Survey of whole air data from the second airborne Biomass Burning and Lightning Experiment using principal component analysis

Yunsoo Choi,[1] Scott Elliott,[2] Isobel J. Simpson,[1] Donald R. Blake,[1] Jonah J. Colman,[2] Manvendra K. Dubey,[2] Simone Meinardi,[1] F. Sherwood Rowland,[1] Tomoko Shirai,[3] and Felisa A. Smith[4]

[1] Hydrocarbon and halocarbon measurements collected during the second airborne Biomass Burning and Lightning Experiment (BIBLE-B) were subjected to a principal component analysis (PCA), to test the capability for identifying intercorrelated compounds within a large whole air data set. The BIBLE expeditions have sought to quantify and understand the products of burning, electrical discharge, and general atmospheric chemical processes during flights arrayed along the western edge of the Pacific. Principal component analysis was found to offer a compact method for identifying the major modes of composition encountered in the regional whole air data set. Transecting the continental monsoon, urban and industrial tracers (e.g., combustion byproducts, chlorinated methanes and ethanes, xylenes, and longer chain alkanes) dominated the observed variability. Pentane enhancements reflected vehicular emissions. In general, ethyl and propyl nitrate groupings indicated oxidation under nitrogen oxide (NO_x) rich conditions and hence city or lightning influences. Over the tropical ocean, methyl nitrate grouped with brominated compounds and sometimes with dimethyl sulfide and methyl iodide. Biomass burning signatures were observed during flights over the Australian continent. Strong indications of wetland anaerobics (methane) or liquefied petroleum gas leakage (propane) were conspicuous by their absence. When all flights were considered together, sources attributable to human activity emerged as the most important. We suggest that factor reductions in general and PCA in particular may soon play a vital role in the analysis of regional whole air data sets, as a complement to more familiar methods.   *INDEX TERMS:* 0322 Atmospheric Composition and Structure: Constituent sources and sinks; 0345 Atmospheric Composition and Structure: Pollution—urban and regional (0305); 0368 Atmospheric Composition and Structure: Troposphere—constituent transport and chemistry; 0365 Atmospheric Composition and Structure: Troposphere—composition and chemistry; 3299 Mathematical Geophysics: General or miscellaneous; *KEYWORDS:* whole air samples, hydrocarbons, halocarbons, principal component analysis, BIBLE-B

1. Introduction

[2] Hydrocarbon detection systems that couple whole air samples with gas chromatography have become capable of resolving a large array of individual chemical species. Several groups can now measure dozens of organic compounds in air retrieved within a single canister [*Apel et al.*, 1999, and references therein]. As an example, in urban, regional and global scale experiments conducted from laboratories at the University of California, Irvine (UCI), the number of reportable species has risen steadily over the last two decades [e.g., *Blake et al.*, 1984, 1992, 1994, 1997, 2003; *Colman et al.*, 2001; *Simpson et al.*, 2001]. Concentrations can be quantified for molecules in all the major hydrocarbon classes including the alkanes, alkenes, alkynes and aromatics. Branched chain and nitrate- or halogen-substituted derivatives can also be identified within the spectra. The UCI group currently monitors on the order of 100 separate volatile organics routinely, from 50–160 samples per aircraft flight or site-visit. As many as 50 flights or visits may occur in a typical working year. Clearly, for this one group alone the data analysis task has become

---

[1]Department of Chemistry, University of California-Irvine, Irvine, California, USA.

[2]Atmospheric and Climate Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

[3]National Space Development Agency, Tokyo, Japan.

[4]Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA.

both a tremendous challenge and an opportunity. The same point can be made with regard to the hydrocarbon measurement community at large.

[3] The UCI group has to date interpreted its whole air measurements primarily through uni- or bivariate comparisons in which concentrations are examined as distributions for single species or plots relating pairs of species [e.g., *Blake et al.*, 1992, 1996, 1997; *Colman et al.*, 1998]. The data sets have reached a size, however, for which multivariate statistical and chemometric strategies become appropriate. A new goal is thus to implement chemically multidimensional approaches to interpret large data sets. Linear algebraic tools are well established for the simultaneous assessment of *n*-species relationships in analytical chemistry [*Hopke*, 1985, 1991; *Sharaf et al.*, 1986; *Massart et al.*, 1988; *Malinowski*, 1991]. However, these tools need to be tailored in order to meet the needs of a laboratory studying atmospheric composition over a variety of scales. It is anticipated that a suite of computations can be developed that will complement the traditional low-dimensionality visualizations and graphics.

[4] Recent participation by UCI in the second Japanese-sponsored Biomass Burning and Lightning Experiment (BIBLE-B) has coincided with planning of the multivariate effort. In the present work we adopt BIBLE-B measurements as a platform for the initial application of large data set factor reduction calculations. In particular, principal component analysis (PCA) [*Blifford and Meeker*, 1967; *Mulaik*, 1972; *Hopke*, 1985; *Malinowski*, 1991] is assessed as a survey tool for tropospheric whole air measurements (hydrocarbons, halocarbons, alkyl nitrates and sulfur compounds) obtained on aircraft campaigns. Previous studies have applied PCA on a local scale in city pollution studies [*Henry and Hidy*, 1979; *Thurston and Spengler*, 1985; *Statheropoulos et al.*, 1998; *Yu and Chang*, 2000]. This investigation is the first full attempt by our group to apply multivariate statistics to a regional whole air data set, and advantages and drawbacks of the approach are discussed.

2. Mission and Whole Air Measurements

[5] Two BIBLE aircraft campaigns have now been completed, sponsored by the Japanese National Space Development Agency (NASDA) and involving analytical atmospheric chemistry groups from around the world. A Gulfstream jet aircraft has served both times as the mission platform. Its primary competency is at high altitudes, and BIBLE has exploited this skill by focusing on free tropospheric air chemistry. An early goal of BIBLE was to characterize the chemistry of Indonesian biomass burning plumes. As a result, flight tracks have generally taken the aircraft south from Japan along the Pacific Rim, with intensive campaigns radiating from tropical headquarter sites. During BIBLE-A, key flights began and ended on the island of Java, and major burning plumes were intercepted over Kalimantan. Although the Gulfstream is not designed for low altitude work, descent over the open Java Sea yielded striking marine signatures in the UCI whole air measurements. The fingerprint of the dense human population of Java was also evident. The economy of the island is gasoline-driven and based on automotive transportation, such that vehicle exhaust could readily be identified in the
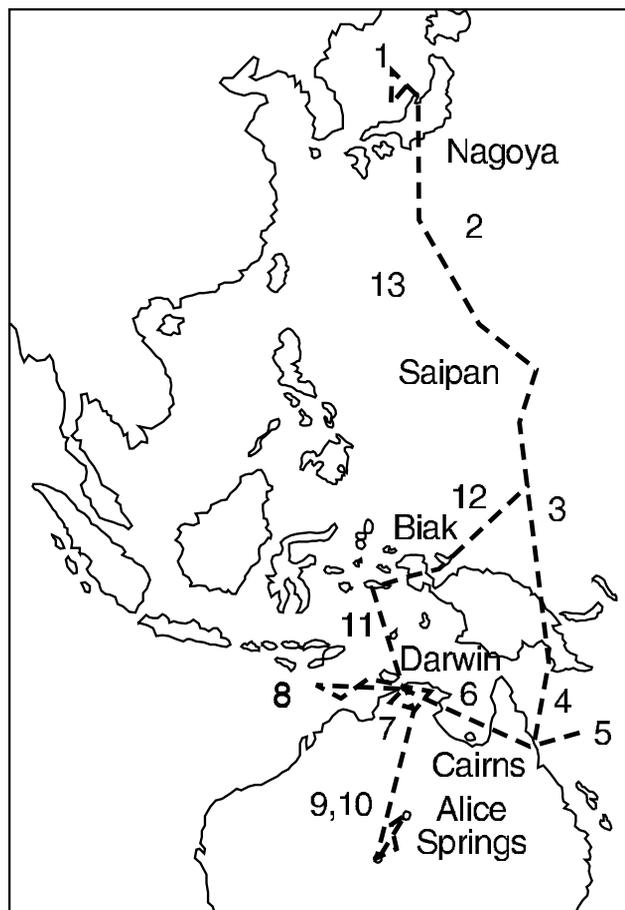

Figure 1. Flights 1–13 of the second Biomass Burning and Lightning Experiment (BIBLE-B), conducted from August–September, 1999.

proximity of Java. The BIBLE-A mission is described in detail in a set of papers by the groups involved [e.g., *Elliott et al.*, 2003; *Kondo et al.*, 2003].

