Appendix 7: DQO Planning Document

Revision 8.3

Prepared for: US Environmental Protection Agency

Clean Air Status and Trends Network

Quality Assurance Project Plan

Revision 8.3

Appendix 7:

DQO Planning Document

October 2015

Clean Air Status and Trends Network

Data Quality Objectives

Planning Document

Prepared for:

U.S. Environmental Protection Agency

Office of Air and Radiation

Research Triangle Park, NC

Prepared by:

AMEC E&I, Inc.

Gainesville, Florida

Data Quality Objectives

Planning Document

1.0 State the Problem

1.1 Planning Team Members

The Clean Air Status and Trends Network (CASTNET) Data Quality Objectives (DQO) planning team is comprised of the AMEC E&I, Inc. (AMEC) Project Manager, Quality Assurance (QA) Supervisor, operations managers, and other personnel with appropriate expertise as needed, and the U.S. Environmental Protection Agency (EPA) QA Officer and Technical Monitors. The DQO decision makers are the AMEC Project Manager, QA Supervisor, and operations managers, together with the EPA Project Officer, QA Officer, and monitors. This planning team will develop and refine CASTNET DQO to support and maintain CASTNET project objectives. The decision makers have the ultimate authority to make final decisions based on the recommendations of the planning team.

1.2 Problem Description/Background

When the U.S. Congress amended the Clean Air Act in 1990, Title IV (Acid Deposition Control Program) mandated a significant reduction in the emissions of sulfur and nitrogen oxides, primarily from the electric utility industry. Titles IV and IX of the Clean Air Act Amendments (CAAA) required that the environmental effectiveness of the Acid Deposition Control Program be assessed through environmental monitoring. This monitoring was required for gauging the impact of emission reductions on air pollution, atmospheric deposition, and the health of affected human populations and ecosystems.

Prior to CASTNET, EPA operated the National Dry Deposition Network (NDDN), which was established in 1986. As with CASTNET, the objective of the NDDN was to obtain field data to establish patterns and trends of dry deposition at approximately 50 sites throughout the United States. The approach adopted by NDDN was to estimate dry deposition using measured air pollutant concentrations and modeled deposition velocities (V_d) estimated from meteorological, land use, and site characteristic data. Since four to five years of data had been collected using the site locations, sampling methodology and frequencies, and equipment types established under NDDN, the same project design was used as the basis for CASTNET. CASTNET became operational in mid-1991. NDDN was incorporated into CASTNET at that time.

1.3 Resources

Published technical studies indicate that using NDDN as a guide/basis for CASTNET was a proper and cost effective strategy, especially in light of the data previously collected by NDDN. Clarke *et al.* (1997)

demonstrated the accuracy and precision of CASTNET/NDDN monitoring data and Holland *et al*. (1998) demonstrated CASTNET/NDDN trend measurement sensitivity.

2.0 Identify the Decision

CASTNET's primary goal is to function effectively as a national, long-term deposition monitoring network that provides information for assessing the effectiveness of current and future emission reductions mandated under the Clean Air Act. To meet this goal, the CASTNET program was designed to fulfill the following objectives:

1. To monitor the status and trends in air quality and atmospheric deposition;
2. To provide atmospheric data on the dry deposition component of total acid deposition, rural ground-level ozone, and other forms of atmospheric pollution that enter the environment as particles and gases; and
3. To assess and report on geographic patterns and long-term, temporal trends in ambient air pollutant concentrations and acid deposition.

The network design was developed based on the assumption that dry deposition can be estimated mathematically using ambient concentrations and meteorological inputs.

2.1 DQO Trends Study Objective

The objective for trends in atmospheric sulfur and nitrogen species is a 10 percent minimum detectable trend after 10 years with a 95 percent level of confidence. This DQO was established based on the study published by Holland *et al.* (1998) that utilized CASTNET data for 34 sites in the eastern United States. Analysis of data from 1989 through 1995 demonstrated that a 10 percent trend could be detected after 10 years of data collection with a 95 percent level of confidence for sulfur dioxide (SO_2), particulate sulfate (SO_4^{2-}), and nitrogen [nitric acid (HNO_3) + particulate nitrate (NO_3^-)]. Analyses were conducted using generalized additive models (GAM) to estimate percent change per year in mean monthly concentrations. Unlike the usual linear models, GAM allow the data to suggest the form of the model. GAM were used rather than linear models to account for variables such as meteorology and seasonality. Confidence was evaluated by iterative deletion of one month of data from the total for a given site. The model estimate for a certain month using all collected data was compared with the model estimate for the same month with its data removed. The study showed that a yearly trend of less than 1.0 percent could be detected with 95 percent confidence. In other words, there is a 95 percent probability of detecting a minimum trend of 10 percent after 10 years at any particular site, for SO_2, SO_{-4}^{2} and nitrogen (N). The objective for trends in CASTNET data is to detect, at minimum, a 1.0 percent annual trend in concentrations after 10 years of data collection.

