final

Source Region Identification Using Kernel Smoothing

Ronald Henry[a], Gary A. Norris[b], Ram Vedantham[b], Jay R. Turner[c]*

[a]Department of Civil & Environmental Engineering, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA, 90089-2531 USA

[b]U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park NC 27711

[c]Energy, Environmental and Chemical Engineering Department, Washington University, Campus Box 1180, One Brookings Drive, St. Louis, MO 63130-4899 USA

*Communicating Author

Resubmitted to *Environmental Science & Technology*, November 21, 2008

Abstract

As described in this paper, Nonparametric Wind Regression is a source-to-receptor source apportionment model that can be used to identify and quantify the impact of possible source regions of pollutants as defined by wind direction sectors. It is described in detail with an example of its application to SO_2 data from East St. Louis, IL. The model uses nonparametric kernel smoothing methods to apportion the observed average concentration of a pollutant to sectors defined by ranges of wind direction and speed. Formulae are given for the uncertainty of all of the important components of the model, and these are found to give nearly the same uncertainties as blocked bootstrap estimates of uncertainty. The model was applied to data for the first quarter (January, February, and March) of 2003, 2004, and 2005. The results for East St. Louis show that almost 50 percent of the average SO_2 concentration can be apportioned to two 30° wide wind sectors containing a zinc smelter and a brewery; a nearby steel mill did not appear to have a significant impact on SO_2 during this period.

Keywords: Air pollution, source apportionment, statistics, receptor model, nonparametric regression, St. Louis, East St. Louis, sulfur dioxide, zinc smelter, brewery

Introduction

Traditional source-oriented air quality models calculate the impact of sources using emissions inventories, meteorological data, chemical reactions, and dispersion modeling. In contrast, receptor-oriented models, such as the chemical mass balance or multivariate receptor models, apportion the contributions of various air pollution sources based on chemical composition profiles of sources and observed composition of pollutant samples collected at monitoring site. A new hybrid source-receptor model is described that seeks to locate and quantify local sources of air pollution through nonparametric regression of 1-hour average atmospheric concentrations of a pollutant on hourly resultant of wind speed and direction. For this reason, it is called Nonparametric Wind Regression (NWR). The NWR method described here can, under appropriate assumptions described below, apportion source regions of a pollutant to multiple sources without use of chemical fingerprints or emissions inventories. Although this is demonstrated using routine monitoring data for a primary pollutant SO_2 from East St. Louis, Illinois, USA, the model is also applicable to more general source types (such as a combination of multiple primary pollutants). It is also worth noting that this scheme can use higher resolution data (for instance, 5-minute date) with the only requirement that the meteorological data associated with the data be similarly highly resolved.

A previous paper introduced the use of nonparametric regression of a pollutant concentration on wind direction to accurately estimate the location of the source [1]. A later paper demonstrated that additional information on the location and nature of a source can be determined by nonparametric regression using both wind speed and direction [2]. The only other method at all comparable to NWR is the Conditional Probability Function (CPF) approach. A comparison of CPF and the earlier nonparametric methods mentioned above is found in [3].

The previous work was essentially qualitative, showing the location and relative strength of local sources. This paper develops a quantitative model that can estimate the weighted pollutant concentration in a range of wind directions and wind speeds and the uncertainty of this estimate.

Nonparametric Wind Regression Methodology

The first step in NWR is to calculate the expected value of the pollutant as a function of wind speed and direction using standard nonparametric regression formulae [2]. In this paper the dependent variable is the hourly averaged concentration of a pollutant C at the receptor site and the predictor variables are the hourly resultant wind direction θ and wind speed u. The average concentration of a pollutant for a particular wind speed and direction pair (θ, u) is calculated as a weighted average of the concentration data in a window around (θ, u) represented by smoothing parameters σ and h using a weighting function $K(\theta, u, \sigma, h) = K_1(\theta, \sigma)K_2(u,h)$, also known as the kernel function. The expected value of C given θ and u is estimated by:

$$E(C \mid \theta, u) = \frac{\sum_{i=1}^{N} K_1\left(\frac{(\theta - W_i)}{\sigma}\right) K_2\left(\frac{(u - U_i)}{h}\right) C_i}{\sum_{i=1}^{N} K_1\left(\frac{(\theta - W_i)}{\sigma}\right) K_2\left(\frac{(u - U_i)}{h}\right)} \tag{1}$$

where C_i U_i, and W_i are the observed concentration of a particular pollutant, resultant wind speed and direction, respectively, for the i-th observation in a time period starting at time t_i; N is the total number of observations.

The simplest kernel function is a constant defined as 1 between -1/2 and 1/2 and zero elsewhere. In this case, the denominator in Eq. 1 is just equal to the number of data points in the intervals given by the smoothing parameters, and Eq. 1 just reduces to a simple moving average. The disadvantage of a simple moving average is the lack any significant smoothing, which usually makes it difficult to determine the general wind direction and speed of significant peaks in the concentration values. This can hinder identification of local sources [1]. Two well-known choices for the kernel functions, and the ones used in this work, are the Gaussian kernel given by

$$K_1(x) = (2\pi)^{-1/2} \exp(-0.5x^2), \quad -\infty < x < \infty \tag{2}$$

and the Epanechnikov kernel

$$\begin{aligned} K_2(x) &= 0.75(1 - x^2), && -1 < x < 1 \\ &= 0 && \text{otherwise.} \end{aligned} \tag{3}$$

The Gaussian kernel is used for wind direction since it is defined over an unbounded range, i.e. a circular range. The Epanechnikov kernel is used for wind speed because it is the simplest bounded kernel with a central peak. Other choices than the Gaussian and Epanechnikov kernel functions are possible, but this article is intended as demonstration of an idea using kernel functions. Also, there is no evidence in the applied statistical literature that the results are particularly sensitive to the choice of the kernel function. In fact, our experience is that the results are insensitive to the choice.

