NISS

The Effect of Temporal Aggregation in Models to Estimate Trends in Sulfate Deposition

Patricia E. Styer

Technical Report Number 19
August, 1994

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Although the information in this document has been funded wholly or in part by the United States Environmental Protection Agency under assistance agreement #CR819638-01-0 to the National Institute of Statistical Sciences, it may not necessarily reflect the views of the Agency and no official endorsement should be inferred.

The Effect of Temporal Aggregation in Models to Estimate Trends in Sulfate Deposition

Patricia E. Styer *

National Institute of Statistical Sciences

Abstract

This research investigates the effect of temporal aggregation in regression models used to measure long-term trends in the wet deposition of sulfate. I propose a set of gamma regression models that utilize precipitation and meteorological data collected on a variety of time scales. Specifically, I examine models that fit daily-level precipitation chemistry to daily-level meteorological covariates, weekly-level precipitation chemistry to weekly-level covariates, and weekly-level precipitation chemistry to daily-level covariates using historical data collected at daily monitoring sites, with artificial aggregation to create weekly-level data. Empirical results show that there can be small differences among the estimates of long-term trend in sulfate deposition under the three aggregation schemes, as well as a loss of precision with aggregation. Using a jackknifing procedure to obtain estimates of the standard errors of the differences in parameter estimates, I conclude that there is no significant difference in the estimation of long-term trends using weekly-level data.

Key words: gamma regression, environmental monitoring

---

*Supported by the EPA Cooperative Agreement CR 819638-01 and the National Science Foundation Grant DMS-9208758.

1 Introduction

The purpose of this research is to compare estimates of trends in sulfate deposition from models that incorporate precipitation chemistry and meteorological covariates on different time scales. It is motivated by recent legislation that mandates substantial reductions in sulfur dioxide emissions at considerable costs, and thus highlights the need for more accurate measurements of the impact of emissions reductions (NAPAP, 1993). This research is also motivated by a second public policy issue regarding funding priorities for environmental monitoring. Until a few years ago, precipitation monitoring in the United States included networks that collected precipitation samples on either a daily or a weekly basis. Most of the daily monitoring sites have been shut down, and there is limited funding to maintain the sites still in operation. It is of interest to examine whether or not the availability of daily level monitoring data can enhance the detection of trends in the wet deposition of sulfate.

To improve the assessment of the relationship between emissions and deposition, and to investigate the effect of having weekly precipitation monitoring data, I propose a set of gamma regression models that utilize precipitation and meteorological data collected on a variety of time scales. Specifically, I examine models that fit daily precipitation chemistry to daily meteorological covariates, weekly precipitation chemistry to weekly-level covariates, and weekly precipitation chemistry to daily-level covariates. In this paper, I present results only from an empirical comparison using precipitation data from four monitoring sites in the MAP3S network. This empirical analysis demonstrates that the temporal aggregation of precipitation chemistry and meteorological covariates can influence the values of the estimated trends in deposition, but the differences are not statistically significant for the data analyzed here. Additionally, the use of weekly-level covariates introduces only small increases in the standard error estimates for the trend parameter. For the purpose of long-term trend measurement at individual sites, even taking into account adjustments for meteorology, weekly-level monitoring data appears to be sufficient.

2 Model Formulation

Meteorological conditions influence all stages of the transport, conversion, and deposition of sulfate. While these meteorological conditions are complex and change constantly, it is necessary to introduce some simplifications to permit the incorporation of meteorological information for many years of data at many monitoring locations. For this analysis, I assume that it is minimally adequate to incorporate meteorological information on the time scale of days—even though local weather conditions can change drastically within a single day. Hence, I will use a daily-level model as the standard basis for comparison.

The regression models presented in this paper are based on a type of generalized linear model that I refer to as gamma regression models (McCullagh and

Nelder, 1989). Since deposition amounts are always positive and tend to have an asymmetric distribution with a long right tail, I use the gamma distribution to model both daily and weekly sulfate deposition. Thus, the model for daily sulfate deposition is that S_i, the sulfate deposition on day i, is gamma-distributed with density

$$f(s_i) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu_i}\right)^{\nu} s_i^{\nu-1} \exp\left(\frac{-\nu}{\mu_i} s_i\right).$$

The expected value for the sulfate deposition on day i is

$$\mu_i = \mathrm{E}(S_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}),$$

where $\mathbf{x}_i$ includes a seasonal factor, a term to permit the estimation of long-term trend, and daily-level meteorological covariates, including precipitation amount. Since these models include the precipitation amount, modeling deposition is equivalent to modeling concentration. More details about the other meteorological covariates are provided in the appendix.

