25th ANNUAL NATIONAL CONFERENCE ON MANAGING ENVIRONMENTAL QUALITY SYSTEMS

APRIL 24-27, 2006

Marriott Renaissance, Austin, Texas

Technical Papers

Methodological Advances in Quality Measurement and Assessment

- P.Lacayo, Confidence Intervals Revisited - 1:00 PM
- S.Chen, Monitoring the Long-Memory Air Quality Data Using ARFIMA Model - 1:30 PM
- L.Camalier, Combining Classical Analyses with Practicality: A Different Look at the STN Versus IMPROVE Network - 2:00 PM

Confidence Intervals Revisited

Dr Pepi Lacayo. U.S. EPA

Confidence intervals (CI's) are so basic to most sciences that some times they are not treated with the proper care. This most often occurs when CI's are estimated from small sample sizes. This paper discusses an approaches to reporting nominal and actual cover. Also discussed are confidence intervals vs tolerance intervals, vs prediction intervals. The paper is expository.

Monitoring the Long-Memory Air Quality Data Using ARFIMA Model

Su-Tsu Chen
Department of Statistics, National Cheng-Kung University, Tainan, Taiwan, R.O.C. 70101
Science Education Center, Fooyin University, Kaohsiung Hsien, Taiwan, R.O.C. 831
Jeh-Nan Pan
Department of Statistics, National Cheng-Kung University, Tainan, Taiwan, R.O.C. 70101

Abstract

Control chart is commonly used in the industry to help ensure stability of manufacturing process and quality of products. Control chart is also used to monitor the environmental data, such as industrial waste or effluent of manufacturing processes, however, it has to be modified if the environmental data exhibit the property of long memory. In this paper, the control chart for fractionally integrated autoregressive moving-average (ARFIMA) model is proposed to monitor the long-memory air quality data. In addition, empirical examples of control chart using autoregressive integrated moving-average (ARIMA) and ARFIMA models are compared. The result shows that control charts using ARFIMA model could be more appropriate than those using ARIMA model.

Keywords: long-memory, control chart, ARFIMA model, ARIMA model.

1. Introduction

Control chart is a useful statistical tool in quality control and improvement to monitor the variation of the key characteristics of products and to detect assignable causes that affect manufacturing or other environmental process. By applying control charts to environmental data, the change of the model of air quality can also be studied. Unlike the traditional control charts in which observations are assumed to be independent, observations of air quality usually have autocorrelations. Therefore, control charts should be used with modifications. Pan and Chen (2004) found that PM10 and O3 of air quality in the Taipei city follow ARIMA models and then compared the performance of four control charts for monitoring autocorrelated air pollution data to select the most appropriate one.

When applying control charts to autocorrelated data, it is assumed commonly that the data could be approximated by a statistical model and white noises. After fitting appropriate model to the data, the residuals are calculated. If the model is suitable, residuals should be independently and identically normally distributed and control charts can be applied to residuals then. If the control charts give a signal, the process will be intervened and necessary corrective actions need to be taken.

According to Chan and Hwang (1996), PM10 is the most important pollutant that deteriorates the air quality in Taiwan. Therefore, PM10 with long memory is the main pollutant discussed in this paper. The objective of this study is to develop a procedure of applying ARFIMA models to monitor long-memory air quality data. We show that natural logarithm of PM10 in southern Taiwan follows ARFIMA models instead of ARIMA models. Comparison of the suitability of applying ARIMA and ARFIMA models to air quality data are also conducted.

.

2. Development of Control Charts for Long- Memory Data

If data are not independent, data are fitted with a suitable model before control charts are used. ARIMA models are commonly used to fit autocorrelated data. Because of their simplicity and flexibility, they became very popular in applied time series analysis (Beran, 1994). The definition of ARIMA process is as following:

Definition 2.1 Let Z_t be a process such that

$$\Phi_p(B)(1-B)^d Z_t = \theta_0 + \Theta_q(B)a_t \qquad (1)$$

where d is positive integer or zero, B is the backshift operator ($BZ_t=Z_{t-1}$), ε_t is white noise, and $\Phi_p(B) = (1-\phi_1 B - \cdots - \phi_p B^p)$ and $\Theta_q(B) = (1-\theta_1 B - \cdots - \theta_q B^q)$ share no common factors. Then Z_t is called an ARIMA(p, d, q) process.

If a process is not stationary, some order of difference of the process is usually taken to make it stationary before control charts are applied.