[6] Flight paths for the second experiment in the series, BIBLE-B, are shown in Figure 1. The mission was conducted from August 28 – September 14, 1999. The flight geography is similar to that of BIBLE-A, but with the port of Darwin, Australia (Northern Territory) acting as the primary tropical hub. The BIBLE-B experiment constituted a transect of Northeast Asian, Asian Pacific, Indonesian and Australasian air masses. Detailed meteorology of the Western Pacific Rim has been described in key papers of the Pacific Exploratory Mission (PEM) West series [*Bachmeier et al.*, 1996; *Merrill*, 1996; *Merrill et al.*, 1997] and more recently in BIBLE-A work [*Elliott et al.*, 2003]. Briefly, atmospheric interchange between the North Pacific and Asia is driven by geostrophic winds aloft, and by monsoon (onshore) flow at the surface in summer. Therefore, in the present work we expect transit flights from and to Japan to display continental influences at upper levels, whereas the distributions of maritime species should be reflected nearer the surface. Moving southward, a transition to the tropical tropospheric environment is typically observed at about 25°N [*Blake et al.*, 1997; *Elliott et al.*, 2003]. Indonesian convection is associated with more thorough trace gas

mixing, and a monsoonal component may also lead to interhemispheric transfer [*Kao*, 1948; *Thompson*, 1951].

[7] Instruments on board the Gulfstream included detectors for ozone, the gas phase nitrogen oxides and the bulk aerosol. In addition, the Gulfstream carried UCI apparatus for filling multiple 2-L stainless steel canisters with whole air samples [e.g., *Blake et al.*, 1992, 1997]. Between 50–60 canisters were filled per flight. The filled sample containers were shipped immediately back to the home laboratory at UCI. Following standard group protocol, several aliquots were taken from each can. These were distributed to multiple gas chromatographs/detectors designed to separate and quantify halocarbons, alkyl nitrates, dimethyl sulfide (DMS), and families of hydrocarbons such as the alkanes or alkenes. Flame ionization, electron capture and mass spectroscopy were all employed as means of detection. Processing the chromatographic results includes a quality control of the baselines. The analytical procedure allows almost 200 different species to be identified and quantified. On a typical flight at low to middle altitudes, many dozens of compounds are present at levels above their detection limit. Complete analytical details are presented by *Colman et al.* [2001].

3. Univariate and Bivariate Interpretation

[8] The examination of whole air measurements from an aircraft mission typically begins with a collection of plots that show species concentration versus day of year or flight number. This allows background values and enhancements over them to be quickly identified. We have performed such an inspection for the complete suite of BIBLE-B measurements. Measured concentrations show the lowest variability for transit flights in which there are extended flight legs in the upper troposphere. By contrast, the concentrations of sea-air tracers are enhanced for flights with segments in the marine boundary layer, and in general at low latitudes because of the influence of active convective systems near Indonesia. Urban tracers are most evident for flights emanating from the larger cities (Nagoya, Cairns, Darwin). Biomass burning products such as methyl chloride (CH_3Cl) reach a maximum over the interior of the Australian Northern Territory. Such concentration versus time plots are univariate in that they embody the measurements independently for each detected species. Although several compounds may be plotted together, the human eye must then search for linkages within the overall patterns.

[9] A second step in the assessment involves plotting mission data in bivariate chemical spaces, with plot markers for the individual flights and for key parameters such as altitude. Ethane is often chosen as the unifying correlate (*x*-axis) because of its intermediate residence time on the order of weeks in the free troposphere [*Singh and Zimmerman*, 1992]. This is long enough for it to be readily detectable, but short enough for significant variability to exist in its distribution. Convenient source tracers are plotted against ethane in the 2-D concentration space, and a quadrant strategy is applied to the visual inspection of each graph. For example, the upper right quadrant of any particular graph is inspected for doubly high concentrations. A typical series of *y*-axis test compounds includes CH_3Cl and methyl iodide (CH_3I), followed by tetrachloroethene (C_2Cl_4). Cor-

relation between CH_3Cl and ethane indicates biomass burning emissions. High CH_3I concentrations located in the upper left hand quadrant show marine boundary layer input. Tetrachloroethene is a unique indicator for industrial air because it is believed to have only anthropogenic sources from its use as an industrial solvent. Among the BIBLE-B data, flights directed from Darwin south and southeastward showed both large CH_3Cl and ethane concentrations, indicating the interaction of these air masses with biomass burning plumes. At the same time, CH_3I levels were well above their background values and were independent of ethane, indicating that oceanic influence must also have been present.

[10] In final analyses, numerous concentration versus flight time, altitude and concentration plots are constructed [e.g., *Blake et al.*, 1997; *Elliott et al.*, 2003]. Linear regressions may be performed, and the slopes of concentration-concentration plots may be used to estimate emission ratios [*Blake et al.*, 1996]. We note that in the manipulations that have been discussed in this section, relationships are not sought between more than two chemical substances simultaneously.

4. Theoretical Considerations

[11] Our goal here is to apply factor analysis techniques to survey the large trace gas data set represented by BIBLE-B whole air sampling. Specifically we will employ the method of principal component analysis because it is simple and robust. In this section we outline the linear algebraic theory behind PCA using *Van de Geer* [1971], *Natusch and Hopke* [1983], *Hopke* [1985], *Norusis* [1986], *Press et al.* [1986], and *Malinowski* [1991]. Collectively these references overview multivariate statistics in general, the matrix algebra that must be understood in elementary applications, geometric perspectives that can be helpful in visualization, numerical approaches to the required calculations, and finally the special concerns of the chemist. The reader is referred to *Van de Geer* [1971], *Thurston and Spengler* [1985], and *Seinfeld and Pandis* [1998] for additional details about the PCA process, including the mathematical equations that are used.

[12] Multivariate statistical analyses normally derive from the classical data matrix, in which samples are represented by rows, and variables by columns [*Hopke*, 1985; *Malinowski*, 1991]. We will term this arrangement *X*. It may be preprocessed for statistical convenience through several conventions. The elements in the matrix may be demeaned, then normalized to the (square root of) the sample number and finally to a standard deviation. By scaling to the square root of the total sample number, premultiplication by the transpose gives the (square, symmetric) variance-covariance set. If standardization is employed, the variances are all unity and the covariances lie between zero and one. The resulting square data set is termed the correlation matrix. In atmospheric chemical applications, standardization has the advantage of enabling comparison among disparate concentration levels. For example, tropospheric mixing ratios of the gases we measure span the range from parts per trillion by volume (pptv) to parts per million by volume (ppmv).

[13] There are several ways in which to interpret the geometric structure of a data matrix [*Van de Geer*, 1971].

For example, individual elements may be conceived of as sample points in a variable space, or vice versa. In the latter case, the variance-covariance matrix gives a product of squared vector lengths and the cosine of the intervening angle. Along the diagonal of the matrix, the cosine is one. Length is the standard deviation. Under standardization, lengths are one and off-diagonal elements provide the cosine explicitly.

[14] In general, multivariate analysts assess the structure of X by applying methods of matrix and linear algebra [*Van de Geer*, 1971; *Norusis*, 1986]. The variables may be modeled as linear combinations of one another (multiple regression), or of manifold layers of precursors (path analysis). In factor analysis we seek unobserved quantities that explain the variation in the data set. They are treated as linear combinations of the observed values. Principal component analysis is sometimes defined as factor analysis building upon the demeaned X; principal factor analysis, is in turn the subset of PCA which adopts standardization [*Van de Geer*, 1971].

