Transport in the subtropical lowermost stratosphere during the Cirrus Regional Study of Tropical Anvils and Cirrus Layers–Florida Area Cirrus Experiment

Jasna V. Pittman,[1,2] Elliot M. Weinstock,[1] Robert J. Oglesby,[3] David S. Sayres,[1] Jessica B. Smith,[1] James G. Anderson,[1] Owen R. Cooper,[4,5] Steven C. Wofsy,[1] Irene Xueref,[1,6] Cristoph Gerbig,[1,7] Bruce C. Daube,[1] Erik C. Richard,[8] Brian A. Ridley,[9] Andrew J. Weinheimer,[9] Max Loewenstein,[10] Hans-Jurg Jost,[11,12] Jimena P. Lopez,[11] Michael J. Mahoney,[13] Thomas L. Thompson,[8] William W. Hargrove,[14] and Forrest M. Hoffman[14]

Received 28 July 2006; revised 17 October 2006; accepted 21 November 2006; published 20 April 2007.

[1] We use in situ measurements of water vapor (H_2O), ozone (O_3), carbon dioxide (CO_2), carbon monoxide (CO), nitric oxide (NO), and total reactive nitrogen (NO_y) obtained during the CRYSTAL-FACE campaign in July 2002 to study summertime transport in the subtropical lowermost stratosphere. We use an objective methodology to distinguish the latitudinal origin of the sampled air masses despite the influence of convection, and we calculate backward trajectories to elucidate their recent geographical history. The methodology consists of exploring the statistical behavior of the data by performing multivariate clustering and agglomerative hierarchical clustering calculations and projecting cluster groups onto principal component space to identify air masses of like composition and hence presumed origin. The statistically derived cluster groups are then examined in physical space using tracer-tracer correlation plots. Interpretation of the principal component analysis suggests that the variability in the data is accounted for primarily by the mean age of air in the stratosphere, followed by the age of the convective influence, and last by the extent of convective influence, potentially related to the latitude of convective injection (Dessler and Sherwood, 2004). We find that high-latitude stratospheric air is the dominant source region during the beginning of the campaign while tropical air is the dominant source region during the rest of the campaign. Influence of convection from both local and nonlocal events is frequently observed. The identification of air mass origin is confirmed with backward trajectories, and the behavior of the trajectories is associated with the North American monsoon circulation.

Citation: Pittman, J. V., et al. (2007), Transport in the subtropical lowermost stratosphere during the Cirrus Regional Study of Tropical Anvils and Cirrus Layers–Florida Area Cirrus Experiment, *J. Geophys. Res.*, *112*, D08304, doi:10.1029/2006JD007851.

1. Introduction

[2] In order to predict the response of the global climate system to thermal and chemical changes resulting from forcing by infrared-active species, it is necessary to understand the mechanisms that transport air masses between the troposphere and the stratosphere year-round. Transport of chemical constituents plays a crucial role in explaining issues such as UV dosage at the surface, distribution of greenhouse gases like water vapor and carbon dioxide in the upper troposphere and lower stratosphere (UT/LS), and

[1]Departments of Earth and Planetary Sciences and of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA.
[2]Now at NASA Marshall Space Flight Center, Huntsville, Alabama, USA.
[3]Department of Geosciences, University of Nebraska, Lincoln, Nebraska, USA.
[4]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.
[5]Also at Earth System Research Laboratory, NOAA, Boulder, Colorado, USA.
[6]Now at Laboratoire des Sciences du Climat et de l'Environment, Commissariat l'Energie Atomique, Gif-Sur-Yvette, France.
[7]Now at Max Planck Institute for Biogeochemistry, Jena, Germany.

0148-0227/07/2006JD007851

[8]Earth System Research Laboratory, NOAA, Boulder, Colorado, USA.
[9]Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, Colorado, USA.
[10]NASA Ames Research Center, Moffett Field, California, USA.
[11]Bay Area Environmental Research Institute, Sonoma, California, USA.
[12]Now at NovaWave Technologies, Redwood City, California, USA.
[13]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.
[14]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

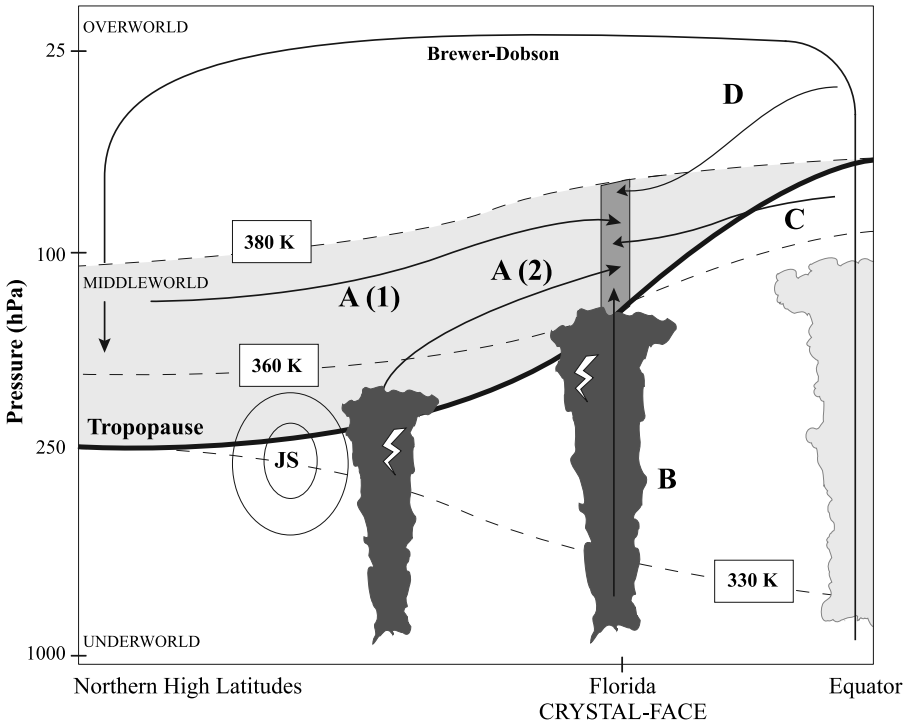

Figure 1. Possible transport pathways of air into the lowermost stratosphere over Florida (dark gray box): equatorward transport of stratospheric air (A(1)), equatorward transport of convective air injected into the stratosphere at higher latitudes (A(2)), local convection (B), poleward transport of air from the tropical transition layer (C), and poleward transport of air that descended diabatically from the tropical stratosphere (D). JS represents the approximate location of the polar jet stream over North America during July 2002. The subtropical jet stream was not observed during this time of the year. The light gray shaded area represents the lowermost stratosphere.

heterogeneous ozone loss in the tropopause region. The summertime circulation, particularly in the stratosphere, has been studied less extensively than the other seasons, because it has been regarded as quiescent because of the weakening of the large-scale circulation driven by thermal forcing and wave activity [*Plumb*, 2002]. Several studies, however, have shown that this season can be important, because deep convection over land can have a direct and significant impact on the chemical composition of the extratropical lower stratosphere [*Jost et al.*, 2004; *Dessler and Sherwood*, 2004].

[3] In this study, we use an Eulerian approach to qualitatively explore transport pathways in the subtropical lowermost stratosphere during the summer by (1) using a novel combination of nonparametric and parametric techniques along with tracer-tracer correlations to identify the origin of sampled air masses and (2) backward trajectory calculations to elucidate the responsible transport pathways. For this purpose, we use in situ measurements of H_2O, O_3, CO_2, CO, NO, and NO_y, obtained aboard NASA's WB-57 aircraft during the Cirrus Regional Study of Tropical Anvils and Cirrus Layers–Florida Area Cirrus Experiment (CRYSTAL-FACE) that was based out of Key West, Florida during July 2002.

[4] Earlier observations of H_2O vapor [*Brewer*, 1949] and O_3 [*Dobson*, 1956] in the stratosphere revealed the existence of a large-scale meridional circulation that couples the troposphere and the stratosphere. In this Brewer-Dobson circulation, tropospheric air ascends into the stratosphere primarily in the tropics, travels poleward, and descends into the troposphere at high latitudes, where it can eventually return to the tropics. Isentropic surfaces, or surfaces of constant potential temperature, can be used to divide the troposphere and the stratosphere into three regions: the overworld, the middleworld, and the underworld [*Hoskins*, 1991]. These regions are delineated in Figure 1. In the overworld and underworld, air is considered to be entirely in the stratosphere and troposphere, respectively. In the middleworld, air is considered to be in the troposphere at low latitudes and in the stratosphere at high latitudes. The stratospheric side of the middleworld is known as the lowermost stratosphere, which is the focal region in this study. The lowermost stratosphere is bounded from below by the local tropopause, and from above by the tropical tropopause. We use the long-standing definition of the tropical tropopause to be the 380 K isentrope. Exchange of air between the lowermost stratosphere and the troposphere is controlled by smaller-scale transient processes [*Holton et al.*, 1995]. These short-lived processes that can develop in the extratropics include blocking highs, cutoff lows, and tropopause folds.

[5] Figure 1 depicts possible transport pathways that can bring air to the subtropical lowermost stratosphere over southern Florida represented by the dark gray box. Pathway

A corresponds to equatorward transport of high-latitude stratospheric air (HLS). This pathway can carry either pure stratospheric air that descended diabatically from the overworld at high latitudes (pathway A(1)), or a mixture with tropospheric air that was injected into the stratosphere via middle- or high-latitude convection (pathway A(2)). Pathway B corresponds to direct penetration of local convection. Pathway C corresponds to poleward isentropic transport of air from the Tropical Tropopause Layer (TTL) across the steeply sloped subtropical tropopause. In this study, we adopt the definition of the TTL presented by *Gettelman and Forster* [2002], which is the region bounded from below by the level of neutral buoyancy (340 to 350 K) and from above by the cold-point tropopause (380 to 390 K). Pathway D corresponds to isentropic transport from the tropical lower stratosphere (TLS) followed by diabatic descent into the subtropical lowermost stratosphere.

[6] Several studies have reported evidence of each one of these transport pathways in the UT/LS at various locations and times using measurements of different chemical tracers. For instance, analysis of ozone profiles during CRYSTAL-FACE showed layers of higher ozone in the subtropical lowermost stratosphere resulting from large-scale equatorward transport of high-latitude air into the subtropics (pathway A(1)) [*Richard et al.*, 2003]. Summertime observations showed tropospheric air reaching into the lowermost stratosphere via convection (pathway A(2) and B) [*Poulida et al.*, 1996; *Vaughan and Timmis*, 1998; *Jost et al.*, 2004; *Ray et al.*, 2004]. Late spring in situ aircraft measurements [*Dessler et al.*, 1995; *Hintsa et al.*, 1998] as well as modeling studies [*Chen*, 1995; *Dethof et al.*, 2000; *Stohl et al.*, 2003] have provided evidence for poleward isentropic transport of air from the TTL (pathway C). Last, a combination of satellite and aircraft measurements of ozone showed that diabatic descent into the lowermost stratosphere (pathway D) can happen during the summer, though it is strongest during late winter and early spring [*Prados et al.*, 2003].

[7] The methodology followed in this study consists of three stages. The first stage consists of objectively clustering the data set and identifying the dominant modes of variability by performing a statistical analysis in three different steps. In the first step, we use a nonparametric, nonhierarchical clustering technique to construct six dimensional clusters based on in situ measurements of H_2O, O_3, CO_2, CO, NO, and NO_y obtained in the subtropical lowermost stratosphere during CRYSTAL-FACE. In the second step, we perform nonparametric agglomerative hierarchical clustering to group the clusters based on similarities to each other in their chemical characteristics. In the third step, we do a parametric Principal Component Analysis (PCA) in which we project the data set onto its first three eigenvectors. We plot the results, color coded by group assignment, in order to examine the cluster groups in a space composed of the most relevant modes of variability in the data set. This three-step statistical analysis has been applied previously to identify distinct tropical convective profiles using the vertical structure information obtained from various measurements from the Tropical Rainfall Measuring Mission (TRMM) [*Boccippio et al.*, 2005]. The second stage consists of identifying the source regions by comparing the mean mixing ratios of the six chemical tracers in each cluster with reference values for each source region and demonstrating consistency in air mass origin between the statistically derived results and the physically based tracer-tracer correlation plots. The third and final stage consists of calculating backward trajectories in order to investigate the recent geographical history of the air masses ultimately sampled in the subtropical lowermost stratosphere. The trajectories are also used to confirm the identification of source regions obtained in the second stage. We will show that the monsoon circulation over North America is critical in order to explain not only the origin but also the elevated humidity observed of the air sampled in the lowermost stratosphere during CRYSTAL-FACE. As *Dunkerton* [1995] pointed out, monsoon circulation can be an important driver of horizontal and vertical transport of air reaching the UT/LS region when atmospheric wave activity is otherwise weak. This study strongly supports that conclusion.

[8] The paper is structured as follows. Section 2 explains the usefulness of the chemical tracers chosen and describes the data sets used in this study. Section 3 describes the statistical techniques, tracer-tracer correlations, and trajectory model. Section 4 discusses the results obtained from applying the methodology described in section 3. Section 5 summarizes the major findings of this study.

2. Data

2.1. Chemical Tracers

[9] We use six different chemical tracers in this study, namely H_2O, O_3, CO_2, CO, NO, and NO_y. These tracers are chosen because they can provide the following information and discrimination in the lowermost stratosphere, our region of interest. H_2O can distinguish wet, convectively influenced air from otherwise very dry stratospheric air. O_3, while having a significant gradient between the troposphere and the stratosphere, can also provide information on the mean age of stratospheric air: older stratospheric air has higher mixing ratios than younger stratospheric air. CO_2 can also provide valuable information on the age of air, since its mixing ratio is affected by both the positive secular trend observed in the atmosphere and its seasonal cycle. CO, when found at high concentrations, is a good indicator of the presence of recently injected tropospheric air into the lowermost stratosphere, since it is produced primarily in the troposphere by human activity such as biomass burning. NO is the most unequivocal evidence of recently injected convective air when found at higher mixing ratios than background levels, since its main source in the UT/LS is lightning. Last, NO_y, which represents the sum of all reactive nitrogen species ($NO_y = NO + NO_2 + NO_3 + 2N_2O_5 + HNO_3$ + peroxyacetyl nitrate (PAN) + ...), can provide information on the altitudinal and latitudinal origin of air, as the tracer increases with increasing altitude and latitude, and along with NO, it can provide information on the age of the convective intrusion into the stratosphere. In this study, we use NO_y measurements in the gas phase only. The use of all these tracers together will ultimately allow us to distinguish a larger variety of air masses more clearly by exploiting the unique information that each tracer can provide.

