Winter 12-10-2017

Multivariate Statistical Analyses of Air Pollutants and Meteorology in Chicago During Summers 2010–2012

Katrina Binaku
Loyola University Chicago, kbinaku@luc.edu

Martina Schmeling
Loyola University Chicago

Author Manuscript

This is a pre-publication author manuscript of the final, published article.

Recommended Citation

Multivariate Statistical Analyses of Air Pollutants and Meteorology in Chicago During Summers 2010–2012

Katrina Binaku, Martina Schmeling[*]

Department of Chemistry & Biochemistry, Loyola University Chicago
1032 West Sheridan Road, Chicago, Illinois 60660, U.S.A.

*Corresponding Author:

Dr. Martina Schmeling

Loyola University Chicago

1032 West Sheridan Road

Chicago, IL 60660

Phone: +1 (773) 508-3124

Fax: +1 (773) 508-3086

Email: mschmel@luc.edu

Abstract

Aerosol, trace gas, and meteorological data were collected in Chicago, Illinois during 2010–2012 summer air studies. Ozone, nitrogen oxides, acetate, formate, chloride, nitrate, sulfate, and oxalate concentrations as well as temperature, wind speed, wind direction, and humidity data were explored by both principal component analysis (PCA) and canonical correlation analysis (CCA). Multivariate statistical techniques were applied to uncover existing relationships between meteorology and air pollutant concentrations and also reduce data dimensions.

In PCA, principal components (PC) revealed a relationship of ozone and nitrate concentrations with respect to temperature and humidity, coupled with transport of species from the south in relation to the sampling site (PC1). PC2 was a measure of secondary aerosols but also suggested acetate and formate presence was a result of primary emissions or transport. Both PC3 and PC4 contained residual information with the former representing days of lower pollution and the latter representing northerly wind transport of chloride, nitrate, and ozone to the sampling site.

In CCA, three canonical functions were statistically significant. The first indicated high temperature and low wind speed had a strong linear relationship ozone, oxalate, and nitrogen oxides concentrations whereas the second function showed a strong influence of wind direction on acetate, formate, and chloride concentrations. Residuals of temperature, wind speed, trace gases, and oxalate also were in the second function. The only new information in the third function was humidity. Overall, PCA and CCA bring forth multi-variable relationships, not represented in descriptive statistics, useful in understanding pollution variability.

Keywords: Air quality, canonical correlation analysis, multivariate statistics, principal component analysis, urban air pollution

1. Introduction

A large amount of data accumulates during extensive air pollution studies. As pollution signatures are region specific due to differences in natural and anthropogenic sources as well as weather patterns, a multitude of pollution monitoring campaigns have been completed across the world (Chan and Yao 2008; Chow et al. 2006; Deshmukh et al. 2013; Gianini et al. 2012; Molina et al. 2010; Samek et al. 2017; Vasconcellos et al. 2007). Air pollution in Chicago has been studied (Cooke and Wadden 1981; Fosco and Schmeling 2006, 2007; Hu et al. 2008; Lee et al. 1993; Scheff et al. 1984; Simcik et al. 1999) but few have analyzed data beyond descriptive statistics. With a resident count of approximately 2.7 million, Chicago, Illinois is considered the third-largest city in the United States (City of Chicago 2013) and has had a long history of industrial presence. With these characteristics and Chicago's unique geographical location on the southwestern end of Lake Michigan, pollution in this major urban city is important to study.

Multivariate statistical techniques are used to discover underlying relationships buried in large data sets that are not found using traditional descriptive methods. Redundant information is eliminated, making the data more manageable for interpretation. Relationships between meteorology and air pollutants are important as both emission sources and weather regimes influence air quality episodes via transport and transformation of air pollutants across regions. Both principal component analysis (PCA) and canonical correlation analysis (CCA) have been

applied to air pollution data collected from various locations (Binaku et al. 2013; Binaku 2014; Braga et al. 2005; Liu et al. 2016; Statheropoulos et al. 1998; Vasconcellos et al. 2007; Yu and Chang 2006).

Principal component analysis is a multivariate technique used to reduce data dimensions and determine relationships across one dataset (Wilks 2011). PCA is applied to the covariance or the correlation matrix of data, of which the latter is preferred when original variables are measured on different scales (Wilks 2011). In PCA, existing variables are transformed into a new set of variables, which are linear combinations of the original variables. Linear combinations, or principal components (PCs), capture the maximum variance within original data without repeating information, resulting in a fewer number of new variables explaining original data variability (Wilks 2011). Derived coefficients called loadings are used to interpret the influence of original variables in a linear combination (Wilks 2011). The first derived principal component (PC1) contains the maximum variability of the original data. The second component (PC2) describes variability not already extracted by PC1. All PCs are orthogonal and uncorrelated to one another (Wilks 2011). An eigenvalue for each PC is used to determine PCs to retain for in-depth interpretation (Wilks 2011). Kaiser's Rule states to retain all PCs whose eigenvalues are greater than one. Additionally, a Scree plot shows eigenvalue versus corresponding principal component number; a change in slope between points on the plot determines the cut-off range for interpreting PCs (Wilks 2011). Score plots are visualizations of original data observations in new variables' space and can be used to identify patterns and clusters (Johnson and Wichern 1998; Wilks 2011). Extensive information on PCA can be found in both Wilks (2011) and Johnson and Wichern (1998).

Canonical correlation analysis derives linear relationships that exist between two datasets, each containing multiple variables (Wilks 2011). One set contains variables classified as predictor or independent variables, while the other dataset of variables are the response or dependent variables (Hair et al. 1992). Similar to PCA, new variables in the form of linear combinations (also called canonical variates) are derived. Canonical weights derived for each original variable expose the maximum correlation present between linear combinations of the predictor and response variables. The larger a canonical weight, the more influential the associated original variable. Each pair of linear combinations is called a canonical function or canonical variate pair (Hair et al. 1992). The first canonical variate pair encompasses the maximum correlation found in the original data (Hair et al. 1992). Each successive canonical variate pair represents maximum correlation between variables without repeating information explained in preceding variate pairs. This results in uncorrelated, orthogonal canonical variate pairs. Plots projecting new variables' scores are used to identify outliers or groups (Hair et al. 1992). More details on CCA can be found in Johnson and Wichern (1998).

