Comparison of 700-hPa NCEP-R1 and AMIP-R2 Wind Patterns over the Continental United States Using Cluster Analysis

ELLEN J. COOTER, JENISE SWALL, AND ROBERT GILLIAM

*Atmospheric Sciences Modeling Division, Air Resources Laboratory, National Oceanic and Atmospheric Administration, Research Triangle Park, North Carolina**

(Manuscript received 5 July 2006, in final form 8 December 2006)

ABSTRACT

Clustering techniques are adapted to facilitate the comparison of gridded 700-hPa wind flow patterns spanning the continental United States. A recent decade (1985–94) of wind component data has been extracted from two widely used reanalysis datasets: NCEP-R1 and the NCEP–Department of Energy (DOE) Atmospheric Model Intercomparison Project, phase two (AMIP-R2). Metrics and measures are identified that facilitate the identification and comparison of large-scale wind flow. Comparison of the cluster results reveals dominant wind patterns common to both datasets as well as three types of reanalysis model differences: 1) relatively minor numerical differences; 2) differences produced by model corrections or parameterization changes, such as snow mask, snow depth, and moisture flux; and 3) systematic differences, such as orography, overocean radiation fluxes, and overocean data assimilation. A second analysis examines the frequency of 700-hPa wind patterns associated with key ozone-season (May–September) synoptic settings. Association of 1990–94 daily maximum 1-h ozone levels with these patterns across the United States follows documented meteorological dependencies. Above-average ozone levels in the Midwest and mid-Atlantic are associated with transitional anticyclone and easterly flow synoptic patterns (39.2% of ozone-season days) while above-average ozone levels across the southern United States are associated with the westward extension of the Bermuda high circulation (14.8% of ozone-season days). Below-average ozone levels throughout most of the eastern United States are associated with frontal passages and migratory anticyclones (29.6% of ozone-season days).

1. Background

The U.S. Environmental Protection Agency–National Oceanic and Atmospheric Administration (EPA/NOAA) Climate Impact on Regional Air Quality (CIRAQ) project is assessing the impact of present-day and future (circa 2050) climate on regional ozone and particulate matter ($PM_{2.5}$) in North America *via* the Community Multiscale Air Quality (CMAQ) model (Byun and Schere 2006). This assessment requires a decade of methodologically consistent current and future climate conditions to drive the air quality simulations. The development of methods and metrics to facilitate the careful analysis of biases in data produced by the climate model is a critical aspect of the CIRAQ project. Previous climate model evaluations most often focus on simulated surface variables, such as temperature and precipitation (e.g., Christensen et al. 2004; Kunkel and Liang 2005; Leung et al. 2004; Liang et al. 2004). More detailed analysis with explicit links to air quality include, among others, Hogrefe et al. (2004), Leung and Gustafson (2005), and Mickley et al. (2004). One potentially informative climate variable that is rarely included in global climate model evaluations is the 700-hPa wind field. Atmospheric flow patterns at 700 hPa reflect the location and movement of synoptic-scale features (e.g., cyclones and anticyclones), which, in turn, control large-scale surface temperature and precipitation patterns as well as the transport and transformation of air pollutants.

Gridded global reanalysis datasets are commonly used as a base against which historical and near-present-day global climate model simulations are compared. Reanalysis variables that are strongly influenced by assimilated data should agree most closely with ob-

* In partnership with the U.S. Environmental Protection Agency.

Corresponding author address: Ellen Cooter, NOAA/ARL/ASMD, mail drop E243-04, Research Triangle Park, NC 27711.
E-mail: ellen.cooter@noaa.gov

DOI: 10.1175/2007JAMC1527.1

servations and should show similarly good agreement across reanalysis models (Kistler et al. 2001). Data assimilation, however, is commonly performed every 6 h, while rawinsonde data, the primary source of upper-air information, are available only every 12 h. Those factors can lead to uncertainty in the characterization of large-scale patterns of atmospheric transport arising from differences in reanalysis model numerics, physics, and parameterizations. Identification of reanalysis similarities and differences regarding the simulation of wind flow patterns known to be closely linked to surface air quality can aid the diagnosis of air quality model projections, provide focus to global climate model evaluation, and support future climate model improvements.

The present study adapts two clustering techniques to facilitate the identification and comparison of gridded meteorological 700-hPa wind flow patterns that are 1) dominant in space and/or time and 2) represent statistically distinct meteorological regimes spanning the continental United States. Comparison metrics and measures are used to explore reanalysis model simulation of large-scale wind patterns and the ability of key 700-hPa wind flow regimes to accurately reflect regional surface ozone level tendencies.

2. Data and methodology

a. The gridded reanalysis data

Ten years, 1985–94, of zonal (u) and meridional (v) wind component data were acquired from two widely used gridded reanalysis datasets: National Centers for Environmental Prediction (NCEP; Kalnay et al. 1996), hereinafter referred to as NCEP-R1 or R1, and the NCEP–Department of Energy (DOE) Atmospheric Model Intercomparison Project, phase two (AMIP-II) (Kanamitsu et al. 2002), hereinafter referred to as AMIP-R2 or R2. Both models characterize observed conditions across the globe at a spatial resolution of 2.5° (T62) and 28 vertical levels through a combination of assimilated observations and physically based models. R2 follows R1 in development. Its purpose is to correct known problems in R1 and to serve as a basic verification dataset for the AMIP-II (Kistler et al. 2001). The R2 global analyses are made using an updated (relative to R1) forecast model, updated data assimilation system, improved diagnostic outputs, and corrections of known R1 processing problems. Variables included in the datasets are classified into three data types, depending on the strength of influence of assimilated observations: A, B, and C. Type A variables are most strongly influenced by assimilated data and type C variables are largely model derived. Both zonal (u) and meridional (v) winds are reported as A variables (Kalnay et al. 1996).

b. Dominant 700-hPa wind component analysis

A hierarchical cluster method, Ward's minimum variance (Ward 1963), is used to identify groups of 700-hPa wind component data containing patterns that are dominant in space and/or time over the continental United States (CONUS). This application closely follows the approach reported in Cohn et al. (2001). Five-day periods, hereinafter denoted "pentads," are constructed for each calendar day for each of four seasons, with the first, third, and fifth days of each pentad making up a reanalysis realization. This results in a matrix of pentad vectors $\mathbf{P}^j$, where $j = 1, \ldots, 2016$ elements (the 2 u and v components $\times$ 336 grid nodes $\times$ 3 days) for N_i days (rows), $i = 1, \ldots, 4$ seasons. Cohn et al. (2001) conclude that defining five clusters per season results in the best explanation of variations in surface level visibility and photochemical pollution potential. Therefore, five 700-hPa clusters per season are retained here as well.

Comparison of clusters across datasets was not part of the original Cohn analysis and so measures are needed that facilitate comparison. To this end, the pentads in a season are first clustered using Ward's method, and each cluster mean is determined. Next, we compute the sum of squares (SOS) difference between pentad cluster means across datasets [Eq. (1)]:

$$\mathrm{SOS}_{AB} = \sum_{j=1}^{2016} (\overline{P}^j_A - \overline{P}^j_B)^2, \tag{1}$$

where $\overline{P}^j_A$ is mean pentad cluster element j of R1 cluster A and $\overline{P}^j_B$ is mean pentad cluster element j of R2 cluster B. We then compute the Pearson's correlation r for each cluster pair across datasets [Eq. (2)]:

$$r_{AB} = \frac{1}{(2016-1)} \sum_{j=1}^{2016} \left[\frac{(\overline{P}^j_A - \overline{P}_A)}{S_A} \frac{(\overline{P}^j_B - \overline{P}_B)}{S_B} \right], \tag{2}$$

where S_A is the standard deviation of mean R1 cluster A, $\overline{P}_A$ is the mean across all elements of mean R1 pentad A, S_B is the standard deviation of mean R2 cluster B, and $\overline{P}_B$ is the mean across all elements of mean R2 pentad B.

Cluster pairs that produce the smallest SOS differences and largest positive correlations are considered further. Table 1 contains the SOS differences and the correlation values for the five winter-season Ward's clusters. The smallest SOS values and largest positive correlation values are in boldface. A cluster pair with a unique row and column maximum r and minimum SOS

TABLE 1. Winter-season example of correlation and SOS matrices for R1 and R2 clusters. Boldface indicates unambiguous cluster pairs.