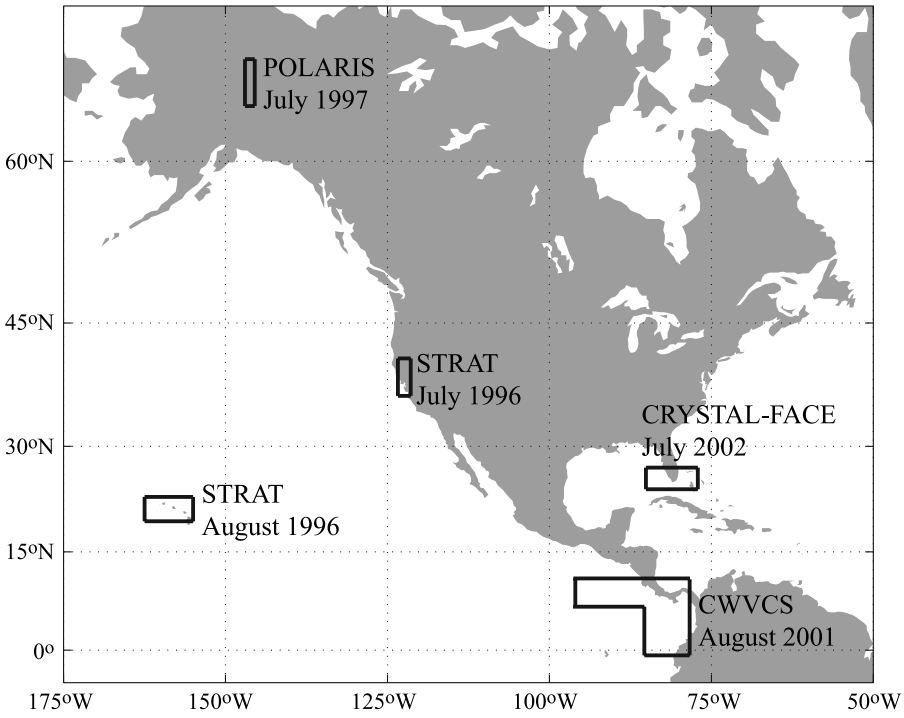

Figure 2. Geographical location of all aircraft campaigns used in this study.

[10] In addition to the above information obtained from each chemical tracer, the chemical lifetimes of these tracers allow us to separate observed changes in mixing ratios due to chemistry from changes due to transport. All tracers except NO have chemical lifetimes ranging from several months to several years in the lower stratosphere. NO_x (= NO + NO_2) has a chemical lifetime of about 1 week in the UT/LS during the summer [*Ridley et al.*, 1996]. Once NO is produced by lightning, the partitioning of NO_x will favor NO during the day because of photolysis of NO_2. Therefore observations of high NO in the UT/LS during the day would be indicative of the presence of tropospheric air that was lofted by strong convection. A recent study showed that large enhancements in NO (up to 50 times larger than background values) were indeed observed during CRYSTAL-FACE and that they were associated with lightning activity from local thunderstorms and possibly transport from the boundary layer [*Ridley et al.*, 2004]. For the remaining tracers, observed changes in mixing ratios during the campaign are more likely the result of changes in transport, which can happen in timescales shorter than the tracers' chemical lifetimes.

2.2. CRYSTAL-FACE

[11] CRYSTAL-FACE was based out of Key West, Florida and included 12 flights over 27 days from 3 to 29 July 2002, excluding the two ferry flights between Houston, TX and Key West, FL [*Jensen et al.*, 2004]. We restrict our analysis to data collected in the lowermost stratosphere over the southern Florida region between 24°N and 27°N and 77°W and 85°W as shown in Figure 2. We exclude the flight of 29 July from this analysis because of the existence of two tropopauses: one at approximately 12 km, which was over 2 km below the average location of the local tropopause and another at approximately 16 km, which was coincident with the location of the 380-K isentrope during this campaign. The variability in the location of the tropopause throughout this flight made it difficult to specify the boundaries for the lowermost stratosphere.

[12] The tracer measurements used are: H_2O measured by Lyman-α photofragment fluorescence using the Harvard hygrometer, with a ±5% uncertainty [*Weinstock et al.*, 1994], O_3 measured by dual-beam UV absorption, with a ±3% uncertainty [*Proffitt and McLaughlin*, 1983], NO and NO_y measured by catalytical reduction of NO_y to NO followed by detection of NO through the chemiluminescence reaction with O_3, with an uncertainty of ±(7% + 15 pptv) for NO and ±(9% + 15 pptv) for NO_y [*Ridley et al.*, 1994], CO measured by mid-IR absorption using a tunable diode laser, with a ±3% uncertainty [*Loewenstein et al.*, 2002], and CO_2 measured by nondispersive IR absorption, with a ±0.1 ppmv uncertainty [*Daube et al.*, 2002]. All these instruments flew on NASA's WB-57F. The tracer data analyzed in this study are averaged over 10 s giving a horizontal resolution of 1.8 km on average along the WB-57's flight track. Only observations that reported measurements for all of the six chemical tracers are considered. Because of the large spatial and temporal range of values measured for these tracers during the campaign, none of the above measurement uncertainties are significant enough to compromise the conclusions of this study.

[13] We use the location of the local tropopause reported by the Microwave Temperature Profiler (MTP) instrument [*Denning et al.*, 1989]. This instrument uses the World Meteorological Organization (WMO) definition of the tropopause as the lowest level at which the lapse rate ($-dT/dz$) decreases to 2 K/km or less, and the average lapse rate within the next 2 km does not exceed 2 K/km. Location of

Table 1. Location of Measurements and References for Tracer Values Used to Represent Each Source Region Considered in This Study[a]

Source Region (Pathway)	Tracer(s)	Isentropes, K, or Lat-Lon, deg	References
HLS (A)	all	360–380	POLARIS
Convection (B)	O_3, CO	345–360	*Poulida et al.* [1996]
Convection (B)	O_3, NO_y, NO	350–370	*Ridley et al.* [2004]
Convection (B)	CO_2	13–42°N, 59–155°W	surface measurements
TTL (C)	H_2O	0–15°N	NCEP/NCAR reanalysis
TTL (C)	H_2O, O_3	360–380	STRAT, CWVCS
TTL (C)	NO_y, NO, CO	360–380	STRAT, CRYSTAL-FACE
TTL (C)	CO_2	360–380	surface measurements, *Boering et al.* [1996]
TLS (D)	H_2O, O_3	380–420	STRAT, CWVCS
TLS (D)	NO_y, NO	380–420	STRAT, CRYSTAL-FACE
TLS (D)	CO	380–420	STRAT
TLS (D)	CO_2	380–420	surface measurements, *Boering et al.* [1996]

[a]The STRAT profiles used are from the flights based out of Hawaii only. The CRYSTAL-FACE profiles used are from the 20020707 and 20020709 flights south of 15°N. HLS stands for high-latitude stratosphere, TTL stands for tropical tropopause layer, and TLS stands for tropical lower stratosphere. Pathways are shown in Figure 1.

the local tropopause using in situ measurements of pressure and temperature made by the pressure-temperature instrument, which has uncertainties of 0.1 hPa and 0.5 K [*Thompson and Rosenlof*, 2003], and application of the WMO definition yielded results that agreed well with those reported by MTP.

[14] We restrict this analysis to clear-air data sampled in the lowermost stratosphere. Using calculated saturation water vapor mixing ratios obtained from in situ pressure and temperature measurements and in situ ice water measurements obtained from the Harvard Total Water and the Harvard Water Vapor instruments, we identify and remove occasional cloud data sampled in our region of interest. Consequently, all references to H_2O in this study are to H_2O in the vapor phase only. For the lower boundary of the lowermost stratosphere, we use the local tropopause observed during each flight. The location of the local tropopause varied from a minimum potential temperature of 360 K (14.4 km or 147 hPa) to a maximum of 365 K (15.6 km or 121 hPa). For the upper boundary of the lowermost stratosphere, we use the tropical tropopause corresponding to the 380-K isentrope, which was found to be between 15.2 km (127 hPa) and 16.2 km (110 hPa) over southern Florida during the campaign.

2.3. Source Regions

[15] In order to put the CRYSTAL-FACE measurements in context, we use in situ measurements from three previous aircraft campaigns. These are: Stratospheric Tracers of Atmospheric Transport (STRAT) based out of California in July 1996 and out of Hawaii in August 1996, Photochemistry of Ozone Loss in the Arctic Region in Summer (POLARIS) based out of Fairbanks, Alaska in July 1997, and Clouds and Water Vapor in the Climate System (CWVCS) based out of Costa Rica in August 2001. Specific locations of these campaigns are shown in Figure 2. The STRAT data are used to obtain reference values for tropical source regions and to investigate circulation in the stratosphere on the west side of the monsoon. The POLARIS and CWVCS data are used to obtain reference values for the possible high-latitude and tropical source regions for air transported into the subtropical lowermost stratosphere via the pathways shown in Figure 1.

[16] The majority of the tracer measurements obtained during these campaigns were made by the same instruments (with attendant uncertainties) used during CRYSTAL-FACE. A few measurements, however, were obtained from different instruments. During the STRAT and POLARIS campaigns, NO and NO_y were measured by the NOAA Aeronomy Laboratory reactive nitrogen instrument [*Fahey et al.*, 1989] with uncertainties of ±(6% + 4 pptv) for NO and ±(10% + 100 pptv) for NO_y, and CO was measured by the Aircraft Laser Infrared Absorption Spectrometer [*Webster et al.*, 1994] with a ±5% uncertainty. During CWVCS, O_3 was measured by the Harvard ozone instrument instead [*Weinstock et al.*, 1986]. All instruments flew aboard NASA's ER-2 during STRAT and POLARIS, and aboard NASA's WB-57 during CWVCS.

[17] The tracers used in this study, except for CO_2, do not have significant trends in the stratosphere that can compromise the conclusions of this study. While the spatial coverage of the in situ aircraft data is limited, several studies have shown consistent correlations in the stratosphere of long-lived tracers such as O_3, CO_2, NO_y, and N_2O collected during aircraft campaigns that took place at different latitudes and longitudes between 1987 and 1997 and excellent agreement between aircraft measurements and output from chemical transport models [*Murphy et al.*, 1993; *Fahey et al.*, 1996; *Strahan et al.*, 1998; *Strahan*, 1999]. This consistency and level of agreement serve as evidence that measured tracer-tracer correlations show little longitudinal dependence, and can be used to represent different latitudinal regions of the stratosphere. On the basis of the results from these studies, we create reference tables for each source region, and we check for consistency with measured tracer-tracer correlations and climatologies [see *Strahan*, 1999; *Richard et al.*, 2003]. We do not find compromising departures between our reference profiles and the climatologies that can have a significant effect on our results. The location and ranges of tracer mixing ratios for the source regions used in this study are reported in Tables 1 and 2, respectively.

[18] In addition to in situ aircraft data, we also use surface measurements provided by NOAA's Climate Monitoring and Diagnostics Laboratory Carbon Cycle and Greenhouse Gases (CMDL CCGG) group. In particular, we use CO_2 measurements collected at various locations in the tropics

Table 2. Representative Values of Trace Gases From Each of the Four Source Regions Considered in This Study[a]

Source Region (Pathway)	H_2O, ppmv	O_3, ppbv	CO_2, ppmv	CO, ppbv	NO, pptv	NO_y, pptv
HLS (A)	10–4	300–600	370.5–368.7	35–21	250–170	2000–2800
Convection (B)	10s–100s	60–120	369–374.5	80–140	100–2000	200–4000
TTL (C)	12–4.8	40–160	373.8–372.8	70–45	150–400	200–1300
TLS (D)	9–4.5	150–400	373.4–371.4	45–25	250–500	500–1500

[a]All ranges except the ones for the convective source are listed as observed values at the bottom of the region followed by observed values at the top of the region. Each region is specified in Table 1. The CO_2 ranges in the TTL and TLS represent the extreme values from the seasonal cycle of this tracer within each source region. These values might not necessarily occur at the edges of these source regions. Pathways are shown in Figure 1.

and midlatitudes. The mixing ratios reported in Table 2 are adjusted or estimated in order to account for this tracer's secular trend and seasonal cycle in the atmosphere. In the case of POLARIS data, we adjust the 1997 measurements to what they would be in 2002 using the secular trend of 1.5 ppmv/year. For the tropical source regions, we estimate the mixing ratios using the mean monthly averaged surface measurements at Mauna Loa and Samoa and the transport times through the TTL and the tropical lower stratosphere observed by *Boering et al.* [1996].

[19] Because tropospheric H_2O is very variable, it is difficult to provide constraints for the contribution of this tracer from the troposphere. Instead of relying on previous measurements, we calculate saturation mixing ratios using the Clausius-Clapeyron equation and tropopause pressure and temperature obtained from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) Reanalysis data set. A warm bias in NCEP/NCAR tropical temperatures was accounted for by subtracting 2 K from the temperature field [*Shah and Rind*, 1998]. The reported H_2O values for the TTL source region shown in Table 2 correspond to the minimum and maximum saturation mixing ratios at the local tropopause between the equator and 15°N and 0° and 180°W during July 2002. In convective sources, H_2O can vary by one or two orders of magnitude, which makes it difficult to constrain. The actual contribution of H_2O vapor from a convective source is controlled by various factors, such as temperature where ice particles from the convective system detrain, strength of the convection, and microphysical conditions. Because of the large variability, we do not provide a specific range in Table 2.

3. Methodology

3.1. Statistical Analysis

[20] The statistical analysis is carried out in three steps. The goal of the first step is to identify "natural" groupings of the data set based on the chemical composition of the air masses using methods that do not rely on prior knowledge or constraints. For this purpose, multivariate cluster analysis is performed on clear-air observations in the lowermost stratosphere. Each air mass observation consists of six tracer measurements, which serve as the six coordinates in data space. Since tracer mixing ratios are reported in different units, we use standardized anomalies instead of the actual mixing ratios for each tracer value. The standardized anomaly is calculated by subtracting the mean value for the campaign from each observation and dividing this difference by the standard deviation. Multivariate clusters are created using the *k*-means method, which is nonparametric and nonhierarchical [*Hargrove and Hoffman*, 2004; *Hoffman et al.*, 2005]. Briefly, in this method the only parameter chosen a priori is k, the number of clusters, and no other a priori knowledge is used. The procedure starts with a set of k seed centroids that are delineated in principal component space to obtain a reasonable first guess. These centroids are a widely distributed subset of the observations and the final results do not depend on the initial guesses. In the first pass, every observation in PC space is assigned to the closest centroid. Then at the end of the pass, the centroid position is recomputed to be the mean of each coordinate of all of the observations classified to that centroid. A new iteration starts using the recomputed centroids. This iterative process continues until fewer than 0.5% of the observations change cluster assignment from the last iteration. At this point, the group assignment has been stabilized and the classification process has converged. The fact that clusters are independently redefined at each new iteration makes this method nonhierarchical. The main criterion used for the selection of k is based on the distribution of the data among the clusters. Ideally, clusters should contain neither too many nor too few observations. We tested various values for k, namely 5, 15, and 20, to produce an acceptable distribution. We found $k = 15$ to give the best distribution of number of observations among clusters.

[21] The goal of the second step is to explore the similarities and differences among the 15 clusters. For this purpose, we perform agglomerative clustering, which is nonparametric but hierarchical [*Wilks*, 2006]. Similar to the multivariate clustering, this method also uses standardized anomalies instead of actual mixing ratios. In this technique, we use the Euclidean metric to calculate distances between observations in the n-dimensional space (in our case $n = 6$). The six-dimensional space where the distances are calculated is composed of the standardized anomalies for each of the six tracers. We will refer to this space as the tracer space. When clusters contain more than one observation, several schemes can be used to determine the actual distance between two clusters (groups of observations). The most common schemes are: single-linkage if the distance is the minimum distance between an observation from one cluster and an observation from the other cluster, average-linkage if the distance is the average distance between all possible pairs of observations in the two clusters, and complete-linkage if the distance is the maximum distance between an observation from one cluster and an observation from the other cluster. We tested these three linkage schemes, and we found that all approaches yielded essentially the same underlying structure. We choose the complete linkage scheme, since it is based on a more stringent criterion for grouping clusters [cf. *Wilks*,

2006, p. 552]. Once the distances between all clusters have been determined, we perform the agglomerative hierarchical clustering. This is an iterative process where at each step the pair of clusters (or groups of clusters) that reside the closest to each other in the six-dimensional tracer space are merged. This process of creating new groups at each step is continued until all clusters are eventually merged into one group (i.e., the original data set). The results of the agglomeration are presented in a tree diagram, or dendrogram, illustrating the hierarchy of the sets of groups. In the dendrogram, clusters that are closer to each other in tracer space, indicating that they share similar tracer mixing ratios, show up as members of the same branch of the tree. Clusters with similar tracer mixing ratios could be associated with similar origins.