Equation (1) represents the general structure of a linear combination, for CCA or PCA:

(1) $\text{Linear combination}_x = b_{1,x}C_1 + b_{2,x}C_2 + b_{3,x}C_3 + \dots b_{n,x}C_n$

The symbols C_1, C_2, $C_3 \dots C_n$ represent each original variable; b denotes the derived weight assigned to each original variable while subscript "n" signifies that a canonical weight value is different for each variable. The subscript "x" symbolizes which successive linear combination is derived (1st, 2nd, 3rd, etc.) and that weights are different in each linear combination.

In this study, Chicago, Illinois air pollutant and meteorological data collected during the summers of 2010, 2011, and 2012 was analyzed using both PCA and CCA. The purpose of data analysis was to reduce data dimensions and identify the presence of both pollutant-pollutant and pollutant-meteorological variable relationships. These results then would be compared to PCA, CCA results from a summer 2002–2004 air pollutant and meteorological dataset. The summer 2010–2012 air pollution studies complimented a previous study completed at the same site during summers 2002–2004, of which aerosol, trace gas, and meteorological data were collected to assess air pollution and the local lake breeze as well as method development of aerosol analysis (Fosco and Schmeling 2006, 2007). Having completed multivariate analysis on the 2002–2004 summer data (Binaku et al. 2013), PCA and CCA relationships between those data and the summer 2010-2012 data were assessed to identify similarities and differences in pollutant-pollutant or pollutant-meteorological variable relationships.

2. Experimental Details

2.1 Description of data collection

Aerosol, trace gas, and meteorological data were collected July–August during the summers of 2010, 2011, and 2012. Instrumentation was located at Loyola University Chicago's Air Station (LUCAS), atop a 60-meter tall on-campus building. The university is situated in a residential area in the Rogers Park neighborhood of Chicago, Illinois. The sampling site is 200 meters west of Lake Michigan shoreline and approximately 13 kilometers north of Chicago's downtown center, shown in Figure 1.

Fig. 1 Map of Chicago, surrounding suburbs, and the northwest Indiana region. The air sampling station (yellow pin, LUCAS) and its relation to Chicago's downtown area (black square) is depicted. Several industry point source pollutant locations in the vicinity are also shown. (Google 2013)

Pollution studies were completed on weekdays during the summer season, as Chicago is most susceptible to lake breezes and secondary pollution is expected to be its highest during this period. Mixing ratios of ozone (O_3) and nitrogen oxides ($NO-NO_2-NO_x$) were measured with 1-minute resolution using Thermo 49C and 42C Analyzers, respectively (Thermo Environmental Instruments, Inc., Franklin, MA). Meteorological parameters were recorded with 15-minute resolution using a Vantage Pro[TM] Weather Station (Davis Instruments, Hayward, CA). Both trace gas and weather instrumentation were online 24-hours each day. Pre-cleaned 47mm quartz fiber filters held in a Teflon filter holder attached to a vacuum pump were used in aerosol collections. Two aerosol samples were collected per day. The first aerosol sampling segment, defined as A, was from 0700 to 1000h local time (LT). Segment B, the second aerosol sampling, was from 1100 to 1300h local time. Aerosol collections were designed to capture short-term pollution development, pollutants on non-lake breeze days, and pollutant concentrations before and during a local lake breeze. It was previously observed in summer 2002–2004 studies at the same location that a lake breeze began most frequently between 1000–1100h (Fosco and Schmeling 2006, 2007). Therefore, a 1-hour gap between segments A and B was inserted so segment B collections could reflect aerosol concentrations during or after a lake breeze. New quartz fiber filters were used in each sampling segment. After collection, filters were stored in individual petri dishes and frozen. Prior to analysis, filters were equilibrated to room temperature. Then each filter was submerged in 4mL of Nanopure H_2O and sonicated for 20 minutes. The aqueous filter extract was analyzed for both cations and anions using a Metrohm 761 Compact Ion Chromatograph (IC) with conductivity detection (Metrohm USA, Inc., Riverview, FL).

2.2 Raw data and treatment

Sixty-six weekdays of air pollution and meteorological data was collected over all three summers. Temperature (degrees Celsius, °C), wind direction (degrees, °), wind speed (meters per second, $m\ s^{-1}$), and relative humidity (percent, %) were included in statistical analyses along with water-soluble anions quantified (microgram per cubic meter, $\mu g\ m^{-3}$) in aerosol samples. These included acetate (CH_3COO^-), formate ($CHOO^-$), chloride (Cl^-), nitrate (NO_3^-), sulfate (SO_4^{2-}), and oxalate ($C_2O_4^{2-}$). Both trace gases were also included, forming a dataset of 132 observations for 8 air pollutant variables and 4 meteorological variables. Aerosol, trace gases, and meteorology data had different resolution due to respective instrument sampling frequency. Trace gas mixing ratios and meteorology data were averaged to match the period of segment A (0700–1000h) and B (1100–1300h) aerosol collections.

All variables were standardized by calculating z-scores, subtracting a variable's average from respective variable observations and dividing the resulting value by standard deviation. Ions' observations mostly below detection limit were not included in this statistical study. An entire day's pollution and meteorology observation is removed in statistical programs if one variable's value is missing for that particular day. Matrix sample size drastically decreases if many observations are below detection limit. Other studies substituted "missing" data by inserting a method or instrumentation detection limit or variable's average value as an "observed" value (Statheropoulos et al. 1998; Yu and Chang 2006; Zhou et al. 2007). This was not done in our study, as derived relationships may be skewed as a result. Out of 132 total observations, 44 were omitted due to missing values. Meteorology were independent (predictor) variables whereas air pollutants were the dependent (response) variables. The computer program used to apply PCA and CCA was SAS® 9.3 Software (SAS Institute Inc 2010). Minitab® 16 Statistical Software (Minitab 2010) was used to generate score plots of PCA and CCA results.

