NISS

Description of the Gulf States Ozone Monitoring Network and Decomposition into Subnetworks

J. Andrew Royle, Peter Bloomfield, Doug Nychka and Qing Yang

Technical Report Number 22
August, 1994

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

Description of the Gulf States Ozone Monitoring Network and Decomposition into Subnetworks*

J. Andrew Royle Peter Bloomfield Doug Nychka
Qing Yang

National Institutes of Statistical Sciences

and

Department of Statistics
North Carolina State University

Report: 12 August, 1994

Abstract

We describe a 118 station Gulf States ozone monitoring network and decompose this network into five spatially cohesive subnetworks using rotated principal components analysis. We provide empirical justification for this particular decomposition and perform some preliminary analyses of ozone in each of the subregions. Unadjusted (for meteorology) trends in network average maximum ozone among the subregions are different, ranging from nearly no trend in the "Louisiana" subregion to -17%/decade in the "Houston" subregion. The standard errors of the estimated trends are homogeneous across subregions and we suggest a model for ozone under which this may further justify our subnetwork specification. Estimated trends in the network maxima for each of the subregions produces results somewhat different than the analysis on the network average maxima. Future modeling work (e.g. meteorological adjustment of ozone) will primarily focus on ozone from these individual subnetworks.

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

Keywords: Ozone trends, Spatial sampling, Principal components analysis.

Contents

List of Tables

List of Figures

1 Introduction

In this report we describe a 118 station ozone monitoring network of the Gulf states region, composed of the states of Texas, Louisiana, Mississippi, Alabama, and Florida. We perform a principal components analysis on a 20 station subset of the 118 station network that has broad spatial and temporal coverage. This principal components analysis suggests five spatially cohesive subregions, which we then associate 69 of the stations with to form 5 ozone monitoring subnetworks. The 69 stations used were those that had greater than or equal to 100 days in common with the stations used in the principal components analysis. Two of these networks contain the stations surrounding Dallas and Houston, TX, and the remaining 3 have no tight geographic focus. In Section 5 we provide empirical justification for our network specification based on the within and among region correlations between stations. In Section 6 we estimate unadjusted trends in ozone for each of the 5 subnetworks, and discuss a model for ozone under which these results further suggest that our subregions are meaningful. The trends are estimated for both the network average maxima calculated from a median polish and the network maximum.

Continuing and future work will involve modeling ozone using local meteorology in some or all of these subnetworks.

2 Description of the Data

The Gulf states ozone network is comprised of 118 stations in Texas, Louisiana, Mississippi, Alabama, and Florida. These stations are described in Table 1. The locations of the 118 stations are shown in Figure 1.

We used the data at each station between 1 April and 31 October (the ozone 'season') from 1981 to 1991. Two stations with codes 481130044 and 481130069 were found to have identical longitude and latitude coordinates, but non-overlapping temporal data. These were assumed to be the same station, but had changed from a 'residential' classification to a 'commercial' classification. Since this would not affect most analyses, we combined them into a single station. If future work were to focus on whether there are land use effects on ozone, then this will have to be acknowledged, and the two stations treated separately.

3 Principal Component Analysis on the Gulf States Network

Due to the large geographic coverage of the Gulf states network, we would like to reduce the network to a few spatially cohesive subnetworks. These networks will likely be more homogeneous in terms of local meteorology than the network as a whole, therefore grouping these stations together makes sense considering our objectives of building meteorology based models to ozone.

We used a principal components analysis to identify homogeneous ozone regions of the whole network. In the Chicago rural ozone network analysis of Bloomfield, Royle and Yang (1993*b*) a principal component analysis revealed three reasonably distinct clusters of ozone stations. These were analyzed separately in unpublished work and were found to be reasonably distinct in terms of their observed relationships between ozone and meteorology and had quite different ozone trends. The principal component analysis was performed using the daily maxima at stations with a suitable number of daily observation. The daily maxima were calculated for each station using all steps of imputation as was done in the Chicago urban ozone analysis discussed in Bloomfield, Royle and Yang (1993*a*).

Because a PCA would require no missing data for the day by station data matrix, we reduced the number of stations to a suitable subset of the 118 stations that gave reasonable spatial coverage, while also maintaining reasonable temporal

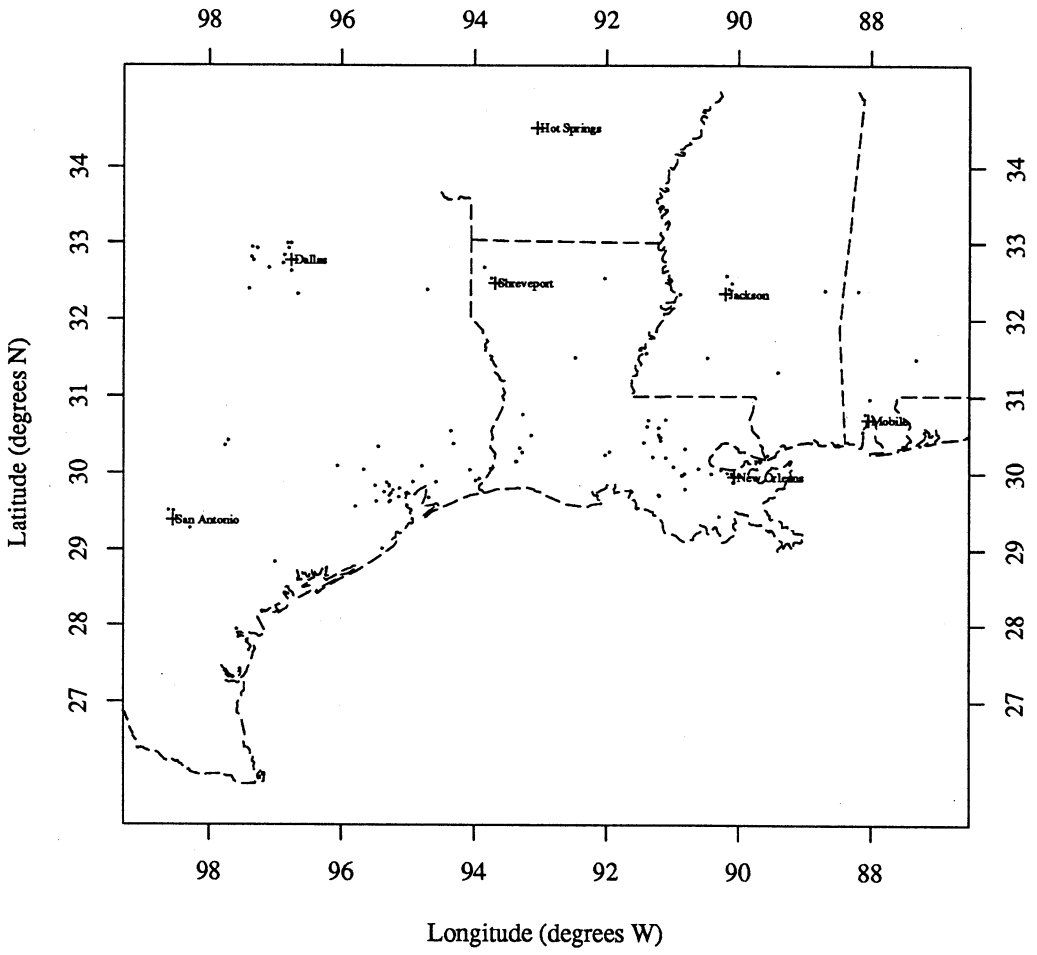

Figure 1: Locations of 118 Gulf States ozone monitoring stations.

Table 1: The ozone monitoring stations. "AIRS" is the EPA air quality data base. "MSA" is the Metropolitan Statistical Area identifier for the station location. Dates of first and last observations are given in "yymmdd" form. First letter of Code is: R-residential, I-industrial, C-commercial, A-agricultural, M-mobile, F-forested. Second letter of Code is: S-suburban, U-urban, R-rural.

