Scientific
Research
Publishing

Application of Multivariate Geostatistics in Environmental Epidemiology: Case Study from Houston, Texas

Faye Anderson

Independent Scholar, Port Arthur, TX, USA
Email: andersonfaye7@gmail.com

Abstract

This study represents an example of investigating the associations between the joint exposure to ozone (O3) and particulate matter of sizes less than or equal to 2.5 micrometers in aerodynamic diameter (PM2.5) and cardiovascular disease (CVD) emergency room (ER) visits and chronic obstructive pulmonary disease (COPD) ER visits using multivariate geostatistics in Houston, Texas, from 2004 to 2009. Analyses showed lack of strong pair-wise association among the predictors of O3, PM2.5, wind speed, relative humidity, and temperature. Whereas CVD and COPD ER visits exhibited a strong positive correlation. Both outcomes drastically increased from 2006 possibly due to immigration from neighboring locations. Parametric testing showed that the data differed significantly between the years. Multivariate multiple regression results on the 2009 data showed that PM2.5, relative humidity, and temperature were significant to both CVD and COPD ER visits. Codispersion coefficients were constant which justified the assumption of intrinsic correlation. That is, our predictors had strong influence on the spatial variability of CVD and COPD ER visits. This multivariate geostatistics approach predicted an increase of 34% in CVD ER visits and 24% increase in COPD ER visits, which calls for more attention from policy makers. The use of multivariate geostatistics analyses enabled us to successfully detect the effects of risk factors on both outcomes.

Keywords

Multivariate Geostatistics, COPD, CVD, Harris, Houston, PM2.5, Ozone

1. Introduction

Assessing causal inferences between exposure to air pollution and respiratory and cardiovascular diseases

provides epidemiological evidence regarding the adverse health effects of air pollution. Most of relevant studies quantify the possible causality through association between the pollutants and the health outcomes due to exposure to pollutants either singularly or jointly. However, the effect of local air pollution versus neighboring levels is not fully explored yet [1] [2], which points to the importance of studying community-related characteristics which could help in estimating the causal inference between air pollution and its respiratory and cardiovascular adverse health effects. Application of multivariate geostatistical techniques can assist in this regards because it has the advantage of making use of the spatial correlations between observations pertinent to neighborhoods to predict values at unsampled places [3] [4].

This study capitalizes on previous studies which assess the effect of ozone and particulate matter on CVD and COPD ER visits in Harris County, Texas from 2004 to 2009 [5] [6] by studying the effect of joint exposure to air pollutants and other factors like temperature and relative humidity on both health outcomes instead of univariate modeling. The use of multivariate geostatistics has been useful in other applications like mining, but with the increased quality of spatial health data, their relevance is gradually increasing. This study presents a bivariate example with two dependent variables (CVD and COPD ER visits, which could be replicated to incorporate more dependent variables. This can provide immeasurable benefits to epidemiologists and policy makers by incorporating more outcomes that have been documented to be associated, thus, representing a more realistic situation than modeling the effects on one outcome at a time.

2. Data and Methods

Daily ozone, PM2.5, wind speed, wind direction, temperature, and relative humidity were obtained from EPA's AirData website [7]. Harris County daily emergency room visits for COPD and CVD from 2004 to 2009 data comes from the Emergency Room Utilization Study, University of Texas School of Public Health [8]. COPD ER visits are those with a primary diagnosis of chronic obstructive pulmonary disease (COPD) including chronic bronchitis (ICD-9 codes 490 - 491), emphysema (ICD-9 code 492), asthma (ICD-9 code 493), bronchiectasis (ICD-9 code 494), and chronic airway obstruction (ICD-9 code 496) [9]. CVD emergency room visits are those with primary diagnosis of heart failure (ICD-9 code 428), heart rhythm disturbances (ICD-9 codes 426 - 427), cerebrovascular events (ICD-9 codes 430 - 438), ischemic heart disease (ICD-9 codes 410 - 414, 429), and peripheral vascular disease (ICD-9 codes 440 - 448) [10]. Harris County zip codes from Texas State Data Center [11]. R version 3.1.1 [12] was used to conduct the statistical analyses. O3 and PM2.5 measurements were interpolated in zip codes where no observations were sampled, using kriging. Statistical analyses were conducted using R version 3.1.1.

