NISS

Evaluation of Data from Three Meteorologic Stations for the Parametric Modeling of Rural Ozone

Laura J. Steinberg and Peter Bloomfield

Technical Report Number 17
June, 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.

Evaluation of Data from Three Meteorologic Stations for the Parametric Modeling of Rural Ozone*

Laura J. Steinberg and Peter Bloomfield

National Institute of Statistical Sciences

and

Department of Statistics
North Carolina State University

June, 1994

Abstract

Ozone concentrations recorded by a network of rural monitoring stations surrounding Chicago have previously been modeled as a non-linear function of meteorology recorded in Chicago (NISS Technical Report #5). Because some of the stations are actually closer to Detroit or St. Louis, it may be appropriate to include meteorology from these cities in the model. Investigation showed, however, that model performance is not improved by these additional data. Another possibility is that the stations are not similarly affected by Chicago meteorology, and that the non-linear model may best be fit using the meteorology of Chicago, St. Louis, or Detroit for three geographically-determined subnetworks of the stations. When this was

*Research supported by the U.S. Environmental Protection Agency under Cooperative Agreement #CR819638-01-0.

done, the model was shown to fit reasonably well for each of the subnetworks. The resulting meteorologically-adjusted trend estimates for Chicago, St. Louis, and Detroit are -2.0%/decade, -12.5%/decade, and +0.5%/decade, respectively. The differences in the trend parameters are only significant for the St.Louis/Detroit comparison. This implies that the meteorologically-adjusted trends in ozone concentration vary between these two subnetworks but that no conclusion can be drawn from this analysis regarding the differences in trend between Chicago and St. Louis or between Chicago and Detroit.

1 Introduction

In NISS Technical Report #5, Bloomfield, Royle and Yang (1993) modeled network typical ozone for the rural area around Chicago using data from twelve ozone-monitoring stations located in Illinois, Wisconsin, Indiana, and Michigan. Bloomfield *et al.* developed a non-linear model which predicts network typical ozone concentration as a function of meteorology, year, and season. The meteorological data included surface data from O'Hare International Airport in Chicago and upper air data from Peoria, Il. Both the ozone-monitoring stations and the meteorological stations are shown in Figure 1 of NISS Technical Report #5.

The scale of the rural network suggests that meteorological conditions throughout the network may differ from monitoring station to monitoring station. The original model assumed that the Chicago meteorological data would adequately summarize the meteorological conditions in the network; the models explored in this supplement incorporate data from two additional monitoring stations (St. Louis and Detroit) located on the boundaries of the rural network. The purpose of this investigation is twofold:

1. to determine if model performance can be significantly improved by including St. Louis and Detroit meteorological data in the model, and

2. to determine if meteorologically-adjusted trends in ozone concentration vary between regions defined by the three major urban centers in the rural network.

2 Preliminary Analysis

The first step in the analysis of the additional meteorological data is to verify that the functional form of the non-linear model specified in Equation 6 of NISS Technical Report #5 is appropriate when meteorological data from St. Louis or Detroit were used in place of O'Hare meteorological data. Figures 10, 11, 12, and 13 from NISS Technical Report #5 were duplicated using St. Louis and Detroit meteorological data instead of Chicago meteorological data. The resulting figures are shown in Figure 1 through Figure 8. Comparing these with the figures in NISS Technical Report #5, it is clear that the form of the relationships between network typical ozone and the meteorological variables are similar, regardless of which meteorological station's data is used.

Parameter values fit using Equation 6 might be expected to differ, however, depending on which meteorological station's data is used in Equation 6. Parameter estimates for γ, the trend parameter, and R^2 values are given in Table 1. These estimates are derived by fitting Equation 6 with meteorological data from Chicago, St. Louis, or Detroit. R^2 values are highest for Chicago data. A possible explanation for this is the closer proximity of most stations to Chicago than to St. Louis or Detroit. The trend parameter is negative when Chicago and St. Louis data is used, but positive when Detroit data is used. In all cases, the change per decade is on the order of only ± 1–4%, and is barely significant statistically.