[15] In the data geometry whereby samples are represented as rows and variables as columns, we may ask whether the axes assigned by our chemical approach are the most natural. Are there other orthogonal vector bundles that concentrate variation in a more intuitive manner? Mathematically, we seek new axes such that their direction cosines premultiply to give the identity matrix, while projections onto them maximize the variance. We note that lengths of the data set vectors and the angles between them must be conserved. The problem may be configured under the rules of matrix differentiation and Lagrange (undetermined) multipliers [*Van de Geer*, 1971]. The eigen problem format falls out, and variance is apportioned into the eigenvalues [*Malinowski*, 1991]. The characteristic vectors form the basis of the transformation from the sample/variable space to that of the new axes. Symbolically, $XG = Y$, where matrix G is dimensioned with rows representing the observed variables and columns representing the derived axes, which are the 'factors'. We note that the columns g are actually weightings for the original variables, as they are brought together in linear combinations to constitute the factor scores (transformed coordinates). The factor matrix F consists of the elements of G along with a reflection of the eigenvalues. Our survey of the BIBLE-B data set interprets the columns f based on the relative values (loadings) for different gases. The variance apportionment contained in the vector of eigenvalues ranks the factors both numerically and relatively (see Tables 3–7). Significant factors must possess eigenvalues greater than one, or in other words, they should explain more of the variance than do individual species [*Van de Geer*, 1971].

[16] To enhance the interpretability of the principal component analysis, we have applied a standard Varimax rotation, which is an optimization of kurtosis (distribution spread) across the factor matrix. A transformation T is determined such that the sum of the fourth powers of all the elements in the rotated matrix is maximized. Key eigen properties of the system (e.g., the communalities) remain unaffected by the Varimax rotation. The process derives from original work by statistical psychologists dating to the 1950s [*Kaiser*, 1958, 1959; *Van de Geer*, 1971]. Indeed, chemists owe many of their most powerful factor analysis tools to social scientists confronted with large, unwieldy data sets [*Hopke*, 1985; *Malinowski*, 1991].

[17] We have already argued that from the perspective of pure geometry, X can be considered as a set of points or vectors equal to the number of samples, distributed through the variable space [*Van de Geer*, 1971]. Principal component analysis runs orthogonal vectors through the space, which pick up variation most naturally. Usually only a limited number of uncorrelated factors (5 or 6) are required to explain almost all the variance in a data set of 15 or more intercorrelated variables [*Thurston and Spengler*, 1985]. In other words, high communality may be achieved through only a few factors, that is, in highly reduced dimensionality. In our current survey we choose X in all cases such that the number of samples exceeds the number of variables (Section 6). An excess number of samples relative to variables is required in order to obtain a statistically meaningful (stable) calculation [*Thurston and Spengler*, 1985].

5. Computations

[18] All the statistical calculations were conducted within our UCI laboratory on a small personal computer equipped with an Intel Pentium III processor. Its operating system is Linux, the international Unix shareware form. The data base server is set up through MySQL, which is efficient, friendly and supports Structure Query Language. Initial analytical chemistry results are stored following the gas chromatography analysis (Hewlett Packard instruments). Spectral information is thus readily available for manipulation by our factor analysis software.

[19] The programming language used was C. Algorithms for performing the required linear algebraic operations (for example in determining the eigenvectors and eigenvalues) were extracted from convenient literature and library sources including Matlab and the numerical recipes series [*Press et al.*, 1986, 1992; *O'Haver*, 1989]. In this study we have avoided the use of conventional commercial statistics packages. In the past, we have applied the Statistics Packages for the Social Sciences (SPSS) system [*Norusis*, 1986; *Malinowski*, 1991]. Although the occasional small bug may exist in such programs, they have usually been reliable in our experience. However, the major advantages of self-programming are an enhanced theoretical understanding and the ability to fine tune, as for example to source receptor and chemical kinetic relationships that are specific to atmospheric chemistry [*Thurston and Spengler*, 1982; *Hopke*, 1985]. Details of the specific algorithms adopted here are given by *Natusch and Hopke* [1983].

[20] Concentration measurements were expressed as volume-to-volume mixing ratios. For individual BIBLE-B flights, the mixing ratios were partitioned into data matrices of the form X, dimensioned as samples (rows) and species (columns) (Section 4). Missing data are only rarely generated by the existing UCI system. However, zero concentrations are sometimes encountered for shorter-lived gases and in particular for the olefins at higher altitudes. As a result, flight data columns were subjected to an informal test for distribution normalcy. For standard deviations larger than the mean value, a given species was removed from the data set. Among other advantages, this procedure reduces the number of species and extracts excessive zero concen-

Table 1. Hydrocarbon Source Fingerprints Used in the BIBLE Interpretation[a]

Compound	LPG1	LPG2	Burning	Autos	NG	Waste	China	Veg
Methane	0	0	10	10	20	5000		
Ethane	1	1	1	1	1	1	1	+
Propane	100	100	0.25	0.3	0.1		0.4	+
n-Butane	60	0.2	0.1	2	0.05		0.1	
i-Butane	30	1	0.01	0.3			0.1	
n-Pentane	1	<0.01	0.03	3			0.05	
i-Pentane	3	0.03	0.01	1			0.05	
Ethene	0	0	1	8			0.1	+
Ethyne	0	0	0.5	6			0.5	
Benzene	0	0	0.4	4			0.2	
CH$_3$Cl			0.1	low				
CH$_3$Br			0.001					
CH$_3$I			0.0001					

[a]All concentrations are normalized to those of ethane. Where data are available, sources are arranged in order of increasing methane-to-ethane ratio. References are itemized in the text. LPG1, liquefied petroleum gas sampled in Mexico City; LPG2, liquefied petroleum gas sampled in Los Angeles; Burning, biomass burning; NG, natural gas; Waste, waste treatment in landfills and sewers; China, a plume tracked to eastern China as point of origin; Veg, terrestrial vegetation.

trations. The standard Varimax rotation (kurtosis maximization) was applied lastly, as described by *Van de Geer* [1971] per earlier references [e.g., *Kaiser*, 1958, 1959].

6. Results

[21] Elements in the columns of the factor matrix are the principal component loadings (weightings to the new orthogonal axes). Factor identifications are made by comparison with regional atmospheric chemistry source profiles (Table 1), supplemented by hydrocarbon and halocarbon lifetime information. We note that the source profiles in Table 1 are not specific to Asia. Excluding several plume fingerprints, we have been able to study only surrogates for hydrocarbon generation processes actually taking place on the Asian continent. The need for Asian source studies is discussed at length by *Elliott et al.* [2003]. As a further aid to interpretation, a qualitative but fairly comprehensive list of global sources for selected hydrocarbons is given in Table 2.

[22] Eigenvalues and factor loadings were computed for all the BIBLE-B flights. In this paper we present results from four of the more interesting flights. In each case, three to five factors that contribute most heavily to the overall variance, as judged by eigenvalue order, are presented (Tables 3–6). Loadings are given for the most heavily weighted species on either side of zero (positive and negative). Thus, we present both correlations (positive loadings) and anticorrelations (negative loadings). The four flights fall conveniently into conceptual pairs. Flights 2 and 13 are outbound and return transits from Nagoya to Saipan that pass through 25°N, and Flights 9 and 10 probe the Australian continental interior (Figure 1).

[23] In a local urban study in Santiago, we found that an excess of around 25 samples relative to species was required to yield stable PCA results. For the four BIBLE-B flights, the number of samples exceeded the number of species by 23, 16, 26 and 26 for Flights 2, 13, 9 and 10, respectively (Tables 3–6). In each case, we performed a sensitivity test whereby we decreased the excess number of samples to between 6 and 16, by adding variables. The results for Flights 9 and 10 remained stable, that is, the top three principal components were almost identical in both cases. Interestingly, the results for Flight 13 were also stable, even though it has the smallest number of excess samples out of the four flights. In the case of Flight 2, the first principal component remained the same, and the second and third components (which have similar eigenvalues) were exchanged. During BIBLE-B, it is possible that stable results were achieved despite a relatively small sample number excess because the sampled air masses were less complicated than in a heavily polluted city.

