Public Reporting of an Air Quality Index Using Continuous PM$_{2.5}$ Monitoring Data

Steven M. Bortnick and Basil W. Coutant
Battelle Memorial Institute, Columbus, OH

Timothy Hanley
U.S. Environmental Protection Agency, Research Triangle Park, NC

INTRODUCTION
The air quality index (AQI) is a single index ranging from 0 to 500 for reporting air quality to the general public. It is a national index, meaning that the values and associated level of health concern are the same throughout the United States. The AQI is based on ambient air concentrations of ozone, fine particulate matter (PM$_{2.5}$), coarse and fine particulate matter (PM$_{10}$), carbon monoxide, sulfur dioxide, and nitrogen dioxide. It measures how clean or polluted the air is in terms of an associated health risk within a few hours or days after breathing polluted air. The AQI scale is designed to transform the different ambient pollutant concentrations into ranges with similar associated health risks, from “Good” to “Hazardous.”

Regarding PM$_{2.5}$, the AQI scale was developed using measurements that correspond to federal reference method (FRM) monitors. However, statistical linear regression models can be applied to relate non-FRM PM$_{2.5}$ measurements, such as continuous monitoring (CM) measurements, with FRM measurements for the purpose of reporting the AQI. The non-FRM CM measurements can then be transformed via the model to remove any bias relative to FRM measurements.

Several studies have investigated relationships between measurements from different types of PM samplers and one recent study compares measurements between FRM PM$_{2.5}$ monitors and non-FRM monitors. These studies indicate varying degrees of success depending on the continuous monitor used, the concentration and species of the aerosol, and effects of meteorological conditions such as ambient temperature and relative humidity. The conclusion about whether a comparison is good or bad is typically based on a qualitative assessment of the slope of the regression, the intercept of the regression, and the correlation coefficient. People may draw different conclusions about whether a comparison is good or bad because the decision is subjective. Of considerable importance is the quality of the model used to relate the CM and FRM measurements. The use of a poor model could result in misleading AQI reporting in the form of incorrectly claiming either better or worse air quality.

This article describes a quantitative measure of adequacy for statistical linear regression models relating FRM and CM PM$_{2.5}$ measurements for the purpose of reporting the AQI. To guide appropriate model development of the relationship between FRM and CM measurements, we make use of the data quality objectives (DQO) process, a seven-step strategic planning approach, based on the scientific method, for determining the appropriate data type, quality, quantity, and synthesis. The decision about whether a relationship developed between an FRM and a CM is...
adequate depends on the number of paired samples, the correlation, and the size of the errors considered acceptable.

Next, this article provides two case studies in which models that meet the specified level of adequacy are developed. The case studies include the metropolitan statistical areas (MSAs) of Greensboro/Winston-Salem/High Point, NC, henceforth referred to as NC; and Davenport/Moline/Rock Island, IA/IL, henceforth referred to as IA/IL. In both case studies, the FRM was a Partisol-Plus 2025 sequential sampler, manufactured by Rupprecht & Patashnick Co., and the non-FRM was a Tapered Element Oscillating Microbalances (TEOM) with a size selective inlet to remove particles larger than 2.5 microns (µm). The TEOM in NC was continuously operated at 50 °C, and the TEOM in IA/IL was operated at 50 °C during the summer and 30 °C during the winter.

POLICY
The current PM$_{2.5}$ state and local agency air monitoring network is comprised of filter-based FRM samplers. The definition and operation of PM$_{2.5}$ FRMs is explicitly defined in 40 CFR Part 50 Appendix L (1997). These samplers provide measurements that are consistent and repeatable both temporally and geographically. However, due to the need for post-sampling laboratory analyses, the measurements are not provided in a timely manner for use in AQI reporting. Weeks can pass before the concentrations for a given day are reported. In contrast, non-FRM monitoring methods, such as CM methods, can provide for timelier reporting of PM$_{2.5}$ data. Timeliness is important in order for agencies to advise the public about the air they are breathing so that they can adjust their activities and therefore their exposure. Hence, 40 CFR 58 (2000) requires all MSAs with a population of at least 350,000 to report daily air quality to the general public. Furthermore, 40 CFR Part 58 Appendix D (2000) specifically requires “continuous fine particulate monitoring” in each metropolitan area with a population of at least one million.