Although ARIMA models are popular, they are not suitable to model data with long memory. ARFIMA models was proposed by Granger and Joyeux (1980) and Hosking (1981) to fit long-memory data, it is similar to Equation (1) except d is a real number between –0.5 and 0.5. There are several possible definitions of the property 'long memory.', for example, any discrete time series process with autocorrelation function (ACF), $\rho(h)$, at lag h possesses long memory if the quantity $\lim_{n\to\infty}\sum_{h=-n}^{n}|\rho(h)|$ is nonfinite.

When time series data have long memory, control charts for ARFIMA models are needed. The residual control charts of ARFIMA models are proposed in this study. Processes are modeled by ARFIMA model first, control charts are applied to the residuals then. There are Phase I and Phase II of constructing control charts. Phase I of constructing control charts for ARFIMA models are as follows:

1. Collect historical air quality or environmental data.
2. Fit the data collected in step 1 with an appropriate model. Check the suitability of the model. After a proper model is selected, residuals can be calculated.
3. Establish control limits for the residuals.
4. Delete any residuals fallen beyond the control limits and estimate parameters of control charts again.
5. Reestablish control limits for the residuals again.
6. Repeat step 4 and step 5 until there are no outliers/out-of-control signals.

If models and parameters of processes are known in advance, then the control limits could be calculated and one can bypass the Phase I. The control limits established in Phase I are used to monitor processes in Phase II. Most control charts for autocorrelated processes do not follow the procedures mentioned above. A literature survey would undoubtedly reveal that distinction between these two phases is lacking in most papers (Faltin et al., 1997). There seems to be a tendency to focus on Phase I, although this is usually not explicitly stated.

If a control chart can detect the change of parameters earlier than other control charts, it will be a better choice. Generally speaking, $\overline{X}$ chart would be a better choice if the larger mean shift is concerned. If the smaller mean shift is to be detected, then either EWMA or CUSUM chart can do the job well.

3. Comparison of ARFIMA and ARIMA Models

3.1 Example of Using the PM10 Data of Nantsz

Nantsz station is a surveillance stations located in the southern Taiwan where is known for a long history of public protest for pollution. The hourly air quality data of PM10 collected by Nantsz station between 1999 and 2002 is discussed and a total 725 observations were recorded during 1999 to 2000 and 712 observations were recorded between 2001 and 2002.

In Phase I, the 725 observations gathered during 1999 and 2000 are treated as historical data. A natural logarithm of PM10, denoted by ln(PM10), has successfully achieved the goal of stabling the variance without deseasonalizing and detrending the raw data. It is found that ARFIMA(0,d,1) is suitable for ln(PM10). After model fitting, a diagnostic of residuals are performed to check the suitability of residual control chart.

To monitor the change of residuals, EWMA chart is suggested to be used since it is known for its sensitivity to detect small sustained shift of process and its robustness to non-normal data. Assume residual at t-th time is r_t. The control statistic of EWMA residual control chart can be written as Equation 2.

$$Y_t = (1-\lambda)Y_{t-1} + \lambda r_t \qquad (2)$$

When EWMA chart is used, the parameter λ and the in-control ARL should be decided first. A different size of mean shift needs different λ. If a smaller mean shift is concerned, a smaller λ needs to be used. The parameterλof residual EWMA chart is set to be 0.1 and control limits are set to have in-control ARLs 370.8 according to Montgomery (2001). With applications of EWMA statistics in Phase I, the appropriate model can be written as Equation (3):

$$(1\text{-}B)^{0.47}\,(\ln(PM10_t) - 4.34) = (1 + 0.16B)\,\varepsilon_t\,. \qquad (3)$$

In Phase II, EWMA residual control chart constructed in Phase I is applied to the residuals as shown in Fig. 1.

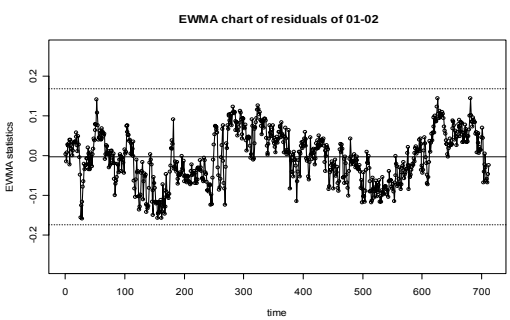


Figure 1. EWMA chart for the residuals of Nantsz's ln(PM10) data in Phase II using ARFIMA model.

