1 DQA in the context of data life cycle.

<u>Step 1. Review DQOs and Sampling Design.</u> Review the DQO outputs to assure that they are still applicable. If DQOs have not been developed, specify DQOs before evaluating the data (e.g., for environmental decisions, define the statistical hypothesis and specify tolerable limits on decision errors; for estimation problems, define an acceptable confidence probability interval width). Review the sampling design and data collection documentation for consistency with the DQOs observing any potential discrepancies.

The $PM_{2.5}$ DQOs define the primary objective of the $PM_{2.5}$ ambient air monitoring network ($PM_{2.5}$ NAAQS comparison), translate the objective into a statistical hypothesis (3-year average of annual mean $PM_{2.5}$ concentrations less than or equal to 15 µg/m^3 and 3-year average of annual 98th percentiles of the $PM_{2.5}$ concentrations less than or equal to 35 µg/m^3), and identify limits on the decision errors (incorrectly conclude area in non-attainment when it truly is in attainment no more than 5% of the time, and incorrectly conclude area in attainment when it truly is in non-attainment no more than 5% of the time).

The CFR contains the details for the sampling design, including the rationale for the design, the design assumptions, and the sampling locations and frequency. If any deviations from the sampling design have occurred, these will be indicated and their potential effect carefully considered throughout the entire DQA.

<u>Step 2. Conduct Preliminary Data Review.</u> Review QA reports, calculate basic statistics, and generate graphs of data. Use this information to understand the structure of the data and identify patterns, relationships, or potential anomalies.

A preliminary data review will be performed to uncover potential limitations of using the data, to reveal outliers, and generally to explore the basic structure of the data. The first step is to review the quality assurance reports. The second step is to calculate basic summary statistics, generate graphical presentations of the data, and review these summary statistics and graphs.

***Review Quality Assurance Reports**. Palookaville will review all relevant quality assurance reports that describe the data collection and reporting process. Particular attention will be directed to looking for anomalies in recorded data, missing values, and any deviations from standard operating procedures. This is a qualitative review. However, any concerns will be further investigated in the next two steps.*

***Calculation of Summary Statistics and Generation of Graphical Presentations**. Palookaville will generate prominent summary statistics for each of its primary and QA samplers. These summary statistics will be calculated at the quarterly, annual, and three-year levels and will include only valid samples. The summary statistics are:*

> *Number of samples, mean concentration, median concentration, standard deviation, coefficient of variation, maximum concentration, minimum concentration, interquartile range, skewness and kurtosis.*

These statistics will also be calculated for the percent differences at the collocated sites. The results will be summarized in a table. Particular attention will be given to the impact on the statistics caused by the observations noted in the quality assurance review. For example, Palookaville may evaluate the influence of a potential outlier by evaluating the change in the summary statistics resulting from exclusion of the outlier.

Palookaville will generate graphics to present the results from the summary statistics and show the spatial continuity over the sample areas. Maps will be created for the annual and three-year means, maxima, and interquartile ranges for a total of 6 maps. The maps will help uncover potential outliers and will help in the network design review. Additionally, basic histograms will be generated for each of the primary and QA samplers and for the percent difference at the collocated sites. The histograms will be useful in identifying anomalies and evaluating the normality assumption in the measurement errors.

Step 3. Select the Statistical Test. Select the most appropriate procedure for summarizing and analyzing the data, based upon the reviews of the performance and acceptance criteria associated with the DQOs, the sampling design, and the preliminary data review. Identify the key underlying assumptions that must hold for the statistical procedures to be valid.