3. Results and Discussion

Table 1 presents descriptive statistics for the collected variables. Analyzed variables included O3 and PM2.5 lagged at two and one day as well. Data exhibited high shifts of skewness and kurtosis from 0 and 3, respectively, which are the characteristic values of normal distribution. Therefore, the variables were of non-symmetric distributions, with long tails and several outliers. The total number of CVD ER visits during the six years was 96,596. Those visits kept increasing with time, starting with 3309 visits in 2004 and ending with 30,765 in 2009,

Table 1. Summary statistics for the dependent and independent variables.

	Ozone (ppb)	PM2.5 (µg/m³)	COPD	CVD	Wind speed (KPH)	Relative Humidity (%)	Temperature (F)
Mean	115.25	11.83	49.20	51.78	4.50	71.28	70.44
Median	25.25	10.82	48.00	50.00	4.28	72.20	72.60
Standard Dev.	292.85	5.02	26.51	33.91	2.28	11.69	12.07
Kurtosis	16.89	1.79	−0.85	−1.50	−0.37	1.36	−0.52
Skewness	3.92	1.13	0.31	0.01	0.40	−0.27	−0.65
Minimum	1.88	0.01	0.00	1.00	0.09	29.21	33.37
Maximum	2,515.97	40.87	131.00	130.00	12.66	143.65	89.04

a greater than 900% increase (**Figure 1**). Modeling the association of each visit to temperature, relative humidity, wind speed, wind direction, age, ethnicity, gender, day of the week, PM2.5 of the same day, PM2.5 of the previous day, O3 of the same day, and O3 of the previous day showed that the following variables were significant: visits on Friday, Saturday, Sunday and Tuesday, being African American male, wind speed, O3 of the same day, PM2.5 of the same day and previous day [6]. The total number of COPD ER visits was 95,765. Similar to CVD ER visits. COPD ER visits had positive significant associations with the being African American male, relative humidity, O3 of the same day, and PM2.5 of the same day [5]. Pair-wise correlations between the variables are given in **Table 2**. The most notable associations include the strong positive association between CVD and COPD ER visits of 0.74 and the negative association between temperature and COPD ER visits of −0.27.

The number of CVD and COPD ER visits was then summed up by date and zip code. Fisher's F test to test the null hypothesis of non-changing variances of each outcome between the years, gave p-values less than the significance level of 0.05. The same result was obtained for t test. Hence we reject the null hypothesis of equal variances and means respectively. That is, neither CVD nor COPD ER cases for Houston during those years can be combined as if they were coming from one statistical distribution. Consequent analyses were conducted on the 2009 data. Multivariate multiple regression results showed that the model is statistically significant and that lagged ozone and PM2.5 concentrations were not statistically significant to both outcomes. Interestingly, only PM2.5, relative humidity and temperature were significant to both CVD and COPD ER visits (**Table 3**). The effect
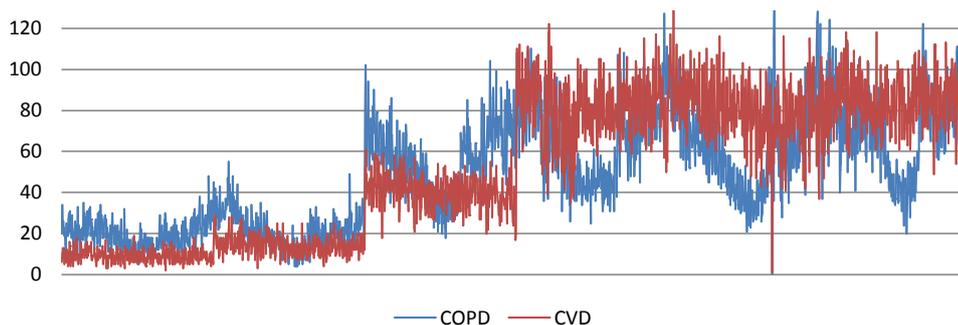

Figure 1. Daily Numbers of CVD and COPD ER visits from January 2004 to December 2009 in Houston, Texas.

Table 2. Pearson pair-wise associations.

	Ozone	PM2.5	COPD	CVD	Wind Speed	Relative Humidity	Temperature
Ozone	1.00						
PM2.5	0.05	1.00					
COPD	0.07	−0.12	1.00				
CVD	0.17	−0.02	0.74	1.00			
Wind Speed	−0.05	−0.13	0.04	0.00	1.00		
Relative Humidity	−0.04	−0.03	−0.03	−0.08	0.06	1.00	
Temperature	0.16	0.24	−0.27	0.04	−0.09	0.16	1.00

Table 3. Summary of type of association investigated by dependent variable.