Correlation						
		R2				
R1		1	2	3	4	5
	1	***0.95***	0.05	−0.60	−0.05	−0.39
	2	−0.23	−0.27	**0.80**	0.07	−0.56
	3	−0.05	**0.48**	−0.28	**0.81**	−0.23
	4	−0.63	−0.04	**0.35**	−0.55	**0.53**
	5	−0.24	−0.26	−0.28	−0.09	**0.75**
SOS						
		R2				
R1		1	2	3	4	5
	1	***22***	399	641	1079	641
	2	618	641	**108**	886	748
	3	456	**238**	438	**274**	490
	4	667	440	**216**	1242	**182**
	5	471	493	444	981	**116**

location is italicized. We then return to the nonstandardized observations to compute mean cluster wind speed, $(u^2 + v^2)^{1/2}$, and direction, $\arctan(u/v)$, for the highlighted cluster means. These values are mapped and the difference between the grid cell cluster mean wind speeds is computed to more easily identify both the level of visual similarity across the domain, as well as to highlight particular features that are characterized differently in space or time. Differences in component magnitude, regardless of sign, are displayed as wind speed differences. If wind components across datasets are of similar sign and change proportionally to one another across the dataset, then, no significant difference in direction will be indicated. These patterns are then considered in light of cluster uncertainty (section 2d) and known reanalysis model differences.

c. Wind component frequency analysis

Simulation of specific large-scale weather patterns is another important aspect of reanalysis model performance. The association between these patterns and air quality varies with geographic region (Lehman et al. 2004) and may, as in the case of the northeastern United States, relate to the frequency and timing of frontal passages or to persistent patterns that promote stagnation and extended periods of photochemical activity, as in the case of the southeastern United States (Rao et al. 2003).

Eder et al. (1994) apply a two-stage clustering technique to a wide range of surface and upper-air meteorological variables at Birmingham, Alabama, for a decade of summer ozone seasons (May–September). An average linkage clustering technique is first used to identify statistically distinct meteorological regimes. Decadal frequencies are then determined via a k-means nonhierarchical clustering algorithm. The k-means method developed by MacQueen (1967) uses an iterative approach that allows for the reclassification of days after they have been grouped into a cluster, thus refining the final cluster solution. Days are assigned to the cluster with the nearest seed values. After all days have been assigned, the centroid of each cluster is recalculated and then used as the new cluster seed. These steps are repeated until the change in new seed values converge to approximately zero. Eder et al. (1994) identify seven homogeneous synoptic patterns (Table 2) that are later paired with coincident daily maximum 1-h ozone observations. For this present analysis, R1 and R2 700-hPa daily (as opposed to pentad) u and v component data for the example dates listed in Table 2 are provided as "seeds" to a k-means clustering algorithm, which is then applied to the full R1 and R2 datasets.

d. Cluster uncertainty

The overall goal of any cluster analysis is to discover a category structure or set of "natural groups" that fits the observations (Anderberg 1973). The classification process may be complicated by imperfect class definitions, overlapping categories, and random variations in the observations. These factors cannot only impact the assignment of observations to a group, but in doing so, can impact the outcome of later comparison across dataset groups. One means of minimizing the effect of random variations and improving class definition is to standardize the data across the variables prior to clustering. In the present case, we will be comparing the cluster results across datasets, so that the data are standardized across datasets as well—that is, joint standardization. Outlier values can also distort the identification of natural data groupings. Ward's method is particularly sensitive to outlier values but, given the geographic extent of our CONUS domain, these outlier values could contain physically meaningful information, and so they are retained.

Once the data have been standardized and the choice has been made to retain outliers, there are two alternative ways the clustering algorithm can be applied to multiple datasets: joint clustering, which provides all the data from both datasets to the algorithm simultaneously, and alternatively, the independent application of the clustering algorithm to each jointly standardized dataset. If the two datasets are identical, then cluster assignments will be identical regardless of the clustering alternative chosen. The principal advantage of performing joint clustering is that the algorithm is allowed to identify which standardized observations are most

TABLE 2. Description of the ozone-season synoptic patterns (after Eder et al. 1994).

Synoptic setting	Description	Example date
Bermuda high (BH)	Westward extension of the quasi-permanent Bermuda high pressure cell across the southern United States with extensive cloud cover and rain to the left of the circulation center.	5 Aug 1983
Southwesterly flow (SWF)	The Bermuda high is positioned eastward of the BH location.	20 May 1982
Migratory anticyclone (MA)	An anticyclone that moves southward out of Canada, usually following a cyclone and its associated cold front. Characterized by northerly winds resulting in cool temperatures, low cloud cover, and high insolation along the leading edge of high pressure.	5 Sep 1984
Cold-frontal passage (CFP)	This pattern is characterized by a well-defined wind shift and decreasing temperatures, and dewpoints behind the front.	8 Sep 1981
Transitional anticyclone (TA)	An anticyclone that generally forms in western Canada. Once formed, it typically traverses the northern sections of the United States in response to its proximity to the summer jet stream. During the summer, however, it often dissociates from the jet stream, stalls, and becomes vertically coupled with an upper-level anticyclone. Once this occurs, a rapid transition from a cold-core to a warm-core anticyclone begins.	17 Sep 1982
Cold-core easterly flow (CCEF)	A strong anticyclone off the northeast coast results in moist easterly flow off the Atlantic Ocean. Dewpoints increase, cloud cover is high, and insolation is low.	2 Aug 1986
Warm-frontal passage (WFP)	Omega (blocking) high pressure is present at 700 hPa and higher, with precipitation, cool temperatures, and low ozone levels near the upper-level cyclones and low moisture and clear skies near the upper-level anticyclone.	25 May 1990
*Warm-core easterly flow (WCEF)	A ridge of high pressure over Florida results in strong moist easterly flow off the Atlantic Ocean.	12 Aug 1994*

* Pattern and date added to Eder et al. (1994) set during the course of present analysis.

similar, thereby minimizing or eliminating the need for more qualitative comparison metrics. Recall, however, that the overall goal of cluster analysis is to identify natural groups in each dataset. The Ward's analysis cluster mean is recalculated after each assignment so that the joint cluster mean is "conditioned" by the presence of both datasets and imposes class definitions that are not necessarily correct for either dataset alone. Independent clustering allows each dataset to direct the evolution of the cluster mean and should result in more correct, although still imperfect, cluster definition.

The final factor impacting cluster assignment and subsequent comparison of cluster means is cluster overlap (misclassification). This influence is minimized for the synoptic analysis because the clustering objective identifies statistically distinct meteorological categories (section 2c). In contrast, the Ward's objective considers only measures of similarity and largely ignores cluster separation. The presence of overlapping clusters was explored using canonical discriminant analysis (SAS Institute Inc. 1999; Wilks 1995). Results of this analysis indicate that misclassification is present in both datasets and all seasons and is likely to lead to imperfect cluster definition. It also reveals that the likelihood of misclassification is essentially the same across datasets for spring, summer, and autumn seasons. Winter-season misclassification is substantially greater in R2 than in R1. This difference most likely relates to R2 corrections and modifications to R1 simulations, the majority of which target the winter season. No adjustment is made to correct for the effect of misclassification, but results of alternative climatological analyses are used whenever possible to confirm that physically meaningful mean wind flow patterns have been identified. Quantitative use of Ward's analysis cluster frequencies is strongly discouraged.

3. Results

a. Dominant spring-season wind patterns

Correlation and SOS analysis of the spring season, 1 March through 31 May, identifies five cluster pairs to be compared (Table 3). Pentad correlations range from 0.40 to 0.89. Average grid cell wind speeds for all clusters range from 1 to 16 m s^{-1}. Domainwide mean wind speed bias, computed as (R2 − R1), ranges from −0.81 to +1.46 m s^{-1}. The sign of wind speed bias is consistent throughout most mean cluster pentads, but the bias varies across clusters. A range of 700-hPa wind observation uncertainty estimated for data collected from the mid-1980s to mid-1990s of 2.4–6.1 m s^{-1} suggests that differences less than 1.0 m s^{-1} are not statistically sig-

TABLE 3. Spring-season correlations, SOS, and mean wind speed bias for R1 and R2 mean cluster pairs.

R1 cluster	R2 cluster	Pentad correlation	SOS	Mean wind speed bias ($m\,s^{-1}$) (R2 − R1) Day 1	Day 3	Day 5
1	5	0.40	209	+1.49	+1.46	+1.17
2	3	0.89	47	−0.10	+0.11	+0.17
3	1	0.71	102	+0.29	+0.26	+0.26
4	4	0.84	78	−0.44	−0.41	−0.51
5	2	0.81	143	−0.56	−0.81	−0.60

nificant (Parrish and Derber 1992). Visual inspection of wind speed and direction difference maps indicates good agreement regarding wind direction, and there is no consistent geographic wind speed bias across cluster pairs. A t test on the unclustered jointly standardized R1 and R2 data indicates that a null hypothesis (H_0) of equal u wind component means cannot be accepted ($\alpha > 0.1$) and so, while similar, spring-season R1 and R2 700-hPa wind data are not statistically identical but are not necessarily exclusively the product of random variations.