[22] The goal of the third step is to provide context for the interpretation and understanding of the clustering results. For this, we use PCA, which is a parametric approach. In this technique, the eigenvalues and eigenvectors of a correlation matrix constructed using the standardized anomalies of a data set are calculated [*Wilks*, 2006]. In our case, the data set consists of the original, nonclustered air mass observations obtained throughout the campaign, where each observation is composed of six tracer measurements. The eigenvectors represent the orthogonal modes of variability in the data and their corresponding eigenvalues give the relative importance of each eigenvector. The first eigenvector explains the largest mode of variability, and each successive eigenvector explains the next largest mode of variability not accounted for by previous eigenvectors. Eigenvectors are commonly referred to as empirical orthogonal functions (EOFs) when they characterize spatial patterns of variability at specific points in time. Since at each point in time, our data set contains only one observation in space, to avoid confusion we will simply refer to them as eigenvectors rather than EOFs. Of the six eigenvectors obtained using this technique, the first three are kept, since their eigenvalues account for 93% of the total variance (51%, 28%, and 14%, respectively). Subsequently, we calculate the principal component time series for each of the three eigenvectors, which we will simply refer to as PC1, PC2, and PC3, respectively. We calculate each PC by summing the product of the standardized anomalies of the data and the corresponding eigenvector over all six tracers for each air mass observation obtained during the campaign. Hence these times series represent the temporal evolution throughout the campaign of the contribution of each eigenvector (or mode of variability) to the total variance. We then plot these orthogonal PCs against each other and color code each element of the PC on the basis of the cluster group assignment of the observation that it represents [*Boccippio et al.*, 2005]. This approach allows us to explore the behavior of the cluster groupings in the context of independent modes of variability in the data.

3.2. Tracer-Tracer Correlations

[23] *Plumb and Ko* [1992] showed that long-lived tracers have compact and nearly linear correlation plots in the stratosphere with a slope reflecting the influence of sources and sinks. Correlation plots have been used effectively in previous studies to help distinguish air originating from different source regions [*Hintsa et al.*, 1998; *Hoor et al.*, 2002; *Ray et al.*, 2004].

[24] In this study, we analyze correlation plots of different tracers with respect to H_2O for several reasons. This tracer has a strong gradient across the tropopause with low values in the stratosphere and high values (one to three orders of magnitude larger) in the troposphere, thereby allowing us to identify convectively injected air into the stratosphere. H_2O has a lifetime of 100 years in the lower stratosphere [*Brasseur and Solomon*, 1986] making it a good tracer for large-scale transport. This tracer has a well documented seasonal cycle and a stratospheric source. On an annual basis, seasonal changes in stratospheric H_2O over the tropics closely track the saturation mixing ratios set by the temperatures of the tropical tropopause [*Mote et al.*, 1996; *Weinstock et al.*, 2001]. The main stratospheric source of H_2O is oxidation of CH_4. For every molecule of CH_4 destroyed, two molecules of H_2O are produced and this contribution is obvious in air masses that have spent more than 3.8 years in the stratosphere and are located north of 20°N and above 440 K [*Dessler et al.*, 1994; *Hurst et al.*, 1999]. For a timescale equivalent to the period of the campaign, changes in H_2O due to the seasonal cycle and CH_4 oxidation can be neglected.

[25] In addition to analyzing tracers versus H_2O (tracer:H_2O) correlation plots, we also analyze CO_2:O_3 and NO_y:O_3 plots. These three tracers are long-lived and exhibit significant vertical and latitudinal gradients in the stratosphere. Both O_3 and NO_y increase with increasing latitude and altitude in the lower stratosphere whereas CO_2 decreases with increasing latitude. Therefore correlation plots using these tracers can provide latitude and altitude information of potential source regions.

[26] On the basis of our current understanding of sources and sinks in the troposphere and stratosphere, we expect to find the following tracer-tracer correlations of air masses in the stratosphere. NO:H_2O and CO:H_2O should show positive correlations, because all these tracers are mainly produced in the troposphere and all of them are then removed by different mechanisms (e.g., dehydration, photolysis, chemistry) once the air reaches the stratosphere. NO_y:O_3 should also show a positive correlation, because these tracers are both lower in the troposphere and higher in the stratosphere where they are produced as a result of photolysis and chemistry. O_3:H_2O and NO_y:H_2O should show negative correlations instead, because while the troposphere is a major source for H_2O, O_3 and NO_y are mainly generated in the stratosphere except for the case of tropospheric convection where lightning can be a significant source of NO_y (via NO_x production). The correlations of CO_2 versus other tracers in tropospheric air can be complicated by the seasonal cycle of CO_2 and its different tropospheric sources (e.g., maritime versus continental, tropics versus high latitudes). For stratospheric air, CO_2:O_3 should show a negative correlation, where CO_2 is decreasing in older air because of its secular trend while O_3 is increasing because of photochemical production. The CO_2:H_2O correlation is further complicated by the seasonal cycle of H_2O and convection into the UT/LS. In general, all tracer:H_2O correlations can breakdown in the presence of

convective air, because convection does not have a unique effect on H_2O.

3.3. Trajectory Model

[27] Backward trajectories for selected flights during the CRYSTAL-FACE campaignare calculated using the Forward and Backward trajectory (FABtraj) model that has been verified against satellite imagery, trace gas measurements and the widely used FLEXTRA model [*Cooper et al.*, 2004a]. The model has been used for studying air mass transport through synoptic-scale midlatitude cyclones and for simulating air mass transport within the lower stratosphere and stratospheric intrusions [*Cooper et al.*, 2004b].

[28] The FABtraj model is a diabatic model that uses NCEP Final Analyses (FNL) wind fields and a linear interpolation scheme in space and time. The meteorological data set used is available every 6 hours and has a horizontal grid resolution of $1^\circ \times 1^\circ$ and a vertical resolution of 21 levels between 1000 and 100 hPa. The vertical levels are prescribed in sigma (or terrain-following) coordinates. Vertical motion in the FABtraj model is driven solely by the vertical wind component of the FNL analyses, which is undefined at pressures lower than 100 hPa. This constraint on vertical motion at low pressures, however, does not compromise the conclusions of our study, because it affects only a very small number of trajectories (e.g., 1 out of 196 initialized on 16 July and 6 out of 80 initialized on 23 July). Convection is therefore not parameterized and it is not resolved in subgrid scale in this model. The effects of convection, however, are parameterized in the FNL.

[29] Summertime convection can be strong enough to transfer air masses from the troposphere into the stratosphere, a process that involves crossing of isentropes. Given the location and the season of CRYSTAL-FACE, we choose to use a trajectory model that allowed this crossing instead of using adiabatic trajectories. Additionally, air masses that are advected back in time for more than 2–3 days are very likely to experience changes in temperature during the summer, especially when traveling over or through regions of strong and frequent convective activity over land. This condition affects the entropy of the air mass and makes its trajectory nonadiabatic.

4. Results and Discussion

4.1. Source Region Identification

[30] Table 2 shows reference values for the six chemical tracers used in this study in each of the four possible source regions identified in Figure 1. Each source region can be qualitatively identified on the basis of (1) the correlations of chemical tracer abundance observed in Table 2 (e.g., high tracer 1 and low tracer 2 in source A) and (2) our understanding of sources and sinks described in section 3.2.

[31] We will show next how different tracers can help discriminate different source regions. Older stratospheric air coming from high latitudes (pathway A in Figure 1) is clearly characterized by higher O_3 and NO_y, and lower CO_2. When this air is not contaminated by convection, H_2O, CO, and NO are expected to be low. Younger stratospheric air coming from tropical latitudes, on the contrary, is characterized by lower O_3 and NO_y, and higher CO_2 (pathways C and D in Figure 1). Similarly, when not contaminated by convection, H_2O, CO, and NO are expected to be low but CO and NO would be higher than in nonconvective older stratospheric air. Convective air found in the lowermost stratosphere is clearly characterized by higher H_2O (pathway B in Figure 1). Good indicators of recent convective injection into the UT/LS region (i.e., less than 1 week old because of the chemical lifetime of NO) are higher CO, NO and NO_y (driven by higher NO). Convective injection into the UT/LS region older than a week will be characterized by elevated H_2O and possibly CO, and low NO due to its chemical loss over time.

4.2. Statistical Analysis

[32] The multivariate clustering technique is used to construct 15 distinct clusters based on the six chemical tracers chosen for this study. Figure 3 shows the centroid values and ranges for each tracer and each cluster. This clustering technique attempts to minimize multivariate variability and construct clusters with clear and unique interpretations. In some cases, however, we might find the cluster interpretation to be ambiguous. Therefore it is important to examine and consider not only the centroid values themselves, but also the range and distribution of values of all cluster members relative to the centroid value. This point will be important later on when we explain the contents of Table 4.

[33] The results of the agglomerative clustering are summarized in the dendrogram shown in Figure 4. To the right of the dendrogram, the percent occurrence of each cluster based on the air masses sampled by the aircraft during the campaign is presented. The closer the clusters are in tracer space (x axis in Figure 4a), the more similarities these clusters share in terms of tracer values and hence presumed origin, and vice versa. The largest distance is found to be between cluster 15 (or C15) and the remaining clusters. According to the results shown in Figure 3, this cluster can be easily and uniquely identified as air dominated by recent convection based on the highest mixing ratios of NO. As we decrease the distance in tracer space in the dendrogram (i.e., move to the left along the x axis), we notice the emergence of new branches. These different branches suggest the existence of structure in the data. The dashed vertical line is plotted at a distance where five distinct branches occur. Those branches or groupings are color coded as follows: group 1 or G1 (with C2, C9, C3, C8, C4, and C7) is coded yellow, G2 (with C6, C10, C11, C1, and C12) is coded green, G3 (with C5 and C14) is coded black, G4 (with C13) is coded red, and G5 (with C15) is coded blue (see Figure 4). This classification will constitute the basis for the analysis presented from this point forward. Groups G1 and G2 both contain a relatively large number of clusters; the dendrogram suggests these can also be divided into subgroups by moving even farther to the left along the x axis. The subgroups within each group presumably share similar origins, but we will show how their distinctions are based on subtle but important differences in the fractional contributions of the dominant source regions and in some cases, the altitudinal location of the clusters.

[34] The grouping of clusters presented in Figure 4 is the result of the complete linkage agglomerative hierarchical clustering scheme chosen. The main difference between these results and the ones obtained using other linkage

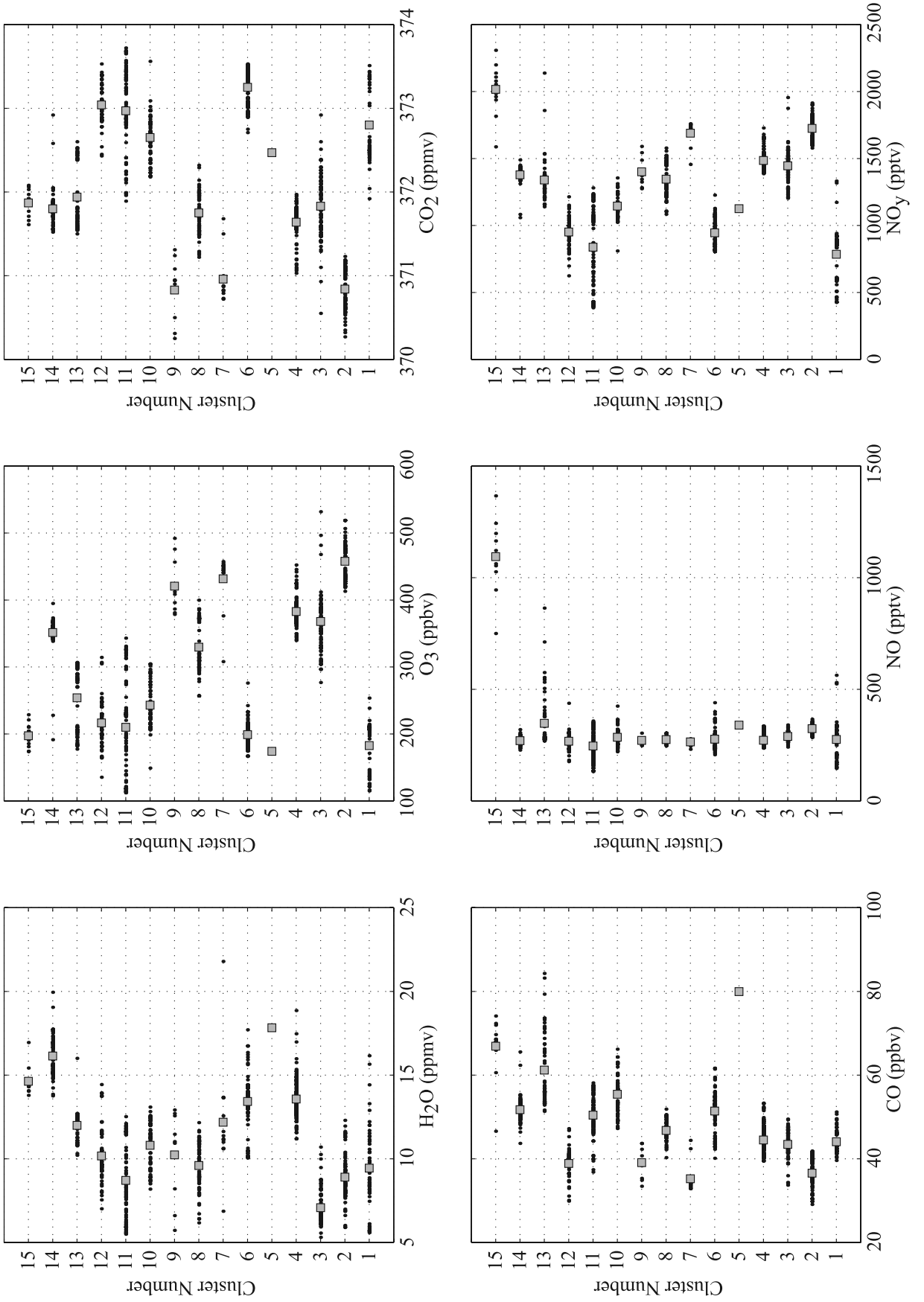

Figure 3. Multivariate clusters constructed using six chemical tracers. Ranges of tracer values are shown in black, and centroid values are represented by the gray squares for each cluster.

