AN ANALYSIS OF HOURLY ACID DEPOSITION DATA

P. Switzer

Technical Report No. 101
AN ANALYSIS OF HOURLY ACID DEPOSITION DATA

BY

PAUL SWITZER

TECHNICAL REPORT NO. 101
SEPTEMBER 1986

STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH (SIMS)

PREPARED UNDER A GRANT TO SIMS FROM
ENVIRONMENTAL PROTECTION AGENCY (EPA)
SLOAN FOUNDATION
NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
An Analysis of Hourly Acid Deposition Data

Paul Switzer

1. Introduction

The variability over time of precipitation acidity can be very substantial. Based on time series of a few years' duration, seasonal variations have been documented as well as year-to-year variability of average levels. See, for example, Bilonick and Nichols (1983), Eynon and Switzer (1983), Rodhe and Granat (1984). These acidity variations are largely driven by meteorological variation as opposed to pollutant emissions variability. The data records for variability studies of this type have mainly been either monthly averages or single precipitation event averages.

Some of the observed variability in acidity is related to precipitation volume. Generally, one might expect lower concentrations for larger precipitation volumes under a scavenging hypothesis. This hypothesis says that the acidity of precipitation results in part by raindrops or snow particles gathering acidic particles in the air column. Examples comparing precipitation volume and pH at several monitoring sites are described in Eynon and Switzer (1983). The data described there are on a precipitation event basis, i.e., a single pH value for each event at each monitoring site. An attempt to infer scavenging rates from concomitant atmospheric and precipitation sulfate is given in Davidson et al. (1986), also on an event basis.

Precipitation data on an hourly basis have become available for an experimental monitoring site at Brookhaven National Laboratory for the period June 1976 to September 1983; see Raynor and Hayes (1984). These data might allow for a more careful study of scavenging models. Single precipitation events of sufficient duration might show enough hourly variability to infer useful information regarding scavenging. In particular, one might hope to partition the total acidity for a given precipitation event into a component which reflects air column scavenging, interpreted as a local effect, and a second component which reflects the acidification of the cloud water before precipitation, interpreted as an imported effect.
2. Data

The hourly data for precipitation events used in this investigation were obtained through the Brookhaven National Laboratory Chemistry Sampling Program. It seemed useful to select a relatively homogeneous subset of the data amenable to statistical analysis and an acidity variable likely to be affected by scavenging. Discussions with Paul Michael, atmospheric chemist at BNL, led to the choice of nitrate as the acidity variable and to the following criteria for selection of precipitation events:

i) The precipitation event should be of at least 7 hours duration.

ii) Consider only warm front events.

iii) Precipitation should be in the form of snow or rain-mixed snow.

iv) Data for nitrate concentration should be available for each hour.

Since this was to be an exploratory and somewhat speculative investigation the first four events satisfying all the criteria were selected for analysis. The dates and durations of these precipitation events are given below and graphs of the hourly precipitation volume and nitrate concentration for each event are shown in Figure 1 and Figure 2, respectively.

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>December 26, 1976</td>
<td>8 hours</td>
</tr>
<tr>
<td>2</td>
<td>January 7, 1977</td>
<td>7 hours</td>
</tr>
<tr>
<td>3</td>
<td>January 14, 1977</td>
<td>9 hours</td>
</tr>
<tr>
<td>4</td>
<td>January 20, 1979</td>
<td>20 hours</td>
</tr>
</tbody>
</table>

The general trend of decreasing nitrate concentration can be seen in the data plots. The notation we use is:

\[ V_t = \text{precipitation amount (millimeters) for hour } t \]
\[ Y_t = \text{nitrate concentration in precipitation for hour } t \]
\[ \text{micro-equivalents/liter} \]
\[ t = 1, 2, \ldots \text{ length of event in hours.} \]