6.1. Transit Flights Over the Western Pacific

[24] Flight 2 was an outbound flight that deployed from Nagoya and was flown south over the western Pacific to Saipan, and Flight 13 was the return flight flown from Saipan to Nagoya (Figure 1). Both transits were flown at high altitude (12.5 km for Flight 2, and 13 km for Flight

Table 2. Qualitative Source Apportionment for Selected Hydrocarbons, Based Largely on Global Budgeting Exercises[a]

Hydrocarbon	Sources
Methane	Managed and natural wetlands, enteric fermentation, landfills, human and animal waste, natural gas, coal combustion
Ethane	Fossil fuel combustion, natural gas, liquefied petroleum gas, biomass burning, the ocean, grasses and other vegetation
Propane	Natural gas, liquefied petroleum gas, biomass burning, the ocean, grasses and other vegetation
n-Butane	Fossil fuel combustion, natural gas, liquefied petroleum gas, biomass burning, the ocean
i-Butane	Fossil fuel combustion, natural gas, liquefied petroleum gas, biomass burning, the ocean
n-Pentane	Fossil fuel combustion, natural gas, liquefied petroleum gas
i-Pentane	Fossil fuel combustion, natural gas, liquefied petroleum gas
Ethene	Fossil fuel combustion, biomass burning, the ocean, vegetation
Propene	Fossil fuel combustion, biomass burning, the ocean
Ethyne	Fossil fuel combustion, biomass burning
Benzene	Fossil fuel combustion, biomass burning, solvents
Toluene and Xylenes	Fossil fuel combustion, solvents, biomass burning

[a]Adapted from *Cicerone and Oremland* [1988], *Singh and Zimmerman* [1992], *Guenther et al.* [1995], *Schlesinger* [1997], and our compilation in Table 1.

Table 3. Five Highest Positive and Negative Factor Loadings for BIBLE-B Flight 2, Using a Principal Component Analysis As Outlined in the Text[a]

Loading Order	Factor 1 Loading	Factor 2 Loading	Factor 3 Loading
Positive 1	+0.97 (Ethyne)	+0.94 (n-PrONO$_2$)	+0.92 (DMS)
Positive 2	+0.97 (CO)	+0.77 (2-BuONO$_2$)	+0.87 (CH$_3$I)
Positive 3	+0.96 (CH$_4$)	+0.66 (i-PrONO$_2$)	+0.50 (CH$_2$Br$_2$)
Positive 4	+0.96 (CHCl$_3$)	+0.29 (CH$_2$BrCl)	-
Positive 5	+0.96 (CH$_2$Cl$_2$)	-	-
Negative 1	−0.42 (MeONO$_2$)	−0.09 (H-2402)	−0.15 (C$_2$Cl$_4$)
Negative 2	−0.40 (CH$_3$I)	−0.08 (DMS)	−0.11 (i-Pentane)
Negative 3	−0.30 (DMS)	−0.02 (Cyclopropane)	−0.09 (H-2402)
Negative 4	−0.29 (H-2402)	−0.02 (CH$_3$CCl$_3$)	−0.03 (C$_2$HCl$_3$)
Negative 5	−0.27 (CH$_2$BrCl)	-	−0.02 (Toluene)

[a]Flight 2 was an outbound transit flight that deployed from Nagoya and was flown south over the western Pacific to Saipan. Fifty-five samples were assessed for 32 species. The eigenvalues for factors 1, 2, and 3 are 14.5, 2.3, and 2.3, respectively. These three components respectively represent 45%, 7%, and 7% of the variance in the Flight 2 data set. Compound abbreviations are given in the text.

13), with no mid-flight excursions into the boundary layer. Key loadings derived for Flights 2 and 13 appear in Tables 3 and 4 (take-off and landing data were excluded from the analysis).

[25] Flight 2 is dominated by a combustion signature (ethyne and CO), which in turn is anticorrelated with the set of maritime tracers. Chloroform (CHCl$_3$) and methylene chloride (CH$_2$Cl$_2$) in positions +4 and +5, respectively, are members of a series of industrial solvents that also includes trichloroethylene (C$_2$HCl$_3$), C$_2$Cl$_4$ and methyl chloroform (CH$_3$CCl$_3$). In our experience these species occur in doublets, triplets and sometimes collectively with regard to urban plumes. Taken together, the combustion markers and solvents are consistent with air of Korean or Japanese origin, as opposed to air from mainland China [Elliott et al., 2003]. The most likely source is the mission headquarters city of Nagoya. However, given the limited fingerprints presently available (Table 1), combustion types cannot be segregated. The first factor for Flight 13 also reflects urban/industrial pollutants, including hexane, C$_2$HCl$_3$ and xylenes. Although C$_2$Cl$_4$ and heptane do not appear in the top five loadings for the first factor of Flight 13, they were almost as highly loaded.

Table 4. As in Table 3 but for Flight 13, a Transit Flight Deployed From Saipan and Flown North Over the Western Pacific Ocean to Nagoya[a]

Loading Order	Factor 1 Loading	Factor 2 Loading	Factor 3 Loading
Positive 1	+0.99 (Hexane)	+0.96 (CH$_2$Br$_2$)	+0.90 (CH$_3$Cl)
Positive 2	+0.99 (C$_2$HCl$_3$)	+0.93 (CHBr$_3$)	+0.71 (Ethane)
Positive 3	+0.99 (Propene)	+0.88 (MeONO$_2$)	+0.70 (CH$_4$)
Positive 4	+0.99 (o-Xylene)	-	-
Positive 5	+0.99 (p-Xylene)	-	-
Negative 1	−0.03 (CH$_3$Cl)	−0.60 (CH$_4$)	−0.13 (n-PrONO$_2$)
Negative 2	−0.03 (CH$_4$)	−0.17 (Ethane)	−0.12 (CHBr$_3$)
Negative 3	−0.02 (F-114)	−0.16 (CCl$_4$)	−0.07 (CH$_3$CCl$_3$)
Negative 4	-	−0.11 (CH$_3$Br)	−0.06 (CCl$_4$)
Negative 5	-	−0.09 (F-114)	−0.06 (2-BuONO$_2$)

[a]Fifty-three samples were assessed for 37 species. The eigenvalues for factors 1, 2, and 3 are 22.8, 3.7, and 2.4, respectively. These three components respectively represent 62%, 10%, and 7% of the variance in the Flight 13 data set.

Table 5. As in Table 3 but for Flight 9, a Transit Flight Over the Australian Continent Flown South From Darwin to Alice Springs[a]

Loading Order	Factor 1 Loading	Factor 2 Loading	Factor 3 Loading
Positive 1	+0.91 (CH$_2$Br$_2$)	+0.96 (Ethyne)	+0.96 (n-Butane)
Positive 2	+0.89 (CHBr$_3$)	+0.93 (Benzene)	+0.96 (i-Butane)
Positive 3	+0.79 (CHBrCl$_2$)	+0.93 (CO)	+0.91 (n-Pentane)
Positive 4	+0.71 (CH$_3$I)	+0.91 (Ethane)	+0.78 (Propane)
Positive 5	-	+0.64 (CH$_3$Cl)	+0.64 (CHCl$_3$)
Negative 1	−0.80 (CH$_2$Cl$_2$)	−0.22 (H-1211)	−0.17 (CH$_3$Cl)
Negative 2	−0.80 (CH$_4$)	−0.15 (F-113)	−0.17 (CCl$_4$)
Negative 3	−0.42 (H-1211)	−0.13 (n-Pentane)	−0.08 (Toluene)
Negative 4	−0.24 (CCl$_4$)	−0.11 (MeONO$_2$)	−0.05 (CO)
Negative 5	−0.20 (H-2402)	−0.08 (CH$_2$Cl$_2$)	−0.02 (CH$_3$CCl$_3$)

[a]Fifty-seven samples were assessed for 31 species. The eigenvalues for factors 1, 2, and 3 are 6.0, 5.0, and 4.2, respectively. These three components respectively represent 19%, 16%, and 14% of the variance in the Flight 9 data set.