3. Results

3.1 Principal component analysis

PCA was applied using the correlation matrix of data described in section 2.2. The first four principal components (PCs 1–4) were retained according to Kaiser's Rule. A scree plot shown in Figure 2 validates interpreting PCs 1–3. However, the slope between PC3 and PC4 suggests PC4 is not important. Because a scree plot is subjective, PC4 was retained. All four PCs accounted for roughly 70% of original data variance. Principal component loading values are shown in Table 1. Eigenvalue and cumulative percent variance explained by each PC is also presented. Only loading values $\pm$ 0.300 or greater were interpreted.
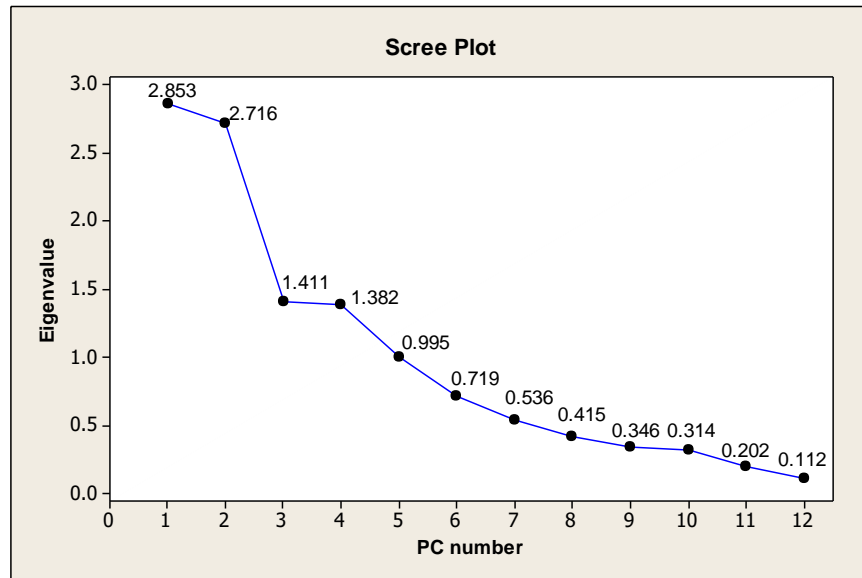

Fig. 2 Plot of eigenvalue versus corresponding principal component (PC) number, referred to as a scree plot

	PC1	PC2	PC3	PC4
Acetate	0.006	0.402	0.408	- 0.003
Formate	- 0.063	0.411	0.476	- 0.016
Chloride	- 0.006	0.379	- 0.275	0.310
Nitrate	0.418	- 0.020	- 0.276	0.317
Sulfate	0.184	0.360	- 0.351	0.133
Oxalate	0.197	0.391	0.186	0.260
Ozone	0.351	- 0.318	0.113	0.426
Nitrogen oxides	0.157	0.248	- 0.430	- 0.517
Wind speed	0.092	- 0.234	0.188	- 0.235
Wind direction	0.407	0.125	0.041	- 0.448
Temperature	0.507	- 0.070	0.184	- 0.076
Humidity	- 0.411	0.076	- 0.168	0.082
Eigenvalue	2.853	2.716	1.411	1.382
Cumulative Variance	23.8%	46.4%	58.2%	69.7%

Table 1 Principal component (PC) loading values, corresponding eigenvalues, and cumulative percentage of original data variance explained in the PCs

The first principal component (PC1) explains 23.8% of original data variability. This PC is a measure of nitrate (0.418), ozone (0.351), temperature (0.507), wind direction (0.407), and humidity (– 0.411), as these variables have the largest loading values. The sign of loading values indicate similarity in behavior of all positively loaded variables but an inverse relationship between humidity and the other variables. Both nitrate and ozone are secondary in nature, produced in the atmosphere by chemical reactions. The main pathway for tropospheric ozone production is nitrogen dioxide (NO_2) photolysis; nitrate ions are mainly a result of the daytime reaction of hydroxyl radicals and NO_2, forming gaseous nitric acid which can then undergo dry or wet deposition. Neutralization by ammonia, $NH_{3\,(g)}$, and deposition in the form of ammonium nitrate, $NH_4NO_{3\,(s,\,aq)}$, is another pathway for nitrate aerosol production (Finlayson-Pitts and Pitts 2000). PC1 is a measure of processed air, late morning or early afternoon, after rush hour emissions have decreased. At this time of day, local temperatures rise (large positive weight), aiding in secondary aerosol production. Winds originating from a large degree (135°, SE; 180°, S; 225°, SW) or southerly direction contribute to the measured pollutants via transport. The weight of nitrate and ozone suggests dependence of concentrations with respect to originating wind direction.

Contrary to literature, PC1 suggests nitrate aerosol increases with high temperatures and low humidity. According to Finlayson-Pitts and Pitts (2000), volatility of ammonium nitrate increases when humidity lessens and temperatures rise. Equilibrium is shifted from ammonium nitrate to gaseous nitric acid and ammonia in this case (Finlayson-Pitts and Pitts 2000). A similar result to PC1 was found when applying multivariate statistics to summer 2002–2004 data taken at the same location (Binaku et al. 2013). Binaku et al. (2013) suggested that organonitrate compounds play a role in the atmospheric signature of nitrate in Chicago. Many urban studies have measured organonitrate compounds in the organic aerosol fraction. The formation of organonitrates occurs both day and night, the product of peroxy radical (RO_2) and nitric oxide (NO) or nitrate radical and alkene reactions, respectively (Finlayson-Pitts and Pitts 2000). The results of our recent study suggest that the nitrate signature has not changed in Chicago since summers 2002–2004.

Principal component 2 has an eigenvalue of 2.716 and explains 22.6% of original data variance. PC2 is a measure of acetate (0.402), formate (0.411), chloride (0.379), sulfate (0.360), and oxalate (0.391). Ozone residuals (– 0.318) not expressed in PC1 are negatively weighted in principal component 2. All measured organic acid anions and two inorganic ions have similar behavior in PC2. There is evidence that formate and acetate both originate in the atmosphere due to direct emissions or secondary formation, sources of which are location dependent (Finlayson-Pitts and Pitts 2000). If Pearson correlations between formate, acetate, and ozone exist, it is indicative that photochemistry is favored over direct emissions (Finlayson-Pitts and Pitts 2000; Khwaja 1995; Souza et al. 1999). PC2 suggests photochemical formation is not the main source of formate and acetate ions in the Chicago region, as the negative sign of ozone's loading value suggests low ozone mixing ratios are present in cases of moderate to high formate and acetate concentrations. Sulfate, which is primarily formed via aqueous oxidation of sulfur dioxide, has a loading value similar to oxalate, suggesting local oxalate originates due to cloud, fog aqueous reactions. However, sulfate is also produced via gas-phase sulfur dioxide oxidation; it is possible that the organic acid anion species similarly originate in the gas phase, through oxidation of volatile organic compounds (Finlayson-Pitts and Pitts 2000). The commonality of organic acid anions and sulfate loading values also indicate transport of species from other regions is important. Earlier studies found that long-range transport of sulfate affects the Midwest region

(Cooke and Wadden 1981; Lee and Hopke 2006; Scheff et al. 1984). This could also be true for sulfate and chloride aerosol in this study.