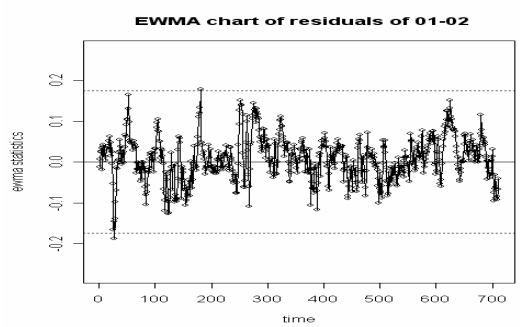


Figure 2. EWMA chart for the residuals of Nantsz's ln(PM10) data in Phase II using ARIMA model.

Fig. 1 indicates that no residual is out of control, which implies that ln(PM10) data in Phase II is likely to follow the similar pattern of Phase I. Thus, we may conclude that there is no evidence that the air quality of PM10 at Nantsz in Phase II is different from Phase I. This means that the air quality in Nantsz area has not been improved from 1999 to 2002 period. Further corrective actions need to be done.

If the long-term autocorrelations were ignored, then the most commonly used models to fit time series are ARIMA models. In contrast with the ARFIMA model, ARIMA model is compared for assessing the suitability of model selection. It is found that ARIMA (0, 1, 2) model could fit the air quality data of Nantsz from 1999 to 2000. With a proper Box-Cox transformation($\lambda=2$) the residuals are normally distributed, then EWMA control chart is applied to these transformed residuals. After performing the procedures in Phase I, the appropriate model can be written as Equation (4).

$$\mathbf{(1-B)\ ln(PM10_t)= -.0003 +(1-0.4031B -0.2883B^2)}\ \varepsilon_t . \qquad \mathbf{(4)}$$

In Phase II, we use Equation (4) to fit ln(PM10) of Nantsz collected between 2001 to 2002. Despite of the fact that residuals could not been transformed to be normally distributed with Box-Cox method, EWMA control charts shown in Fig. 2 are applied to monitor the residuals without transformation. Fig. 2 indicates that there are two points out of control and its pattern is different from Fig. 1. This Nantsz example demonstrates that the ARFIMA model is more appropriate than ARIMA model. False alarms would occur if one selects a wrong ARIMA model instead of using ARFIMA model.

3.2 Example of Using the PM10 data of Tsoying

For comparison with Nantsz's data, the PM10 data collected by Tsoying station, which is near Nantsz station, is discussed. A total 729 observations were recorded during 1999 to 2000 and 717 observations were recorded between 2001 and 2002.All the procedures are similar to Nantsz example. It is found that at the end of Phase I, the appropriate model can be written as Equation (5):

$$(1- 0.17B+0.17\ B^2)\ (1-B)^{0.49}\ (\ln(PM10_t) - 4.27) = \varepsilon_t . \qquad (5)$$

In Phase II, EWMA control chart, shown in Fig. 3, indicates that one residual is out of control, which implies that ln(PM10) data in Phase II is likely to follow the similar pattern of Phase I except one day, in that day air quality of Tsoying has abruptly changed.

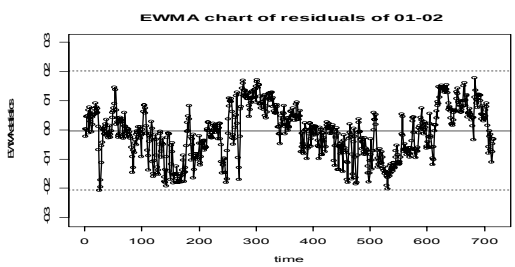


Figure 3. EWMA control chart for the residuals of Tsoying's data in Phase II using ARFIMA model

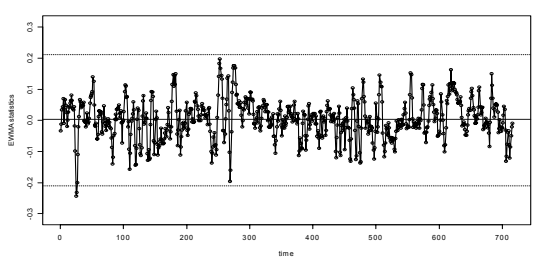


Figure 4. EWMA control chart for the residuals of Tsoying's data in Phase II using ARIMA model

If the long-term autocorrelations were ignored, then the most commonly used models to fit time series are ARIMA models. At the end of Phase I the appropriate model can be written as Equation (6).

$$(1-B)\ \ln(PM10_t)= -.0002 +(1-0.6855B +0.1171B^2)\ \varepsilon_t . \qquad (6)$$

In Phase II, EWMA control chart, shown in Fig. 4, indicates that there are two points out of control and its pattern is different from Fig. 3. This demonstrates that the ARFIMA model is more appropriate than ARIMA model and false alarms would occur if one selects a wrong ARIMA model instead of using ARFIMA model.