[26] The second factor for Flight 2 and the third factor for Flight 13 represent combustion, but of potentially different sorts for each flight (automotive and biomass burning, respectively). Short-lived secondary products load heavily for Flight 2, whereas longer-lived tracers (e.g., CH$_3$Cl, ethane, CH$_4$) show high loadings for Flight 13. We hypothesize that plumes of differing ages were encountered. The alkyl nitrates in Flight 2 suggest the oxidation of automotive emissions, whereas CH$_3$Cl and ethane in the Flight 13 loadings suggests influence from biomass burning.

[27] The third factor of Flight 2 and the second factor of Flight 13 are both positive for tracers of sea-air transfer. For Flight 2, the short-lived oceanic tracers DMS and CH$_3$I loaded together in Factor 3, while the longer-lived marine tracers methyl nitrate (MeONO$_2$), bromoform (CHBr$_3$) and methylene bromide (CH$_2$Br$_2$) loaded together in Factor 4 (not shown). Previous studies have also found associations between DMS and CH$_3$I [e.g., Bassford et al., 1999] and between alkyl nitrates and brominated marine species [e.g., Atlas et al., 1993]. For Flight 13, the compounds of oceanic origin include CH$_2$Br$_2$, CHBr$_3$ and MeONO$_2$. Dimethyl sulfide and CH$_3$I were not included as variables in the original Flight 13 analysis, but when added as variables in the PCA calculation, we found that they also loaded onto the second factor. The source associated with Factor 2 of Flight 13 is therefore air from the marine

Table 6. As in Table 3 but for Flight 10, a Transit Flight Over the Australian Continent Flown North From Alice Springs to Darwin[a]

Loading Order	Factor 1 Loading	Factor 2 Loading	Factor 3 Loading
Positive 1	+0.95 (Ethane)	+0.90 (Propene)	+0.87 (DMS)
Positive 2	+0.92 (Propane)	+0.86 (Ethene)	+0.86 (Isoprene)
Positive 3	+0.89 (CO)	+0.79 (Toluene)	+0.64 (CHBr$_3$)
Positive 4	+0.88 (Ethyne)	+0.60 (n-PrONO$_2$)	+0.59 (CH$_3$I)
Positive 5	+0.79 (Benzene)	-	+0.52 (CHCl$_3$)
Negative 1	−0.27 (MeONO$_2$)	−0.42 (1,2-DCE)	−0.33 (CH$_2$Cl$_2$)
Negative 2	−0.09 (F-11)	−0.36 (CH$_2$Cl$_2$)	−0.29 (1,2-DCE)
Negative 3	−0.08 (F-12)	−0.35 (F-114)	−0.17 (F-12)
Negative 4	−0.08 (HCFC-142b)	−0.21 (CH$_4$)	−0.16 (F-114)
Negative 5	−0.06 (CH$_2$Br$_2$)	−0.12 (F-113)	−0.07 (HCFC-142b)

[a]Fifty-seven samples were assessed for 31 species. The eigenvalues for factors 1, 2, and 3 are 6.0, 4.8, and 3.8, respectively. These three components respectively represent 20%, 15%, and 12% of the variance in the Flight 10 data set.

boundary layer that had been recently influenced by oceanic emissions.

[28] Pollution sources that are notably absent from the high loading lists in Tables 3 and 4 deserve mention as well. In our group we are sensitized to the propane-dominated Liquefied Petroleum Gas fingerprint based on studies conducted in Latin America, where LPG leakage is an ongoing problem [*Blake and Rowland*, 1995; *Chen et al.*, 2001]. By contrast, no LPG factors have been obtained in our Asian work. With the transition of the Chinese and related economies through the energy market penetration sequence, propane emissions should become more important, as LPG is among the cleaner of the fossil fuels.

6.2. Transit Flights Over Australia

[29] The BIBLE series was designed in part to study atmospheric chemistry of the ongoing and growing Southeast Asian biomass burn phenomenon [*Sawa et al.*, 1999; *Elliott et al.*, 2003]. On a round trip into the Australian interior (Flights 9 and 10 to and from Alice Springs, respectively), PCA was used to identify clear burning signatures. When take-off and landing are excluded, the sampled altitudes ranged from the boundary layer to 13 km during Flight 9, and from 9–13 km during Flight 10.

[30] The Northern Territory of Australia is a strong global biomass combustion source [*Hurst et al.*, 1994], and its terrestrial ecosystems range from forest to savannah to desert. During the BIBLE-B campaign, the area was shifting meteorologically from the dry into the wet season. The factor analysis results for Flights 9 and 10 show that the variance (represented as eigenvalues) is quite evenly distributed across the factors (Tables 5 and 6). This may be attributable to sampling along a transect of biogeochemical regimes, ranging from maritime to desert. Sea-air transfer tracers (brominated species and CH_3I) dominate the first factor of Flight 9 (Table 5). Industrial emissions such as CH_2Cl_2 and H-1211 account for the anticorrelation suite. Because current emissions of H-1211 are strongly associated with the Asian continent [*Fraser et al.*, 1999], inter-hemispheric exchange appears to have influenced the sampled air masses. The strong negative CH_4 loading may reflect modest Australian production and consumption of natural gas [*Water Resources Institute* (*WRI*), 1992]. On the return from Alice Springs during Flight 10, a classic biomass burning signature showed the largest variance (Table 6). The correlating species include the lightest alkanes (ethane, propane), as well as CO and triple bonded compounds and aromatics (ethyne, benzene). These species are relatively long lived as a class, which may indicate that the sampled plume had aged such that more reactive species were removed, or that the more reactive species loaded onto another factor.

[31] The second factor recorded for Flight 9 is similar to the first factor for Flight 10, but with the addition of CH_3Cl (Table 5). Biomass burning is the largest known source of CH_3Cl [*Butler*, 2000] and CH_3Cl is considered one of the more reliable burning fingerprint species. Small negative values for *n*-pentane may point to automotive activity in or between Darwin and Alice Springs, or to transport from the developed islands of Indonesia. For Flight 10, the second factor shows influence from a mix of tracers. The short-lived burning tracers propene and ethene appear to confirm

Table 7. As in Table 3, but for All BIBLE-B Flights[a]

Loading Order	Factor 1 Loading	Factor 2 Loading	Factor 3 Loading
Positive 1	+0.99 (*m*-Xylene)	+0.98 (Propadiene)	+0.88 (2-BuONO$_2$)
Positive 2	+0.99 (*p*-Xylene)	+0.98 (Ethene)	+0.73 (*i*-PrONO$_2$)
Positive 3	+0.98 (*o*-Xylene)	+0.97 (Propene)	+0.71 (*n*-PrONO$_2$)
Positive 4	+0.98 (C_2Cl_4)	+0.97 (Benzene)	+0.55 (CH_3Br)
Positive 5	+0.98 (CH_2Cl_2)	+0.95 (Ethyne)	+0.49 (1,2-DCE)
Negative 1	−0.02 (CH_2Br_2)	−0.19 (H-1211)	−0.07 (F-113)
Negative 2	−0.02 (CH_3Cl)	−0.16 (F-113)	−0.06 (F-12)
Negative 3	−0.02 (F-114)	−0.15 (F-12)	−0.06 (H-2402)
Negative 4	−0.01 (Cyclopropane)	−0.13 (HCFC-142b)	−0.06 (Octane)
Negative 5	−0.01 (CH_4)	−0.10 (H-2402)	−0.03 (MeONO$_2$)

[a]574 samples were assessed for 53 species. The eigenvalues for factors 1, 2, and 3 are 15.8, 12.0, and 4.3, respectively. These three components respectively represent 33%, 25%, and 9% of the variance in the BIBLE-B data set.

the interpretation of Flight 10 factor one, in which it was suggested that the more reactive burning species may have loaded onto another factor. *n*-Propyl nitrate is produced by hydrocarbon oxidation under nitrogen oxide rich conditions, and it has a small oceanic source [*Atkinson et al.*, 1982; *Blake et al.*, 2003]. Overall, the second factor of Flight 10 appears to primarily reflect short-lived biomass burning species. The anticorrelating species point to industry and natural gas usage, with likely transport from the northern hemisphere.