The notation suppresses dependence on the event number, i.e., there are a pair of hourly time series \( V_t, Y_t \) for each of the four events. The units of micro-equivalents per liter of nitrate were calculated as Total \( N \) in \( NO_3^- + NO_2^- \) in parts per million, multiplied by the conversion factor 71.38; one micro-equivalent of \( H^+ \) is the same as one part per million \( H^+ \).
FIGURE 1
THE FOLLOWING RESULTS ARE FOR:

ID = 5.00

ID = 8.00

ID = 7.00

ID = 15.0

THE FOLLOWING RESULTS ARE FOR:

JANUARY 7, 1977

JANUARY 14, 1977

JANUARY 20, 1979

FIGURE 2
The goal of the analysis of these short time series for precipitation nitrate concentrations and precipitation volume is to estimate two non-time-dependent quantities for each precipitation event:

\[
\alpha = \text{concentration of nitrate in cloud water (micro-equivalents/liter)}
\]

\[
\beta = \text{total scavengable nitrate in the air column (micro-equivalents/meter-squared)}.
\]

Figure 3 illustrates the general scheme in cartoon form.
3. The Scavenging Model

A basic scavenging model asserts that each unit of precipitation surface area scavenges a fixed proportion of the available nitrate aerosol in the air column through which the precipitation falls. This model has the following simple description: imagine a meter square air column, then

\[ \frac{d\beta_t}{dS_t} \propto \beta_t \]

where

\[ \beta_t = \text{nitrates aerosol remaining in the air column after } t \text{ hours of precipitation} \]
\[ (\text{micro-equivalents/meter-square}) \]
\[ S_t = \text{total surface area of precipitation particles falling during the first } t \text{ hours} \]
\[ (\text{millimeters/meter-square}) \]

Thus for each precipitation event the two relevant event-specific functions of time are \( \beta_t \) and \( S_t \). The notation suppresses dependence on the event number. More explicitly we may write

(1) \[ \beta_t = \beta \cdot \gamma^{S_t} \]

where

\[ \beta_t = \text{nitrates aerosol in the air column at the beginning of the precipitation event} \]
\[ (\text{micro-equivalents/meter-squared}) \]
\[ \gamma = \text{a time independent rate parameter referred to as the scavenging efficiency, } 0 < \gamma < 1. \]

The total nitrate scavenged by precipitation particles falling during hour \( t \) is then \( \beta_{t-1} - \beta_t \). Hence the concentration of scavenged nitrate in precipitation for hour \( t \) of the precipitation event may be expressed as

(2) \[ \frac{\beta_{t-1} - \beta_t}{V_t} = \beta \cdot \frac{(\gamma^{S_{t-1}} - \gamma^{S_t})}{V_t} \equiv \beta \cdot X_t(\gamma), \text{ say,} \]

\[ \text{(micro-equivalents/liter)} \]

where, as before \( V_t \) is the precipitation volume in millimeters for hour \( t \), and we make use of expression (1) above.

The total concentration of nitrate in precipitation will be a sum of the contribution due to air column nitrate aerosol scavenging given by (2) and the contribution due to nitrate in the cloud water. This total concentration \( Y_t \) for hour \( t \) of the precipitation event, a measured quantity, may then be written as

(3) \[ Y_t = \alpha + \beta \cdot X_t(\gamma) + \epsilon_t \]
where $\alpha$ is the time-independent cloud-water concentration of nitrate, $\beta$ is the time-independent total initial nitrate in the air column, $X_t(\gamma)$ is defined in (2), and $\epsilon_t$ is a residual accounting for measurement error and inadequacy of the model assumptions (such as time-independence of $\alpha$ and $\beta$). It is convenient to think of the $\epsilon_t$ as zero-mean random variables. There would be a separate model of the type (3) for each precipitation event.

The quantities $V_t, Y_t$ are available in the data set for each hour of each precipitation event. The goal is to estimate the initial nitrate values $\alpha, \beta$ representing, respectively the 'imported' cloud water contribution and the 'local' air column contribution. The scavenging efficiency $\gamma$ is regarded as a nuisance parameter.