Figure 1a illustrates the R1 and R2 cluster pair with the lowest SOS and largest r value. The pattern resembles the Great Basin trough pattern reported in Davis and Walker (1992) in which winds are northwesterly over the West Coast but shift to westerly or southwesterly over Nevada. The polar front is displaced to the south. During spring, this cluster occurs most frequently in May and represents a transition pattern from

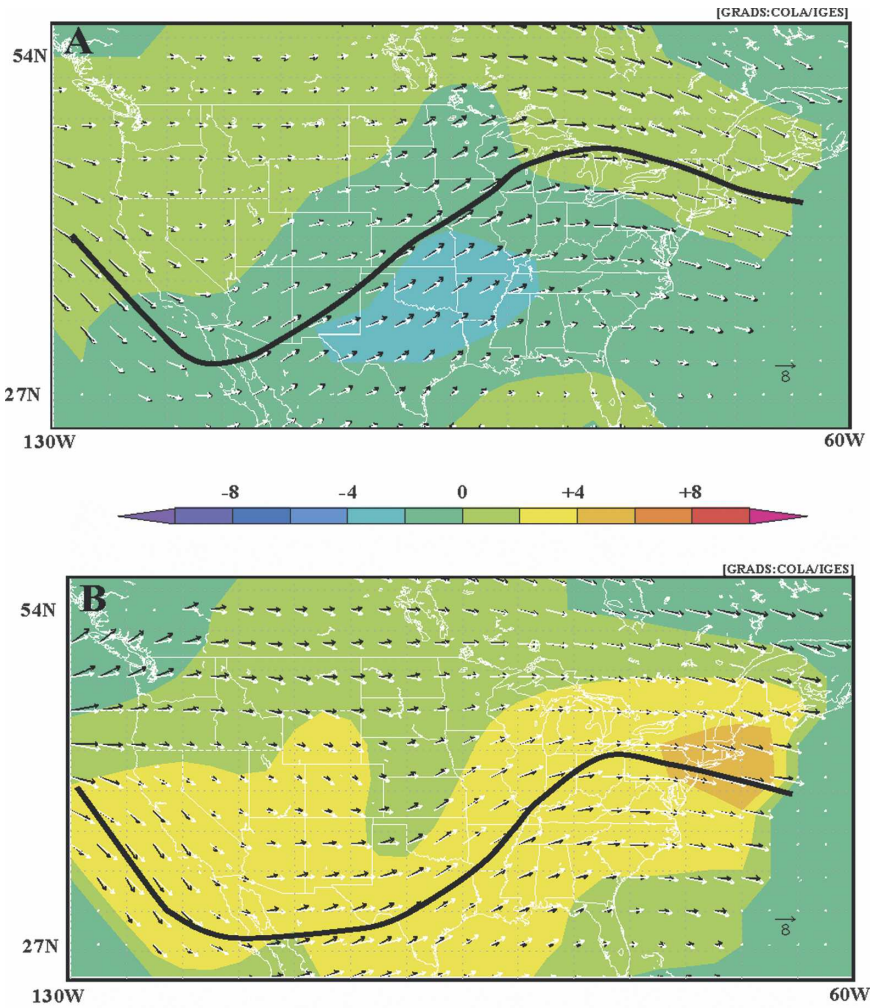

FIG. 1. The 700-hPa 1800 UTC (a) NCEP-R1 cluster 2 and AMIP-R2 cluster 3 and (b) NCEP-R1 cluster 1 and AMIP-R2 cluster 5 spring-season transport pattern differences. Black arrows represent NCEP-R1 wind directions. White arrows represent AMIP-R2 wind directions. Arrows are scaled to wind velocity. Velocity differences ($m\,s^{-1}$) are computed as (R2−R1). Solid line highlights the centerline of the trough feature.

cold spring to warmer summer conditions. Other dominant patterns include the westward extension of the Bermuda high circulation (R1 cluster 3 and R2 cluster 1), a strong subtropical jet with southwesterly flow (R1 cluster 5 and R2 cluster 2), and a strong ridge over the Rockies and trough over the eastern United States reminiscent of the meridional phase of the Pacific–North American (PNA) teleconnection pattern (R1 cluster 4 and R2 cluster 4) (Davis and Walker 1992; Eder et al. 1994; Leathers et al. 1991). The spatial location of the PNA teleconnection on time scales of months to seasons is a function of the position of the mean stationary waves. The resulting mean flow over the Pacific–North American sector is characterized by a trough in the east-central North Pacific Ocean, a ridge over the Rocky Mountains, and a trough over eastern North America (Leathers et al. 1991). SOS, r, mean bias, and visual inspection suggest that differences between R1 and R2 datasets for these four patterns are small.

The weakest spring-season cluster assignment (Fig. 1b) also resembles a Great Basin trough pattern. In this case, dataset disagreement involves the amplitude of the wave feature and maximum wind speeds within the jet stream core, which produce wind speed biases 2 to 3 times that of the other cluster pattern means and most likely account for the t-statistic differences noted above.

b. Dominant summer-season wind patterns

Correlation and SOS analyses of the summer season, 1 June through 31 August, identify two unambiguous and four ambiguous (boldface) cluster pairs to be compared (Table 4). An unambiguous cluster pair is one in which there is a unique row and column maximum r and minimum SOS location (Table 1). An ambiguous cluster pair is identified by multiple row and/or column maximum r and minimum SOS locations, all of which are examined visually. Table 4 mean pentad correlations range from 0.55 to 0.75. Average grid level wind speeds for all clusters range from 1 to 12 m s^{-1}. Domainwide wind speed biases are not likely to be significant, that is, less than 1 m s^{-1}. While wind directions also show excellent agreement, t-test results for the jointly standardized unclustered data indicate that a null hypothesis of equal R1 and R2 v component variance cannot be accepted.

The strongest correlation results are associated with R1 pattern 5 and R2 pattern 4 (Fig. 2), which resemble the cold-frontal passage pattern reported by Eder et al. (1994). The relatively large SOS reflects differences in the position of cyclonic circulation in the northeast and anticyclonic circulation in the south-central United States. R1 pattern 4 and R2 pattern 3 resemble the westward extension of the Bermuda high pattern reported in Eder et al. (1994) in the southern United States and the western monsoon pattern described in Davis and Walker (1992). The western monsoon pattern reflects the northernmost position of the polar jet and evidence of a weak trough off the West Coast, as well as the typical anticyclonic circulation over the southern United States and transport of Gulf moisture into the southwest. The remaining, more ambiguous associations are various manifestations of the "normal" position of anticyclonic circulation associated with the Bermuda high, western monsoon, and "dry" or pre/postmonsoon circulations (Davis and Walker 1992; Eder et al. 1994). While the location of summer-season anticyclonic circulation somewhere in the southern United States is predictable, the exact position of the center of circulation and the geographic extent of its influence is temporally variable. This is reflected in significant cluster overlap within each dataset and the t-test variance results across datasets.

TABLE 4. Summer-season correlations, SOS, and mean wind speed bias for R1 and R2 mean cluster pairs. Boldface indicates an ambiguous cluster association.

R1 cluster	R2 cluster	Pentad correlation	SOS	Mean wind speed bias (m s^{-1}) (R2 − R1) Day 1	Day 3	Day 5
1	**1**	**0.67**	**167**	**−0.24**	**−0.27**	**−0.24**
2	**1**	**0.68**	**103**	**−0.34**	**−0.20**	**−0.03**
3	**2**	**0.60**	**84**	**+0.27**	**+0.17**	**+0.08**
3	**5**	**0.55**	**157**	**−0.21**	**−0.06**	**−0.05**
4	3	0.70	77	−0.07	0.00	−0.05
5	4	0.75	247	−0.02	−0.10	−0.06

c. Dominant autumn-season wind patterns

Correlation and SOS analyses of the autumn season, 1 September through 30 November, identify three unambiguous and three ambiguous (bold text) R1 and R2 cluster pairs for comparison. The t-test results on the unclustered jointly standardized data suggest no significant difference between R1 and R2 u or v wind component means or variances. Table 5 pentad correlations range from 0.66 to 0.91. Average grid cell wind speeds range from 1 to 18 m s^{-1}. Domainwide mean wind speed bias ranges from −0.61 to +1.28 m s^{-1}. The sign of wind speed bias is consistent throughout the pentad in most cases but varies in direction (positive or negative) across cluster pairs.