O3 (0d lag)	O3 (1d lag)	PM2.5 (0d day lag)	PM2.5 (1d lag)	Relative Humidity	Temperature
Dependent variable: COPD					
√		√		√	
Dependent variable: CVD					
		√	√	√	
Dependent variables: CVD & COPD					
		√		√	√

of these three factors on the mechanisms of toxicity for respiratory and cardiovascular disease can be attributed to the oxidative stress leading to inflammation, which generates the physiological processes evolving as respiratory and/or cardiovascular symptoms like narrowing of airways, shortness of breath, wheezing, cough, and the ability of particles to penetrate the lung wall accumulating in the pulmonary interstitium between the lung and the bloodstream [13]-[15]. As for heat's effect on the health outcomes, research shows that hot-humid conditions are more challenging for patients with respiratory and cardiovascular diseases because heat and humidity affects the body's ability to preserve normothermia [16] [17]. The increase in humidity combined with particulate matter pollution allows for the delivery of the pollutant deeper into the lung [18]. Both heat and humidity exacerbate respiratory and cardiovascular conditions [19].

As for the spatial association between CVD and COPD ER visits across the different zip codes of Houston, the codispersion coefficients were constant with a variance of 0.001 (**Figure 2**). These are the ratios of the cross versus the direct variograms which can capture spatial association that is not possible to calculate following conventional methods. Therefore, we can assume the presence of intrinsic correlation whereby the model of intrinsic correlation is entirely specified by its spatial structure, and by the variance-covariance matrix [3] [20] [21]. Tests of normality lead to rejection of the null hypothesis of normality for CVD and COPD ER visits. Therefore, they were transformed using the logarithmic function. In order to predict future numbers of ER visits for each of the zip codes variograms were estimated, which exhibited anisotropy. This indicates that the independent variables had strong influence on the spatial variability of the two health outcomes. Directional variograms were estimated to characterize the spatial variability in different directions. Fitted variogram parameters (nugget, range, sill, smoothness parameter, practical range) were used in ordinary kriging [22]. Cross validation proved reliability of the approach used. This approach predicts that CVD and COPD ER visits will increase by more than 34% and 24% respectively. These predictions can be compared to future ER visits from following years. Knowing that Houston is one of the top cities with unpaid ER visits [23], this high percent carries multi-fold responsibilities: The need for more sophisticated interventions and resources in order to manage the continually increasing number of visits.

The persistent increase in CVD and COPD ER visits (**Figure 3**) in Houston, Texas could be attributed to variations in data coverage and completeness as well as socioeconomic factors including immigrations due to severe weather disasters where the majority of movers are usually those who belong to the traditionally under-served under-represented groups [5] [6]. The previous models were univariate, *i.e.*; the joint effects of the risk factors were investigated to one health outcome at a time. While these analyses provided good prediction because of inclusion of multipollutants (O3 of the same day, O3 of the previous day, PM2.5 of the same day, and PM2.5 of the previous day), they did not investigate the joint effect of the risk factors on both outcomes. This study allowed us to see how CVD and COPD ER visits are related once control variables have been included (through the examination of the correlation of the residuals), and that we can test the effect of ozone and PM2.5 both across CVD and COPD ER visits jointly [24]. **Figure 4** gives a proposed decision tree for applying multivariate geostatistics in environmental epidemiology, which was conducted in this study. Grayed boxes were not followed.
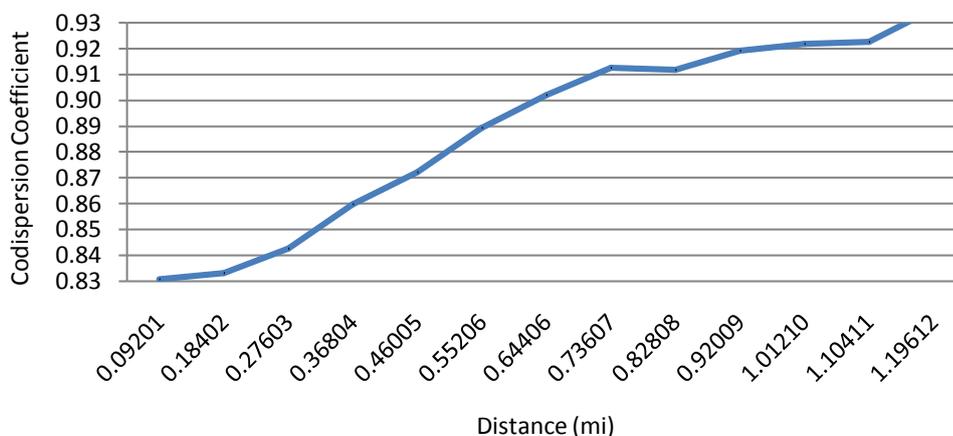