In the expression (2) the time series $S_t$ of hourly cumulative precipitation surface contact area is not observable and can only be approached via a surrogate series. The simplest surrogate for precipitation surface area would be a quantity proportional to precipitation volume, i.e.,

$S_t - S_{t-1} = c \cdot V_t$.

In the model (2,3) the generic proportionality constant $c$ could be absorbed into the nuisance parameter $\gamma$ by replacing $\gamma$ with $\gamma^c$, so that $c$ would not need to be separately determined. For example, the relation (4) would be reasonable if the relative frequency distribution of water particle sizes and shapes did not change during the course of a rainfall event. However, it has been observed that larger hourly precipitation volumes are accompanied by a higher relative frequency of larger water particles. An example of an empirical relationship is the Marshall-Palmer-Gunn formula given by

$S_t - S_{t-1} = c \cdot V_t^\delta$ where

$\delta = 0.79$ for rain particles

$\delta = 0.52$ for snow particles,

see Gunn and Marshall (1958) and Pruppacher and Klett (1978), Chapter 2. Once again, the generic proportionality constant $c$ could be absorbed into the scavenging efficiency parameter $\gamma$. While they are widely used, such empirical formulas could only hold approximately and only for a fixed time step. The shortcoming is evident when one considers doubling the time step, say; in that case the incremental surface area $S_t - S_{t-2}$ should be $c(V_t + V_{t-1})^\delta$ which is not the same as $c(V_t^\delta + V_{t-1}^\delta)$.

Despite its shortcomings the Marshall-Palmer-Gunn formula is here adopted to obtain surrogate surface air measures. Since the particular precipitation events used as illustrations in this paper are all snowfall events, the area-to-surface conversion
exponent $\delta = 0.52$ is used. Hence, we have

\begin{equation}
S_t = \sum_{i=1}^{t} V_t^{0.52}
\end{equation}

for substitution into (2), with the proportionality constant $c$ being absorbed by the parameter $\gamma$. 
4. Parameter Estimation

The real goal of this demonstration is the estimation of the model parameters \( \alpha, \beta \) with their respective interpretations as imported and local nitrate contributions to precipitation. For each precipitation event there would be such a parameter pair to estimate. If the scavenging efficiency parameter \( \gamma \) were a specified constant then model (3) is in the form of a simple linear regression for each multiple-hour precipitation event. Straightforward estimates of \( \alpha, \beta \) would then be obtainable with at least pro forma estimates of their standard errors.

While the nuisance parameter \( \gamma \) introduces interesting complexity to the estimation problem, this aspect will not be seriously explored here. Instead a device was used to estimate \( \gamma \) in a direct fashion from the basic data series \( Y_t, V_t \) of hourly precipitation nitrate concentration and hourly precipitation volume. The estimate of \( \gamma \) so obtained was treated as a fixed constant thereafter and the parameters of interest \( \alpha, \beta \) were estimated as in a linear regression model.

The device used to estimate \( \gamma \) directly is to write model (3) for the precipitation nitrate concentration in the apparent form of a second order autoregression, viz.,

\[
Y_t = \hat{Y}_t(\gamma) + \tau_t(\gamma), \quad t \geq 3, \quad \text{where}
\]

\[
\hat{Y}_t(\gamma) = Y_{t-1} - R_t(\gamma) \cdot [Y_{t-1} - Y_{t-2}],
\]

\[
R_t(\gamma) = [X_t(\gamma) - X_{t-1}(\gamma)] / [X_{t-2}(\gamma) - X_{t-1}(\gamma)],
\]

with \( X_t(\gamma) \) defined in (2). The salient part of this representation is that \( \tau_t(\gamma) \) has mean zero for every \( \alpha, \beta, \gamma, \) and \( t \geq 3 \), and that \( \hat{Y}_t(\gamma) \) does not need an estimate of \( \alpha, \beta \). For any \( \gamma, \hat{Y}_t(\gamma) \) is computable directly from the observed concentrations \( Y_{t-1}, Y_{t-2} \) and the observed precipitation volumes \( V_t, V_{t-1}, V_{t-2} \). These facts are verified by direct substitution.