According to 40 CFR 58 Appendix G (2000), particulate matter measurements from non-FRM or non-federal equivalent method (FEM) monitors may be used for the purpose of reporting the AQI. Example non-FRM/FEM particulate matter methods include TEOM, nephelometers, beta attenuation monitors (BAMs), and continuous ambient mass monitors (CAMMs). The primary advantage of these non-FRM/FEM methods is their timely, time-resolved, cost-efficient measurements. However, owing to the different method of measuring particulate matter, their concentrations may differ from those of an FRM/FEM. For example, typical operating procedures for the TEOM involve heating the inlet to 30–50 °C to reduce the amount of particle-bound water. Such heating may bias the measurement when there is volatile particulate matter present. Due to these potentially different estimates of particulate matter, 40 CFR 58 Appendix G (2000) requires that non-FRM/FEM measurements be used for AQI reporting only once a statistical linear relationship has been established with an FRM/FEM.

PLANNING
The U.S. Environmental Protection Agency's (EPA) Order 5360.1 CHG 1 provides agency requirements for the conduct of quality management practices, including quality assurance (QA) and quality control (QC), for all environmental data collection and environmental technology programs performed by or for EPA. The primary goal of the agency-wide quality system is to ensure that environmental programs and decisions are supported by data of the type and quality needed and expected for their intended use, and that decisions involving the design, construction, and operation of environmental

![Figure 1. Scatterplots of FRM vs. CM measurements in NC and IA/IL, on the natural log scale.](image1)

![Figure 2. Time series of [ln(FRM) – ln(CM)] for NC (top) and IA/IL (bottom).](image2)
technology are supported by appropriate quality-assured engineering standards and practices.

EPA’s policy for the quality system is established in EPA Order 5360.1 CHG 1 with implementation requirements given in the EPA Quality Manual. This manual also incorporates the specifications and guidelines of ANSI/ASQC E4 and contains elements from the International Standards Organization (ISO) 9000 series of quality management standards. The quality system is comprised of three structural levels: policy, organization/program, and project. This structure promotes consistency among quality systems at the higher management levels, while allowing project managers the flexibility necessary to adapt EPA’s quality system components to meet the individual, and often unique, needs of their work.

The third level, the project level, comprises components that are applied to projects within an organization that require data collection or generation. There are three integrated phases within the project level: planning, implementation, and assessment. The planning stage begins with a requirement that a systematic planning process, based on the scientific method, must be used to develop acceptance or performance criteria for the collection, evaluation, or use of environmental data. In particular, the DQO process is a planning tool to clarify and document the specific use of environmental data for making decisions and to ensure efficiency in collecting data that meets the needs of decision-makers. It divides the process into seven steps. Each step and the activities associated with them are considered in the sections that follow. The process, however, is not entirely linear. Some iteration between steps is often required, because the results of the later steps may impact earlier ones. Ideally, the final three steps of the DQO process should be done in parallel.

**State the Problem (Step 1)**
The purpose of this step is to concisely define the problem at hand. **Problem statement:** It is desired to use continuous PM$_{2.5}$ measurements for the purpose of reporting an AQI in, for example, NC or IA/IL. According to 40 CFR 58 Appendix G (2000), these data may be used for this purpose if a linear relationship between continuous measurements and reference or equivalent PM$_{2.5}$ method measurements can be established by statistical linear regression. Therefore, a model relating FRM and continuous PM$_{2.5}$ measurements, possibly adjusting for meteorological data, is required.

**Identify the Decision (Step 2)**
The purpose of this step is to clearly define the decision statement the study will attempt to resolve. **Decision statement:** Is the statistical linear model relating FRM PM$_{2.5}$ to continuous PM$_{2.5}$ measurements acceptable for transforming continuous measurements for the purpose of reporting, for example, the NC or IA/IL AQI? If so, then the continuous PM$_{2.5}$ data, along with the model, can be used to report the AQI. If not, then data should not be used to report the AQI.

