Meteorological Adjustment of Chicago, Illinois, Regional Surface Ozone Observations with investigation of Trends

Joel H. Reynolds David Caccia Paul D. Sampson
Peter Guttorp

NRCSE

Technical Report Series

NRCSE-TRS No. 025

Meteorological Adjustment of Chicago, Illinois, Regional Surface Ozone Observations with investigation of Trends

Joel H. Reynolds David Caccia Paul D. Sampson
Peter Guttorp

June 11, 1999

Abstract

Meteorological adjustment of surface ozone records from 10 monitors in the Chicago region was conducted by using the SVD to generate univariate network summaries of the ozone and surface temperature networks. The summaries were then used in a constructing a regression model to remove the influence of meteorology on surface ozone. Remaining temporal trends were investigated using generalized additive models; no significant trend was found. This lack of trend is discussed in comparison to other published analyses of Chicago area ozone data, resulting in guidelines for future meteorological adjustment investigations.

1 Introduction

Tropospheric ozone is known to reduce both human and vegetation health as well as decrease visibility ([NRC 91], [Wax 91]). In response to this, the EPA has instituted air quality standards, with recently proposed revisions [EPA 95], that each state must meet or else face noncompliance. In order for a state to develop plans for controlling ozone exceedences, it must first account for the influence of meteorological conditions conducive to ozone formation. Historical trends due to changing precursor emissions can then be investigated in the meteorologically-adjusted ozone record [NRC 91].

Meteorologically-adjusted ozone is obtained by statistically modeling the association between surface ozone observations and meteorological variables generally collected nearby [NRCSE 98]. Most trend analyses model meteorological data from a single location with ozone observations from a single monitor [NRCSE 98]. In cases where a spatial network of ozone monitors is available, the literature contains two basic approaches to modelling the association between ozone and meteorology [NRCSE 98]: independently model the dependence of ozone response for each monitor or model the dependence of an ozone network summary (e.g., median daily 1 hour maximum ozone from the network [Bloom 96]). Fitting ozone monitor- specific models directly to local surface meteorology ignores any information content on regional dynamics of meteorology and ozone in the simultaneous analysis of the whole ozone network (Figure 1); using an ozone network summary captures more of this regional information. The literature contains no trend analyses associating ozone

network observations with a spatial network of meteorology observations, besides that of [Reynolds 98] which the current analysis extends. If spatial meteorological observations are available, using data from only a single location ignores information on regional-scale meteorological influences on ozone production.

Surface ozone observations from the Chicago, Illinois, U.S.A., metropolitan region have been used to demonstrate a number of statistical approaches to meteorological adjustment ([Bloom 96], [Smith 93], [Milan 98], [Niu 96], [Huang 99]). The approaches vary in statistical methodology (respectively: nonlinear regression, extreme events models, time-series filtering, nonlinear additive time series models, CART models) and spatial range of ozone response (single monitor or network summary). The analysis reported here records our effort to model the regional-scale association between the Chicago area ozone monitor network and selected meteorology variables (see Section 2), including a network of surface temperature monitors, with the ultimate goal of assessing the presence of long-term ozone trends remaining after meteorological adjustment.

The analysis has three stages: (1) determination of univariate summaries of the ozone and surface temperature networks which capture the dominant patterns of association between the two spatial fields, (2) regression modeling of the univariate regional ozone summary derived in stage (1) as a response to the univariate surface temperature summary and other meteorological measurements, and (3) assessment of long-term trend by including in the stage (2) model a nonparametric regression component describing long-term trend . The approach differs from previous analyses in a number of major details. In stage (1), Canonical Covariance Analysis is used to derive the univariate ozone network and surface temperature network summaries. This is undertaken as an interative process, allowing selection of the most informative subset of surface temperature monitors to overcome varying data quality (see Section 4.3). In stage (2), model selection is peformed using best subsets regression [Ryan 97]. In stage (3), long-term trend is assessed by extending the linear regression model derived in Stage (2) to a generalized additive model by including a nonparametric regression of long-term trend. This allows assessment of long-term trend without assuming a potentially inappropriate linear form [NRCSE 98].

2 Data Sources, Selection and Description

2.1 Ozone data

The ozone data consists of hourly measurements (in ppm) from a subset of ten ozone monitors in the Chicago, Illinois, region selected from the EPA AIRS database (Aerometric Information Retrieval System, http://www.epa.gov/docs/airs/airs.html). A monitor was selected if it had over 80% complete records for the 1981 - 1991 time period (Table 1, Figure 2). The number of active monitors on any given day varied from 4 to 10 with a daily average of 9.3 monitors (Table 2). The ozone season was defined as April 1 to october 31, following [Bloom 96].

Maximum 1 hour average ozone readings were calculated as daily summaries of each monitor's record. Following the EPA Region X guidelines, these daily maximum are calculated over the interval 9 am - 9 pm LST.

Figure 1: Relationship of co-occuring daily maximum 1 hour ozone readings across each pair of monitors in the ozone network.

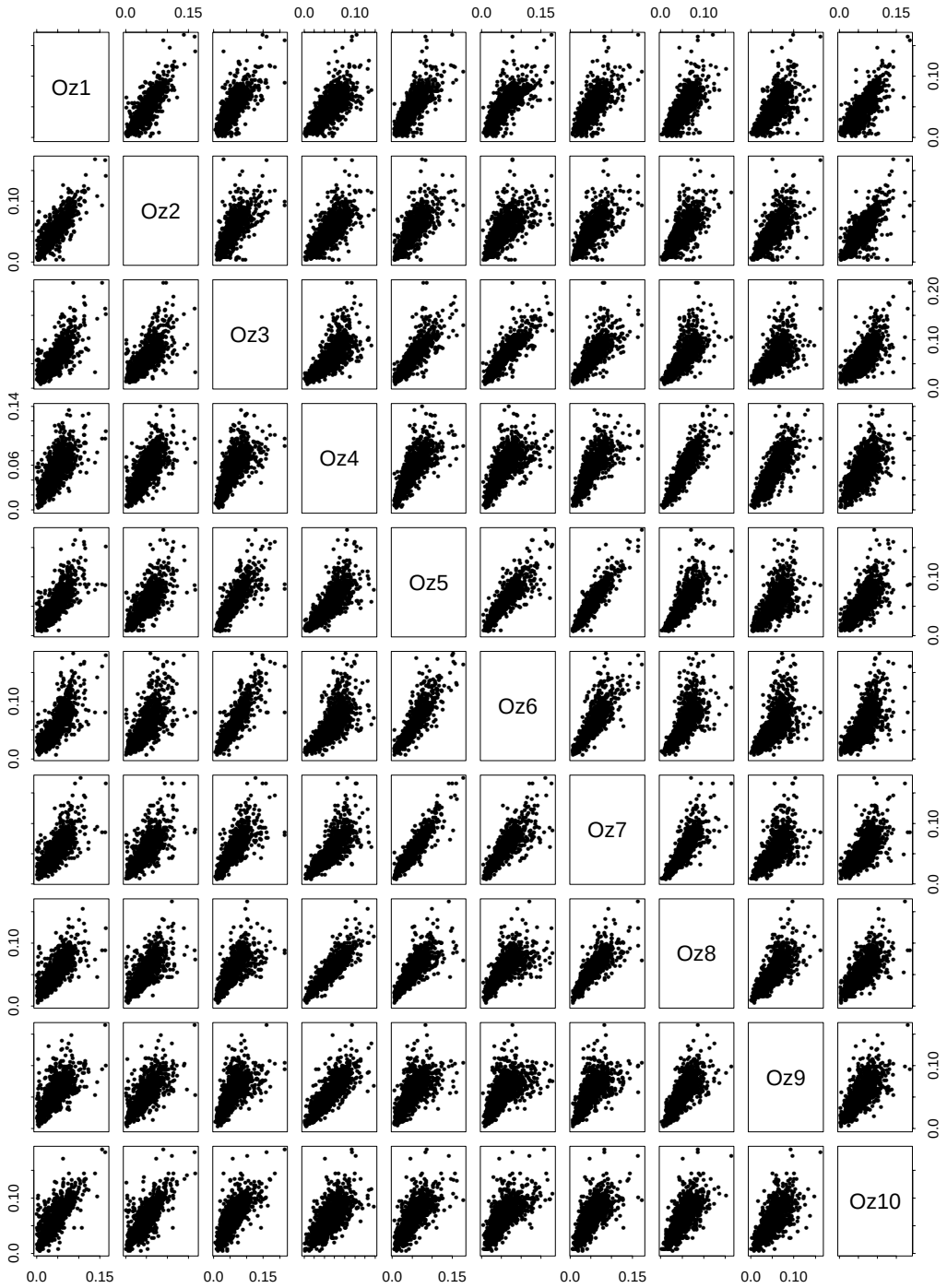


Table 1: EPA A.I.R.S. Chicago region ozone monitors used in the analysis.

Monitor	AIRS ID	Latitude	Longitude	Record Period	% Complete[a]
Oz1	170310050	41.7072	87.5686	1/82 - 12/91	83.2
Oz2	170311002	41.6164	87.5578	1/81 - 10/90	82.3
Oz3	170317002	42.0622	87.6736	1/81 - 12/91	97.7
Oz4	170890005	42.0494	88.2731	1/81 - 12/91	96.9
Oz5	170970001	42.1772	87.8647	1/81 - 12/91	95.8
Oz6	170971002	42.3869	87.8414	1/81 - 12/91	96.5
Oz7	170973001	42.2906	87.9817	1/81 - 12/91	96.1
Oz8	171110001	42.2211	88.2411	1/81 - 12/91	96.7
Oz9	171971008	41.5761	88.0547	1/81 - 12/91	93.3
Oz10	180892008	41.6394	87.4936	1/81 - 4/90	92.4

[a]Refers strictly to April 1 - October 31 ozone seasons, 1981 - 1991.

Figure 2: Selected Chicago, Illinois region ozone monitors (Oz*) and surface meteorology sites (T*). See Table 1 for ozone and Table 3 for surface meteorology monitor details. Lake Michigan forms the boundary shown in the upper right of the map. Dashed lines denotes shared boundary of small reanalysis grids (see Table 5).