AIRS ID	Lat	Lon	MSA	Code	First and Last Dates		State
010970003	30.77	88.088	5160	RS	820303	921129	AL
010970025	30.541	88.124	5160	IR	810113	820227	AL
010970028	30.958	88.028	5160	IS	810101	921129	AL
010990001	31.487	87.327	0	CS	811231	821029	AL
011190002	32.364	88.202	0	FR	911015	921129	AL
120330004	30.533	87.2	6080	IS	811231	921230	FL
120330018	30.368	87.271	6080	MS	811231	921230	FL
220050003	30.083	90.983	760	IR	830607	840916	LA
220110002	30.492	93.144	0	IR	890402	921230	LA
220111001	30.767	93.267	0	–	830601	831019	LA
220150008	32.534	93.75	7680	CU	811231	921230	LA
220170001	32.676	93.86	7680	AR	810114	921230	LA
220190001	30.152	93.363	3960	IS	830109	830925	LA
220190002	30.143	93.372	3960	IS	830930	921230	LA
220190008	30.262	93.284	3960	–	920930	921230	LA
220191003	30.326	93.323	3960	IS	810122	920929	LA
220330003	30.419	91.183	760	RU	811231	921230	LA
220330004	30.461	91.188	760	CU	811231	891230	LA

continued on next page

Table 1 (continued)

AIRS ID	Lat	Lon	MSA	Code	First and Last Dates		State
220330006	30.464	91.188	760	CU	900107	920730	LA
220330008	30.7	91.1	760	–	921231	921230	LA
220330009	30.461	91.177	760	–	920802	921230	LA
220331001	30.587	91.207	760	AR	811231	921230	LA
220430001	31.5	90.471	0	AR	890103	921230	LA
220430101	31.5	92.472	0	–	810112	821003	LA
220470002	30.2	91.1	0	RR	811231	921230	LA
220470006	30.208	91.292	0	–	921231	921209	LA
220470007	30.4	91.425	0	–	921231	921230	LA
220510003	29.996	90.174	5560	CS	830601	831020	LA
220511001	30.043	90.275	5560	RS	820425	921230	LA
220512001	29.883	90.083	5560	RS	841231	890320	LA
220550003	30.233	92.017	3880	CS	830602	920604	LA
220550004	30.267	91.95	3880	–	920610	921230	LA
220570001	29.433	90.3	3350	CR	830515	840910	LA
220570002	29.792	90.804	3350	RS	890315	921230	LA
220630002	30.312	90.813	760	–	921231	921230	LA
220710005	29.951	90.075	5560	CU	811231	821230	LA
220710012	29.994	90.103	5560	RU	810201	921230	LA
220730002	32.533	92.033	5200	IU	810427	921230	LA
220770001	30.685	91.367	0	AR	881115	921230	LA
220770002	30.6	91.383	0	AR	830608	840916	LA
220870002	29.982	89.999	5560	RS	810310	921230	LA
220890003	29.984	90.411	5560	IR	910228	921230	LA
220890100	29.984	90.41	5560	RS	830509	861230	LA

continued on next page

Table 1 (continued)

AIRS ID	Lat	Lon	MSA	Code	First and Last Dates		State
220930001	29.972	90.859	0	IR	830531	840925	LA
220930002	29.994	90.82	0	IR	880930	921230	LA
220950002	30.058	90.608	5560	IR	810212	921230	LA
221010002	29.7	91.2	0	IS	830601	840913	LA
221010003	29.715	91.21	0	IS	880930	921230	LA
221210001	30.502	91.21	760	CS	830309	921230	LA
280010004	31.561	91.391	0	CU	910228	921129	MS
280450001	30.231	89.563	0	IR	910228	921129	MS
280490010	32.388	90.141	3560	CS	811231	921129	MS
280590005	30.381	88.494	6025	IS	840229	920530	MS
280590006	30.377	88.534	6025	–	920813	921129	MS
280730002	31.319	89.409	0	CR	910228	921129	MS
280750002	32.372	88.703	0	RU	910228	921129	MS
280890001	32.466	90.111	3560	RS	810105	821129	MS
280890002	32.568	90.184	3560	AR	880314	921129	MS
281490004	32.323	90.887	0	CS	910228	921129	MS
480290001	29.282	98.296	7240	AR	811231	810430	TX
480290032	29.514	98.621	7240	RS	810716	921230	TX
480290036	29.503	98.54	7240	RS	811231	921230	TX
480290041	29.429	98.494	7240	CU	811231	810505	TX
480391003	29.008	95.392	1145	RS	811231	921230	TX
480710900	29.879	94.921	3360	AR	810731	870227	TX
480710902	29.882	94.576	3360	AR	850205	861230	TX
480710903	29.675	94.676	3360	AR	850207	861230	TX
480850004	32.989	96.823	1920	RS	860702	861005	TX

continued on next page

Table 1 (continued)

AIRS ID	Lat	Lon	MSA	Code	First and Last Dates		State
481130044	32.826	96.864	1920	RS	811231	860420	TX
481130045	32.926	96.808	1920	RU	811231	921230	TX
481130052	32.719	96.891	1920	RS	810131	871230	TX
481130055	32.617	96.757	1920	–	820331	921230	TX
481130069	32.826	96.864	1920	CU	860707	921230	TX
481130086	32.987	96.759	1920	–	860423	860929	TX
481131047	32.774	96.553	1920	–	810818	811022	TX
481390082	32.326	96.664	1920	MR	910228	911030	TX
481570004	29.565	95.799	3360	CS	901231	901226	TX
481671002	29.399	94.933	2920	RS	811231	921230	TX
481830001	32.382	94.713	4420	AR	811231	921230	TX
481990002	30.381	94.316	840	CU	891231	921230	TX
482010024	29.875	95.326	3360	RS	811231	921120	TX
482010026	29.802	95.125	3360	AS	810101	881230	TX
482010029	30.039	95.675	3360	AR	810409	921230	TX
482010038	29.838	95.286	3360	RS	810121	820509	TX
482010039	29.612	95.279	3360	RS	811231	840423	TX
482010046	29.827	95.284	3360	RS	830504	921229	TX
482010047	29.835	95.496	3360	RS	810110	921230	TX
482010051	29.624	95.474	3360	RS	811231	921230	TX
482010059	29.706	95.281	3360	RS	860107	921230	TX
482010062	29.631	95.267	3360	–	840510	921230	TX
482011003	29.679	95.131	3360	–	810618	921230	TX
482011034	29.771	95.222	3360	RS	811231	921230	TX
482011035	29.733	95.257	3360	IS	811231	921230	TX

continued on next page

Table 1 (continued)

AIRS ID	Lat	Lon	MSA	Code	First and Last Dates		State
482011037	29.752	95.361	3360	CU	811231	921230	TX
482017001	29.572	95.017	3360	CS	811231	891230	TX
482450009	30.039	94.075	840	RS	811231	921230	TX
482450010	29.922	93.926	840	RS	810607	841230	TX
482450011	29.894	93.988	840	RS	861231	921230	TX
482510002	32.387	97.403	0	AR	900219	901030	TX
482570001	32.761	96.312	1920	AR	910228	911030	TX
482910089	30.086	94.783	0	AR	881231	881204	TX
483390088	30.338	95.452	0	AR	881231	881204	TX
483550019	27.789	97.433	1880	RS	811231	810407	TX
483550022	27.834	97.553	1880	RS	811231	810407	TX
483550025	27.764	97.433	1880	R-	810924	921230	TX
483550026	27.835	97.557	1880	–	840625	921230	TX
483611001	30.058	93.764	840	RS	811231	921230	TX
484090002	27.94	97.589	1880	A-	810511	840610	TX
484390005	32.664	97.1	2800	AR	811231	810614	TX
484391002	32.805	97.357	2800	CU	811231	921229	TX
484391003	32.761	97.329	2800	CU	811231	810503	TX
484392002	32.944	97.353	2800	AR	810210	820805	TX
484392003	32.922	97.279	2800	AR	820916	921230	TX
484530003	30.422	97.706	640	FR	810719	921230	TX
484530014	30.356	97.758	640	RS	811231	921230	TX
484570101	30.544	94.346	0	AR	891231	920530	TX
484690003	28.838	97.005	8750	CS	890331	921230	TX
484730001	30.094	96.066	3360	AR	900114	901230	TX

coverage. Essentially we constructed the data matrix for use in the PCA by maximizing the product of its dimensions. This produced a 20 station network shown in Figure 2. These 20 stations had 757 days in common. These days did not include any observations from 1981, 1982, or 1983 and included 55, 90, 106, 127, 131, 149, and 99 days in the years 1984 to 1991 respectively.