The CAAA Title IV Control Program mandated a 10-million ton reduction from 1980 emissions for SO_2 and a 2-million ton reduction for NO_x. In 1980, SO_2 emissions were measured at 26 million short tons1. A 10-million ton reduction from 1980 levels would be equal to an approximate 38 percent decrease. If this reduction had been achieved in 1991 when CASTNET started, it would indicate a decrease of 4 percent per year from 1980 levels. NO_x emissions were measured at 23 million short tons in 1980[2]. If achieved, a 2-million ton reduction would be equal to an approximate 8.7 percent decrease in 1991 or about 0.9 percent per year. Holland *et al.* (1998) demonstrated that SO_2 and N trends in airborne concentrations could be detected at 1.0 percent per year with a 95 percent level of confidence for sites in the eastern United States. For western sites, low site density and low concentrations prevent extrapolation of this result. However, since the U.S. Congress is the de facto decision-maker as regards the reductions required by CAAA Title IV, the 4 percent SO_2 and 1 percent NO_x decisions still apply for western sites. More data from western sites, including better geographic coverage and coverage of meteorological conditions, are needed to make a reasonable determination of the sensitivity of trend calculations for this region.

2.2 Additional DQO

Spatial patterns are also desired for policy decision-making. Initial study into formulation of a spatial pattern DQO was performed by Dr. William Tucker of AMEC. The Technical Memorandum resulting from this initial study is attached as Appendix A.

Uncertainties in the computer model have not been sufficiently quantified to determine a precise DQO for deposition flux.

3.0 Identify the Inputs to the Decision

Parameters used by the computer model for CASTNET/NDDN are listed in the following table:

Measurement Parameter	Medium	Method
Wind Speed	Continuous Ambient Monitoring	Anemometer
Wind Direction	Continuous Ambient Monitoring	Wind Vane
Sigma Theta	Continuous Ambient Monitoring	Wind Vane
Relative Humidity	Continuous Ambient Monitoring	Thin Film Capacitor
Solar Radiation	Continuous Ambient Monitoring	Pyranometer
Precipitation	Continuous Ambient Monitoring	Tipping Bucket Rain Gauge Weighing Rain Gauge
Ambient Temperature	Continuous Ambient Monitoring	Platinum RTD

[1] http://www.epa.gov/oar/emtrnd94/tres.pdf

[2] Ibid.

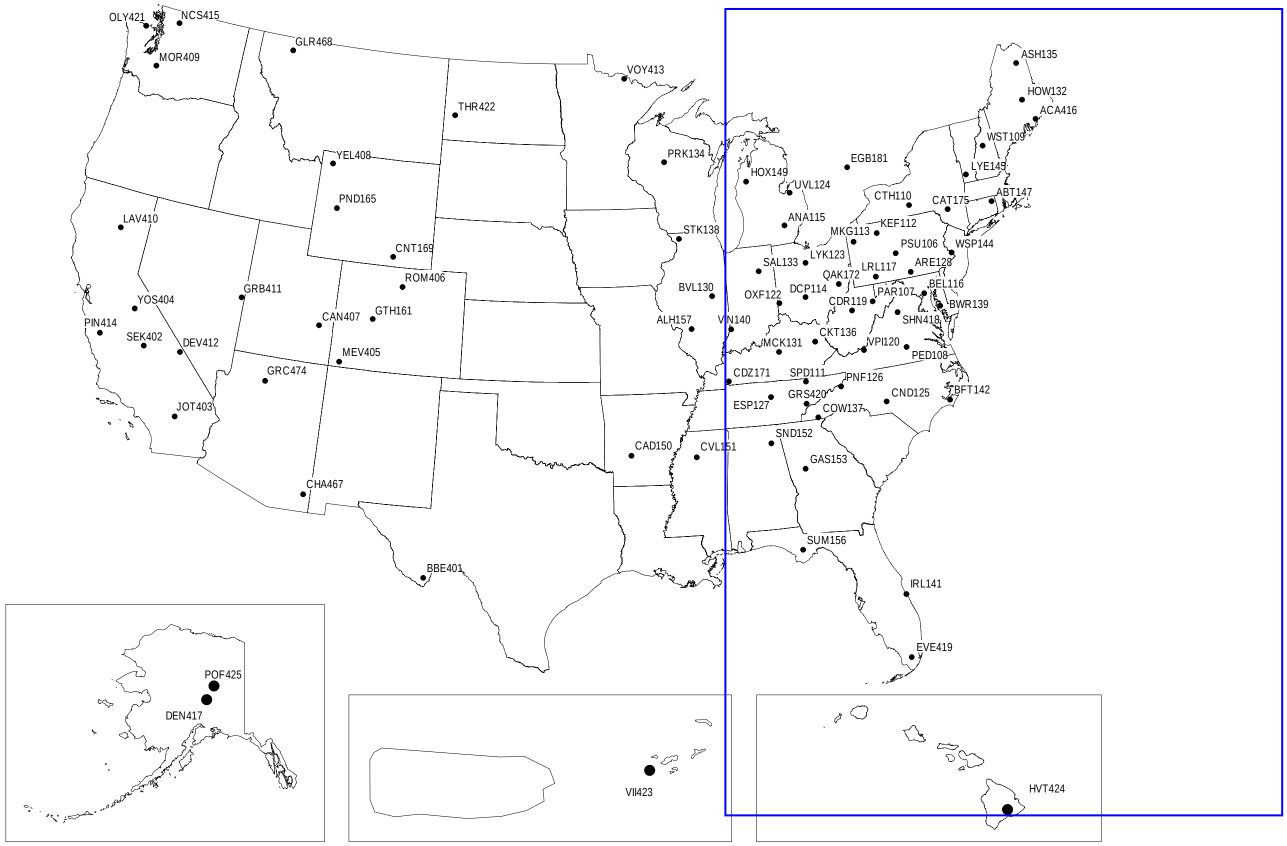
Surface Wetness	Continuous Ambient Monitoring	Conductivity Bridge
O_3	Continuous Ambient Monitoring	Ultraviolet Absorbance
Filter Pack Flow*	Continuous Ambient Monitoring	Mass Flow Controller
Ammonium (NH_4^+)	Filter Pack Samples	Automated Colorimetry
Sodium (Na^+)	Filter Pack Samples	ICAP-AE
Potassium (K^+)	Filter Pack Samples	ICAP-AE
Magnesium (Mg^{2+})	Filter Pack Samples	ICAP-AE
Calcium (Ca^{2+})	Filter Pack Samples	ICAP-AE
Nitric Acid (HNO_3)	Filter Pack Samples	Ion chromatography
Nitrate (NO_3^-)	Filter Pack Samples	Ion chromatography
Sulfate (SO_4^{2-})	Filter Pack Samples	Ion chromatography