The only new information derived in PC3 is the loading of nitrogen oxides (– 0.430). The pollutant was not highly weighted in the earlier principal components. Residual loadings of acetate (0.408), formate (0.476), and sulfate (– 0.351) not expressed in PC1 or PC2 are weighted in PC3. The associated eigenvalue is 1.411 and PC3 explains 11.8% of variance in the original data. PC3 is a measure of moderate acetate, formate concentrations and low sulfate, nitrogen oxide concentrations. Formate and acetate loadings were discussed in PC2 and the residual information in PC3 shows similar relationship between the two pollutant species. With meteorology not loaded in PC3, there is no distinction of weather conditions or their potential role on weighted pollutant information for this component.

PC4 consists of residual information. PC4 expresses 11.5% of original data variance and has an eigenvalue of 1.382. Residuals of chloride (0.310), nitrate (0.317), ozone (0.426), nitrogen oxides (– 0.517), and wind direction (– 0.448) play a large role in PC4. Loading values suggest PC4 is a measure of moderate concentrations of secondary pollutants transported by winds originating from a small degree (0°, N; 45°, NE; 90°, E) or northerly direction.

Score plots are shown in Figure 3. Observations (scores) are distinguished by sample year. Collections in 2010, 2011, and 2012 are represented by a circle (blue), square (orange), and diamond (black), respectively. In the top plot of Figure 3, PC2 versus PC1, there are two main clusters of points. Scores highly positive on PC1 (cluster no. 1) are influenced by high temperature and low humidity, in addition to high ozone and nitrate concentrations and large degree with respect to wind direction. These scores are from segment B collections in 2011 and 2012. For example, no. 119 and 123 were very warm days (34.1 and 32.7°C, respectively) with high ozone mixing ratios (77.51 and 93.37 ppb, respectively) and low relative humidity (45%), combined with westerly (270°) winds. Conversely, cluster no. 2 at the negative end of PC1 encompasses scores from 2011 and 2012 segment A collections and several from 2011 segment B samples. Scores 41, 56, 65, and 89 correspond to observations of low nitrate (0.315–1.108 μg m^{-3}) and temperature (18.9–23.5°C), low ozone (26.72–36.02 ppb), high humidity (70–90%), and winds originating from a small degree (northerly to easterly, 0–135°). This resulted in humidity dominantly affecting the location of scores. Along the vertical axis (PC2), summer 2012 observations are all weighted negatively, while both summers 2010 and 2011 scores are scattered about the axis. There are two outliers with large positive scores along PC2, no. 10 and 11, corresponding to summer 2010 measurements. Low ozone mixing ratios combined with high sulfate and above average chloride, acetate, and formate ion concentrations resulted in a high overall score on PC2, as all of the variables listed have positive loading values with the exception of ozone.

In the score plot of PC3 versus PC1 (middle graph, Figure 3), interpreting PC3 is emphasized as PC1 was already discussed. Along PC3, acetate, formate, sulfate, and nitrogen oxides influence score location. Scores with high sulfate and nitrogen oxides but low acetate and formate concentrations are on the negative end of PC3. Generally, the opposite is true for scores located on the positive end of PC3. To illustrate, no. 44 and 46 are at the opposite ends of PC3. Both observations are from 2011 segment A collections. The main influence on no. 44 is low mixing ratios of nitrogen oxides (16.85 ppb), combined with high formate (5.379 μg m^{-3}) and acetate (2.879 μg m^{-3}) concentrations. On the contrary, the negative location of no. 46 is a result of high nitrogen oxides (54.33 ppb), above average sulfate (11.974 μg m^{-3}) and low formate (0.079 μg m^{-3}) and acetate (0.521 μg m^{-3}) concentrations.