The first 6 modes of the PCA on the 20 station sub-network explain approximately 45.4, 14.4, 6.5, 5.5, 3.8, and 3.6 percent of the variation in this 20 station subnetwork respectively. The singular values and some related quantities are shown in Table 2.

We used the varimax rotation of the first 6 modes of the PCA to help facilitate the interpretation of these 6 modes. The station loadings of the first 6 rotated components are shown in Table 3. Spatial interpolation of these loadings using `interp` (Becker, Chambers and Wilks (1988)) for each of the 6 modes are shown in Figures 3, 4 and 5. A zero-contour line is given to facilitate interpretation.

The image plots of the PCA loadings indicate some reasonable, geographically cohesive, initial subregions that correspond roughly to the Houston area stations ("Houston" subregion), the stations of the eastern portion of the network ("Eastern" subregion), the Dallas area stations ("Dallas" subregion), the stations of the southwestern network ("Southwest" subregion), the Louisiana stations ("Louisiana" subregion), and the sixth mode weighs heavily only on a single station west of Houston. This station will be examined in detail in Section 4.

We will define our 6 subregions to be the stations associated with each of these 6 dominant modes. Thus, each subregion is characterized by the principal component time series which represents it.

4 Specification of the Subnetworks

This principal component analysis uses very little of the available data to break the Gulf Coast region into smaller cohesive subregions. A problem with this is that each of the subregions then has only a small number of stations available with which to build suitable network summaries for modeling purposes. To resolve this, we correlated the daily maxima for all stations with the first 6 principal component time series. We then assigned each station to one of the 6 regions based on the correlation. Each station was assigned to the subregion whose principal component time series it was most highly correlated. We only used stations which had a minimum of 100 days in common with the data used in

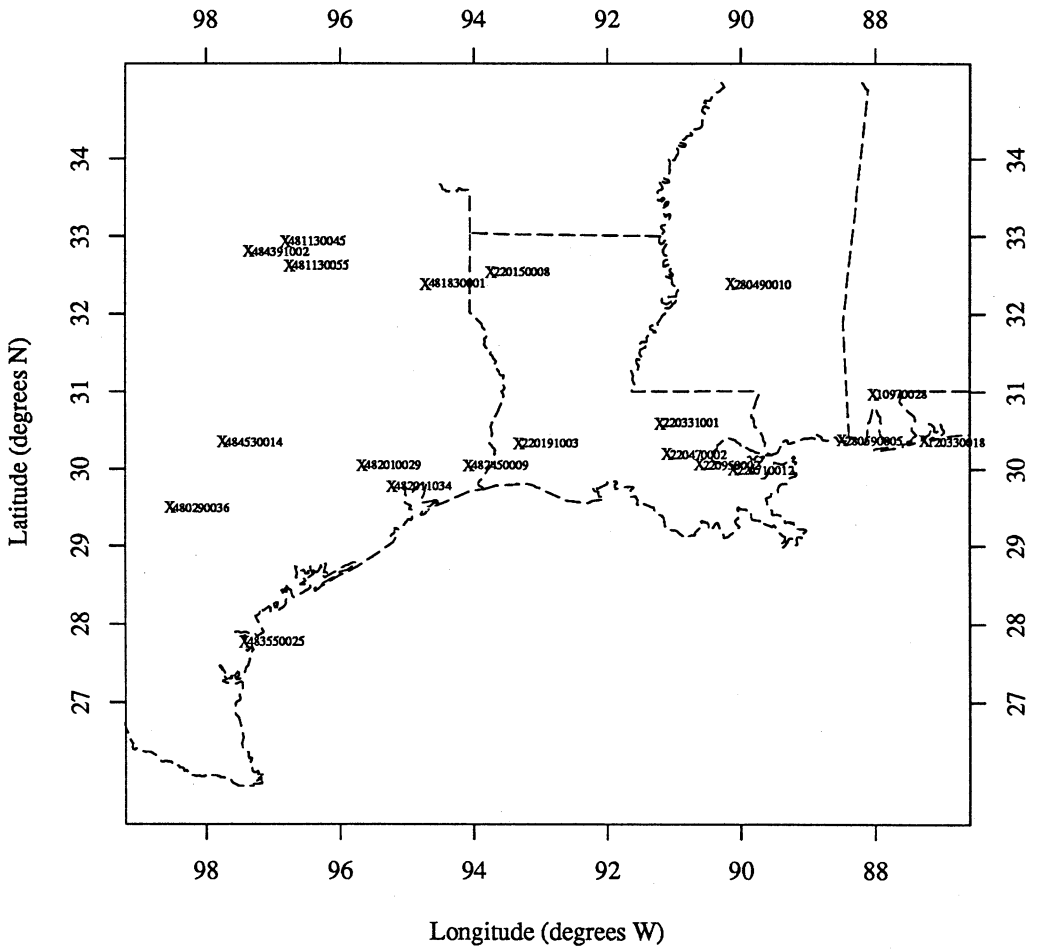

Figure 2: Locations of 20 Gulf States stations used in PCA.

Table 2: Results of principal components analysis for the 20 station network.

Singular value	Percent of variance	Cumulative percent
1920.2	45.43	45.43
1081.4	14.41	59.84
725.7	6.49	66.33
668.2	5.50	71.83
553.4	3.77	75.60
536.5	3.55	79.15
472.5	2.75	81.90
444.4	2.43	84.33
414.1	2.11	86.44
406.9	2.04	88.48
375.5	1.74	90.22
354.3	1.55	91.77
348.1	1.49	93.26
329.6	1.34	94.60
311.9	1.20	95.80
295.5	1.08	96.88
286.0	1.01	97.89
265.7	0.87	98.76
238.9	0.70	99.46
209.9	0.54	100.00

Table 3: Station loadings for the first six rotated principal components.

Station	Component					
	1	2	3	4	5	6
481830001	0.1983	-0.0444	0.2990	-0.0691	-0.1390	-0.0558
280590005	0.1153	-0.4579	0.0217	-0.0602	-0.1870	-0.0041
481130045	0.2150	-0.0357	0.5516	-0.0162	-0.0306	-0.0708
280490010	0.0935	-0.2325	0.1260	0.0383	-0.2012	-0.1664
220950002	0.1448	-0.3496	-0.0083	-0.0707	-0.3452	-0.0666
220470002	0.1901	-0.2489	0.0344	-0.0821	-0.4022	-0.1642
220331001	0.0757	-0.1558	0.0390	0.1182	-0.4703	-0.2779
220710012	0.0962	-0.3336	0.0282	-0.0056	-0.2784	0.0412
220150008	0.1880	-0.0857	0.2349	-0.0059	-0.2247	-0.1525
480290036	0.1265	-0.0045	0.2453	-0.4395	-0.0043	-0.0498
220191003	0.2046	-0.0322	0.0942	-0.1304	-0.3427	-0.0077
484391002	0.1413	0.0357	0.4846	-0.2202	-0.0144	-0.0156
482011034	0.6354	-0.0826	0.1178	-0.0179	-0.0556	-0.0750
484530014	0.1494	0.0046	0.2440	-0.4339	-0.0419	-0.1163
120330018	0.1135	-0.4600	0.0044	-0.0160	-0.1533	0.0277
482010029	0.3728	-0.0183	0.0609	-0.1736	-0.1322	-0.8263
483550025	0.0462	-0.0399	0.0653	-0.6029	0.0061	-0.2643
10970028	0.0436	-0.4223	0.0229	0.0585	-0.0970	-0.1372
482450009	0.3214	0.0528	0.0370	-0.3087	-0.3158	0.1753
481130055	0.1480	-0.0466	0.3794	-0.1555	-0.0389	-0.0754