In general, the covariance structure of the above residuals \( \tau_t(\gamma) \) will be complicated and will depend on \( \gamma \) even if the original residuals \( \epsilon_t \) of model (3) are taken to be uncorrelated and of constant variance. However, we should obtain a reasonable estimate of \( \gamma \) by minimizing a distance criterion between the \( Y_t \) series and the \( \hat{Y}_t(\gamma) \) series. For example, one may use the value of \( \gamma \) which minimizes

\[
d(\gamma) = \sum_{t=3} |Y_t - \hat{Y}_t(\gamma)|.
\]

The minimizing value of \( \gamma \) for the above criterion was found for each of the four precipitation events described in Section 2. These four estimates of \( \gamma \) ranged between 0.40 and 0.70 with rather flat minima. However, for purposes of illustrating the
methodology, it was decided to use a common value of $\gamma = 0.60$ for all four examples and thereafter to treat $\gamma$ as a fixed constant in model (3) for purposes of estimating a separate $\alpha, \beta$ for each precipitation event.

From (1) and (2) we have the following interpretations: $\gamma = 0.60$ means that 40% of the available nitrate in the air column is scavenged by 1.0 millimeters of precipitation falling in one hour. More generally, if $V$ millimeters of precipitation fall in one hour (as snow) then the scavenging model states that the fraction of the available nitrate scavenged from the air column is

$$ (1 - \gamma^{V^{0.52}}) \times 100\%. $$

(9)

A short table is given below using, for illustration, the estimated $\gamma = 0.6$.

| one-hour precipitation intensity (mm): | 1.0 | 2.0 | 4.0 | 8.0 |
| scavenged fraction (% nitrate):        | 40% | 52% | 65% | 78% |

As a further example, if there were 8.0 mm of precipitation (snow) falling as 2.0 mm per hour over a four hour period then the calculated precipitation surface area is proportional to $4 \times (2.0)^{0.52}$ and the scavenged fraction obtained from (1) is 95%. This compares with 78% from the table above for 8.0 mm all falling in one hour.

With the scavenging parameter treated as a fixed constant, the model (3) for the measured hourly nitrate concentrations in precipitation is a linear model in the parameters $\alpha, \beta$ of interest. The parameter $\alpha$, interpreted as the time-invariant cloud-water nitrate concentration, has units of micro-equivalents/liter; the parameter $\beta$, interpreted as the initial total nitrate in the air column, has units of micro-equivalents/meter-square.

The two parameters $\alpha, \beta$ were estimated for each of the four example precipitation events using ordinary least squares. However, the data for the first hour of each event was more or less anomalous and was not used in fitting of the model. The four sets of parameter estimates $\hat{\alpha}, \hat{\beta}$ are shown in Table 2 below.
Table 2. Model parameter estimates

<table>
<thead>
<tr>
<th>Precipitation Event</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\beta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 26, 1976</td>
<td>4.5 (1.3)</td>
<td>126.8 (16.1)</td>
</tr>
<tr>
<td>January 7, 1977</td>
<td>48.4 (17.8)</td>
<td>384.0 (101.7)</td>
</tr>
<tr>
<td>January 14, 1977</td>
<td>19.3 (2.1)</td>
<td>105.8 (24.0)</td>
</tr>
<tr>
<td>January 20, 1979</td>
<td>2.2 (0.4)</td>
<td>22.8 (1.9)</td>
</tr>
</tbody>
</table>

If the residuals in model (3) are taken to be uncorrelated and homoscedastic and \( \gamma \) a fixed constant, then the nominal standard errors of the regression coefficients are those shown in parentheses in the table above.
5. Estimates of Nitrate Deposition

The parameter estimates $\hat{\alpha}$, $\hat{\beta}$ were used to calculate the fitted values $\hat{Y}_i = \hat{\alpha} + \hat{\beta} \cdot X_i(\gamma)$ of the time series of precipitation nitrate concentrations. The explanatory variable $X_i(\gamma)$, for $\gamma = 0.60$, is calculated from the corresponding observed precipitation volume time series $V_i$ using (2) and (6). The resulting fitted values, $\hat{Y}_i$, for each of the four precipitation events are shown together with the measured values, $Y_i$, in Figure 4.