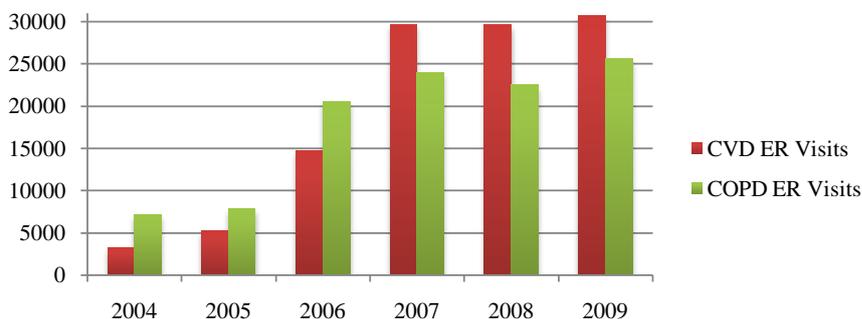**Figure 2.** Codispersion coefficients against distance in miles.

Figure 3. Total number of ER visits for CVD and COPD in Houston, Texas.
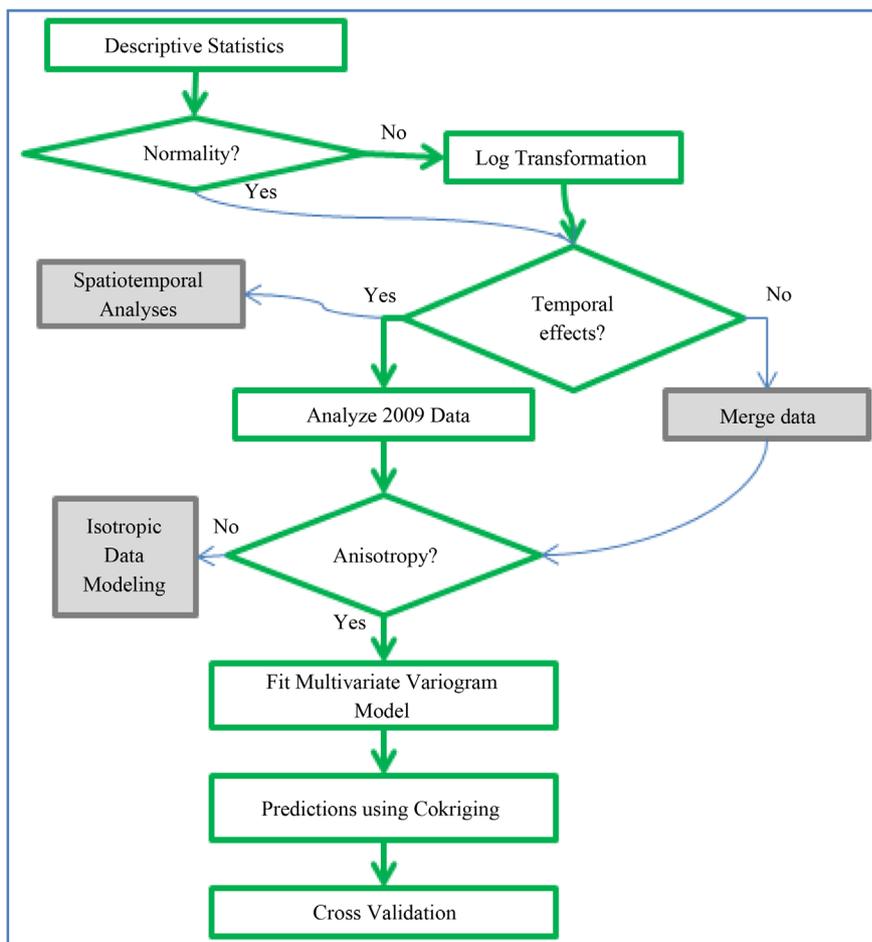

Figure 4. Proposed decision tree for applying multivariate geostatistics in environmental epidemiology.