To apportion the weighted concentration to source areas, the above results must be weighted by the frequency of the winds. Thus, the second step in the NWR model is to calculate the empirical joint probability density of wind speed and direction using the kernel density estimate [4]

$$f(\theta,u)=\frac{1}{N\sigma h}\sum_{i=1}^{N}K_1\left(\frac{(\theta-W_i)}{\sigma}\right)K_2\left(\frac{(u-U_i)}{h}\right) \tag{4}$$

The final step is to estimate the fraction of the weighted pollutant concentration associated with wind speed u in the closed interval $U = [u_1\ u_2]$ and wind direction in the interval $\Theta = [\theta_1\ \theta_2]$ by multiplying $f(\theta,\ u)$ and $E(C|\theta,u)$ and integrating over the appropriate ranges of wind speed of u and a wind direction of θ:

$$S(\Theta,U)=\int_{u_1}^{u_2}\int_{\theta_1}^{\theta_2} f(\theta,u)E(C\mid\theta,u)\,d\theta\, du \tag{5}$$

where $S(\Theta,U)$ is the mean value of the pollutant concentration associated with winds from the sector defined by the intervals U and Θ, which will be referred to as the sector apportionment. If the winds from the intervals U and Θ are associated with the impact of a particular source, then the sector apportionment $S(\Theta,U)$ is an estimate of the average concentration associated with that source; dividing by the average value of the entire data set gives the fraction of the average pollutant concentration associated with winds from intervals U and Θ.

From equations 1 and 4, $S(U, \Theta)$ is estimated from the data by

$$S(\Theta,U) = \frac{\Delta u \Delta \theta}{N\sigma h} \sum_{\theta_k=\theta_1}^{\theta_2} \sum_{u_j=u_1}^{u_2} \sum_{i=1}^{N} K_1\left(\frac{(\theta_k - W_i)}{\sigma}\right) K_2\left(\frac{(u_j - U_i)}{h}\right) C_i \,. \tag{6}$$

Where wind direction is quantized as equally spaced values θ_i from 0 to 360 with spacing $\Delta\theta$, and wind speeds u_i equally spaced values between u_{min} and u_{max} with spacing Δu. Sometimes it is more convenient to work with Cartesian coordinates instead of polar coordinates. If the positive x-axis is east, then the wind speed and direction are replaced with x,y coordinates by

$$\begin{aligned} x &= u \cos\phi, \\ y &= u \sin\phi, \\ \phi &= \pi \operatorname{mod}(450 - \theta, 360)/180 \end{aligned} \tag{7}$$

where the last equation converts the azimuth angle in degrees clockwise from north to the mathematical angle ϕ in radians counterclockwise from +x-axis; here *mod(a,b)* is *a* modulo *b*. In this case, since both predictors are bounded, K_2 in the above equations is taken to be the Epanechnikov kernel rather than the Gaussian kernel, otherwise the equations are the same as above with θ and u replaced by x and y

Uncertainty Estimates

As shown by example later, estimates of the effects of random error can be particularly useful in NWR. Oftentimes the receptor site might be impacted by abnormal values in all three dimensions: wind speed, wind direction and pollutant concentration. A methodology is needed to assess the likelihood that the NWR results for these abnormal events could be real or be attributable to chance.

The variance of $E(C \mid \theta, u)$ is approximated by [2]

$$\operatorname{var}(E(C \mid \theta,u)) = \frac{c_{K_1} c_{K_2} \sum_{i=1}^{N} K_1\left(\frac{(\theta - W_i)}{\sigma}\right) K_2\left(\frac{(u-U_i)}{h}\right)(C_i - E(C \mid \theta,u))^2}{\left[\sum_{i=1}^{N} K_1\left(\frac{(\theta - W_i)}{\sigma}\right) K_2\left(\frac{(u-U_i)}{h}\right)\right]^2}, \qquad (8)$$

where

$$c_{K_i} = \int K_i^2(x)dx, \quad i = 1,2.$$

The variance of $f(\theta,u)$ is estimated as [3]

$$\operatorname{var}(f(\theta,u)) = \frac{c_{K_1} c_{K_2}}{Nh\sigma} f(\theta,u) = \frac{c_{K_1} c_{K_2}}{(Nh\sigma)^2} \sum_{i=1}^{N} K_1\left(\frac{(\theta - W_i)}{\sigma}\right) K_2\left(\frac{(u-U_i)}{h}\right). \qquad (9)$$

By standard propagation of errors, the variance of the product of the two is approximately

$$\begin{aligned} s^2(\theta,u) &= \operatorname{var}(f(\theta,u)E(C \mid \theta,u)) \\ &= \operatorname{var}(f(\theta,u))E(C \mid \theta,u)^2 + \operatorname{var}(E(C \mid \theta,u))f(\theta,u)^2 \\ &+\left(r^2 \operatorname{var}(f(\theta,u)) \operatorname{var}(E(C \mid \theta,u))\right)^{1/2} \end{aligned} \qquad (10)$$

where r^2 in the final term is the correlation of f and $E(C \mid \theta,u)$. In many cases, as in the example below, r^2 is close to 0 and the final term may be ignored.