3 Optimizing the Use of Meteorological Data

In Section 2, Equation 6 was fit using either data from Chicago, St. Louis, or Detroit. In this section, the meteorological data is drawn from any of these three

Met Station	R^2	y (% /decade)	tvalue for y
Chicago	0.797	-1.03	-0.73
St. Louis	0.688	-2.74	-1.71
Detroit	0.652	3.76	2.00

Table 1: Model results for Equation 6 using Chicago, St. Louis, and Detroit meteorological data.

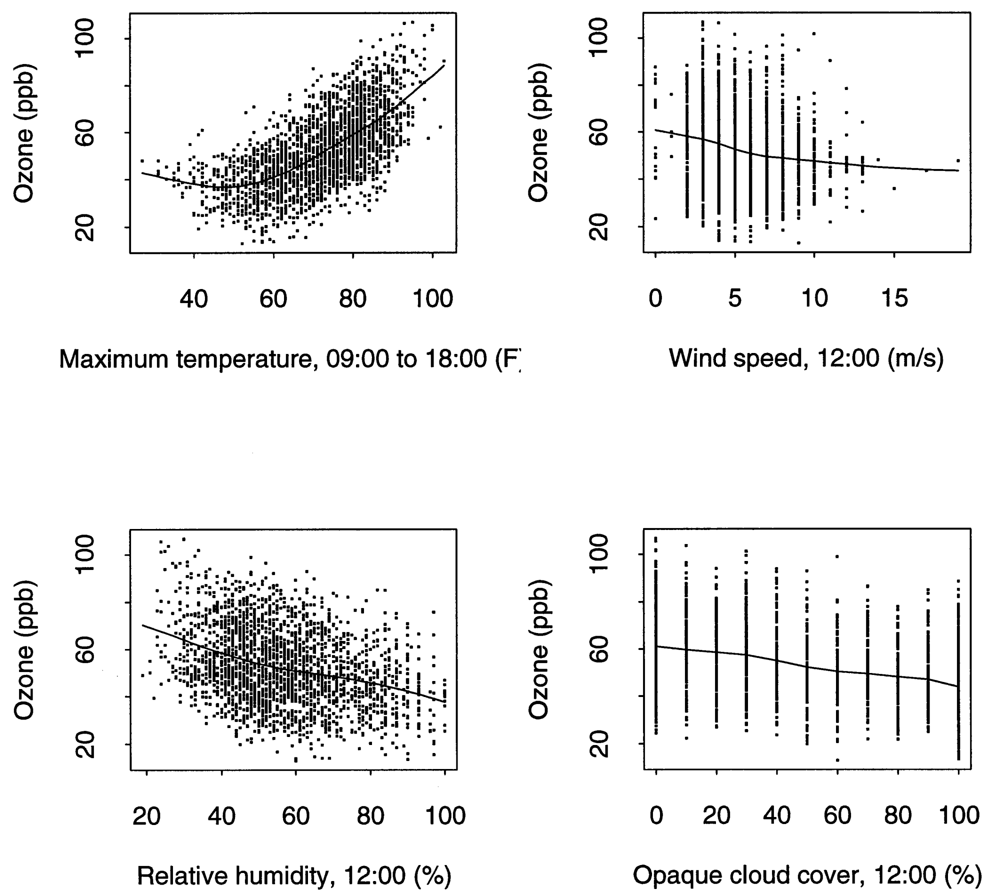

Figure 1: Scatter plots of ozone against temperature, wind speed, relative humidity, and cloud cover using Detroit meteorological data.