Note: *Flow rate is used along with filter pack sample measurements to calculate atmospheric concentrations. The calculated atmospheric concentrations are then used in the model.

ICAP-AE = inductively coupled argon plasma-atomic emission
RTD = resistance-temperature device

4.0 Define The Study Boundaries

The study boundaries for site-by-site concentration trends encompass the entire network, as would the boundaries of future studies to quantify deposition and spatial distribution within the network. However, the western region of the network has a lower site density compared with the eastern region. For this reason, results from the studies cited previously should not be extrapolated to apply to the entire network. More data from western sites, including new sites with additional coverage of geographic and meteorological conditions, are needed to make a reasonable determination of any network-wide trends or spatial distribution DQO. Current CASTNET site locations are indicated on the following map. Only the sites located within and to the east of the line of states extending from Minnesota in the north to Louisiana in the south (outlined in blue) comprise the area where complete DQO for concentration trends and spatial distribution of SO_2 can be established with reasonable rigor.



5.0 Develop a Decision Rule

The accuracy of reported CASTNET concentration data allows for measurement of a minimum annual change of 1.0 percent for SO_2, SO_4^{2-}, and N at a given site in the eastern region with a statistical confidence of 95 percent. Spatial distribution maps for SO_2 are accurate with 90 percent confidence for eastern sites.

6.0 Specify Tolerable Limits on Decision Errors

6.1 Sulfur Dioxide, Particulate Sulfate, and Nitrogen

Limits on decision errors for SO_2, SO_4^{2-}, and N are indicated by the study published by Holland *et al.* (1998) that utilized CASTNET data from 1989 through 1995 for 34 sites in the eastern United States. Analysis of these data demonstrated that an approximate 1.0 percent trend could be detected per year with a 95 percent level of confidence. In other words, there is a 95 percent probability of detecting a minimum trend of 10 percent after 10 years at any particular site, for SO_2, SO_4^{2-}, and N at eastern sites.

6.2 Ozone

Limits on decision errors for ozone (O_3) are based on analysis of historical O_3 calibration results at eastern sites. All calibrations were performed using EPA traceable standards, which provide a good indication of analyzer accuracy. The data show that 98 percent of all calibrations on record from January 1989 through October 2001 were within the established ±10 percent criterion (i.e., calibration curve slopes were between 0.90 and 1.10) and 97 percent were within the ±5 percent criterion (slopes were between 0.95 to 1.05). Calibration results for all collocated sites (approximately 106, paired) for the same period yielded a measured precision of 3 percent. Using the following propagated uncertainties:

Analyzer accuracy = 5 percent or 0.05
Network precision = 3 percent or 0.03
Total Propagated Uncertainty (TPU) = SQRT$[(0.05)^2 + (0.03)^2]$ = 0.06 percent

These data indicate that trends above 6 percent can be detected after approximately 13 years with a 97 percent confidence level. For the sake of simplicity, the stated DQO for O_3 will match the sulfur and nitrogen species DQO (i.e., 10 percent minimum detectable trend after 10 years with 95 percent confidence).

6.3 Spatial Distribution of SO_2

Spatial distribution maps for SO_2 in the eastern United States show a real pattern with 90 percent confidence that the maximum interpolated value is greater than the minimum interpolated value. This applies to the area shown on the map as a whole. More analysis is needed to establish accuracy for a given locality within the mapped region. The test for local areas will likely involve analysis of absolute errors in the kriging estimates as compared with the estimated geometric means for the region as described in Appendix A.

6.4 Dry Deposition

A DQO for trends in dry deposition is not practical at this time. Although Meyers *et al*. (1998) and Finkelstein *et al*. (2000) demonstrated that the MLM is essentially unbiased for flat, non-forested settings, the uncertainties in the MLM have not been sufficiently quantified for establishing a definitive DQO.

6.5 Other Measurements

As stated previously, more data are needed to quantify accuracy and uncertainties in all measurements at western sites.

In addition, more analysis is needed to determine a spatial patterns DQO for all pollutants. More analysis using kriging is necessary to allow for more accurate extrapolation of spatial distribution data to smaller localized areas.