4. Conclusion

In this paper control charts using ARFIMA model are used to monitor long-memory air quality data. The proposed procedures of applying ARFIMA models to monitor air quality data can also be used for monitoring other long-memory environmental data. Through two empirical examples of air quality of southern Taiwan, one set of ln(PM10) data follows ARFIMA(0, 0.47,1) model while another one follows ARFIMA(2, 0.49,0) model, we have demonstrated that control charts using ARFIMA models are more appropriate than control charts using ARIMA models in monitoring long-memory air quality data. When monitoring autocorrelated data, the meaning of "out-of-control" indicates not only the residuals of processes may deviate from what are assumed, but also the underlying model of the process might be changed.

Due to the complexity of the ARFIMA model, procedures for constructing an appropriate control chart especially in Phase II is more difficult than the traditional autocorrelated control charts. To ensure the efficiency and effectiveness of monitoring environmental data with long memory, it is suggested that an on-line analyzer of the residual control chart using ARFIMA model be developed to help practitioners understand the quality change of the environment, so a timely corrective action can be made.

5. References

Beran J. (1994). *Statistics for long-memory processes*. Chapman & Hall: New York.

Brockwell PJ, Davis RA. (1991). *Time Series: Theory and Methods*. Springer-Verlag: New-York.

Chan CC, Hwang JS. (1996). Selling the blue skies: some reflection on air pollution fee policy in Taiwan. *Journal of the Chinese Institute of Environmental Engineering* **6**, 235-46.

Faltin FW, Mastrangelo CM, Runger GC. (1997). Considerations in the monitoring of autocorrelated and independent data. *Journal of Quality Technology* **29**(2), 131-133.

Granger CWJ, Joyeux R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**, 15-29.

Hosking JRM. (1981). Fractional differencing. *Biometrika* **68**, 165-176.

Pan JN, Chen BD. (2004.) The comparison of environmental control charts for monitoring autocorrelated air pollution data in Taipei area. *Journal of The Chinese Statistical Association* **42**(1), 31-62.

Combining Classical Analyses with Practicality: A Different Look at the STN versus IMPROVE Network Intercomparison Study using Deming Regression and a Conservative Outlier Identification Technique

Louise H. Camalier
EPA Office of Air Quality, Planning and Standards
Air Quality Analysis Division
Mail Drop C304-02
Research Triangle Park, NC 27711
email: camalier.louise@epa.gov
phone: 919-541-0200

Abstract

In a continuing effort to better collect and analyze speciated particulate matter in ambient air, a chemical Speciation Trends Network (STN) of 54 sites was implemented to characterize the primary chemical components of mass in urban areas and to provide data for trends across the United States. The Interagency Monitoring of Protected Visual Environments (IMPROVE) program, established in 1985, supports the Regional Haze Rule and serves as a rural companion network to the STN for urban versus rural comparison purposes.

Data from both the urban STN and rural IMPROVE networks are often used together to help assess urban excess emissions, track trends, and assess improvements in air quality. There are some differences in the sampling and analysis protocols used by these two networks; therefore, it is important for data users to understand the differences and similarities in the data generated. Furthermore, it is important to use this information to gain the knowledge needed to make program improvements. To aid in this understanding, the EPA launched a small network of collocated STN and IMPROVE comparison sites in a few urban/rural areas. Data from these comparison sites has been evaluated in the past to begin exploring the differences, improve our understanding, and educate the data use community.1 This paper takes a different approach to the data exploration and statistical analysis of the comparison data to further explore differences.

The first step of the analysis process was to validate the intercomparison study carbon data. For quality assurance purposes, the Mahalanobis' distance was calculated for both elemental and organic carbon at each site to identify bivariate outliers that were consequently excluded from the analysis. After the data validation process, Deming regression was used to evaluate the relationship between STN and IMPROVE concentrations at each specific site for both carbon types. Deming Regression is a method used for evaluating a relationship between two variables (x and y), allowing error in both. This method develops a proportional relationship between errors and concentration to reasonably estimate the slope and intercept, which proves beneficial as it depresses the dominance of both very high values as well as very low values (below MDL -- where the ratio of uncertainty vs. concentration is high). The Bootstrap re-sampling technique is used to estimate the standard error of the slope and intercept, to test slope and intercept hypotheses, and to assess differences among urban and rural sites.