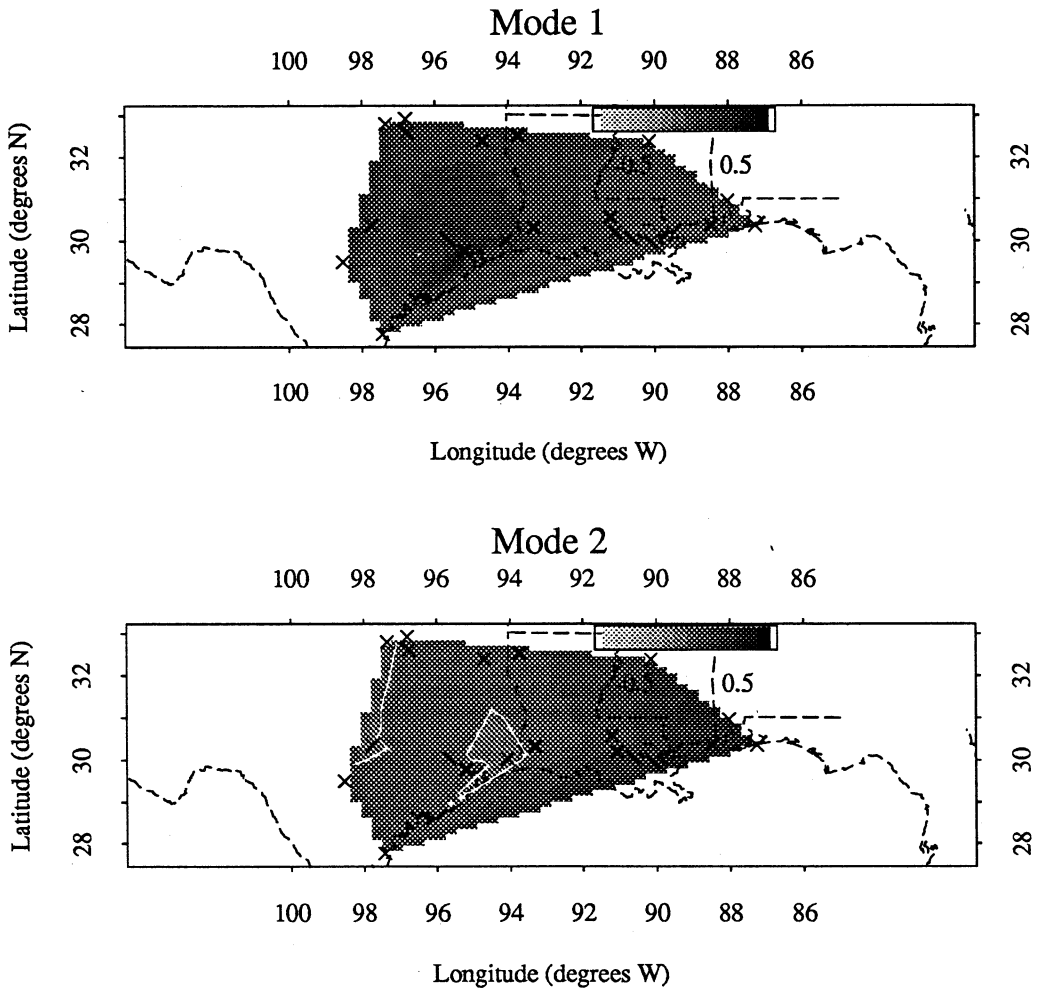

Figure 3: Image plots of the loadings of rotated components 1 and 2 of the 20 station sub-network.

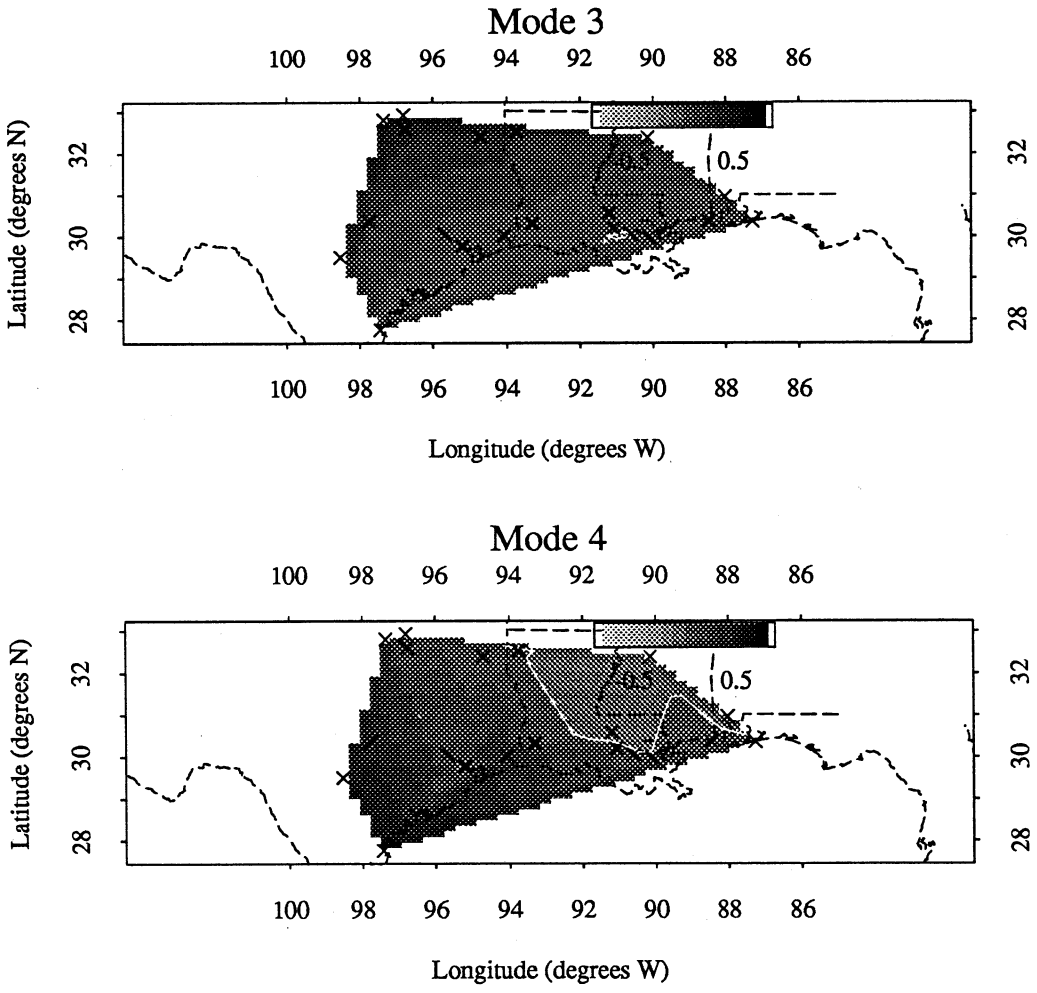

Figure 4: Image plots of the loadings of rotated components 3 and 4 of the 20 station sub-network.