7.0 Optimize the Design

Since NDDN sites were transferred to CASTNET at the beginning of the project, initial network and site design were necessarily driven by the prior design of NDDN and the four to five years of data collection already performed. Site design and sampling methodology have largely been dictated by this and by computer model requirements. Sampling duration and frequency were selected for increased comparability with other networks such as the National Oceanic and Atmospheric Administration’s (NOAA) Atmospheric Integrated Research Monitoring Network (AIRMoN). The selection of the parameters measured and completeness requirements are all model-driven. Factors not driven by model requirements, such as regional site density, may allow for further optimization if research shows that project objectives may still be met. For example, automated sequential samplers may reduce costs if it is determined that filter packs can remain on the tower for a certain period after sampling is complete, thus reducing site operator visits. The spatial pattern estimation, as noted in Dr. William Tucker’s technical memorandum (Appendix A), may be cost-optimized with further research into the level of site densities required per region to achieve a certain minimum accuracy of kriging estimates.

References

Brook, J.R., Di-Giovanni, F., Cakmak, S., Meyers, T.P. 1997. Estimation Of Dry Deposition Velocity Using Inferential Models and Site-Specific Meteorology – Uncertainty Due to Siting of Meteorological Towers. Atmospheric Environment 31, 23, 3911-3919.

Clarke, J.F., Edgerton, E.S., Martin, B.E. 1997. Dry Deposition Calculations for the Clean Air Status and Trends Network (CASTNET). Atmospheric Environment. 31:3667-3678.

Erisman, J.W., Draaijers, G., Duyzer, J., Hofschreuder, P., Van Leeuwen, N., Romer, F., Ruijgrok, W., Wyers, P., Gallagher, M. 1997. Particle Deposition to Forests – Summary of Results and Application. Atmospheric Environment 31, 3, 321-332.

Finkelstein, P.L., Ellestad, T.G., Clarke, J.F., Meyers, T.P., Schwede, D.B., Hebert, E.O., and Neal, J.A. 2000. Ozone and Sulfur Dioxide Dry Deposition to Forests: Observations and Model Evaluation. JGR. 105:D12:15,365-15,377.

Holland, D.M., Principe, P.P., and Sickles, J.E. 1998. Trends in Atmospheric Sulphur and Nitrogen Species in the Eastern United States for 1989-1995. Atmospheric Environment. 33:37.

Meyers, T. P., Finkelstein, P., Clarke, J., Ellestad, T.G., and Sims, P.F. 1998. A Multilayer Model for Inferring Dry Deposition Using Standard Meteorological Measurements. J. Geophys. Res., 103D17:22,645-22,661.

U.S. Environmental Protection Agency (EPA). 1998. Clean Air Status and Trends Network (CASTNET) Deposition Summary Report (1987-1995). EPA/600/R-98/027. OAQPS, Research Triangle Park, NC 27711.

U.S. Environmental Protection Agency (EPA). 1991. GEO-EAS 1.2.1 User's Guide. EPA/600/8-91/008.

Wesely, M.L., Cook, D.R., Hart, R.L., Speer R.E. 1985. Measurements and Parameterization of Particulate Sulfur Dry Deposition Over Grass. Journal of Geophysical Research 90, 2131-2143.

Wesely, M.L., Cook, D.R., Hart, R.L. 1983. Fluxes of Gases and Particles Above a Deciduous Forest in Wintertime. Boundary-Layer Meteorology 27, 237-255.

Appendix A

Technical Memorandum

TO: Tom Lavery
Kemp Howell

DATE: January 23, 2002

Cc: Mary Burnett

FROM: William Tucker

SUBJECT: CASTNET Data Quality Objectives – Spatial Patterns

The purpose of this memorandum is to recommend a technical approach for defining Data Quality Objectives for the Clean Air Status and Trends Network regarding spatial patterns of air quality and deposition.

The objectives of the analysis are:

- Identify quantitative measures that characterize the reliability of inferred spatial patterns. "Spatial Patterns" imply such products as maps that depict concentration and/or deposition isopleths. These isopleths are necessarily the result of interpolating data between fixed stations where these parameters were measured.
- Provide a tool that would facilitate future decision-making regarding optimal station locations. This tool could support decisions to add or remove stations.

Kriging is the preferred statistical tool for supporting the analysis, and is, in effect, the only reasonable approach. Kriging is the only standard spatial interpolation technique that provides a statistically meaningful estimate of the uncertainty in interpolated values. This is the critical feature of kriging that makes it ideal for this analysis. Kriging also has other features/benefits that may be valuable to EPA:

- Many users find that kriging has the capacity to produce smoothed contours that are aesthetically pleasing and tend to "look like" subjective hand-drawn isopleths
- Kriging can specifically account for anisotropy if present in the data. Anisotropy could occur if concentrations are better correlated along a prevailing wind direction and less correlated transverse to the prevailing wind direction.
-

GEO-EAS was used to conduct preliminary analyses. GEO-EAS is EPA-supported public domain software (EPA. 1991. GEO-EAS 1.2.1 User's Guide. EPA/600/8-91/008) and has the required features to support this analysis.

Recommended Analytical Approach

Several, though not all, of the steps in this process were preliminarily tested using SO2 concentrations from 46 stations in the northeastern United States in the 4th quarter of 2000. Lessons learned from the preliminary testing are incorporated into the recommended approach. Findings specifically related to the test-case (4th quarter 2000 SO_2 concentrations) are highlighted with bold text.