Patterns R1 cluster 1 and R2 cluster 2 resemble the dry summer monsoon pattern described in Davis and Walker (1992) in the western United States. In the autumn season this pattern occurs most frequently during September and represents postsummer monsoonal

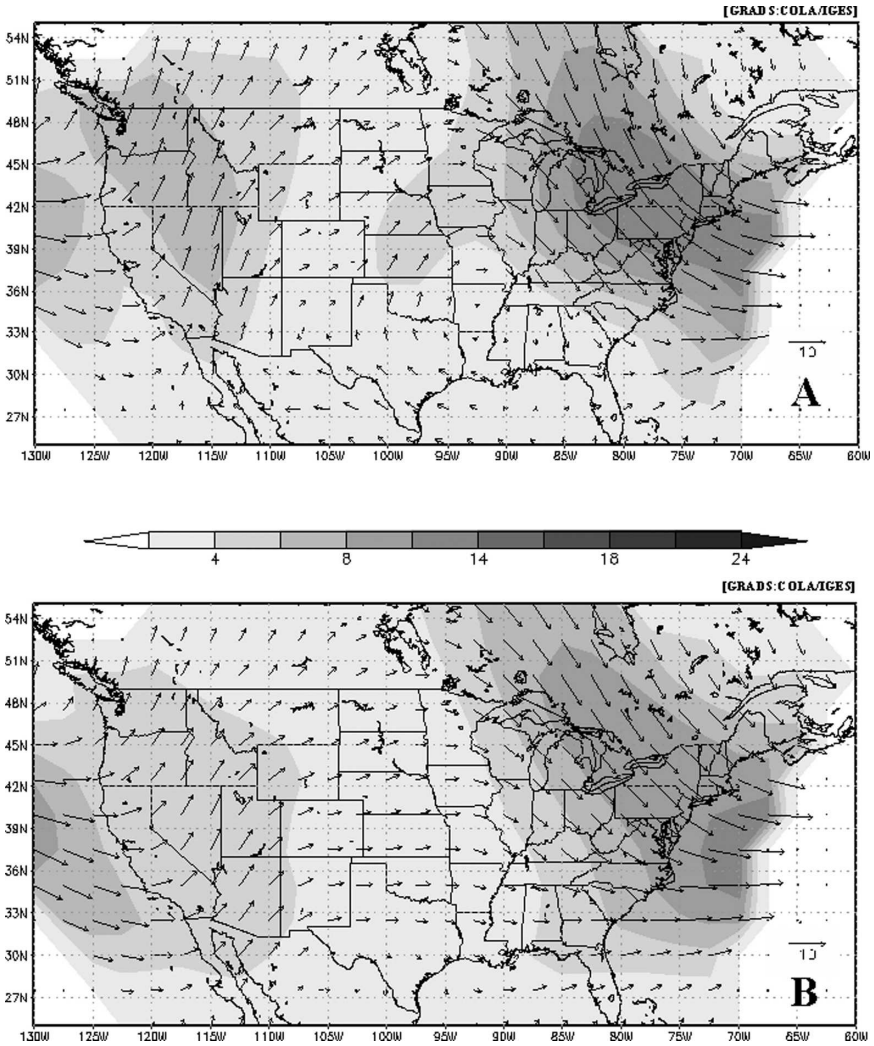

FIG. 2. Summer-season 700-hPa 1800 UTC wind speed (m s^{-1}) and direction for (a) NCEP-R1 cluster 5 and (b) AMIP-R2 cluster 4. Arrows are scaled to velocity.

flow. The polar jet is weak and located far to the north but is slightly stronger than during the moist monsoon pattern. The second unambiguous pattern, R1 cluster 3 and R2 cluster 1, resembles the most common position of the Bermuda high circulation over the southeastern United States (Eder et al. 1994). The last unambiguous cluster pair, R1 cluster 4 and R2 cluster 4, likely reflects the Great Basin trough circulation pattern described by Davis and Walker (1992). They report that this circulation occurs most frequently in the autumn during October and represents a transition from summer- to cooler-season patterns.

TABLE 5. Autumn-season correlations, SOS, and mean wind speed bias for R1 and R2 mean cluster pairs. Boldface indicates an ambiguous cluster association.

R1 cluster	R2 cluster	Pentad correlation	SOS	Mean wind speed bias (m s^{-1}) (R2 − R1) Day 1	Day 3	Day 5
1	2	0.67	68	−0.03	−0.25	−0.08
2	**5**	**0.66**	**193**	**+0.56**	**+1.19**	**+1.28**
3	1	0.91	83	−0.41	−0.46	−0.61
4	4	0.82	69	−0.07	−0.17	−0.05
5	**3**	**0.76**	**178**	**+0.31**	**+0.35**	**+0.24**
5	**5**	**0.75**	**163**	**+0.10**	**+0.14**	**−0.39**

All three ambiguous patterns involve a ridge positioned over the Rocky Mountains and a trough located over the eastern United States that resemble the meridional phase of the PNA teleconnection pattern illustrated in Fig. 3a (Leathers et al. 1991). SOS, r, t-test values, and canonical analysis all suggest random variation in the observations and overlapping cluster assignments strongly influence the appearance of these patterns and their comparisons.

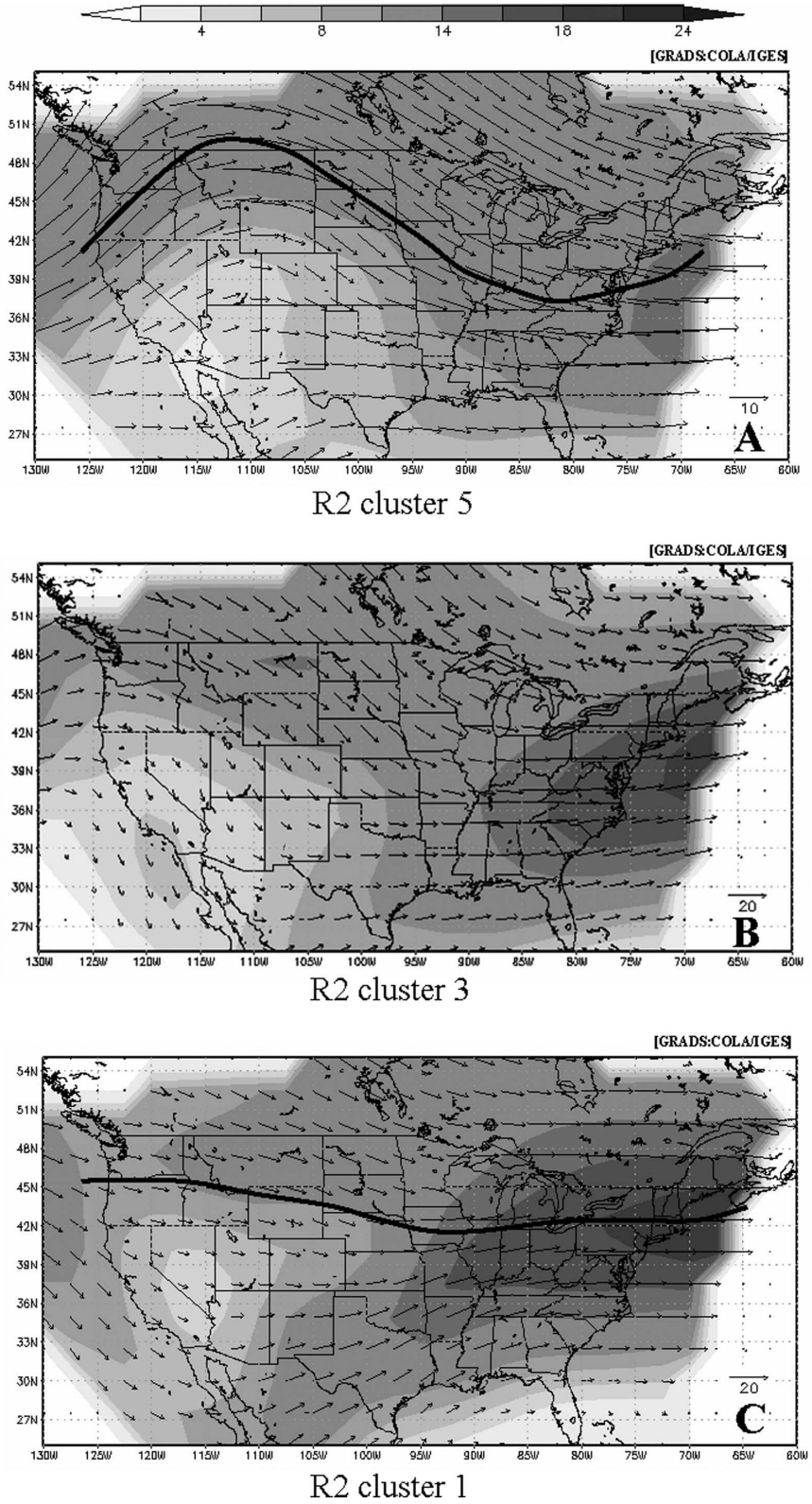

FIG. 3. Winter-season 700-hPa 1800 UTC R2 wind speed ($m\,s^{-1}$) and direction for (a) meridional (positive phase), (b) mean, and (c) zonal (negative phase) PNA patterns. Arrows are scaled to velocity.

TABLE 6. Winter-season correlations, SOS, and mean wind speed bias for R1 and R2 mean cluster pairs. Boldface indicates an ambiguous cluster association.

R1 cluster	R2 cluster	Pentad correlation	SOS	Mean wind speed bias ($m\,s^{-1}$) (R2 − R1) Day 1	Day 3	Day 5
1	1	0.95	22	−0.16	−0.18	−0.26
2	**3**	**0.80**	**108**	**−0.34**	**−0.63**	**−0.71**
3	**2**	**0.48**	**238**	**+1.34**	**+1.35**	**+0.73**
3	**4**	**0.81**	**274**	**+0.45**	**+0.31**	**−0.37**
4	**5**	**0.53**	**182**	**−0.28**	**−0.81**	**−0.81**
5	**5**	**0.75**	**116**	**−0.22**	**−0.35**	**−0.19**

d. Dominant winter-season wind patterns

Correlation and SOS analyses of the winter season, 1 December through 28 February, identify only one unambiguous match, R1 cluster 1 and R2 cluster 1 (Table 6), leaving six ambiguous (boldface) pairs to be compared. Pentad correlations range from 0.35 to 0.95. Average grid cell wind speeds for all clusters range from 1 to 22 $m\,s^{-1}$ in both datasets. Mean domainwide wind speed bias for the Table 6 cluster pairs range from −0.83 to +1.35 $m\,s^{-1}$. In most cases, the sign of wind speed bias is consistent throughout the pentad but varies in direction (positive or negative) across cluster pairs. A t test on the unclustered jointly standardized R1 and R2 data indicates a null hypothesis (H_0) of equal u wind component means cannot be accepted.