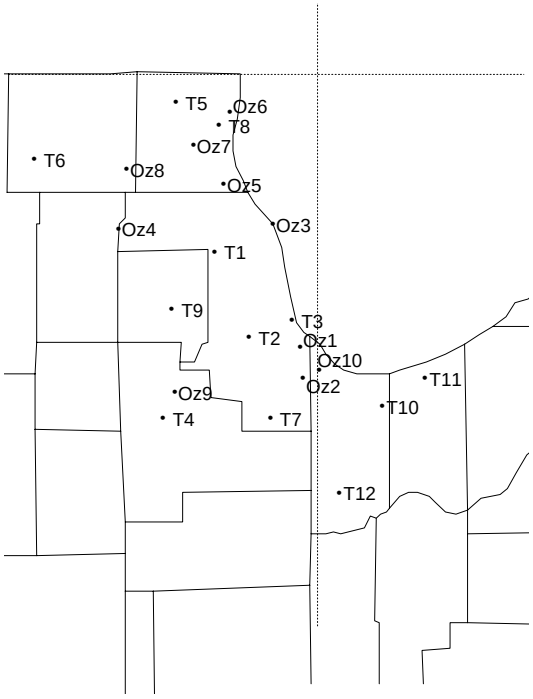


Table 2: Average number of active ozone monitors each season.

Year	Mean No. Monitors
1981	8.5
1982	9.3
1983	9.6
1984	9.3
1985	9.0
1986	9.4
1987	9.7
1988	9.4
1989	9.7
1990	9.7
1991	8.8

2.2 Surface Meteorology Data

Maximum daily temperature (degrees Fahrenheit) from active surface meteorology stations in the region were obtained from the Cooperative Observer Network Database of the National Climatic Data Center (NCDC) [Hydro 95] (Table 3). Only monitors with records over 95% complete for the analysis period were considered, resulting in a network of 12 monitors. In addition, further meteorology variables measured at the O'Hare Aiport were included (Table 4). The O'Hare data was acquired from the NOAA Midwest Climate Center (http://mcc.sws.uiuc.edu/index.html).

The surface meteorology records vary in their completeness (Figure 3). Considered as a network, complete temperature records exist for 85.8% of the analysis period (approximately 100% if considering just O'Hare (T1) and Midway (T2) airports).

2.3 Reanalysis Meteorology Data

Gridded tropospheric and surface boundary meteorological data were obtained from the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Reanalysis Project (http://www.cdc.noaa.gov) as representative measures of regional meteorology characteristics [Kalnay 96]. The reanalysis project uses a state-of-the-art global atmospheric model combined with as complete a database as possible to provide a variety of spatially-gridded meteorological fields.

The predicted fields differ in their spatial and temporal resolution, as well as in their reliability [Kalnay 96]. NCEP provides a four-level classification of variable reliability, though we are only concerned with the first three classifications:

A variables strongly influenced by observed data (and hence most reliable),

B variables partly defined by observations but which are strongly influenced by the model characteristics (and hence should be interpreted with caution),

C variables not directly observed whose values are derived solely from the model (and hence should be interpreted with caution),

Table 3: National Weather Service/NOAA Cooperative Observer Network surface meteorology stations.

Surface Station Label	**Station ID**[a]	**Latitude**	**Longitude**	**% Complete**[b]
T1 (O'Hare Airport)	1549	41:49	87:54	100
T2 (Chicago Midway Airport)	1577	41:44	87:46	100
T3 (Chicago University)	1572	41:47	87:36	99.9
T4	4530	41:30	88:06	96.0
T5	4837	42:35	88:03	98.6
T6	5326	42:15	88:36	98.6
T7	6616	41:30	87:41	96.5
T8	9029	42:21	87:53	97.2
T9	9221	41:49	88:04	96.9
T10	4008	41:32	87:15	100
T11	4244	41:37	87:05	98.9
T12	5174	41:17	87.25	97.4

[a]Cooperative Observer Network ID.
[b]Refers strictly to ozone seasons, 1981 - 1991.

Table 4: Surface meteorology variables from Chicago O'Hare International Airport (surface monitor T1). Variables in lower section of table are derived.

Variable Name	**Description (units)**
RelHum	Relative Humidity (%)
BarP	Barometric Pressure (inches mercury?)
CloudCover	Percent cloud cover (to nearest 10%)
WindSpd	Wind Speed (average? noon? m/s)
WindDir	Wind Direction (average? noon? (radians from North, measured clockwise)
DelBarP	$\mid \text{BarP}_{today} - \text{BarP}_{yesterday} \mid$
DelWindSpd	$\mid \text{WindSpd}_{today} - \text{WindSpd}_{yesterday} \mid$
DelWindDir	$\text{WindDir}_{today} - \text{WindDir}_{yesterday}$

Figure 3: Availability of surface meteorology records, by site (see Table 3), for April 1 - October 31 each year. Black denotes missing records.

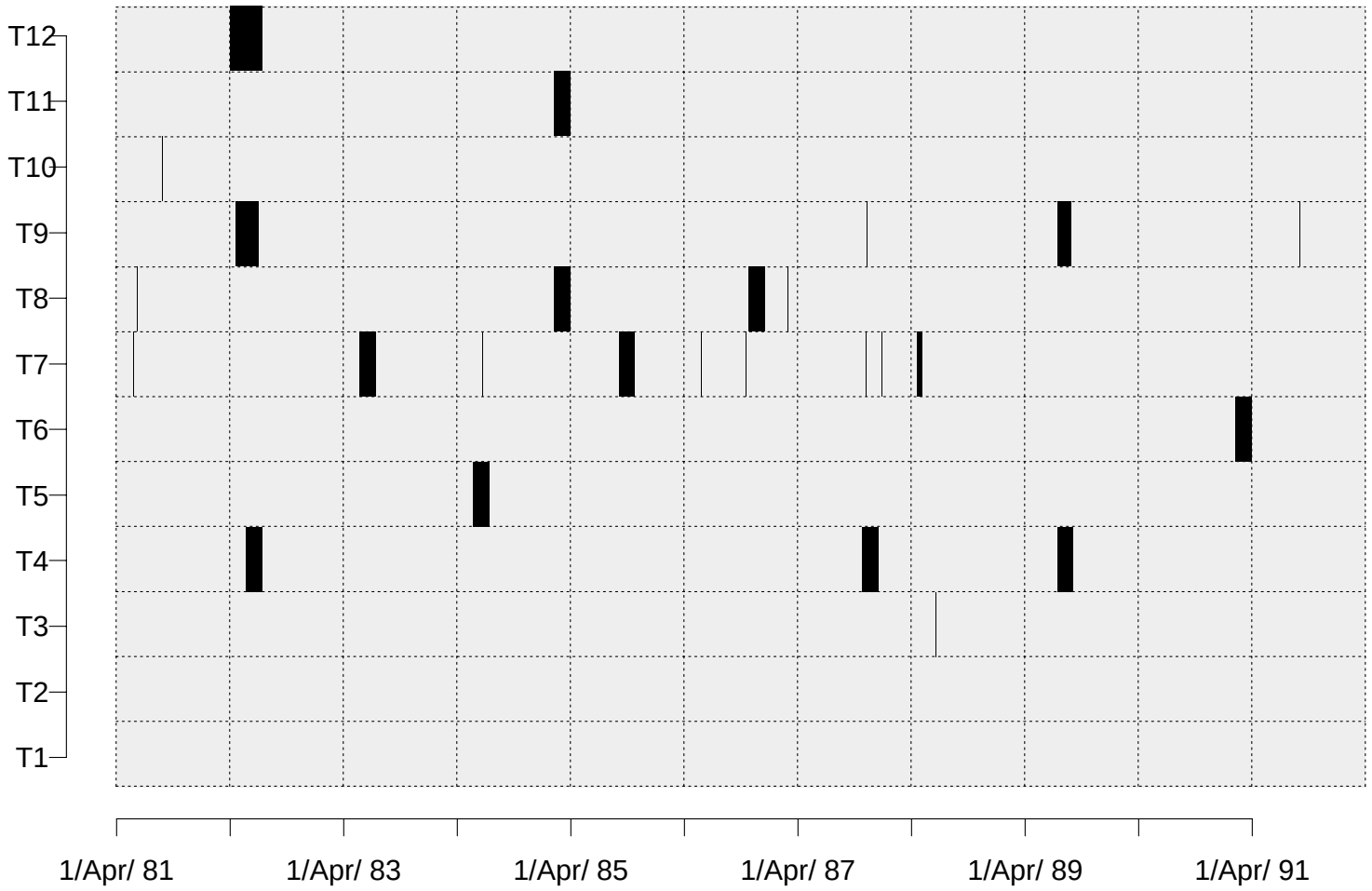


Table 5: NCEP/NCAR Reanalysis Fields Selected for Analysis

Variable	Description (units) [a]	Grid [b]	Reliability [c]
GP1000	Mean[d] geopotential height of 1000 mbar pressure field (m)	L	A
GP850	Mean geopotential height of 850 mbar pressure field (m)	L	A
T1000	Mean temperature of 1000 mbar pressure field (K)	L	A
T850	Mean temperature of 850 mbar pressure field (K)	L	A
T2m[e]	Mean temperature 2 m from surface (K)	S	B
WindU	Mean magnitude of U[f] wind component 10 m from surface (m/s)	S	B
WindV	Mean magnitude of V[g] wind component 10 m from surface (m/s)	S	B

[a] All variables are spatial means over the grid cell.

[b] L - large: 2.5 degree latitude x 2.5 degree longitude grid; S - small: 'T62 Gaussian grid' 1.875 degree latitude x 1.9047 degree longitude.

[c] See Reliability Classifications in text.

[d] Daily means (0000, 1200 GMT).

[e] The region is covered by one large grid cell and two small grid cells, so T2m1 is the value of variable T2m in cell 1 (leftmost cell). Similarly for WindU and WindV.

[f] Eastward component.

[g] Northward component.

D variables obtained from observed data with no dependence on the model.

The meteorological fields selected for analysis, their spatial grid dimension, temporal resolution, and reliability are given in Table 5. The study region was contained completely within one of the larger grid cells and two of the smaller grid cells (Figure 2).

Discussion of known problems or caveats regarding the reanalysis project predictions can be found at http://www.cdc.noaa.gov/cdc/reanalysis/problems.htm. Those of importance to this analysis concern the calculation of the near surface variables T2m, Wind Magnitude, and Wind Direction (Table 5). Previous use of reanalysis data for meteorological adjustment of surface ozone is discussed in [Reynolds 98].