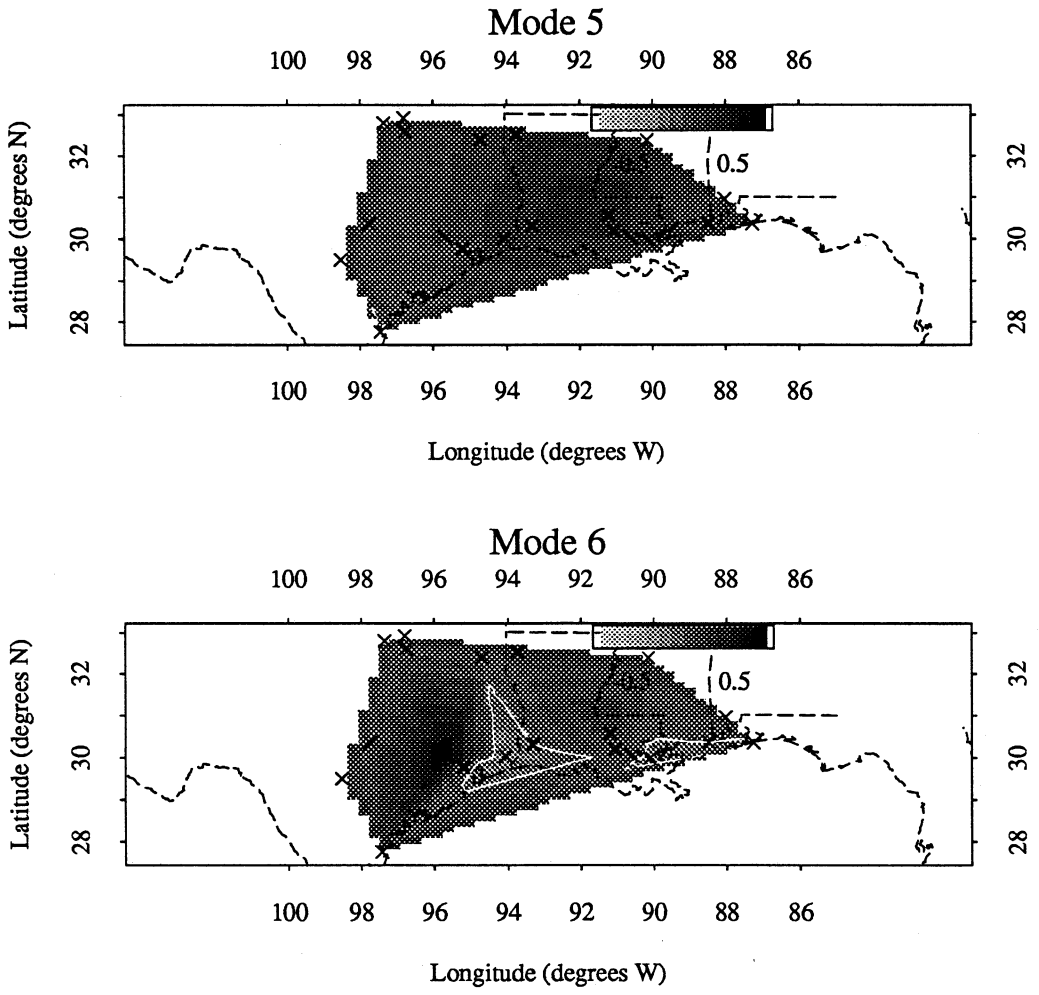

Figure 5: Image plots of the loadings of rotated components 5 and 6 of the 20 station sub-network.

the principal component analysis in order to get fairly precise estimates of the correlations. Although this is fairly arbitrary, and one could use stations that have fewer than 100 days in common, the additional stations would contribute only marginally to the formation of any network ozone summary such as the network typical, or maximum. A total of 69 stations had at least 100 days in common with the data used in the PCA. The locations and region that each station was associated with are shown in Figure 6. The number of stations associated with each subregion is given in Table 4.

Figure 6 shows several stations that are disjoint from the rest of the stations in their respective subregions. In particular, there is a region 4 station northeast of Houston, whereas the rest of the region 4 stations lie southwest of Houston. Two stations that are associated with region 1 are closer to both regions 3 and 5, and two stations of region 2 are mixed in with region 5 stations in southern Louisiana. It may also be desirable to place the single station of region 6 in with the Houston area stations. Table 5 shows the correlations of these 6 stations with the time series associated with each of the 6 regions. It is evident that the difference between the correlation of the region that each of these stations was associated with and the region with which we would like to have the stations associated with due to their spatial location is not large in all cases. The largest difference exists for the station that was assigned to region 6. Here, the correlation between that station's daily maxima and the principal component time series representing region 6 is 0.56 whereas it's correlation with the time series representing region 1 is 0.47. Thus, for the sake of maintaining the spatial integrity of each of these subnetworks we will reassign these 6 stations to the subregion with which they have the second highest correlation. The new allocation of these 69 stations to sub-networks is shown in Figure 7 and the subnetwork allocation is given in Table 6.

It is peculiar that a single station (number 484730001) was the only one to receive such a high loading for the 6th mode. This suggests that something unusual

Table 4: Number of stations in each of 6 sub-regions.

sub-network	1	2	3	4	5	6
number of stations	25	7	7	9	20	1

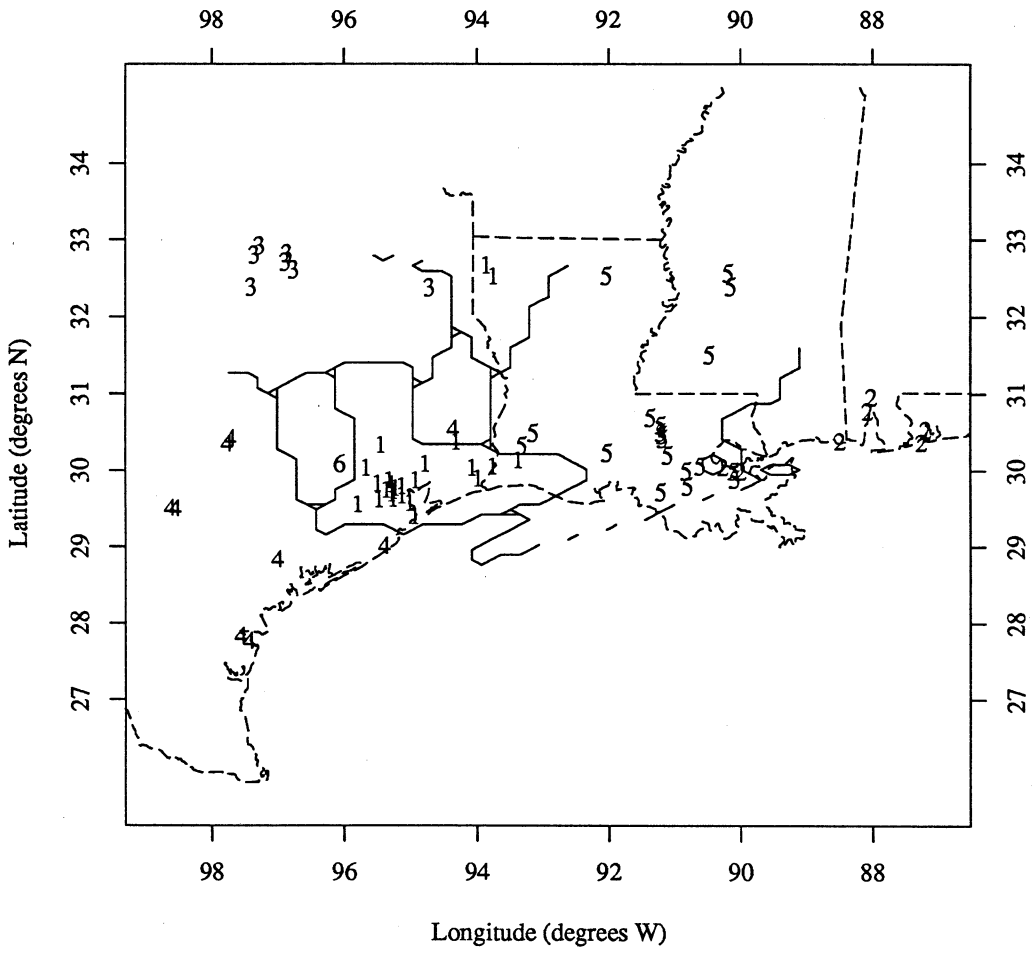

Figure 6: Sub-region association of 69 ozone monitoring stations.