[32] For the third factor of Flight 9, the presence of the butane-pentane isomeric suite is likely a strong reflection of Australian vehicular activity. The negative correlates represent a mix of sources, though they are only weakly loaded. The third factor for Flight 10 is unique in our experience, whereby isoprene groups strongly with traditional marine boundary layer tracers such as DMS, $CHBr_3$ and CH_3I. Isoprene is known to be produced in and emitted from the ocean, but it is surprising to observe such a high loading. Isoprene is very short-lived under normal tropical daytime conditions (lifetime on the order of hours), and one possible interpretation is that a very fresh onshore plume was sampled.

6.3. All BIBLE-B Flights

[33] As a final exercise, a large factor decomposition was conducted for the entire mission (Table 7). The loadings both reiterate and confirm analyses for the paired flights, and they allow the results from these flights to be expanded upon. The factor of largest eigenvalue for the entire mission reflects solvents and industry (xylenes, C_2Cl_4, CH_2Cl_2), and shows the tendency for the xylenes to track one another because of similar emission sources and photochemistry. The pentane and butane vehicular tracers are not listed but they also entered very high in the loadings list. Negative loadings from the first factor are too weak to merit evaluation. The first factor from Table 7 shows a strong resemblance to the first factor of Flight 13 (Table 4). It is an interesting comment on the state of the Asian environment that in an atmospheric chemistry experiment spanning the western Pacific Rim and emphasizing remote samples, urbanization defines the dominant composition mode.

[34] The second factor for the entire mission consists of biomass burning tracers listed roughly in order of increasing

stability. The methyl halides are not shown, but all loaded in the 0.8 range. The anticorrelating species are industrial pollutants. The third factor is the most difficult to interpret. It is dominated by C_3-C_4 alkyl nitrates, which are well-known general tracers for oxidation taking place under nitrogen oxide rich conditions [*Flocke et al.*, 1998, and references therein]. Outflow from lightning storms contains elevated NO_x concentrations and was specifically sought during the BIBLE high altitude flight legs. As a result, one might suspect that this factor represents either processed air or lightning activity. The assessment is confounded, however, by the relatively high loadings of methyl bromide (CH_3Br), which has a variety of natural and anthropogenic sources, and of the industrial tracer 1,2-dichloroethane (1,2-DCE). Overall, the third factor likely represents a mix of influences.

[35] We conclude this section with a brief summary of the factors from the entire mission beyond these first three columns. In one case, CH_4 occurred together with short-lived anthropogenic halogenates, and the sea-air transfer species gave negative values. This may indicate either that CH_4 variability in the Asian biogeochemosphere is dominated by natural gas release, or that interhemispheric exchange plays a strong role. Overall the dominance of natural gas would be somewhat unexpected, given the proximity of large-scale rice cultivation and human waste inputs. However, natural gas leakage often emanates from point sources so that plumes are physically constrained. In addition to the CH_4 factor, purely oceanic factors were also in evidence, with eigenvalues on the order of 3. Conspicuously lacking were fingerprints that may be readily associated with the leakage of liquefied petroleum gas (e.g., propane) or with coal burning. However, propane-dominated factors may begin to appear as the Chinese economy modernizes [*Elliott et al.*, 2003].

7. Discussion

[36] Though a few factor analysis calculations have been reported by our group for urban systems [e.g., *Decker et al.*, 2000], we have not previously applied multivariate statistical analysis to regional/hemispheric measurement sets. Indeed, multivariate approaches have only rarely been applied to large-scale gas phase chemistry data sets [e.g., *Buhr et al.*, 1996a, 1996b]. Elemental composition of the regional aerosol has received considerably more attention [*Hopke*, 1985; *Maenhaut et al.*, 1996; *Miranda et al.*, 1996; *Swietlicki and Krejci*, 1996]. In the case of hydrocarbon measurements, we have normally interpreted these data sets via large numbers of uni- and bivariate comparisons (Section 3). Because a massive number of species permutation pairs exists (given many dozens of species that are detected), the molecules that are plotted may be restricted by chemical family or be based upon fingerprint relationships. Clearly, the tasks of constructing and inspecting graphics using uni- and bivariate techniques is labor intensive.

[37] Relative to this experience, the present survey highlights both strengths and weaknesses of PCA in the study of whole air samples at the regional scale. In terms of advantages, the factor methods are readily configured such that they are very inclusive of species and samples. In one of our runs, the entire BIBLE-B data set was covered in a single computation. The approach is both comprehensive and fast. Factor loadings are collected in a single table from which major, universally dimensioned conclusions may be distilled upon quick inspection. The results can be very concise, for example when marine tracers of several types group closely together and load strongly. At times the patterns are quite thought provoking. Unexpected tracer combinations occur as a matter of course and inspire the consideration of proximal sources and air mass mixing.

[38] The weaknesses of factor analysis are also laid revealed. Closer inspection of some variability groupings requires zooming in on flight legs of low sample number. Degree-of-freedom issues arise unless the measured species set is culled, in which case valuable information is lost [*Rummel*, 1970; *Malinowski*, 1991]. Furthermore, the analyst is eventually compelled to return to low dimensionality statistics in order to confirm identifications that can be made only tentatively using PCA. As a result, we argue that coupling factor analysis with uni- and bivariate calculations constitutes an ideal combination of interpretive tools. For example, PCA may be conducted upon a given large-scale data matrix and used to suggest the plotting of unusual species sets against altitude, flight time or one another. During low dimensionality surveys, such compounds may be discovered only randomly.

[39] Several authors have included meteorological and geographic variables in factor analysis studies of atmospheric composition [*Hopke*, 1985; *Buhr et al.*, 1996a, 1996b]. Under the PCA approach, measurements of all kinds are standardized and rendered dimensionless. With both physical and chemical quantities incorporated into the analysis, the need to zoom in on individual flight legs may become clear as correlations with altitude or wind direction become apparent. A cost associated with this refinement is the introduction of non-chemical variables into a matrix which requires an excess number of samples relative to variables. We have begun testing this refinement using the BIBLE-B data set, and will report the results elsewhere. Our emphasis here has been the internal properties of the whole air data set.

8. Summary and Conclusions

[40] The Japanese sponsored Biomass Burning and Lightning Experiment (BIBLE) was conceived to study the major generation mechanisms for nitrogen oxides in the troposphere along the western Pacific Rim. Second in the series, the BIBLE-B mission was flown in August–September, 1999 and comprised flights from Nagoya headquarters to the Marianas Islands, south to tropical and outback Australia, and out over the open sea and into biomass burning zones. The Gulf Stream aircraft platform was normally operated at upper levels (>10 km), though take-off, landing, boundary layer and connecting legs permitted sampling of the full depth of the troposphere. The full spectrum of stable organic compounds was investigated during BIBLE-B using whole air sampling followed by gas chromatography. More than 500 individual air samples were obtained during BIBLE-B, and 50 to 100 volatile compounds were quantified per flight, even in oxidation-processed high-altitude air.

[41] Interpretation of large whole air data sets can be a complex task. In the present work we tested a multivariate

approach, whereby the BIBLE-B whole air data were subjected to the classic factor reduction known as principal component analysis (PCA). Although we have applied factor analysis to a few urban hydrocarbon matrices in previous work, this constitutes the first attempt to analyze our regional-scale data sets in a multidimensional fashion. Our calculation system is semi-automated and does not derive from commercial packages. During the analysis of the BIBLE-B data set, compounds were culled based on concentration and lifetime criteria, such that the number of species retained was much less than the typical number of samples collected per flight. Major polluting sources ('factors') were identified based on known source fingerprints.

[42] On mid-latitude north-south transects, industrial tracers included small chlorinated species (e.g., $CHCl_3$, CH_2Cl_2), xylenes, and longer chain alkanes (up to C_7). Combustion tracers such as CO and ethyne also loaded heavily. The C_3-C_4 alkyl nitrates varied tightly together and may typify oxidation in the NO_x-rich Asian boundary layer. Urban contributions were attributable to Japan in the vicinity of the base of operations. The suite of sea-air transfer compounds (e.g., CH_3I, DMS, bromochloromethanes) tended to anticorrelate with the industrial suite. There was possible indication that multiple modes of sea-surface hot spots were detectable.

