

Data Quality Objectives for PM Continuous Methods II

Prepared for

National Exposure Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Contract 68-D-00-206

ManTech Environmental Technology, Inc., Team
P.O. Box 12313
Research Triangle Park, NC 27709

A ManTech International Company

Data Quality Objectives for PM Continuous Methods II

by

Paul Mosquin and Frank McElroy
Research Triangle Institute

ManTech Team

*ManTech Environmental Technology, Inc.
Research Triangle Park, NC 27709
Prime Contractor*

*Research Triangle Institute
Research Triangle Park, NC 27709
Subcontractor*

Submitted to

Elizabeth Hunike, Contracting Officer's Representative
Process Modeling Research Branch
Human Exposure and Atmospheric Sciences Division
National Exposure Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Contract 68-D-00-206

Reviewed and approved by

Frank McElroy
Principal Investigator

Christopher Noble
RTI Technical Supervisor

ManTech Environmental Technology, Inc.
P.O. Box 12313
Research Triangle Park, NC 27709

Foreword

This technical report presents the results of work performed by ManTech Environmental Technology, Inc., and RTI International under Contract 68-D-00-206 for the Human Exposure and Atmospheric Sciences Division, National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC. This technical report has been reviewed by ManTech Environmental Technology, Inc., and RTI International and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Contents

Section	Page
Foreword	iii
Figures	v
Tables	vi
1 Introduction	1
2 Review of DQO and WA 76 Model	3
3 Review of Field Samples	5
3.1 Bakersfield Field Data	5
3.2 Detroit Field Data	8
3.3 New York Field Data	9
3.4 Tampa Field Data	10
3.5 Winston-Salem Field Data	12
3.6 Conclusions from Field Data	12
4 Proposed Measurement Error Model	17
5 Performance Bounds for Model Parameters	19
5.1 Bounds for the Precision, the Intercept, and the Slope	19
5.2 Bounds for the Correlation Coefficient	22
6 Parameter Estimation	27
6.1 Maximum Likelihood Estimation	27
6.2 Other Estimators	27
6.3 Correlation and Its Estimation	29
7 Illustration Using Simulated and Field Data Sets	33
7.1 Model Fit Using Simulated Data	33
7.2 Model Fit Using Field Data	35
8 Discussion	39
9 References	41
Appendix A: R Code for Maximum Likelihood Estimation	43
Appendix B: Precision and Variance.	45

Figures

Figure		Page
1	Time series of days with readings for all four samplers from the Bakersfield data set	6
2	Scatterplots of Bakersfield readings for collocated FRM and collocated MetOne BAM samplers	6
3	Scatterplots of Detroit readings for collocated FRM and an R&P TEOM running at 50 °C	8
4	Scatterplot of Detroit TEOM readings vs. FRM readings with points labeled by month	9
5	Scatterplots of New York readings for collocated FRM and an R&P TEOM running at 50 °C	10
6	Time series of days with complete data for the Tampa site	11
7	Scatterplots of Tampa readings for the FRM, the Thermo-Andersen, and the DKK beta gauges	11
8	Scatterplot of Winston-Salem FRM and R&P TEOM readings	12
9	Scatterplots illustrating the effect of an unmeasured variable on the correlation	15
10	Allowable multiplicative and additive biases according to the DQO model	22
11	Left plot is of 100 independent normal $N(50,400)$ variates adjusted for measurement error. Right plot is 100 independent normal $N(50,400)$ variates truncated within $[45,55]$ and then adjusted for measurement error	24
12	Simulated sampler readings vs. true daily values for reference samplers and candidate samplers	33
13	Estimated daily values vs. true daily values	34

Tables

Table		Page
1	Sample Correlations among Bakersfield Samplers	7
2	Bakersfield Parameter Estimates for Candidate Samplers According to the WA 76 Model	7
3	Field Data Estimated Parameter Values Based on Current Equivalency Methods	13
4	Comparison of Quantities Estimated between Current DQO and Equivalency Methods	13
5	Sample Means, Standard Deviations, and 5th and 95th Percentiles for One Million Simulations of \bar{Z} at Varying Sampler Precisions	21
6	Correlations for Single Reference vs. Candidate Samplers for Various τ_U and σ_C	25
7	Correlations for Sample Means of Three Reference vs. Three Candidate Samplers for Various τ_U and σ_C	25
8	Sample Data for Correlation Analysis	31
9	Field Data Estimated Parameter Values after Likelihood Fit and Correlation Analysis	36

1. Introduction

Under this work assignment (WA), field data from the national PM_{2.5} particulate monitoring network was analyzed to further develop methodology presented in the earlier WA 76 (U.S. EPA, 2003). The ultimate goal of both studies is the development of methods for testing whether samplers that are candidates for equivalent method status in the PM_{2.5} network perform acceptably in comparison to current federal reference method (FRM) samplers. These tests are known as equivalency tests, and candidate samplers that pass are designated federal equivalent method (FEM) samplers. A focus of both studies is candidate samplers that can record measurements continuously over time.

A main goal of WA 76 was to compare performance requirements set by existing FRM equivalency tests with field standards required by the data quality objective (DQO) process (U.S. EPA, 2002). This comparison had the goal of determining the extent to which these independently developed standards were consistent. The conclusion of this comparison was that the two standards were largely incompatible, and the resulting recommendation was that the equivalency standards should largely be modified for compatibility with the standards required by the DQO methodology. The DQO standards are that an FRM monitor sampling at a 1-in-6-day frequency should have measurement error precision of no more than 0.1 and multiplicative bias between 0.9 and 1.1.

This WA evaluates the recommendations of WA 76 by applying the recommended methodology to field sampler data. The field data sets are obtained from five sites representing diverse geographic and climatic regions within the network and also varied continuous sampler types. This WA also conducts some exploratory data analysis of the field sampler data to check if the WA 76 methods are adequate or whether further modifications might yet be made.

Principal recommendations of this WA for future equivalency tests are as follows:

1. The basic conclusions of WA 76 should be retained, namely that the equivalency tests be performed using the DQO measurement error model for both reference and candidate samplers. For candidate samplers the DQO model allows estimation of both precision and multiplicative bias, and these parameter estimates can be compared to bounds obtained by DQO grey zone simulation.
2. Additive bias should be introduced to the DQO measurement error model to allow its estimation in equivalency tests and incorporation into DQO simulations.
3. Correlation of candidate and reference sampler readings should be used as a simple but useful measure of model fit.

The last two of these recommendations concern elements of the current equivalency tests that the exploratory data analysis suggests be retained.

This report is divided into eight sections. After the introductory material in section 1, section 2 reviews the DQO and WA 76 measurement error models. Section 3 describes the field data and conclusions that might be drawn from it, and section 4 gives a proposed measurement error model. Section 5 deals with obtaining performance bounds for the model parameters, and section 6 describes approaches to estimating parameters. Application of the resulting methods to simulated and field data is done in section 7. Section 8 provides a discussion.

2. Review of DQO and WA 76 Model

The DQO model consists of both a population model for the generation of true $PM_{2.5}$ daily values and a measurement error model that adjusts these true values to provide simulated readings of field monitors in the national network.

The DQO measurement error model is that for sampler $j=1, \dots, J$ at site $i=1, \dots, I$ at time $k=1, \dots, K$ the recorded measurement Y_{ijk} is equal to

$$Y_{ijk} = \beta_j U_{ik} \varepsilon_{ijk}$$

where U_{ik} is the true particulate concentration at a given site at a given time, β_j is a sampler-specific multiplicative bias, and the ε_{ijk} are independent normal errors $\varepsilon_{ijk} \sim N(1, \sigma_{C,j}^2)$ with $\sigma_{C,j}^2$ being the precision¹ of sampler j . The DQO model thus allows both sampler-specific precisions and biases since it is used to set performance bounds on individual samplers in the field. When they need to be simulated, true daily values U_{ik} are simulated according to the DQO population model, described later in section 5.1.

In adapting the DQO model for use with equivalency tests, it was necessary to specify a separate measurement error model for both the FRM samplers and the candidate samplers. The FRM measurement error model has observed daily reading X_{ijk}^* equal to the product of the true daily value and a normally distributed random error

$$X_{ijk}^* = U_{ik} \varepsilon_{ijk}^*$$

where $\varepsilon_{ijk}^* \sim N(1, \sigma_R^2)$ and the ε_{ijk}^* are independent. That is, it assumes no multiplicative bias for reference samplers in the DQO test, an assumption that is based on historically observed performance of FRM samplers under equivalency test conditions. The precision parameter σ_R^2 is assumed common to all reference samplers.

For candidate samplers, the WA 76 measurement error model has observed daily reading X_{ijk} equal to the product of a sampler-specific multiplicative bias, the true value, and a random error

$$X_{ijk} = \beta_j U_{ik} \varepsilon_{ijk}$$

where $\varepsilon_{ijk} \sim N(1, \sigma_C^2)$ and are independent. Multiplicative biases are allowed to be sampler-specific; however, the precision parameter is common to all candidate samplers.

This model can be fit using an analysis-of-variance-like weighted regression where variances σ_C^2 and σ_R^2 are first estimated, and these estimates are then used as weights in the regression to estimate candidate sampler multiplicative biases β_1, \dots, β_J . Since the goal of the equivalency test is to evaluate overall properties of a type of candidate sampler, random effects β_1, \dots, β_J are considered as a group in evaluating the overall candidate sampler multiplicative bias.

¹The meaning of the term *precision* is detailed in Appendix B.

The WA 76 report thus recommended an equivalency test where candidate samplers were evaluated according to a DQO measurement error model. The report did not recommend the retention of an additive bias (intercept) or correlation from the current equivalency tests. As will now be seen, trends in the field data suggest that these parameters might still be useful in an equivalency test.

3. Review of Field Samples

Field data were provided from five sites that had both FRM and continuous monitors present. These five sites—Bakersfield, Detroit, New York, Tampa, and Winston-Salem—were selected to represent a broad spectrum of geographic regions, climatic regions, and continuous sampler types. Field data for each site consisted of time series of a single or a collocated pair of FRM samplers and of a single or a collocated pair of continuous samplers. Although some of these data should allow a basic evaluation of a proposed equivalency testing method, the data sets themselves were not collected as equivalency tests. For equivalency tests, fewer days (currently a minimum of 10) and more replicates (currently three) of both FRM and candidate samplers would be collected.

Of the sites, the Bakersfield site had both paired FRM and continuous samplers, while New York and Detroit had paired FRM and a single continuous. The Tampa site had a single FRM and two varieties of paired continuous samplers. The Winston-Salem site had only a single FRM and a single continuous. Thus, the Bakersfield site may be considered most representative of an equivalency test, having replicate samplers of both FRM and continuous kind.

Some exploratory data analysis for each of these sites is now presented. For each site, scatterplots of daily readings are given to illustrate association among the readings. Some time series are provided for interpretation of temporal trends.

3.1 Bakersfield Field Data

The Bakersfield samples provide the only data set with both collocated reference and candidate samplers. These replications of both kinds of samplers make this data set most similar to current equivalency tests.

Time series plots² of the Bakersfield readings are given in Figure 1. Trends illustrated by this series are that the continuous MetOne BAM samplers tend to produce higher readings than the FRM samplers, and that samplers of the same type tend to have more similar readings than those of different types.

The consistency of readings within sampler types is further illustrated by the pair-wise scatterplots given in Figure 2. Paired readings within sampler types show relatively high correlation as compared to correlations across sampler types. There appear to be relatively few outliers, with the apparent exception of a couple of smaller FRM readings.

²All PM_{2.5} concentration data in this report are given in µg/m³.