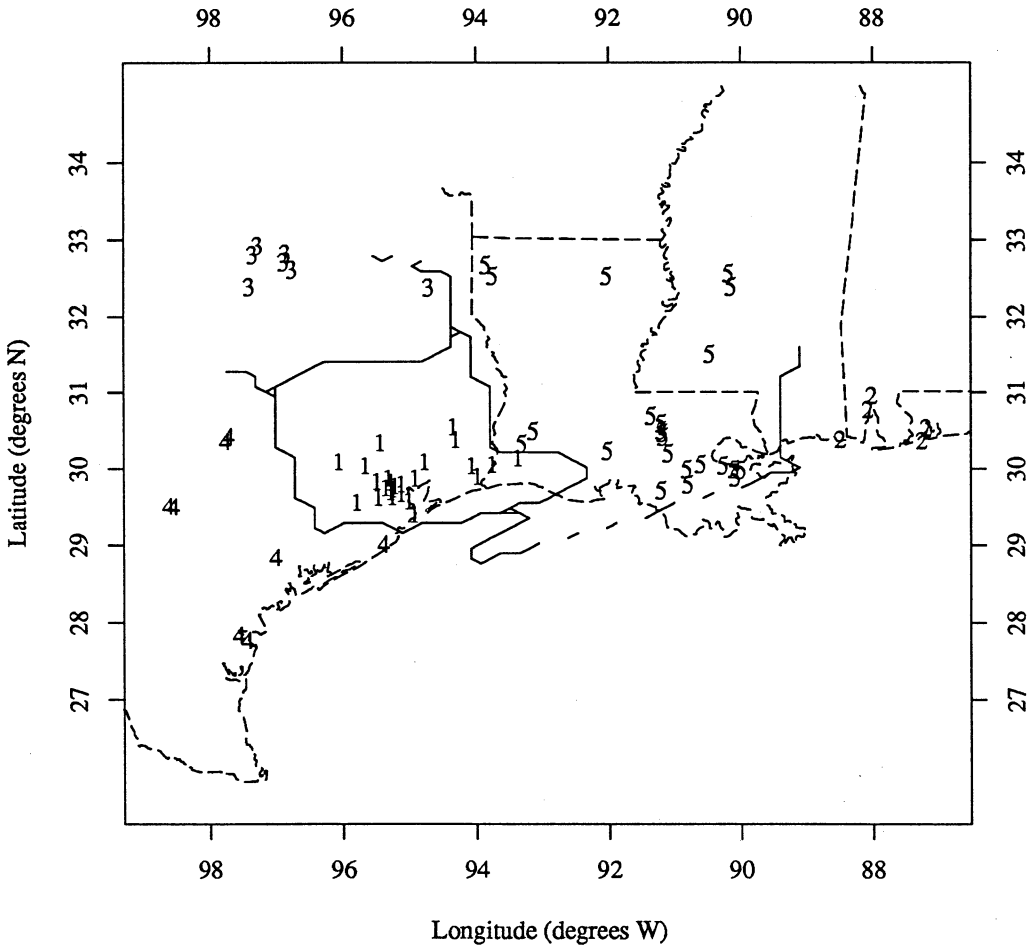

Figure 7: Alternative sub-region association of 69 ozone monitoring stations.

Table 5: Correlations of daily maxima of 6 stations with the PCA time series.

Station	region 1	region 2	region 3	region 4	region 5	region 6
484730001	0.47	-0.10	0.04	0.38	-0.15	0.56
484570101	0.28	0.08	0.15	0.34	0.21	0.16
220150008	0.49	0.14	0.41	0.01	0.44	0.18
220170001	0.43	0.14	0.37	0.04	0.42	0.17
220870002	0.32	0.52	0.04	0.08	0.46	0.03
220511001	0.28	0.54	0.03	0.15	0.51	0.09

has occurred at this site to cause this particular station to be so different than the rest, or that this site is somehow quite distinct from the remainder of the sites in the whole network. This site has data available only for the year 1990. The data from this station is shown in Figure 8, smoothed using the cross-validated smoothing spline estimate. The nearest station, number 482010029, is shown for comparative purposes. Overlaying these plots are 6 degree-of-freedom smoothing splines to better elucidate the annual behavior of ozone for each of these two stations, and the corresponding smoothed version of the subregion 1 ("Houston") network typical maximum. It is evident here that the deviation of the peculiar station (484730001) from the rest of the network is such that at this station ozone moves broadly counter to "Houston" ozone as a whole starting in about May, reaching an annual minimum at about the time the rest of the network achieves it's yearly maximum. The nearby station (482010029) is roughly synchronous with the "Houston" network ozone. In all periods of prolonged high ozone, station 484730001 never achieves as high a level of ozone as the "Houston" region, and in many cases peak ozone is only 40 to 50 percent that of the "Houston" region. The reasons for this difference in behavior, which might include local meteorological effects and instrumental problems, have not been determined. However, the differences on the seasonal scale seem large enough to justify not including this station in the Houston region. In future analyses, consideration will have to be given as to whether to include station 482010029 in the "Houston" subregion ozone, or to discard it all together. For the remaining analyses of this report we have discarded this station, and so are

Table 6: Allocation of the 69 stations to 5 subnetworks.

Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5
220190002	10970003	481130044	480290032	220110002
480710900	10970028	481130052	480290036	220150008
481570004	120330004	481130055	480391003	220170001
481671002	120330018	481830001	483550025	220191003
481990002	280590005	482510002	483550026	220330003
482010024	–	484391002	484530003	220330004
482010026	–	484392003	484530014	220330006
482010029	–	–	484690003	220331001
482010046	–	–	–	220430001
482010047	–	–	–	220470002
482010051	–	–	–	220511001
482010059	–	–	–	220512001
482010062	–	–	–	220550003
482011003	–	–	–	220570002
482011034	–	–	–	220710012
482011035	–	–	–	220730002
482011037	–	–	–	220770001
482017001	–	–	–	220870002
482450009	–	–	–	220930002
482450011	–	–	–	220950002
482910089	–	–	–	221010003
483390088	–	–	–	221210001
483611001	–	–	–	280490010
484570101	–	–	–	280890002
484730001	–	–	–	–

working with a 24 station Houston network.

5 Empirical Justification of Subnetworks

We have assumed up to this point that the rotated principal component analysis identifies meaningful subregions, and that by correlating the station maxima with the principal components, we can associate these stations with regions in a meaningful way. These regions are 'nice' in the sense that they happen to be spatially cohesive, and so it seems logical to assume that the stations within each subregion experience similar meteorology, ozone production mechanisms and ozone and ozone precursor transport. Here we provide some empirical justification that the stations within each subregion are, in some way, similar. Section 6 provides a brief description of the ozone within each subregion.

One way in which to determine if stations are homogeneous within subregions is to examine the correlations between daily station maxima within and among subregions. We would expect that the correlation between stations within a subregion is larger than the correlation between stations among subregions. Intuitively, this should be the case because, by design, the stations within a subregion are highly correlated with a time series which defines that particular region. Table 7 gives the median correlation between the daily maxima of all pairs of stations among and within the 5 subregions. The subregions are listed in the table from east to west, with subregion 2 being the most eastern, and subregion 4 being the most western.

Table 7 shows that the highest correlation between stations occurs between stations within a subnetwork, and this correlation is generally greater than 0.60. Correlations between stations across subnetworks is on average only half this, being as small as 0.07 and as large as 0.49. Given that these subnetworks are spatially cohesive, we would expect these results assuming that the correlation between any two spatial locations is inversely proportional to the distance between locations. The regions in Table 7 are approximately ordered from east to west, and the correlations generally decrease as the distance between regions gets larger.