PC4 is the vertical axis in the bottom graph of Figure 3. Points are scattered along the axis but there are two sets of score extremes. Scores 46, 49, and 88 are weighted highly negative on PC4, while scores 10, 45, 59, 108, and 117 are highly positive on PC4 and far removed from the majority of the scores. A combination of wind direction, trace gas mixing ratios, and the concentrations of both nitrate and chloride recorded on sampling days affect score location along PC4. Large positive scores have lower nitrogen oxides and higher ozone than the highly negative scores. Wind direction for no. 10, 59, 108 and 117 were easterly, southeasterly in contrast to westerly winds for no. 46 and 88 and northerly winds for no. 45 and 49. There was no pattern for chloride and nitrate; their small loading values in comparison to ozone and nitrogen oxides show their influence is not as significant. Overall, clusters and outliers identified in score plots contributed to a better understanding of similarity in the behavior of local pollution and meteorology.
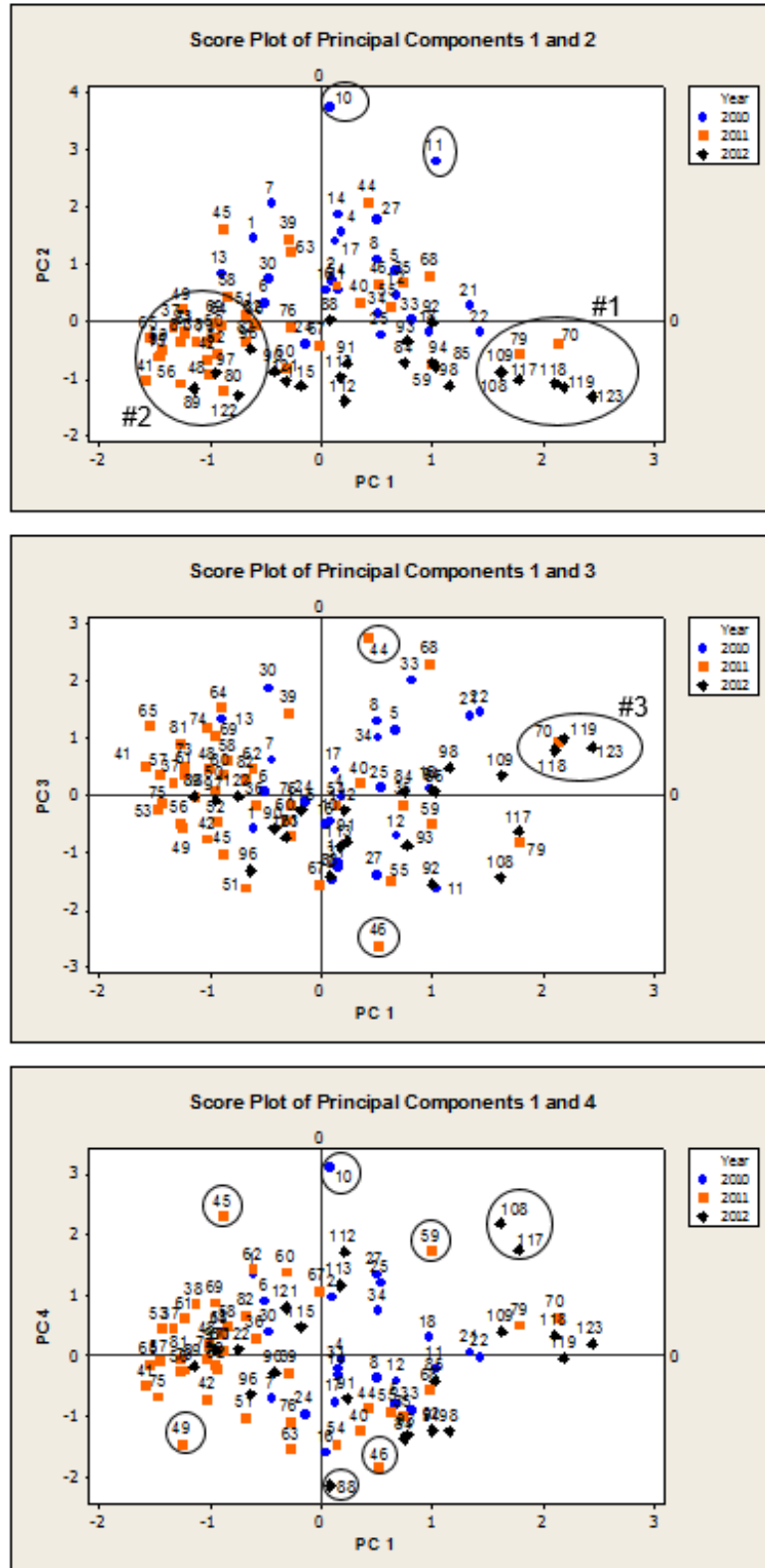
Fig. 3 Score plots projecting original data in derived principal component space. Observations are distinguished by both color, shape of plotting symbol and labeled with row numbers to identify sample dates and segments

10

3.2 Canonical correlation analysis

3.2.1 Canonical functions

The first three canonical functions in CCA are explained herein, as the fourth canonical function is not statistically significant at the 0.05 level. AP1, AP2, and AP3 refer to the canonical variates (linear combinations) derived for air pollution variables. Symbols M1, M2, and M3 represent meteorological parameter canonical variates. Canonical functions, canonical correlations, and canonical weights are reported in Table 2.

Canonical Function	M1, AP1	M2, AP2	M3, AP3
Canonical correlation	0.769	0.623	0.395
Wind speed	- 0.241	- 0.389	0.292
Temperature	0.925	- 0.784	0.279
Wind direction	0.039	1.016	0.641
Humidity	- 0.141	- 0.229	1.002
Acetate	0.214	0.347	- 0.336
Formate	0.011	- 0.502	0.978
Chloride	- 0.109	- 0.300	- 0.060
Nitrate	- 0.207	0.154	0.028
Sulfate	- 0.042	- 0.097	0.387
Oxalate	0.349	0.622	- 0.727
Ozone	1.142	- 0.643	0.128
Nitrogen oxides	0.723	0.457	0.501

Table 2 CCA results including each canonical function and corresponding canonical correlation. Canonical weights for air pollution (AP1, AP2, and AP3) and meteorology (M1, M2, and M3) linear combinations are also shown

The correlation between M1 and AP1, the first canonical variate pair, is high (0.769). Additionally, M1 explains roughly 60.3% of the variance in AP1. The most influential variable in M1 is temperature, with a canonical weight of 0.925. Wind speed is less influential with a smaller weight, – 0.241. Wind direction and humidity canonical weights are near zero, indicating they are negligible in M1. Overall, M1 is a measure of high temperatures and low wind speeds. Ozone, nitrogen oxides, and oxalate have large weights in AP1: 1.142, 0.723, and 0.349, respectively. This canonical function (M1, AP2) shows both temperature and wind speed affect the concentrations of oxalate, ozone, and, nitrogen oxides. High temperatures influence photochemical reactions such as in the production of ozone. In contrast, nitrogen oxides are primarily emitted due to fossil fuel combustion. It is unclear why nitrogen oxides' positive correlation with ozone and oxalate was derived, as it is known the trace gases are anti-correlated. Lack of air mass mixing and transport due to low winds is one possible explanation. A score plot generated by this canonical function is shown in the top graph of Figure 4. The large, positive correlation between M1 and AP1 canonical variates is clear. No trends or clusters are present and there are only a few points that deviate from the majority, namely 117 and 123. Both scores are largely positive on AP1. Observation no. 117 corresponds to July 13, 2012 segment B measurement of which the highest average ozone mixing ratio over all three summers was recorded, 94.43 ppb. Nitrogen oxides (11.46 ppb) and oxalate (0.254 µg m^{-3}) concentrations for the same period were average in comparison to cumulative three-summer data. Score no. 123 corresponds to July 23, 2012 segment B measurement. On this day the second-highest ozone

mixing ratio over all three summers was recorded, 93.37 ppb. Both nitrogen oxides and oxalate concentrations were similar to those observed for no. 117, 11.59 ppb NO_x and 0.277 µg m^{-3} oxalate. As ozone has the strong influence on score location along AP1, the above average mixing ratios for no. 117 and 123 directly contributed to their outlying position on the score plot.