The recommended approach closely follows standard methods for statistical evaluation of data, in general, and standard methods and guidance for applications of kriging, in particular. The GEO-EAS User's Guide (EPA, 1991) is a good example of such guidance. The first step is a general examination of the data set. The data set should be examined to determine appropriate distributional assumptions. Kriging is a parametric statistical method, which relies, in some steps, on assumptions of normality. It is commonly observed that environmental concentration data follow a lognormal distribution. The 4th qtr 2000 SO_2 concentrations were reviewed, station-by-station, using the Shapiro and Wilk W test. These tests indicated that, by station and within a quarter, SO_2 concentrations are lognormal. Further, the uncertainty in central tendency estimates at stations (both means and geometric means) shows generally constant relative standard deviations. Absolute standard deviations, on the other hand, are proportional to the mean concentration. This is a typical characteristic of lognormal data. Parametric statistical calculations should be performed on logarithms of the data, rather than the absolute values. If CASTNET data consistently follow a lognormal distribution, as expected both theoretically and from experience, it may be acceptable to assume the data is lognormal without testing of each data set. If so, kriging analyses should be performed on logarithms of the data values. The resulting values should not be subsequently used in any interpretive analysis or calculation where an average or integral should be used, such as estimation of deposition rates. The geometric means or relative standard deviations produced from lognormal distributions are not accurate measures of time-weighted averages or area-weighted fluxes that would be required for such analyses.

The time interval over which the analysis should be performed is the one that will produce the most reliable spatial pattern. This can be defined as the time interval that produces the smallest standard deviation of the station means of the logarithms of the data. If there were no seasonal variations or long term trends, the entire data set should be used because the large sample size would produce the least uncertainty in the means at each station. On the other hand, strong seasonal variation or trends could lead to actually more variance and uncertainty. This should be tested. Choose either quarters, ½ year (e.g., October through March) or 1 year depending on which yields smallest uncertainty in mean concentration.

Follow GEO-EAS guidance and develop the variogram model. Evaluate potential anisotropy. The 4th quarter 2000 SO_2 results were evaluated with locations specified by latitude and longitude. Logarithms of 4th quarter 2000 SO_2 data were adequately modeled using a Gaussian model with a nugget of 0.1, a sill of

1.4, and a range of 18. The variogram exhibited anisotropy, with a larger range in the east/west direction, implying better correlation between points far apart in the east-west direction, but less correlation if far apart in a north-south direction. Consequently the data were kriged using a Gaussian variogram, a nugget of 0.1, a sill of 1.4, and a range of 22 in the east-west

direction, but a range of 12 in the north-south direction. The east-west range of 22 implies that data were no longer correlated if they were separated by 22 units of longitude. The north-south range implies that data were no longer correlated if they were separated by 12 units of latitude.

The best fit nugget of 0.1 is meaningful and consistent with other characteristics of the data set. The average standard deviation of the mean of the log-transformed data was 0.07. The nugget represents the uncertainty in each data point, which may be due to measurement error. The fact that the nugget is similar to the standard deviation of the typical station means is consistent with this concept.

Figure 1 illustrates the kriged interpolation of the geometric mean concentrations of SO_2 during the 4th quarter of 2000.

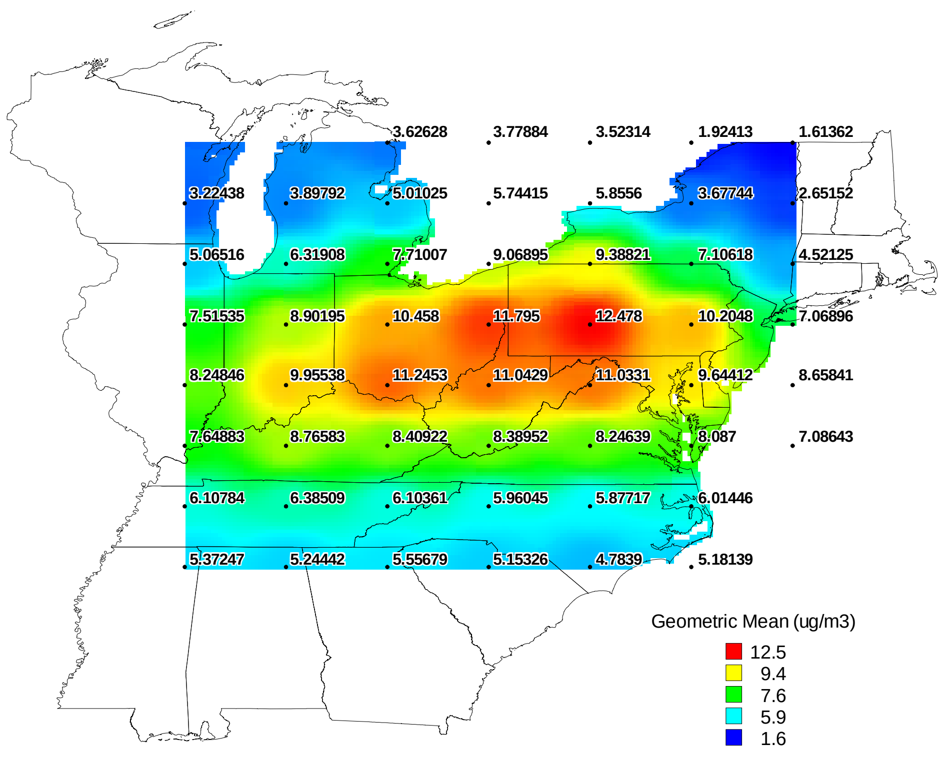


Figure *1*: Kriging Estimated SO_2 Concentration (µg/m^3), 4th Quarter 2000

If the "best" variogram model appears to vary from data set to data set, it may be advisable to select the "best" model in accordance with GEO-EAS guidance, but also, as a sensitivity test, examine all data sets with the same model type, choosing the model that best fits most of the data sets. Variogram model selection is partially an art, and the results can vary somewhat according to decisions made during the course of the application. For this reason verification tests are recommended:

- Compare model predicted values (using ordinary or "block" kriging) with the values observed at the station locations. Compute the root mean square deviation between observed and model interpolated results. This calculation should be done using the log-transformed values. This is analogous to examining the residuals during regression analysis. The residual should be randomly distributed within the model domain, without pattern. If the deviations are similar to, or greater than, the area-weighted average of the kriging relative standard deviation (AWA KRSD) it would imply that the kriging model is not as reliable as might be indicated by the kriging standard deviation alone.