Meridional, mean, and zonal PNA patterns are well represented in the R2 winter dataset (Figs. 3a–c). There is excellent agreement across R1 and R2 datasets regarding the zonal PNA pattern (R1 cluster 1 and R2 cluster 1), good association for the mean PNA pattern (R1 cluster 2 and R2 cluster 3), and a weaker meridional association (R1 cluster 4 and R2 cluster 5).

Although the PNA is an important part of the wintertime climate of the United States, there are other patterns that can be easily identified in the R2 dataset but whose expression is far more ambiguous in the R1 data. Continental polar north-northwest flow (Davis and Walker 1992) is identified by northerly or northwesterly winds at all levels and a strong ridge off the West Coast (Fig. 4a). The zonal-strong jets pattern (both polar and subtropical) is distinguished by the presence of both the subtropical and polar jets, northwesterly winds above 700 hPa, and evidence of a small ridge over Nevada and Utah (Fig. 4b).

e. Synoptic analysis

1) 700-hPa wind patterns

The first cluster analysis identified 700-hPa wind patterns that are persistent in time or that dominate the spatial domain. Regional air quality conditions, however, are also influenced by shorter-lived synoptic events, such as frontal passages or migratory high pressure systems (Eder et al. 2006; Lehman et al. 2004). This second cluster analysis focuses on the identification of specific 700-hPa wind patterns that reflect synoptic situations associated with daily maximum 1-h ozone values that are above or below seasonal (May–September) daily maximum 1-h ozone means. The approach described in section 2c with adaptations as described in section 2d is applied to daily ozone-season 700-hPa wind component data for the period 1985–94.

Eder et al. (1994) consider decade average within-season frequencies as well as seasonal totals. When a similar analysis was performed for the reanalysis datasets, an additional easterly flow synoptic pattern was identified. Lericos et al. (2002) describe an easterly flow pattern, "Subtropical Ridge to the North" (of Florida), and note the confounding influence of cold-core midlatitude high pressure systems, that is, the Eder et al. (1994) easterly flow pattern, which can also produce easterly flow across the southeast. An eighth key synoptic day, 12 August 1994, typical of the subtropical ridge setting, was identified and the new cluster, warm-core easterly flow (WCEF), was added to the analysis. The original easterly flow pattern was renamed cold-core easterly flow (CCEF).

R1 and R2 relative frequencies and interannual variability, estimated by the coefficient of variation (CV), are provided in Table 7. Date matching for the eight synoptic patterns establishes that, over 1530 ozone-season days, cluster assignments differ by only 8% across reanalysis datasets and average ±0.5% of the two dataset mean for individual patterns. The mean 700-hPa wind patterns associated with each key pattern date are provided in Fig. 5. An open circle indicates the location of Birmingham, Alabama, which is the focus of the Eder et al. analysis.

2) 700-hPa pattern association with observed surface ozone

Eder et al. (1994) identify synoptic patterns associated with daily maximum 1-h ozone levels greater than or less than the mean seasonal daily maximum 1-h level at Birmingham, Alabama. A similar analysis was performed with the 700-hPa CONUS reanalysis patterns, and our results were compared with those reported for Birmingham. Hourly ozone observations were obtained from the U.S. EPA's Air Quality System (AQS) [formerly the Aerometric Information Retrieval System (AIRS)] network. Additional information regarding these data can be found at http://www.epa.gov/air/data/aqsdb.html. Data for three AQS locations near Birmingham were processed as described in Eder et al.

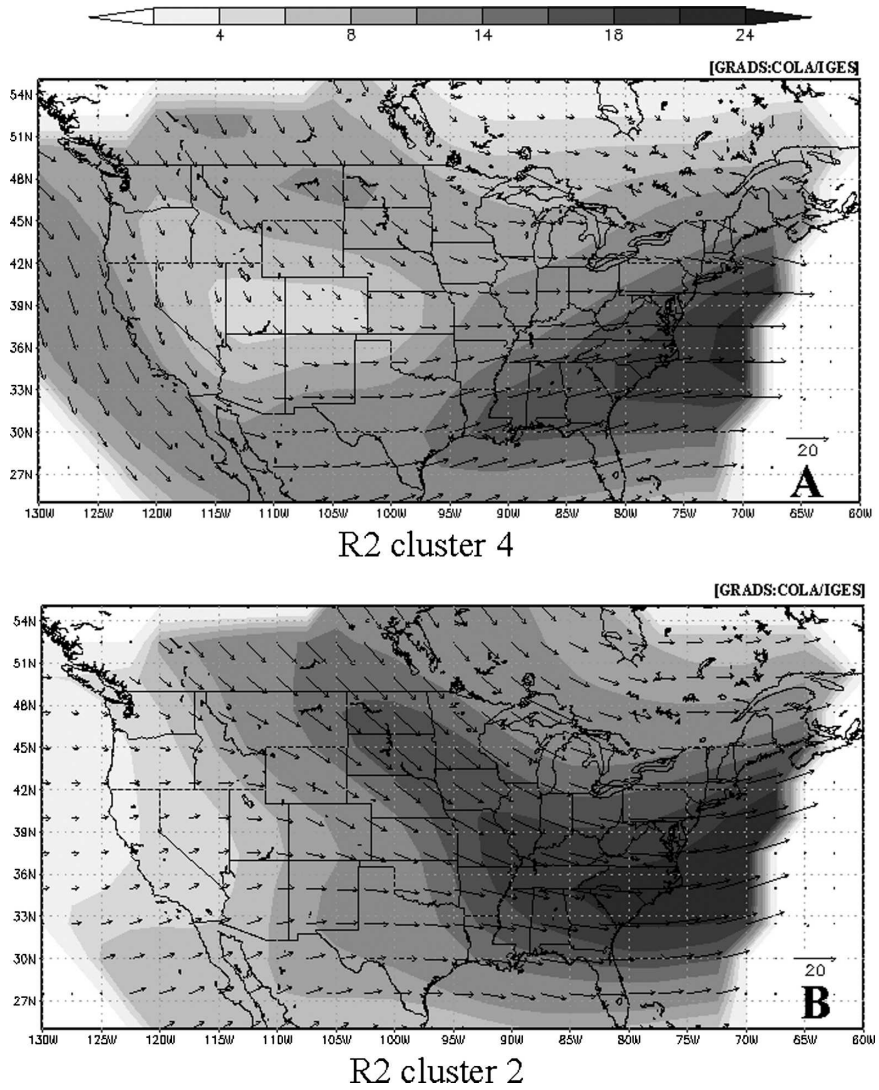

FIG. 4. Winter-season 700-hPa 1800 UTC wind speed (m s^{-1}) and direction for R2 clusters similar to Davis and Walker (1992): (a) continental polar north-northwest flow and (b) zonal-strong jets (both polar and subtropical) patterns. Arrows are scaled to velocity.

(1994) for 1990–94. The 700-hPa cluster dates determined in section 3e(1) were used to assign ozone data (153 days per season × 5 seasons) to specific synoptic clusters.

Figure 6 contains the 1990–94 ozone level anomalies associated with each synoptic pattern across the eastern United States. Anomalies are computed as the difference between the 5-yr average of daily maximum 1-h level at a site for a particular pattern and the overall ozone-season average level. AQS sites located in or around Birmingham are indicated by an open circle. Ozone anomalies for Bermuda high (BH), cold-frontal passage (CFP), transitional anticyclone (TA), and CCEF synoptic patterns agree with those reported by Eder et al. at Birmingham based on a single upper-air sounding location. Reanalysis synoptic pattern-to-ozone relationships for southwesterly flow (SWF), migratory anticyclone (MA), and warm-frontal passage (WFP) do not agree with Eder et al. (1994), and these reanalysis relationships are explored further.

Ozone levels at Birmingham for reanalysis SWF days (Fig. 6b) are slightly higher (+2.3 ppb) than the season mean, while Eder et al. (1994) report a value very much below (−16.3 ppb) the mean. The SWF reanalysis key day follows the Table 2 synoptic description, including southwesterly flow from the Gulf of Mexico into Alabama. Both R1 and R2 mean reanalysis 700-hPa flows for SWF (Fig. 5b), however, place high pressure farther west, virtually eliminating significant southwesterly flow into Alabama. The reduced cloud cover and increased radiation that would result from this shift are

TABLE 7. Synoptic cluster relative frequency (percent) results for summer ozone-season synoptic patterns; CV is the ozone-season frequency coefficient of variation.