2.4 Derived Variables

Measures of meteorological stability were derived from the O'Hare surface data (the Del* variables in Table 4). The seasonal variation in solar intensity was approximated as:

$$Luminosity_{Julian\ day\ i} = sin(latitude\ of\ Chicago + Earth's\ tilt * sin(2\pi * (i - Spring\ Equinox\ date)/365),$$

with latitude set to 48^o and tilt to 23^o. This variable was investigated as a more process-based representation of short-term seasonal trend than a fourier series.

3 Method

3.1 Approach

Exploratory data analyses of the ozone monitor records and the various meteorological variables demonstrate common relationships across the ozone monitor network (e.g., Figure 1 and contents of Appendix A). It is therefore reasonable to pursue an analysis which captures the regional ozone response to meteorology. Preferably, the analysis employed to reveal the dominant patterns of temporal association between the ozone spatial field (defined by the 10 ozone monitors) and the surface temperature spatial field (defined by the 12 surface temperature monitors) will also provide a means of deriving univariate summary measures for each field which approximate this coupling. The association between the summary measures can then be modeled with standard regression techniques, leading to a meteorologically adjusted regional ozone summary. The possibility of a remaining long-term trend in the ozone summary is investigated by including a nonparametric regression component for long-term trend in the meteorological adjustment regression model, producing a generalized additive model [Hastie 90].

3.2 Deriving A Regional Summary

The network median daily 1 hour maximum ozone has been used as a regional ozone summary[1] [Bloom 96]. This summary was suggested by a Principal Component Analysis of the selected ozone network [Bloom 93]. However, we have no reason to assume a priori that this summary will be best for the ozone network selected here. More importantly, it doesn't necessarily follow that this summary will reveal the dominant patterns of association with the surface temperature field, nor is there an obvious a priori regional summary of the surface temperature field.

The temporal cross-covariance and cross-correlation matrices of the ozone x surface temperature fields summarizes the association between each ozone monitor and each surface temperature monitor (see Appendix A, Table 13, and Figure 13), summarizing the patterns of association between these spatial fields. Canonical Covariance Analysis applies the singular value decomposition (SVD) to the cross-covariance matrix to reveal ([Breth 92], [Wall 92]):

- the number of dominant patterns of association among the ozone and surface temperature monitors which combine to optimally approximate the observed cross-covariances[2];

[1] Median polish was used to account for shifting patterns of missing data across the ozone monitors.

[2] See [Breth 92] for a detailed description of this use of the SVD. 'Optimality' here means maximizing the cumulative squared covariance fraction (CSCF): Let $\tilde{C}_{Napprox}$ be the rank N approximation to the cross-covariance matrix reconstructed from the first N SVD patterns, C_{obs} be the observed cross-covariance matrix. Then the

$$CSCF_N = 1 - \frac{\|\tilde{C}_{Napprox} - C_{obs}\|^2}{\|C_{obs}\|^2}$$

where

$$\|C\|^2 = \sum_i \sum_j C_{ij}^2.$$

as well as

- the weight coefficients used to produce the linear combination of ozone monitors and the linear combination of surface temperature monitors which best approximate, respectively, the ozone and surface temperature components of each of these dominant patterns of association[3].

We can thus model the pattern of association between the fields by

1. For each of the dominant patterns, calculating the weighted linear combinations of monitors which approximates the ozone and surface temperature fields[4], (i.e., one time series for each field for each dominant pattern),

2. For each field, combining the most informative N pattern approximations to achieve a univariate approximation to the field's component of association, (i.e., one time series for each field, approximating the coupling between the two spatial fields),

3. modeling the univariate approximation to the ozone field as a response to the univariate approximation of the surface temperature field and other meteorology variables. This produces a residual time series which represents the meteorological-adjusted ozone network summary.

This approach allows us to utilize the full spatial networks of both ozone and surface temperature in deriving our daily ozone network summary and it's association with meteorology.

3.2.1 Surface Meteorology Transformations

As the relationship between maximum daily 1 hour ozone and maximum daily surface temperature is known to be nonlinear [NRC 91], a monotone nonparametric transformation

[3]The weight coefficients $\{w_i\}$, $\{v_j\}$ are chosen to maximize the $\text{cov}(\sum w_i Ozone_i, \sum v_j Surface\_Temperature_j)$ under the restriction that each field's set of weight coefficient vectors (one vector for each pattern) are orthonormal. This differs from Canonical Correlation Analysis (CCA) in this orthonormality restriction on the weight coefficients for the patterns. The restriction to orthonormal patterns allows one to reconstruct an optimal approximation to a field's original time series from the linear combination of selected patterns (where here 'optimal' refers to the property that the residual time series, i.e. the true time series - approximate time series, is orthogonal to each of the patterns which were combined to form the approximation) [Breth 92].

[4]Each pattern is associated with scalar called a singular value, σ. A measure of the importance of the k^{th} pattern is obtained from

$$\frac{\sigma_k^2}{\sum_j \sigma_j^2},$$

which is equal to the fraction of the total squared covariance in the observed cross-covariance matrix explained by the approximation based on this particular pattern. Note,

$$\sum_j \sigma_j^2 = \|C_{Obs}\|^2.$$

See [Breth 92].

of surface temperature was derived to achieve an approximately linear relationship with each ozone monitor. The transformation was developed by first deriving a monotonic nonparametric transformation[5] of surface temperature for each pair (Ozone monitor$_i$, Surface Temperature monitor$_j$) to maximize the $corr$(Ozone monitor$_i$, transformed Surface Temperature monitor$_j$). The transformations were sufficiently consistent across monitor pairs (Figure 4) to suggest a final single transformation could be derived by averaging, across all (Ozone monitor$_i$, Surface Temperature monitor$_j$) pairings, the transformed values at each temperature value with 5 or more observations. A nonparametric smooth[6] was then fit to this average transformed value. This transformation was applied to each surface temperature monitor before generating the cross-covariance matrix for the SVD.

Surface wind direction is an angular measurement and hence needs tranformation to account for its periodicity before being used as a predictor in linear regression. A nonparametric smoothing algorithm was used[7], with a periodic constraint, to find the transformation which maximally correlated transformed surface wind direction with the ozone network summary.

3.2.2 Network Subset Selection

The cross-correlations (Figure 13) suggest that not all surface temperature monitors are equally informative, perhaps due to varying data quality characteristics. The strongest association between fields may utilize only a subset of the surface temperature network. Since the SVD cannot assign weights of zero to uninformative monitors, the most informative surface temperature network subset is found by applying the SVD to various subsets suggested by (Figure 13). The surface temperature subset whose linear combination most highly correlates with its associated linear combination of ozone monitors is selected for the final analysis.

3.3 Calculating Univariate Summaries: Imputing Missing Observations

The calculation of a spatial field's univariate summary from the dominant pattern's weight coefficients requires records be available for every site in the spatial field every day. The data obviously do not satisfy this requirement (Tables 2, 3, Figure 3).

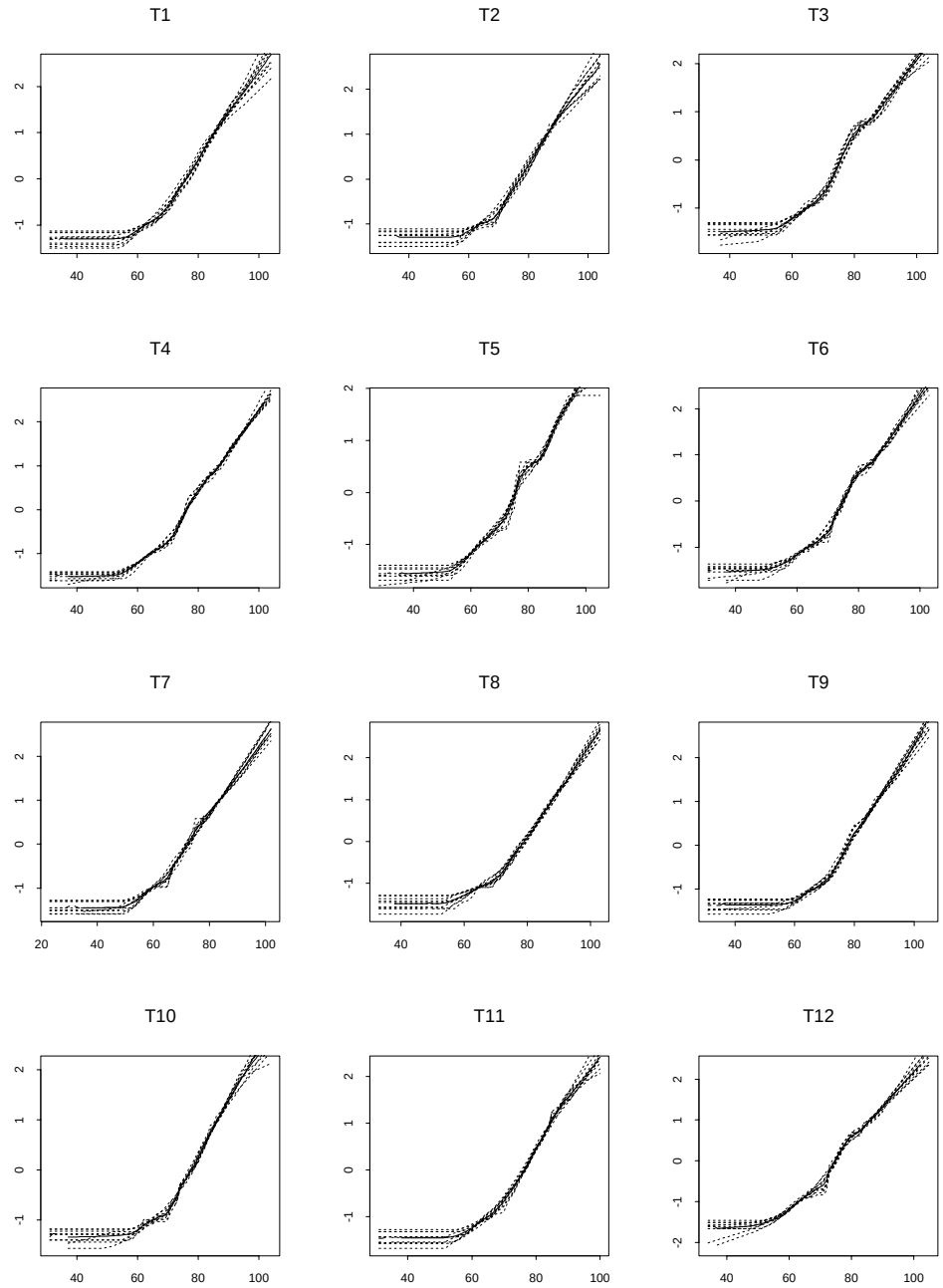
In the case of missing predictor variables (i.e., surface temperature), we refrain from imputing the missing records as we don't want to assume any temporally fixed structure among variables. Days with incomplete surface temperature network records must simply be dropped from the analysis.