Finally, the uncertainty in the sector apportionment $S(\Theta, U)$ is estimated by

$$\operatorname{var}(S(\Theta,U)) = \sum_{\theta_k,\theta_m \in \Theta,\, u_k,u_n \in U} \sum s(\theta_k,u_j)s(\theta_m,u_n)\exp\left(-\frac{1}{2}\left(\frac{\theta_k - \theta_m}{\sigma}\right)^2\right). \qquad (11)$$

The sum in the above equation is taken over all pairs of (θ_k, u_j) and (θ_m, u_n) that are in intervals Θ and U. This last equation is derived from the fact that the variance of a sum of correlated random variables is the sum of all the elements of the covariance matrix. In this case, the covariance is estimated to be the product of the standard deviations times the correlation coefficient. The correlation coefficient between the nonparametric regression estimates is primarily a function of wind direction and is approximated by a Gaussian function with the same smoothing parameter as used in the nonparametric regression. The equation does not have a similar term for wind speed u since concentrations are not highly correlated with wind speed.

All these uncertainty estimates are based on assumptions concerning the independence and distribution of the errors that are not always met by real world data; nonetheless they are usually a realistic measure of the effects of random error. The estimates of these formulae are compared to bootstrap estimates after presentation of the results for St. Louis.

Application to East St. Louis SO_2 Data

Hourly data for criteria pollutants at several sites in the greater St. Louis metropolitan area were retrieved from the Environmental Protection Agency's (EPA) Air Quality System (AQS) database. The data covered the entire period of 2003 through the end of 2005. Where available, hourly resultant wind speed, direction and other meteorological parameters were retrieved for the same period. This paper will focus on SO_2 and meteorological data collected at the 13th and Tudor site in East St. Louis, Illinois. Hourly compliance monitoring data for SO_2 are routinely collected by the Illinois EPA; hourly surface meteorology data (10 m height) were collected under the EPA-funded St. Louis -

Midwest Supersite program. The East St. Louis site is located 3 km east of the City of St. Louis, MO central business district and is separated from it by the Mississippi River. There are several major industrial zones along the river including Sauget to the south/southwest and Granite City to the north. The area is also a major national transportation hub with a large rail yard 2.5 km to the southeast and many nearby heavily traveled limited access highways, bridges, and barge traffic on the river.

The analysis was restricted to the first quarter of years 2003, 2004, and 2005 since the meteorological conditions during a single season are fairly consistent and because this period exhibits advective conditions (the lowest frequency of calm winds) and significant variability in the wind direction. Another reason the first quarter was chosen is because the concentrations of SO_2 are consistently high, as would be expected for a pollutant dominated by local sources during the winter when wind speeds are moderate and mixing heights tend to be lower than other times of the year. The basic statistics of SO_2 for this period are given in Table 1. In the following analysis, only SO_2 values greater or equal to 1 ppb were used; there are 821 values less than this. The following parts of this section present and discuss the SO_2 emissions inventory, distribution of wind speed and direction, the nonparametric regression of SO_2 as a function of wind speed and direction, and the apportionment of the average SO_2 for the period to nearby sources.

Sources of SO_2

Large sources of SO2 in the area are listed in Table 2 and shown in . These are taken from the National Emissions Inventory for 2002 compiled by the US EPA (**http://www.epa.gov/ttn/chief/net/2002inventory.html**). There are three major sources within 10 km of the East St. Louis monitoring site: a zinc smelter, a brewery, and a steel mill. The distance and directions to these from the monitor are given in the table.. The zinc smelter is the closest and the direction to the whole facility ranges between 200° to 225° azimuth. There are other SO_2 sources in this wind sector at various distances from the site, but the smelter is expected to be the most significant contributor to SO_2 at the East St. Louis site due to its close proximity and since it also has a sulfuric acid manufacturing plant. There are also five other large SO_2 sources listed in the table, two oil refineries and three coal-fired power plants. However, these are 25 to 35 km distant, too far to produce high concentrations of SO_2 at the monitoring site. Since the wind and concentration data are 1-hour averages, local sources with travel times less than 1 hour will be most amenable to NWR analysis.

Joint Probability Distribution of Wind Speed and Direction

A wind rose is the traditional way of looking at the joint distribution of wind direction and wind speed. A wind rose of the 1-hr resultant wind speed and direction for the 6504 hours of the 9 first quarter months from 2003 to 2005 is shown in Figure 2. The wind direction bins are 10° wide. A much more detailed view is given by the joint probability distribution of wind speed and direction given in Figure 3. The distribution was estimated by Eq. 4 with a Gaussian kernel for the wind direction with a Full Width at Half Maximum (FWHM) of the of 10 degrees and the FWHM for the Epanechnikov kernel for wind speed being 1.5 kmh^{-1}, with a maximum wind speed of 15 kmh^{-1}. The area encompassed by the white line is the contour of a signal-to-noise ratio of 2, where the variance of the noise is calculated by Eq. 9; areas encompassed by this contour have too few data points to reliably estimate a value for the density.

The distribution has a number of features that impact the nonparametric regression analysis. The predominant winds are from the southeast between 5 and 15 kmh^{-1}; thus, even if there are no large sources in this direction, a relatively large fraction of the average pollutant concentration may be explained by winds from the southeast. Northerly winds are the next most frequent with wind speeds between 5 and 15 km/hr being prevalent. Winds from the northwest exhibit very few hours with wind speeds less than 10 km/hr, which is consistent with occasional strong advection from the northwest in the wintertime. The paucity of lower wind speeds from the northwest is shown by the white contour in Figure 3, which shows that the signal-to-noise ratio of the distribution estimate in this region is less than 2, indicating that there are too few points to estimate the distribution accurately. Another anomaly in the distribution that will be important in the later discussion is in the direction of 225 degrees azimuth, the approximate direction of a nearby zinc smelter. In this direction wind speeds of 5 to 10 km/hr are much more frequent than other wind speeds, particularly low wind speeds. This is also seen in the wind rose but not nearly so clearly.