[43] During round trip flights from Darwin to the Australian interior (Alice Springs) the variance among polluting sources was evenly distributed, reflecting the interception of air masses overlying several distinct ecosystems (coastal and urban through savannah to desert). The bromo- and bromochloromethanes were prominent among the oceanic tracer gases. An interesting mixed marine signature paired isoprene with DMS, $CHBr_3$ and CH_3I, and was anticorrelated with local light industrial emissions. The Australian Northern Territories are known for dry season biomass burning, and plumes were encountered in which the combustion tracers grouped by lifetime (stability).

[44] A final PCA calculation was performed for the entire BIBLE-B data set. Covariance of the xylenes formed a dominant factor, attributable in part to their similarity in emissions, distribution and photochemical properties. The pentane isomers also loaded heavily in the first factor and were indicative of automotive activity. One-third of the PCA variance was attributed to this factor, which appears to be an Asian urban signal. A second factor was interpreted to consist of short-lived burning tracers that anticorrelated with the stable freons. We note that anthropogenic activity was found to dominate the tropospheric variability of the whole air samples, even in a Pacific Rim campaign that crossed the hemispheres and included remote samples. In a third factor, high loadings of C_3-C_4 alkyl nitrates and CH_3Br may reflect lightning activity during the course of the mission. Conspicuous in their absence were factors representative of liquefied petroleum gas leakage, which by analogy with Latin American experience should be on the rise in Asia.

[45] There are advantages and disadvantages to using factor reduction as a mechanism for analysing suites of whole air data. In the past, the relationships among the assayed compounds have been determined through the detailed examination of multiple plots of individual species versus physical flight parameters or one another. Factor analyses are inherently inclusive of many samples and compounds and are extraordinarily concise. In our final computation, for example, the vast majority of the mission data were dealt with simultaneously. Loadings of the components were collected in a tabular form that was readily interpreted. On occasion the loadings pointed to unexpected connections between molecular families. For example, diverse pairings such as xylene-heptane and iso-prene-DMS were observed in the BIBLE-B data set.

[46] Disadvantages to the PCA survey approach are that flight segments generally do not contain a sample number excess relative to the number of resolvable species. As a result, chemical subsets may be constructed based on stability considerations. In many cases, source identification can only be tentatively made, and the analyst must then consult bivariate plots for a confirmation of suspected sources. We suggest that factor reductions in general and principal component analysis in particular may soon play a vital role in analysis of regional hydrocarbon data sets, as a complement to more familiar methods. In the context of UCI measurements, surveying the data using PCA can guide the selection of compounds for low dimensionality plots. We envision future publications beginning with load tables, and proceeding through graphics with families or mixed species groupings presented against flight time or perhaps ethane. It is possible that the eigen modes and loadings will themselves form the basis for illuminating the figures. Hydrocarbon variances also embody oxidation and removal information, and coupling with photochemical models may prove useful in the future.

References

Apel, E. C., J. G. Calvert, T. M. Gilpin, F. C. Fehsenfeld, D. D. Parrish, and W. A. Lonneman, The Nonmethane Hydrocarbon Intercomparison Experiment (NOMHICE): Task 3, *J. Geophys. Res.*, *104*, 26069–26086, 1999.

Atkinson, R., S. M. Aschmann, W. P. L. Carter, A. M. Winer, and J. N. Pitts Jr., Alkyl nitrate formation from the NO_x-air photooxidations of C_2-C_8n-alkanes, *J. Phys. Chem.*, *86*, 4563–4569, 1982.

Atlas, E., W. Pollock, J. Greenberg, L. Heidt, and A. M. Thompson, Alkyl nitrates, nonmethane hydrocarbons, and halocarbon gases over the equatorial Pacific Ocean during SAGA 3, *J. Geophys. Res.*, *98*, 16,933–16,947, 1993.

Bachmeier, A. S., R. E. Newell, M. C. Shipham, Y. Shu, D. R. Blake, and E. V. Browell, PEM-West A: Meteorological overview, *J. Geophys. Res.*, *101*, 1655–1677, 1996.

Bassford, M. R., G. Nickless, P. G. Simmonds, A. C. Lewis, M. J. Pilling, and M. J. Evans, The concurrent observation of methyl iodide and dimethyl sulphide in marine air: Implications for sources of atmospheric methyl iodide, *Atmos. Environ.*, *33*, 2373–2383, 1999.

Blake, D. R., and F. S. Rowland, Urban leakage of liquefied petroleum gas and its impact on Mexico City air quality, *Science*, *269*, 953–956, 1995.

Blake, D. R., V. H. Woo, S. C. Tyler, and F. S. Rowland, Methane concentrations and source strengths in urban locations, *Geophys. Res. Lett.*, *11*, 1211–1214, 1984.

Blake, D. R., D. F. Hurst, T. W. Smith, W. J. Whipple, T. Chen, N. J. Blake, and F. S. Rowland, Summertime measurements of selected nonmethane hydrocarbons in the Arctic and Subarctic during the 1988 Arctic Boundary Layer Expedition (ABLE3A), *J. Geophys. Res.*, *97*, 16,559–16,588, 1992.

Blake, D. R., T. W. Smith, T. Chen, W. J. Whipple, and F. S. Rowland, Effects of biomass burning on summertime nonmethane hydrocarbon concentrations in Canadian wetlands, *J. Geophys. Res.*, *99*, 1699–1719, 1994.

Blake, N. J., D. R. Blake, B. C. Sive, T.-Y. Chen, F. S. Rowland, J. E. Collins Jr., G. W. Sachse, and B. E. Anderson, Biomass burning emissions and vertical distribution of atmospheric methyl halides and other

reduced carbon gases in the South Atlantic region, *J. Geophys. Res.*, *101*, 24,151–24,164, 1996.

Blake, N. J., D. R. Blake, T. Chen, J. E. Collins, G. W. Sachse, B. E. Anderson, and F. S. Rowland, Distribution and seasonality of selected hydrocarbons and halocarbons over the western Pacific basin during PEM-West A and PEM-West B, *J. Geophys. Res.*, *102*, 28,315–28,331, 1997.

Blake, N. J., D. R. Blake, E. Atlas, F. Flocke, and F. S. Rowland, Latitudinal, vertical and seasonal variations of $C_1–C_4$ alkyl nitrates in the troposphere over the Pacific Ocean during PEM-Tropics A and B: Oceanic and continental sources, *J. Geophys. Res.*, *108*(D2), 8242, doi:10.1029/2001JD001444, 2003.

Blifford, I. H., and G. O. Meeker, A factor analysis model of large scale air pollution, *Atmos. Environ.*, *1*, 147–157, 1967.

Buhr, M. P., K. J. Hsu, C. M. Liu, R. Liu, L. Wei, Y. C. Liu, and Y. S. Kuo, Trace gas measurements and air mass classification from a ground station in Taiwan during the PEM West A experiment (1991), *J. Geophys. Res.*, *101*, 2025–2035, 1996a.

Buhr, M. P., D. Sueper, M. Trainer, P. Goldan, B. Kuster, F. Fehsenfeld, G. Kok, R. Shillawski, and A. Schanot, Trace gas and aerosol measurements using aircraft data from the North Atlantic Regional Experiment (NARE 1993), *J. Geophys. Res.*, *101*, 29,013–29,027, 1996b.

Butler, J. H., Better budgets for methyl halides?, *Nature*, *403*, 260–261, 2000.

Chen, T.-Y., I. J. Simpson, D. R. Blake, and F. S. Rowland, Impact of the leakage of liquefied petroleum gas (LPG) on Santiago air quality, *Geophys. Res. Lett.*, *28*, 2193–2196, 2001.

Cicerone, R. J., and R. S. Oremland, Biogeochemical aspects of atmospheric methane, *Global Biogeochem. Cycles*, *2*, 299–327, 1988.

Colman, J. J., D. R. Blake, and F. S. Rowland, Atmospheric residence time of CH_3Br estimated from the Junge spatial variability relation, *Science*, *281*, 392–396, 1998.