- Cross-validation – a standard test of a kriging model – sequentially eliminate each data point from the data base and compare the interpolated value to the missing data point. Compute the root mean square deviation between observed and cross-validated interpolated results. Cross validation is more rigorous than the simple comparison of observed and interpolated results described in paragraph (a) but is substantially more resource-intensive (the kriging analysis must be repeated as many times as there are data points). Consequently, this technique should be applied to one representative data set, and the results interpreted. On the other hand method (a) should be performed for each data set evaluated.

Identification of a Metric for Characterizing Data Quality

The kriging program produces interpolated estimates of the input variable (in the example case, the logarithm of the concentrations) and an estimate of the uncertainty in this parameter, referred to as the kriging standard deviation. In the subject case, this parameter is also in logarithmic (base e) units. Several potential metrics are suggested. First convert the kriging standard deviation from logarithmic units by taking its exponential. For example, if the kriging standard deviation is 0.2 in logarithmic units, then $\exp(0.2) = 1.22$, implying a relative standard deviation (RSD) of 22%. This will be referred to as the kriging relative standard deviation (KRSD). Two metrics are suggested:

(a) the AWA KRSD over the model domain (the network), and

(b) the maximum kriging relative standard deviation (MAX KRSD) within the model domain.

For the example case, the AWA KRSD over the model domain was 20%. The MAX KRSD was 40% and that value occurred near the edge of the network, at Alpena, MI. Relatively large KRSDs will always exist at the boundaries of the network, simply because they are at the boundary and there is no data outside network to “bound” the estimates. These types of errors must be tolerated unless a high density of stations were established along the boundary. This is probably not cost-effective. The KRSDs are shown in Figure 2.

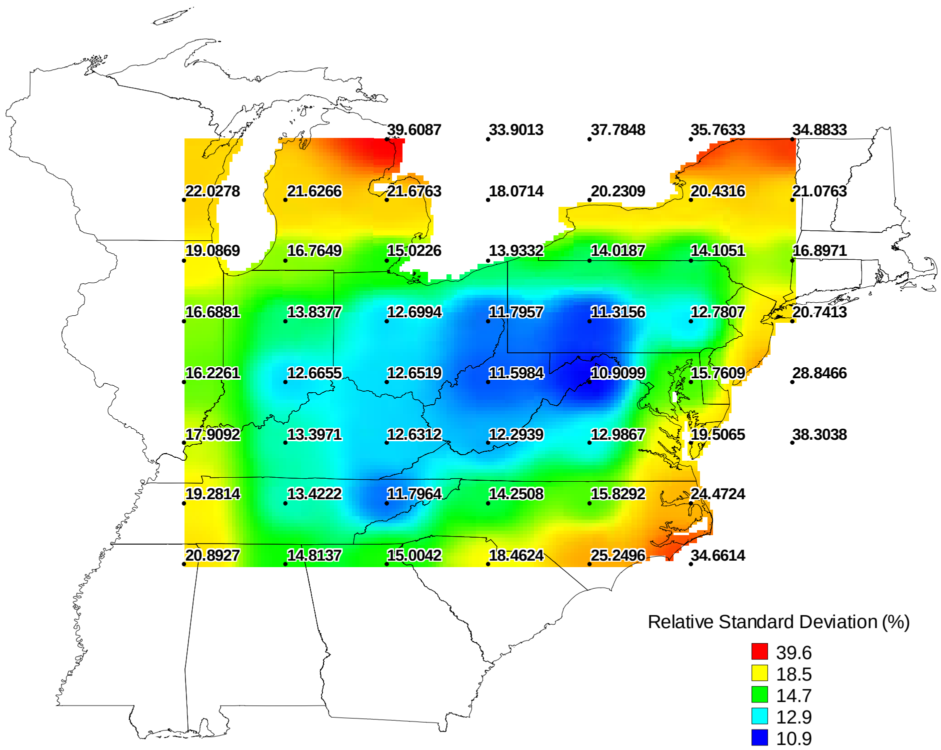


Figure *2*: Kriging Relative Standard Deviation (%),SO_2, 4th Quarter 2000

A second meaningful metric is the kriging absolute standard deviation, KASD, which would be calculated by multiplying the geometric mean by the KRSD. For the test case, results of this calculation are illustrated on Figure 3. The maximum kriging absolute standard deviation (MAX KASD) is approximately 1.6 $\mu g/m^3$ near Atlantic City, NJ, and Cape Hatteras, NC. As was the case with KRSDs, large KASDs occur near the boundary. A large KASD at the boundary is a greater problem than a large KRSD and possible corrective actions may be considered. This could lead to greater density of stations near the boundary, but only in areas with high absolute concentrations. The area-weighted average kriging absolute standard deviation (AWA KASD) over the northeastern United States was 1.1 $\mu g/m^3$.