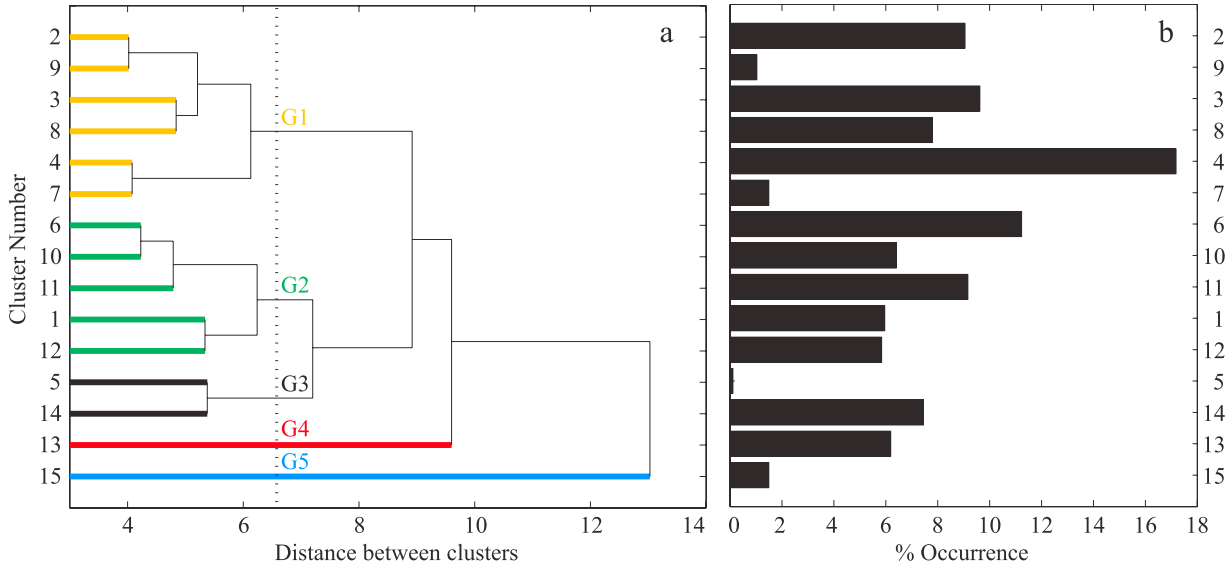

Figure 4. (a) Agglomerative hierarchical clustering of the 15 multivariate clusters. The distance between clusters is calculated in tracer space. The vertical line shown at 6.5 is used to analyze the branches found at this distance. The branches or groups are color coded as shown in the figure. Subgroups are found within G1 and G2 and their distinctions are explained in the text. (b) Percent occurrence of each cluster sampled in the lowermost stratosphere during CRYSTAL-FACE. This distribution shows that 15 is an appropriate number of clusters.

schemes is in the membership of C13 and C14. The different calculation details forced these clusters to change branches. This result implies that these clusters have somewhat ambiguous origins. As will be shown later, these two clusters are made up of members that appear to have differing origins.

[35] Before examining the five groups of clusters in PC space, we deduce a physical interpretation for each of the underlying eigenvectors (hence PCs) shown in Table 3 using the qualitative description of source regions described in section 4.1. The coefficients of these vectors constitute the weight given to each tracer in the calculation of the PCs. In eigenvector 1, the largest positive weights are given to O_3 and NO_y and the largest negative weight is given to CO_2, suggesting that this eigenvector denotes how old on average stratospheric air is. In eigenvector 2, the largest weights, all positive, are given to H_2O, CO, and NO, suggesting that this eigenvector denotes how recent convective influence is. Last in eigenvector 3, the largest positive weight is given to H_2O and the largest negative weight is given to NO, suggesting that this eigenvector denotes how large and/or old convective influences are. While mathematically possible, the large negative NO anomalies required for old convective influence should not physically occur. Instead, this weight reflects the many small negative anomalies that represent the background NO values to which aged convective air will relax.

[36] Figures 5–7 show the vertical profiles of each of the tracers with shading based on the sign of the PC. Figure 5 shows the tracer profiles for PC1. Consistent with the interpretation of eigenvector 1, PC1 denotes the "mean age of air in the stratosphere." When a PC1 value is positive, the air mass observed at that time corresponds to older air presumably originating from higher latitudes in the stratosphere, and when the value is negative, the air mass corresponds to younger stratospheric air presumably originating from tropical latitudes. Because of the small coefficients for H_2O and NO in eigenvector 1, both old and young air masses appear to have varying extents of convective influence. Figure 6 shows the tracer profiles for PC2. Consistent with the interpretation of eigenvector 2, PC2 denotes the "age of the convective influence." When a PC2 value is positive, the observed air mass at that time corresponds to recent convection, and when the value is negative, the air mass corresponds to older convection. Besides an apparent, and not unreasonable, altitude preference of positive PC2 values for lower altitudes, the other two tracers, O_3 and CO_2, have small coefficients in eigenvector 2 implying that there is little correlation between the age of the convective influence and the mean age of stratospheric air. Figure 7 shows the tracer profiles for PC3. Consistent with the interpretation of eigenvector 3, PC3 denotes the "extent of convective influence." When a PC3 value is positive, the air mass observed at that time corresponds to greater and/or older convective influence, and when the value is negative, the air mass corresponds to smaller convective influence. The negative PC sign of the few large NO values associated with a known recent convective event exemplifies the mathematical peculiarity

Table 3. Coefficients of the Three Eigenvectors Used in This Analysis[a]

Eigenvector	H_2O	O_3	CO_2	CO	NO	NO_y
1	−0.0271	0.5548	−0.5533	−0.3067	0.1084	0.5287
2	0.5044	−0.0901	−0.0452	0.5520	0.5978	0.2709
3	0.7853	0.1796	−0.0301	−0.0200	−0.5872	−0.0709

[a]Each coefficient constitutes the weight given to each tracer in the calculation of the PCs shown in Figures 6–8.

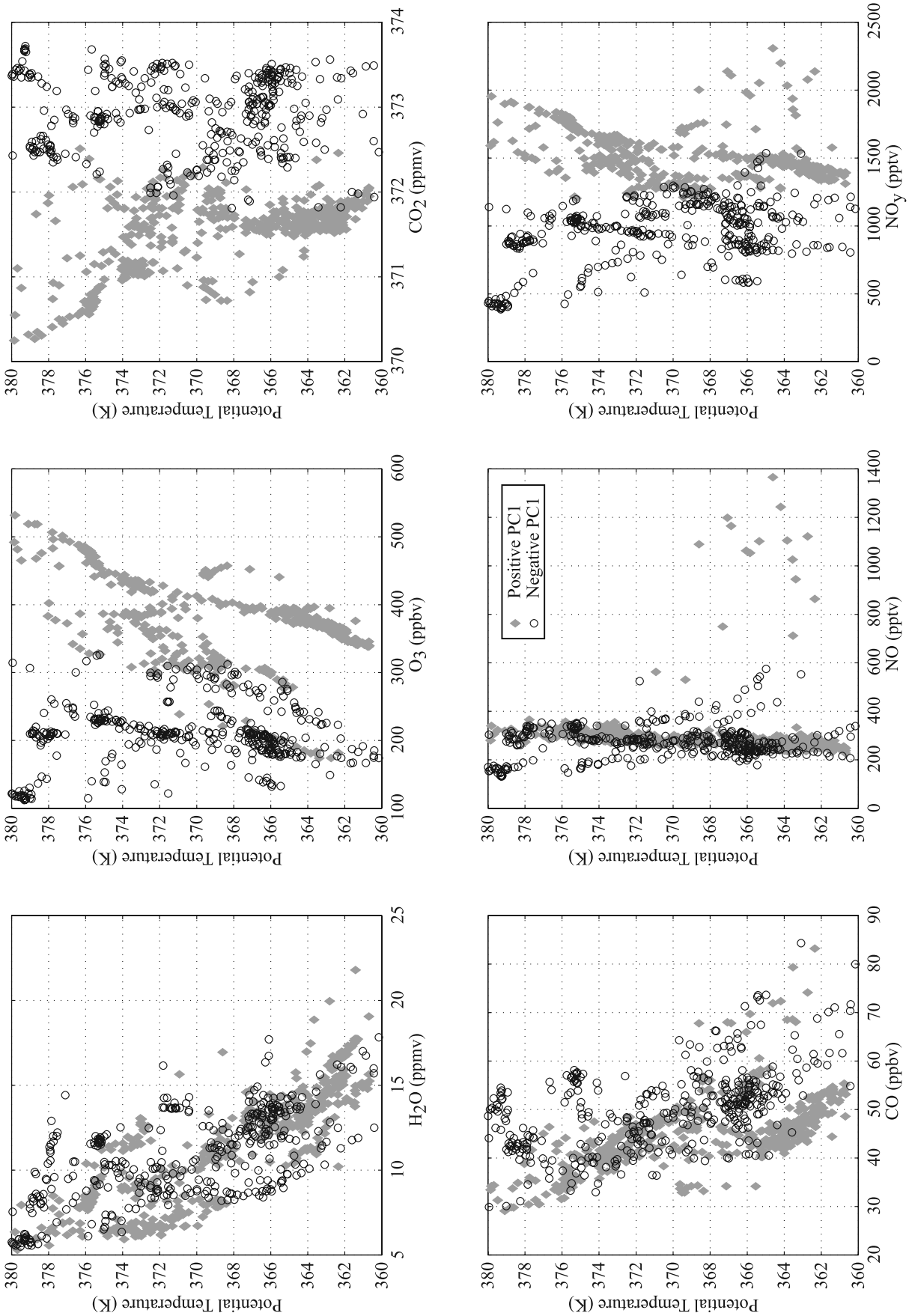

Figure 5. Vertical profiles of the six chemical tracers used in this study shaded on the basis of the sign of PC1. Gray diamonds are for positive PC1, and black circles are for negative PC1. This PC is inferred as representing the mean age of air in the stratosphere. See text for a detailed explanation of the inference.

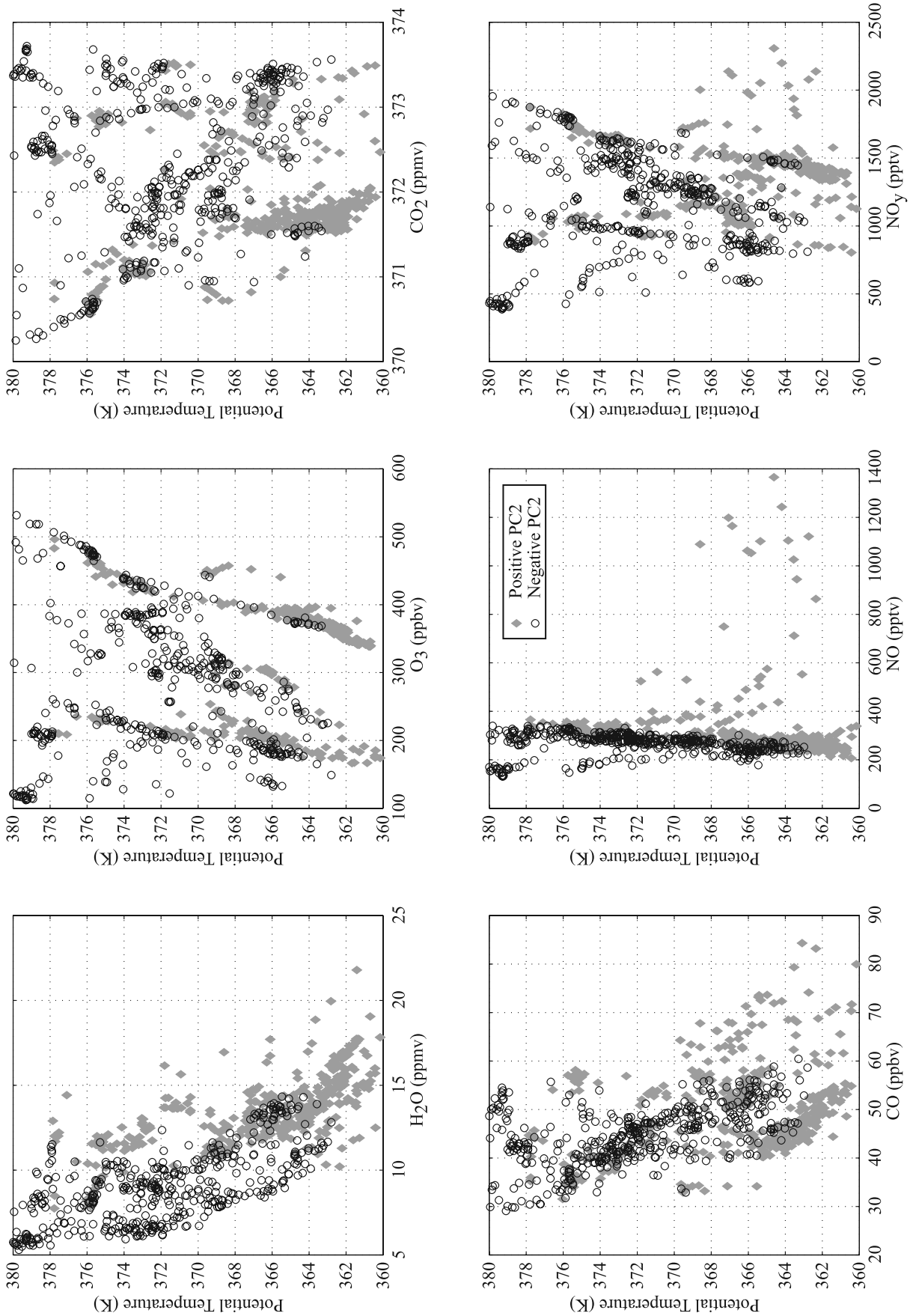

Figure 6. Vertical profiles of the six chemical tracers used in this study shaded on the basis of the sign of PC2. Gray diamonds are for positive PC2, and black circles are for negative PC2. This PC is inferred as representing the age of the convective influence. See text for a detailed explanation of the inference.

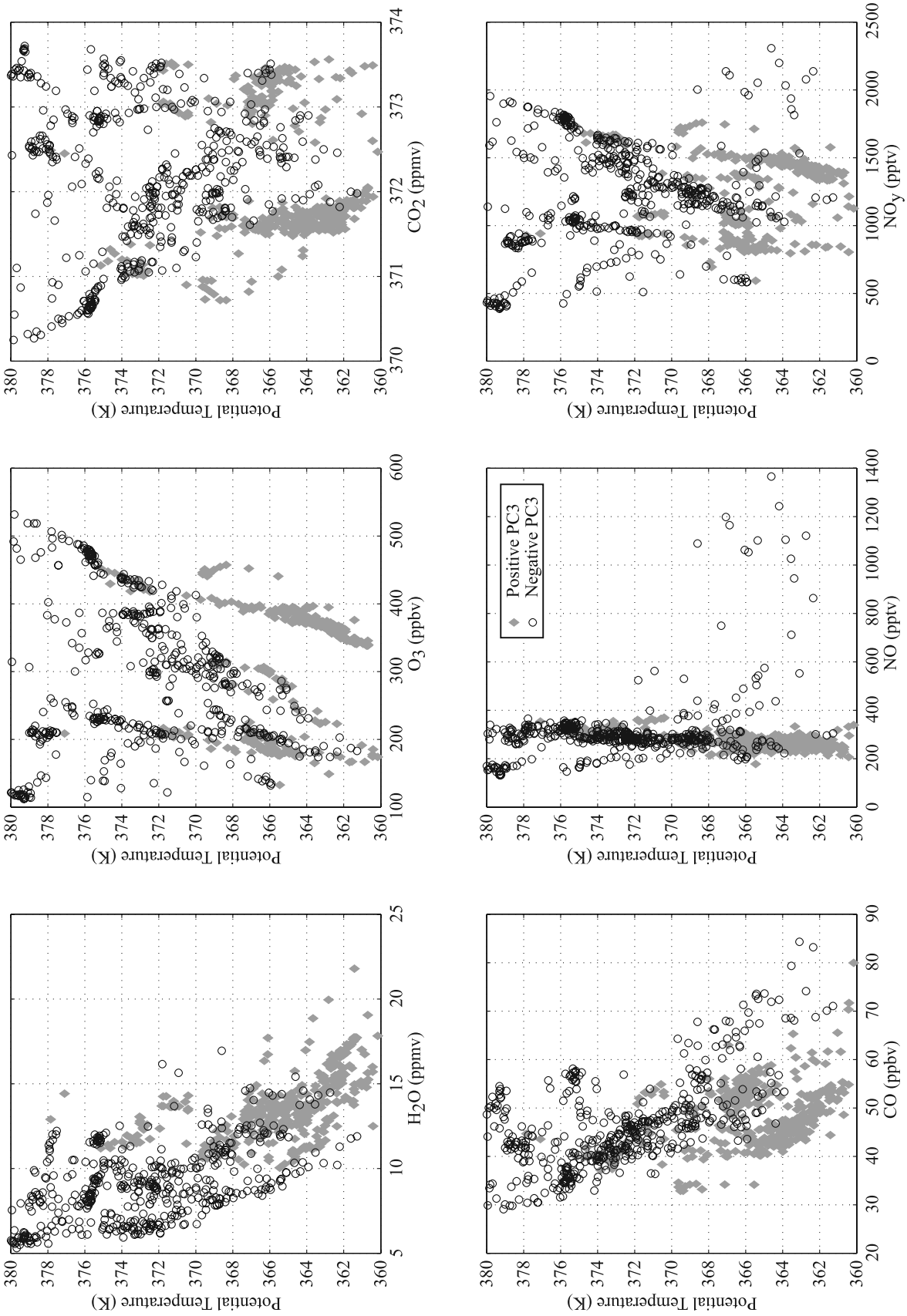

Figure 7. Vertical profiles of the six chemical tracers used in this study shaded on the basis of the sign of PC3. Gray diamonds are for positive PC3, and black circles are for negative PC3. This PC is inferred as representing the extent of the convective influence, which can be related to the latitude of the convective input following results from *Dessler and Sherwood* [2004]. See text for a detailed explanation of the inference.