Synoptic setting	Percentage R1 relative frequency (CV)	Percentage R2 relative frequency (CV)	Wind speed bias ($m\ s^{-1}$) (R2 − R1)
Bermuda high (BH)	16.3 (0.24)	15.2 (0.25)	+0.09
Southwesterly flow (SWF)	14.3 (0.23)	15.4 (0.24)	−0.09
Migratory anticyclone (MA)	11.1 (0.31)	11.7 (0.34)	−0.15
Cold-frontal passage (CFP)	12.1 (0.42)	11.0 (0.41)	−0.07
Transitional anticyclone (TA)	14.1 (0.34)	13.6 (0.35)	−0.15
Cold-core easterly flow (CCEF)	13.2 (0.33)	14.0 (0.34)	−0.09
Warm-frontal passage (WFP)	6.7 (0.55)	6.7 (0.60)	+0.13
Warm-core easterly flow (WCEF)	12.2 (0.25)	12.3 (0.20)	−0.13

more conducive to ozone formation than the key day (Rao et al. 2003), and so it is not surprising that ozone levels associated with this reanalysis pattern are greater than the Eder et al. (1994) analysis.

Ozone levels at Birmingham for reanalysis MA days (Fig. 6c) are lower (−4.3 ppb) than the season mean, while Eder et al. (1994) report a value above the mean (6.0 ppb). The reanalysis 700-hPa wind pattern on the MA key day is once again in agreement with the Table 2 synoptic description and indicates northerly flow over the Birmingham region. The mean 700-hPa reanalysis pattern (Fig. 5c), however, characterizes the presence of surface anticyclonic circulation as a ridge over the central United States. Northerly flow is much weaker than on the key date and is shifted to the northeast reflecting a more northerly polar jet position. Eder et al. (1994) attribute higher 1981–90 MA ozone concentrations for Birmingham to the transport of precursors from high emission areas to the north. Lower 1990–94 ozone levels in the absence of northerly transport emphasize the importance of such precursor transport to regional ozone levels in the southeastern United States.

Figure 6g indicates reanalysis WFP pattern ozone levels at Birmingham are substantially higher (+3.8 ppb above the mean) than those reported in Eder et al. (1994, −7.5 ppb below the mean). Ideally, the warm-

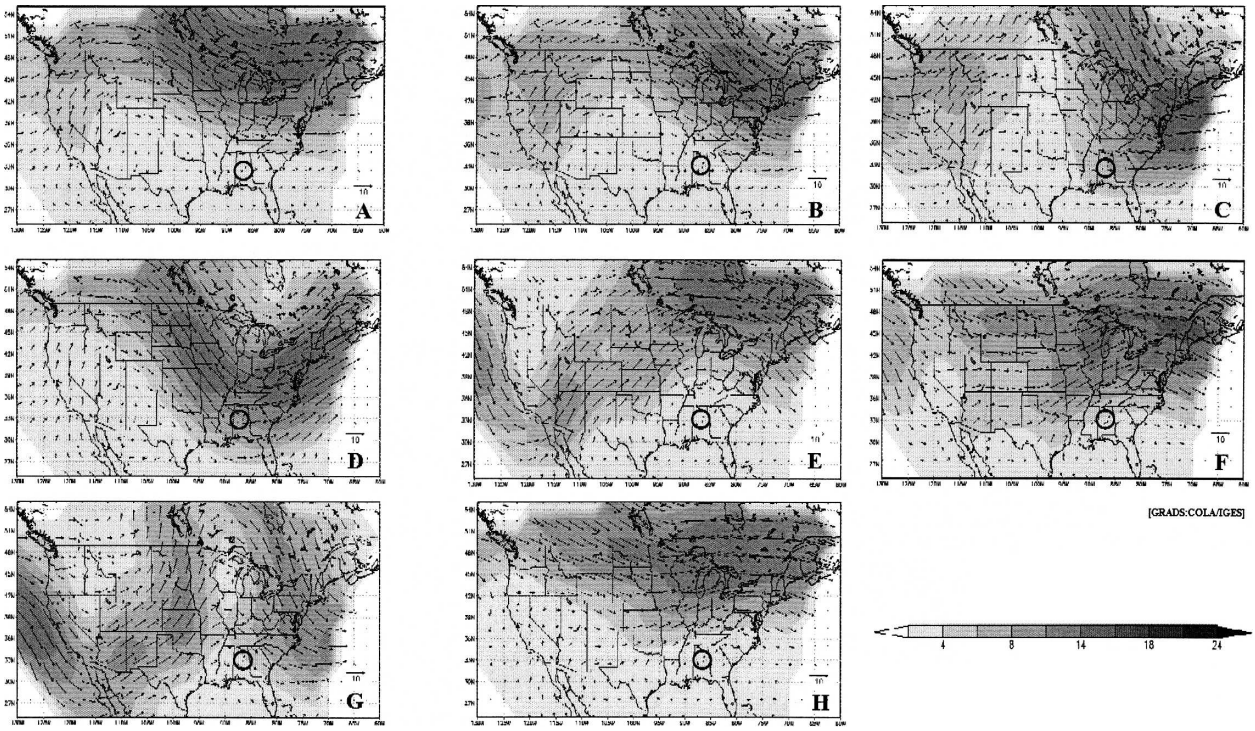

FIG. 5. Mean 700-hPa 1800 UTC wind speed ($m\ s^{-1}$) and direction patterns for (a) BH, (b) SWF, (c) MA, (d) CFP, (e) TA, (f) CCEF, (g) WFP, and (h) WCEF 700-hPa patterns. Circled stations are located in the vicinity of Birmingham, AL. Arrows are scaled to velocity.

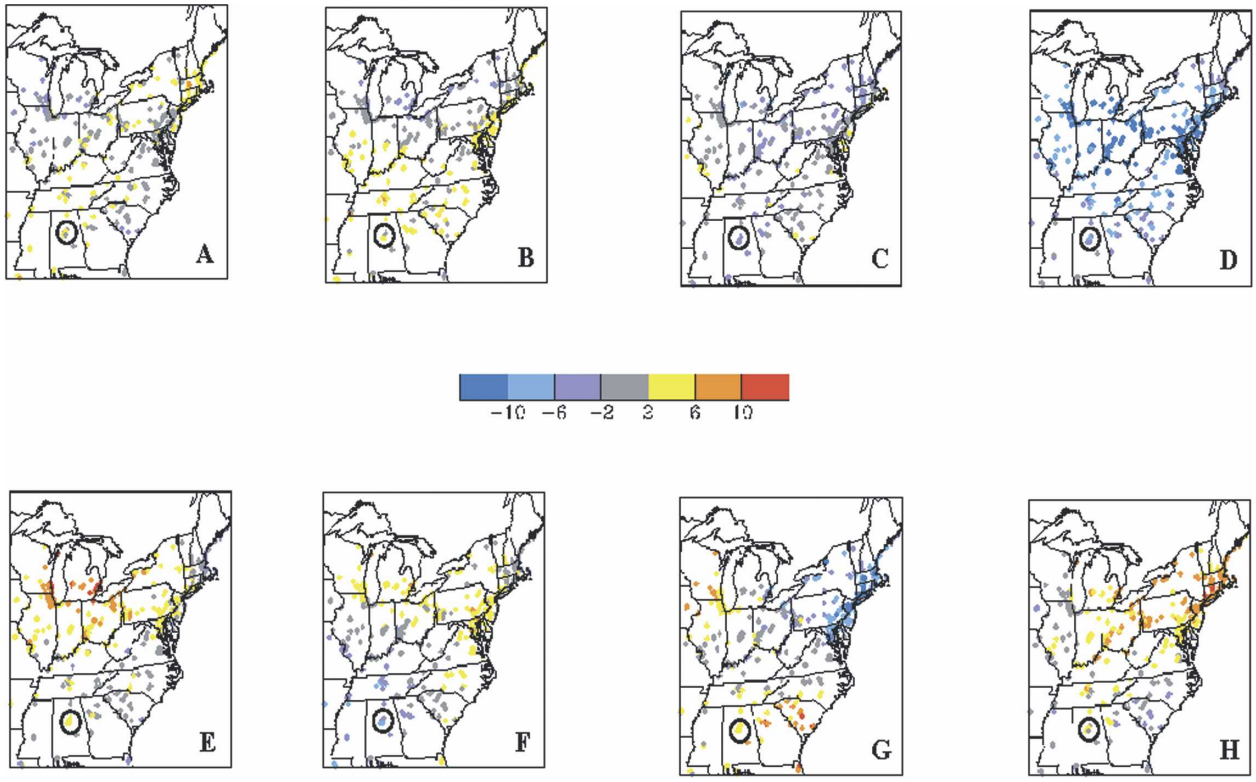

FIG. 6. Mean daily maximum 1-h ozone anomalies (ppb) computed using reanalysis dates for (a) BH, (b) SWF, (c) MA, (d) CFP, (e) TA, (f) CCEF, (g) WFP, and (h) WCEF 700-hPa patterns. An anomaly is computed as the difference between the pattern mean ozone level and the mean over all ozone-season days. Circled stations are located in the vicinity of Birmingham, AL.

frontal passage pattern should be associated with strong southerly flow off the Gulf with higher humidity, clouds, and lower radiation, but the mean 700-hPa reanalysis pattern for WFP (Fig. 5g) indicates the presence of an omega (blocking) high (Bluestein 1993). Regions near the upper-level cyclones tend to experience a persistent combination of precipitation and cool temperatures, while those near the upper-level anticyclone tend to experience drought conditions: low moisture and cloud cover and higher temperatures and radiation. Birmingham lies under anticyclonic influence of the 1985–94 mean pattern, and a +2–6 ppb mean anomaly is estimated. A similar feature is present at 500 hPa on the key day but is not apparent at the surface (http://docs.lib.noaa.gov/rescue/dwm/data_rescue_daily_weather_maps.html). Small changes in eastward or westward position of this blocking feature will significantly impact ozone levels at a point location. The interannual CV of reanalysis WFP pattern frequency is much larger than the other synoptic patterns and it has the lowest decadal frequency. Taken together, these results suggest that, while this can be an important pattern in terms of producing high local-to-regional ozone levels, its contribution to overall decadal ozone concentrations is highly variable.