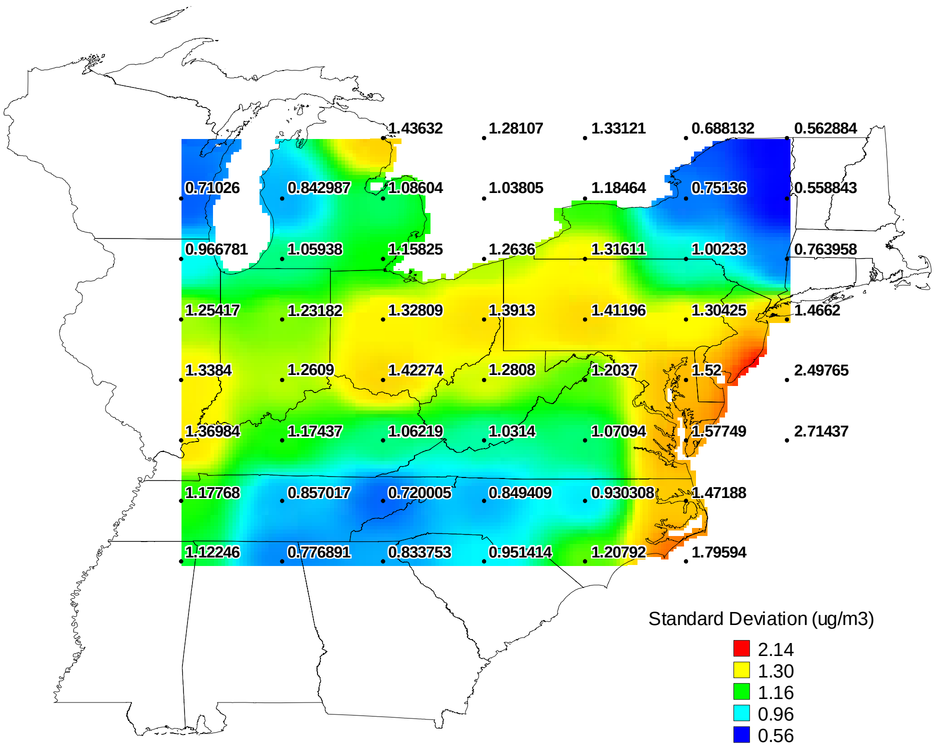


Figure 3: Kriging Absolute Standard Deviation ($\mu g/m^3$), SO_2 4th Quarter 2000

These metrics appear suitable for characterizing the success of the network in defining spatial patterns.

Additional Applications/Benefits of the Kriging Technology

The primary management benefit of performing the kriging analyses would be to provide an objective method to evaluate cost/benefit of alternative network configurations. Existing stations could be prioritized by quantifying the effect that their removal would have on the spatial pattern metric (e.g., removal of station X would increase the MAX KASD from 1.6 μg/m^3 to 1.8 μg/m^3, and the AWA KASD from 1.10 μg/m^3 to 1.16 μg/m^3; while removal of station Y might have a negligible effect. Consequently station Y is more easily sacrificed than station X). Likewise optimum locations of new stations could be identified by identifying points within the domain with large absolute or relative kriging standard deviations.

Does the Current Network Adequately Define Spatial Patterns?

The current analysis provides metrics that can be used in answering this question.

The kriged estimates (shown in Figure 1) represent the best estimate geometric mean on the map at any location without a station. In effect, this describes what you know at places where you did not make a measurement. These exhibit a range from high (12.5 μg/m^3) to low (1.6 μg/m^3) of 10.9 μg/m^3. The AWA KASD is 1.1 ug/m^3. So the error is only 10% of the apparent range. The corollary to this result is that, with statistical confidence, 90% of the apparent spatial pattern is real. Consequently it can be concluded that the apparent spatial pattern is an actual spatial pattern. This quantity is a reasonable measure of the validity of apparent spatial patterns. The formula would be:

$$\text{Spatial Pattern Validity} = 1 - \left(\frac{\text{AWA KASD}}{\text{Max Krige Geometric Mean} - \text{Min Krige Geometric Mean}} \right) = 1 - \left(\frac{1.1}{12.5 - 1.6} \right) = 0.90 = 90\ \%$$

This formula produces a non-dimensional result that can be computed and readily compared across all air quality parameters.

Identification of a Data Quality Objective that Can be Applied to other Parameters and Averaging Times

As previously discussed, the Spatial Pattern DQO should be based on the most accurate definition of the spatial pattern, and should not depend on time. The most reliable definition of the spatial pattern will be achieved when the averaging time is selected so as to produce the smallest intrastation uncertainty in means, specifically, when the standard deviation of the station means of the logarithms of the data within that averaging period is minimal. This is calculated by:

- Take the logarithms of the reported values.
- Determine the standard deviation of the resulting logarithms.

- Divide the standard deviation of the logarithms by the square root of N, the number of samples in the averaging period.

- Identify the averaging period for which the resulting quantity is minimal, across the grid (each station has a resulting standard deviation of the mean – select an averaging period for which the average and/or maximum of these values is at a minimum).