Graphical examples and results from the study at the six locations are presented for comparison of the two networks. Results show a unique look at the STN versus IMPROVE intercomparison carbon data through data validation and the use of a regression technique that accounts for bivariate uncertainty

Introduction

A national monitoring program was implemented by the United States Environmental Protection Agency (EPA) to support the $PM_{2.5}$ National Ambient Air Quality Standard (NAAQS) for gravimetric mass in 1997. In a continuing effort to better collect and analyze speciated particulate matter in ambient air, a chemical Speciation Trends Network (STN) of 54 sites was implemented to characterize the primary chemical components of mass in urban areas and to provide data for trends across the United States. The Interagency Monitoring of Protected Visual Environments (IMPROVE) program, established in 1985, supports the Regional Haze Rule and serves as a rural companion network to the STN for urban versus rural comparison purposes. The IMPROVE network consists of 110 sites put in place to support assessment and tracking of visibility impairment in Class I areas.

Data from both the urban STN and rural IMPROVE networks are often used together to help assess urban excess emissions, track trends, and assess improvements in air quality. There are some differences in the sampling and analysis protocols used by these two networks; therefore, it is important for data users to understand the differences and similarities in the data generated. Furthermore, it is important to use this information to gain the knowledge needed to make program improvements. To aid in this understanding, the EPA launched a small network of collocated STN and IMPROVE comparison sites in a few urban/rural areas. Data from these comparison sites have been evaluated in the past to begin exploring the differences, improve our understanding, and educate the data use community.1 This paper takes a different approach to the data exploration and statistical analysis of the comparison data to further explore differences in carbon.

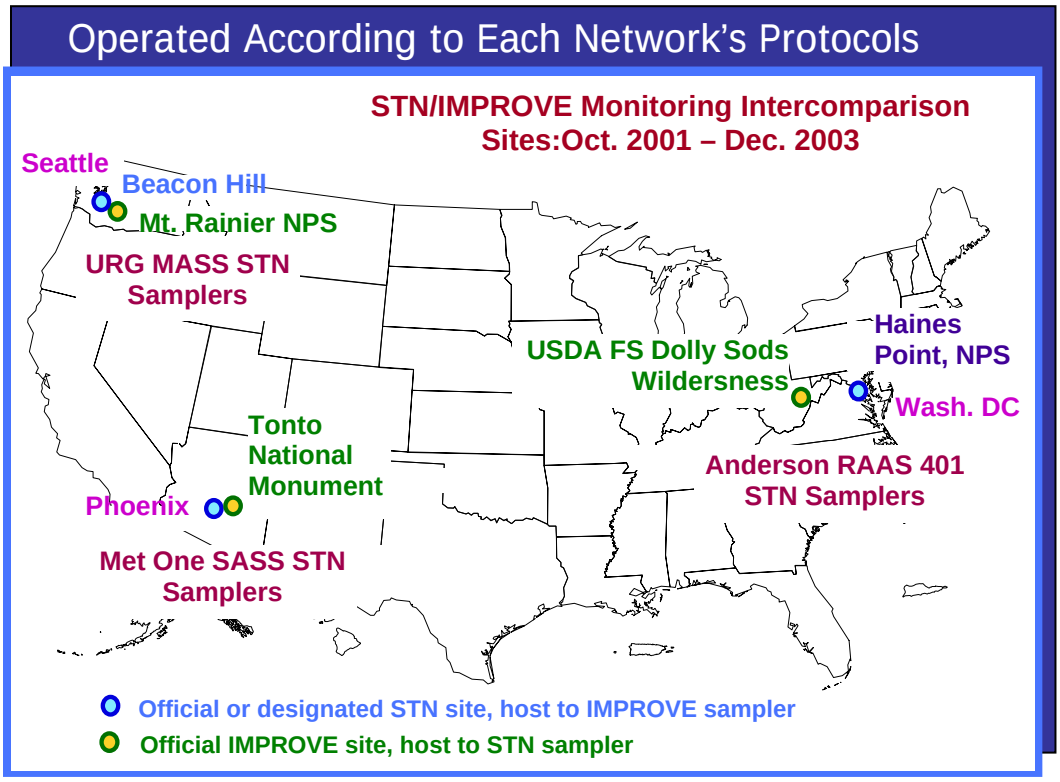
Inherent Network Differences and Initial Study Design

The study design covered three independent geographic regions with six sites. Two sites were established in each geographic area; one in a city and one in a more remote setting, with collocated STN and IMPROVE samplers to monitor both the urban and rural behaviors. The geographic areas covered in the study were the northeast (Washington, D.C.), the northwest (Washington state), and the southwest (Arizona). Each of these geographic regions has a unique pollutant "thumbprint" that is based on its own particular sources and meteorology. While the urban-based sites were located in a major metropolitan area, their rural counterparts were located in a national park or other remote setting. The urban-rural pairing was set in place to allow for an interesting look at urban versus rural concentration levels and aerosol types.