Nonparametric Wind Regression of SO_2

A pollution rose is the traditional way of looking at the relationship of wind direction and pollutant concentrations. The SO_2 pollution rose with 10° bins is shown in Figure 4. From this plot, the prevalence of high concentrations in the region of 200° to 250° is clear, as well as the fact that winds from this direction are not common. However, a much more detailed picture of the relationship of SO_2 and wind direction *and* speed emerges in the NWR plot of SO_2 seen in Figure 5. The white areas in the plot are regions where the signal-to-noise ratio as estimated by Eq. 10 is less than 2, i.e., there are too few points for a reliable estimate. The highest concentrations lie between 200° and 255° azimuth with the highest peak at 221° and wind speed of 10 km/hr, the next highest peak at 221° and 12 km/hr, another peak at 213° and 7 km/hr, and finally a smaller peak at 247° and 12 km/hr. The direction of the peaks at 221° and 213° is consistent with the zinc smelter and its sulfuric acid plant and the small peak at 247° may be associated with the brewery.

Is the peak at 247° really a peak or is it perhaps a statistical variation caused by a few errant, large values? Fortunately, this question can be addressed quantitatively since the NWR produces estimates of the uncertainties, as discussed in detail in the previous section. In this case, the question is restated as: is the difference between the peak and the surrounding background level large compared to the uncertainties in these values. The variance of the difference of two random values is given by

$$\mathrm{var}(X-Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2Cov(X,Y) \qquad (12)$$

In this case X is 30.45 ppb the peak value at azimuth angle 247° with a standard error of 7.17 ppb. Y is 16.9 ppb; the minimum value between this peak and the larger peak at 221° and it has an estimated standard error of 4.93 ppb. The difference X-Y is 13.55 ppb. The uncertainty in this difference is given by $[7.17^2 + 4.93^2 - 2(7.17)(4.93)\exp(-0.5((247-234)/4.2)^2)]^{1/2} = 8.67$. The exponential term estimates the covariance of the two estimates, as explained after Eq. 11; 4.2 is the sigma corresponding to a Gaussian with a full width at half maximum of 15 degrees. The ratio of the difference to the uncertainty is 1.56. This statistic is approximately normally distributed with mean 0 and standard error 1. From the inverse cumulative normal function, the probability of getting a value this large or larger by random variation is about 6 percent. Putting it another way, the peak is different from the background with a confidence level of about 94 percent. In conclusion, the peak at 247° is very probably real and related to the major source in that direction, the brewery.

Sector Source Apportionment

The fraction of the average SO_2 coming from a sector 30 degrees wide and centered on each degree of azimuth is shown in Figure 6 as calculated using the formulae given in the first part of this paper. The 2-sigma error bars are also given. 30° wide sectors were used since 15° is approximately the median value of the standard deviation of the wind direction in 1 hour. In addition to the expected peak in the direction of the smelter, there are lesser peaks to the south and north because the winds frequently come from these directions.

Table 3 gives the sector apportionment for 30° wide sectors centered on the direction to the three nearest major sources. Assigning all the SO_2 from a sector to the source, the smelter accounts for almost 40 percent of the average SO_2 for the period, even though the winds only came from that direction about 7 percent of the time. Also, the average SO_2 concentration for winds from this sector is 20 ppb, much greater than the 4.6 ppb overall average. The winds came from the brewery sector less than 5 percent of the time, but this sector accounted for almost 10 percent of the average SO_2 and an average SO_2 of 10.3 ppb. These are all signs that the known sources of SO_2 in these sectors are impacting the site. The surprise is that there are no high concentrations in the NWR plot seen in the direction of the steel mill. There is nothing in Table 3 to indicate that the sector containing the steel mill has a SO_2 source that is influencing the site. Winds are quite frequent from this direction and yet it explains little of the average SO_2 and the average concentration with winds from this sector is 3.8 ppb, which less than the overall average of 4.6 ppb. Thus, there is no evidence that the steel mill SO_2 emissions have a noticeable effect on the site. Altogether winds from the direction of the two major local sources account for 47.0 ± 11.8 percent or almost half of the average SO_2.

Overall, the NWR method is superior to the pollution rose in identifying nearby sources in that it is not sensitive to the selection of the number and width of the wind direction bins. It is better at separating sources that lie in the same general direction, as with the brewery and zinc smelter sources. The existence of uncertainty estimates for all the parts of the method is important; for example, these were used to show that the peak associated with the brewery is a real peak and not an artifact. While the NWR has many more strengths than the pollution rose, the two have similar weaknesses. The primary one being not a weakness of the methods but of the data; if there are nearby obstructions to the air flow or if for some other reason the wind data not representative of local transport, then both methods will produce erroneous results.

Bootstrap Uncertainty Estimates

Table 3 gives uncertainties of the sector contributions calculated by Eq. 11. However, these estimates may not be reliable as the assumptions inherent in these formulae may not all be valid in this application. Furthermore, the wind speed, direction, and concentrations all have significant positive serial correlation caused by 24-hour diurnal periodicities that may result in under estimation of the uncertainties. The reliability of the calculated uncertainties can be tested by comparison with uncertainties independently estimated by bootstrap methods. However, the serial correlation in the data must be preserved by the bootstrap samples. This was done in this case by randomly resampling blocks of 24 values for a given day, not the individual hours as with a classic bootstrap; this is a version of the blocked bootstrap. In this way, the 24-hour periodicities in the data are preserved in the bootstrap samples, as well as large, positive serial correlations. Fourier analysis of the seasonal data found no significant periodicities greater than 1 day [4].

The bootstrap estimates of the 1-sigma errors for 30-degree sectors for SO_2 are shown in Figure 7, along with the bootstrap estimate of the bias. As evident from this figure, the bias is small compared to the uncertainty, which is an important requirement for the validity of the bootstrap method. The bootstrap estimates of the uncertainties in the wind speed and direction probability density were found to be very close to the uncertainties estimated by Eq. 9. This was also the case for the bootstrap estimates of the uncertainty in the nonparametric regression of pollutant concentrations on wind speed and direction and the uncertainties calculated by Eq. 8. As the bootstrap estimates include serial correlation effects in wind speed, wind direction, pollutant concentrations, it is evident that these have little impact on the estimated uncertainties.