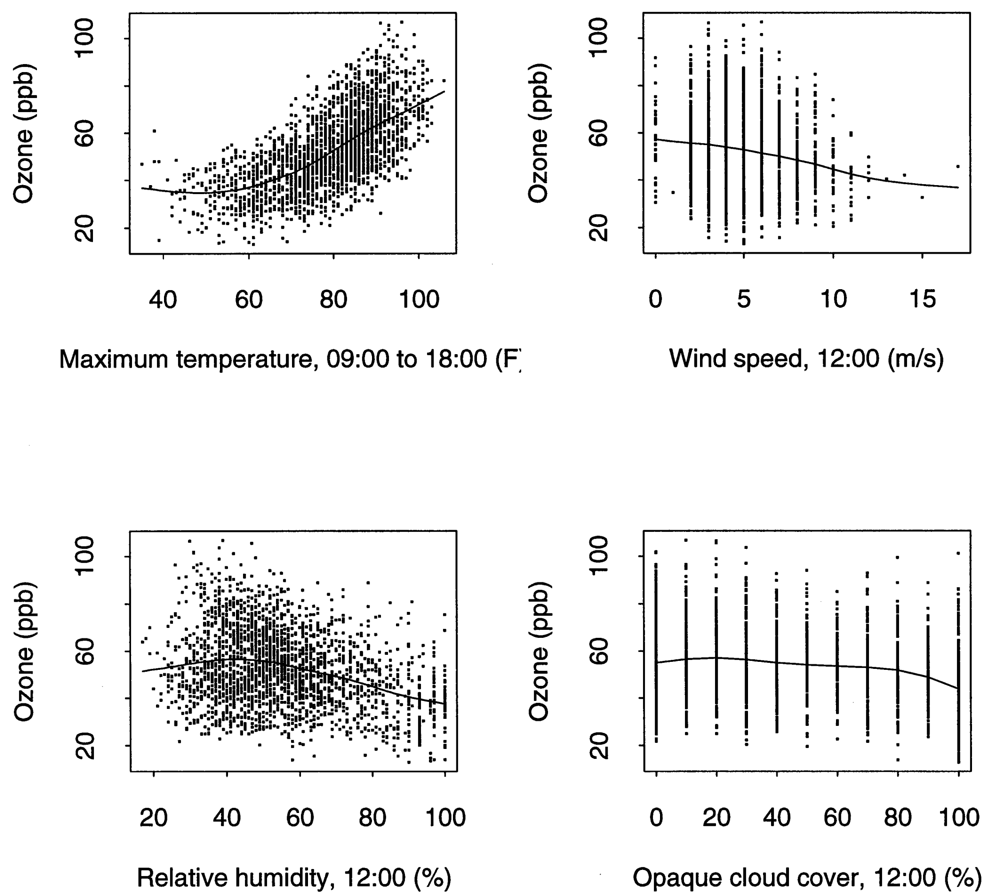

Figure 2: Scatter plots of ozone against temperature, wind speed, relative humidity, and cloud cover using St. Louis meteorological data.

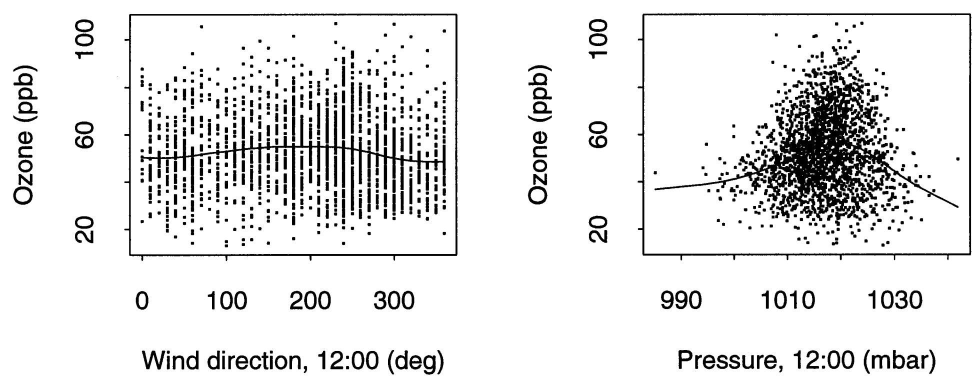

Figure 3: Scatter plots of ozone against wind direction and barometric pressure using Detroit meteorological data.