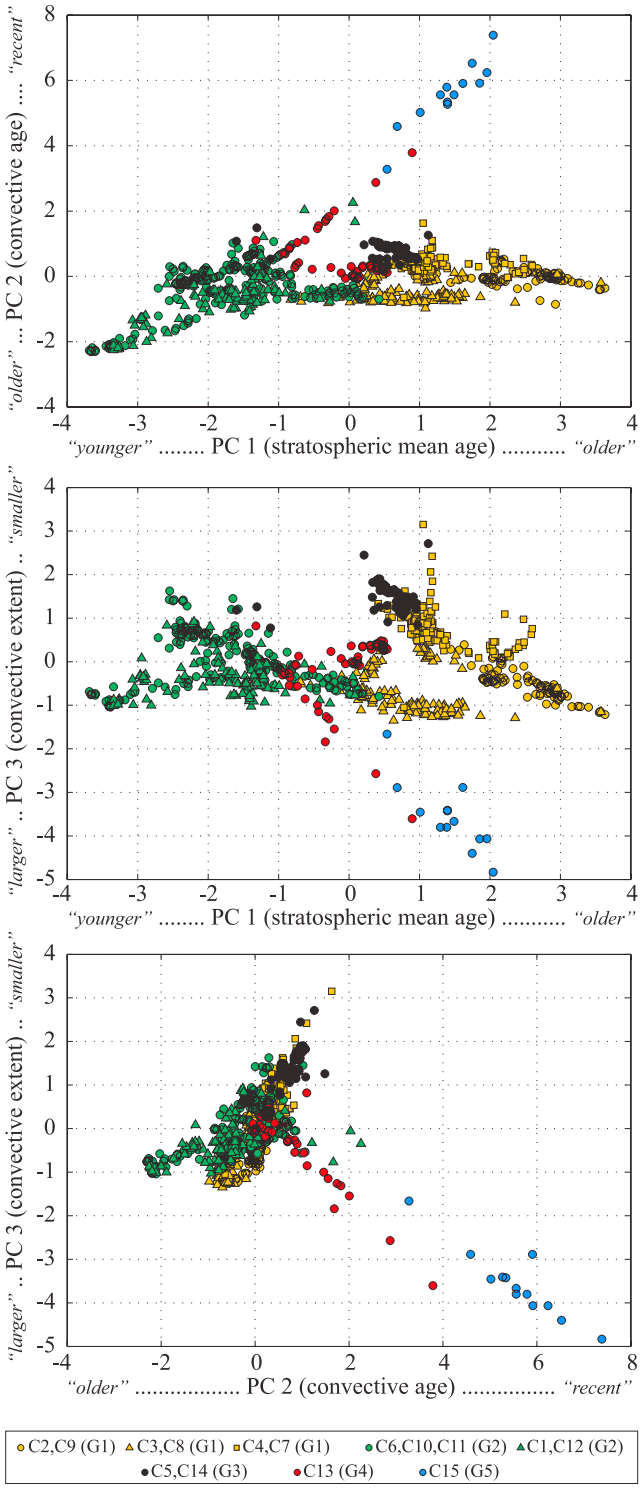

Figure 8. First three PCs color coded on the basis of the groups defined in Figure 4. Legend shows cluster numbers followed by (group number). G1 is referred to as HLS + OC, G2 is referred to as TLS* + OC, G3 is referred to as HLS + TLS* + wet, G4 is referred to as HLS + TLS* + RC, and G5 is referred to as RC. See text for details.

noted above for eigenvector 3. The other tracers, O_3, CO_2, CO and NO_y, have small coefficients in eigenvector 3 implying that there is little correlation between the extent of convective influence and the age of the convection and the mean age of stratospheric air. When removing the air masses with high NO caused by recent convection and consequently the resulting extreme negative values for PC3, the magnitude and sign of PC3 in the remaining air masses are highly dominated by H_2O only (correlation coefficient of 0.93). The more positive the standardized anomaly for H_2O is, the larger the value for PC3 is. In a recent study, *Dessler and Sherwood* [2004] argued that the impact of convective injection on stratospheric H_2O is higher at higher latitudes. They attributed this connection to warmer temperatures, lower relative humidity, and a larger "convective contrast" (i.e., difference in tracer concentration between the convective outflow and the local environment) in the midlatitude stratosphere compared to the tropics. We will use their argument to extend the physical interpretation of PC3 as follows: positive values, which are driven by positive anomalies in H_2O, correspond to a larger extent of convective influence and presumed latitudes of convective injection northward of southern Florida where the convective contrast is larger, and vice versa.

[37] We next examine the five groups of clusters in PC space. The reason for doing this is to look at the correspondence between clusters and PCs in the context of the dominant modes of variability in the data set. This approach is somewhat analogous to the more physically based tracer-tracer correlations, but in this case we use a statistical framework instead. The results are shown in Figure 8.

[38] On the basis of the physical meaning deduced for the underlying eigenvectors of each PC using tracer mixing ratios, and the location of each cluster group in PC space shown in Figure 8, we find the following results. G1 is high-latitude stratospheric air mixed with various amounts of older convective air except for a few members in the C4–C7 couplet that show more recent and larger convective influence from events that presumably occurred at northern latitudes, G2 is tropical air (from the tropopause layer or the lower stratosphere) mixed with various amounts of older convective air, and in some cases even older convective air than G1, from events both north and south of our region of sampling, G3 is a mix of the air masses embodied in both G1 and G2 with more recent and larger convective influence than most members of the previous two groups from events that occurred at northern latitudes, G4 is also a mix of the G1 and G2 air masses, but in some cases with even more recent convective influence than all previous groups, though lesser perhaps because of the injection occurring in the tropics, and finally G5 is high-latitude stratospheric air mixed with the most recent convective air but the least convective influence. The observations that define G5 are characterized by the spikes in NO, which were sampled by the aircraft over active convective regions on the flight of 21 July. In the next section, however, we will show and explain why the assignment of older stratospheric air in G5 based on statistical results is not consistent with physical measurements. On the basis of the dominant source regions identified, we will refer to each group of clusters as follows: G1 as HLS + OC, G2 as TLS* + OC, G3 as HLS + TLS* + wet, G4 as HLS + TLS* + RC, and G5 as RC, where OC

Table 4. Physical Interpretation of Multivariate Clusters Obtained by Comparing Centroid Values and Ranges for Each Tracer and Each Cluster as Shown in Figure 3, to Expected Values Summarized on Table 2[a]

Cluster (Group)	H_2O, ppmv	O_3, ppbv	CO_2, ppmv	CO, ppbv	NO, pptv	NO_y, pptv	Theta (K) ± 1σ	Assigned Dominant Source Regions
2 (1)	8.91	457.91	370.84	36.59	323.67	1724.9	375.0 ± 1.8	HLS + OC
9 (1)	10.24	420.82	370.83	39.11	271.15	1400.4	370.7 ± 6.3	HLS + OC
3 (1)	7.08	367.89	371.83	43.46	288.99	1445.2	372.7 ± 2.6	HLS + OC? TLS + OC?
8 (1)	9.6	329.55	371.75	46.8	274.93	1345.9	369.2 ± 3.4	HLS + OC? TLS + OC?
4 (1)	13.57	382.72	371.64	44.51	271.55	1485.8	365.0 ± 3.0	HLS + OC
7 (1)	12.19	431.81	370.96	35.25	264.27	1689.6	368.5 ± 2.8	HLS + OC
6 (2)	13.43	199.04	373.25	51.43	276.6	945.4	367.6 ± 3.3	TLS + OC
10 (2)	10.81	242.89	372.65	55.43	285.12	1144.6	366.9 ± 2.3	TLS + OC
11 (2)	8.71	209.87	372.97	50.45	246.38	838.39	374.1 ± 4.1	TLS + OC
1 (2)	9.45	182.68	372.8	44.1	275.37	785.77	375.4 ± 5.0	TLS (or TTL?) + OC
12 (2)	10.17	216.49	373.04	38.91	267.68	950.92	372.2 ± 4.0	TLS + OC
5 (3)	17.83	174.1	372.47	79.97	339.51	1125.9	360.2	OC + TLS
14 (3)	16.14	351.47	371.8	51.76	269.81	1378	362.1 ± 0.7	OC + HLS? OC + TLS?
13 (4)	12	254.09	371.94	61.2	347.39	1339.8	365.9 ± 2.1	OC + HLS? OC + TLS?
15 (5)	14.63	197.38	371.87	66.92	1094.4	2016	365.3 ± 1.8	RC

[a]Clusters are listed in the same order as they appear in Figure 4. OC stands for old convection, and RC stands for recent convection.

stands for Old Convection, and RC stands for Recent Convection. For simplicity and to avoid confusion, we use TLS* in the group names to refer to all air coming from the tropics regardless of altitude of origin (i.e., TTL or TLS).

[39] Table 4 shows the tracer mixing ratios and altitude for each centroid. We first compare these values, keeping in mind the ranges of tracers in each cluster shown in Figure 3, to the possible ranges for each source region presented in Table 2. On the basis of all this information, we identify the one (or two) most dominant source region(s) in order to explain the observations. We found that at most two source regions could be identified with confidence using this approach. The possibility of mixtures with more than two source regions cannot be ruled out, but they would be difficult to clearly identify using this method. Those conclusions are summarized in the right-most column of Table 4. We compare the statistically deduced group origins to the physical measurements shown in Table 2 and we find that the statistical approach yields physically consistent results.

[40] In some cases, however, we find that comparing tracer mixing ratios of clusters to possible source regions alone leads to ambiguous results as seen in Table 4. The apparent ambiguity in cluster origin can be elucidated by exploring the underlying statistical behavior of the data. For instance, both C3 and C8 have most members in +PC1 (i.e., older stratospheric air or HLS) but a few lie in −PC1 (i.e., younger stratospheric air or TLS). These few members in −PC1 are thus impacting the centroid values, which are nothing more than the arithmetic mean of the cluster members' coordinates, and therefore producing the ambiguity shown in Table 4. Similarly, C13 and C14 have members in both +PC1 and −PC1 but with a more even distribution of the number of members between the two signs of PC1. These two clusters, C13 (or the HLS + TLS* + RC group) and C14 (which is 98% of the HLS + TLS* + wet group), not only contain members with different ages represented in PC1, but they also differ in PC2 and PC3 space as previously described. Recall multivariate clustering is a purely nonparametric statistical technique; it produces statistically homogeneous results, but this does not guarantee physical homogeneity. Clusters that have less physical homogeneity may change cluster group membership depending on the linkage scheme used in the agglomerative clustering as our results show for the case of C13 and C14.

4.3. Physical Analysis

[41] We now use the statistical groups from the previous section to help analyze the physically based tracer-tracer correlations. Figure 9 shows tracer:H_2O correlations and Figure 10 shows CO_2:O_3 and NO_y:O_3 correlation plots.

[42] Several interesting features emerge in Figure 9. For instance, the stratospheric age distinction between the HLS + OC and the TLS* + OC groups can be identified in these correlation plots. Older stratospheric air with higher O_3, and nonconvective NO_y, and lower CO_2, CO, and NO belongs to the HLS + OC group, while younger stratospheric air with lower O_3, and nonconvective NO_y, and higher CO_2 and CO belongs to the TLS* + OC group. The extent of convective influence can be observed in the large range of H_2O mixing ratios measured in the air masses; this large range is found in both older and younger stratospheric air. Additionally, the tracer-H_2O correlations show that the most recent convection as evidenced by the spike in NO (RC group) is not found in the wettest air. The wettest air, as expected, is found in the air masses with the largest amount of convective influence (the HLS + TLS* + wet group and portions of the HLS + OC group – C4 and C7, specifically).

[43] Figure 9 also shows the differences between subgroups in the HLS + OC group. C2 and C9 contain the largest influence of just old stratospheric air, while C4 and C7 show larger convective influences mixed in with old stratospheric air. C3 and C8 show somewhat "younger" old stratospheric air (i.e., air that did not come from as high of a latitude) as evidenced by reduced O_3 and NO_y and enhanced CO_2 mixing ratios. The differences between the subgroups in the TLS* + OC group, however, are not as distinct as in the HLS + OC group. This result suggests that while the subgroups are found to be statistically different within this group, there is not a clear distinction in the chemical composition of the clusters' members. As shown below, the distinction between these two subgroups appears to be primarily in their altitude, even though altitude was not used as a clustering characteristic.