With regard to the measurement of carbon particulate matter, differences in sampling procedures or methods can have an effect on the resulting measurements. The first difference between the STN and IMPROVE sampling protocols is the nominal flow rate and filter diameter, or filter face velocity. The STN samplers use 46.2-mm filters and operate in a range from 6.7 to 16.7 L/min, depending on sampler type. The IMPROVE samplers use a 25-mm filter for carbon and operate around 22.8 L/min. Both programs use quartz filters for collecting carbon. In this intercomparison study, the STN had an Andersen sampler located in the northeast (Haines Point, MD and Dolly Sods, WV), a MetOne sampler in the southwest (Phoenix and Tonto, AZ), and a

URG sampler in the northwest (Beacon Hill and Mt. Rainier, WA). The Andersen had a flow rate of 7.3 L/min, while the MetOne had a flow rate of 6.7 L/min, and the URG sampler had a flow rate of 16.7 L/min. Another innate difference between the two networks is the temperature in which the sampled filters are shipped to the lab for analysis. STN filters are shipped overnight in coolers with a reusable ice substitute, while IMPROVE filters are shipped via standard postal service under ambient conditions. While both networks collected and analyzed field and trip blanks, only the IMPROVE program actually "corrected" the raw carbon measurement data for artifacts using the blank analyses. This correction makes the data not comparable because the correction for blanks on only one side of the study causes a negative bias in the IMPROVE data versus the STN raw data. There is also an argument against using results from blank analysis as a correction at all because of the high level of uncertainty and the scientific disagreement on the procedures that should be used to estimate sampling artifacts. In addition, the procedures for collecting blank samples for artifact assessment differ between the programs; therefore, "blank-correcting" the STN data to try to reconcile artifacts could be problematic. Of the procedural inconsistencies between the STN and IMPROVE programs, the area that has received the most attention is the difference between the STN and IMPROVE organic carbon (OC) and elemental carbon (EC) measurements as defined by the differing thermal optical analysis protocols. The STN and IMPROVE network use different thermal profiles and optical correction techniques to define the OC and EC split.

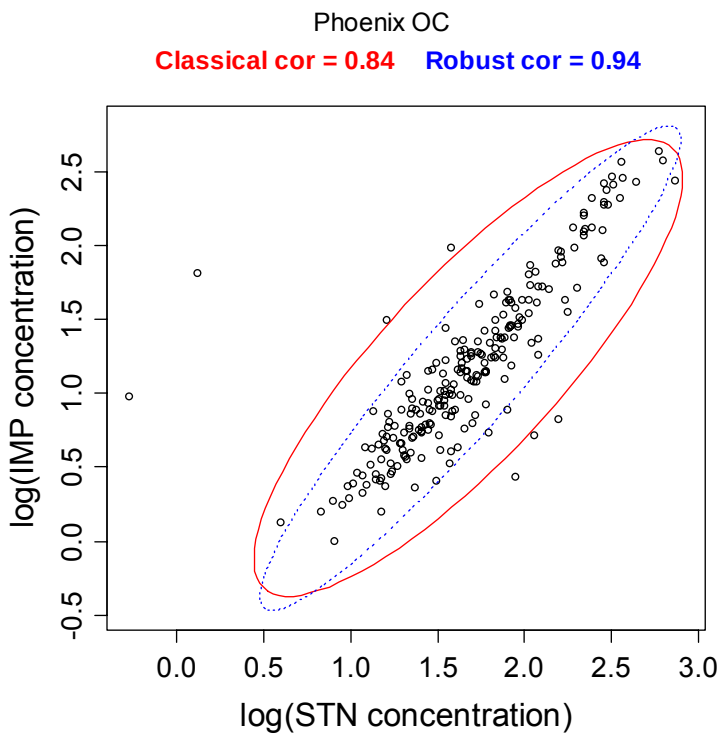
The only factor that remained constant in the study design was the species to be collected and measured (see map to the right). An important variable in the intercomparison study was the geographic differences (both meteorological and varying PM characterization profiles) in the six locations in which the study took place.



Data Treatment and Validation

The database used for the STN/IMPROVE intercomparison analysis was built by the EPA's Office of Research and Development. The database consisted of all data readings, including those below the method detection limit (MDL). This inclusion of data below the MDL is important as it minimizes the introduction of bias into the system. However to properly assess differences, one must begin with the raw data. The calculated blank value that was previously subtracted from the IMPROVE readings (page 3 of this document) was added back onto each datapoint to regain the original "raw data" from which to work. For quality assurance purposes, the Mahalanobis' distance was calculated for each parameter and site to identify bivariate outliers that were consequently excluded from the analysis. The Mahalanobis distance is one of many methods to detect bivariate outliers, and uses the variable populations' covariance matrix to calculate a Mahalanobis distance from the mean of the data. The region of constant Mahalanobis' distance around the mean forms an ellipse in two dimensions in which the most "representative" data points are contained (see figure below). This method proves useful for data filtering, especially when x and y are highly correlated, however one of its assumptions is that the data are normally distributed. Because the distributions of the parameters at each site appeared to be lognormally distributed, the data were log-transformed to improve linearity.