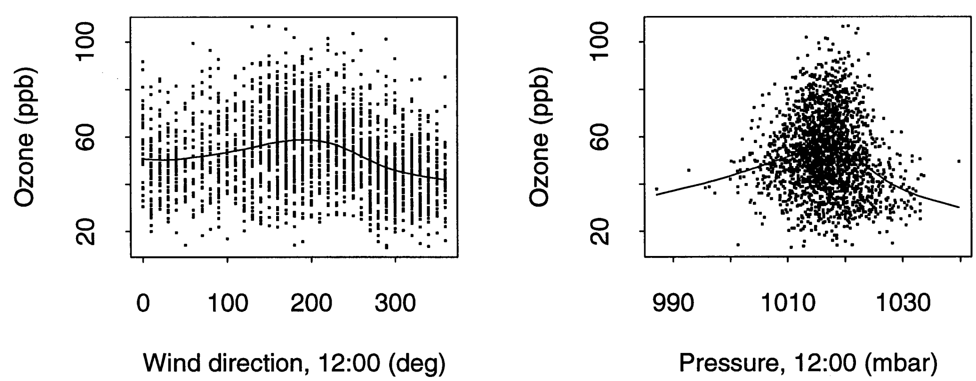

Figure 4: Scatter plots of ozone against wind direction and barometric pressure using St. Louis meteorological data.

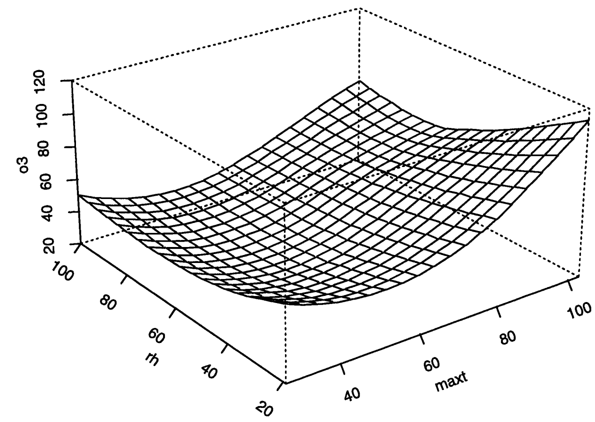

Figure 5: Nonparametric regression surface for ozone against temperature (maximum from 09:00 to 18:00) and noon relative humidity using Detroit meteorological data.

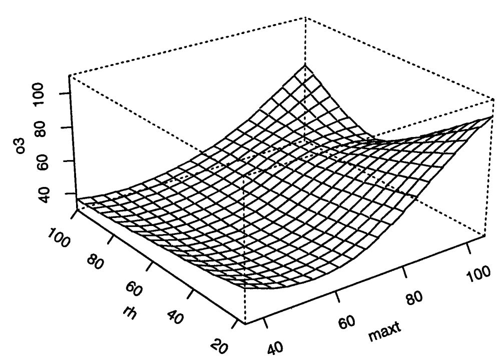

Figure 6: Nonparametric regression surface for ozone against temperature (maximum from 09:00 to 18:00) and noon relative humidity using St. Louis meteorological data.

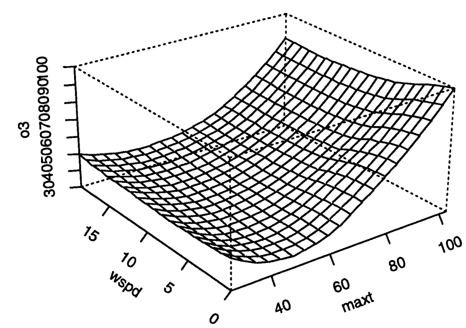

Figure 7: Nonparametric regression surface for ozone against temperature (maximum from 09:00 to 18:00) and noon wind speed using Detroit meteorological data.