If only weekly precipitation chemistry measurements are available, a reasonable model for weekly deposition would be that S_j, the sulfate deposition for the j^{th} week, is gamma distributed with mean

$$\mu_j = \mathrm{E}(S_j) = \exp(\bar{\mathbf{x}}_j'\boldsymbol{\beta}),$$

where $\bar{\mathbf{x}}_j$ includes a seasonal factor, a term to permit the estimation of long-term trend, and a weekly aggregate of meteorological covariates. Under the parameterization that is used for generalized linear models, the sum of the daily depositions do not follow a gamma distribution. With this parameterization, the scale parameter ν is not constant so the sums of the gamma variables are not themselves gamma distributed. Hence, if the daily model is true, there is a slight misspecification in applying the gamma assumption to the weekly monitoring data. If standard diagnostics are applied to the weekly data, however, the gamma distribution appears to be appropriate. Often, there is only one or two days of rain per week, so the total deposition amounts remain skewed.

Finally, a model that uses only weekly measures of sulfate deposition, but daily summaries of meteorology, is given by assuming that the weekly sulfate measurements are gamma distributed with mean

$$\mu_j = \mathrm{E}(S_j) = \sum_{i=1}^{I_j} \exp(\mathbf{x}_{ij}'\boldsymbol{\beta}),$$

where I_j is the number of days with rain in the j^{th} week, and daily measures of meteorology are included in the vector $\mathbf{x}_{ij}$, consisting of meteorological covariates for the i^{th} day of rain during the j^{th} week, as well as components for long-term trend and season. This model utilizes the relationship that the total weekly deposition, S_j, equals the sum of the daily depositions for each day with precipitation. Since it is nonlinear in the systematic component after applying the logarithmic link function, it cannot be fit using standard statistical packages for generalized linear models. Assuming that the weekly aggregated sulfate depositions are gamma-distributed, it can be fit by maximizing the likelihood directly. Again, with the assumption that the daily model with constant coefficient of variation is true, the weekly depositions are not in truth exactly gamma distributed. This model is subject to the same misspecification as the model with only weekly summaries of meteorology.

3 Empirical Evaluation of Aggregation Effects

Fitting similar regression models to monitoring data on different time scales should have two basic effects. First, the underlying parameters from the daily model may not be equivalent to the parameters in the weekly and nonlinear models, so the parameter estimates obtained from aggregated data will not necessarily be estimates of the true daily-level parameters. Second, temporal aggregation introduces a loss of information. The standard errors for the parameter estimates using aggregated data should generally be larger than the standard errors using unaggregated data.

While it is of interest to study the differences in the underlying model parameters directly, the goal of the research presented here is to evaluate the empirical evidence that estimates of long-term trends in sulfate deposition depend on the temporal scale of the monitoring data. I present results from four precipitation monitoring sites in the Eastern United States from the Multistage Atmospheric Power Product Pollutions Study (MAP3S) network located in Whiteface, New York, Ithaca, New York, College Station, Pennsylvania, and Charlottesville, Virginia. In this network, the sample bucket is changed daily during periods of intermittent precipitation and after each episode of heavy precipitation (MAP3S, 1982). The data span just over 12 years, from late 1976 through 1988. The reported sulfate and precipitation values used in this analysis were obtained from the Acid Deposition System (ADS) and include a quality control flag (Watson and Olsen, 1984). I exclude observations that failed the ADS quality control checks.

Although precipitation data collected in MAP3S network are not strictly daily-level observations, for convenience I call them daily-level samples in this analysis and use daily meteorological summaries that correspond to the day that the sample bucket was removed. I then artificially aggregate the precipitation data to obtain weekly-level aggregates. Differences in the parameter estimates and standard error estimates will be evaluated as evidence of the effect of temporal aggregation. Since one goal of this analysis is the improvement of trend detection by using regression models to account for known sources of variability, it is of interest to know if the unavailability of daily monitoring data will diminish the effectiveness of this use of regression analysis.

Table 1 shows the estimated long-term trend parameters and standard errors for the four stations in the MAP3S network using the three models described above, converted to an estimated percent-per-year change in deposition. For illustration, estimates for the other covariates in the regression models at Whiteface, New York, are listed in the appendix. At the first three sites, all three of the models suggest a significant decrease in sulfate deposition, with a typical reduction of about three percent per year. There is a consistent negative trend at the Virginia site, but it is not significantly different from zero in any of the models. Generally, the differences between the estimates from the nonlinear models and daily models are smaller than differences between the weekly and daily models. The standard errors increase as the level of data aggregation increases. They are uniformly smallest for the daily models, with no aggregation, and largest for the weekly models, where both the covariates and the precipitation chemistry are aggregated.