Colman, J. J., A. L. Swanson, S. Meinardi, B. C. Sive, D. R. Blake, and F. S. Rowland, Description of the analysis of a wide range of volatile organic compounds in whole air samples collected during PEM-Tropics A and B, *Anal. Chem.*, *73*, 3723–3731, 2001.

Decker, E. H., S. Elliott, F. A. Smith, D. R. Blake, and F. S. Rowland, Energy and material flow through the urban ecosystem, *Ann. Rev. Energy Environ.*, *25*, 685–740, 2000.

Elliott, S., D. R. Blake, N. J. Blake, M. K. Dubey, F. S. Rowland, B. C. Sive, and F. A. Smith, BIBLE: A whole air sampling as a window on Asian biogeochemistry, *J. Geophys. Res.*, *108*(D3), 8407, doi:10.1029/2001JD000790, 2003.

Flocke, F., A. Volz-Thomas, H.-J. Buers, W. Pätz, H.-J. Garthe, and D. Kley, Long-term measurements of alkyl nitrates in southern Germany: General behavior and seasonal and diurnal variation, *J. Geophys. Res.*, *103*, 5729–5746, 1998.

Fraser, P. J., D. E. Oram, C. E. Reeves, S. A. Penkett, and A. McCulloch, Southern hemispheric halon trends (1978–1998) and global halon emissions, *J. Geophys. Res.*, *104*, 15,985–16,000, 1999.

Guenther, A., et al., A global model of natural volatile organic compound emissions, *J. Geophys. Res.*, *100*, 8873–8892, 1995.

Henry, R. C., and G. M. Hidy, Multivariate analysis of particulate sulfate and other air quality variables by principal components, part I, Annual data from Los Angeles and New York, *Atmos. Environ.*, *13*, 1581–1596, 1979.

Hopke, P. K., *Receptor Modeling in Environmental Chemistry*, Wiley-Intersci., New York, 1985.

Hopke, P. K., An introduction to receptor modeling, *Chemometrics Intel. Lab. Syst.*, *10*, 21–43, 1991.

Hurst, D. F., D. W. T. Griffith, and G. D. Cook, Trace gas emissions from biomass burning in tropical Australian savannas, *J. Geophys. Res.*, *99*, 16,441–16,456, 1994.

Kaiser, H. F., The Varimax criterion for analytic rotation in factor analysis, *Psychometrika*, *23*, 187–196, 1958.

Kaiser, H. F., Computer program for Varimax rotation in factor analysis, *Educ. Psychol. Meas.*, *19*, 413–423, 1959.

Kao, Y. S., General circulation of the lower atmosphere over the Far East, *Mem. Inst. Meteorol. Acad. Sinica*, *16*, 1–7, 1948.

Kondo, Y., M. Ko, M. Koike, S. Kawakami, and T. Ogawa, Preface to Special Section on Biomass Burning and Lightning Experiment (BIBLE), *J. Geophys. Res.*, *108*(D3), 8397, doi:10.1029/2001JD002401, 2003.

Maenhaut, W., I. Salma, and J. Cafmeyer, Regional atmospheric aerosol composition and sources in the eastern Transvaal, South Africa, and impact of biomass burning, *J. Geophys. Res.*, *101*, 23,631–23,650, 1996.

Malinowski, E. R., *Factor Analysis in Chemistry*, John Wiley, New York, 1991.

Massart, D. L., B. G. M. Vandeginste, S. N. Denning, Y. Michotte, and L. Kaufman, *Chemometrics*, Elsevier Sci., New York, 1988.

Merrill, J. T., Trajectory results and interpretation for PEM-West A, *J. Geophys. Res.*, *101*, 1679–1690, 1996.

Merrill, J. T., R. E. Newell, and A. S. Bachmeier, A meteorological overview for the Pacific Exploratory Mission WEST Phase B, *J. Geophys. Res.*, *102*, 28,241–28,253, 1997.

Miranda, J., E. Andrade, A. Lopez-Suarez, and R. Ledesma, A receptor model for atmospheric aerosols from a southwestern site in Mexico City, *Atmos. Environ.*, *30*, 3471–3479, 1996.

Mulaik, S. Z., *Foundations of Factor Analysis*, McGraw-Hill, New York, 1972.

Natusch, D. F. S., and P. K. Hopke, *Analytical Aspects of Environmental Chemistry*, John Wiley, New York, 1983.

Norusis, M. J., *Advanced Statistics*, SPSS, Chicago, Ill., 1986.

O'Haver, T. C., Teaching and learning chemometrics with Matlab, *Chemometrics Intel. Lab. Syst.*, *6*, 95–103, 1989.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran*, Cambridge Univ. Press, New York, 1986.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge Univ. Press, New York, 1992.

Rummel, R. J., *Applied Factor Analysis*, Northwestern Univ. Press, Evanston, Ill., 1970.

Sawa, Y., H. Matsueda, Y. Tsutsumi, J. B. Jensen, H. Y. Inoue, and Y. Makino, Tropospheric carbon monoxide and hydrogen measurements over Kalimantan in Indonesia and northern Australia during October, 1997, *Geophys. Res. Lett.*, *26*, 1393–1396, 1999.

Schlesinger, W. H., *Biogeochemistry*, Academic, New York, 1997.

Seinfeld, J. H., and S. N. Pandis, *Atmospheric Chemistry and Physics*, 1326 pp., John Wiley, New York, 1998.

Sharaf, M. A., D. L. Illman, and B. R. Kowalski, *Chemometrics*, John Wiley, New York, 1986.

Simpson, I. J., J. J. Colman, A. L. Swanson, A. R. Bandy, D. C. Thornton, D. R. Blake, and F. S. Rowland, Aircraft measurements of dimethyl sulfide (DMS) using a whole air sampling technique, *J. Atmos. Chem.*, *39*, 191–213, 2001.

Singh, H. B., and P. Zimmerman, Atmospheric distribution and sources of nonmethane hydrocarbons, in *Gaseous Pollutants: Characterization and Cycling*, edited by J. Nriagu, pp. 177–235, John Wiley, New York, 1992.

Statheropoulos, M., N. Vassiliadis, and A. Pappa, Principal component and canonical correlation analysis for examining air pollution and meteorological data, *Atmos. Environ.*, *32*, 1087–1095, 1998.

Swietlicki, E., and R. Krejci, Source characterisation of the Central European atmospheric aerosol using multivariate statistical methods, *Nucl. Instrum. Methods Phys. Res., Sect. B*, *109–110*, 519–525, 1996.

Thompson, B. W., An essay on the general circulation of the atmosphere over southeast Asia and the west Pacific, *Q. J. R. Meteorol. Soc.*, *77*, 569–597, 1951.

Thurston, G. D., and J. D. Spengler, Source contributions to inhalable particulate matter in metropolitan Boston, *Pap. 82-21.5*, Air Pollut. Control Assoc., Pittsburgh, Pa., 1982.

Thurston, G. D., and J. D. Spengler, A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston, *Atmos. Environ.*, *19*, 9–25, 1985.

Van de Geer, J. P., *Introduction to Multivariate Analysis for the Social Sciences*, W. H. Freeman, New York, 1971.

World Resources Institute (WRI), *World Resources 1992 and 1993*, Oxford Univ. Press, New York, 1992.

Yu, T.-Y., and L.-F.W. Chang, Selection of the scenarios of ozone pollution at southern Taiwan area utilizing principal component analysis, *Atmos. Environ.*, *34*, 4499–4509, 2000.

————————

D. R. Blake, Y. Choi, S. Meinardi, F. S. Rowland, and I. J. Simpson, Department of Chemistry, University of California-Irvine, 573 Rowland Hall, Irvine, CA 92697, USA. (isimpson@uci.edu)

J. J. Colman, M. K. Dubey, and S. Elliott, Atmospheric and Climate Sciences, Los Alamos National Laboratory, Mail Stop D-401, Los Alamos, NM 87545, USA.

T. Shirai, Earth Observation Research Center, National Space Development Agency, 14F Roppongi First Building, 1-9-9 Roppongi, Minato-ku, Tokyo, 106-0032 Japan.

F. A. Smith, Department of Biology, University of New Mexico, 86 Castetter Hall, Albuquerque, NM 87131-1091, USA.