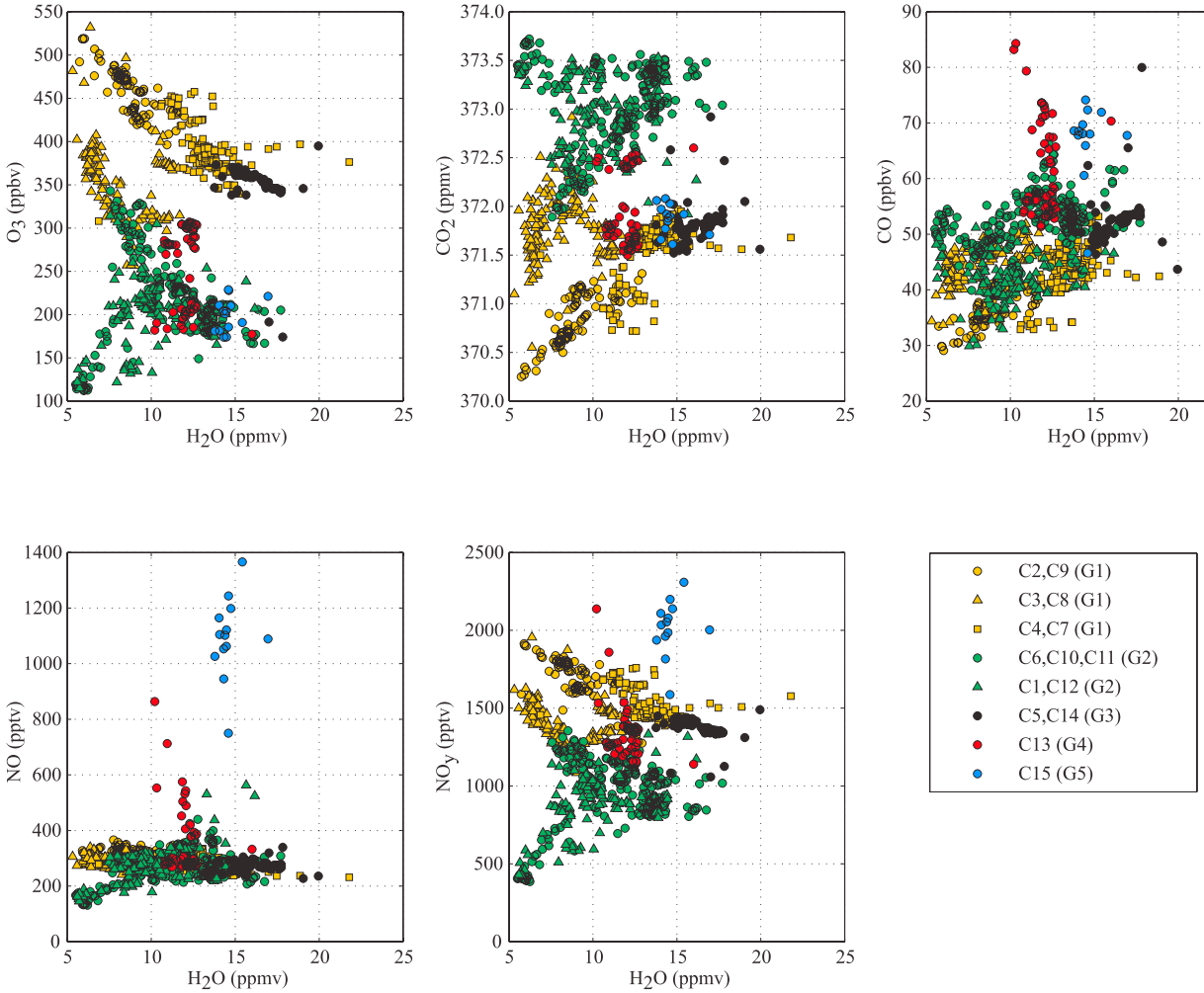

Figure 9. Correlations plots of O_3, CO_2, CO, NO, and NO_y with respect to H_2O. Color coding is based on the statistical definitions of cluster groups obtained from agglomerative hierarchical clustering shown in Figure 4. Legend shows cluster numbers followed by (group number). G1 is referred to as HLS + OC, G2 is referred to as TLS* + OC, G3 is referred to as HLS + TLS* + wet, G4 is referred to as HLS + TLS* + RC, and G5 is referred to as RC. See text for details.

[44] In addition to tracer:H_2O correlations, we examine correlations of CO_2 and NO_y with respect to O_3 in Figure 10. Because of their long lifetimes in the stratosphere, the expected nearly linear and compact correlations are observed in this data set to a large extent. The range of stratospheric age is clearly shown in the progression from the HLS + OC group (older stratospheric air characterized by higher O_3, lower CO_2, and higher NO_y) to the TLS* + OC group (younger stratospheric air characterized by lower O_3, higher CO_2, and lower NO_y). Of particular interest is the group of air masses that depart from the quasi-linear relation, namely a small fraction of the HLS + TLS* + wet group, and all of the HLS + TLS* + RC and the RC groups. In these air masses, we find the most recent convective air as evidenced by higher NO_y, which is driven by high NO, and by lower O_3 and CO_2, which is consistent with air of tropospheric character. While these correlation plots are essentially linear, the scatter around the line indicates that the stratospheric air is not pure, but that it has been influenced by convection in the past (HLS + OC, TLS* + OC, and HLS + TLS* + wet groups).

[45] The air masses that clearly depart from the quasi-linear relations in Figure 10 are of further interest, because they also comprise the group of air masses that reside on a distinct branch away from most of the observations in PC plots shown in Figure 8. In section 4.2, we mentioned that the statistical identification of the RC group (i.e., G5) is not entirely consistent with the tracer-tracer plots. On the basis of Figure 8, the RC group is characterized by older stratospheric air (+PC1) with recent convection (+PC2), but with a smaller than average extent of convective influence (−PC3). When looking at Figures 9 and 10, we find that this group has the most recent and not the largest convective influence as evidenced by the highest NO and high H_2O; however, O_3 mixing ratios clearly indicate that this group has younger stratospheric character instead (i.e., expected negative values for PC1 based on the physical interpretation given to eigenvector 1). Recall that the computation of each PC consists of taking the product of

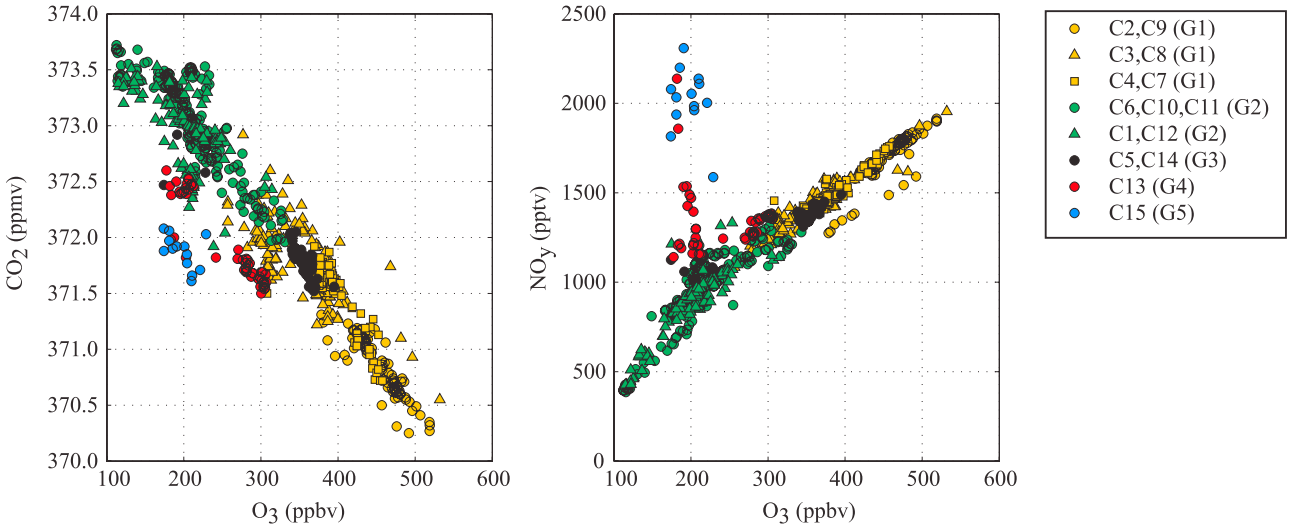

Figure 10. Correlation plots of CO_2 and NO_y with respect to O_3. Color coding is based on statistical definitions obtained from agglomerative hierarchical clustering shown in Figure 4. Legend shows cluster numbers followed by (group number). G1 is referred to as HLS + OC, G2 is referred to as TLS* + OC, G3 is referred to as HLS + TLS* + wet, G4 is referred to as HLS + TLS* + RC, and G5 is referred to as RC. See text for details.

the standardized anomaly of each observation and the coefficient in the eigenvector for the corresponding tracer (shown in Table 3) and summing up these products, in our case a total of six products, one for each tracer used in this study. The most significant tracers in eigenvector 1 are O_3, CO_2, and NO_y. In the case of older stratospheric air, these three tracers have all positive contributions to PC1, but in the case of younger stratospheric air, these three tracers have all negative contributions instead. Careful examination of each term in the products of these three tracers shows negative anomalies along with a positive weight for O_3, negative anomalies along with a negative weight for CO_2, and positive anomalies along with a positive weight for NO_y. Consequently, O_3 has a negative contribution to PC1, but both CO_2 and NO_y have positive contributions to PC1 yielding an overall positive value for PC1 instead of an expected negative value. In the RC group, NO_y is elevated not because the air mass is of older stratospheric character, but because of the convectively generated NO that drives the increase in NO_y as shown in Figures 9 and 10. O_3 is lower than average, because of its tropical origin and not because of significant dilution due to the presence of convective air, which would be inconsistent with the other tracers. CO_2 mixing ratios in this group represent average values and cannot be used to clearly discriminate between old and young stratospheric air. Similar results are obtained when carefully examining the members of the HLS + TLS* + RC group that have positive PC1 values and share the same branch with the RC group in PC space. This inconsistency between the statistical and physical analyses illustrates a shortcoming in the purely statistical technique. While mathematically any combination is possible, an accurate interpretation of the statistical results requires examination of the results in tracer-tracer space and use of multiple tracers to properly discriminate different source regions.

[46] In addition to looking at tracer mixing ratios, it is also valuable to explore the temporal evolution of groups of clusters as a function of altitude. This time series is presented in Figure 11, grouped by the flights used in this study. Recall that, as discussed above, only flights with measurements of all six tracers, in clear air, and in the lowermost stratosphere were considered. Despite the constraints on data selection, the data we were able to use span the entire period of the campaign. Two main dynamical regimes can be observed in Figure 11 based on the dominant presence of the HLS + OC group at the beginning of the campaign, and the dominant presence of the TLS* + OC group during the middle and the end of the campaign. The remaining statistical groups are found throughout the campaign, and their presence implies that air masses with more apparent convective influence from both local and nonlocal (recent and old) events were transported to our region of study during the entire campaign.

[47] Figure 11 also provides insights into the distinction between subgroups. For instance, in the HLS + OC group, the C2–C9 couplet is mostly found at higher altitudes while the C4–C7 couplet is mostly found at lower altitudes. On the other hand, the C3–C8 couplet is found at all altitudes. The subgroup in the TLS* + OC group composed of C6, C10, and C11 differs from the C1–C12 couplet in the same group in that the former is found at lower altitudes while the latter is found at higher altitudes. These altitude differences emerge as a function of the groupings formed on the basis of the tracer characteristics.

4.4. Trajectories

[48] The geographical origin and subsequent transport of the air ultimately sampled over southern Florida is examined by calculating 10-day backward trajectories with an integration time step of 30 min using the FABtraj model. The starting location and time for the trajectories vary depending on the group of clusters analyzed as shown in

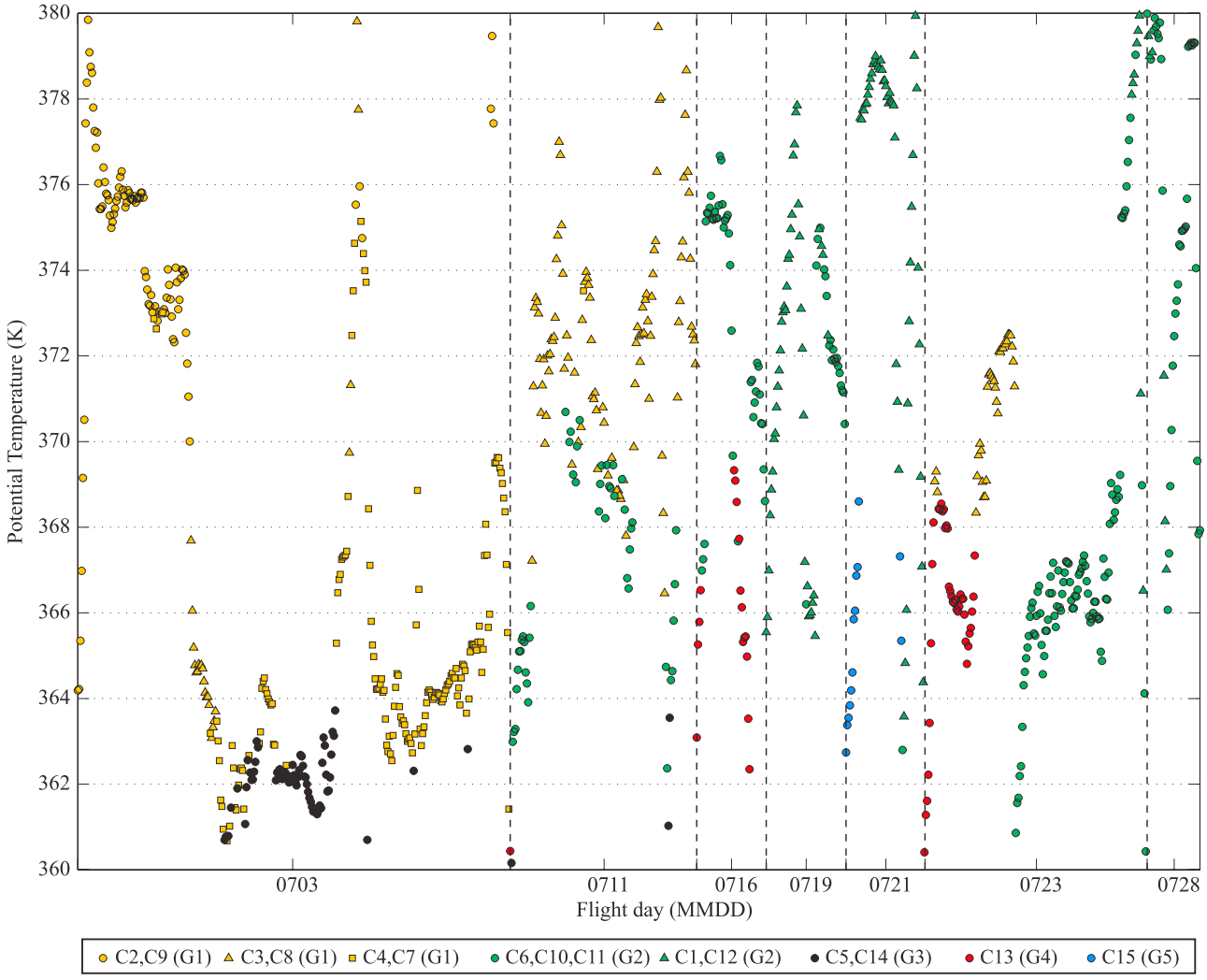

Figure 11. Time series of the altitudinal distribution of each statistical group during each flight used in this study. Legend shows cluster numbers followed by (group number). G1 is referred to as HLS + OC, G2 is referred to as TLS* + OC, G3 is referred to as HLS + TLS* + wet, G4 is referred to as HLS + TLS* + RC, and G5 is referred to as RC. See text for details.

Figure 11. Parcels are initially positioned within a volume that encompasses the latitude, longitude, and altitude of a given group of clusters sampled by the aircraft on the day of interest. This volume extends 55 km in latitude and longitude and 5 hPa in altitude past the limits of the aircraft location. All trajectories are started at the same time for any given day; therefore the presence of different subgroups of clusters at the same altitude during a given flight cannot be resolved. This constraint, however, is not significant to compromise the results of this study. All these parcels start in the lowermost stratosphere, and they are separated by 5 hPa in the vertical and 55 km in the horizontal.