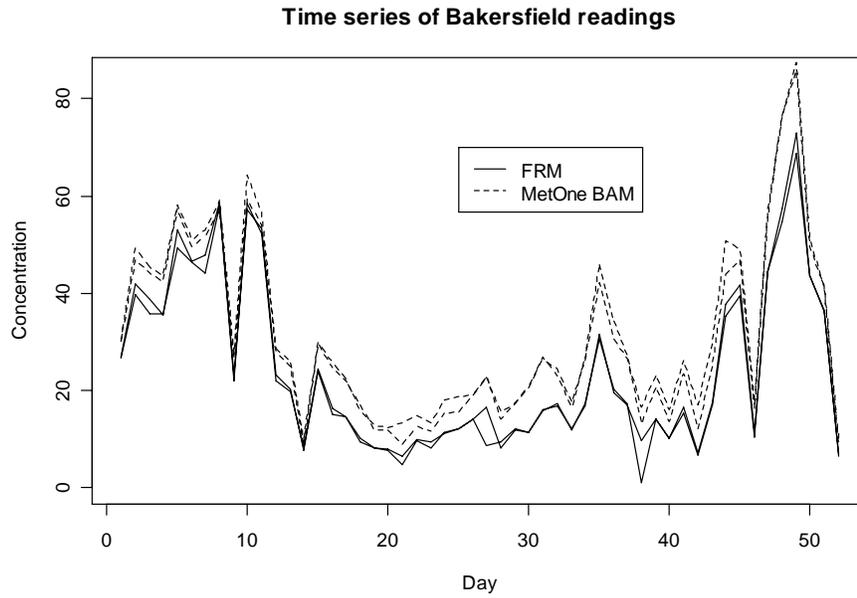


Figure 1. Time series of days with readings for all four samplers from the Bakersfield data set.

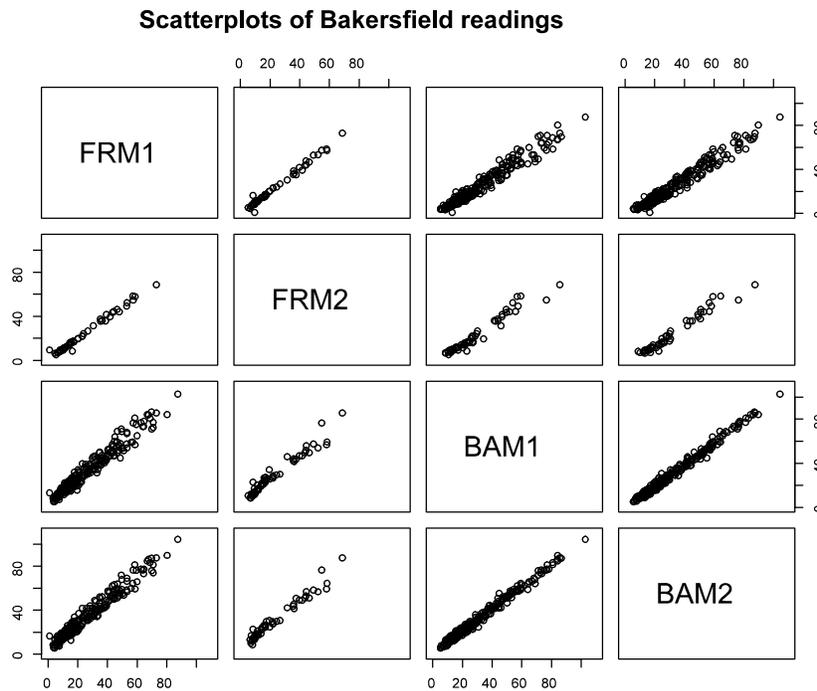


Figure 2. Scatterplots of Bakersfield readings for collocated FRM and collocated MetOne BAM samplers.

That the correlation between sampler types is less than might be expected, given the correlation within sampler types, is suggested by the correlation matrix of Table 1, where the across-type correlation is notably smaller than within type. Note that the average of the across-sampler-type correlations is 0.978, which exceeds the minimally required 0.97 of the current equivalency tests. The correlation of the sample means of the candidate and equivalent samplers equals 0.981, also exceeding the required amount.

Table 1. Sample Correlations among Bakersfield Samplers

	FRM1	FRM2	BAM1	BAM2
FRM1	1.000	0.993	0.978	0.980
FRM2	0.993	1.000	0.974	0.981
BAM1	0.978	0.974	1.000	0.994
BAM2	0.980	0.981	0.994	1.000

Computing the regression statistics of the current equivalency tests gives an intercept of 5.89 and a slope of 1.05. The slope satisfies the requirement that it lie between 0.95 and 1.05, but the intercept exceeds the bounds of -1 and 1. Thus, these field data for the collocated BAM samplers give estimates that do not satisfy current equivalency requirements.

Similarly, the methods of WA 76 can be used to evaluate the Bakersfield data as if they were from an equivalency test. Table 2 gives precisions and biases for the Bakersfield data. Two approaches are taken, with and without outliers. The analysis without outliers removes two observations from the reference sampler data set: paired observations of (16.4, 8.5) and (1.0, 9.5). There did not appear to be any outliers in the candidate data. With outliers removed, both reference and candidate samplers have estimated precisions better than the DQO-required 0.1. The average multiplicative bias of the two samplers is well above 1.1, so the candidate samplers would also fail equivalency tests according to the DQO standards and WA 76 methods.

Table 2. Bakersfield Parameter Estimates for Candidate Samplers According to the WA 76 Model

Parameter	Estimate with Outliers	Estimate without Outliers
Reference precision	0.230	0.050
Candidate precision	0.077	0.077
Average candidate bias	1.350	1.350

3.2 Detroit Field Data

The field data from Detroit were obtained from collocated FRM samplers and a continuous R&P TEOM operating at 50 °C. Scatterplots of the Detroit readings are given in Figure 3 and suggest that the FRM samplers show relatively good within-sampler precision (estimated reference sampler precision is 0.055); however, in comparison with the continuous TEOM sampler, the scatterplots suggest a bivariate mixed distribution. This mixed distribution appears to have two component distributions, each associated with seasonal effects as illustrated by Figure 4, which gives a scatterplot of the TEOM versus the first FRM sampler with the points labeled according to month of the year. Of the two component distributions of Figure 4, one appears to largely represent warmer months, while the other the cooler ones. This mixed distribution might be expected to decrease the correlation, and the average of the correlation coefficients of the two FRM-continuous pairings is 0.82, which is well below the current equivalency requirement of 0.97.

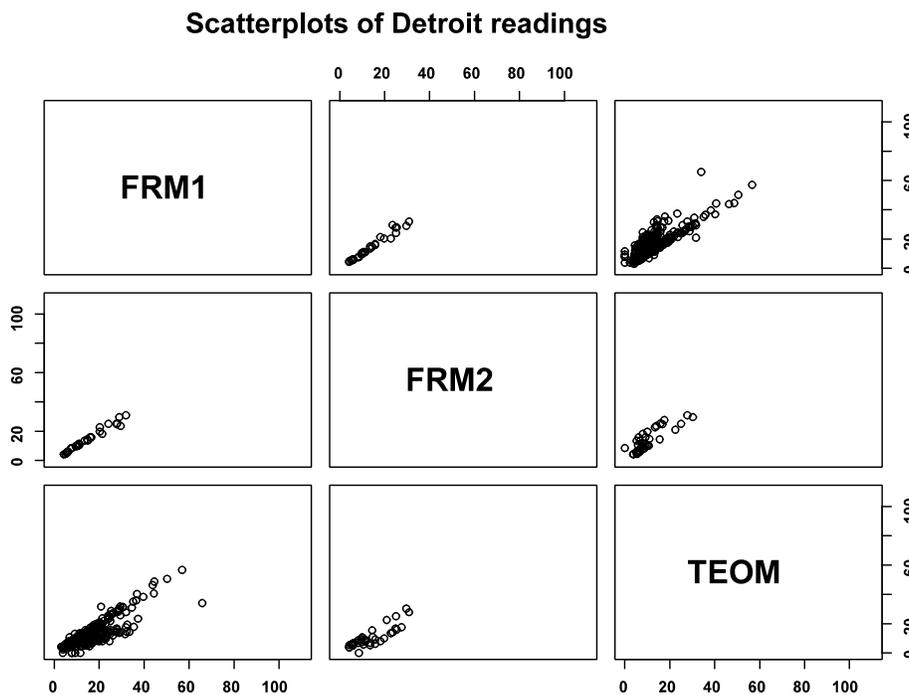


Figure 3. Scatterplots of Detroit readings for collocated FRM and an R&P TEOM running at 50 °C.

Scatterplot of Detroit TEOM vs FRM1

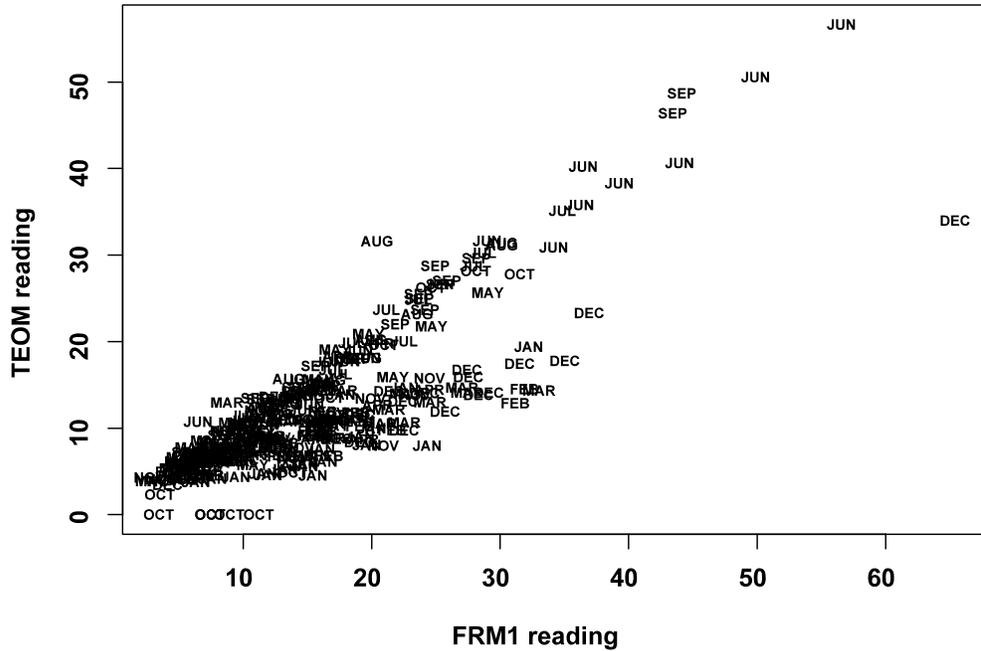


Figure 4. Scatterplot of Detroit TEOM readings vs. FRM readings with points labeled by month.

3.3 New York Field Data

Like the Detroit site, the New York site had two FRM samplers and a single R&P TEOM operating at 50 °C. Scatterplots of the New York readings are given in Figure 5, and as might be expected, show a very similar pattern to Detroit. The FRM readings are very consistent among themselves, but they again suggest a mixed distribution in comparison to the TEOM. The reference sampler precision is estimated as 0.040, and the average correlation across sampler types is 0.945.