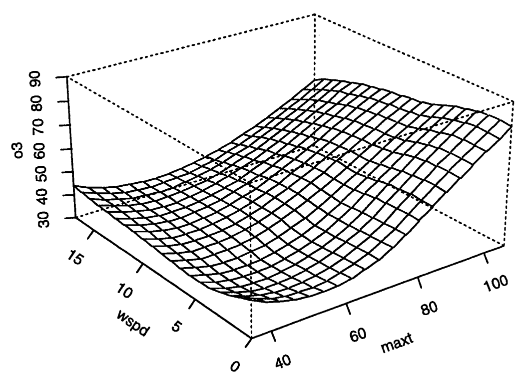

Figure 8: Nonparametric regression surface for ozone against temperature (maximum from 09:00 to 18:00) and noon wind speed using St. Louis meteorological data.

meteorological stations. For each variable, the meteorological variable from the station which provides the greatest increase in R^2 is chosen. In addition, the effect on R^2 of having the same meteorological variable in the model two or three times (once for each station), was tested. Equation 1 shows the resulting equation yielding the highest possible R^2, given the functional form of the model and data from three meteorological stations.

$$
\begin{aligned}
\texttt{o3} \sim\ & (\texttt{mu0} + (\texttt{t0} + \texttt{t1} * (\texttt{maxt} - 60) \\
&\qquad + \texttt{t2} * (\texttt{maxt} - 60)^2 + \texttt{t3} * (\texttt{maxt} - 60)^3 \\
&\qquad + \texttt{stl1} * (\texttt{stlag1} - 60) + \texttt{stl2} * (\texttt{stlag2}) - 60) \\
&\quad * 1/(1 + \texttt{wspd}/\texttt{vh} + \texttt{wspd700}/\texttt{vh700} + \texttt{wlag}/\texttt{vhl})) \\
& * (1 + \texttt{sr} * (\texttt{srh} - 50) + \texttt{dr} * (\texttt{drh} - 50) + \texttt{rl} * (\texttt{rhlag} - 50)) \\
& * (1 + \texttt{sop} * (\texttt{sopcov} - 50)) \\
& * (1 + \texttt{sv} * (\texttt{svis} - 12)) \\
& * (1 + \texttt{m.u} * \texttt{mean.u} + \texttt{m.v} * \texttt{mean.v}) \\
& * (1 + \texttt{y} * (\texttt{year} - 1985)) \\
& + \texttt{a1} * \texttt{cos}(2 * \texttt{pi} * \texttt{year}) + \texttt{b1} * \texttt{sin}(2 * \texttt{pi} * \texttt{year}) \\
& \quad + \texttt{a2} * \texttt{cos}(4 * \texttt{pi} * \texttt{year}) + \texttt{b2} * \texttt{sin}(4 * \texttt{pi} * \texttt{year}). \qquad (1)
\end{aligned}
$$

The variables and parameters in this equation are defined in NISS Technical Report #5 except that some variables in Equation 1 represent meteorological data from St. Louis or Detroit, not from Chicago. These variables begin with an `s` or `d`, respectively. This model yielded an R^2 of 0.812. The trend parameter, `y`, was estimated as 0.8% decade (0.00080), with a standard error of 0.0013 and a tvalue of 0.62.

This model represents only a slight improvement in R^2 over that seen in the simpler model using only Chicago meteorology i.e. R^2 increases from 0.80 to 0.81. A slightly better R^2 is expected when additional variables are added; in this case little seems gained by using the additional data. In both cases, the trend parameter is statistically insignificant. Furthermore, its estimate is highly variable depending on the meteorological data used, possibly resulting from multicollinearity between the predictor variables.