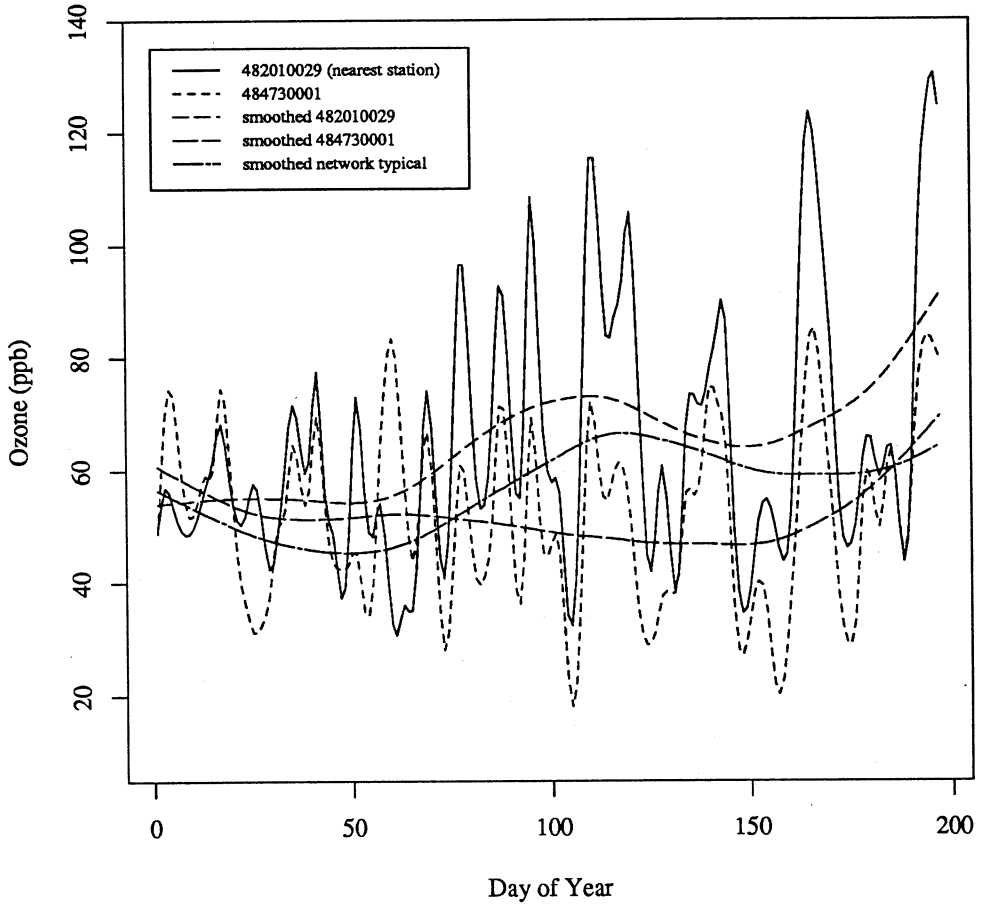

Figure 8: Smoothed ozone from station 484730001 with a nearby station and the network typical ozone.

Table 7: Correlations between stations within and among subregions.

Region	Region 2	Region 5	Region 1	Region 3	Region 4
Region 2	0.61	0.46	0.24	0.12	0.07
Region 5	–	0.58	0.40	0.26	0.22
Region 1	–	–	0.57	0.43	0.44
Region 3	–	–	–	0.66	0.49
Region 4	–	–	–	–	0.66

6 Description of Ozone Within the Subregions

These subregions differ somewhat in both the network typical and network maxima over time. The 'network typical' value is that computed by performing a median polish on the matrix of daily station maxima as was done in Bloomfield et al. (1993*a*). Figure 9 shows the smoothed network typical of each of these networks over time. Smoothing was done using a 6 degree of freedom smoothing spline. Figure 10 shows the corresponding smoothed network maxima. These plots show that Houston ozone is considerably higher than the remaining regions, the difference being more striking in the plot of the daily maxima. Also, the daily maximum ozone is lowest in the Eastern and Southwestern subregions.

We can calculate the unadjusted ozone trends for each of these subregions by fitting a model with seasonal components and a linear trend term. We used the annual and semi-annual frequencies of a Fourier series to represent the seasonality of ozone as was done in Bloomfield et al. (1993*a*). Figure 11 shows the median smoothed daily ozone over all years for each of the 5 subregions. Smoothing was done using a 6 degree-of-freedom smoothing spline. Ozone over the April to October period is strongly bimodal with peaks in ozone occurring in May and July or August for most regions. There is variability in both strength of the seasonality (which we can define as the seasonal range of ozone), and the timing of the modes across subregions.

For the network typical ozone, the trend and standard errors for each of the subregions are given in Table 8. The standard errors adjusted for serial correlation

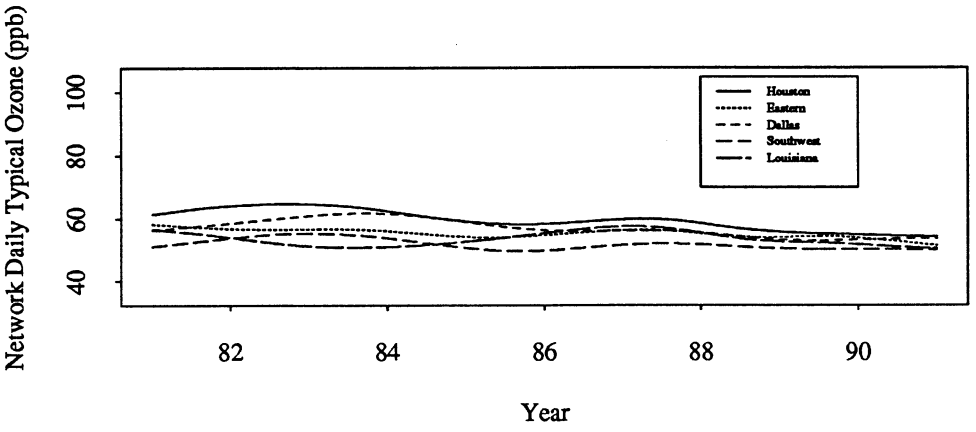

Figure 9: Network typical for each of the 5 Gulf states sub-regions.

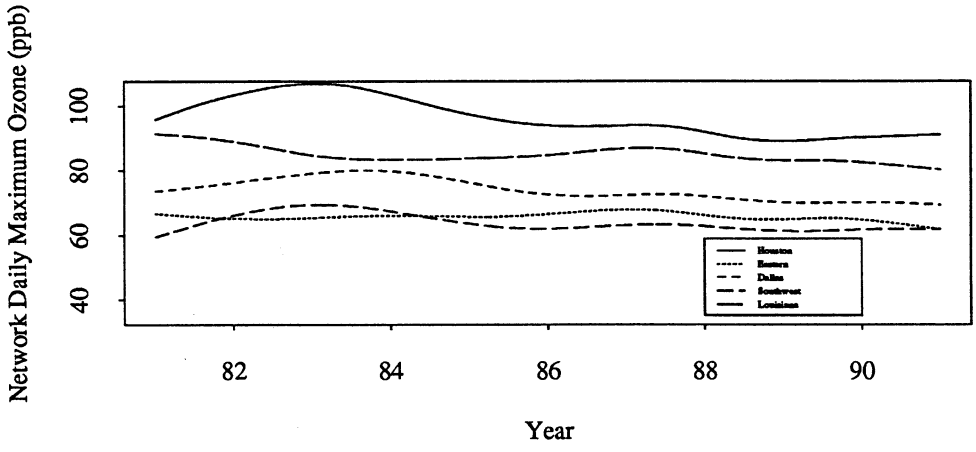

Figure 10: Network maxima for each of the 5 Gulf states sub-regions.

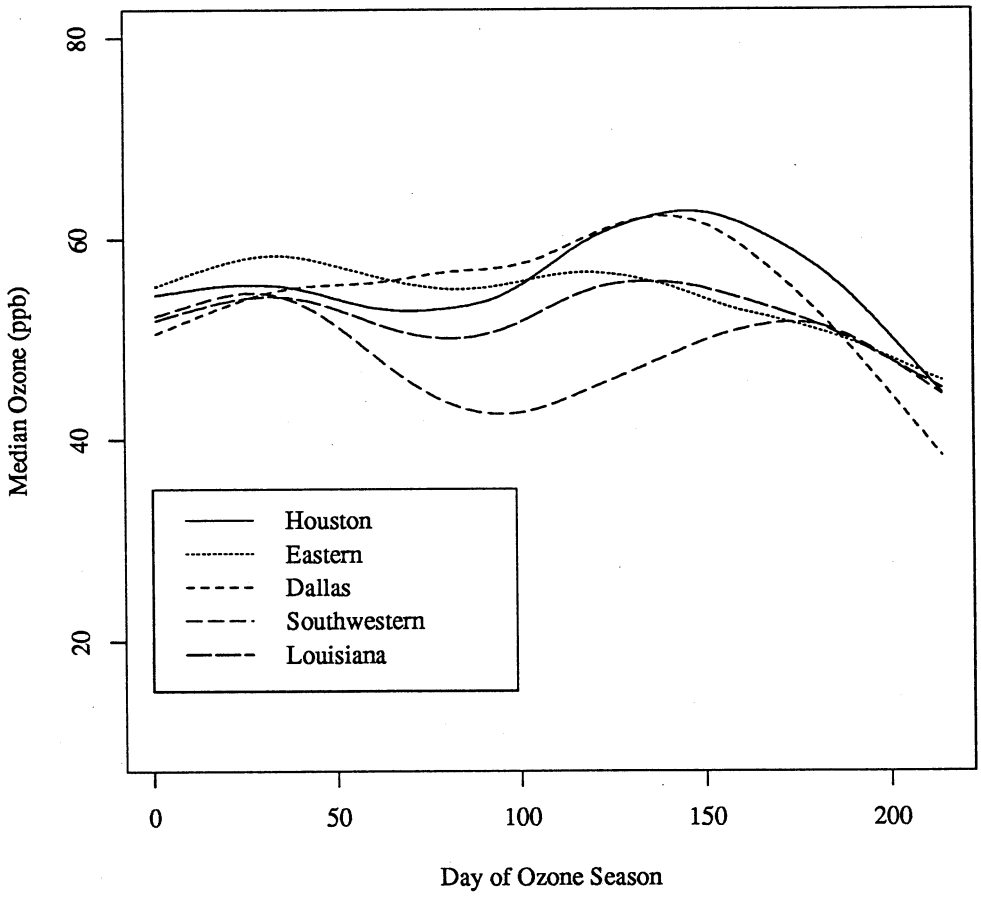

Figure 11: Daily median ozone from 1980 to 1991 for the 5 subregions.