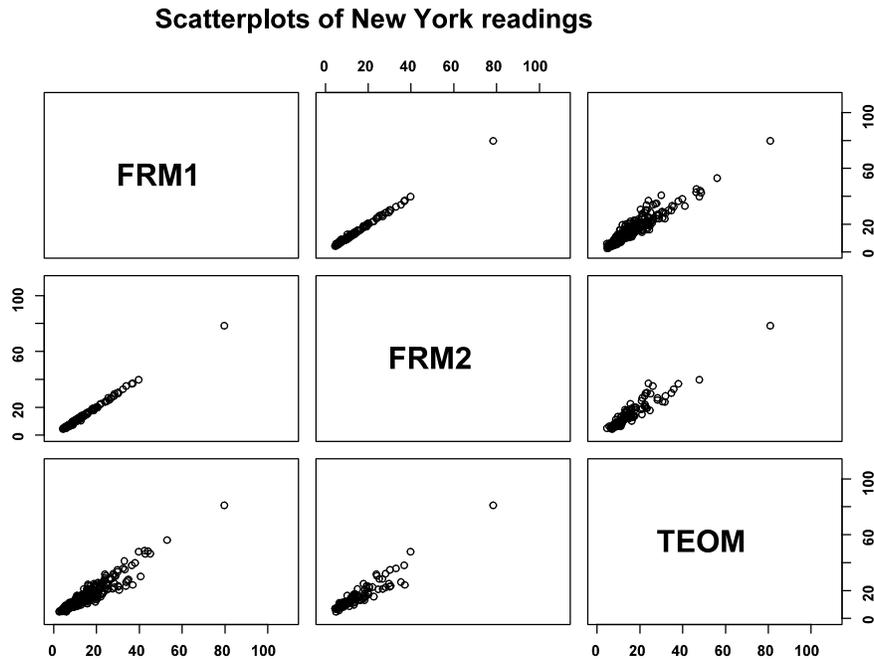


Figure 5. Scatterplots of New York readings for collocated FRM and an R&P TEOM running at 50 °C.

3.4 Tampa Field Data

The Tampa site had a single reference sampler and two types of paired continuous samplers: paired DKK beta gauges and paired FH-62 Thermo-Andersen (TA) beta gauges. A time series plot of the Tampa readings for all days without missing observations is given in Figure 6. The time series shows that the DKK and TA samplers tend to have measurements larger than the FRM sampler, with the DKK samplers typically having the largest readings. The within-sampler-type precision appears to be good for the TA samplers but poor for the DKK samplers. These effects are reflected by their estimated precisions, which are 0.061 for the TA samplers and 0.190 for the DKK samplers.

Scatterplots for the Tampa readings are given in Figure 7. As expected from the time series, a relatively high level of correlation is apparent among the TA samplers, with less correlation among DKK samplers or across sampler types. The average of the sample correlation of the TA samplers with the FRM sampler is 0.933, while that of the DKK samplers with the FRM sampler is 0.909.

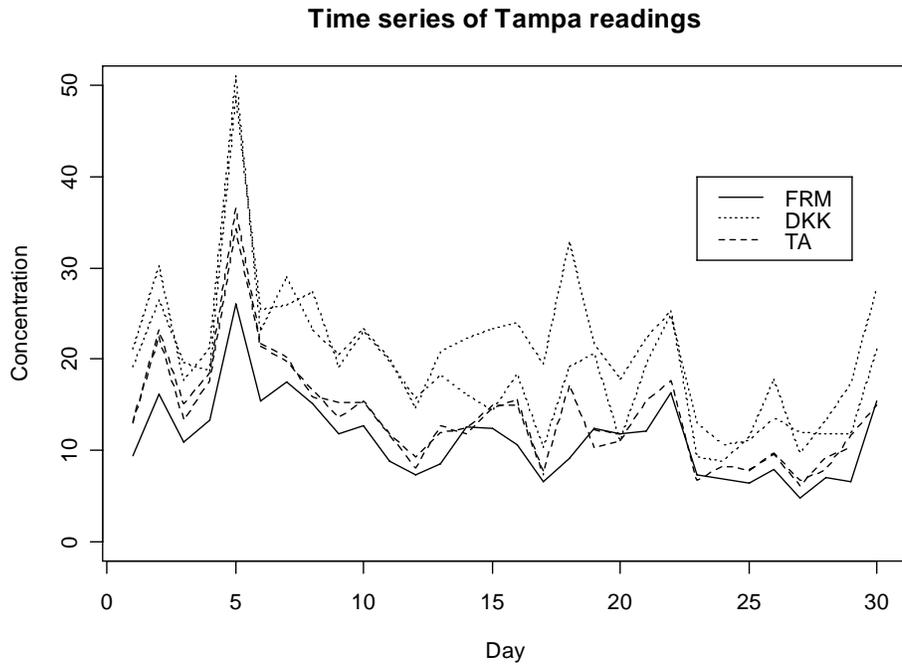


Figure 6. Time series of days with complete data for the Tampa site. Series are federal reference method (FRM) samplers, DKK beta gauge samplers, and Thermo-Andersen (TA) beta gauges.

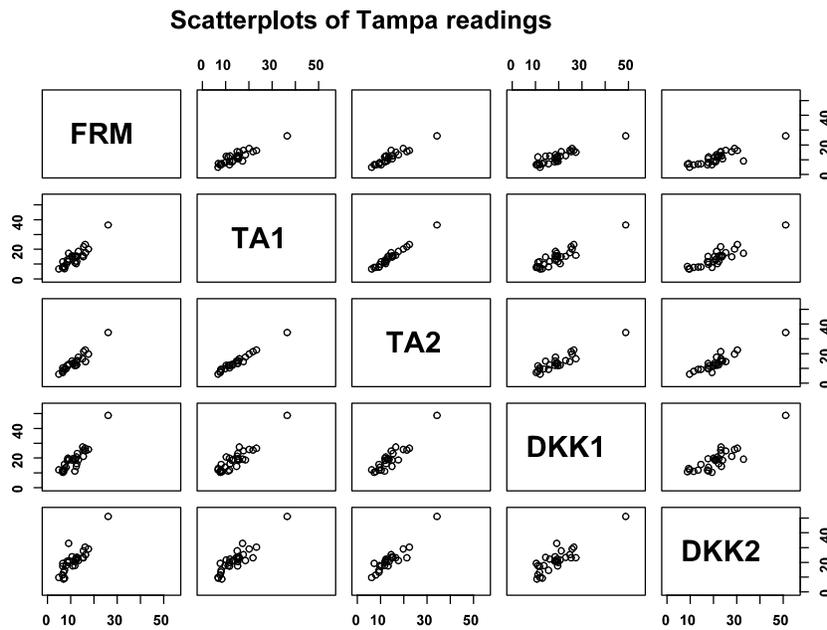


Figure 7. Scatterplots of Tampa readings for the FRM, the Thermo-Andersen, and the DKK beta gauges.

3.5 Winston-Salem Field Data

Data from the Winston-Salem site comprise readings from only a single time series of FRM and a single time series of R&P TEOM operating at 50 °C and are presented in the scatterplot in Figure 8. Unlike the more northerly Detroit and New York readings, there is little suggestion of a mixed distribution in the paired observations (with the possible exception of three observations). This might be due to the more moderate seasonal conditions at this site. The sample correlation of the two samplers is 0.979.

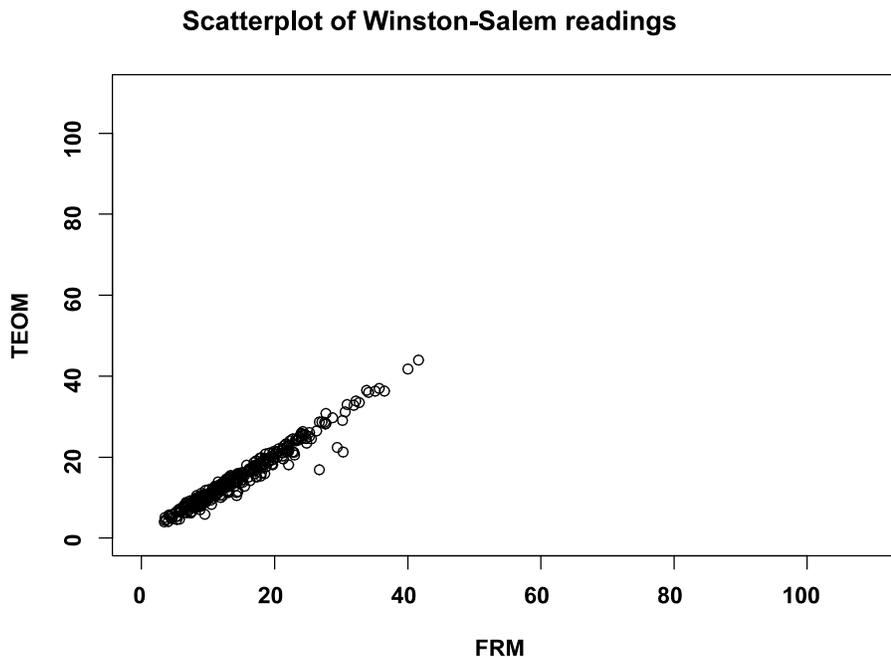


Figure 8. Scatterplot of Winston-Salem FRM and R&P TEOM readings.

3.6 Conclusions from Field Data

Current equivalency tests place constraints on the estimated correlation, regression intercept, and slope. According to current equivalency methodology, the estimated intercept of the Bakersfield readings exceeded its allowed bounds, while the other two statistics were within bounds. Alternatively, according to the WA 76 methodology, the precisions are acceptable, while the multiplicative biases are not. Thus, both the current equivalency methods and the suggested WA 76 methods would not consider the Bakersfield continuous samplers equivalent.

A summary of results from all equivalency tests as performed according to current methods is given in Table 3. It can be seen from the table that all field samplers would fail current equivalency tests for the observed statistics.

Table 3. Field Data Estimated Parameter Values Based on Current Equivalency Methods

Site	Continuous Sampler Type	Number of Reference Samplers/ Continuous Samplers/Days Complete Observations	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$
Bakersfield	BAM	2/2/52	5.89	1.05	0.981
Bakersfield ¹	BAM	2/2/50	5.55	1.06	0.982
Detroit	TEOM	2/1/31	0.36	0.70	0.826
New York	TEOM	2/1/111	1.86	0.92	0.945
Tampa	DKK	1/2/30	2.64	1.54	0.905
Tampa	TA	1/2/25	0.04	1.23	0.936

¹After removal of two FRM sampler outliers as described in section 3.1.

The differences between the DQO/WA 76 and current equivalency methods are summarized in Table 4, which gives the quantities measured as well as the required ranges by each approach. The only common parameter among them is the multiplicative bias.

Table 4. Comparison of Quantities Estimated between Current DQO and Equivalency Methods

Parameter	DQO/WA 76	Requirement	Current Equivalency Tests	Requirement
Precision	Yes	≤ 0.1	No	—
Multiplicative bias ($\mu\text{g}/\text{m}^3$)	Yes	≥ 0.9 and ≤ 1.1	Yes	≥ 0.95 and ≤ 1.05
Additive bias ($\mu\text{g}/\text{m}^3$)	No	—	Yes	≥ -1 and ≤ 1
Correlation of means	No	—	Yes	≥ 0.97

After analysis of the field data, both the precision and the slope remain useful quantities to measure, as they still minimally describe the multiplicative error model, and the observed data (at least for FRM samplers) appeared to be consistent with this basic measurement error model.

Although the WA 76 report did not recommend the inclusion of an additive bias parameter, the estimated non-zero intercept of the Bakersfield continuous samplers suggests that for some samplers at some sites the addition of an intercept parameter might lead to a better fitting model. Including an intercept does not fundamentally change the measurement error model, as the current DQO model is simply the zero intercept special case. The intercept is not currently part of the DQO methodology, however, and so adding it to the equivalency tests suggests perhaps introducing it into the DQO methodology as well.

The field data also suggest that correlation can be a useful parameter to estimate in an equivalency test. The Bakersfield readings suggested a somewhat lower across-sampler-type

correlation than might be expected, based on the within-type correlations, and the readings from Detroit and New York revealed a likely mixed distribution of measurement errors, where this mixed distribution allowed low across-sampler-type correlations. Correlation thus might be retained in the tests as a check on model adequacy, as is now described.