The primary objective for the $PM_{2.5}$ mass monitoring is determining compliance with the $PM_{2.5}$ NAAQS. As a result, the null and alternative hypotheses are:

$$H_0: X \leq 15\,\mu g/m^3 \text{ and } Y \leq 35\,\mu g/m^3$$
$$H_A: X > 15\,\mu g/m^3 \text{ or } Y > 35\,\mu g/m^3$$

where X is the three-year average $PM_{2.5}$ concentration and Y is the three-year average of the annual 98th percentiles of the $PM_{2.5}$ concentrations recorded for an individual monitor. The exact calculations for X and Y are specified in 40 CFR Part 50, Appendix N. The null hypothesis is rejected; that is, it is concluded that the area is not in compliance with the $PM_{2.5}$ NAAQS when the observed three-year *average of the annual arithmetic mean concentration exceeds 15.05 µg/m^3 or when the observed three-year average of the annual 98th percentiles exceeds 35.5 µg/m^3. If the bias of the sampler is ± 10% and the precision is within 10%, then the error rates (Type I and Type II) associated with this statistical test are less than or equal to 5%. The definitions of bias and precision will be outlined in the following step.*

<u>Step 4. Verify Assumptions of Statistical Test.</u> Evaluate whether the underlying assumptions hold, or whether departures are acceptable, given the actual data and other information about the study.

The assumptions behind the statistical test include those associated with the development of the DQOs in addition to the bias and precision assumptions. The method of verification will be addressed in this step. Note that when less than three years of data are available, this verification will be based on as much data as are available.

The DQO is based on the annual arithmetic mean NAAQS*. For each primary sampler, Palookaville will determine which, if either, of the $PM_{2.5}$ NAAQS concentration is violated. In the DQO development, it was assumed that the annual standard is more restrictive than the 24-hour standard. If there are any samplers that violate ONLY the 24-hour NAAQS, then this assumption is not correct. The seriousness of violating this assumption is not clear. Conceptually, the DQOs can be developed based on the 24-hour NAAQS and the more restrictive bias and precision limits selected. However, Palookaville will assume the annual standard is more restrictive, until proven otherwise.*

Normal distribution for measurement error. *Assuming that measurement errors are normally distributed is common in environmental monitoring. Palookaville has not investigated the sensitivity of the statistical test to violate this assumption; although, small departures from normality generally do not create serious problems. Instead, Palookaville will evaluate the reasonableness of the normality assumption by reviewing a normal probability plot, and calculating the Shapiro-Wilk W Test statistic (if sample size less than 50) or calculating the Kolmogorov-Smirnoff Test statistic (if sample size greater than 50). All three techniques are provided by standard statistical packages. If the plot or statistics indicate possible violations of normality, Palookaville may need to determine the sensitivity of the DQOs to departures in normality.*

Decision error can occur when the estimated 3-year average differs from the actual (true) 3-year average. This is not really an assumption as much as a statement that the data collected by an ambient air monitor is stochastic, meaning that there are errors in the measurement process, as mentioned in the previous assumption.

The limits on precision and bias are based on the smallest number of required sample values in a 3-year period. In the development of the DQOs, the smallest number of required samples was used. The reason for this was to ensure that the confidence was sufficient in the minimal case; if more samples are collected, then the confidence in the resulting decision will be even higher. For each of the samplers, Palookaville will determine how many samples were collected in each quarter. If this number meets or exceeds 12, then the data completeness requirements for the DQO are met.

The decision error limits were set at 5%. If the other assumptions are met, then the decision error limits are less than or equal to 5%.

Measurement imprecision was established at 10% coefficient of variation (CV). For each sampler, Palookaville will review the coefficient of variation calculated in Step 2. If any exceed 10%, Palookaville may need to determine the sensitivity of the DQOs to larger levels of measurement imprecision.

Table 18-1 will be completed during each DQA. The table summarizes which, if any, assumptions have been violated. A check will be placed in each of the row/column combinations that apply. Ideally, there will be no checks. However, if there are checks in the table, the implication is that the decision error rates are unknown, even if the bias and precision limits are achieved. As mentioned above, if any of the DQO assumptions are violated, then Palookaville will need to reevaluate its DQOs.

Achievement of bias and precision limits. *Lastly, Palookaville will check the assumption that at the 3-year level of aggregation, the sampler bias is within ± 10% and precision is < 10%. The data from the collocated samplers will be used to calculate quarterly, annual, and 3-year bias and precision estimates even though it is only the 3-year estimates that are critical for the statistical test.*

Since all the initial samplers being deployed by Palookaville will be FRMs, the samplers at each of the collocated sites will be identical method designations. As such, it is difficult to determine which of the collocated samplers is closer to the true $PM_{2.5}$ concentration. Palookaville will calculate an estimate of precision. A bias measure will also be calculated, but it can only describe the relative difference of one sampler to the other, not definitively indicate which sampler is closer to the "true" value. The following paragraphs contain the algorithms for calculating precision and bias. These are similar, but differ slightly, from the equations in 40 CFR Part 58, Appendix A.