**Identify the Inputs to the Decision (Step 3)**
The purpose of this step is to identify the informational inputs needed to resolve the decision statement and determine the inputs that require environmental measurements. The list of environmental measurements required for this study include FRM PM$_{2.5}$ daily measurements, continuous PM$_{2.5}$ hourly measurements, and, possibly, meteorological data such as temperature. At the most basic level, an MSA will require a set of days for which both FRM PM$_{2.5}$ and CM PM$_{2.5}$ measurements have been obtained from sites within the MSA. Such information is vital to developing a model relating the two measures. Ideally, (1) a large number of days will be available, including data spanning at least one year, (2) at least some of the FRM and CM data will be colocated, and (3) meteorological data will be available for model improvement, if necessary. For example, both NC and IA/IL have monitoring sites producing colocated FRM and CM measurements. In many cases, these data will be available in EPA’s aerometric information retrieval system (AIRS) database. In some cases, data will be accessible from an MSA’s own data archive.
Define the Boundaries of the Study (Step 4)
The purpose of this step is to define the spatial and temporal boundaries covered by the decision statement. The population of interest is daily FRM/FEM PM$_{2.5}$ concentrations for the MSA, measured in $\mu$g/m$^3$. The MSA is the geographical area within which the decision statement is to be applied. Data permitting, some MSAs might develop models specific to subregions within the MSA; hence, the spatial scale of decision-making could be anywhere from an MSA subregion surrounding the site(s) used to develop the model up to the entire MSA itself.

The time frame to which the decision applies will be up to individual MSA’s decision-makers. It is recommended that an acceptable model should be checked for accuracy and updated (if necessary) annually, if not quarterly or even more frequently. Hence, the temporal bounds of decision-making might range from a few days, if a model is continuously updated or replaced, to one year, if the MSA’s decision-makers feel the model is still accurate one year after development.

Develop a Decision Rule (Step 5)
The purpose of this step is to define the parameter of interest that quantifies the decision, specify the action level, and integrate previous DQO outputs into a single statement that describes a logical basis for choosing among alternative actions. Since the overall purpose of this exercise is to develop an acceptable model that relates FRM and CM measurements, it was determined that the parameter of interest is R$^2$. In general, R$^2$ measures the strength of the model fit to the data. In this case, R$^2$ is interpreted as the correlation between measured and modeled FRM data.

It is important to keep in mind that in the current context, a decision is to be made based on estimating the model’s true R$^2$ value, a rather uncommon activity in practice. In most applied contexts, the sample statistic $r^2$ obtained from software regression output is interpreted as the true R$^2$ value, when in fact it is only an estimate of the true unknown value. Under a hypothesis testing scenario, accepting or rejecting a model based on a true R$^2$ action level of 0.60 will be seen to be equivalent to requiring a sample $r^2$ value equal to 0.80, a model adequacy threshold common to many applied data analysts. Hence, the action level around which a model will be deemed acceptable was determined to be the value of R$^2$ equal to 0.60. “If...then...” statement: If the true R$^2$ value of the statistical linear regression model relating FRM and continuous PM$_{2.5}$ measurements within, for example, NC or IA/IL, is greater than 0.60, then continuous PM$_{2.5}$ data can be used, along with the model, to report the AQI for the next 90-day to one-year period. Otherwise, the model in its current form is not acceptable and should not be used for AQI reporting.

Specify Tolerable Limits on Decision Errors (Step 6)
The purpose of this step is to specify tolerable limits on decision errors. The decision as to whether the model is acceptable may be statistically formalized as the following hypothesis test:

$$H_0: R^2 \leq 0.60 \text{ vs. } H_a: R^2 > 0.60 \quad (1)$$

where R (and, hence, R$^2$) can theoretically range from 0.0 (i.e., no relation between actual and modeled FRM PM$_{2.5}$ measurements)
to 1.0 (i.e., perfect correlation between actual and modeled FRM PM\(_{2.5}\) measurements). Along with the above hypothesis statement, three additional parameters must be specified to formally accept or reject the model; namely, the false rejection decision error rate (\(\alpha\)), the false acceptance decision error rate (\(\beta\)), and the size of the “gray region” in decision-making (\(\Delta\)). The false rejection decision error rate (\(\alpha\)) specifies the maximum probability of claiming the model is adequate (\(R^2 > 0.60\)) when, in fact, it is not (\(R^2 \leq 0.60\)). The false acceptance decision error rate (\(\beta\)) specifies the maximum probability of claiming the model is not adequate (\(R^2 < 0.60\)) when, in fact, it is (\(R^2 > 0.60\)). The size of the gray region in decision-making (\(\Delta\)) specifies an area, starting at \(R^2 = 0.60\) up to \(R^2 = (0.60 + \Delta)\), within which somewhat higher false acceptance decision error rates (\(\beta\)) are considered tolerable. For the purposes of this article, values of \(\alpha = 0.05\), \(\beta = 0.30\), and \(\Delta = 0.25\) will be used.