The field data analyzed here suggest that the correlation, and potentially the intercept, allow an evaluation of DQO model adequacy within the equivalency tests. They test for model adequacy both by generalizing the standard DQO model and by allowing for the detection of the influence of additional, unmeasured variables on sampler performance. To illustrate, consider the WA 76 candidate sampler measurement error model in the presence of unmeasured variables Z_{ik} , $i = 1, \dots, I$, $k = 1, \dots, K$ that are added to the candidate sampler readings, so at a given site on a given day

$$X_{ijk} = \beta U_{ik} \varepsilon_{ijk} + Z_{ik}.$$

These unmeasured variables might have any distribution (including mixed distributions such as seen in Detroit and New York), but suppose that they have independent normal distributions $Z_{ik} \sim N(5, U_{ik}^2 / 100)$. The model then corresponds to $X_{ijk} = \alpha + \beta U_{ik} \varepsilon_{ijk} + Z_{ik}^*$ where $\alpha = 5$ and $Z_{ik}^* \sim N(0, U_{ik}^2 / 100)$. Thus, a first effect of this variable is that its expectation can be absorbed into the intercept, and the intercept therefore does necessarily represent an inherent property of a candidate sampler.

A scatterplot of simulated measurements for two reference and two candidate samplers in the presence of this unmeasured variable model is given in Figure 9. From this figure, it can be seen that the correlation across sampler types is less than within types, an effect which is due to the presence of the unmeasured variables. This is a similar situation to that in Bakersfield, where such variables may have affected the correlation.

The data analysis then leads to the recommendation that parameters for additive bias, multiplicative bias, precision, and correlation be included in the equivalency tests. Precision and multiplicative bias are the basic parameters of the sampler measurement error model, with the additive bias allowing a somewhat more general model. In addition to correlation, additive bias can also be used to assess the fit of the basic model, with lack of fit potentially due to the presence of unmeasured variables or some other functional misspecification of the model.

Error model with missing variable

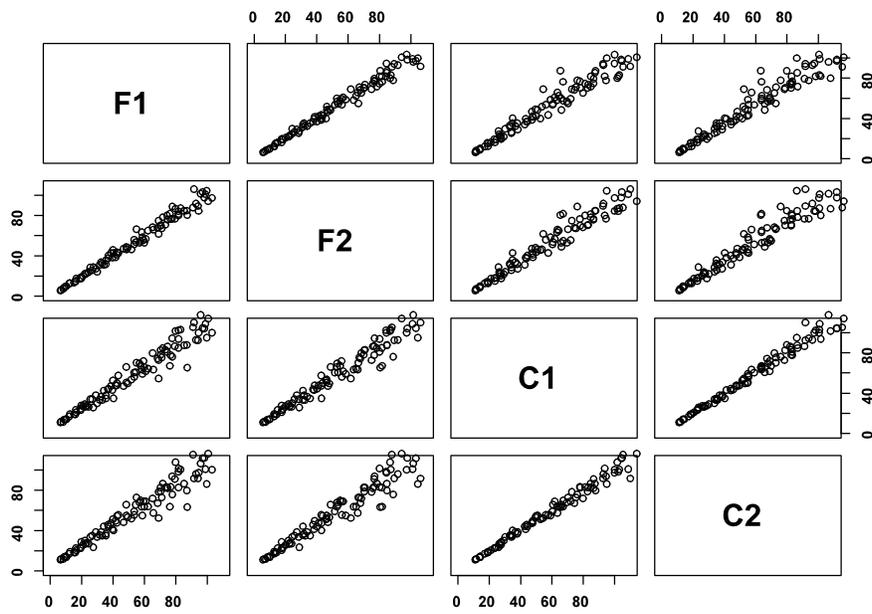


Figure 9. Scatterplots illustrating the effect of an unmeasured variable on the correlation.

4. Proposed Measurement Error Model

The model for reference sampler measurement error of sampler j at time k at site i has reference sampler reading X_{ijk}^* given as

$$X_{ijk}^* = U_{ik} \varepsilon_{ijk}^*$$

where U_{ik} is the true daily value and ε_{ijk}^* is the multiplicative measurement error. The errors are assumed independent and normally distributed $\varepsilon_{ijk}^* \sim N(1, \sigma_R^2)$ with mean equal to 1 and precision σ_R .

For the candidate samplers, the measurement error model is more general, with candidate sampler reading X_{ijk} given as

$$X_{ijk} = \alpha + \beta U_{ik} \varepsilon_{ijk}$$

where α is an intercept allowing for additive bias, β is a slope allowing for multiplicative bias, and the ε_{ijk} are independent errors that are normally distributed $\varepsilon_{ijk} \sim N(1, \sigma_C^2)$ with mean 1 and precision σ_C .

5. Performance Bounds for Model Parameters

Each equivalency test parameter must fall within some range of values that allows the candidate sampler to exhibit equivalent performance. Current bounds based on DQO methodology or equivalency tests have been given in Table 4. The parameters to be estimated and their bounds differ for the two methods of the table, and the recommendation of WA 76 was that future bounds be based on the DQO methodology.

That bounds be based on DQO methods is again recommended in this report, and the DQO methodology can be extended to provide bounds for the intercept in addition to the current slope and precision. This extension is described in section 5.1.

A bound for correlation can be obtained as a function of the bounds for precision and of parameters of the PM_{2.5} population process, which generates true daily means (in the DQO methodology this would be the sine curve combined with the daily variation). Methods to obtain bounds for the correlation coefficient are given in section 5.2.

5.1 Bounds for the Precision, the Intercept, and the Slope

Current bounds for the precision and slope are determined by DQO grey-zone simulation. This simulation is described in this section and is extended here to include an intercept. For specified population mean, multiplicative bias, and precision, the current DQO simulations simulate many realizations of sampler three-year means to estimate an error rate equal to the proportion of times that the simulated three-year mean lies on the other side of 15.05 µg/m³ from the true simulation mean.

Each simulation run proceeds as follows:

1. Generate $3 \times 365 = 1095$ days of mean particulate readings according to the DQO sinusoidal population model of yearly mean 15.

$$\mu_k = 15 + 10.238 \sin\left(\frac{2\pi j}{365}\right) \quad k = 1, \dots, 1095$$

2. Find the true daily particulate concentration for each day V_k as $V_k = \mu_k e_k$ where the e_k are independent lognormal errors of mean 1 and standard deviation 0.8. The three-year realized population mean is then $\bar{V} = \sum_k V_k / 1095$.
3. Divide each year into four quarters and select 12 days at random from each quarter according to a 1 in 6 sampling scheme whose first day is the first day of the quarter. Only 12 days are selected per quarter to allow for 25% missing data. A total of $n = 48 \times 3 = 144$ days is selected from the 1095 available with labels given as s_1, \dots, s_{144} .
4. Convert the realized process of mean \bar{V} into one with realized mean of interest μ^0 (e.g., $\mu^0 = 12.2$, $\mu^0 = 18.8$) by dividing through by \bar{V} and multiplying by μ^0 . The adjusted true particulate concentrations for the DQO simulation days are then

$$U_i = \frac{\mu^0 \mu_{s_i}}{\bar{V}} \quad i = 1, \dots, n$$

5. To allow for sampler measurement error, generate independent normally distributed measurement errors ε_i of mean 1 and precision (standard deviation) σ . Set the observed reading equal to $X_i = \beta U_i \varepsilon_i$ where β is a sampler multiplicative bias. This is the DQO measurement error model with multiplicative bias and multiplicative error.
6. Compute the observed three-year mean \bar{X} as the sample mean of the X_i .

Simulations are run repeatedly N times to generate many three-year sample means $\bar{X}^{(1)}, \dots, \bar{X}^{(N)}$, after which the error rate for a true mean of μ^0 with sampler multiplicative bias β and precision σ is estimated as the proportion of times that simulation sample means are generated on the other side of 15.05 from μ^0 .

The goal here is to simulate the grey-zone for multiple values of μ^0 , β , σ , and α where α is an additive bias for the measurement error model, that is, $X_i = \alpha + \beta U_i \varepsilon_i$. The approach taken is to note that in the simulations the observed daily readings X_i are equal to

$$\begin{aligned} X_i &= \alpha + \frac{\beta \mu^0 \mu_{s_i} e_{s_i} \varepsilon_i}{\bar{V}} \\ &= \alpha + \beta \mu^0 Z_i \end{aligned}$$

where

$$Z_i = \frac{\mu_{s_i} e_{s_i} \varepsilon_i}{\bar{V}}$$

is the random amount, which only depends on the parameter σ . Distributions of

$$\bar{Z} = \sum_i Z_i / n$$

can be simulated for each σ , and their 5th and 95th percentiles, $\bar{Z}_{.05}$ and $\bar{Z}_{.95}$, estimated. The estimated percentiles of the distribution of \bar{X} , $\bar{X}_{.05}$, and $\bar{X}_{.95}$ are then obtained as

$$\bar{X}_{.05} = \alpha + \beta \mu^0 \bar{Z}_{.05} \quad \text{and}$$

$$\bar{X}_{.95} = \alpha + \beta \mu^0 \bar{Z}_{.95}.$$

Running 100,000 simulations of the \bar{Z} process for a range of values of σ gives the results presented in Table 5. As expected, the average of the \bar{Z} is very close to 1, with standard deviation increasing moderately as the precision increases. The estimated $\bar{Z}_{.05}$ is approximately equal to 0.89 over the range of σ used, while the estimated $\bar{Z}_{.95}$ is approximately 1.12.

Table 5. Sample Means, Standard Deviations, and 5th and 95th Percentiles for One Million Simulations of \bar{Z} at Varying Sampler Precisions

Precision (σ)	Sample Mean of \bar{Z}	Sample Standard Deviation of \bar{Z}	Estimated $\bar{Z}_{.05}$	Estimated $\bar{Z}_{.95}$
0.05	1.000	0.070	0.890	1.119
0.06	1.000	0.070	0.890	1.119
0.07	1.000	0.070	0.890	1.119
0.08	1.000	0.070	0.890	1.119
0.09	1.000	0.070	0.890	1.120
0.10	1.000	0.070	0.889	1.120
0.11	1.000	0.071	0.889	1.120
0.12	1.000	0.071	0.889	1.121
0.13	1.000	0.071	0.888	1.121
0.14	1.000	0.071	0.888	1.122
0.15	1.000	0.072	0.887	1.122
0.20	1.000	0.073	0.885	1.125

Thus, it appears that the percentiles can be approximated as

$$\bar{X}_{.05} = \alpha + 0.89\beta\mu^0$$

and

$$\bar{X}_{.95} = \alpha + 1.12\beta\mu^0$$

over all σ of interest. For example, using the current DQO lower bound $\mu^0 = 12.2$, $\alpha = 0$, $\beta = 1.1$, then

$$\bar{X}_{.95} = 1.12(1.1)(12.2) = 15.03,$$

and if using the current DQO upper bound $\mu^0 = 18.8$, $\alpha = 0$, $\beta = 0.9$, then

$$\bar{X}_{.05} = 0.89(0.9)(18.8) = 15.06.$$

The results given in Table 5 suggest that precision is not very important as far as three-year average attainment decisions are concerned. Its bounds might be set for reasons other than DQO attainment. Currently the bound is 0.10, although it could rise to 0.15 or even 0.20 with little effect on the grey zone.

Given that precision might largely be ignored, the results of Table 5 also suggest allowable ranges for α and β when considered jointly. For a fixed grey-zone boundary and fixed percentile value (15.05), we can solve for α in terms of β (or vice versa), for example,

$$\alpha = \bar{X}_{.05} - 0.89\beta\mu^0 = 15.05 - 0.89\beta 18.8 = 15.05 - 16.73\beta$$

$$\alpha = \bar{X}_{.95} - 1.12\beta\mu^0 = 15.05 - 1.12\beta 12.2 = 15.05 - 13.66\beta.$$

The two lines thus described by current grey-zone bounds are plotted in Figure 10. The region between them gives all allowable values of α and β such that the DQO requirements are met. Two notable special cases are if $\alpha = 0$, then $0.9 \leq \beta \leq 1.1$, and if $\beta = 1$, then $-1.7 \leq \alpha \leq 1.4$.