Acknowledgement

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here under contract

EP06D000578 to Ronald C. Henry. It has been subjected to Agency review and approved for publication.

Dedication

The authors dedicate this work to our friend and colleague the late Dr. Charles W. Lewis of the U.S. EPA. Chuck was a supporter of this research and continues to be an inspiration.

References

1. Henry, R. C., Y-S Chang, C. H. Spiegelman 2002. Locating Nearby Sources of Air Pollution by Nonparametric Regression of Atmospheric Concentrations on Wind Direction. *Atmospheric Environment* **36**: 2237-2244.

2. Yu, K.N.; Y.P. Cheung; T. Cheung; and R. C. Henry 2004. Identifying the impact of large urban airports on local air quality by nonparametric regression. *Atmospheric Environment* **38**:4501-4507.

3. Kim, E and P. K. Hopke 2004. Comparison between conditional probability function and nonparametric regression for fine particle source directions. *Atmospheric Environment* **38**:4667-4673.

4. David W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons, Inc., New York, 1992.

5. Dilmaghani S.; Henry I.; Soonthornnonda, P.; Christensen E. R., and Henry, R.C. 2007. Harmonic Analysis of Environmental Time Series with Missing Data or Irregular Sample Spacing. *Environmental Science & Technology,* **41**, 7030-7038. DOI: 10.1021/es0700247

Tables

Table 1. Basic Statistics of 1-hr Sulfur Dioxide for First Quarter 2003 – 2005 (ppb)

Number	Mean	Median	Max	Min	Standard Deviation
6504	4.56	2	229	0	9.45

Table 2. Facilities Near East St. Louis (ESL) Site with Emissions of Sulfur Dioxide ≥ 1000 Tons per Year for 2002

Facility Name	Emissions (tons/yr)	Distance from ESL site (km)	Azimuth from ESL site
Big River Zinc Corp	1,378.6	1.5	216°
Anheuser-Busch Inc-St. Louis	6,249.9	4.6	249°
National Steel Corp - Granite City Div.	5,002.1	9.8	019°
Premcor Refining Group	1,607.9	25.5	014°
ConocoPhillips Co.	12,745.0	26.2	018°
Amerenue-Meramec Plant – Electric Utility	16,446.5	28.0	213°
Dynegy Midwest Generation Inc– Electric Utility	8,798.0	28.1	005°
Amerenue-Sioux Plant– Electric Utility	45,959.7	35.7	342°

Table 3. Sector Apportionment for ESL Sulfur Dioxide for First Quarter 2003 – 2005.

Putative Source	**Sector Center (Degrees Azimuth)**	**Percentage of Wind from Sector**	**Percentage of Average SO_2**	**1-sigma Error**	**Average Concentration (ppb)**
Zinc Smelter	216°	6.8	37.8	5.6	20.0
Brewery	249°	4.5	9.2	2.0	10.3
Steel Mill	019°	12.8	10.9	1.3	3.8
	Sum Smelter & Brewery		47.0	5.9	

Figures

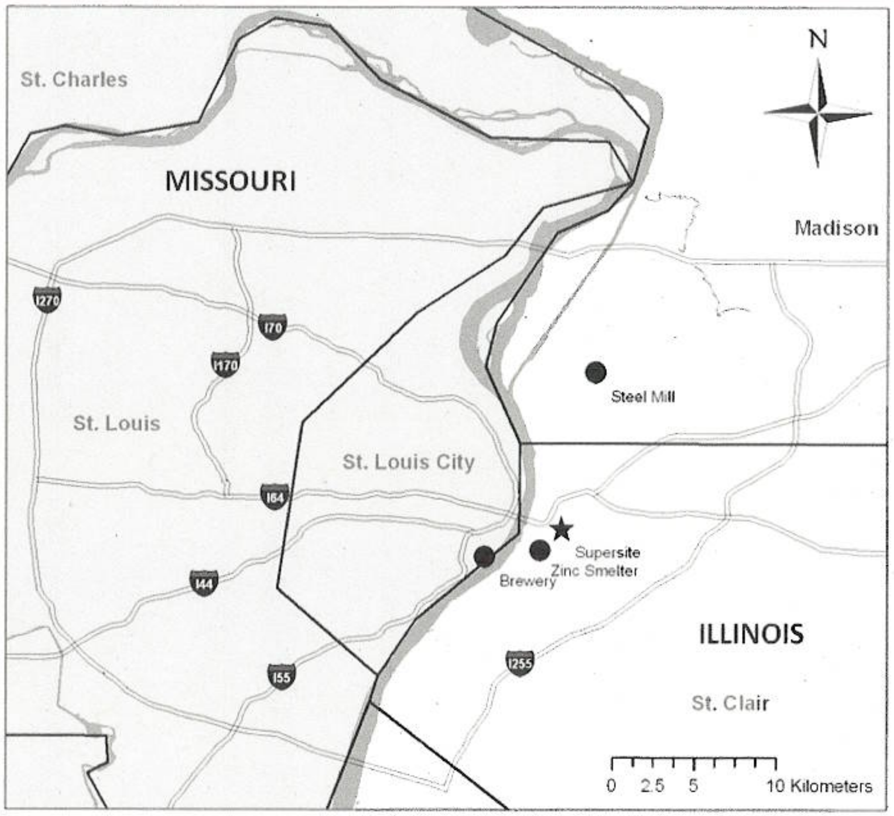

Figure 1 Locations of local sources of SO_2.