This is not true in the case of missing response variables (ozone records). The underlying assumption of a regional-scale analysis is that ozone is a function of regional meteorology, and, therefore, on a given day there is a discernible relationship among the ozone sites as they are all being driven by the same regional meteorology. Statistically, we are assuming a temporally stationary spatial covariance matrix, $\sum_{ozone}$, across the ozone network. We can impute missing ozone monitor records for a given day from the ozone monitors observed that day by regression equations constructed from the estimated covariance matrix, $\hat{\sum}_{ozone}$.

[5] The 'super smoother' function, supsm, in S-Plus 3.4 [MathSoft 96].

[6] Specifically, a b-spline with 5 df [**?**].

Figure 4: Monotonic nonparametric transformation of each surface temperature monitor with each ozone monitor. The transformations in each plot are derived to maximize the correlation of a specific ozone monitor with the surface temperature monitor labelling each plot (dashed curves). The solid curve denotes the spline smooth of the average transformed values for each observed temperature. The x-axis displays untransformed maximum surface temperature (degrees Fahrenheit), the y-axis the transformed value specified only to relative scale. Note the strong consistency of transformations among monitors.



In the presence of missing data, the covariance matrix is estimated using the EM algorithm [Little 86]. Technical details are provided in Appendix B. Note that the imputed ozone values are only used in the construction of the daily ozone network summary from the SVD's weight coefficients; the cross-covariance matrix used to generate the weight coefficients is based solely on the observed values.

3.4 Modeling Issues

The need for ozone scale transformations is investigated as part of the meteorological adjustment model diagnostics. Transforming ozone requires subsequent recalculation of the surface temperature transformation, the cross-covariance matrix, the SVD, and EM imputation for calculating the univariate ozone summary. Model selection is performed by using best subsets regression[7] to determine the most promising models in terms of *adjusted* R^2 [Ryan 97]. Models over a range of sizes are then compared in terms of overall fit, residual characteristics (autocorrelation, normality, homoscedasticity, etc.), and model structure complexity.

Short-term trend was modeled both as a four component fourier series (annual and semi-annual terms, following [Bloom 96]) and as a single component approximating the seasonal change in relative solar intensity at Chicago (*Luminosity*, described in Section 2.4). Both representations produced comparable reductions in residual squared error, so *Luminosity* was used as it involved fewer parameters.

The agreement between the observed and predicted 95^{th} percentiles each summer was investigated to assess model adequacy. Specifically, the 95^{th} percentile for season i was predicted as follows:

1. Estimate the model coefficients using all observations except those from season i.
2. Predict season i's daily ozone response using this fitted model.
3. Add to each predicted daily response a randomly selected residual from the Step 1 model fitting. Repeat this step B times, resulting in an empirical prediction distribution of B values for each day in the season.
4. Combine daily empirical distributions, producing an empirical prediction distribution for the whole season consisting of B**length of season* values.
5. Select the 95^{th} percentile from this distribution.
6. Repeat Steps 1 - 5 for each year.

3.5 Investigation of Long-Term Trend

Once a regression model has been developed for meteorological adjustment of the univariate ozone summary, the most appropriate method for assessing long-term trend is by incorporating a trend term into the model [NRCSE 98]. In order to avoid a priori assuming a specific parametric form for the trend, trend significance is initially investigated using nonparametric

[7] As implemented in the leaps function in S-Plus v. 3.4, [MathSoft 96].

regression in a generalized additive model framework [Hastie 90]. If temporal dependence in the residuals undermines the asymptotic properties of the analysis of deviance, significance of the long-term trend component can be assessed by using the block bootstrap [Davison 97] on the meteorological adjustment residuals to generate a 95% pointwise confidence envelope of nonparametric trend regressions under the hypothesis of *no* long-term trend. If either analysis suggests that the trend component is informative, then an appropriate parametric form may be developed to approximate the trend.

3.6 Summary of Analysis Steps.

The complete analysis sequence is:

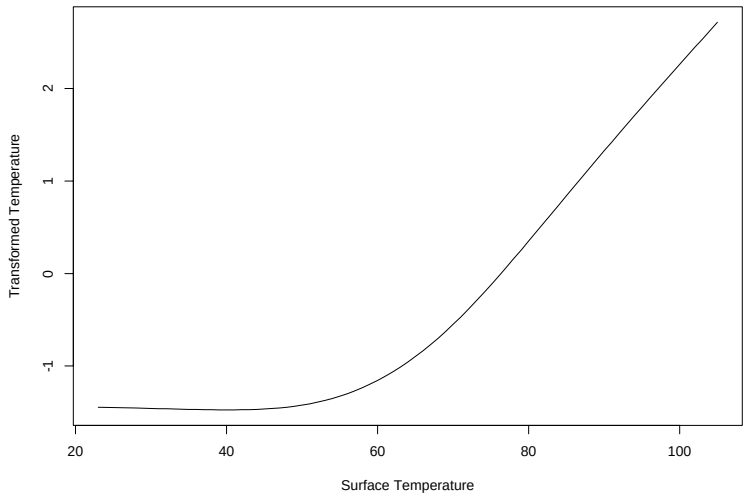
1. Derive linearizing transformation for surface temperature records.
2. Calculate ozone x transformed surface temperature cross-covariance matrix.
3. Calculate SVD of cross-covariance matrix.
4. Investigate singular values to determine number of dominate coupling patterns among fields.
5. Impute missing ozone observations using the EM algorithm.
6. Calculate univariate composite summaries of dominant ozone (comp.03) and surface temperature (comp.Temp) patterns.
7. Repeat steps (3) - (6) on selected submatrices of the cross-covariance matrix, determining the most informative subset of the surface temperature monitor network.
8. Use best subsets regression of comp.03 on comp.Temp and other meteorological predictors, including a seasonal trend term, to select the best meteorological adjustment model.
9. Conduct model diagnostics, investigating the need for ozone scale transformation or nonlinear representations of meteorological predictors. If a transformation of ozone is require, restart the analysis using suitably transformed ozone records.
10. Add a nonparametric regression component for long-term trend into the selected meteorological adjustment model, and assess its significance using either analysis of deviance or the block bootstrap.

4 Results

4.1 Ozone Scale Transformations

Initial analysis revealed that residuals from meteorological adjustment of the univariate ozone summary based on the natural scale (raw) daily 1 hour maximum ozone measurements were positively-skewed. Logarithmic and square root transformations of the ozone monitor records

Figure 5: Final monotonic nonparametric transformation of surface temperature. Derivation described in Section 3.2.1. The x-axis displays untransformed maximum surface temperature (degrees Fahrenheit), the y-axis the transformed value specified only to relative scale.



were investigated. Logarithm transformations overcorrected, producing a negatively skewed residual distributions. All subsequent results reported are from the final analysis using square root transformed ozone records (see Figure 8 for result of transformation).

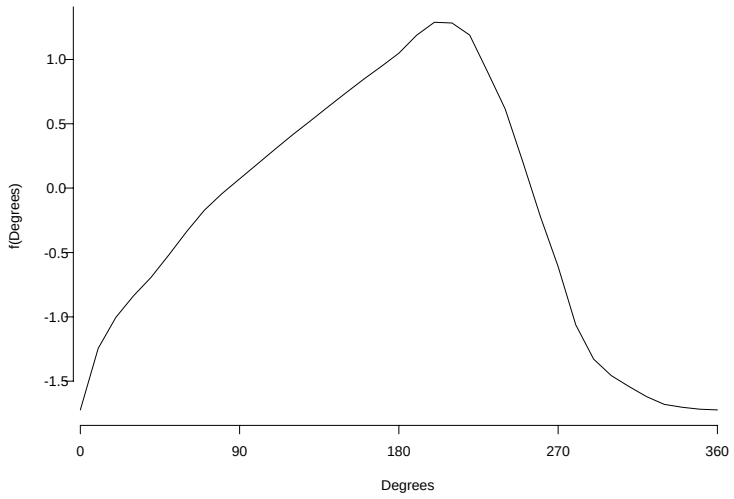
4.2 Linearizing Transformations

The transformations across ozone x surface temperature monitor pairs were highly consistent, reinforcing a regional perspective of the fields' association. A final single transformation was therefore derived (Figure 5). The cross-covariance and cross-correlation matrices of the square-root ozone network and linearized surface temperature network displayed little spatial variability (Appendix A). The periodic nonparametric transformation of surface wind direction is shown in Figure 6.

4.3 Surface Temperature Network Subset Selection

As suggested by Figure 13, the inclusion of surface temperature monitors T3 - T10 (Table 3) reduced the association between the SVD-derived univariate summaries. Presumably this is due to a drop in monitor record quality as one moves away from the regional airport monitors (T1 - O'Hare, T2 - Midway) to the other cooperative network monitors. All further results are from the analysis of the square-root transformed ozone network and the transformed surface temperature subnetwork consisting of monitors T1 and T2 (Table 3).

Figure 6: Periodic nonparametric transformation of surface wind direction. The x-axis displays untransformed wind direction (measure clockwise, from north), the y-axis displays the transformed value specified only to relative scale.



4.4 Singular Value Decompositions

The SVD of the cross-covariance matrix revealed one dominant pattern of association between the ozone field and the surface temperature field. The rank one approximation to the observed cross-covariance matrix derived from this dominant pattern accounts for 99.99% of the observed matrix's cummulative squared covariance. I.e., only a single set of univariate summaries is needed to capture the association between the two fields.

The weight coefficients of this dominant pattern reveal the relative contribution of each monitor to the network composite summaries (Table 6). The lack of variation in coefficients across a network reveals little spatial variation in the strength of association between the variable and the opposing field. The univariate composite summary of the ozone network is therefore approximately proportional to the mean network ozone observation (square-root daily 1 hour maximum).

4.5 Imputation and Composite Calculations

Missing ozone observations were imputed, following the procedure outlined Section 3.3 and Appendix B, to allow calculation of the daily ozone network summary, comp.03. The imputation regressions were derived from the estimated covariance matrix of the square-root transformed ozone network (Appendix B, Table 14).