[49] In order to examine the origin of each group, trajectories are initialized on: 3 July for the HLS + OC group, 23 July for the TLS* + OC group, 11 July for the HLS + TLS* + wet group, and 16 July for the HLS + TLS* + RC group. Subgrid convection is not resolved in this model; therefore trajectory calculations for the RC group are not possible, nor are they necessary because this group represents a known local convective event that can be easily identified in GOES 8 images when the flight track is overlaid. Plots of the geographical location of the air masses during the trajectory calculation along with the vertical extent of the trajectories for each of the four groups are shown in Figure 12.

[50] Overall, the results of the trajectory calculations are consistent with the results obtained from tracer composition grouped using the statistical techniques. The calculation for the HLS + OC group shows the old (high-latitude) stratospheric air component (consistent with high O_3), and the anticyclonic flow associated with the North American Monsoon responsible for the equatorward transport of these air masses as observed in Figure 12a. The calculation for the HLS + TLS* + wet group shows results similar to the HLS + OC group, but it differs in that the air masses do not reach as far north (qualitatively consistent with lower O_3 than in the HLS + OC group) and have more of a component from the subtropical Pacific as observed in Figure 12c. This calculation also shows that air masses are not coming from a preferred altitude, but instead from a large range of

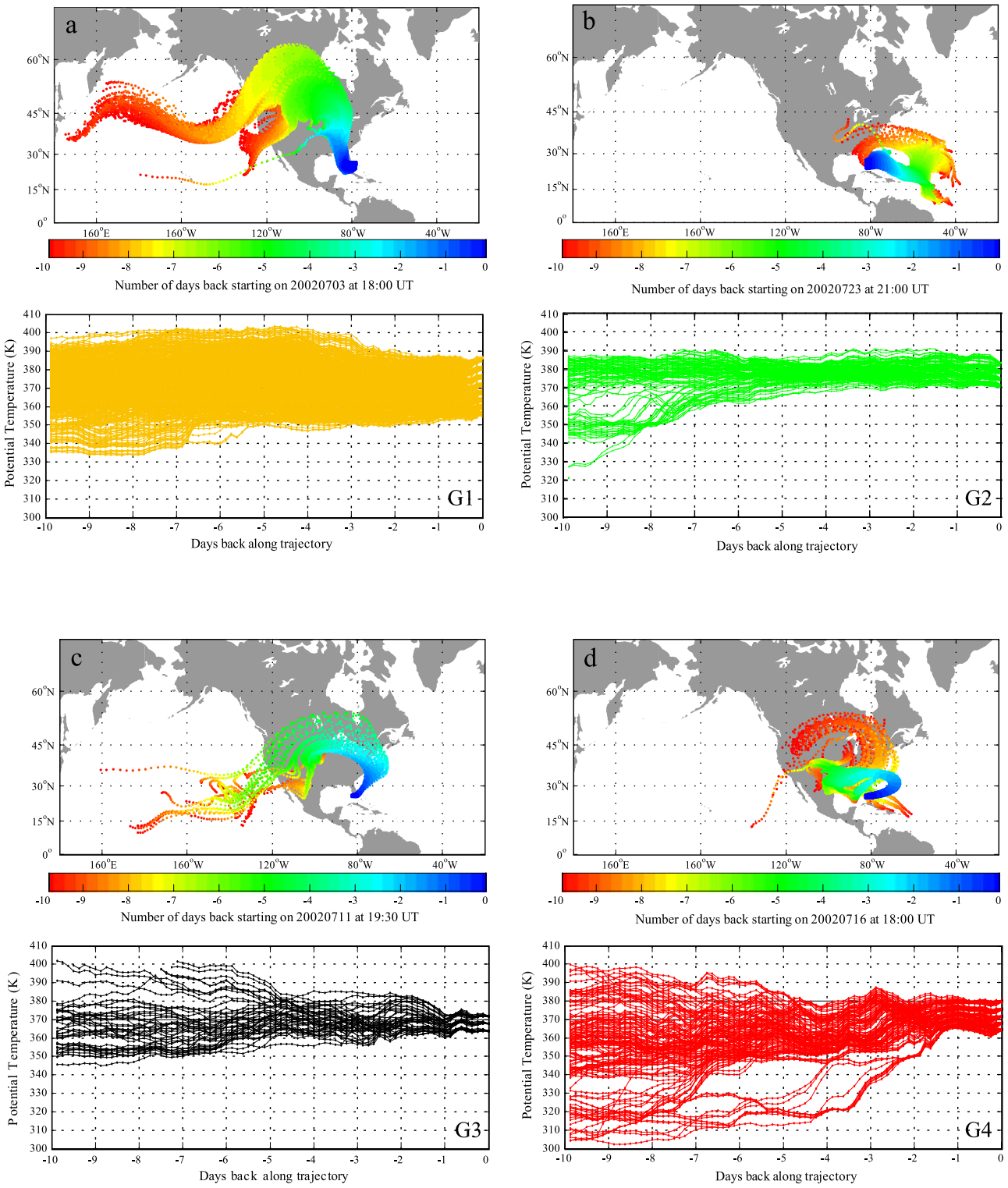

Figure 12. Results from 10-day backward trajectory calculations initialized at the altitudes and days shown in Figure 11 that correspond to four of the five cluster groups. Two plots are shown for each group. The top plot is a map with the location of the air masses along the trajectories. The colors represent the numbers of days back in time from the start of the calculation. The bottom plot shows the vertical motion of the air masses along the trajectories. (a) G1 or the HLS + OC group, (b) G2 or the TLS* + OC group, (c) G3 or the HLS + TLS* + wet group, and (d) G4 or the HLS + TLS* + RC group. See text for details.

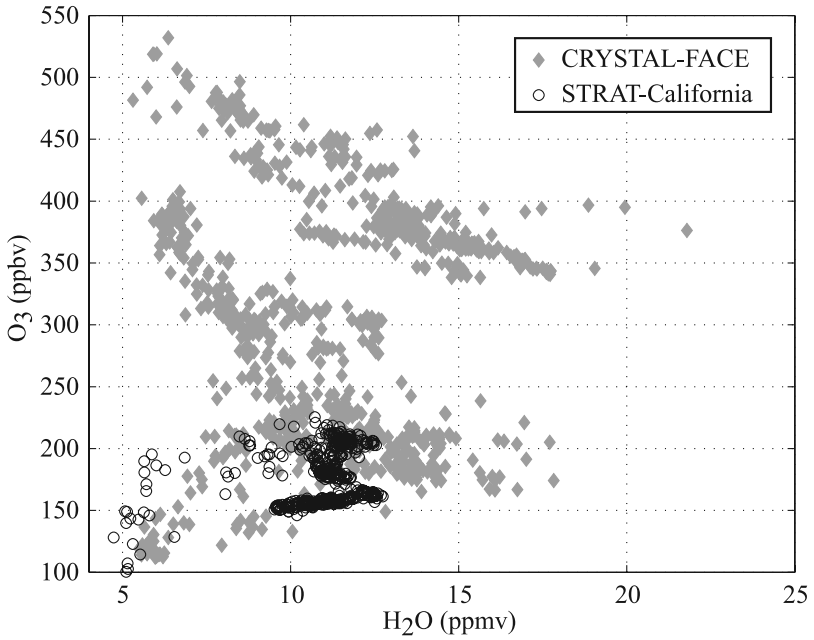

Figure 13. H_2O:O_3 correlation plot using in situ aircraft measurements in the lowermost stratosphere obtained over southern Florida during CRYSTAL-FACE in July 2002 (gray diamonds) and over California during STRAT in July 1996 (black circles). This plot illustrates the difference in chemical composition between the eastern side (CRYSTAL-FACE) and the western side (STRAT) of the North American monsoon.

altitudes. Next, the calculation for the HLS + TLS* + RC group shows the different source regions previously inferred, namely high and tropical latitudes as well as convection as observed in Figure 12d. Of particular interest, this group shows a significant influence from the lower and middle troposphere over the south central US, and from the boundary layer over the Caribbean, Gulf of Mexico, and Florida, which are all convective-rich environments during this time of the year. These results are consistent with higher NO and CO, and high H_2O.

[51] We now turn to the TLS* + OC group, which has the most complex origins. The calculation started on the day chosen to best represent the TLS* + OC group shows the tropical component of air originating from the TTL and/or the bottom of the lower stratosphere over the Atlantic as observed in Figure 12b. These results are qualitatively consistent with O_3 lower than in the HLS + OC group. Members of the TLS* + OC group are also found during the flights of 11 and 16 July, which were chosen to describe the HLS + TLS* + wet and the HLS + TLS* + RC groups, respectively. The trajectories initialized on these two flight days show some air masses originating from tropical and subtropical latitudes over the Pacific (11 July case) and mostly over the Atlantic (16 July case) before being transported to southern Florida. Air masses originating over the Pacific also travel poleward to higher latitudes before reaching Florida. The trajectory calculations thus show that some members of the TLS* + OC group have a significant tropical and subtropical origin and that the pathways traveled from the tropical Pacific or Atlantic are variable during the campaign. In addition to examining the horizontal motion of the air masses in the TLS* + OC group sampled on 23 July, we can also investigate their vertical motion. We find a significant number of the air masses that remain in the middle and upper troposphere initially over Florida and later over the subtropical and tropical Atlantic, a convective-rich environment, for several days prior to arriving to the location of sampling. The rest of the air masses show more of a quasi-adiabatic flow originating from the TTL and TLS over the tropical Atlantic consistent with lower O_3 and higher CO_2 than in the HLS + OC group.

[52] Backward trajectories show that the required equatorward transport of high-latitude stratospheric air is associated with the upper level anticyclone that developed over North America in July 2002, in agreement with the conclusions from the work of *Richard et al.* [2003]. As an important extension to that previous work, we associate the upper level anticyclone with the North American Monsoon (NAM) that develops over the continent every summer. This type of circulation is characterized by a poleward flow of air on the western side and an equatorward flow of air on the eastern side of the monsoon region as shown by *Dunkerton* [1995]. Because of the distinct chemical composition of the tropical and high-latitude UT/LS, we can use correlation plots of O_3 and H_2O to confirm the expected impact of monsoon circulation on the distribution of these two chemical tracers in the UT/LS region.

[53] Figure 13 shows the O_3:H_2O correlation plot using in situ aircraft data collected over the eastern side of the NAM (i.e., Florida during CRYSTAL-FACE) and over the western side of the NAM (i.e., California during STRAT) in the lowermost stratosphere. In addition to their location of sampling, these two aircraft campaigns shared similar zonal boundaries (i.e., the jet stream on the north side, and the subtropical tropopause on the south side). During the summertime, the subtropical jet stream

is weak because of a significant decrease in the temperature gradient between the tropics and the midlatitudes. This condition implies that transport across the tropopause in the subtropics is not inhibited as is the case during the winter. A climatology study of mixing along the subtropical tropopause as revealed by Rossby Wave Breaking (RWB) events showed that the mixing was indeed more frequent during the summertime and in particular over regions east of the upper level subtropical high associated with monsoon circulation, which together with the weak PV gradients observed during the summer provide favorable conditions for these mixing events [*Postel and Hitchman*, 1999]. During both campaigns, NCEP-assimilated horizontal winds showed the presence of only one jet stream located north of 40°N during STRAT, and between 45°N and 55°N during CRYSTAL-FACE. The observations shown in Figure 13 are consistent with the expected contrast between west and east in the lowermost stratosphere, namely drier air with lower O_3 due to poleward transport of tropical air on the west (black circles) versus wetter air due to transport over continental convective-rich regions and higher O_3 due to equatorward transport of high-latitude stratospheric air on the east (gray diamonds) in agreement with the results of *Gettelman et al.* [2004] and *Dessler and Sherwood* [2004]. Therefore summertime transport in the lowermost stratosphere can no longer be thought of as strictly meridional (i.e., 2-D); the zonal component of the flow also becomes critical at least where influenced by monsoon circulation consistent with the results of *Dunkerton* [1995].

4.5. Utility of the Method

[54] The combination of cluster analysis, PCA, tracer-tracer correlations, and trajectory calculations provides an objective methodology in which a large data set of in situ aircraft observations can be reduced to few, natural groupings that capture the most significant variability of the data set. In addition to qualitatively describing a data set, quantification of the resulting clusters is desired. We have developed a simple mixing model that uses in situ measurements and tracer profiles from selected source regions to quantify the contribution of each of those source regions (E. M. Weinstock et al., Quantifying the impact of the North American monsoon and deep midlatitude convection on the subtropical lowermost stratosphere using in situ measurements, submitted to *Journal of Geophysical Research*, 2007, hereinafter referred to as Weinstock et al., submitted manuscript, 2007). The model calculations are performed on tracer mixing ratios averaged over several flights and using prescribed convective profiles that best represent the cumulative convective perturbation.

[55] The methodology presented in this study can be used as a data filter for the model. We find consistency in the composition of the air mass between the qualitative results presented in this study and the quantitative results obtained from the mixing model. For instance, model results for C2, which is a cluster member of the HLS + OC group, show 80–90% high-latitude stratospheric air, 10–15% tropical stratospheric air, and about 3% convective air. Similarly, model results for C6, which is a cluster member of the TLS* + OC group, show 80% tropical stratospheric air, 15% high-latitude stratospheric air, and 3% convective air. Furthermore, the modeled results of the composition of air masses do not vary significantly within each cluster member and the characterizations of the air mass for different cluster members of the same group (e.g., C2 versus C3 in the HLS + OC group) are consistent as well. This consistency between expected fractions of source regions based on the objective clustering and the actual values obtained from the model constitutes an independent way of validating the model results.

[56] In addition to validating the model results, this methodology could be used to differentiate convective inputs in an objective way and provide a better, more constrained fit to the convective fraction. This can be accomplished by separately analyzing groups of air masses that have been presumably impacted by convective events of different character rather than assigning one convective source term to represent multiple convective events.

5. Conclusions

[57] In this study we qualitatively identified the latitudinal origin and subsequent transport pathways of air sampled in the subtropical lowermost stratosphere during CRYSTAL-FACE using an objective methodology based on statistical analysis, tracer-tracer correlation plots with six different chemical tracers and backward trajectory calculations. We found that the four possible source regions illustrated in Figure 1 were appropriate to describe the transport processes affecting our region of study.

[58] Despite the relatively small size of the data set, the statistical techniques used and the chosen number of clusters proved to be adequate for this task. We are able to statistically identify and differentiate the presumed dominant source regions for the air ultimately sampled during CRYSTAL-FACE. The majority of the clusters do not have unique origins, which implies that there is significant horizontal and/or vertical mixing taking place prior to the arrival of these air masses to our region of sampling. PCA shows that most of the variability in the data set can be explained primarily by the mean age of air in the stratosphere, followed by the age of the convective influence, and last by the extent of the convective influence, which can be associated with the latitude of convective injection. While the statistical approach is objective, it is important to remember that it is a purely mathematical technique, and an accurate interpretation can only be accomplished by incorporating information based on actual mixing ratios and tracer-tracer correlations plots as we found to be the case for the RC group and members of the HLS + TLS* + RC group.

[59] We find that different source regions are dominant during different periods of time. At the beginning of the campaign, the dominant signature is of high-latitude stratospheric air while the rest of the campaign is dominated by the presence of tropical air. On the basis of comparisons with measurements from previous campaigns, we also find that the majority of the air masses with tropical signature originate from the TLS rather than the TTL. In some cases, however, air masses have properties of consistent with either the upper TTL or the lower TLS. Because this boundary is not very distinct in terms of tracer mixing

ratios, we cannot exclude the possibility of TTL air present in our region of study.