This process was tested for SO_2 concentrations measured during the period October 1999 through September 2000. Three averaging periods were tested, quarters, half years (October through March), and the full year. It was found that the standard deviations of the logarithms (result of step 2) were steady when the averaging period was increased from the quarter to the half year, but increased significantly over the full year. The result of step 3 (after dividing by the square root of N) was that the full year produced a more reliable estimate of intrastation means (lower standard deviation of the means). Based on these results either the half-year (Summer vs. Winter) or the full year should be selected as the appropriate averaging period for SO_2.

In contrast to the test case, it is recommended that the kriging process be applied to all the available data, characterizing the entire network, rather than just the northeastern United States.

After performing the kriging analysis for the appropriate averaging period, determine whether the maximum interpolated logarithmic mean is significantly greater than the minimum interpolated logarithmic mean, using the t test. Recognize that all variances entering into this analysis are actually standard deviation of the means rather than standard deviations of the population of values.

For the 4th quarter 2000 SO_2, the maximum interpolated logarithmic mean is 2.524 (corresponding to the maximum interpolated geometric mean of 12.48, i.e., 12.48 = e2.524). The standard deviation of this logarithmic mean is 0.107 (corresponding to the KRSD of 11.3%, i.e., 0.113 = e0.107 – 1). The minimum interpolated logarithmic mean on the network is 0.478 (corresponding to the minimum interpolated geometric mean of 1.61), and its standard deviation is 0.299 (corresponding to its KRSD of 34.9%). The test is whether the maximum is significantly greater than the minimum, a one-tailed test. The test statistic is (2.524 – 0.478) ÷ (0.1072 + 0.2992)1/2 = 2.046 ÷ 0.318 = 6.434. This implies that the high concentration is greater than the low concentration with more than 99.9% confidence, and the spatial pattern is real.

Although it is more appropriate to perform this calculation on the logarithms, because the underlying data is lognormal, a similar result is obtained if the geometric means and the KASDs are used, and this may be helpful to illustrate the idea. Expressed in concentration units, the maximum interpolated value is 12.48 ± 1.41, the minimum interpolated value is 1.61 ± 0.56. The high value is significantly greater than the low value. If these results were characterizing a normal variate, the test statistic would be (12.48 – 1.61) ÷ (1.412 + 0.562)1/2 = 10.87 ÷ 1.52 = 7.15, i.e., the test statistic by this inappropriate method is practically the same as the test statistic calculated from the lognormal distribution. In either case the maximum is significantly greater than the minimum, implying that the observed spatial pattern is real.

There may be parameters or averaging periods that exhibit less pronounced spatial patterns than the test data set. It is assumed that parameters that have less than a factor of 2 between the high and low interpolated values, have a negligible spatial pattern. It is not necessary to accurately and definitively define the spatial pattern for parameters with such negligible spatial patterns.

The recommended Spatial Pattern DQO following this procedure can be stated as:

> Where the maximum interpolated value within the network exceeds the minimum interpolated value by a factor of 2, the difference is statistically significant with 90 % confidence. The quantitative test is that the maximum interpolated value will be shown to be greater than the minimum interpolated value with 90 % confidence. Specifically, this means that the test statistic must exceed 1.28.

Identification of Locations where Additional Stations Should be Located

The above-described DQO characterizes whether apparent spatial patterns are statistically significant, looking at the network footprint as a whole. It is not useful for identifying local areas where interpolation errors are unacceptably large.

A simple criterion would be that the KRSD should not exceed a specified value. For example, if the KRSD exceeds 52%, then the interpolated geometric mean is not reliable, with 90 % confidence, to within a factor of 2. This result is obtained from the logarithmic standard deviations as follows:

- KRSD = 52%, implies that the kriging standard deviation in natural logarithm units is ln(1.52) = 0.418;
- To be 90% confident that the true concentration is within a factor of 2 of the interpolated estimate, then $\ln(2) \div 0.418 > 1.645$ (this is tcritical for a two-tailed test, with infinite degrees of freedom; the approximation at infinite degrees of freedom is appropriate because the number of data that were used to estimate the interpolated value is very large).
- $\ln(2) \div 0.418 = 0.693 \div 0.418 = 1.66$, which is greater than, but approximately equal to, 1.645.

In the test case, the MAX KRSD = 39.6%, so all interpolated values are accurate to within a factor of 2, with 90% confidence.

This statistically simple test, however, may not appropriately represent the needs of EPA and other stakeholders. For example, an area with very low concentrations could have a high KRSD without compromising the utility of the network. One may still be highly confident, in such a case, that concentrations are well below thresholds expected to cause adverse effects. Additional stations in such areas would not be warranted.

Consider a criterion based on absolute errors. When the KASD anywhere in the network approaches the maximum geometric mean within the network, then the local error is large relative to any apparent spatial pattern. In the test case, the MAX KASD was 1.6 $\mu g/m^3$ (on the mid-Atlantic seaboard) while the maximum geometric mean was 12.5 $\mu g/m^3$ in central Pennsylvania. The MAX KASD is 13% of the maximum interpolated concentration. Wherever the KASD exceeds 25% of the maximum geometric mean, the error at that location is presumed to compromise the validity of the apparent spatial pattern, and additional stations at or near such locations should be considered. This presumptive and preliminary DQO is defined as:

> MAX KASD anywhere within the network should not exceed 25% of the maximum geometric mean within the network

As previously discussed, the statistical significance of this criterion is not obvious, so it is a subjectively defined criterion. Further experience with statistical evaluation of the network could lead to modifying the numerical value associated with this criterion. Nonetheless, areas with high KASD should be priorities for siting of new stations.