Table 6: SVD weight coefficients for the dominant pattern of association.

Monitor	Coefficient	Monitor	Coefficient
Oz1	0.334	T1	0.705
Oz2	0.313	T2	0.710
Oz3	0.335		
Oz4	0.295		
Oz5	0.322		
Oz6	0.346		
Oz7	0.302		
Oz8	0.272		
Oz9	0.284		
Oz10	0.350		

4.6 Model Selection for Meteorological Adjustment of Ozone Network Composite

Best subsets regression, selecting main effects and two-way interactions from {comp.Temp, *Luminosity*, and the variables in Tables 4 & 5}, suggests a range of models with from 3 to 5 terms (Figure 7, Table 7). The 3 and 5 term models suggested by the *adjusted* R^2 analysis were investigated. The additional terms in the models of order higher than 5, though producing slight increases in *adjusted* R^2, were insignificant even in the face of conservative standard error estimates.

Three term model:

$$comp.03 = \beta_0 + \beta_1 RelHum + \beta_2 Luminosity + \beta_3 comp.Temp * RelHum \quad (1)$$

with the parameter estimates and fit summaries given in Table 8.

Five term model:

$$\begin{aligned} comp.03 &= \beta_0 + \beta_1 RelHum + \beta_2 WindSpd + \beta_3 WindDir \\ &\quad + \beta_4 Luminosity + \beta_5 comp.Temp * RelHum \end{aligned} \quad (2)$$

with the parameter estimates and fit summaries given in Table 9.

4.7 Model Assessment

For both models, the residuals show no trends or heteroskedasticity and are approximately normally distributed (Figures 8 & 9) except for the top 7% - unsurprisingly, as these are the most extreme events. The 5 term model does not substantially improve model fit (Tables 8 & 9, Figure 10 - only the 1990 fit is improved), nor reduce the short-term residual autocorrelation (Table 10). Both models display similar long-term patterns in the residuals (Figure 11). Further analyses are recorded for both models, but Model 1 is the preferred choice for reasons of parsimony.

The correlation among predictors (Table 11) may result in some multicollinearity in parameter estimates (Tables 8 & 9), potentially raising concern for inflated standard error

Figure 7: Best *adjusted* R^2 as a function of model size as resulting from the best subsets regression procedure. Note model size includes intercept coefficient.

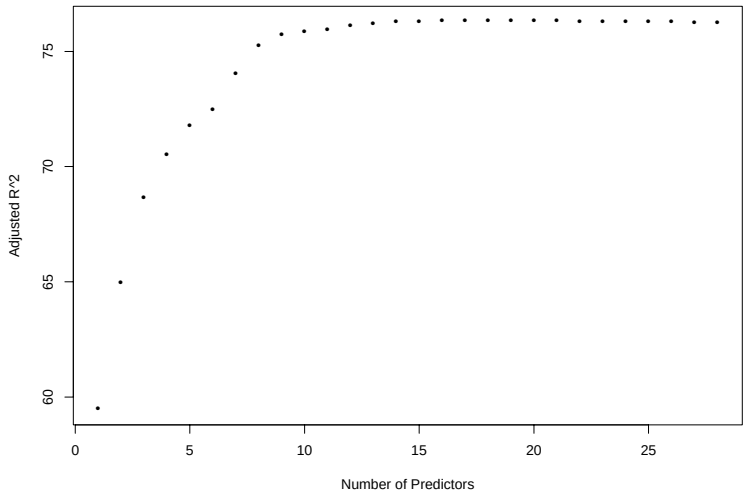


Table 7: Best adjusted R^2 by model size.

Size	**adjusted R^2**	**Additional Terms**
2	0.59	comp.Temp X RelHum
3	0.65	+ RelHum
4	0.69	+ Luminosity
5	0.70	+ WindSpd
6	0.72	+ WindDir
7	0.72	+ comp.Temp
8	0.74	+ 2mT2
9	0.75	+ WindU1
10	0.76	+ comp.Temp X Luminosity

Table 8: Coefficient Estimates & Correlations, Model 1

Coef	**Value**	**std.error**
β_0	0.645	0.0168
β_1	-0.0027	0.0001
β_2	0.2692	0.0162
β_3	0.0010	0.00002
Res. Std. Error	0.0783 on 2348 df	
Multiple R^2	0.687	
F-statistic	1715 on (3, 2348) df	p-value = 0

	β_0	β_1	β_2
β_1	-0.57		
β_2	-0.83	0.03	
β_3	0.30	0.15	-0.45

Figure 8: Diagnostic plots for Model 1:(A) observed versus predicted composite ozone summary, (B) residuals versus predicted composite ozone summary, with nonparametric smooth (see Figure 11 for residuals versus time), (C) residual normality plot, with reference line, (D) residual autocorrelation plot for remaining short-term time structure (see Table 10).

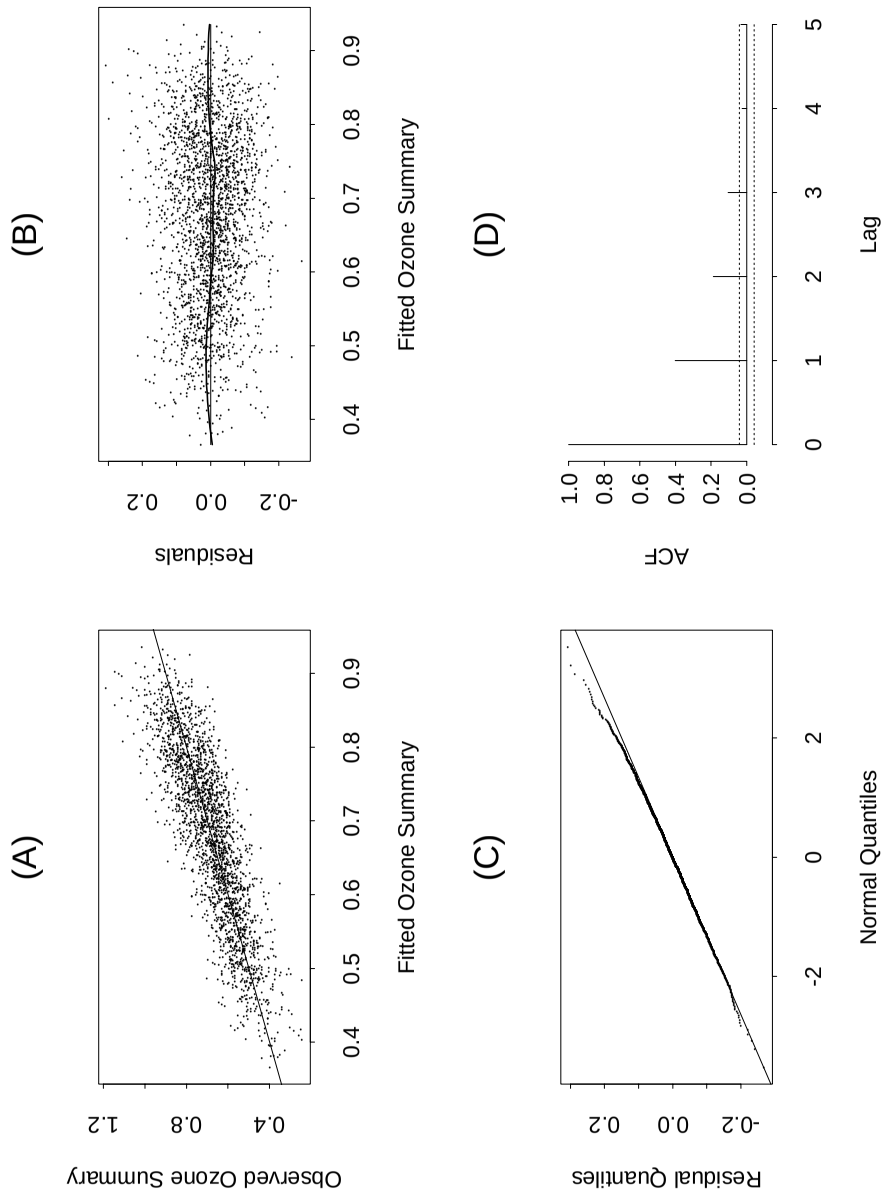


Figure 9: Diagnostic plots for Model 2:(A) observed versus predicted composite ozone summary, (B) residuals versus predicted composite ozone summary - plotted smooth exaggerates structure due to bias at endpoints (see Figure 11 for residuals versus time), (C) residual normality plot, (D) residual autocorrelation plot for remaining short-term time structure (see Table 10).