Table 18-1 Summary of Violations of DQO Assumptions

Site	Violate 24-Hour Standard ONLY?	Measurement Errors Non-Normal?	Data Complete? (∃ 12 samples per quarter)	Measurement CV > 10%?
Primary Samplers				
A1				
A2				
A3				
A4				
B1				
QA Samplers				
A1				
B1				

Before describing the algorithm, some ground work is necessary. When less than three years of collocated data are available, then the three-year bias and precision estimates must be predicted. Palookaville's strategy for accomplishing this will be to use all available quarters of data as the basis for projecting where the bias and precision estimates will be at the end of the three-year monitoring period. Three-year point estimates will be computed by weighting the quarterly components, using the most applicable of the following assumptions:

1. *Most recent quarter's precision and bias are most representative of what the future quarters will be.*
2. *All previous quarters precision and bias are equally representative of what the future quarter's will be.*
3. *Something unusual happened in the most recent quarter, so the most representative quarters are all the previous ones, minus the most recent.*

Each of these scenarios results in weights that will be used in the following algorithms. The weights are shown in Table 18-2 where the variable Q represents the number of quarters for which observed bias and precision estimates are available. Note that when Q=12, that is, when there are bias and precision values for all of the quarters in the three-year period, then all of the following scenarios result in the same weighting scheme.

Table 18-2 Weights for Estimating Three-Year Bias and Precision

Scenario	Assumption	Weights
1	Latest quarter most representative	w_q = 12-(Q-1) for latest quarter, w_q = 1 otherwise
2	All quarters equally representative	w_q = 12/Q for each quarter
3	Latest quarter unrepresentative	w_q = 1 for latest quarter, w_q = 11/(Q-1) otherwise

In addition to point estimates, Palookaville will develop confidence intervals for the bias and precision estimates. This will be accomplished using a re-sampling technique. The protocol for creating the confidence intervals are outlined in Box 18.1.

Box 18.1 Method for Estimating Confidence in Achieving Bias and Precision DQOs

Let Z be the statistic of interest (bias or precision). For a given weighting scenario, the re-sampling will be implemented as follows:

1. Determine M, the number of collocated pairs per quarter for the remaining 12-Q quarters (default is M=15 or can use M=average number observed for the previous Q quarters.
2. Randomly select with replacement M collocated pairs per quarter for each of the future 12-Q quarters in a manner consistent with the given weighting scenario.
Scenario 1: Select pairs from latest quarter only.
Scenario 2: Select pairs from any quarter.
Scenario 3: Select pairs from any quarter except the latest one.
Result from this step is "complete" collocated data for a three-year period, from which bias and precision estimates can be determined.
3. Based on the "filled-out" three-year period from step 2, calculate three-year bias and precision estimate, using Equation 1 where w_q = 1 for each quarter.
4. Repeat steps 2 and 3 numerous times, such as 1000 times.
5. Determine P, the fraction of the 1000 simulations for which the three-year bias and precision criteria are met. P is interpreted as the probability that the sampler is generating observations consistent with the three-year bias and precision DQOs.

The algorithms for determining whether the bias and precision DQOs have been achieved for each sampler follow:

Bias Algorithm

1. For each measurement pair, estimate the percent relative bias, d_i.

$$d_i = \frac{Y_i - X_i}{(Y_i + X_i)/2} \times 100\,\%$$

where X_i represents the concentration recorded by the primary sampler and Y_i represents the concentration recorded by the collocated sampler.

2. Summarize the percent relative bias to the quarterly level, $D_{j,q}$, according to

$$D_{j,q} = \frac{1}{n_{j,q}} \sum_{i=1}^{n_{j,q}} d_i$$

where $n_{j,q}$ is the number of collocated pairs in quarter q for site j.