4 Analysis of Ozone Concentrations by Subnetwork

4.1 Subnetwork Parameter Estimation

In order to determine whether there are regional differences in the trend parameter, the network was broken into three subnetworks. Each meteorological station was assigned the ozone monitoring stations closest to it. Referring to Figure 1 of NISS Technical Report #5, Detroit was assigned stations 260812001, 260370001, and 261611001. St. Louis was assigned stations 171192007, 181630013, and 170190004. The remaining six stations were assigned to the Chicago subnetwork.

Median polish was applied to the daily maximum concentrations at the stations comprising each subnetwork. The daily maxima were decomposed as described in Section 4.2 of the NISS Technical Report #5 in order to generate subnetwork averages. Then, Equation 6 was fit to each of the three subnetwork averages, using the meteorological data from the assigned meteorological station. The resulting parameter estimates are shown in Tables 2-4. Trend parameter estimates of -1.8%/decade, -12.5%/decade[1], +0.5%/decade for Chicago, St. Louis, and Detroit, respectively, were found.

Although the same parameters Tables 2-4 appear to be significant for each subnetwork, the actual parameter estimates vary considerably. Since the trend parameter is of particular interest, the sampling distributions of the three trend parameter estimates are investigated further to determine if the trend parameters are significantly different from each other.

4.2 Jackknifed Parameter Estimation

The standard errors of the trend parameters were estimated more carefully by jack-knifing the parameter estimates. As described in NISS Technical Report

[1]Two other stations, 171431001 and 180970042, might have been assigned to the St. Louis network, in addition to the three chosen. These two stations and station 171192007 (already included in the St. Louis network) are nearly equidistant from Chicago and St. Louis. The sensitivity of the St. Louis regression results to the selection of stations comprising the St. Louis network was evaluated by running the non-linear regression for various combinations of the five possible stations. When all five were included in the network, the trend was estimated at -11.2%/decade, with a standard error of 1.3%/decade and an R^2 of 0.76. For all combinations evaluated, the trend coefficient was between -11.2%/decade and -15.2%/decade. Standard errors varied from 1.3%/decade to 1.5%/decade. Thus, the trend parameter estimates for St. Louis are not appreciatively effected by changes in the set of stations included in the St. Louis network.

	Value	Std. Error	t value
mu0	4.822584e+01	2.4181940640	19.9429175
t0	-1.266424e+00	7.5104633418	-0.1686213
st1	1.885272e+00	0.4101875614	4.5961225
st2	5.798177e-02	0.0127528133	4.5465867
st3	-6.585176e-04	0.0002327260	-2.8295832
stl1	-2.874123e-02	0.1821958825	-0.1577490
stl2	-7.085818e-01	0.2212026912	-3.2033146
svh	1.261629e+01	4.6212672133	2.7300506
svh700	1.603694e+01	5.1925617088	3.0884452
svhl	4.135691e+00	1.3877691997	2.9800998
sr	-1.855135e-03	0.0001788347	-10.3734607
srl	-3.486660e-03	0.0003234547	-10.7794367
sop	-3.553612e-04	0.0001464648	-2.4262563
sv	-7.186089e-03	0.0006628527	-10.8411557
sm.u	-5.563326e-03	0.0014046904	-3.9605351
sm.v	1.887897e-03	0.0017318879	1.0900805
sy	-1.254575e-02	0.0013367562	-9.3852164
sa1	-1.164921e+01	1.8187800565	-6.4049595
sb1	2.219269e+00	0.5598250883	3.9642192
sa2	-2.185041e+00	0.8943415266	-2.4431837
sb2	-1.712316e+00	0.6002454929	-2.8526924
$R^2 = 0.76$			

Table 2: Parameter values for the St. Louis subnetwork.