The decision between using a robust or classical Mahalanobis' distance is subjective and depends on the nature of the data and the level of strictness. The more "outliers" identified and excluded, the higher the correlation between x and y becomes. The robust correlation will always be higher than the classical correlation, however choosing a robust Mahalanobis' fitting strictly based on this characteristic may not always achieve the most reasonable result.
Careful consideration was given to the robust Mahalanobis' elliptical region, however the classical Mahalanobis' fitting was more conservative and allowed for more uncertainty in STN and IMPROVE data values, which is best in terms of the highly variable nature of speciated ambient particulate matter.



The data from each parameter and site combination was filtered for significant bivariate outliers by calculating the classical Mahalanobis' distance, excluding data at the α=.01 rejection level. This method and rejection region results in a very conservative filter that allows for more natural variability while avoiding artificial improvement of the correlation and unnecessary data loss.

Deming Regression and the Estimation of Two-Dimensional Error

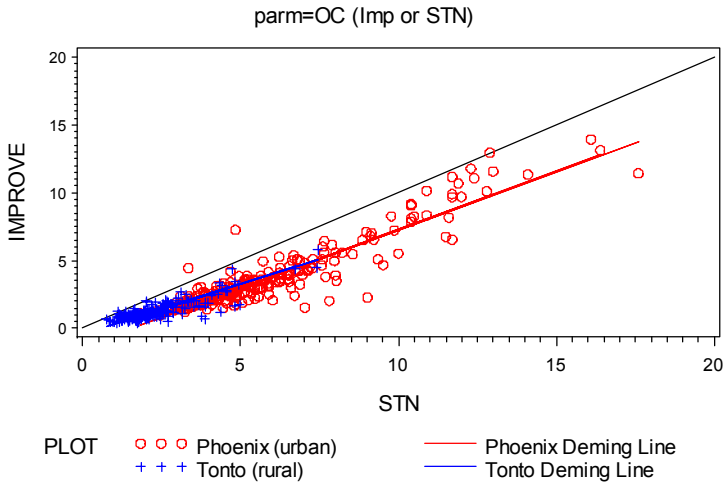
After the data validation process, Deming Regression was used to evaluate the relationship between STN and IMPROVE concentrations at each specific site for each pollutant. Deming Regression is a method used for evaluating a relationship between two variables (x and y), allowing error in both.2

As Deming Regression is based on error in both variables, an appropriate two-dimensional error structure needed to be created to best calculate the linear x-and-y fit. A function was created to best approximate the actual measurement error for both IMPROVE and STN, and consisted of an artifact (intercept) variability along with a reasonable estimate of proportional error due to concentration (slope). The artifact variability component of the error was based on the estimation of the standard deviation of the trip and field blank values. The variability of the blank values provides a good idea of "intercept error" while the median of relative percent differences (RPD) of collocated measurements within sampler type provides a representative and feasible slope-related error.

Based on the individual behaviors of method measurements, IMPROVE and STN should ideally have unique error structures. A personalized error structure for IMPROVE was not possible however, because of the lack of availability of variability data for IMPROVE. This resulted in the decision to use the standard deviation of the blank value for the STN samplers as a surrogate for the error associated with the IMPROVE blank. The substitution of the standard deviation from the STN blanks for the standard deviation from the IMPROVE lab blanks proved to be the best estimation of IMPROVE blank variability because the STN blank value incorporated error from both the lab and sampling (field) setting (field and trip blanks). In addition, collocated measurements were only available for STN samplers. The "within method" variability procured from the collocated STN data was used to approximate the "within method" variability for IMPROVE. The error structure for IMPROVE and STN is defined below:

error(y) = std(STN blank value by area) + median(RPD of colloc STN monitors by param)*x
error(x) = std(STN blank value by area) + median(RPD of colloc STN monitors by param)*y

(where y is the IMPROVE concentration and x is the STN concentration)



The slope and intercept was calculated for each parameter and site combination (with 99% confidence intervals on both) using the Deming Regression method with error in both x and y estimated by the specified functions above (see figure at left for the resulting Deming fit). Setting both x and y to follow the same error structure results in the standard error ratio of the Deming Regression to be 1. This causes neither STN nor IMPROVE to hold more precedence as the independent variable.