and heteroscedastic errors using the method of Gallant (1987) are also given. The R^2 of these fits is given to allow relative comparisons of the quality of the fit of this very reduced model. The estimated trends and standard errors for each of the subnetwork maxima are given in Table 9

It is obvious here that the trend in ozone is highly variable among these subregions, with the largest (in magnitude) trend occurring in the Houston area. For the network typical, the unadjusted trend for the Houston subregion is -16.4%/decade, and this is highly significant. In contrast, the Louisiana subnetwork has an estimated trend of near zero. Also worthy of note here is the apparent difference in seasonality of the Dallas subregion relative to the other subnetworks. The R^2 of this fit, containing only trend and seasonal components, is 0.104 for the Dallas subnetwork. And Figure 11 shows that the decline in ozone at the end of the season is much more dramatic than in the other subregions and also the seasonality of Dallas ozone is less bimodal than elsewhere. The other four subnetworks achieve R^2s of less than one half that of Dallas, and in the case of the Louisiana subnetwork, near one fifth this value. For Louisiana though, this small R^2 is partially due to the insignificance of the trend. However this also suggests little seasonality for the Louisiana subnetwork at least as far as can be modeled in this fashion. The results for the subnetwork maxima are comparable except that the trend in the Louisiana maxima is considerably larger in magnitude than the network typical value, and the trend in the Eastern stations maxima is near zero.

The standard errors of these trend estimates from Table 8 and Table 9 are remarkably homogeneous given that each subnetwork is composed of quite different numbers of stations. We can pose a reasonable model that, assuming the model to be true, the homogeneity of these standard errors suggests that the subregion specification is reasonable. This model was used to model trends in Dobson Total Ozone records in Bloomfield, Oehlert, Thompson and Zeger (1983), and a similar model was used by Oehlert (1993) to model sulfate wet deposition. This model for ozone over time at the kth station in the jth subregion is:

$$Ozone_{t,j,k} = \mu_{t,j,k} + \alpha_t + \beta_{t,j} + \gamma_{t,j,k} \tag{1}$$

Here α, β and γ are assumed to be stationary time series with internal autocorrelations but no cross-correlations. The series α_t is present in data from all stations in the *whole* network, $\beta_{t,j}$ is present in data from all stations in the jth subregion, and $\gamma_{t,j,k}$ is a station-specific noise series. We will assume that the station series all have the same serial covariances and the region series all have the same serial

Table 8: Unadjusted trends and standard errors for the subregion typicals computed conventionally and adjusted for heteroscedasticity and serial correlation.

Region	Trend	Conventional		Adjusted		
		Std. error	t Value	Std. error	t Value	R^2
Houston	-0.0164	0.00269	-6.10	0.00377	-4.36	4.6
Eastern	-0.0053	0.00203	-2.64	0.00311	-1.72	3.3
Dallas	-0.0121	0.00215	-5.59	0.00316	-3.81	10.4
Southwestern	-0.0069	0.00217	-3.18	0.00332	-2.08	4.6
Louisiana	-0.0003	0.00210	-0.13	0.00323	-0.08	2.2

Table 9: Unadjusted trends and standard errors for the subregion maxima computed conventionally and adjusted for heteroscedasticity and serial correlation.

Region	Trend	Conventional		Adjusted		
		Std. error	t Value	Std. error	t Value	R^2
Houston	-0.0170	0.00279	-6.10	0.00381	-4.46	5.3
Eastern	-0.0005	0.00215	-0.21	0.00308	-0.15	3.3
Dallas	-0.0116	0.00207	-5.61	0.00290	-4.00	11.5
Southwestern	-0.0082	0.00234	-3.53	0.00360	-2.29	2.9
Louisiana	-0.0054	0.00201	-2.71	0.00274	-1.99	5.3

covariances. Now, the estimated trend for a region has sampling variance from all three levels, but the part that is propagated from the station series goes down with the number of stations since the station series have the same internal covariances and are not cross-correlated. The part that comes from the region series would have the same magnitude in each region. Thus, if we observe standard errors that are more constant than inverse to the number of stations in a subregion, it suggests that the region series dominate the calculation. In our case, Table 8 shows standard errors for the trend that are nearly the same suggesting that there exists underlying regional effects dominating these standard errors. This adds further justification for our subregion specification of Section 5. Although this model is acknowledged to be highly simplified, it is useful for gaining insight and interpreting results such as we have here.

Presumably these subnetworks also differ in their local meteorology, however data for all subregions is unavailable to us at this time to make this determination.

7 Conclusions

We have described a network of 118 ozone monitoring stations in the states of Texas, Louisiana, Alabama, Mississippi, and Florida (the 'Gulf States' region). Sixty-nine of these were found to have enough temporal data to provide meaningful results, and were allocated to 5 subnetworks using a rotated principal components analysis. These subnetworks represent the "Houston", "Dallas", "Louisiana", "Eastern" and "Southwestern" portions of the Gulf States region.

The ozone among stations within subnetworks is more similar than ozone among stations across subnetworks, and the correlation between stations among subregions decreases approximately inversely to distance between subregions. This provides some justification for the given subnetwork specification and also suggests that there exists strong regional effects that play a role in determining ozone characteristics across the Gulf States region.

The 5 geographically cohesive subregions that these networks represent were found to have different ozone characteristics as measured by trend and seasonality. The estimated trend (unadjusted for meteorology) was negative in all five subregions, being the most negative in the Houston and Dallas subregions where the estimates are -16.4%/decade and -12.1%/decade respectively. The estimated trend in the Louisiana subregion was estimated to be only -0.03%/decade and was quite insignificant. The standard errors of the estimated trends are more sim-

ilar than would be expected based on the different numbers of stations in each subnetwork. This supports the notion that there are underlying dominant regional effects contributing to differences in ozone across subregions. The seasonality of ozone is distinctly bimodal in all subregions (but less so in the Dallas subregion), and differs in the timing of the modes and strength of the seasonality across subregions. The seasonality appears to be strongest in the Dallas subregion and weakest in the Louisiana subregion.

The decomposition of the Gulf States region into spatially cohesive subregions provides a starting point for the study of ozone in some or all of the individual subregions. Because each of these regions is more homogeneous with respect to ozone and meteorology than the Gulf States region as a whole, construction of meteorology based ozone models will be simplified and more meaningful, and results of modeling exercises will be more interpretable in the context of local meteorological conditions and local ozone transport and production mechanisms.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). *The New S Language*, Advanced Books and Software. Pacific Grove, California: Wadsworth.

Bloomfield, P., Oehlert, G., Thompson, M. L. and Zeger, S. (1983). 'A frequency domain analysis of trends in Dobson Total Ozone records', *Journal of Geophysical Research* **88**, 8512–8522.

Bloomfield, P., Royle, J. A. and Yang, Q. (1993*a*), Accounting for meteorological effects in measuring urban ozone levels and trends, Technical Report 1, National Institute of Statistical Sciences.

Bloomfield, P., Royle, J. A. and Yang, Q. (1993*b*), Rural ozone and meteorology: Analysis and comparison with urban ozone, Technical Report 5, National Institute of Statistical Sciences.

Gallant, A. R. (1987). *Nonlinear Statistical Models*. New York: Wiley.

Oehlert, G. (1993). 'Regional trends in sulfate wet deposition', *Journal of the American Statistical Association* **88**(422), 390–399.