To investigate the significance of the differences in the trend estimates between the daily and weekly models, and to remove the effect of any possible short term autocorrelation, I calculated jackknifed estimates and standard errors leaving out data in blocks of one-month periods. The jackknifed estimates for the trend parameters, listed in Table 2, are very similar to the usual estimates obtained without jackknifing. The jackknifed standard errors are slightly larger than the standard errors calculated from the usual asymptotic covariance matrix listed in Table 1. More importantly, though, jackknifing provides a way to estimate the standard errors for the differences in the estimates from the daily and weekly models. The final column of Table 2 shows the jackknifed differences and standard errors, permitting a formal comparison of the trend estimates from the two models. None of the differences are significantly different from zero. This result also holds when the jackknife standard errors are calculated by leaving data out in blocks of individual years and four-month seasons.

4 Summary and Conclusions

This research is motivated by the need to understand the implications of restricting precipitation monitoring efforts to weekly-level sampling protocols. Specifically, I investigate how temporal aggregation can influence the estimation of the effect of reductions in sulfur dioxide emissions. I fit models incorporating measures of season, meteorological covariates, and emissions to precipitation data and compare the empirical estimates of the parameters and standard errors. This empirical investigation shows that there can be small differences between the parameter estimates and standard errors in the unaggregated and aggregated processes. When estimates of standard errors are included in the analysis, there is no significant difference in trend estimates among the models using different levels of aggregation.

5 Appendix

The regression models include meteorological covariates measuring temperature and wind conditions. As suggested in other studies, I use 850 millibar wind measurements, selected to be concurrent with the precipitation event, as potential measures of both transport and local weather conditions (Styer and Stein, 1992). The wind data used here are from a subset of the National Meteorological Center (NMC) gridded upper air data (Atmos. Sci., Univ. of Wash., 1990). The particular data set employed has a grid resolution of two degrees latitude by four degrees longitude, so I selected the grid point closest to the monitoring site. Information about wind is incorporated using the zonal (U) and meridional (V) components, which provide information about both wind speed and direction.

The regression equations also include a measure of seasonally adjusted temperature, represented by T. In this analysis, the temperature variable is also extracted from the 850 millibar measurements using NMC gridded upper air data. For each MAP3S site analyzed, I used the 850 millibar temperature measurements from the closest grid point, taken to be the measure of daily temperature for that precipitation event. Specifically, I calculate T_i as the daily temperature for the i^{th} precipitation event minus the average daily temperature for the month during which the i^{th} precipitation event began. The average daily temperatures for each month are estimated from all of the years for which monitoring data are available.

In addition to the temperature and wind covariates, I also include a covariate showing the number of days since the last precipitation event (D). Since precipitation activity cleanses the air of pollutants, sulfate concentration should be lower during periods of frequent precipitation episodes and should increase as the time span between precipitation episodes increases. And finally, I include an index variable corresponding to year to provide an estimate of long-term trend in deposition. The yearly index variable is represented by Y in the regression equation.

The full specification of the daily-level regression model is the following: daily sulfate concentration is assumed to be gamma-distributed with mean

$$\begin{aligned}\mu_i &= \exp(a_1 + \sum_{m=2}^{12} a_m M_{mi} + b_p \log P_i + b_u U_i + b_v V_i \\ &\quad + b_t T_i + b_d D_i + b_y Y_i),\end{aligned}$$

where, for the i^{th} precipitation event, M_{mi} is an index variable for the month in which the precipitation event begins, P_i is the total precipitation amount in millimeters, U_i and V_i are the zonal and meridional components of wind, T_i is a measure of temperature in degrees Celsius, D_i is the number of days since the last precipitation event, and Y_i is the index variable for year.

For the weekly-level model, it is necessary to create weekly-level measures of the meteorological covariates. Here, I used the total weekly precipitation and the mean weekly temperature. For the wind components, I used the wind measurements corresponding to the day of the week with the greatest amount of precipitation. For the day count, I took the first day of the week with rain and counted the number of days since the previous precipitation event. I also included an additional covariate, labeled $nday$, counting the number of days in the week with precipitation. The monthly and yearly index variables were also coded using the date of the first precipitation event of each week. For the nonlinear model, the daily-level meteorological covariates are used in conjunction with the weekly-level precipitation chemistry. The daily and weekly models are fit using the glm function in the S language (Chambers and Hastie, 1993). For the nonlinear model, I use the function ms and obtain parameter estimates by maximizing the assumed likelihood function directly.