[60] Very moist air is ubiquitously present during the entire campaign and throughout the subtropical lowermost stratosphere. Enhancements in H_2O were associated with a local convective event in one case only (21 July) as also evidenced by the significant spike in NO. The rest of the time, air masses with up to 22 ppmv of H_2O (clear convective signature) contained less than 400 pptv of NO, which suggests nonlocal convective air that came from the tropics or that was injected in the UT/LS region at higher latitudes and it was subsequently transported over the course of a few days to a week into our region of sampling.

[61] Backward trajectory calculations confirmed the latitudinal origin of the air masses sampled during the campaign. The trajectories also showed that monsoon circulation plays a critical role in the distribution of key radiative species such as H_2O and O_3 in the UT/LS. Therefore the canonical view of a purely meridional (i.e., 2-D) large-scale circulation linking the stratosphere and the troposphere cannot be treated as universal. Over monsoon regions, both meridional and zonal transport pathways need to be considered as well as the longitude of the meridional transport, which can be either poleward or equatorward.

[62] Last, the methodology presented in this study was used to compare results with a mixing that provides the fractional contribution of each source region (Weinstock et al., submitted manuscript, 2007). We found consistent results between the qualitatively identified source regions obtained in this study and the model-derived contribution of each source region.

[63] **Acknowledgments.** We are very grateful to the WB-57 flight and ground crews as well as the CRYSTAL-FACE mission planners for their hard work in making this campaign possible. We thank the NOAA Climate Monitoring and Diagnostics Laboratory for providing surface measurements. We thank D. Boccippio for providing insightful comments on the statistical analysis presented in this work. We are also grateful to P. Robertson, T. Miller, and the anonymous reviewers for their helpful comments and suggestions. J. V. Pittman acknowledges the NASA Earth System Science Graduate Student Fellowship (grant NGT5-3040) for their support during the course of this research project. This work was funded in part by the NASA Upper Atmospheric Research Program. Work performed by M.J.M. at the Jet Propulsion Laboratory, California Institute of Technology, was done under contract with NASA.

References

Boccippio, D. J., W. A. Petersen, and D. J. Cecil (2005), The tropical convective spectrum: Part I: Archetypal vertical structures, *J. Clim.*, *18*, 2744–2769.

Boering, K. A., S. C. Wofsy, B. C. Daube, H. R. Schneider, M. Loewenstein, J. R. Podolske, and T. J. Conway (1996), Stratospheric mean ages and transport rates from observations of carbon dioxide and nitrous oxide, *Science*, *274*, 1340–1343.

Brasseur, G., and S. Solomon (1986), *Aeronomy of the Middle Atmosphere: Chemistry and Physics of the Stratosphere and Mesosphere*, Dordrecht, Boston, Mass.

Brewer, A. W. (1949), Evidence for a world circulation provided by measurements of helium and water vapor in the stratosphere, *Q. J. Roy. Meteorol. Soc.*, *75*, 351–363.

Chen, P. (1995), Isentropic cross-tropopause mass-exchange in the extratropics, *J. Geophys. Res.*, *100*, 16,661–16,673.

Cooper, O., et al. (2004a), A case study of trans-Pacific warm conveyor belt transport: The influence of merging airstreams on trace gas import to North America, *J. Geophys. Res.*, *109*, D23S08, doi:10.1029/2003JD003624.

Cooper, O. R., et al. (2004b), On the life-cycle of a stratospheric intrusion and its dispersion into polluted warm conveyor belts, *J. Geophys. Res.*, *109*, D23S09, doi:10.1029/2003JD004006.

Daube, B. C. J., K. A. Boering, A. E. Andrews, and S. C. Wofsy (2002), A high-precision fast-response airborne CO_2 analyzer for in situ sampling from the surface to the middle stratosphere, *J. Atmos. Ocean Technol.*, *19*, 1532–1543.

Denning, R. F., S. L. Guidero, G. S. Parks, and B. L. Gary (1989), Instrument description of the airborne microwave temperature profiler, *J. Geophys. Res.*, *94*, 16,757–16,765.

Dessler, A. E., and S. C. Sherwood (2004), The effect of convection on the summertime extratropical lower stratosphere, *J. Geophys. Res.*, *109*, D23301, doi:10.1029/2004JD005209.

Dessler, A. E., E. M. Weinstock, E. J. Hintsa, J. G. Anderson, C. R. Webster, R. D. May, J. W. Elkins, and G. S. Dutton (1994), An examination of the total hydrogen budget of the lower stratosphere, *Geophys. Res. Lett.*, *21*, 2563–2566.

Dessler, A. E., E. J. Hintsa, E. M. Weinstock, J. G. Anderson, and K. R. Chan (1995), Mechanisms controlling water vapor in the lower stratosphere: A tale of two stratospheres, *J. Geophys. Res.*, *100*, 23,167–23,172.

Dethof, A., A. O'Neill, and J. Slingo (2000), Quantification of the isentropic mass transport across the dynamical tropopause, *J. Geophys. Res.*, *105*, 12,279–12,293.

Dobson, G. M. B. (1956), Origin and distribution of the polyatomic molecules in the atmosphere, *Proc. R. Soc. London, Ser. A*, *236*, 187–193.

Dunkerton, T. J. (1995), Evidence of meridional motion in the summer lower stratosphere adjacent to monsoon regions, *J. Geophys. Res.*, *100*, 16,675–16,688.

Fahey, D. W., K. K. Kelly, D. M. Murphy, G. V. Ferry, M. Loewenstein, K. R. Chan, L. R. Poole, and J. C. Wilson (1989), In situ measurements of total reactive nitrogen, total water, and aerosol in a polar stratospheric cloud in the Antarctic, *J. Geophys. Res.*, *94*, 11,299–11,315.

Fahey, D. W., et al. (1996), In situ observations of NO_y, O_3, and the NO_y/O_3 ratio in the lower stratosphere, *Geophys. Res. Lett.*, *23*, 1653–1656.

Gettelman, A., and P. M. D. Forster (2002), A climatology of the tropical tropopause layer, *J. Meteorol. Soc. Jpn.*, *80*, 911–924.

Gettelman, A., D. E. Kinnison, T. J. Dunkerton, and G. P. Brasseur (2004), Impact of monsoon circulations on the upper troposphere and lower stratosphere, *J. Geophys. Res.*, *109*, D22101, doi:10.1029/2004JD004878.

Hargrove, W. W., and F. M. Hoffman (2004), The potential of multivariate quantitative methods for delineation and visualization of ecoregions, *Environ. Manage.*, *34*, S39–S60.

Hintsa, E. J., et al. (1998), Troposphere-to-stratosphere transport in the lowermost stratosphere from measurements of H_2O, CO_2, N_2O and O_3, *Geophys. Res. Lett.*, *25*, 2655–2658.

Hoffman, F. M., W. W. Hargrove, D. J. Erickson III, and R. J. Oglesby (2005), Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models, *Earth Interact.*, *9*, EI110, doi:10.1175/EI110.1.

Holton, J. R., P. H. Haynes, M. E. McIntyre, A. R. Douglass, R. B. Rood, and L. Pfister (1995), Stratosphere-troposphere exchange, *Rev. Geophys*, *33*, 403–439.

Hoor, P., H. Fischer, L. Lange, J. Lelieveld, and D. Brunner (2002), Seasonal variations of a mixing layer in the lowermost stratosphere as identified by the CO-O_3 correlation from in situ measurements, *J. Geophys. Res.*, *107*(D5), 4044, doi:10.1029/2000JD000289.

Hoskins, B. J. (1991), Towards a PV-Theta view of the general circulation, *Tellus, Ser. A*, *43*, 27–35.

Hurst, D. F., et al. (1999), Closure of the total hydrogen budget of the northern extratropical lower stratosphere, *J. Geophys. Res.*, *104*, 8181–8200.

Jensen, E., D. Starr, and O. B. Toon (2004), Mission investigates tropical cirrus clouds, *Eos Trans. AGU*, *85*, 45.

Jost, H. J., et al. (2004), In situ observations of mid-latitude forest fire plumes deep in the stratosphere, *Geophys. Res. Lett.*, *31*, L11101, doi:10.1029/2003GL019253.

Loewenstein, M., H. Jost, J. Grose, J. Eilers, D. Lynch, S. Jensen, and J. Marmie (2002), Argus: A new instrument for the measurement of the stratospheric dynamical tracers, N_2O and CH_4, *Spectrochim. Acta A*, *58*, 2329–2345.

Mote, P. W., K. H. Rosenlof, M. E. McIntyre, E. S. Carr, J. C. Gille, J. R. Holton, J. S. Kinnersley, H. C. Pumphrey, J. M. Russell, and J. W. Waters (1996), An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor, *J. Geophys. Res.*, *101*, 3989–4006.

Murphy, D. M., D. W. Fahey, M. H. Proffitt, S. C. Liu, K. R. Chan, C. S. Eubank, S. R. Kawa, and K. K. Kelly (1993), Reactive nitrogen and its correlation with ozone in the lower stratosphere and upper troposphere, *J. Geophys. Res.*, *98*, 8751–8773.

Plumb, R. A. (2002), Stratospheric transport, *J. Meteorol. Soc. Jpn.*, *80*, 793–809.

Plumb, R. A., and M. K. W. Ko (1992), Interrelationships between mixing ratios of long lived stratospheric constituents, *J. Geophys. Res.*, *97*, 10,145–10,156.

Postel, G. A., and M. H. Hitchman (1999), A climatology of Rossby wave breaking along the subtropical tropopause, *J. Atmos. Sci.*, *56*, 359–373.

Poulida, O., R. R. Dickerson, and A. Heymsfield (1996), Stratosphere-troposphere exchange in a midlatitude mesoscale convective complex, *J. Geophys. Res.*, *101*, 6823–6836.

Prados, A. I., G. E. Nedoluha, R. M. Bevilacqua, D. R. Allen, K. W. Hoppel, and A. Marenco (2003), POAM III ozone in the upper troposphere and lowermost stratosphere: Seasonal variability and comparisons to aircraft observations, *J. Geophys. Res.*, *108*(D7), 4218, doi:10.1029/2002JD002819.

Proffitt, M. H., and R. J. McLaughlin (1983), Fast-response dual-beam UV-absorption ozone photometer suitable for use on stratospheric balloons, *Rev. Sci. Instrum.*, *54*, 1719–1728.

Ray, E. A., et al. (2004), Evidence of the effect of summertime midlatitude convection on the subtropical lowermost stratosphere from CRYSTAL-FACE tracer measurements, *J. Geophys. Res.*, *109*, D18304, doi:10.1029/2004JD004655.

Richard, E. C., K. C. Aikin, E. A. Ray, K. H. Rosenlof, T. L. Thompson, A. Weinheimer, D. Montzka, D. Knapp, B. Ridley, and A. Gettelman (2003), Large-scale equatorward transport of ozone in the subtropical lower stratosphere, *J. Geophys. Res.*, *108*(D23), 4714, doi:10.1029/2003JD003884.

Ridley, B. A., J. G. Walega, J. E. Dye, and F. E. Grahek (1994), Distributions of NO, NO_x, NO_y, and O_3 to 12-km altitude during the summer monsoon season over New Mexico, *J. Geophys. Res.*, *99*, 25,519–25,534.

Ridley, B. A., J. E. Dye, J. G. Walega, J. Zheng, F. E. Grahek, and W. Rison (1996), On the production of active nitrogen by thunderstorms over New Mexico, *J. Geophys. Res.*, *101*, 20,985–21,005.

Ridley, B. A., et al. (2004), Florida thunderstorms: A faucet of reactive nitrogen to the upper troposphere, *J. Geophys. Res.*, *109*, D17305, doi:10.1029/2004JD004769.

Shah, K. P., and D. Rind (1998), Comparing upper tropospheric and lower stratospheric temperatures: Microwave sounding unit, radiosonde, COSPAR International Reference Atmosphere, and National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis monthly mean climatologies, *J. Geophys. Res.*, *103*, 31,569–31,592.

Stohl, A., et al. (2003), Stratosphere-troposphere exchange: A review, and what we have learned from STACCATO, *J. Geophys. Res.*, *108*(D12), 8516, doi:10.1029/2002JD002490.

Strahan, S. E. (1999), Climatologies of lower stratospheric NOy and O_3 and correlations with N_2O based on in situ observations, *J. Geophys. Res.*, *104*, 30,463–30,480.

Strahan, S. E., A. R. Douglass, J. E. Nielsen, and K. A. Boering (1998), The CO_2 seasonal cycle as a tracer of transport, *J. Geophys. Res.*, *103*, 13,729–13,742.

Thompson, T. L. and K. H. Rosenlof (2003), Accuracy and precision of the NOAA Aeronomy Laboratory pressure temperature instrument on the NASA WB-57F, paper presented at CRYSTAL-FACE Science Team Meeting, Natl. Aeronaut. Space Admin., Salt Lake City, Utah.

Vaughan, G., and C. Timmis (1998), Transport of near-tropopause air into the lower midlatitude stratosphere, *Q J. Roy. Meteorol. Soc.*, *124*, 1559–1578.

Webster, C. R., R. D. May, C. A. Trimble, R. G. Chave, and J. Kendall (1994), Aircraft (ER-2) Laser Infrared Absorption Spectrometer (ALIAS) for in-situ stratospheric measurements of HCl, N_2O, CH_4, NO_2, and HNO_3, *Appl. Opt.*, *33*, 454–472.

Weinstock, E. M., C. M. Schiller, and J. G. Anderson (1986), In situ stratospheric ozone measurements by long path UV absorption–Developments and interpretation, *J. Geophys. Res.*, *91*, 5237–5248.

Weinstock, E. M., E. J. Hintsa, A. E. Dessler, J. F. Oliver, N. L. Hazen, J. N. Demusz, N. T. Allen, L. B. Lapson, and J. G. Anderson (1994), New fast-response photofragment fluorescence hygrometer for use on the NASA ER-2 and the Perseus remotely piloted aircraft, *Rev. Sci. Instrum.*, *65*, 3544–3554.

Weinstock, E. M., E. J. Hintsa, D. B. Kirk-Davidoff, J. G. Anderson, A. E. Andrews, R. L. Herman, C. R. Webster, M. Loewenstein, J. R. Podolske, and T. P. Bui (2001), Constraints on the seasonal cycle of stratospheric water vapor using in situ measurements from the ER-2 and a CO photochemical clock, *J. Geophys. Res.*, *106*, 22,707–22,724.

Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, Elsevier, New York.

J. G. Anderson, B. C. Daube, D. S. Sayres, J. B. Smith, E. M. Weinstock, and S. C. Wofsy, Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA.

O. R. Cooper, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA.

C. Gerbig, Max Planck Institute for Biogeochemistry, D-07745 Jena, Germany.

W. W. Hargrove and F. M. Hoffman, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.

H.-J. Jost, NovaWave Technologies, Redwood City, CA 94065, USA.

M. Loewenstein, NASA Ames Research Center, Moffett Field, CA 94035, USA.

J. P. Lopez, Bay Area Environmental Research Institute, Sonoma, CA 95476, USA.

M. J. Mahoney, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA.

R. J. Oglesby, Department of Geosciences, University of Nebraska, Lincoln, NE 68588, USA.

J. V. Pittman, NASA Marshall Space Flight Center, Huntsville, AL 35805, USA. (jasna.pittman@nasa.gov)

E. C. Richard and T. L. Thompson, Earth System Research Laboratory, NOAA, Boulder, CO 80305, USA.

B. A. Ridley, and A. J. Weinheimer, Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, CO 80305, USA.

I. Xueref, Laboratoire des Sciences du Climat et de l'Environment, Commissariat l'Energie Atomique, F-91191 Gif-Sur-Yvette Cedex, France.