The second derived canonical variate pair (M2, AP2) has a moderate canonical correlation (0.623). The amount of AP2 variance explained by M2 is 26.5%. Both acetate (0.347) and formate (– 0.502) are significantly weighted in AP2. The negative sign indicates an inverse relationship between species. Residual influence of ozone (– 0.643), nitrogen oxides (0.457), and oxalate (0.622) not explained by AP1 are present in this function. Overall, AP2 is a measure of high acetate, oxalate, and nitrogen oxides along with low formate and ozone concentrations. With respect to M2, wind direction has a canonical weight of 1.016. Temperature (– 0.784) and wind speed (– 0.389) residuals not explained in M1 are weighted in M2. This canonical function distinguishes different directions of wind and the effect on air pollutants. In cases of a large wind direction, the concentrations of acetate, oxalate, and nitrogen oxides are high while ozone and formate concentrations are low. To the west and south of the sampling site, there are major roadways, an airport, and suburban, agricultural activity while to the north is residential which may explain wind direction, pollution variability. Low wind speed also contributes to accumulation of local pollutants in the atmosphere.

The middle graph in Figure 4 shows the score plot of AP2 versus M2. Several scores in this function deviate from the majority (no. 11, 17, 18, 39, 49, and 88). Scores 17 and 88 are highly positive on M2 and AP2 and correspond to segment A measurements on August 6, 2010 and July 9, 2012, respectively. Northwesterly winds resulted in a large, positive score along M2, due to wind direction's canonical weight. Wind speed and temperature minimally influenced the overall calculated score. Two observations (no. 18 and 39) along the positive end of M2 are weighted negative on AP2. Similar to no. 17 and 88, large wind direction, high temperature, and low wind speed contributed to the scores' positive location on M2. High formate and average ozone concentrations influenced no. 18 and 39 the most along AP2. Score 11 corresponds to segment A measurements taken on July 27, 2010. Along M2, no. 11 is near zero while its location along AP2 is distinctly positive. A combination of above average nitrogen oxides, acetate, and oxalate concentrations contributed to the high positive score of no. 11 along AP2. Observation 49, an outlier, is negative along M2 due to a major influence of observed low temperature and high wind speed. With respect to AP2, the observed July 25, 2011 segment A measurement (no. 49) had low ozone and high nitrogen oxides mixing ratios in addition to average oxalate and acetate concentrations.

The third statistically significant canonical function has a low canonical correlation of 0.395. Roughly 7.7% of the variance in AP3 is explained by M3. The only new information in AP3 is sulfate's canonical weight (0.387). Acetate, formate, oxalate, and nitrogen oxide residual correlations not expressed earlier are weighted in AP3. In M3, humidity has a large positive canonical weight (1.002) as well as wind direction residuals (0.641). This function measures the direct relationship between formate, sulfate, nitrogen oxides, wind direction, and humidity, in addition to an inverse relationship between mentioned meteorology with acetate and oxalate. The bottom graph in Figure 4 shows the third canonical function. The location of several outliers (no. 10, 14, 34, 44, and 68) along M3 is dependent on humidity observed during sampling. Recorded humidity for no. 34 and 14 was low (30% and 58%), compared to observations 44 and 68 (77% and 70%, respectively). No. 44 and 68 had the first and second highest observed formate

concentration in the entire study, respectively, affecting their location along AP3 as formate is the most influential canonical weight.
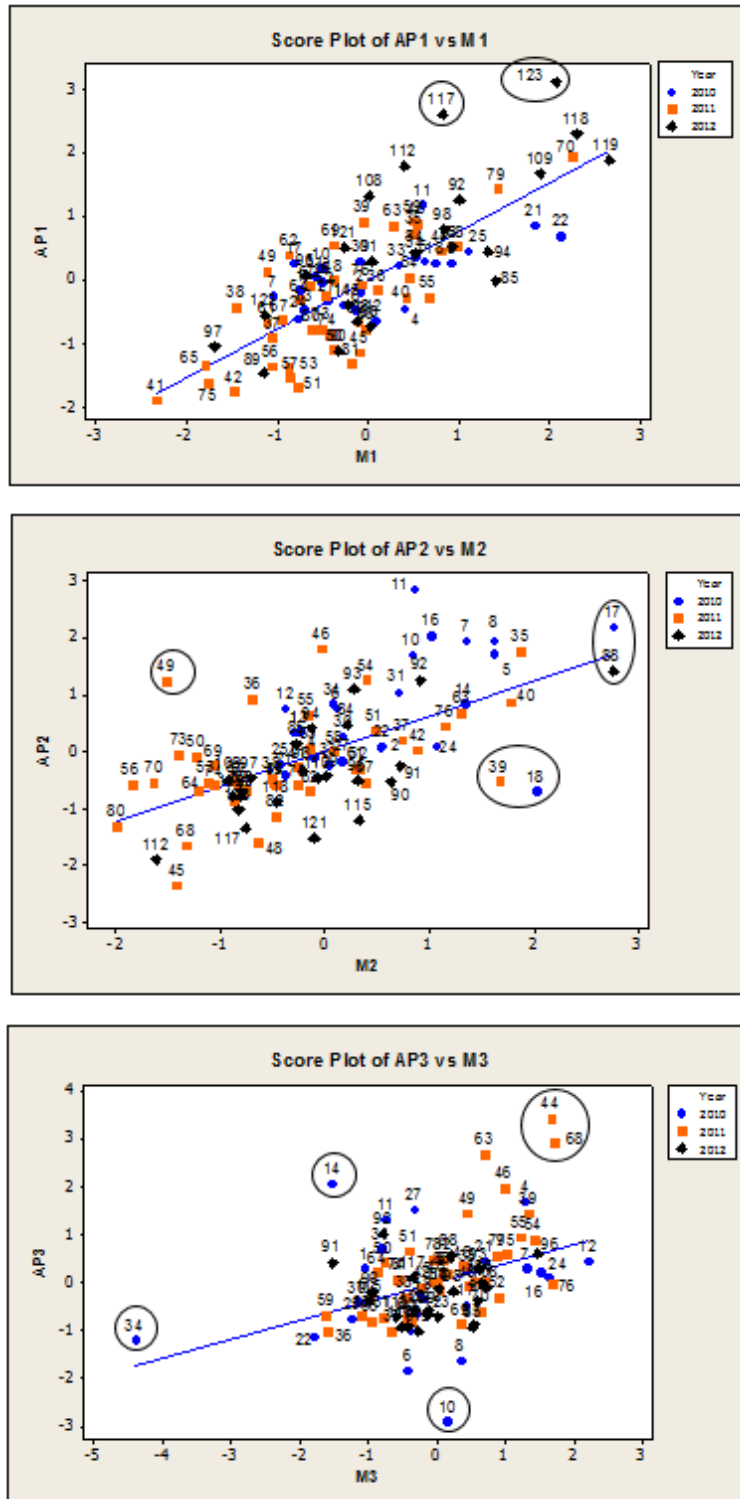