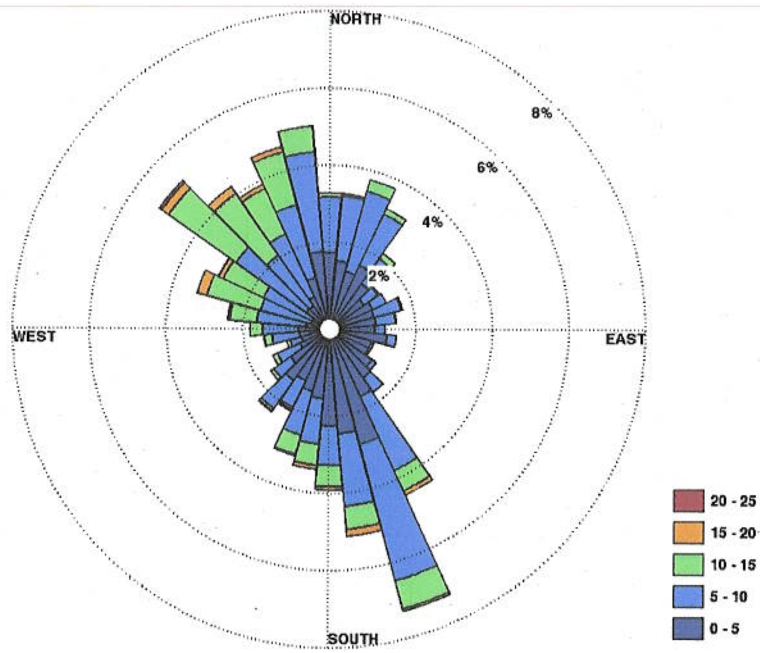

Figure 2. Wind rose at E. St. Louis for 9 winter months from 2003 to 2005. Units in the legend are km/h.

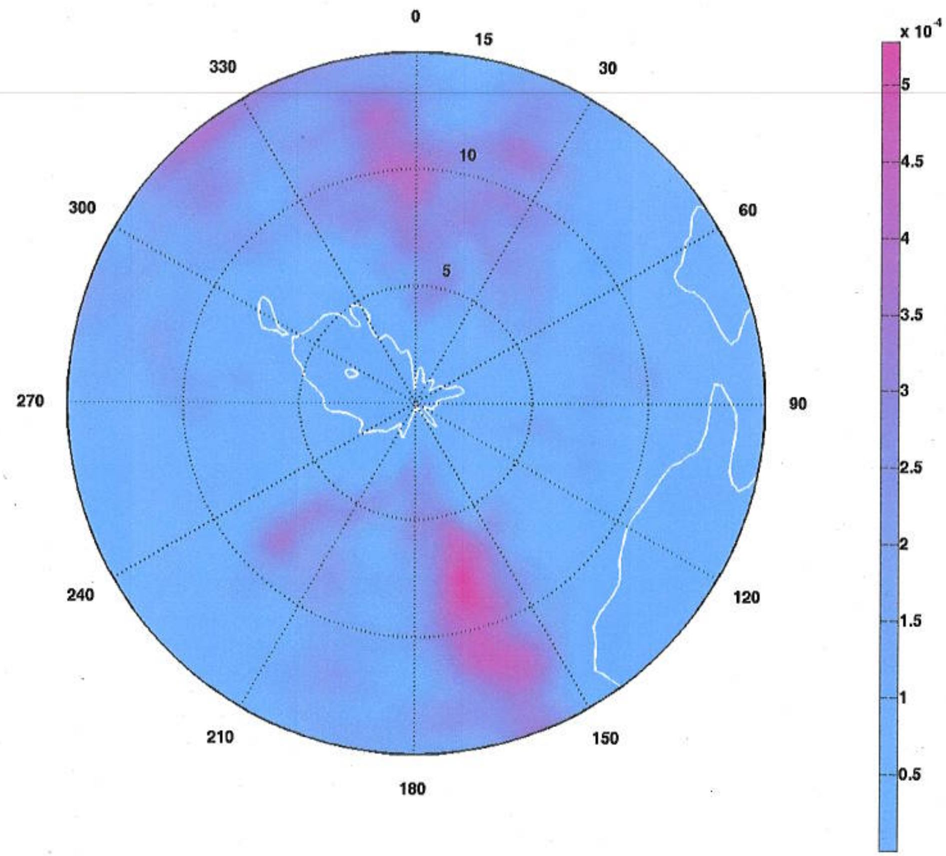

Figure 3. Joint probability distribution of wind speed and direction for 9 winter months from 2003 to 2005 for hours with $SO_2 \geq 1$ ppb and the standard deviation of the wind direction is less than 25 degrees. North is to the top and East to the right; the radial axis is km/h, maximum 15. The values inside the white contours have a signal-to-noise ratio of 2 or less.

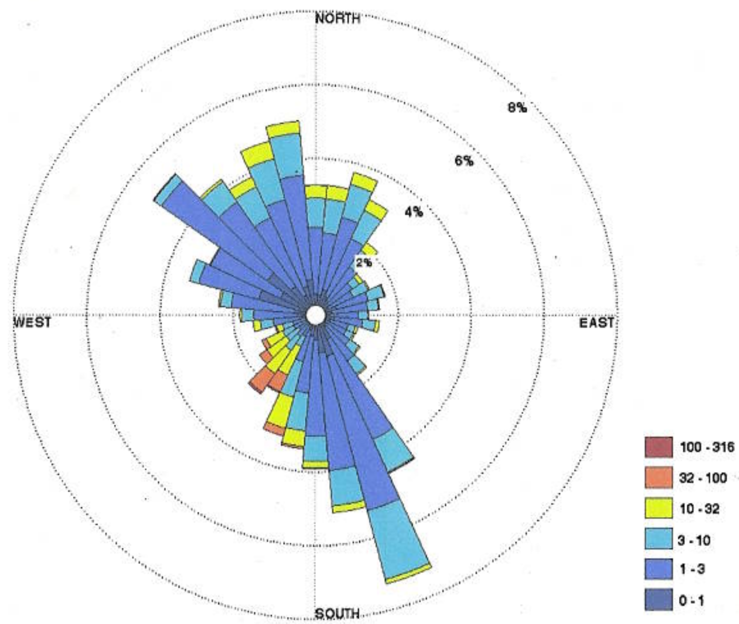

Figure 4 SO_2 pollution rose at E. St. Louis for 9 winter months from 2003 to 2005. Units in the legend are ppb.