4. Discussion

R1 and R2 reanalysis 700-hPa wind flow differences have been identified that reflect 1) minor (nonsignificant) dataset differences and uncertainty related to an absence of distinct, nonoverlapping clusters; 2) larger season-specific differences that can be traced back to error correction, algorithm changes, or parameter modification; and 3) systematic model biases. Small wind flow differences can occur from numerical or computational instabilities or minor parameter and algorithmic changes. They can also produce overlapping clusters that can further complicate the identification of distinct patterns. Comparison with previous analyses, however, reveals surprisingly consistent pattern characterizations in spite of significant cluster overlap. Larger, season-specific pattern differences were noted during summer (section 3b) and winter (section 3d) seasons. During the summer season, *t*-test results for the unclustered data suggest the presence of significant wind component differences across datasets. Although its source is not apparent from the Ward's analysis, 20-yr mean August sea level pressure data suggest the R1 reanalysis shifts dominant summertime anticyclonic circulation slightly westward of its R2 position (Kalnay et al. 1996;

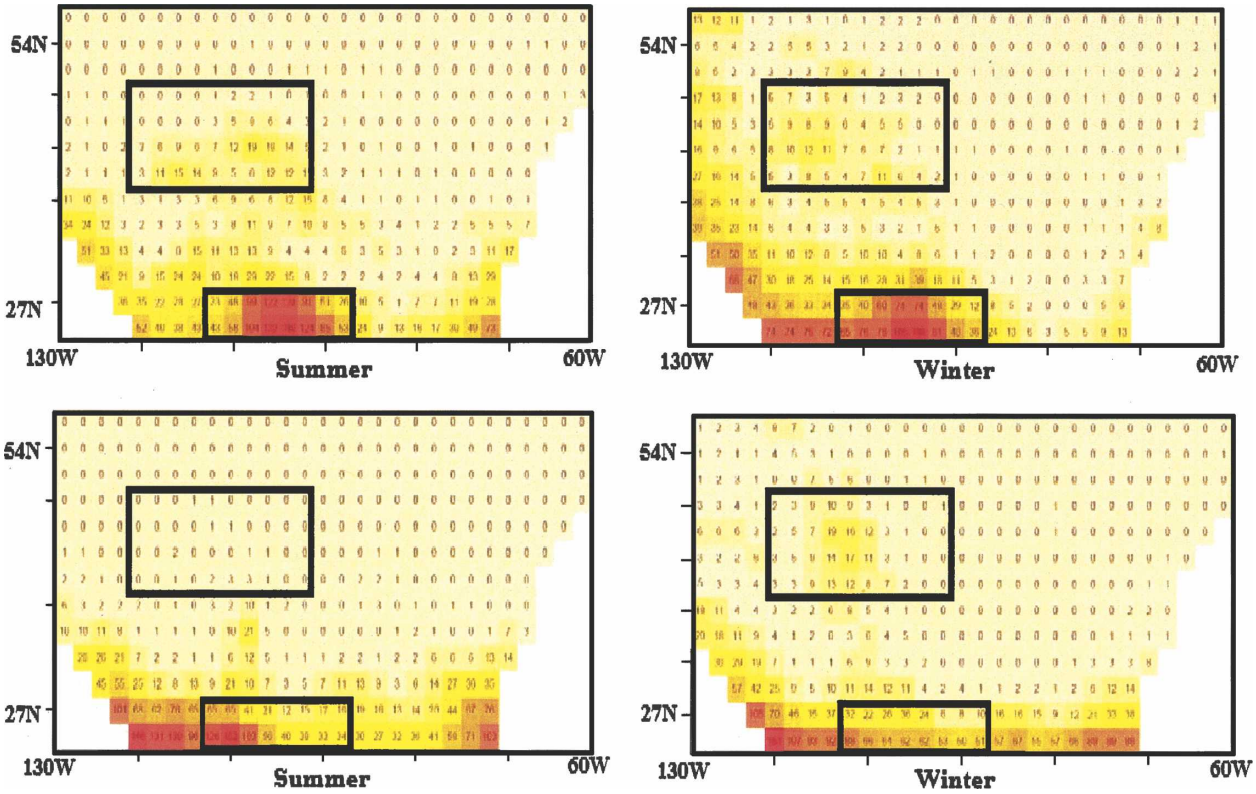

FIG. 7. Number of days for which R1–R2 differences are extreme for (a) summer-season u component, (b) winter-season u component, (c) summer-season v component, and (d) winter-season v component. Boxed areas correspond to boxed areas in Fig. 8.

Kanamitsu et al. 2002; Rutledge et al. 2006). Additional detail is provided by Table 7, which indicates more frequent R1 westward extension of the Bermuda high circulation than does R2. If all Bermuda high related patterns are considered, that is, sum of BH and SWF, this distinction disappears. It appears that the presence of overlapping clusters does not prevent the Ward's analysis from correctly identifying Bermuda high similarities but masks the detection of more subtle differences between normal and westward extension of the circulation. Major changes in the R2 treatment of surface moisture and convective and boundary layer parameterization reportedly reduce R2 precipitation over the southern United States (Kanamitsu et al. 2002) and could be associated with R1 and R2 simulation of Bermuda high circulation mean location and interannual variability.

PNA characterization differences noted during the winter season most likely derive from two R2 modifications to the R1 model. First, R2 corrects a known R1 error, that is, snow cover corresponding to 1973 was used during every R1 model year between 1974 and 1994. The effect of this change should be most easily noted near the surface over regions where the snow cover mask normally varies (Kistler et al. 2001). A second change is improved high latitude precipitation, surface air temperature, and surface fluxes in R2 as a result of implementation of a "spectral snow" correction. These differences appear to have the greatest impact on mean and meridional PNA pattern simulation (e.g., Fig. 3).

Systematic differences that persist throughout the entire year are also present in the 700-hPa reanalysis data (e.g., Fig. 7). The plotted values are the number of days for which grid cell R1 and R2 component differences fall above the upper 0.5% tail (>0.64 standard deviations in summer, >0.57 standard deviations in winter) or below the lower 0.5% tail (<−1.01 standard deviations in summer, <−0.62 standard deviations in winter) of the jointly standardized R1 and R2 data distribution (see section 2e). Orange to red colors indicate a higher frequency of occurrence. Differences over the Rocky Mountains and northern Mexico most likely reflect differences in R1 (mean) and R2 (smoothed) orography (Fig. 8). The largest u and v component differences are in the proximity of greatest elevation gradient (boxed areas). Overall, R1 elevations are greater than R2. The effect of these differences extends westward to the Pacific coast during the wintertime, when mean 700-hPa heights are at their lowest point and may not be located above the planetary boundary layer. This westward extension could also reflect snow mask differences dis-

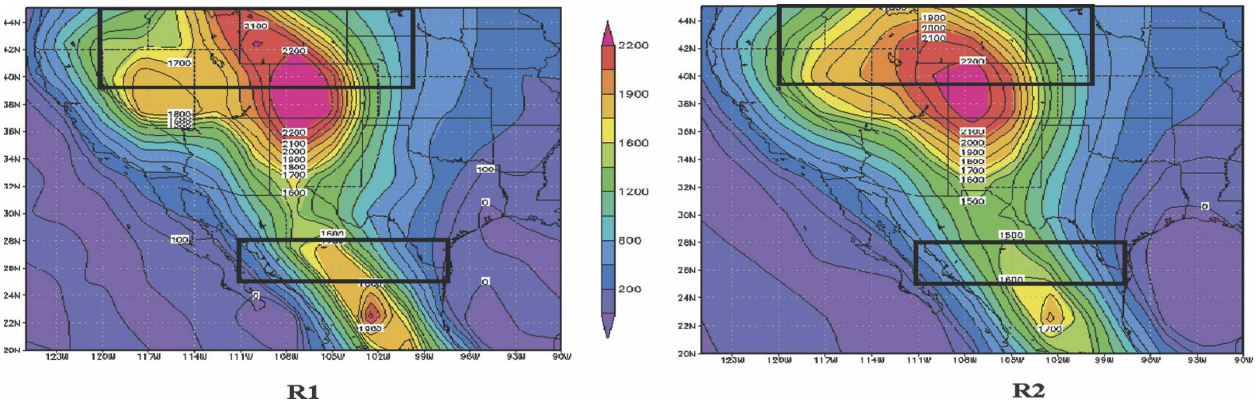

FIG. 8. NCAR-R1 and AMIP-R2 elevation maps (m). Boxed areas highlight regions of significant elevation gradient difference (Kalnay et al. 1996; Kanamitsu et al. 2002; Rutledge et al. 2006).