**Optimize the Design for Obtaining Data (Step 7)**

The purpose of this step is to identify a resource-effective data collection design for generating data that are expected to satisfy the DQOs. In the current context, the nature of the data collection design is predominantly determined by the existing infrastructure an MSA, such as NC or IA/IL, has in place to develop a model. How many FRM or CM monitoring station sites have been collecting data? Are there colocated FRM and CM measurements, as in the cases of NC and IA/IL? Which sites should be used in developing a model (e.g., colocated core FRM and CM sites only)? Assuming these questions are answered, the major design issue remaining is the minimum number of days required for adequate model development (i.e., the sample size requirement). Along the way to determining the sample size requirement, the practical rule for deciding whether the model is “good enough” is developed (i.e., the model adequacy requirement). The full statistical rationale for determining the sample size and model adequacy requirements is provided by Bortnick, Eberly, and Coutant.\(^{13}\)

The resulting sample size (\(n\)) requirement may be summarized as follows:

\[
 n \geq \left[ \frac{z_\alpha - z_{1-\beta}}{\frac{1}{2} \ln \left( \frac{1 + \sqrt{0.60 + \Delta}}{1 - \sqrt{0.60 + \Delta}} \right) - 1.0317} \right]^4 + 3 \tag{2}
\]

where \(z_\alpha\) is the \(\alpha\)th percentile of the standard normal statistical distribution, and \(z_{1-\beta}\) is the \((1-\beta)\)th percentile of the standard normal statistical distribution. Likewise, the resulting model adequacy (\(r^2\)) requirement may be summarized as follows:

\[
 r^2 \geq \frac{\exp(2c) - 1}{\exp(2c) + 1} \tag{3}
\]

RESULTS

To relate CM data with FRM data, hourly CM measurements were averaged to produce a daily value. Consistent with 40 CFR Part 50 Appendix H (2000), a criterion of 75% completeness, 18 out of 24 hr, was used to determine valid CM daily averages (specified in DQO Step 3).\(^{10}\) The data were all well above the method detection limit. For the purposes of this article, we present only the results for the colocated CM and FRM data of NC and IA/IL. Figure 1 displays scatterplots of FRM measurements versus daily average CM measurements for NC and IA/IL, with a 1-1 line overlaid. A bias between the two measurements will show up as a clustering of the data away from the 1-1 line, a phenomenon slightly more noticeable for the IA/IL data.

**Model Development**

In both cases, NC and IA/IL, the data were first transformed via the natural log function to better satisfy the normality assumption of DQO Step 7. The correlation between FRM and
CM data was very strong in NC, requiring only a simple log-linear regression model \(\ln(\text{FRM}) = \beta_0 + [\beta_1 x \ln(\text{CM})] + \text{residual error}\) to satisfy the model quality requirements of DQO Step 7. The IA/IL data, however, required an additional temperature term in the log-linear model \(\ln(\text{FRM}) = \beta_0 + [\beta_1 x \ln(\text{CM})] + [\beta_2 x \text{temp.}] + \text{residual error}\) to account for seasonal differences between FRM and CM measurements.

Specifically, the simple log-linear regression model in IA/IL yielded an \(r^2\) of 0.69, whereas adding daily average temperature to the model improved \(r^2\) to 0.86. The seasonal differences observed in IA/IL may have been due to the fact that the TEOM in IA/IL was operated at 50 °C during the summer and 30 °C during the winter. One model adjustment that was considered to account for this observed effect in the data was to include an indicator, or mean shift variable, for season (i.e., mode of operation). However, including temperature as a continuous covariate proved to be a superior model adjustment. Figure 2 demonstrates why a slightly more complicated model was required in IA/IL. For both the NC and IA/IL data, the figure plots the time series of the difference between FRM and CM measurements on the natural log scale, with a smooth trend overlaid. (The smooth trend was estimated with the loess function in S-Plus.)

The NC time series of FRM-CM differences exhibits no clear seasonal pattern, whereas the IA/IL differences oscillate between positive and negative values depending on season. A plausible explanation for this seasonal pattern is the effect of temperature on one or both measuring devices. Hence, a temperature term was added to the IA/IL model to adjust for the seasonally dependent differences between FRM and CM measurements in this case.

Table 1 provides regression summary statistics for the final models developed in the two case studies. Notice the very strong relationship, \(r^2 = 0.96\), between FRM and CM daily values in NC. In both cases, NC and IA/IL, the sample size \((n \geq 213)\) and model adequacy \((r^2 \geq 0.81)\) results were considered sufficient evidence, according to DQO Step 7, that the models are appropriate for use in reporting the AQI.