	Value	Std. Error	t value
dmu0	3.212896e+01	5.0627135387	6.3461945
dt0	1.493950e+01	4.8809001221	3.0608084
dt1	9.159206e-01	0.1669437904	5.4864014
dt2	2.815600e-02	0.0051269356	5.4917803
dt3	-3.229412e-04	0.0001318817	-2.4487185
dtl1	-1.706516e-01	0.0989114618	-1.7252967
dtl2	-1.241131e-01	0.0854504339	-1.4524573
dvh	1.751294e+01	5.4210985096	3.2305147
dvh700	2.600297e+01	7.7815699835	3.3416093
dr	-2.559032e-03	0.0005340779	-4.7914965
drl	-2.615183e-03	0.0004987212	-5.2437780
dop	-1.353720e-03	0.0001933426	-7.0016656
dv	-4.821178e-03	0.0006851433	-7.0367440
dm.u	1.885847e-03	0.0020448969	0.9222213
dm.v	1.747116e-02	0.0021701022	8.0508486
dy	4.704085e-04	0.0005510932	0.8535916
da1	-1.069233e+01	2.3158709232	-4.6169788
db1	6.314236e+00	1.8834872434	3.3524179
da2	-3.883034e+00	1.1170061823	-3.4762870
db2	-1.980749e+00	0.7646862200	-2.5902762
$R^2 = 0.64$			

Table 3: Parameter values for the Detroit subnetwork.

	Value	Std. Error	t value
mu0	4.149370e+01	1.6891804747	24.5643950
t0	8.108885e+00	5.0439693024	1.6076396
t1	2.203887e+00	0.4027063795	5.4726894
t2	5.976653e-02	0.0114882676	5.2023975
t3	-1.129425e-03	0.0002754693	-4.1000015
tl1	-6.734444e-02	0.1537696480	-0.4379566
tl2	-3.905979e-01	0.1455719520	-2.6831945
vh	1.224696e+01	4.1522552342	2.9494721
vh700	1.195894e+01	2.9050624839	4.1165875
vhl	4.448293e+00	1.2924575702	3.4417321
r	-7.688939e-04	0.0001893492	-4.0607182
rl	-2.676460e-03	0.0003626424	-7.3804383
op	-1.255534e-03	0.0001394104	-9.0060231
v	-4.573817e-03	0.0007498218	-6.0998711
m.u	1.168707e-02	0.0015277943	7.6496348
m.v	8.623802e-03	0.0015938083	5.4108152
y	-1.960144e-03	0.0015685593	-1.2496463
a1	-8.366353e+00	1.4512403552	-5.7649670
b1	4.670013e+00	0.4987926216	9.3626348
a2	-2.951809e+00	0.7731566556	-3.8178662
b2	-1.458809e+00	0.5772103672	-2.5273430
$R^2 = 0.77$			

Table 4: Parameter values for the Chicago subnetwork.

#5, jack-knifed estimates for the trend parameter are created by averaging yearly pseudo-value estimates, y_i, where:

y_i=Y(estimate from all data) - (Y-1)(leave one out estimate)

where Y is equal to the number of years (11) and $i = 1 \ldots Y$. The "estimate from all the data" is the trend parameter in Tables 2-4, and the "leave out one estimate" is obtained from fitting the non-linear model without year i. Yearly pseudo-values for each subnetwork are shown in Table 5. These may be thought of as eleven observations on the trend parameter, and their mean as an estimator of the trend parameter. The standard error of the mean provides an estimate of the standard error of either the original parameter estimate or the jack-knifed estimate.

The standard errors obtained through jack-knifing for each of the three subnetworks are shown in Table 6. Comparing these with the standard errors in Tables 2-4, it is seen that they show little change except for the Chicago subnetwork, where the standard error increases 360%.