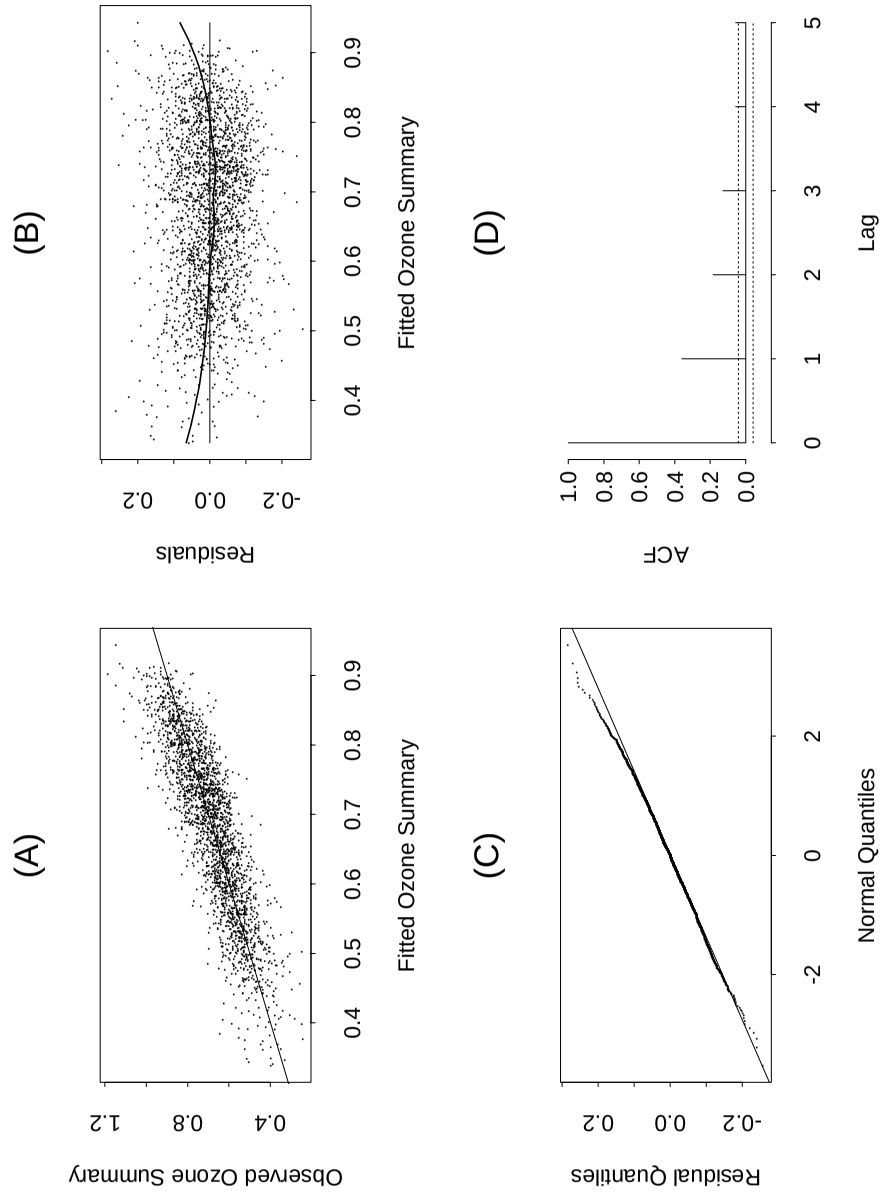


Figure 10: Observed (solid) and predicted (dashed) seasonal 95^{th} percentiles. See Section 3.4 for description of prediction calculation. Vertical axis is derived ozone network summary.

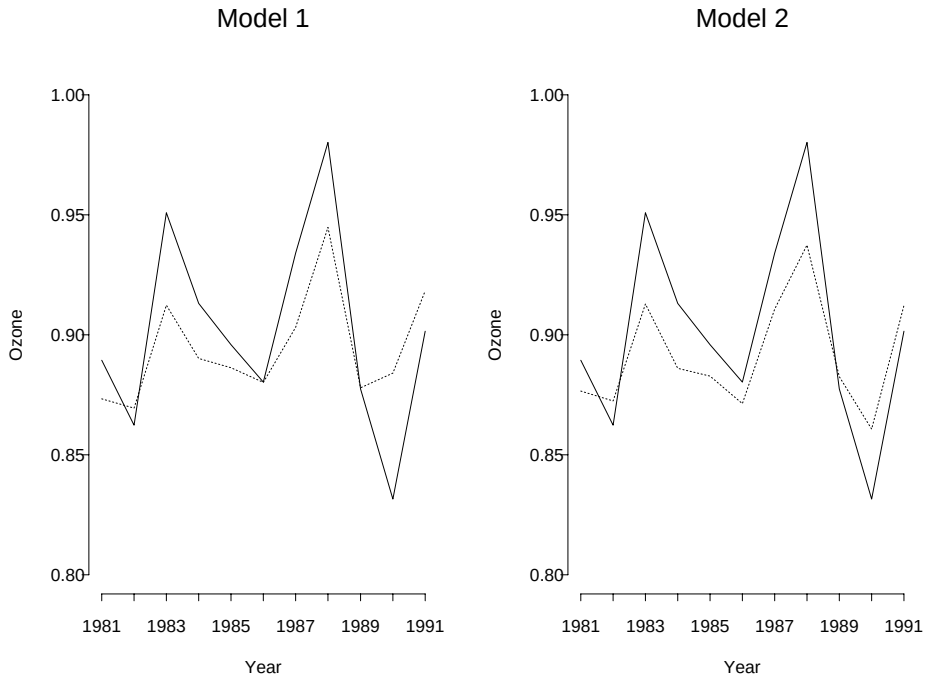


Figure 11: Residual time plot with loess smooth to reveal long-term trends.

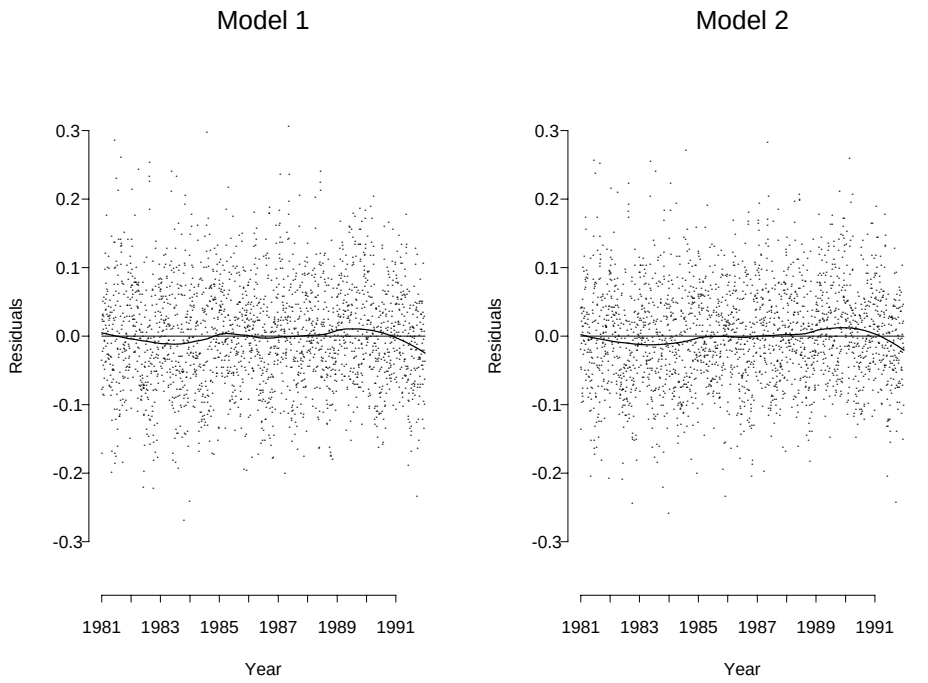


Table 9: Coefficient Estimates & Correlations, Model 2

Coef	Value	std.error
β_0	0.6699	0.0169
β_1	-0.0028	0.0001
β_2	-0.0071	0.0005
β_3	0.0179	0.0017
β_4	0.3138	0.0158
β_5	0.0008	0.00002
Res. Std. Error	0.0743 on 2346 df	
Multiple R^2	0.7181	
F-statistics	1195on (5,2346) df	p-value = 0

	β_0	β_1	β_2	β_3	β_4
β_1	-0.57				
β_2	-0.26	0.08			
β_3	-0.18	0.02	-0.08		
β_4	-0.80	0.04	-0.05	0.24	
β_5	0.27	0.14	0.25	-0.45	-0.49

estimates [Belsey 80] and biased coefficient tests. However, the goal of the analysis is to adjust comp.03 for meteorology. Therefore prediction, not parameter testing, is our focus and this is unaffected by multicollinearity. For this reason no statistical tests of the coefficients are given. Note that it is clear the models explain a substantial component of the variation in the ozone network composite.

Further issues preventing undue reliance on coefficient tests (standard errors) and omnibus tests (F-statistics) include

- The large sample sizes involved, which drive down the standard error estimates (making statistical significance more likely and less informative). Effect size, rather than statistical significance, should be the focus if the influence of specific meteorological variables is the goal.

- The presence of positive short-term correlation in the residuals (Table **??**) produces a conservative bias in the residual standard error estimates and hence in any significance tests.

The short-term autocorrelation has little influence on long-term trends. Therefore, no further steps were taken to adjust for the short-term autocorrelation remaining in the residuals.

4.8 Investigation of Long-term Trends

The nonparametric smooth of the residuals from either meteorological adjustment model suggests an insubstantial long-term component (Figure 11). Rather than assume an inappropriate functional form to investigate long-term trend, our initial analysis fit a nonparametric regression component for long-term trend as a function of year to both models[8]. The fitted components, distinctly non-monotonic (Figure 12) and therefore poorly captured by a linear trend term, make little contribution to either model (Table 12), even with the conservative tests resulting from the positive short-term autocorrelation.

[8]The smoothing spline option in S-Plus' general additive model function, gam, was used.

Table 10: Residual Autocorrelations for investigated models.

Model	**1 day lag**	**2 day**	**3 day**	**4 day**
3-term model	0.402	0.188	0.105	0.027
5-term model	0.360	0.183	0.129	0.057

Figure 12: Nonparametric regression component for long-term trend with 95% pointwise confidence bands (unadjusted for positive short-term residual autocorrelation). The vertical axis reveals the contribution of the long-term trend component to the overall model of composite ozone.

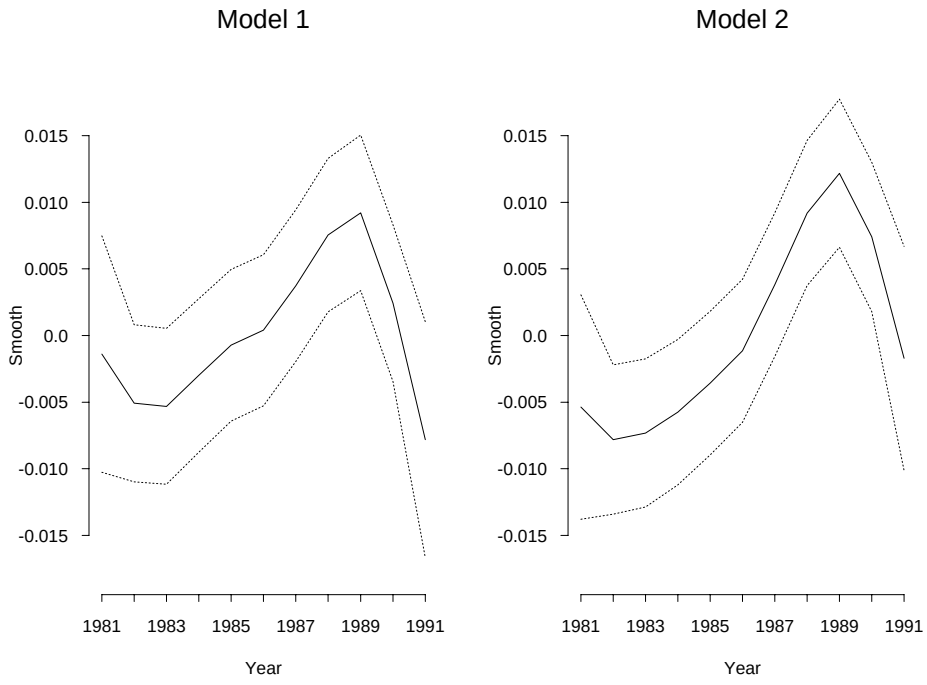


Table 11: Cross-Correlations among selected meteorological variables.