Fig. 4 Score plots of canonical functions. In each graph, score symbols are differentiated by year of collection (different shape, color of plotting symbol) and also labeled with the observation row number used to identify collection samples and dates

3.2.2 Canonical loading values

Canonical loadings reveal information masked by canonical weights due to multicollinearity between original variables (Hair et al 1992). Loading values $\pm$ 0.400 or greater were interpreted and are displayed in Table 3. AP1 is a measure of high ozone and nitrate concentrations. M1 is a measure of high temperatures and low humidity, as well as, wind directions with moderately large degree or southerly direction (135°, SE; 180°, S; 225°, SW), shown through the canonical loading values (T: 0.960, H: – 0.656, and D: 0.600). This was identified in PCA as well, with temperature, nitrate, and ozone inversely related to humidity. Wind direction and humidity were partialled out in the canonical weights.

Canonical loadings in AP2 indicate that nitrogen oxides (0.674) and oxalate (0.430) have a positive correlation with AP2 while ozone has a strong, negative correlation (– 0.608). Both wind speed and wind direction are correlated with M2, – 0.472 and 0.618, respectively. The air pollutant canonical loadings confirm the known anti-correlation between nitrogen oxides and ozone. The inverse relationship expressed between ozone and oxalate suggests photochemistry is not the main pathway for this organic acid anion. In other studies, oxalate was measured in the exhaust of automobiles (Kawamura and Kaplan 1987) and its relationship to nitrogen oxides and AP2 in this study is suggestive of auto exhaust origins as well. Furthermore, the canonical loading correlations within AP2 and M2 suggest morning conditions of low wind speed originating from the west affect loaded pollutant species.

Canonical loadings of formate, sulfate, and nitrogen oxides indicate correlation with AP3. This is similar to information found in the canonical weights in AP3. Wind direction and wind speed are similarly influential in M3, 0.415 and 0.400, respectively. Humidity is also positively correlated with M3. These canonical loadings confirm relationships found in the canonical weights of the third canonical function.

	AP1	AP2	AP3
Acetate	0.125	0.342	0.084
Formate	- 0.031	0.074	0.470
Chloride	- 0.061	0.071	0.060
Nitrate	0.596	- 0.011	- 0.002
Sulfate	0.147	0.258	0.363
Oxalate	0.303	0.430	- 0.264
Ozone	0.710	- 0.608	- 0.269
Nitrogen oxides	0.249	0.674	0.549
	M1	**M2**	**M3**
Wind speed	0.017	- 0.472	0.400
Temperature	0.960	- 0.180	0.173
Wind direction	0.600	0.618	0.415
Humidity	- 0.656	- 0.208	0.568

Table 3 Canonical loading values, the correlation between original observed variables and respective canonical variates. Values are presented for original air pollutant variables and canonical variates AP1-AP3 as well as meteorological parameters with corresponding variates M1-M3

4. Discussion

Data collected during the summer months of 2010–2012 of which CCA and PCA were applied to represents an extension to initial measurements of air pollutants and meteorology completed in summers 2002–2004 at the same location. Multivariate statistics applied to the first study's data validated the robustness of CCA and PCA (Binaku et al. 2013). Differences between data matrices of both studies were inclusion of acetate, formate, and chloride along with the absence of ammonium and calcium data for summers 2010–2012 as well as an increase in the number of measurement days. Several relationships between variables derived in both three-year summer studies' data indicate consistency in certain local pollutant concentrations and weather.

With respect to PCA, PC1 identified relationships between nitrate, ozone, and temperature in both three-summer studies. PC2, PC3, and PC4 represented different relationships in both 3-summer studies. In 2002–2004, insight on meteorological relationships was revealed in PC2 and PC3 (Binaku et al. 2013) while in 2010–2012 only pollutants' relationships were derived. In CCA, strong linear relationships between temperature, ozone, and oxalate identified in canonical weights along with the influence of nitrate on the first air pollutant linear combination were derived in both three-year studies. Temperature and humidity canonical loadings for the first meteorological linear combination were also consistent in both studies. However, 2002–2004 relationships showing moderate wind speed, humidity, and temperature residuals' inverse relationship with ozone and nitrogen oxides in the second canonical function (Binaku et al. 2013) differed from (M2, AP2) in 2010–2012. In 2002–2004, moderate temperature, wind direction, and humidity along with low wind speed were linearly correlated with high ammonium, nitrate, and oxalate with low sulfate and ozone concentrations (Binaku et al. 2013) but the exact relationships were not derived in the recent study. Similarities in the third canonical function of both studies include residuals of humidity, wind speed, and wind direction influencing both oxalate and sulfate concentrations. Other residual pollutant information present was exclusive to the former or latter study, not both. Variability between both studies is mainly due to different weather conditions and pollutants, as pollutant-pollutant relationships affect derived components. Analyzing statistical results of both studies shows that local relationships are not consistent and change over time. This highlights the importance of completing long-term pollution studies as emission sources, regulations, and pollutants change over time.

5. Conclusion

Underlying information between air pollutants and meteorology in Chicago, Illinois was explored using both PCA and CCA, which proved to be effective in reducing data dimensions and uncovering relationships between variables. In PCA, the first four principal components were significant, accounting for roughly 70% of original data variance. PC1 is a measure of local processed air masses from late morning, early afternoon hours with winds originating from the southerly direction. PC2 is a measure of both inorganic and organic ions and their inverse relationship with ozone residuals. PC3 and PC4 contain an abundance of residual information. Loadings in PC3 are indicative of morning or afternoons with low pollutant concentrations, whereas PC4 suggests several secondary pollutants are transported via northerly winds.

In CCA, three statistically significant canonical functions were derived. Through interpreting canonical weights it was found that temperature, oxalate, nitrogen oxides, and ozone have moderately strong, positive linear correlations. Low wind speed also influenced these

variables. The second canonical function describes the influence of wind direction and low wind speed on both acetate and formate, as well as pollutant residuals. AP3 and M3 were a measure of residual pollutant information and the positive, linear relationship with humidity and wind direction residuals. Overall, several underlying relationships between meteorology and pollutant concentrations were found which are useful in establishing local air pollution trends over time and use meteorological relationships in a preliminary predictive manner for pollution variability.