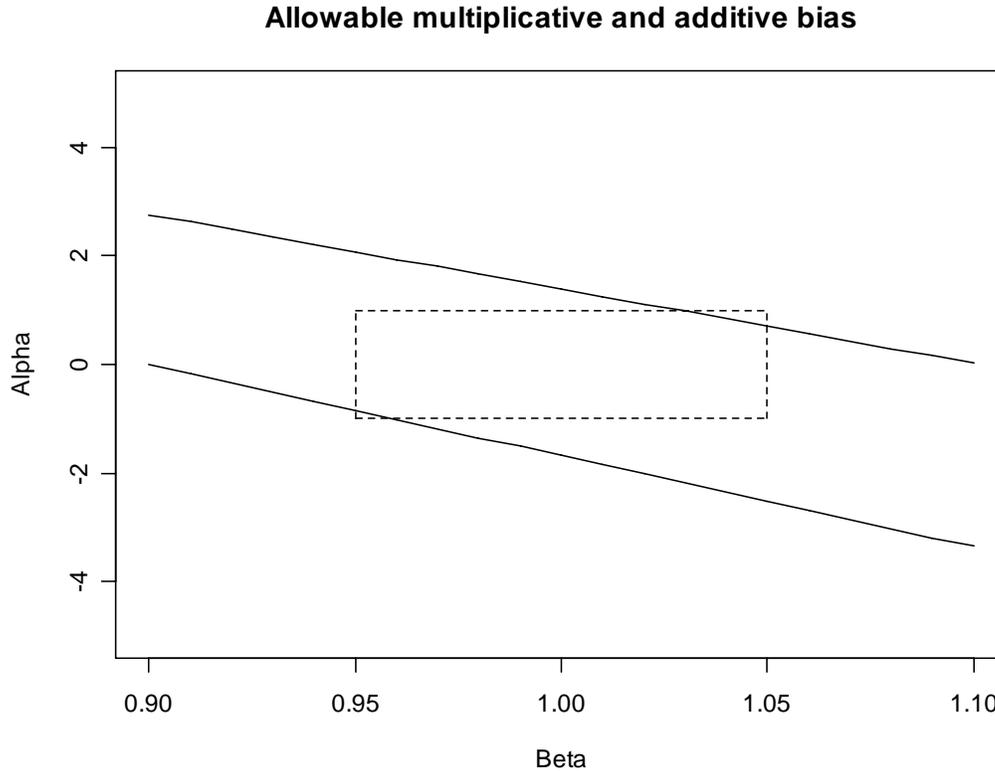


Figure 10. Allowable multiplicative and additive biases according to the DQO model. The upper line is for a true mean of 12.2, and values below it are acceptable. The lower line is for a true mean of 18.8 and values above it are acceptable. The dashed region contains acceptable values under current equivalency tests.

5.2 Bounds for the Correlation Coefficient

Correlation is recommended for inclusion in the equivalency tests for evaluation of model fit. However, it is more difficult to assess than the other parameters since its value depends not only on measurement error variances (σ_R^2 and σ_C^2) but also on moments of the population process, which generates the true daily means. To demonstrate this, recall that for measured values of reference samplers

$$X_{ijk}^* = U_{ik} \varepsilon_{ijk}^*$$

while for candidates

$$X_{ijk} = \alpha + \beta U_{ik} \varepsilon_{ijk}$$

Dropping indices for clarity (and since U is now drawn randomly according to the sampling scheme and population process), consider a pair of reference and candidate measurements X^* , X collected on the same day at the same site, with true value U and errors ε^* , ε , respectively. The covariance of paired reference and candidate measurements is then

$$\begin{aligned}
 \text{Cov}[X^*, X] &= \text{Cov}[U\varepsilon^*, \alpha + \beta U\varepsilon_i] \\
 &= \beta \text{Cov}[U\varepsilon^*, U\varepsilon] \\
 &= \beta(E[U\varepsilon^*U\varepsilon] - E[U\varepsilon^*]E[U\varepsilon]) \\
 &= \beta(E[U^2]E[\varepsilon^*]E[\varepsilon] - E[U]E[\varepsilon^*]E[U]E[\varepsilon]) \\
 &= \beta(E[U^2] - E[U]^2) \\
 &= \beta\sigma_U^2
 \end{aligned}$$

and the variance of the reference measurements equals

$$\begin{aligned}
 \text{Var}[X^*] &= \text{Var}[U\varepsilon^*] \\
 &= E[U^2(\varepsilon^*)^2] - E[U\varepsilon^*]E[U\varepsilon^*] \\
 &= E[U^2]E[(\varepsilon^*)^2] - E[U]^2 \\
 &= E[U^2](\sigma_R^2 + 1) - E[U]^2 \\
 &= \sigma_U^2 + \sigma_R^2 E[U^2]
 \end{aligned}$$

and similarly the variance of the candidate measurements equals

$$\begin{aligned}
 \text{Var}[X] &= \text{Var}[\alpha + \beta U\varepsilon] \\
 &= \beta^2(\sigma_U^2 + \sigma_C^2 E[U^2])
 \end{aligned}$$

so that the correlation of reference and candidate measurements equals

$$\begin{aligned}
 \text{Corr}[X^*, X] &= \frac{\text{Cov}[X^*, X]}{\sqrt{\text{Var}[X^*]\text{Var}[X]}} \\
 &= \frac{\beta\sigma_U^2}{\sqrt{(\sigma_U^2 + \sigma_R^2 E[U^2])\beta^2(\sigma_U^2 + \sigma_C^2 E[U^2])}} \\
 &= \left[\left(1 + \sigma_R^2(1 + \tau_U^{-2})\right) \left(1 + \sigma_C^2(1 + \tau_U^{-2})\right) \right]^{-\frac{1}{2}}
 \end{aligned}$$

where τ_U is the population coefficient of variation, the ratio of the population standard deviation to the population mean

$$\tau_U = \frac{\sigma_U}{\mu_U} .$$

Thus, the correlation only depends on σ_C^2 , σ_R^2 and population process parameter τ_U , and does not depend on either α or β .

It can also be shown that if \bar{X}^* is the mean of J^* reference samplers on a given day and \bar{X} is the mean of J candidate samplers on the same day, then the correlation of the sample means is

$$\text{Corr}[\bar{X}^*, \bar{X}] = \left[\left(1 + \frac{\sigma_R^2}{J^*} (1 + \tau_U^{-2}) \right) \left(1 + \frac{\sigma_C^2}{J} (1 + \tau_U^{-2}) \right) \right]^{-\frac{1}{2}}$$

To illustrate the effect of the population distribution on the correlation, consider the two plots of Figure 11. In the left plot, 100 independent normal random variates are drawn from a distribution of mean 50 and standard deviation 20. These true daily values are then adjusted for measurement error precision of candidate samplers of 0.10 and reference samplers of 0.05. In the right plot, the same process is repeated except that only realizations within the interval [45,55] are accepted for measurement error adjustment. The right plot might represent results if a non-representative sample of yearly days were selected for an equivalency test.

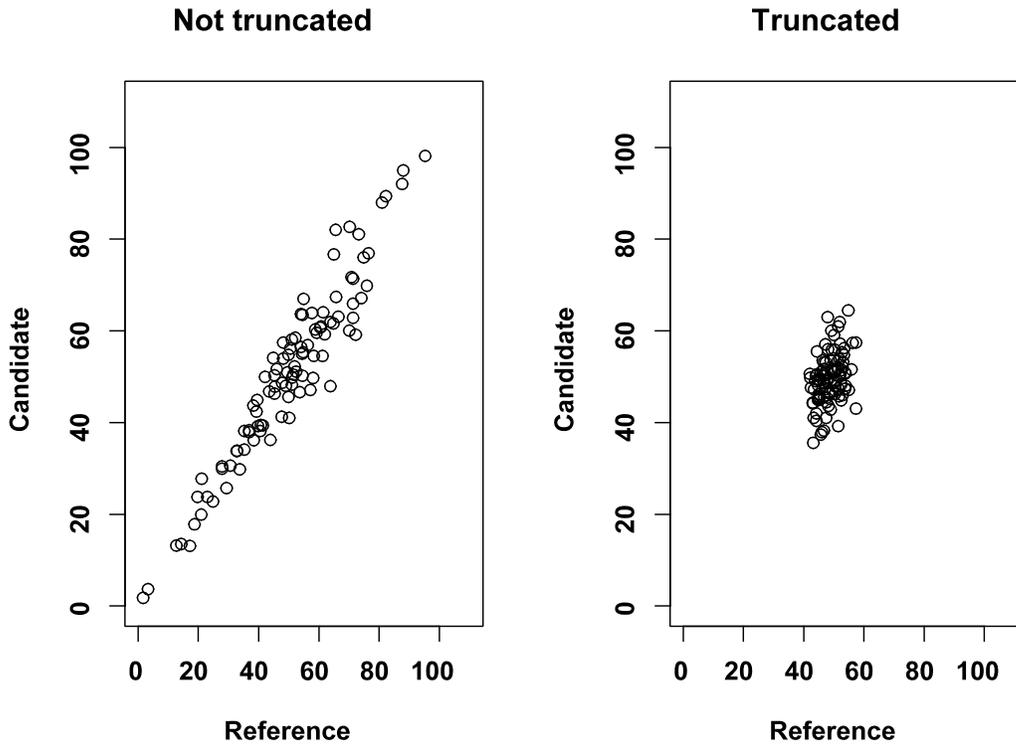


Figure 11. Left plot is of 100 independent normal $N(50,400)$ variates adjusted for measurement error. Right plot is 100 independent normal $N(50,400)$ variates truncated within [45,55] and then adjusted for measurement error.

Calculating the sample correlation in each case gives 0.96 for the left plot, and 0.34 for the right plot. Thus, the sampling scheme selected for daily values can significantly affect the value of the correlation coefficient. The true mean and variance for the untruncated population are 50 and

400, while for the truncated population they are 50 and 8.23. Assuming reference sampler precision of 0.05 and candidate sampler precision of 0.10, the theoretical values of the correlation for each population are therefore 0.96 and 0.38, respectively, based on population values for τ_U of

$$\sqrt{400} / 50 = 0.4 \text{ and } \sqrt{8.23} / 50 = 0.057 .$$

Theoretical correlations for various combinations of the population precision-mean ratio τ_U and the candidate precision σ_C are provided in Table 6. From the table, it can be seen that provided τ_U is about 0.5 or greater the correlation should be 0.97 or greater for candidate sampler precision of 0.10 or less. Correlation increases rapidly as the precision becomes large relative to the mean.

Table 6. Correlations for Single Reference vs. Candidate Samplers for Various τ_U and σ_C .
Reference sampler precision is set equal to 0.05.

τ_U	σ_C											
	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15	0.20
0.01	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01
0.02	0.14	0.12	0.10	0.09	0.08	0.07	0.07	0.06	0.06	0.05	0.05	0.04
0.04	0.39	0.35	0.31	0.28	0.25	0.23	0.21	0.20	0.18	0.17	0.16	0.12
0.10	0.80	0.77	0.73	0.70	0.66	0.63	0.60	0.57	0.54	0.52	0.49	0.40
0.20	0.94	0.93	0.91	0.90	0.88	0.86	0.85	0.83	0.81	0.79	0.77	0.68
0.25	0.96	0.95	0.94	0.93	0.92	0.91	0.89	0.88	0.86	0.85	0.83	0.76
0.33	0.98	0.97	0.96	0.96	0.95	0.94	0.93	0.92	0.91	0.90	0.89	0.83
0.50	0.99	0.98	0.98	0.98	0.97	0.97	0.97	0.96	0.95	0.95	0.94	0.91
1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.96
2.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.97

Table 7 provides correlations calculated in the same way as Table 6 except that they are based on averages of three reference samplers versus three candidate samplers. Similar trends are apparent as in Table 6, although the correlations are somewhat higher due to the decrease in the variance of the sample means.