cussed previously and R2 modified treatment of snowpack (Kanamitsu et al. 2002). Systematic overocean differences in Fig. 7 most likely derive from two model differences. First, the R2 model modifies R1 overocean radiation fluxes by, among other changes, reducing oceanic albedo by ~50%. Kanamitsu et al. (2002) also note that some R2 differences exist in the upper-air height and temperature analysis over Northern Hemisphere oceans, where most of the observations are from satellites. Relatively minor changes in the location of strong Pacific and, to a lesser extent, southern Atlantic features could easily produce the large Fig. 7 wind component differences. Both these factors, that is, uncertainty in treatment/location of large systems in the eastern Pacific and modified flow regimes over the mountainous western United States, could be involved in the PNA simulation differences discussed previously. Analyses involving R2 PNA simulation are rare, but additional factors contributing to R1 and R2 PNA simulation differences may be suggested by diagnostic R1 studies, such as Feldstein (2002).

5. Summary

Statistical clustering methods have been used to identify and to describe well-documented CONUS differences between 700-hPa wind component data sampled from two gridded reanalysis datasets. Previously published regional analyses have been expanded to the CONUS to identify large-scale synoptic patterns associated with above and below season-average observed ozone levels. Small numerical differences, larger season-specific reanalysis differences, and systematic reanalysis model biases were identified. Model differences that influence the character and location of the Bermuda high circulation are reflected in the ozone-season synoptic analysis but are masked by imperfect cluster definitions in the dominant pattern results. Systematic differences, most likely related to changes in the treatment of orography, overocean data assimilation, and overocean radiation flux, have limited geographic extent but may influence the simulation of the PNA in the western United States significantly.

The strongest reanalysis agreement regarding dominant 700-hPa patterns occurs during the spring and autumn seasons. Summer patterns show consistent levels of agreement, but there is significant uncertainty regarding our ability to clearly associate specific R1 patterns with those in R2. This is not surprising since summer patterns tend to change slowly through time, making the identification of distinct clusters more difficult. The principle source of winter-season wind flow pattern differences lies in the characterization of the mean and meridional phases of the PNA. While it can be debated that one reanalysis model is "better" than another at simulating such features, both reanalyses continue to see wide use. Comparison across these two datasets should be a reasonable representation of the breadth of gridded reanalysis models in current use.

While strongly overlapping clusters preclude the quantitative application of Ward's analysis cluster frequencies, the alternative distinct synoptic clusters offer ample opportunity for such comparisons. Synoptic analysis results, coupled with AQS surface ozone observations for the eastern United States, indicate that MA and CFP patterns are associated with daily maximum 1-h ozone levels during the ozone season (May–September) that are below season-mean values. WFP patterns are also associated with near or below season-mean ozone levels throughout the eastern United States, with the exception of the southeastern coastal plain and west of Lake Michigan. Above season-mean ozone levels in the Midwest, mid-Atlantic, and, to a limited degree, New England are associated with TA, CCEF, and WCEF patterns. SWF patterns are associ-

ated with above season-average ozone levels throughout the south, while BH patterns limit above season-mean values to a band paralleling the Smoky and Appalachian Mountains and New England. These findings are in agreement with previously published studies regarding the dependence of surface ozone levels on meteorological conditions (Eder et al. 1994; Lehman et al. 2004; Rao et al. 2003).

Acknowledgments. The authors thank Dr. Brian Eder, NOAA/ARL/AMD in Research Triangle Park, North Carolina, and Dr. Richard Cohn, Constella Health Sciences Team, Constella Group, Inc., Durham, North Carolina, for their input regarding the adaptation of their spatial analysis research. We also thank Dr. Christian Hogrefe, Atmospheric Sciences Research Center, University at Albany, for his assistance in preparing the ozone data for synoptic analysis. The research presented here was performed under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to the NOAA Air Quality and Global Climate Programs. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

REFERENCES

Anderberg, M. R., 1973: *Cluster Analysis for Applications*. Probability and Mathematical Statistics Series, Vol. 19, Academic Press, 359 pp.

Bluestein, H. B., 1993: Observations and theory of weather systems. *Synoptic-Dynamic Meteorology in Midlatitudes,* Vol. 2, Oxford University Press, 79–82.

Byun, D. W., and K. L. Schere, 2006: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.,* **59,** 51–77.

Christensen, N. S., A. W. Wood, N. Voisin, D. P. Lettenmaier, and R. N. Palmer, 2004: The effects of climate change on the hydrology and water resources of the Colorado River basin. *Climatic Change,* **62,** 337–363.

Cohn, R. D., B. K. Eder, S. K. Leduc, and R. L. Dennis, 2001: Development of an aggregation and episode selection scheme to support the Models-3 Community Multiscale Air Quality Model. *J. Appl. Meteor.,* **40,** 210–228.

Davis, R. E., and D. R. Walker, 1992: An upper-air synoptic climatology of the western United States. *J. Climate,* **5,** 1449–1467.

Eder, B. K., J. M. Davis, and P. Bloomfield, 1994: An automated classification scheme designed to better elucidate the dependence of ozone on meteorology. *J. Appl. Meteor.,* **33,** 1182–1199.

——, D. Kang, R. Mathur, S. Yu, and K. L. Schere, 2006: An operational evaluation of the Eta-CMAQ air quality forecast model. *Atmos. Environ.,* **40,** 4894–4905.

Feldstein, S. B., 2002: Fundamental mechanisms of the growth and decay of the PNA teleconnection pattern. *Quart. J. Roy. Meteor. Soc.,* **128,** 775–796.

Hogrefe, C., and Coauthors, 2004: Simulating changes in regional air pollution over the eastern United States due to changes in global and regional climate and emissions. *J. Geophys. Res.,* **109,** D22301, doi:10.1029/2004JD004690.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83,** 1631–1643.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247–267.

Kunkel, K. E., and X.-Z. Liang, 2005: GCM simulations of the climate in the central United States. *J. Climate,* **18,** 1016–1031.

Leathers, D. J., B. Yarnal, and M. A. Palecki, 1991: The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. *J. Climate,* **4,** 517–528.

Lehman, J., K. Swinton, S. Bortnick, C. Hamilton, E. Baldridge, B. Eder, and B. Cox, 2004: Spatio-temporal characterization of tropospheric ozone across the eastern United States. *Atmos. Environ.,* **38,** 4357–4369.

Lericos, T. P., H. E. Fuelberg, A. I. Watson, and R. L. Holle, 2002: Warm season lightning distributions over the Florida Peninsula as related to synoptic patterns. *Wea. Forecasting,* **17,** 83–98.

Leung, L. R., and W. I. Gustafson Jr., 2005: Potential regional climate change and implications to U.S. air quality. *Geophys. Res. Lett.,* **32,** L16711, doi:10.1029/2005GL022911.

——, Q. Yun, X. Bian, W. M. Washington, J. Han, and J. D. Roads, 2004: Mid-century ensemble regional climate change scenarios for the western United States. *Climatic Change,* **62,** 75–113.

Liang, X.-Z., L. Li, K. E. Kunkel, M. Ting, and J. X. L. Wang, 2004: Regional climate model simulation of U.S. precipitation during 1982–2002. Part I: Annual cycle. *J. Climate,* **17,** 3510–3529.

MacQueen, J. B., 1967: Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press, 281–297.

Mickley, L. J., D. J. Jacob, B. D. Field, and D. Rind, 2004: Effects of future climate change on regional air pollution episodes in the United States. *Geophys. Res. Lett.,* **31,** L24103, doi:10.1029/2004GL021216.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747–1763.

Rao, S. T., J.-Y. Ku, S. Berman, K. Zhang, and H. Mao, 2003: Summertime characteristics of the atmospheric boundary layer and relationships to ozone levels over the eastern United States. *Pure Appl. Geophys.,* **160,** 21–55.

Rutledge, G. K., J. Alpert, and W. Ebisuzaki, 2006: NOMADS: A climate and weather model archive at the National Oceanic and Atmospheric Administration. *Bull. Amer. Meteor. Soc.,* **87,** 327–341.

SAS Institute Inc., 1999: SAS OnlineDoc, version eight. [Available online at http://v8doc.sas.com/sashtml.]

Ward, J. H., 1963: Hierarchical groupings to optimize an objective function. *J. Amer. Stat. Assoc.,* **58,** 236–244.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. International Geophysics Series, Vol. 59, Academic Press, 467 pp.