Compliance with Ethical Standards

The authors declare that they have no conflict of interest. No external funding was applied to this research project.

6. References

Binaku KL (2014) Investigation of Chicago air pollution and meteorological influence on pollution development: physical measurements and statistical analyses. Ph.D. Dissertation, Loyola University Chicago, Chicago, IL

Binaku K, O'Brien T, Schmeling M, Fosco T (2013) Statistical analysis of aerosol species, trace gasses, and meteorology in Chicago. Environ Monit Assess 185(9):7295–7308

Braga CF, Teixeira EC, Meira L, Wiegand F, Yoneama ML, Dias JF (2005) Elemental composition of PM_{10} and $PM_{2.5}$ in urban environment in South Brazil. Atmos Environ 39(10):1801–1815

Chan CK, Yao X (2008) Air pollution in megacities in China. Atmos Environ 42:1–42

Chow JC, Chen L-WA, Watson JG, Lowenthal DH, Magliano KA, Turkiewicz K, Lehrman DE (2006) $PM_{2.5}$ chemical composition and spatiotemporal variability during the California Regional $PM_{10}/PM_{2.5}$ Air Quality Study (CRPAQS). J Geophys Res doi: 10.1029/2005JD006457

City of Chicago (2013) City of Chicago: Facts & statistics. http://www.cityofchicago.org/city/en/about/facts.html. Accessed 29 July 2013

Cooke MJ, Wadden RA (1981) Atmospheric factors influencing daily sulfate concentrations in Chicago air. J Air Pollut Control Assoc 31(11):1197–1199

Deshmukh DK, Deb MK, Suzuki Y, Kouvarakis GN (2013) Water-soluble ionic composition of $PM_{2.5-10}$ and $PM_{2.5}$ aerosols in the lower troposphere of an industrial city Raipur, the eastern central India. Air Qual Atmos Health 6(1):95–110

Finlayson-Pitts BJ, Pitts Jr. JN (2000) Chemistry of the upper and lower atmosphere: Theory, experiments, and applications. Academic Press, New York

Fosco T, Schmeling M (2007) Determination of water-soluble atmospheric aerosols using ion chromatography. Environ Monit Assess 130:187–199

Fosco T, Schmeling M (2006) Aerosol ion concentration dependence on atmospheric conditions in Chicago. Atmos Environ 40:6638–6649

Gianini MFD, Gehrig R, Fischer A, Ulrich A, Wichser A, Hueglin C (2012) Chemical composition of PM_{10} in Switzerland: An analysis for 2008/2009 and changes since 1998/1999. Atmos Environ 54:97–106

Google (2013) Google earth. earth.google.com/web. Accessed 9 August 2013

Hair Jr. JF, Anderson RE, Tatham RL, Black WC (1992) Canonical correlation analysis. In Multivariate data analysis with readings. Macmillan Publishing Company, New York, pp 193–222

Hu D, Martinez A, Hornbuckle KC (2008) Discovery of non-Aroclor PCB (3,3'-Dichlorobiphenyl) in Chicago air. Environ Sci Tech 42(21):7873–7877

Johnson RA, Wichern DW (1998) Applied multivariate statistical analysis. Prentice Hall, Upper Saddle River, New Jersey

Kawamura K, Kaplan IR (1987) Motor exhaust emissions as a primary source for dicarboxylic acids in Los Angeles ambient air. Environ Sci Tech 21:105–110

Khwaja HA (1995) Atmospheric concentrations of carboxylic acids and related compounds at a semiurban site. Atmos Environ 29:127–139

Lee JH, Hopke PK (2006) Apportioning sources of $PM_{2.5}$ in St. Louis, MO using speciation trends network data. Atmos Environ 40:S360–S377

Lee HS, Wadden RA, Scheff PA (1993) Measurement and evaluation of acid air pollutants in Chicago using an annular denuder system. Atmos Environ 27A(4):543–553

Liu Z, Li N, Wang N (2016) Characterization and source identification of ambient VOCs in Jinan, China. Air Qual Atmos Health 9(3):285–291

Minitab Inc (2010) Minitab® 16 Statistical Software [Computer Software]. State College, PA

Molina LT, Madronich S, Gaffney JS, Apel E, de Foy B, Fast J, Ferrare R, Herndon S, Jimenez JL, Lamb B, Osornio-Vargas AR, Russell P, Schauer JJ, Stevens PS, Volkamer R, Zavala M (2010) An overview of the MILAGRO 2006 campaign: Mexico City emissions and their transport and transformation. Atmos Chem Phys 10:8697–8760

Samek L, Stegowski Z, Furman L, Fiedor J (2017) Chemical content and estimated sources of fine fraction of particulate matter collected in Krakow. Air Qual Atmos Health 10(1):47–52

SAS Institute Inc (2010) SAS 9.3 Statistical Software [Computer Software]. Cary, NC

Scheff PA, Wadden RA, Allen RJ (1984) Quantitative assessment of Chicago air pollution through analysis of covariance. Atmos Environ 18(8):1623–1631

Simcik M, Eisenreich SJ, Lioy PJ (1999) Source apportionment and source/sink relationships of PAHs in the coastal atmosphere of Chicago and Lake Michigan. Atmos Environ 33:5071–5079

Souza SR, Vasconcellos PC, Carvalho LRF (1999) Low molecular weight carboxylic acids in an urban atmosphere: Winter measurements in Sao Paulo City, Brazil. Atmos Environ 33:2563–2574

Statheropoulos M, Vassiliadis N, Pappa A (1998) Principal component and canonical correlation analysis for examining air pollution and meteorological data. Atmos Environ 32(6):1087–1095

Vasconcellos PC, Balasubramanian R, Bruns RE, Sanchez-Ccoyllo O, Andrade MF, Flues M (2007) Water-soluble ions and trace metals in airborne particles over urban areas of the state of Sao Paulo, Brazil: Influences of local sources and long range transport. Water Air Soil Poll 186:63–73

Wilks DS (2011) Statistical methods in the atmospheric sciences. Academic Press, New York

Yu TY, Chang IC (2006) Spatiotemporal features of severe air pollution in northern Taiwan. Environ Sci Pollut R 13(4):268–275

Zhou F, Guo H, Liu L (2007) Quantitative identification and source apportionment of anthropogenic heavy metals in marine sediment of Hong Kong. Environ Geol 53:295–305