Table 7. Correlations for Sample Means of Three Reference vs. Three Candidate Samplers for Various τ_U and σ_C . Reference sampler precision is set equal to 0.05.

τ_U	σ_C											
	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15	0.20
0.01	0.11	0.09	0.08	0.07	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.03
0.02	0.32	0.28	0.25	0.23	0.20	0.19	0.17	0.16	0.15	0.14	0.13	0.10
0.04	0.66	0.61	0.57	0.53	0.49	0.46	0.43	0.41	0.38	0.36	0.34	0.27
0.10	0.92	0.91	0.89	0.87	0.85	0.83	0.81	0.79	0.77	0.75	0.72	0.63
0.25	0.99	0.98	0.98	0.98	0.97	0.97	0.96	0.95	0.95	0.94	0.94	0.90
0.33	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.97	0.97	0.96	0.96	0.94
0.50	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.97
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99
2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99

The results of this section show that the correlation between reference and candidate samplers expected if the measurement error models are true depends on both the moments of the population process that generates the true daily values and the precisions of each type of sampler. Methods to estimate the correlation expected if the model were true and also the sample correlation are given later in section 6.3.

6. Parameter Estimation

Methods to estimate the parameters of interest in an equivalency test are now provided. Note that, due to resource limitations in completion of this work, it has not been possible to investigate statistical issues relating to sampling error, such as interval estimation or hypothesis tests.

6.1 Maximum Likelihood Estimation

The measurement error model of section 4 can be fit numerically according to the maximum likelihood approach. Once the model has been specified, the natural logarithm of the corresponding likelihood function can be maximized numerically to obtain maximum likelihood estimates for the parameters of interest.

For the model considered here the likelihood is a product of $(IJK)^2$ normal densities based on the assumptions that the X_{ijk}^* are independent and normally distributed $N(U_{ik}, U_{ik}^2 \sigma_R^2)$, and the X_{ijk} are independent and normally distributed as $N(\alpha + \beta U_{ik}, \beta^2 U_{ik}^2 \sigma_R^2)$.

The challenge in using likelihood methods for this problem is that the number of parameters to be estimated is not only the four of α , β , σ_C , and σ_R , but also the IK parameters for each of the true daily $PM_{2.5}$ values U_{ik} . The dimension of the parameter space over which maximization must be done can therefore be very large, increasing with both days and sites. These additional parameters are known statistically as nuisance parameters and can either be numerically integrated out of the likelihood or maximized in addition to the parameters of interest. The approach taken here will be to maximize the full model, and, remarkably, the optimization procedures used converge for all data sets considered here, up to 111 days of observations (a 115-dimensional space). In comparison, a standard equivalency test should have between 30 and 45 days of observations.

Starting values for the maximization are taken as the day means for the reference samplers, an intercept $\hat{\alpha}_0$ of zero, a slope $\hat{\beta}_0$ of 1, reference precision $\hat{\sigma}_{R0}$ of 0.05, and candidate precision $\hat{\sigma}_{C0}$ of 0.10. Code for performing the maximization in the R computing language is provided in Appendix A.

6.2 Other Estimators

In addition to maximum likelihood estimation of model parameters, the precisions of both reference and candidate samplers can be estimated using simple estimators. These estimators are useful in that they guard against situations where the model is misspecified, possibly with additional sources of error present such as unmeasured variables (as described in section 3.6). They do this by relying only on the candidate or reference data, depending on which precision is being measured. This is more important for candidate samplers than for reference samplers, as the reference sampler model is well supported by historical data and so the likelihood estimate (particularly if restricted to reference data alone) should be reasonable.

In general, the model considered here is of an additive intercept with multiplicative error:

$$X_{ijk} = \alpha + \beta U_{ik} \varepsilon_{ijk}$$

where the ε_{ijk} are independent errors that might be assumed to be normally distributed $\varepsilon_{ijk} \sim N(1, \sigma_C^2)$ with mean 1 and precision σ_C . The goal is to develop a simple estimator for σ_C , which is difficult here since applying a log transform does not in this case make the model additive due to the additive intercept.

The approach taken here will be to note that it is expected that the squares of both precisions will be well less than 1, i.e., $\sigma_R^2 \leq 0.05^2 = .0025 < 1$ and $\sigma_C^2 \leq 0.2^2 = .04 < 1$, and for random variable ε with $E[\varepsilon] = 1$ and $Var[\varepsilon] = \sigma^2$ small, the Taylor expansion of function $h(\varepsilon)$ around $E[\varepsilon]$ is approximately

$$h(\varepsilon) \approx h(E[\varepsilon]) + h'(E[\varepsilon])(\varepsilon - E[\varepsilon])$$

where h' is the first derivative of h , assuming it exists. If $h(\varepsilon_{ijk}) = \log(\alpha + \beta U_{ik} \varepsilon_{ijk})$, then the expansion can be written

$$h(\varepsilon_{ijk}) \approx \log(\alpha + \beta U_{ik}) + \frac{\beta U_{ik}}{\alpha + \beta U_{ik}} (\varepsilon_{ijk} - 1)$$

and so

$$\begin{aligned} Var_{ik}[h(\varepsilon_{ijk})] &\approx \left(\frac{\beta U_{ik}}{\alpha + \beta U_{ik}} \right)^2 Var[\varepsilon_{ijk}] \\ &= \frac{1}{\left(1 + \frac{\alpha}{\beta U_{ik}} \right)^2} \sigma^2 \end{aligned}$$

This result has two consequences:

1. For reference samplers, $\alpha \equiv 0$ and so

$$Var_{ik}[h(\varepsilon_{ijk})] \approx \sigma^2.$$

Thus, the sample variance of the log-transformed reference sampler measurements at a given site on a given day is a closely approximating estimator of σ_R^2 . That is,

$$\tilde{\sigma}_R^2 = \frac{1}{IK} \sum_k S_k^2$$

where S_k^2 is the sample variance of the log reference sampler measurements $\log(X_{ijk}^*)$ at site i at time k . The precision estimator is then $\tilde{\sigma}_R = \sqrt{\tilde{\sigma}_R^2}$.

2. For candidate samplers, $\alpha \neq 0$ in general and so

$$\frac{\alpha}{\beta U_{ik}}$$

is also nonzero. However, as U_{ik} gets large,

$$\frac{\alpha}{\beta U_{ik}}$$

becomes small and so the approximation is closer for days with larger average readings. So an estimator might be

$$\tilde{\sigma}_C^2 = \frac{1}{IK} \sum_k S_k^2$$

where S_k^2 is the sample variance of the log candidate sampler measurements $\log(X_{ijk})$ at site i at time k restricted to days with larger reference sampler readings, perhaps over 20. The candidate sampler precision is then estimated as

$$\tilde{\sigma}_C = \sqrt{\tilde{\sigma}_C^2}$$

An alternative approach might use substitution of α and β from the likelihood estimation, although that is not studied here. Note that, provided α is not too far from zero, the estimator should be reasonably good and some idea as to what value α takes can be obtained from the likelihood estimation.

6.3 Correlation and Its Estimation

In addition to estimating parameters of the measurement error model, we are also interested in estimating the correlation between candidate and reference samplers and comparing the estimated value to a correlation expected if the model were true. As described previously, the estimated correlation coefficient can provide a measure of model fit.

The correlation between reference and candidate samplers can be estimated with data from an equivalency test; however, it is difficult to interpret without knowing what value might be expected were the model correct. This expected value can, however, be estimated if a consistent estimator of $\tau_{U,i}$, the coefficient of variation at site i , can be obtained. It turns out that the reference sampler data allow such an estimator.

Using the approach in section 5.2, it can be shown that for collocated reference samplers j, j'

$$Cov_{i,jj'}[X_{ijk}^*, X_{ij'k}^*] = \sigma_U^2 \quad j \neq j'$$

and so the sample covariance among reference samplers is an estimate of the population process variance. Since the population process mean $\mu_{U,i}$ at site i can be estimated as the sample mean of

the reference sampler values at that site, the population coefficient of variation at site i can therefore be estimated as

$$\begin{aligned}\hat{\tau}_{U,i} &= \frac{\sqrt{\hat{\sigma}_{U,i}^2}}{\hat{\mu}_{U,i}} \\ &= \frac{\sqrt{\frac{2}{J(J-1)} \sum_m \sum_{n \neq m} \hat{Cov}_{i,mn}(x_{imk}^*, x_{ink}^*)}}{\frac{1}{JK} \sum_{jk} x_{ijk}^*}\end{aligned}$$

where $\hat{Cov}_{i,mn}[x_{imk}^*, x_{ink}^*]$ is the sample covariance between reference samplers m and n at site i . Thus, the population coefficient of variation at a given site is estimated as the square root of the average of sample covariances among all pairings of reference samplers at the site divided by the sample mean of the reference sampler readings at that site. Using this estimated coefficient of variation, a target correlation $\hat{\rho}_{0,i}$ for the means of reference and candidate samplers at the site is then estimated as

$$\hat{\rho}_{0,i} = \left[\left(1 + \frac{\sigma_{R,0}^2}{J^*} (1 + \hat{\tau}_{U,i}^{-2}) \right) \left(1 + \frac{\sigma_{C,0}^2}{J} (1 + \hat{\tau}_{U,i}^{-2}) \right) \right]^{\frac{1}{2}}$$

where J^* and J are the number of reference and candidate samplers at the site. In estimating this target correlation, values for $\sigma_{R,0}$ and $\sigma_{C,0}$ can be obtained either as sample estimates or as the largest allowable values, for example, $\sigma_R = 0.05$ and $\sigma_{C,0} = 0.10$. If sample estimates are used, it is recommended that variance estimates based on Taylor approximations, and not on maximum likelihood, be used, since the Taylor approximation estimates are not affected by the types of unmeasured variables described earlier.

To illustrate the method, suppose that 10 days of observations were available for two collocated reference samplers and a candidate sampler for the first, non-truncated population of Figure 11. For this simulated sample given in Table 8, the sample mean among paired reference samplers is 33.5. The sample covariance among reference samplers is 433.4, and the estimated τ_U is therefore 0.62. The target correlation is then found to be 0.980, assuming precisions of 0.05 for reference and 0.10 for candidate samplers. The sample correlation of the mean of the reference measurements versus the candidate measurements is 0.977.

Table 8. Sample Data for Correlation Analysis

Day	FRM1	FRM2	Candidate
1	50.6	52.2	53.5
2	14.6	15.3	16.4
3	41.8	42.3	48.9
4	46.3	41.2	45.5
5	38.4	34.6	33.7
6	31.2	29.4	29.6
7	08.5	07.0	07.3
8	68.1	77.0	64.0
9	03.8	03.7	03.8
10	34.2	28.8	38.3

7. Illustration Using Simulated and Field Data Sets

This section applies the methods of this report to evaluate the performance of reference and candidate samplers for both simulated and field data sets.

7.1 Model Fit Using Simulated Data

A simulated data set was generated according to the design of the proposed equivalency tests. For a single site, three reference and three candidate samplers had measurements simulated for 30 days. These measurements were simulated according to the measurement error model of section 4, with sampler parameter values set as $\alpha = 1$, $\beta = 0.95$, $\sigma_R = 0.05$, and $\sigma_C = 0.10$. True daily values (the U_{1k}) were generated independently from a normal distribution of mean 50 and standard deviation of 25, with any negative values discarded until a total of 30 true daily values were obtained. Figure 12 plots the simulated values for reference and candidate samplers against the true daily values.