The Bootstrap technique was then used

to appropriately estimate the standard error of the slopes and intercepts to come up with the most reasonable slope and intercept estimates. Bootstrapping is a re-sampling technique that was run 500 times for each parameter and site combination. The results of the Bootstrapping produce the standard error estimate to which one can use in order to test slope and intercept hypotheses, and to assess differences among urban and rural sites.

Organic Carbon and Elemental Carbon Results and Explanations

The table below displays results from slope hypothesis testing and assessments of rural vs. urban differences for organic carbon and elemental carbon from the initial intercomparison study data.

		Arizona -- slope				Washington D.C. -- slope				Washington State -- slope			
Category	**Parameter**	**Urban**	**Rural**	**SE**	**t-value**	**Urban**	**Rural**	**SE**	**t-value**	**Urban**	**Rural**	**SE**	**t-value**
Carbon	**EC**	0.937	1.249	0.14	-2.18	1.020	1.062	0.08	-0.54	1.037	1.863	0.11	**-7.70**
	OC	*0.861*	*0.742*	0.04	**2.83**	*0.757*	*0.804*	0.04	-1.09	*1.088*	*0.915*	0.04	**4.86**
Total	**PM2.5**	1.025	*1.147*	0.05	-2.36	*0.967*	*0.935*	0.01	2.23	*1.039*	*0.922*	0.03	**3.80**

Legend:
Value: slope was found to be significantly different from 1 (based on a null hypothesis test: slope is equal to 1)
Value: rural/urban differences were found between species at an area (based on a null hypothesis test: both slopes are equal)

Organic Carbon (OC) is a difficult species to quantify due to positive and negative sampling artifacts, lack of thermal optical analysis method consistency, and lack of method calibration materials. The STN and IMPROVE quartz filters are analyzed by different thermal optical methods, which differ in temperature profile, analysis time, laser correction for char, and artifact correction procedures. The two programs also face differing sampling method flow rates and subsequent filter face velocities which consequently affect the negative artifact. In addition, neither program addresses positive artifacts through the use of gas-phase organic denuders. Significant urban/rural slope differences were not apparent for OC for sites in Arizona (Phoenix and Tonto) and Washington D.C. (Haines Point and Dolly Sods), however these slopes were all below one, indicating that IMPROVE consistently measures lower OC than STN. This discrepancy is likely a result of the differing flow rates/filter face velocities of the sampler types. Because IMPROVE's flow rate is higher, a negative bias is created on OC when compared to lower flow rate samplers. This flow rate difference is not present in Washington state (Seattle and Mt. Rainier) which is most likely due to the STN URG sampler flow rate of 16.7 L/min, which is more equivalent to that of the IMPROVE method (22 L/min).

Elemental Carbon (EC) is similarly difficult to quantify depending on the source profile for the area in which the particulate is measured. No significant slope differences are seen for EC in urban areas, as these sources are mainly suspected to be composed of diesel and gasoline emissions (where EC concentrations are larger). The only significant slope difference occurs in Mt. Rainier, in urban Washington State, where the slope is ~1.86. Other scientists have identified a factor of 2 difference for IMPROVE EC.3,4 The proposed theory behind this difference is the contribution of woodsmoke in this area and low values of elemental carbon.

The results from this and other studies were used to further understand differences between STN and IMPROVE carbon measurements and to support decisions made regarding the institution of a consistent sampling and analysis method for OC and EC in the EPA Speciation network.

References

1. Solomon, P., Klamser-Williams, T., Egeghy, P., J. Rice, and Crumpler, D. "STN/IMPROVE Comparison Project; Preliminary Results," at the MARAMA/MANE-VU meeting on the Carbonaceous Aerosol; Baltimore, MD; January 2004.
2. Cornbleet, J., Gochman, N. "Incorrect Least-Squares Regression Coefficients in Method-Comparison Analysis," *Clinical Chemistry*, Vol. 25, No. 3, pp. 432-438, 1979.
3. Chow, J., et.al. "The DRI Thermal/Optical Reflectance Carbon Analysis System: Description, Evaluation and Applications in U.S. Air Quality Studies," *Atmospheric Environment*, Vol. 27a, No. 8, pp. 1185-1201, 1993, Great Britain.
4. Watson, J.G., Chow, J.C., Antony Chen, L.-W. "Summary of Organic and Elemental Carbon/Black Carbon Analysis Methods and Intercomparisons," *Aerosol and Air Quality*, Vol. 5, No. 1, pp. 65-102, 2005.