The significance of the differences between these three trend parameters was tested assuming that the pair-wise differences between the yearly pseudo-values were normally distributed with unknown means. Tests of the null hypotheses that the means are 0 ppb resulted in rejection for the St. Louis/Detroit comparison, but not for the Chicago/St. Louis and Chicago/Detroit comparisons. Table 7 shows the p-value associated with each test. The p-value for the St. Louis/Detroit comparison is small enough so that even under a simultaneous test of the three hypotheses, the null hypothesis for this comparison would still be rejected. Bonferroni simultaneous ninety-five percent confidence intervals for the differences (in %/decade) are also shown in Table 7. The results in Table 7 indicate that the meteorologically-adjusted trends in ozone concentration differ between St. Louis and Detroit, but there is no evidence to suggest that the meteorologically-adjusted trend in Chicago is different from that occurring in St. Louis or Detroit.

5 Conclusions

The functional relationships between Chicago meteorological data and network typical ozone concentrations which were described in NISS Technical Report #5 also apply to Detroit and St. Louis meteorological data. Network typical ozone concentrations are better modeled using Chicago meteorological data than

	Chicago	St. Louis	Detroit
1981	-0.004401584	-0.01460325	0.0007404935
1982	0.015398416	-0.01520325	-0.0005275065
1983	-0.005321584	-0.02280325	0.0004474935
1984	0.001288416	-0.00660325	-0.0021475065
1985	0.002268416	-0.01240325	0.0003184935
1986	-0.001731584	-0.01160325	-0.0002735065
1987	0.006948416	-0.01330325	-0.0006505065
1988	0.008648416	-0.01860325	0.0001334935
1989	0.006498416	-0.00430325	0.0020304935
1990	-0.008021584	-0.00540325	0.0019684935
1991	-0.055631584	-0.01480325	0.0011834935

Table 5: Yearly trend parameter pseudo-values for the three subnetworks.

	Chicago	St. Louis	Detroit
standard error	5.652e-03	0.168e-02	3.657e-04

Table 6: Jack-knifed standard errors of the trend parameter for the three subnetworks.

	Chicago-Detroit	Chicago-St. Louis	St. Louis-Detroit
$\hat{D}$	-0.00243	0.01059	-0.01302
standard error of $\hat{D}$	0.005786116	0.005761198	0.001671462
p-value	0.571	0.1267	<0.001
confidence interval (%/decade)	(-18.9,14.0)	(-5.8,27.0)	(-17.8,-8.3)

Table 7: P-values and simultaneous confidence intervals for D, the difference in the means of the trend parameter. $\hat{D}$ is the estimate of D obtained by differencing the trend parameters in Tables 2–4.

St. Louis or Detroit meteorological data. The performance of the non-linear model developed in NISS Technical Report #5 is only marginally improved when Chicago, St. Louis, and Detroit meteorological data are all included in the model. When the rural network is divided into three subnetworks grouped around Chicago, Detroit, and St. Louis, the non-linear model performs better for the Chicago and St. Louis subnetworks than for the Detroit subnetwork.

The estimate of the trend parameter, y, is extremely variable depending on which meteorological data sets are used, and which subnetwork is modeled. In most cases, the standard error of the estimate of y is quite high, indicating a possible lack of significance of this variable in the model and collinearity in the variables.

When Chicago, Detroit, or St. Louis meteorological data are used individually to estimate y for the entire rural network, y values of -1.03 %/decade, 3.76 %/decade, and -2.74 %/decade, respectively, are found. The model which combines data from all three meteorological stations is fit with a trend parameter of 0.8%/decade. Subnetwork models for Chicago, St. Louis, and Detroit yielded parameter estimates of -2.0 %/decade, -12.5 %/decade, and +0.5 %/decade, respectively. Pairwise comparisons of the subnetwork trend parameters using standard errors obtained through jackknifing indicate that the meteorologically-adjusted trends in ozone concentrations differ between St. Louis and Detroit. This analyis was inconclusive regarding differences in trends between Chicago and St. Louis, and Chicago and Detroit.

References

Bloomfield, P., Royle, A. and Yang, Q. (1993). Rural Ozone and Meteorology: Analysis and Comparison with Urban Ozone, *Technical Report No. 5*. National Institute of Statistical Sciences, P.O.Box 14162, Research Triangle Park, NC 27709-4162.