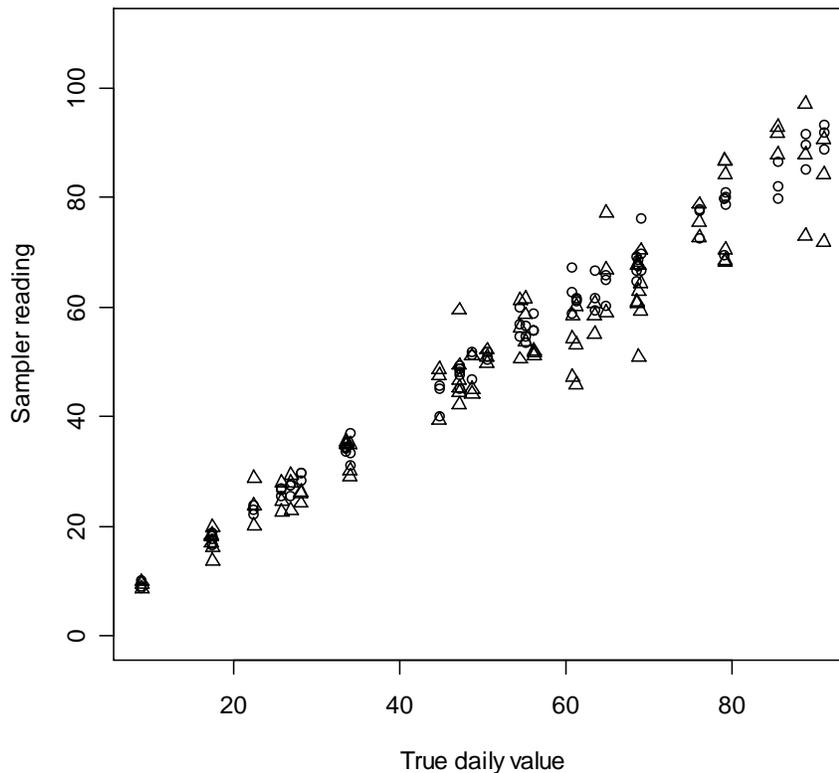


Figure 12. Simulated sampler readings vs. true daily values for reference samplers (circles) and candidate samplers (triangles).

Fitting the model using the likelihood method gives maximum likelihood estimates of $\hat{\alpha} = 0.598$, $\hat{\beta} = 0.955$, $\hat{\sigma}_R = 0.052$, and $\hat{\sigma}_C = 0.097$. Figure 13 plots the estimated daily values against the true daily values. This plot shows estimated mean daily values lying close to a 45-degree line through the origin, as expected.

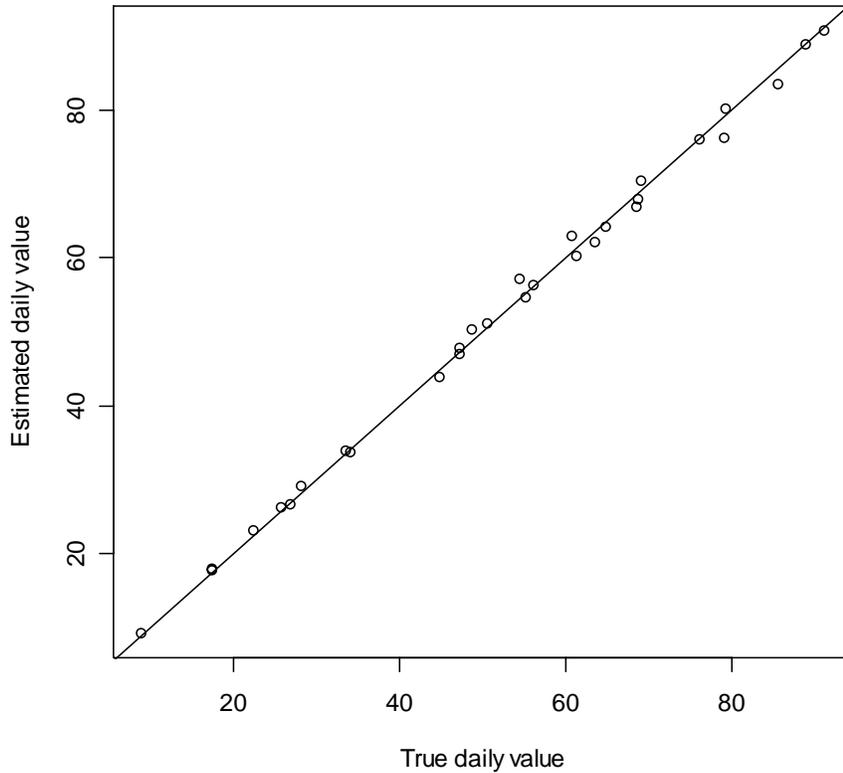


Figure 13. Estimated daily values vs. true daily values. A line of slope 1 through the origin is included for reference.

For comparison, the model was also fit using only the reference sampler data. In this case, the reference sampler precision is estimated as $\hat{\sigma}_R = 0.049$, close to its estimate under the full model using both reference and candidate sampler data.

Using the Taylor approximation estimators $\tilde{\sigma}_R$ and $\tilde{\sigma}_C$ of section 6.3, estimates are $\tilde{\sigma}_R = 0.059$ and $\tilde{\sigma}_C = 0.095$ over all days, or $\tilde{\sigma}_C = 0.097$ if restricting to days with reference means greater than 20. Note that, although these estimates are close to the true values, the likelihood estimates would be preferred in this case as the measurement error model is known to hold exactly for the simulated data.

For correlation, the population coefficient of variation for the 30 observed days is $\tau = 0.41$ and the estimated population coefficient of variation (using only the reference sampler measurements) is $\hat{\tau} = 0.40$. The target correlation of sample means can then be estimated as either $\hat{\rho}_0 = 0.985$ (using the largest allowable $\sigma_{R,0}^2 = 0.05$ and $\sigma_{C,0}^2 = 0.10$), $\hat{\rho}_0 = 0.986$ using the estimated precisions from the likelihood model, or $\hat{\rho}_0 = 0.985$ using the estimated precisions based on Taylor approximation. The sample correlation of the day means of the reference sampler measurements versus the day means of the candidate sampler measurements is $\hat{\rho} = 0.988$.

7.2 Model Fit Using Field Data

To evaluate the method's performance with real data, the estimators described here were applied to the field data of Bakersfield, Detroit, New York, and Tampa. The resulting estimates are given in Table 9.

The likelihood method was applied to obtain estimates of α , β , σ_C , and σ_R . It was also applied to the reference sampler data to obtain an estimate of σ_R based only on that data. Doing so provides some insight into the effect of the assumed model structure on the estimate when candidate sampler data are included.

Taylor approximation estimates are obtained for σ_C and σ_R , with two sets of estimates obtained for σ_C : the first based on the full candidate sampler data, as would be appropriate if it was thought reasonable to assume that $\alpha = 0$ (perhaps supported by a hypothesis test or confidence interval), and the second after truncating the candidate data above a somewhat arbitrary threshold. This second estimate reduces the effect of a non-zero α on the estimate.

Average sample correlations of reference and candidate samplers are obtained and compared to expected values as computed using reference sampler data. These expected values are computed after first estimating τ_U and then obtaining values for σ_C and σ_R . Values for precision are specified either as fixed target values ($\sigma_R = 0.05$ and $\sigma_C = 0.10$) or as Taylor approximation estimates $\tilde{\sigma}_R$ and $\tilde{\sigma}_C$, which are preferred over the maximum likelihood estimates as they are more robust to model misspecification.

Not all statistics could be estimated for all sites since there are fewer replicate samplers than would be expected in an equivalency test. Sites with only one reference sampler do not allow estimation of τ_U or computation of the Taylor variance $\tilde{\sigma}_R$ or of the maximum likelihood estimate $\hat{\sigma}_R$ based only on the reference data. Sites with only one candidate sampler do not allow computation of the Taylor estimator $\tilde{\sigma}_C$, and consequently no estimate of a target correlation $\hat{\rho}_0$ based on estimated precisions. Current plans for equivalency tests call for three replicate candidate and reference samplers at each site, so all statistics could be computed for the resulting data set, if needed.

Table 9. Field Data Estimated Parameter Values after Likelihood Fit and Correlation Analysis

Site	Continuous Sampler Type	Number of Reference Samplers/ Continuous Samplers/ Days Complete Observations	$\tilde{\sigma}_R$	$\hat{\sigma}_R$ FRM Data	$\hat{\sigma}_R$ Full Data	$\tilde{\sigma}_C$ Candidate Data ²	$\tilde{\sigma}_C$ Reduced Candidate Data ²	$\hat{\sigma}_C$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\tau}$	$\hat{\rho}$	$\hat{\rho}_0$ Est ⁴	$\hat{\rho}_0$ Fixed ⁵
Bakersfield	BAM	2/2/52	0.245	0.104	0.123	0.082	0.035 (22) ³	0.115	3.03	1.17	0.709	0.981	0.971	0.993
Bakersfield ¹	BAM	2/2/50	0.051	0.036	0.037	0.080	0.035 (22) ³	0.165	4.09	1.13	0.693	0.982	0.998	0.993
Detroit	TEOM	2/1/31	0.056	0.040	0.045	NA	NA	0.367	1.09	0.71	0.574	0.826	NA	0.991
New York	TEOM	2/1/111	0.041	0.030	0.030	NA	NA	0.181	0.93	1.03	0.676	0.945	NA	0.993
Tampa	DKK	1/2/30	NA	NA	0.002	0.181	0.102 (7) ³	0.242	1.97	1.72	NA	0.905	NA	NA
Tampa	TA	1/2/25	NA	NA	0.002	0.061	0.059 (6) ³	0.193	3.07	1.00	NA	0.936	NA	NA

1. After removal of two FRM sampler outliers as described in section 3.1.

2. Candidate data are all candidate observations. Reduced candidate data are those whose daily reference means exceeds a prespecified threshold.

3. Value in parentheses is number of reference days of mean greater than the threshold. The threshold was set at 20 in Bakersfield and 15 in Tampa.

4. Estimated precisions are based on Taylor approximation estimators $\tilde{\sigma}_R$ and $\tilde{\sigma}_C$ (reduced candidate data).

5. Fixed precisions assume $\sigma_R = 0.05$ and $\sigma_C = 0.10$.

The Bakersfield data set was fit twice, once for all days with complete data (both reference and continuous samplers) and once for a reduced data set that had two apparent FRM outliers removed. Removal of the outliers had the effect of strongly reducing the estimated reference method precision and increasing the target correlation based on estimated precisions. The data thus illustrate the importance of carefully checking for the presence of outliers.

Regarding the estimates of reference sampler precision by both Taylor approximation and maximum likelihood, the estimates are largely consistent with exception of the Bakersfield data with outliers present and the Tampa TA data. The Tampa TA data provide a maximum likelihood estimate of $\hat{\sigma}_R = 0.002$ based on only a single reference sampler and the full data set. This might be a special case where the likelihood favors making the reference samplers extremely precise relative to the candidate data whose variation is more directly measured. For the remaining sites, Taylor and likelihood estimates are similar, with the likelihood being somewhat smaller.

The candidate sampler precision estimates were uniformly larger than those for reference samplers at each site. Furthermore, likelihood estimates tended to be larger than Taylor estimates. That the likelihood estimates are larger might be due to model misspecification, as the model requires that all variation present in the data be absorbed into its parameter estimates. This is good if the model is believed (as it is in the DQO methodology); however, for departures from the model (due, for example, to outside sources of variation such as seasonality), the Taylor estimates may provide a more accurate representation of the precision of candidate measurements at a given site on a given day. The improved precision is due to the Taylor estimates being based solely on the candidate data.