The fitted version of model (3) may now be used to separately estimate the total nitrate deposition over a whole precipitation event, due respectively to the initial concentration of nitrate in cloud water and the initial total nitrate in the air column. One might then interpret these two deposition estimates as estimates of imported versus local source contributions to nitrate deposition.

The first of these estimated deposition quantities, due to cloud water nitrate, will be the product of the initial concentration estimate $\alpha$ and the measured total rainfall volume, viz.,

\[
\text{deposition (cloud source)} = \hat{\alpha} \cdot \sum V_i \quad \text{micro-equivalents/meter-square}
\]

\[
(10) \quad \text{deposition (air column source)} = \hat{\beta} \cdot (1 - \gamma^{\sum V_i^{0.52}}) \quad \text{micro-equivalents/meter-square.}
\]

The air column source deposition estimate in (10) above will usually be close to $\hat{\beta}$ except for very short precipitation events. Even for the January 7, 1977 event, $\hat{\beta}$ differs from the estimated scavenged nitrate by less than 5%.

Table 3 below shows cloud and air column estimated contributions to the nitrate deposition for each of the four precipitation events described in Section 2, using (10) above and the parameter estimates of Table 2. The sum of the two contributions should, in principle, equal the total measured deposition given by the volume weighted sum of measured hourly concentrations, viz.,

\[
(11) \quad \sum (V_i \cdot Y_i)
\]

which is also shown in Table 3. The discrepancy between the measured total deposition and the sum of the two estimated contributions to deposition is due to the structure imposed by the model. The fact that the discrepancies seem small is somewhat reassuring.
Table 3. Estimated cloud and air column contributions to deposition

<table>
<thead>
<tr>
<th>Event Date</th>
<th>Duration (hours)</th>
<th>Precipitation total (mm)</th>
<th>Deposition from cloud</th>
<th>Deposition from air</th>
<th>Deposition total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec 26, 1976</td>
<td>8</td>
<td>12.6</td>
<td>166</td>
<td>57</td>
<td>126</td>
</tr>
<tr>
<td>Jan 7, 1977</td>
<td>7</td>
<td>5.2</td>
<td>524</td>
<td>250</td>
<td>370</td>
</tr>
<tr>
<td>Jan 14, 1977</td>
<td>9</td>
<td>12.2</td>
<td>341</td>
<td>235</td>
<td>105</td>
</tr>
<tr>
<td>Jan 20, 1979</td>
<td>20</td>
<td>95.1</td>
<td>227</td>
<td>209</td>
<td>22</td>
</tr>
</tbody>
</table>

(micro-equivalents/meter-square)

It is interesting to note that the events of December 26, 1976 and January 14, 1977, of comparable duration and total precipitation, have total nitrate depositions differing by a factor of two and estimated cloud-water nitrate contributions differing by a factor of four. For large precipitation events, such as the January 20, 1979 event, one would expect most of the deposition of nitrate to be contributed by cloud water nitrate since the available nitrate in the air column is largely scavenged early and is then overwhelmed by the continuing cloud water contribution.

This example of an attempt to model hourly data in a simple fashion via a scavenging model is still incomplete in many respects. In particular, work is continuing on the problem of the uncertainty of the scavenging efficiency parameter and the effects of this uncertainty on final deposition contribution estimates. The question of which model parameters should be pooled across precipitation events and which should be event specific is not well resolved. Finally, the question of the possible time variability of cloud water nitrate concentrations has not been addressed. So in many respects this work must be regarded as preliminary; its objective is to show how information relating to simple physical processes might be extracted from hourly data.
References