	RelHum	comp.Temp	Luminosity	WindSpd	WindDir	T2m2	WindV1	comp.Temp*RelHum
comp.Temp	-0.137							
Luminosity	-0.115	0.457						
WindSpd	-0.040	-0.233	-0.081					
WindDir	-0.091	0.406	-0.004	-0.033				
T2m2	0.158	0.846	0.504	-0.199	0.233			
WindV1	-0.224	0.087	-0.055	0.088	-0.081	-0.089		
comp.Temp*RelHum	-0.184	0.984	0.462	-0.239	0.399	0.813	0.106	
Luminosity*comp.Temp	-0.133	0.992	0.383	-0.239	0.412	0.817	0.106	0.971

Table 12: Analysis of Deviance of long-term trend contribution, Models 1 & 2 . Approximate p-values are from χ^2 with df = Change in df.

Model	**Residual df**	**Residual Deviance**
Model 1	2348.00	13.89
Model 1 + smooth(Year)	2344.00	13.78
Change	4.00	.11
Approximate p-value	.99	
Model 2	2346.00	12.96
Model 2 + smooth(Year)	2342.00	12.81
Change	4	.15
Approximate p-value	.99	

5 Discussion

5.1 Regional Signal & Meteorological Adjustment

The analysis reveals a strong regional-scale response in ozone, both in the uniformity across ozone monitors of the linearizing transformation of surface temperature (Figure 4) and in the uniformity across ozone monitors of the dominant coupling pattern's weight coefficients (Section 4.4). The unequal signal strength across surface temperature monitors is likely due to varying monitor quality (Figure 13).

There is a relatively simple coupling between the ozone and surface temperature networks: a single pattern of association linking comp.03, approximately proportional to the average square-root transformed maximum daily 1 hour ozone monitor observation (Table 6), with comp.Temp, approximately proportional to the average maximum daily surface temperature at O'Hare and Midway airports. Comp.O3, being approximately a regional average, is similar to the network average arrived at from a PCA of a larger Chicago ozone network [Bloom 93], an analysis making no direct use of meteorological fields .

A previous analysis of Chicago region ozone modelled the untransformed ozone network median by a fairly complex nonlinear parametric model, mainly multiplicative in structure (it has an additive seasonal component) [Bloom 96]. The surface meteorology variables available were similar to those utilized here, though direct upper air measurements were employed rather than Reanalysis output. Though not directly comparable due to differing ozone network summaries, Models 1 and 2 achieve roughly comparable performance and fit as the 21 parameter nonlinear parametric model (which gave an R^2=0.80).

5.2 Prediction

The models fit for meteorological adjustment do not lend themselves to use for prediction. Prediction requires specification of the short-term temporal structure in the residuals so that this can be accounted for and the prediction error reduced. More fundamentally, the goal of prediction would presumably be to forecast extreme ozone events. Models 1 and 2 were not developed to capture the association between *extreme* ozone events and meteorology, they were developed to capture mean associations between ozone and meteorology. The difference

in goal necessitates a difference in model structure and hence long-term trend conclusions (see Section 5.3). Note that changes to either the ozone or the surface temperature networks would require repeating the full analysis as the underlying cross-covariance matrix would change.

5.3 Long-Term Trend Investigation

The current analysis revealed a non-monotonic yet insignificant long-term trend; previous analyses for the region have returned differing conclusions, as summarized below. Unless stated otherwise, the ozone network analyzed is a superset of the monitors in Table 1 (see [Bloom 96]) and the analysis period consists of the April - October ozone seasons from 1981 - 1991 (as in the current study).

1. [Bloom 96] modelled network median daily 1 hour maximum ozone (untransformed) as a 21 parameter nonlinear parametric function of meteorology, with a multiplicatively incorporated linear trend term. I.e.,

$$Ozone = f(Meteorology) * (1 + \beta_{trend} * Year) + f(Seasonal\ Component).$$

   The resulting negative long-term trend coefficient (an approximate decline in ozone level of 2.7% / decade) was not statistically significant. The authors remarked that the fairly wide confidence interval on the trend coefficient implied that "even strong trends may not be detected as statistically significant results for some years."

2. [Smith 93] modelled exceedance over a threshold separately for three Chicago area monitors (two of which are in the current analysis, monitors Oz3 and Oz6 (Table 1)) and for the maximum of the network analyzed in [Bloom 96]. The logit of the probability of exceedance was modelled as a linear function of meteorological covariates and, in some cases, an indicator of the previous day's exceedance status. Using a threshold of 120 ppb, all three stations displayed a negative linear trend in exceedances, but only two were statistically significant (including monitor Oz3); using a threshold of 100 ppb, all three stations displayed a negative linear trend, but only one was statistically significant (monitor Oz3). Further analysis of this station revealed a significant negative quadratic trend in exceedances (at 120 ppb), a result that also held for the network maximum. In both cases, a slight increase in the frequency of exceedances of daily 1 hour maximum ozone occurred up to 1984 or 1985, followed by a decrease.

3. [Milan 98] investigated long-term trend in the natural logarithm of daily 1 hour maximum of a Chicago area monitor (Oz1 in Table 1) from 1984 - 1995 by decomposing the ozone and meteorological predictor time series into short-term and seasonal & long-term components using the Kolmogorov-Zurbenko filtering technique (an iterative moving average). The different time scale components of ozone were then regressed against the associated time scale components of selected meteorological variables (short-term: temperature, dew-point depression; seasonal & long-term: solar radiation, specific humidity). The residuals from the different time scale meteorological adjustments were combined to reconstruct a meteorologically adjusted ozone time series. Fitting a linear

time trend to this series resulted in a significant positive trend. This result remains if the analysis is applied to the 1981 - 1991 period used in the current study (David Caccia, pers. comm.). While not strictly comparable due to differences in response variables, it is of interest to note that their long-term trend does not show the pattern of decline in the period 1984 - 1991 found by [Smith 93] at a different monitor in the region.

4. [Niu 96] modelled daily maximum ozone from 1983 - 1992 as a nonlinear additive function of meteorology with an associated auotoregressive moving average time series model for the error component. The variance of the error component model was further modelled as a log-linear function of meteorology. A linear long-term trend was fit, resulting in a significant negative trend.

5. [Huang 99] fit a CART model to the network maximum daily 1 hour maximum ozone for the 16 Chicago area monitors, generating 'clusters' of meteorological and ozone comparable days. Within each meteorological/ozone cluster, they fit a separate linear model of ozone response as a function of meteorology, including a linear time trend in each model. Thus, they were able to demonstrate an interaction between time trend in network maximum daily ozone and meteorological condition. They found that while the clusters showed predominately negative temporal trends, the clusters most conducive to extreme ozone events showed, for the most part, the most negative trends. In summary, they showed that there appears to be a significant decrease in the occurrence of extreme ozone events, but no significant decrease in moderate magnitude ozone events (though this trend is negative).

To summarize, three analyses report negative trends ([Smith 93], [Niu 96],, [Huang 99]), one reports a positive trend [Milan 98], and two report no significant trends ([Bloom 96], the current analysis).

The results of [Smith 93] can be reconciled with the non-significant trends by noting the distinction between trends in *extreme events* versus in the *average ozone baseline*. The former is looking at patterns of increasing or decreasing frequency of extreme threshold exceedances by the network maximum, which is distinct from the patterns of changes in the average network ozone level. While, all things remaining the same, shifts in average level are sufficient to produce shifts in exceedance frequency, they are not necessary for such shifts to occur. The lack of significant trends when a lower threshold was used to define exceedance can be interpreted as an agreement with the lack of significant trends in the average ozone level. The distinction between network maximum and network average may also explain the difference in results of [Niu 96] and ([Bloom 96], the current analysis). The recent results of [Huang 99] support this conclusion.

The method of [Milan 98] was applied to the ozone network summary derived in Section 4.5 (David Caccia, pers. comm.). The resulting non-significant long-term trend estimate was comparable to that seen in Figure 12. Thus, the contrast in conclusions reached by [Milan 98] versus ([Bloom 96], the current analysis) appears to arise from a contrast in analyzing the behavior of a single monitor rather than a composite network response.

The non-monotonicity of the observed long-term trend (Figure 12) points out the dangers of automatically investigating long-term trend using a linear representation. Though the the

component was not found to significantly contribute to model performance, it is not difficult to imagine a situation where the presence of a significant long-term dependence component could be masked by implicitly assuming a linear form for the investigation.

6 Conclusion

Canonical Covariance Analysis of the Chicago region ozone and transformed surface temperature networks revealed a simple association between the two fields, with the ozone field summary being approximately proportional to the network average of square-root transformed maximum daily 1 hour ozone observations. A distinct drop among the surface temperature monitors in strength of association with ozone suggested substantial differences in data quality among the surface temperature network. Linear regression models for meteorological adjustment of the regional ozone summary were selected using best subsets regression; the resulting models were comparable in performance to those of previous researchers [Bloom 96]. Rather than assume a linear long-term trend structure, nonparametric regression was used to investigate long-term trend in the meteorologically adjusted regional ozone summary; no significant trend was found. The results were compared to those of other analyses of the region. The comparison clarified the need for standards which specifically define the statistic used for ozone trend investigation, the increased senstivity of extreme statistics to means in terms of detecting trends, the need for consideration of the differences between single-site and regional network analyses, and the dangers of implicity assuming a linear trend structure. The composite variables, raw data, and S-plus functions used in the analysis are available from the authors.

References

[Belsey 80] Belsley, D. A., E. Kuh, R. E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley and Sons, New York.

[Bloom 96] Bloomfield, P., J. A. Royle, L. J. Steinberg, Q. Yang. 1996. Accounting for meteorological effects in measuring urban ozone levels and trends. Atmospheric Environment 30(17): 3067 - 3077.

[Bloom 93] Bloomfield, P., J. A. Royle, Q. Yang. 1993. Accounting for meteorological effects in measuring urban ozone levels and trends. Technical Report 1, National Institute of Statistical Science.

[Breth 92] Bretherton, C. S., C. Smith, J. M. Wallace. 1992. An intercomparison of methods for finding coupled patterns in climate data. Journal of Climate, 5: 541 - 560.