However, Taylor estimates based on the full candidate data were typically larger than those based on the reduced data. This is the opposite of what was expected given the Taylor expansion of section 6.2 (leaving smaller values in should lead to smaller variance estimates). Further analysis (not shown) reveals larger variances of daily log observations for days of small reference sampler means. This suggests an even more general reference and candidate sampler measurement error model than the one given here, namely, one that has both an additive and multiplicative error. Allowing for such a model might be seen as a topic for future work. Regardless, based on this observed effect, we are currently left without a good estimator for candidate precision for use with the correlation tests. The likelihood estimator may be too large due to model misspecification, and the Taylor estimate makes too few model assumptions. Based on the limited field data, it would seem that the full Taylor estimator might be preferred for use with correlation; however, lacking solid justification for its use will suggest using fixed precision values.

Estimates of slope and intercept at all sites reveal that most have candidate slope-intercept combinations that lie outside of ranges given in section 5.1. An exception would be the New York data set.

Sample correlation estimates between day means of reference and candidate samplers were found to lie between 0.826 and 0.981, while population coefficient of variations were estimated between 0.574 and 0.709. The estimated target correlations were all over 0.99 if fixed sampler and reference precisions were used. Excluding outliers from the Bakersfield data and using Taylor precision estimates gave a target correlation of 0.998. Both Detroit and New York show correlations well below their targets, while Bakersfield is somewhat close.

8. Discussion

This report has analyzed field data for sites with both continuous and reference samplers to evaluate and suggest modifications to the DQO/WA 76 methodology for equivalency testing. The recommendation of the report is that additive bias, multiplicative bias, precision, and correlation each be included in the testing methods. Two steps in developing tests have been given: (1) estimators have been provided for each of these parameters, and (2) methods have been given to determine what ranges the parameters they estimate should fall in to consider a candidate sampler equivalent.

The completion of these two steps allows the development of equivalency tests similar in approach to those currently used. The current tests have an experimental design (three samplers of each kind, minimum of 10 days, etc.) and statistics that are computed based on the observed data for comparison to prespecified values. If these statistics fall within allowable bounds, then the sampler is equivalent. In a similar fashion, this report recommends statistics for use in a test and provides methods to determine what values the statistics should take for large samples. Under current proposed methods, these values are directly compared to sample statistics to perform the test.

Although not part of the current equivalency methods, some additional steps could be followed to statistically complete the methodology. These would be either to simulate or find test distributions for hypotheses of interest or to obtain interval estimates for parameters considered here. Interval estimates would be compared to allowable ranges of values to check if overlap existed. Given hypothesis tests or interval estimators, sample sizes for adequately powerful tests might be obtained.

Four parameters are recommended for equivalency testing: candidate sampler precision, candidate sampler additive bias (intercept), candidate sampler multiplicative bias (slope), and correlation of reference sampler and candidate sampler day means. It is recommended that the first three of these be estimated according to maximum likelihood and the correlation be estimated as it is currently.

Maximum likelihood methods offer the advantage that they more closely model the assumed error structure in the data. As described in WA 76, the implied model for the current equivalency tests least-squares fit is simple linear regression, which assumes that predictors are measured without error—an assumption that does not hold for these data since there is error in the reference measurements. Having error in the predictors for a regression leads to biased parameter estimators with the slope being “attenuated” to be flatter than it should be and the intercept estimate biased accordingly. This may be problematic for the estimate of the intercept, since it is estimated as a value along the regression line lying outside the range of the data. It is also required to lie within a relatively tight range, as seen in section 5.1. The other assumption of simple linear regression, which does not hold here, is that the errors are of constant variance along the length of the regression curve. The main effect of this difference is that various tests and interval estimates based on simple linear regression assumptions are not appropriate.

Likelihood methods also allow for future extensions to the methods to obtain interval estimates and hypothesis tests. This can be done according to well-established theory.

The primary disadvantage of likelihood methods is that they are somewhat more complicated to implement and require greater sophistication on the part of the modeler. Code is provided in the appendix that should allow their estimates to be obtained, and this should be evaluated for ease of use.

Each of the four tested parameters is now discussed in turn, with general recommendations for their incorporation into the equivalency test design.

The additive bias or intercept parameter was introduced into the measurement error model due to evidence of a non-zero intercept in the Bakersfield data. It is tested in current equivalency tests using the simple linear regression approach. As was shown in section 3.6, it can represent either a property of the sampler or a property of the particular site, so equivalency tests at multiple sites should be required to determine if estimated intercepts are site-specific or not. The recommended approach to its estimation is the maximum likelihood approach. Like the slope, the precision of its estimate improves with daily measurements that are well dispersed.

The multiplicative bias or slope parameter currently exists in both current equivalency methods and the DQO methods. Like the intercept, it can be estimated using the maximum likelihood approach. Also like the intercept, the estimate benefits from having reference and sample measurements that are well dispersed.

The precision parameter describes the variation in sampler measurement error. It can be estimated according to either maximum likelihood or Taylor approximation, with the likelihood approach being preferred. Neither of the estimators is entirely satisfactory, however, as the assumed model has been shown to not hold exactly (section 7.2), with errors being not strictly multiplicative. With that said, the resulting estimates should be close enough and make similar assumptions to current methods used to estimate precision with field data. Furthermore, precision is the least important of the parameters as far as DQO equivalency is concerned, and for moderate values it has little effect on the grey zone (section 5.1). As the Bakersfield data illustrated, its estimation is sensitive to outliers, so care must be taken to identify and remove those from the final data.

The recommendation here is to estimate the correlation following the current approach—that is, as the correlation of reference and candidate day means. This report has shown how its target value for a given sample can be estimated (section 5.2), and this target depends on the reference and candidate sampler precisions as well as the coefficient of variation of the underlying population $PM_{2.5}$ process. In estimating the target value, there is a choice of whether to substitute fixed worst-case precisions or to use the precision sample estimates. Both approaches were evaluated using the field data, where for the estimated population coefficient of variations, all led to target correlations of .99 or more (discarding outliers). The estimated target might then be adjusted in an ad hoc manner (say, by subtracting 0.01 or 0.02) to allow for some sampling error. Further study could provide more rigorous methods.

9. References

U.S. EPA. 2002. DQO Companion, Version 1.0 Users's Guide. EPA Work Assignment 5-07.

U.S. EPA. 2003. Data Quality Objectives for PM Continuous Methods, TR-4423-03-08. Research Triangle Park, NC: ManTech Environmental Technology, Inc.

Appendix A: R Code for Maximum Likelihood Estimation

This appendix provides the software code needed to estimate the general measurement error parameters α , β , σ_C , and σ_R using the method of maximum likelihood.

The code is in the R language, a high-quality, freely available data analysis package based on the S language. The software can be found at www.r-project.org for most major operating systems. The commercially available S-Plus is also based on S, and the code given here should run in S-Plus with only minor modification, if any. The code is run by starting the R program and then by cutting and pasting code (after relevant modifications) into the R command window.

```
### Load field data
filepath <- "M:\\WA97\\Data\\" # MS-Windows directory
filename <- "Bakersfield-2002.csv" # Filename
# File is a comma delimited text file,
# with separate columns of data for each sampler,
# variable labels are at top of each column
# in this sample code, the six variables have
# labels: bkF1,bkF2,bkF3,bkC4,bkC5,bkC6

dat <- read.table(paste(filepath,filename,sep=""),sep=",",header=T)
J <- 3 # Number of samplers of each type
K <- 30 # Number of days measurement
# restrict to only days of complete data
x <- dat$bkF1+dat$bkF2+dat$bkF3+dat$bkC4+dat$bkC5+dat$bkC6
x <- !is.na(x)
dat <- dat[x,]
xref <- rep(0,J*K)
xcnd <- rep(0,J*K)
for(i in 1:K){
  xref[(i-1)*J+1] <- dat$bkF1[i]
  xref[(i-1)*J+2] <- dat$bkF2[i]
  xref[(i-1)*J+3] <- dat$bkF3[i]
  xcnd[(i-1)*J+1] <- dat$bkC4[i]
  xcnd[(i-1)*J+2] <- dat$bkC5[i]
  xcnd[(i-1)*J+3] <- dat$bkC6[i]
}
u0 <- (dat$bkF1+dat$bkF2+dat$bkF3)/3

## Code for estimation using just reference sampler data
sdr0 <- .05
initp <- c(sdr0,u0)
# index to allow matching each day with correct mean
uidx <- rep((1+1):(K+1),rep(J,K))
f <- function(p){
  rmeans <- p[uidx]
  return(-J*K*log(p[1])-sum(log(rmeans)
    )-.5*sum(((xref-rmeans)/(p[1]*rmeans))^2))
}
```

```
# First value returned is estimated reference precision
# Remaining values are day means
optim(initp,f,control=list(fnscale=-1,maxit=100000))

## Estimation both reference and candidate samplers
a0 <- 0
b0 <- 1
sc0 <- .1
sr0 <- .05

initp <- c(sr0,sc0,a0,b0,u0)
uidx <- rep((1+4):(K+4),rep(J,K))

f <- function(p){
  rmeans <- p[uidx]
  return(-J*K*log(p[1])-sum(log(rmeans))- .5*sum(((xref-rmeans)/(p[1]*rmeans))^2)
        -J*K*log(p[2])-sum(log(rmeans))-J*K*log(p[4])
        - .5*sum(((xcnd-(p[3]+p[4]*rmeans))/(p[2]*p[4]*rmeans))^2))
}

# First value returned is estimated reference precision
# Second value is estimated candidate precision
# Third value is additive bias (intercept)
# Fourth value is multiplicative bias (slope)
# Remaining values are estimated day means
optim(initp,f,control=list(fnscale=-1,maxit=100000))
```

Appendix B: Precision and Variance

In this report, precision is used synonymously with standard deviation for the multiplicative model. This usage has developed through their equivalence for models considered by the DQO methodology and for models of WA 76. With the addition of an additive bias (intercept) to the model, the relationship is no longer exact, although the usage has been retained for consistency. These relationships are described in more detail below.

For measurements X_1, \dots, X_n of mean μ , variance σ^2 , the population coefficient of variation (precision) is defined as

$$CV = \frac{\sqrt{\sigma^2}}{\mu} = \frac{\sigma}{\mu}$$

which is often estimated as

$$\hat{CV} = \frac{S}{\bar{X}}$$

where S is the sample standard deviation and \bar{X} is the sample mean. Note that the CV is unitless since μ and σ have the same units.

For daily measurements X_1, \dots, X_n of daily true value U (realization of population process), we assume measurement error model

1. $X_k = U \varepsilon_k$, or
2. $X_k = \alpha + \beta U \varepsilon_k$

In the case of 1, the population mean (for that day) is U and the variance is $U^2 \sigma^2$, so the coefficient of variation is

$$CV = \frac{\sqrt{U^2 \sigma^2}}{U} = \frac{U \sigma}{U} = \sigma$$

and so the CV, which is unitless, happens to equal σ , which has units.

In the case of 2, the population mean is $\alpha + \beta U$ and the variance is $U^2 \sigma^2$, so the coefficient of variation is

$$CV = \frac{\sqrt{\beta^2 U^2 \sigma^2}}{\alpha + \beta U} = \frac{\sigma}{1 + \frac{\alpha}{\beta U}}$$

which equals σ if $\alpha = 0$. Therefore, when the intercept is non-zero, the standard deviation does not exactly equal the coefficient of variation (although if $\alpha \neq 0$ then the CV might not be what one would want to measure, since it then depends on U leading to a different CV for each possible true daily value. In such a situation, σ is probably more useful).