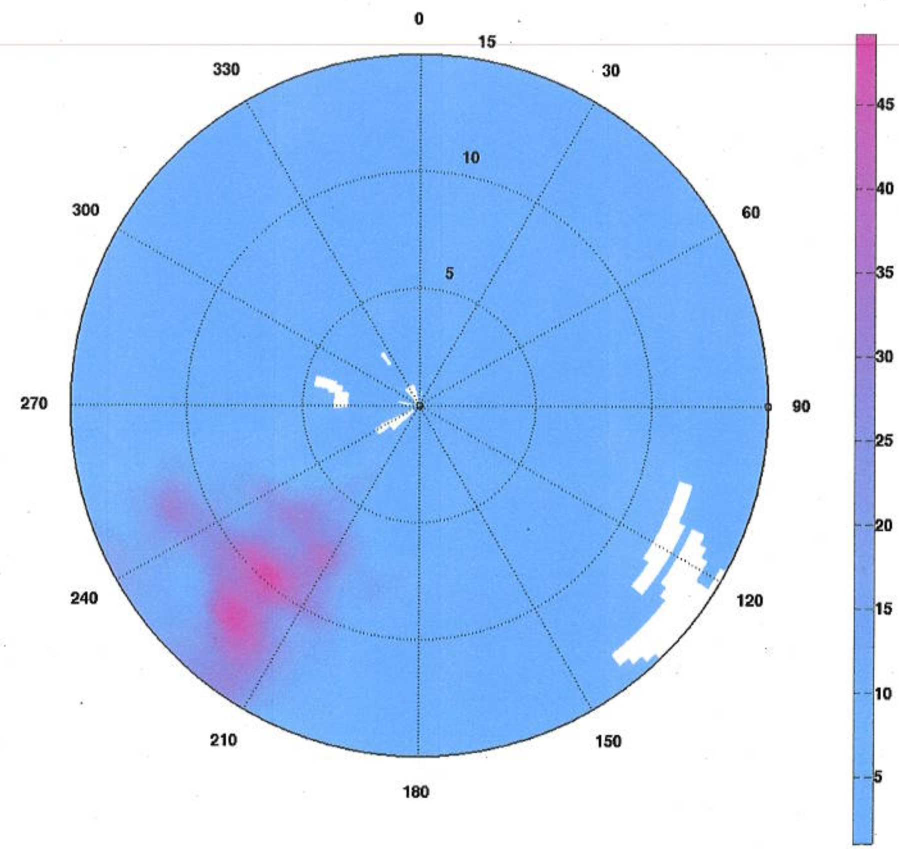

Figure 5. NWR plot of average SO_2 (in ppb) as a function of wind speed and direction for the winter months of 2003-2005. White areas have signal-to-noise less than 2.