[Davison 97] Davison, A. C., D. V. Hinkley. 1997. *Bootstrap Methods and their Application.* Cabridge University Press, England.

[EPA 95] Environmental Protection Agency. 1995. Review of the National Ambient Air Quality Standards for Ozone: Assessment of Scientific and Technical Information. OAPQS Staff Paper, EPA - 452/R.

[Hastie 90] Hastie, T. J., R. J. Tibshirani. 1990. *Generalized Additive Models*. Chapman and Hall, Boca Raton, Florida.

[Huang 99] Huang, L.-S., R. L. Smith. 1999. Meteorologically-dependent trends in urban ozone. Environmetrics 10, 103 - 118.

[Hydro 95] Hydrosphere Environmental Databases: NCDC Summary of the Day. 1995. Hydrosphere Data Products, Inc. 1002 Walnut, Suite 200, Boulder, CO. 80302.

[Kalnay 96] Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, R. Jenne, D. Joseph. 1996. The NCEP/NCAR 40-year Reanalysis Project. Bull. American Meteorological Society, 77(3): 437-471.

[Little 86] Little, R. J. A., D. B. Rubin. 1986. *Statistical Analysis with Missing Data*. Wiley and Sons, New York.

[MathSoft 96] Decision Analysis Products Division, MathSoft, Inc. 1996. S-Plus version 3.4.

[Milan 98] Milanchus, M. L., S. T. Rao, I. G. Zurbenko. 1998. Evaluating the Effectiveness of Ozone Managaement Efforts in the presence of Meteorological variability. J. Air & Waste Management Assoc. 48: 201 - 215.

[Niu 96] Niu, X.-F. 1996. Nonlinear additive models for environmental time series, with applications to ground-level ozone data analysis. J. of the American Statistical Assoc., 91: 1310 - XXX.

[NRCSE 98] National Research Center for Statistics and the Environment. Review of statistical methods for meteorologically adjusting surface ozone observations. Draft technical report.

[NRC 91] National Research Council. 1991. Rethinking the ozone problem in urban and regional air pollution. National Academy Press, Washington, D. C.

[Reynolds 98] Reynolds, J. H., B. Das, P. Sampson, P. Guttorp. 1998. Meteorological Adjustment of Western Washington and Northwest Oregon Surface Ozone observations with investigation of trends. NRCSE Technical Report NRCSE-TRS No. 15.

[Ryan 97] Ryan, T. P. 1997. *Modern Regression Methods*. Wiley and Sons, New York.

[Smith 93] Smith, R. L., L.-S. Huang. 1993. Modeling High Threshold Exceedances of Urban Ozone. Technical Report No. 6, National Institute of Statistical Sciences.

[Wall 92] Wallace, J. M., C. Smith, C. Bretherton. 1992. Singular value decomposotion of wintertime sea surface temperature and 500-mb height anomalies. Journal of Climate, 5:561-576.

[Wax 91] Waxman, H., G. Wetstone, P. Barnett. 1991. Roadmap to Title I of the Clean Air Act Amendments of 1990: bringing blue skies back to America's cities. Environmental Law, 21(4): 1843 - 1946.

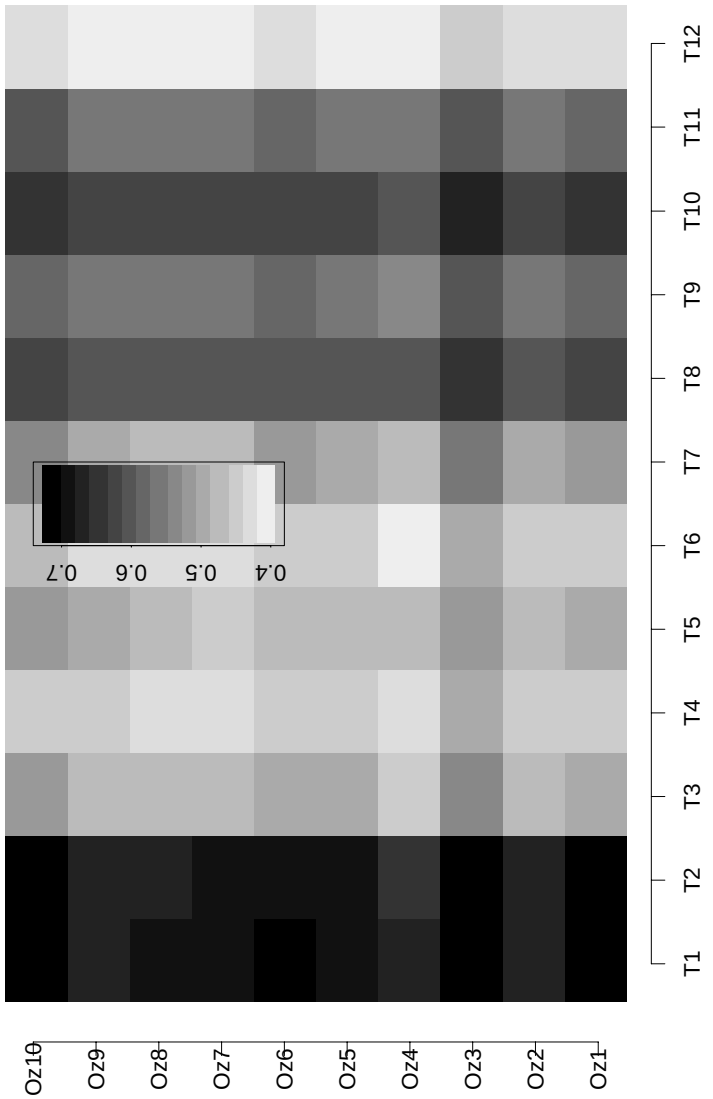
A Exploratory Analysis of Regional Ozone Signal

Pairwise scatterplots of the ozone monitor records display common regional ozone patterns (Figure 1). Similarly, cross-correlations between each ozone monitor and each transformed surface temperature monitor display common associations across the region (Figure 13 ,Table 13).

Table 13: Square-Root Ozone and transformed Surface Temperature network cross-correlations

	T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**	**T11**	**T12**
Oz1	0.70	0.71	0.50	0.45	0.49	0.46	0.52	0.62	0.58	0.64	0.59	0.43
Oz2	0.66	0.66	0.48	0.46	0.48	0.45	0.49	0.60	0.56	0.62	0.57	0.43
Oz3	0.72	0.71	0.53	0.49	0.52	0.49	0.55	0.64	0.61	0.66	0.61	0.46
Oz4	0.66	0.66	0.46	0.43	0.47	0.42	0.47	0.60	0.55	0.60	0.55	0.40
Oz5	0.70	0.69	0.49	0.44	0.47	0.45	0.49	0.61	0.57	0.63	0.57	0.42
Oz6	0.70	0.70	0.50	0.45	0.48	0.45	0.52	0.61	0.57	0.63	0.58	0.42
Oz7	0.69	0.68	0.47	0.43	0.46	0.42	0.48	0.60	0.56	0.62	0.56	0.40
Oz8	0.69	0.68	0.47	0.44	0.48	0.43	0.48	0.61	0.56	0.62	0.57	0.41
Oz9	0.66	0.67	0.47	0.45	0.49	0.43	0.49	0.61	0.55	0.62	0.56	0.41
Oz10	0.71	0.71	0.51	0.46	0.51	0.47	0.53	0.62	0.59	0.65	0.60	0.43

Figure 13: Absolute cross-correlations between daily maximum 1 hour ozone and transformed maximum daily surface temperature from each monitor pair. Note the common pattern of cross-correlation demonstrated by the dominant horizontal striations.



B Imputation of Missing Ozone Records

See [Little 86] for a thorough discussion of imputation using the EM algorithm. Here we assume that the ozone monitor records for the network follow a multivariate normal distribution,

$$\begin{pmatrix} Oz1 \\ Oz2 \\ . \\ . \\ . \\ Oz10 \end{pmatrix} N(\mu_{dayofseason}, \sum_{Ozone}).$$

The procedure follows five steps:

1. Estimate the seasonal mean trend at each ozone site,

$$\hat{\mu}_{Oz1,April1} = \sum_{years} Oz1_{April1,yeari}/n$$

   .

2. Remove the seasonal mean trend at each site, centering the network distribution around the zero-vector.

3. Use the EM algorithm to calculate $\hat{\sum}_{Ozone}$, (Table 14).

4. For each monitor with missing observations, and each combination of monitors with observations available on those days that the monitor of interest was unobserved, calculate the linear regression equation

$$Oz_{missing,day\ i} = \beta_{0,missing} + \sum_{observed\ j} \beta_{j,missing} Oz_{observed\ j,dayi}$$

   from $\hat{\sum}_{ozone}$.

5. Use equations from the previous step to impute the unobserved records.

The estimated covariance matrix for the locally-standardized ozone monitors is given below.

Table 14: Square-Root Ozone network covariance matrix, estimated by EM algorithm.

	Oz1	**Oz2**	**Oz3**	**Oz4**	**Oz5**	**Oz6**	**Oz7**	**Oz8**	**Oz9**	**Oz10**
Oz1	0.00275	0.00219	0.00205	0.00182	0.00204	0.00214	0.00189	0.00172	0.00176	0.00217
Oz2	0.00219	0.00258	0.00195	0.00179	0.00193	0.00194	0.00181	0.00165	0.00178	0.00209
Oz3	0.00205	0.00195	0.00251	0.00184	0.00216	0.00233	0.00200	0.00173	0.00170	0.00208
Oz4	0.00182	0.00179	0.00184	0.00211	0.00188	0.00190	0.00183	0.00179	0.00178	0.00184
Oz5	0.00204	0.00193	0.00216	0.00188	0.00236	0.00229	0.00211	0.00178	0.00167	0.00201
Oz6	0.00214	0.00194	0.00233	0.00190	0.00229	0.00266	0.00216	0.00180	0.00171	0.00209
Oz7	0.00189	0.00181	0.00200	0.00183	0.00211	0.00216	0.00217	0.00177	0.00163	0.00191
Oz8	0.00172	0.00165	0.00173	0.00179	0.00178	0.00180	0.00177	0.00181	0.00159	0.00175
Oz9	0.00176	0.00178	0.00170	0.00178	0.00167	0.00171	0.00163	0.00159	0.00206	0.00182
Oz10	0.00217	0.00209	0.00208	0.00184	0.00201	0.00209	0.00191	0.00175	0.00182	0.00271