### AQI Reporting

Once an acceptable model is obtained, it may be used along with CM measurements to provide timelier AQI reporting. Since the model relates daily averages, using it to report a rolling 24-hr average AQI would appear more appropriate than reporting the most recent hour’s worth of data. While a rolling 24-hr average includes 24 hr of past data, it remains a more timely reporting mechanism because it includes the most recent hour’s worth of data. This fact is highlighted in Figures 3 and 4, which display two weeks of AQI reporting using once daily FRM measurements versus hourly CM measurements averaged over the most recent 24-hr period. For comparison, the figures display AQI values calculated from both transformed (i.e., model adjusted) and untransformed CM measurements.

The AQI values reported in Figures 3 and 4 were calculated from NC data and IA/IL data, respectively, during the two-week period of February 13, 2000, to February 27, 2000. The model adjustment for the IA/IL data is clearly more significant than that for the NC data, indicating a stronger bias between the CM and FRM measurements at the IA/IL monitoring site relative to the NC site. Consider the 15 midnight-to-midnight averaging periods during this two-week period and compare the AQI classification (0–50 = Good, 51–100 = Moderate) using FRM measurements to that using either raw CM (untransformed) or model-transformed CM (modeled FRM) measurements. In NC, the resulting AQI classifications match in 14 out of 15 cases, regardless of whether raw or model-transformed CM measurements are used. In IA/IL, one of the 15 days is missing the FRM measurement, thus there are only 14 periods. The resulting AQI classifications match in 12 out of 14 cases when model-transformed CM measurements are used, but only eight out of 14 cases when raw CM measurements are used.

### CONCLUSIONS

EPA’s DQO process was employed to develop a measure of adequacy for deciding whether a statistical linear regression model that relates FRM and continuous \(\text{PM}_{2.5}\) measurements is sufficient for use in AQI reporting. Next, MSA-specific models were developed for NC and IA/IL that relate FRM \(\text{PM}_{2.5}\) measurements with CM measurements. Note that the DQO steps are not specific to an MSA, but the model development is. For this article, the purpose of the models is to allow for the use of CM measurements in providing timelier reporting of an AQI. However, other applications requiring daily FRM-like \(\text{PM}_{2.5}\) measurements can benefit from these models. Since the models were based on recent data specific to the time of their development, their validity should be checked on a periodic basis. How often an MSA checks and updates its model will depend on how varied the monitors in its area tend to be.

### Table 1. Regression summary statistics for final models in NC and IA/IL.

<table>
<thead>
<tr>
<th>MSA</th>
<th>Days of Data: N</th>
<th>Intercept: (\beta_0) (s.e.)</th>
<th>(\ln(\text{CM})) Effect: (\beta_1) (s.e.)</th>
<th>Temp. Effect: (\beta_2) (s.e.)</th>
<th>RMSE</th>
<th>(r^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>227</td>
<td>-0.114 (0.036)</td>
<td>1.054 (0.013)</td>
<td>NA</td>
<td>0.104</td>
<td>0.96</td>
</tr>
<tr>
<td>IA/IL</td>
<td>214</td>
<td>-0.269 (0.078)</td>
<td>1.280 (0.036)</td>
<td>-0.027 (0.002)</td>
<td>0.225</td>
<td>0.86</td>
</tr>
</tbody>
</table>
It will probably take at least one quarter’s worth of data to make any significant change, unless there are changes to the operating procedures of the CM.

ACKNOWLEDGMENTS
This research was funded in part by EPA under Contract No. 68-D-98-030. It has not been subjected to EPA peer review and may not necessarily reflect the views of the agency. No official endorsement should be inferred. We wish to thank G. Johannesson for invaluable data management, summary, and analysis support. We also would like to thank S. Eberly, L. Weinstock, T. Tamanini, E. Michel, G. Denniston, and B. Lambeth for their comments and contributions in developing the models and in understanding the practicalities of AQI reporting.

REFERENCES

About the Authors
Steve Bortnick received his master’s degree in statistics in 1996 and obtained a Ph.D. in statistics in 1999, both from Ohio State University. He is currently a principal research scientist with Battelle Memorial Institute, Columbus, OH. Basil Coutant received his Ph.D. in mathematics from Ohio University in 1990 and a master’s in statistics from North Carolina State University in 2000. He is currently a research scientist with Battelle Memorial Institute. Timothy Hanley received his master’s degree in environmental management in 1997 from Duke University. He is currently an environmental scientist with EPA.