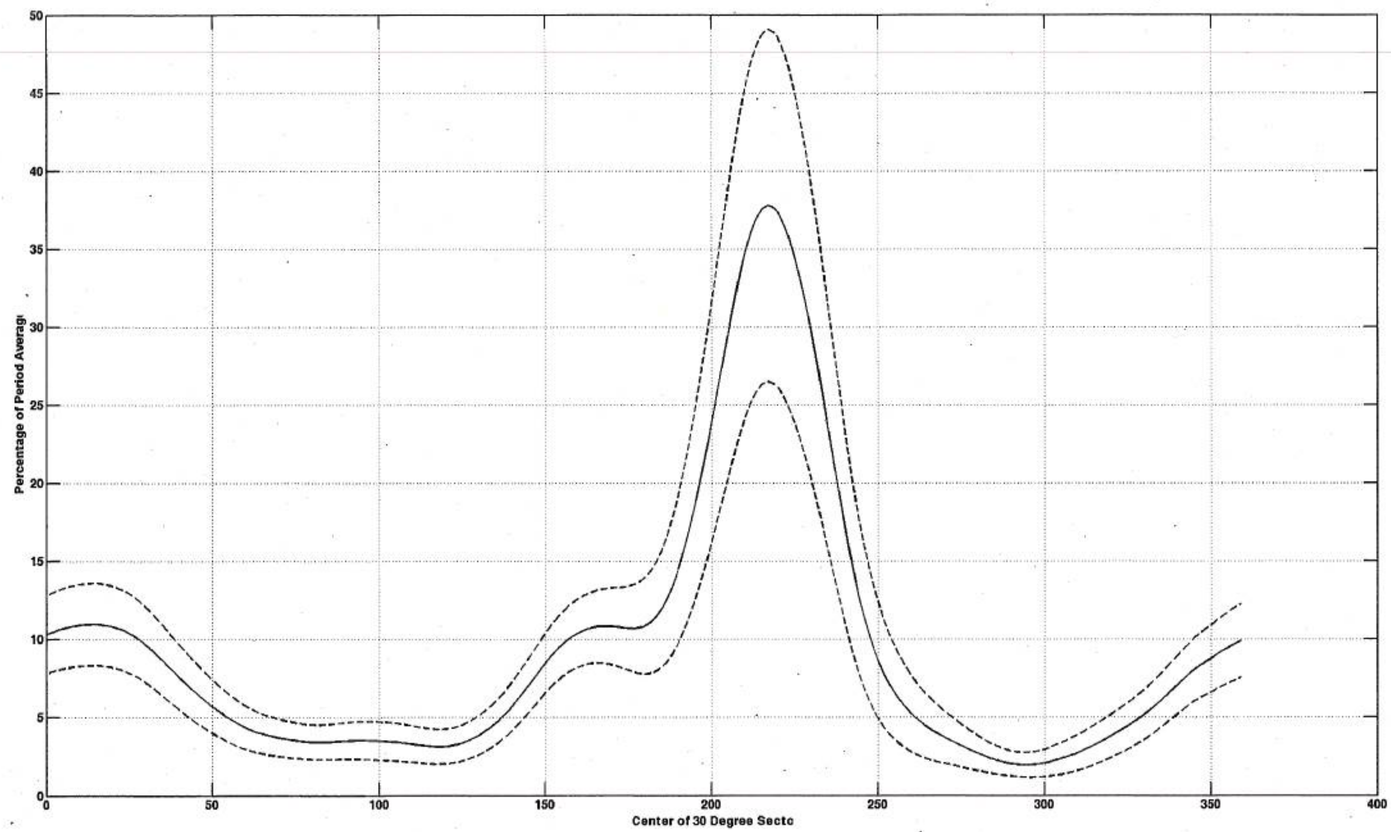

Figure 6. The percentage of the average SO_2 concentration associated with 30-degree wide sectors centered at each degree of azimuth and wind speed ≤ 15 km/hr with 2-sigma error bars.

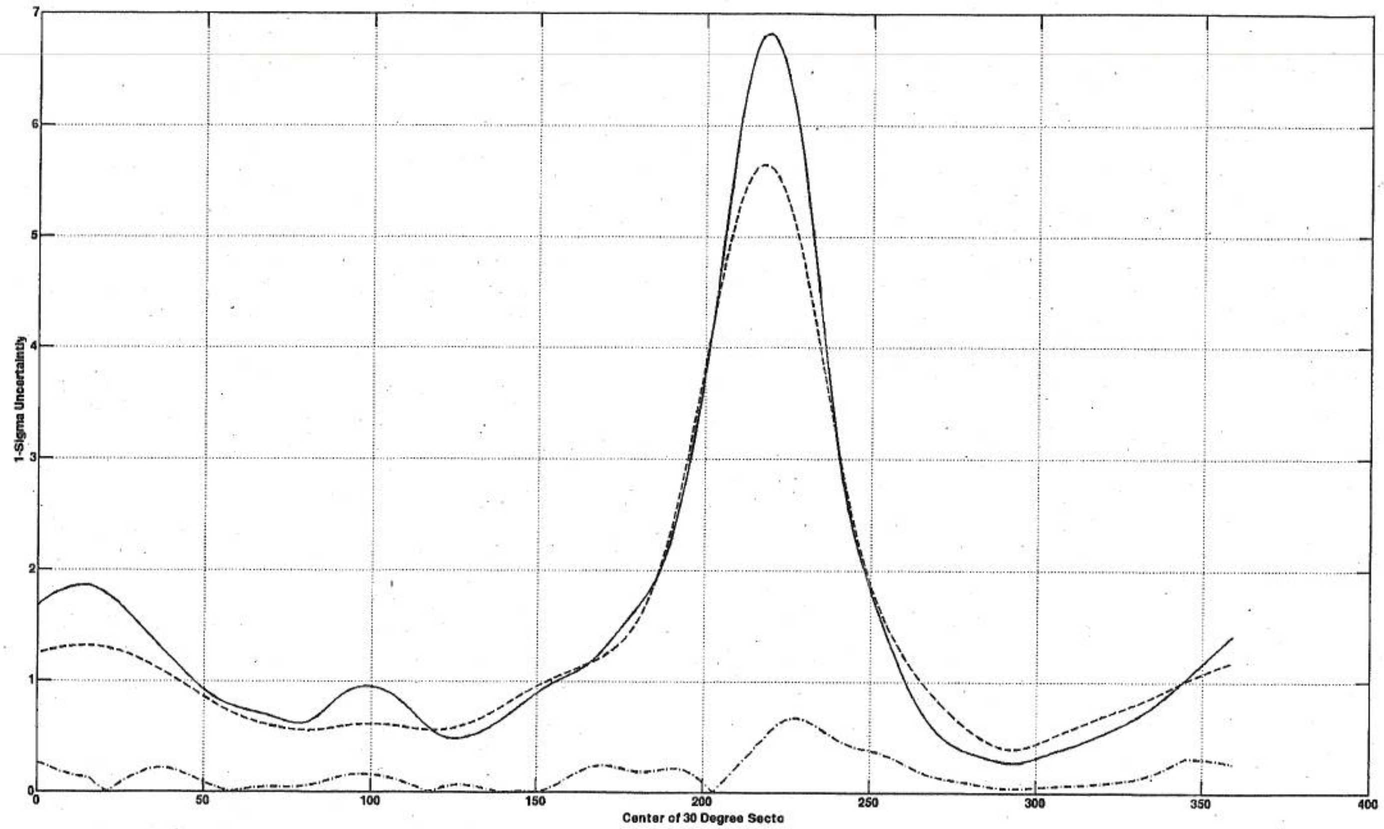

Figure 7. Blocked bootstrap 1-sigma uncertainties (solid line) and bias (dot-dash line) for 30 degree sectors in Fig. 7. The dashed line is the uncertainty calculated by Eq. 11. The values are a percentage of the average SO_2 for the entire period.