3. Summarize the quarterly bias estimates to the three-year level using

$$\hat{D}_j = \frac{\sum_{q=1}^{n_q} w_q \, D_{j,q}}{\sum_{q=1}^{n_q} w_q} \qquad \text{Equation 18-1}$$

where n_q is the number of quarters with actual collocated data and w_q is the weight for quarter q as specified by the scenario in Table 18-2.

4. Examine $D_{j,q}$ to determine whether one sampler is consistently measuring above or below the other. To formally test this, a non-parametric test will be used (Wilcoxon Signed Rank Test), which is described in EPA QA/G-9S[2]. If the null hypothesis is rejected, then one of the samplers is consistently measuring above or below the other. This information may be helpful in directing the investigation into the cause of the bias.

Precision Algorithm

1. For each measurement pair, calculate the coefficient of variation, cv_i,

$$cv_i = \frac{|d_i|}{\sqrt{2}}$$

2. *Summarize the coefficient of variation to the quarterly level, $CV_{j,q}$, according to*

$$CV_{j,q} = \sqrt{\frac{\sum_{i=1}^{n_j} CV_i^2}{n_{j,q}}}$$

where $n_{j,q}$ is the number of collocated pairs in quarter q for site j.

3. *Summarize the quarterly precision estimates to the three-year level using*

$$\hat{CV}_j = \sqrt{\frac{\sum_{q=1}^{n_q} \left(w_q CV_{j,q}^{\ 2}\right)}{\sum_{q=1}^{n_q} w_q}} \qquad \textit{Equation 18-2}$$

where n_q is the number of quarters with actual collocated data and w_q is the weight for quarter q as specified by the scenario in Table 24-2 (reference to Model QAPP).

4. *If the null hypothesis in the Wilcoxon Signed Rank Test was not rejected, then the coefficient of variation can be interpreted as a measure of precision. If the null hypothesis in the Wilcoxon Ssigned Rank Test was rejected, the coefficient of variation has both a component representing precision and a component representing the (squared) bias.*

Confidence in Bias and Precision Estimates

1. *Follow the method described in Box 18.1 to estimate the probability that the sampler is generating observations consistent with the three-year bias and precision DQOs. The re-sampling must be done for each collocated site.*

Summary of Bias and Precision Estimation

The results from the calculations and re-sampling will be summarized in Table 18-3. There will be one line for each site operating a collocated sampler.

Table 18-3 Summary of Bias and Precision

Collocated	Three-year Bias Estimate (Equation. 1)	Three-year Precision Estimate (Equation. 2)	Null Hypothesis of Wilcoxon Test Rejected?	*P* (Box 18-1)
A1				
B1				

Step 5. Draw Conclusions from the Data. Perform the calculations required for the statistical test and document the inferences drawn as a result of these calculations. If the design is to be used again, evaluate the performance of the sampling design.

Before determining whether the monitored data indicate compliance with the $PM_{2.5}$ NAAQS, Palookaville must first determine if any of the assumptions upon which the statistical test is based are violated. This can be easily checked in Step 5 because of all the work done in Step 4. In particular, as long as

- *in Table 18-1, there are no checks, and*
- *in Table 18-3,*
  - R *the three year bias estimate is in the interval [-10%,10%], and*
  - R *the three year precision estimate is less than or equal to 10%*

then the assumptions underlying the test appear to be valid. As a result, if the observed three-year average $PM_{2.5}$ concentration is less than 15 $\mu g/m^3$ and the observed three-year average 98th percentile is less than 35 $\mu g/m^3$, the conclusion is that the area seems to be in compliance with the $PM_{2.5}$ NAAQS, with an error rate of 5%.

If any of the assumptions have been violated, then the level of confidence associated with the test is suspect and will have to be further investigated.

DQA without DQOs

Even though DQOs, based upon the EPA G-4 guidance, have not been developed for all criteria pollutants, a process very similar to this approach was originally used[4]. In addition, monitoring organizations collect enough types of QA/QC data to estimate the quality of their data and should be able to express the confidence in that information.

[4] Curran, Thomas C. et.al., "Establishing Data Quality Acceptance Criteria for Air Pollution Data" Transactions of the 35 Annual Conference of the American Society for Quality Control (May 27-29,1981)