The estimates for the three models are listed in Table 3. For a discussion of the interpretation of these parameters, see Styer(1994).

6 References

Atmospheric Sciences, University of Washington & Data Support Section, NCAR (1990). *National Meteorological Center Grid Point Data Set: Version II.* University of Washington.

Chambers, J.M. & Hastie, T. J. (1993). *Statistical Models in S.* Chapman & Hall, New York.

MAP3S (1982). The MAP3S/RAINE Precipitation Chemistry Network: statistical overview for the period 1976-1980. *Atmospheric Environment* **16**, 1603-1631.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, New York.

NAPAP (1993). 1992 Report to Congress. National Acid Precipitation Assessment Program. Washington, DC.

Styer, P.E. (1994). An illustration of the use of generalized linear models to measure long-term trends in the wet deposition of sulfate. NISS Technical Report No. 18. National Institute of Statistical Sciences, Research Triangle Park, NC.

Styer, P.E. & Stein, M.L. (1992). Acid deposition models for detecting the effect of changes in emissions: an exploratory investigation utilizing meteorological variables. *Atmospheric Environment* **26A**, 3019-3028.

Watson, C.R. and Olsen, A.R. (1984). *Acid Deposition System (ADS) for statistical reporting: system design and user's code manual.* EPA-600-8-84-023, U.S. Environmental Protection Agency, Research Triangle Park, N.C.

Table 1
Estimates and Standard Errors for Long-Term Trend
Percent per Year Change in Deposition

Site	Daily	Weekly	Nonlinear
Whiteface, NY	-3.17 (0.69)	-3.28 (0.82)	-3.10 (0.75)
Ithaca, NY	-3.63 (0.65)	-4.39 (0.72)	-4.11 (0.68)
College Station, PA	-2.90 (0.51)	-2.60 (0.66)	-2.72 (0.62)
Charlottesville, VA	-1.05 (0.82)	-0.67 (0.92)	-0.68 (0.83)

Table 2

Jackknifed Estimates of Trends and Differences

Site	Daily	Weekly	Weekly-Daily
Whiteface, NY	-3.22 (0.78)	-3.38 (0.90)	-0.15 (0.55)
Ithaca, NY	-3.58 (0.66)	-4.34 (0.70)	-0.77 (0.42)
College Station, PA	-3.17 (0.66)	-2.90 (0.75)	0.26 (0.51)
Charlottesville, VA	-1.10 (1.00)	-0.82 (1.07)	0.28 (0.42)

Table 3
Estimates and Standard Errors for Regression Coefficients at Whiteface, NY

Coefficients	Daily Model	Weekly Model	Nonlinear Model
Intercept (a_1)	2.895(0.127)	3.013(0.147)	2.946(0.133)
Feb-Jan (a_2)	0.331(0.127)	0.407(0.134)	0.418(0.124)
Mar-Jan (a_3)	0.624(0.115)	0.708(0.122)	0.685(0.112)
Apr-Jan (a_4)	1.026(0.118)	1.000(0.127)	1.032(0.118)
May-Jan (a_5)	1.160(0.115)	1.238(0.125)	1.291(0.114)
Jun-Jan (a_6)	1.223(0.114)	1.268(0.128)	1.274(0.117)
Jul-Jan (a_7)	1.162(0.116)	1.264(0.128)	1.282(0.117)
Aug-Jan (a_8)	1.381(0.114)	1.437(0.127)	1.467(0.117)
Sep-Jan (a_9)	0.925(0.116)	1.062(0.127)	1.050(0.117)
Oct-Jan (a_{10})	0.744(0.116)	0.815(0.127)	0.808(0.115)
Nov-Jan (a_{11})	0.428(0.114)	0.560(0.124)	0.572(0.114)
Dec-Jan (a_{12})	0.294(0.115)	0.417(0.126)	0.331(0.116)
log precip (b_p)	0.691(0.021)	0.641(0.030)	0.656(0.027)
temperature (b_t)	0.036(0.003)	0.033(0.005)	0.033(0.006)
u (b_u)	-0.023(0.003)	-0.012(0.002)	-0.021(0.004)
v (b_v)	0.014(0.008)	0.012(0.002)	0.014(0.004)
day count (b_d)	0.033(0.008)	0.012(0.008)	0.024(0.009)
nday (b_n)		0.147(0.033)	
yearly trend (b_y)	-0.031(0.006)	-0.032(0.008)	-0.031(0.007)