4. Conclusion

This paper presents an approach to study the multivariate spatial nature between health outcomes and multipollutants using multivariate geostatistics. The example presented modeled the effect of ozone, PM2.5, relative humidity, wind speed, and temperature on both CVD and COPD ER visits at the same time. Now that we answered questions like the effect on both outcomes simultaneously, we can address more complex questions like the effect on more health outcomes that have been historically associated to each other. Further investigation is recommended in order to address potential issues like computational complexity and bias [22].

References

[1] Vanderweele, T.J. and Vansteelandt, S. (2009) Conceptual Issues Concerning Mediation, Interventions and Composition. *Statistics and Its Interface*, **2**, 457-468. http://dx.doi.org/10.4310/SII.2009.v2.n4.a7

[2] Robins, J.M. and Greenland, S. (1992) Identifiability and Exchangeability for Direct and Indirect Effects. *Epidemiology*, **3**, 143-155. http://dx.doi.org/10.1097/00001648-199203000-00013

[3] Wackernagel, H. (2013) Multivariate Geostatistics: An Introduction with Applications. Springer Science & Business Media.

[4] Goovaerts, P. (2006) Geostatistical Analysis of Disease Data: Accounting for Spatial Support and Population Density in the Isopleth Mapping of Cancer Mortality Risk Using Area-to-Point Poisson Kriging. *International Journal of Health Geographics*, **5**, 52. http://dx.doi.org/10.1186/1476-072X-5-52

[5] Anderson, F., Carson, A., Whitehead, L. and Burau, K. (2015) Age, Race and Gender Spatiotemporal Disparities of COPD Emergency Room Visits in Houston, Texas. *Occupational Diseases and Environmental Medicine*, **3**, 1-9. http://dx.doi.org/10.4236/odem.2015.31001

[6] Anderson, F., Carson, A., Whitehead, L. and Burau, K. (2014) Spatiotemporal Analysis of the Effect of Ozone and Fine Particulate on CVD Emergency Room Visits in Harris County, Texas. No. December, 87-99.

[7] EPA (2015) AirData. http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html#Daily

[8] Begley, C., Burau, K. and Courtney, P. (2011) ER Utilization Study. https://sph.uth.edu/research/centers/chsr/hsrc/

[9] CDC (1998) COPD. http://www.cdc.gov/niosh/pdfs/98-157-d.pdf

[10] AHA (2007) Heart Disease and Stroke Statistics. http://sociedades.cardiol.br/socerj/area-cientifica/Heart_Disease_and_Stroke_Statistics_2007.pdf

[11] TSDC (2015) No Title. http://txsdc.utsa.edu/txdata/shapefiles/

[12] Core, R. (2014) R: A Language and Environment for Statistical Computing.

[13] Duffin, R., Mills, N.L. and Donaldson, K. (2007) Nanoparticles—A Thoracic Toxicology Perspective. *Yonsei Medical Journal*, **48**, 561-572. http://dx.doi.org/10.3349/ymj.2007.48.4.561

[14] Mills, N.L., Miller, M.R., Lucking, A.J., Beveridge, J., Flint, L., Boere, J.F., Fokkens, P.H., Boon, N., Sandstrom, T., Blomberg, A., Duffin, R., Donaldson, K., Hadoke, P.W.F., Cassee, F.R. and Newby, D.E. (2011) Combustion-Derived Nanoparticulate Induces the Adverse Vascular Effects of Diesel Exhaust Inhalation. *European Heart Journal*, **32**, 2660-2671. http://dx.doi.org/10.1093/eurheartj/ehr195

[15] Donaldson, K., Mills, N., MacNee, W., Robinson, S. and Newby, D. (2005) Role of Inflammation in Cardiopulmonary Health Effects of PM. *Toxicology and Applied Pharmacology*, **207**, 483-488. http://dx.doi.org/10.1016/j.taap.2005.02.020

[16] Wilker, E.H., Yeh, G., Wellenius, G.A., Davis, R.B., Phillips, R.S. and Mittleman, M.A. (2012) Ambient Temperature and Bioniarkers of Heart Failure: A Repeated Measures Analysis. *Environmental Health Perspectives*, **120**, 1083-1087. http://dx.doi.org/10.1289/ehp.1104380

[17] Hosenpud, J.D. and Greenberg, B.H. (1994) Congestive Heart Failure. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-1-4613-8315-4

[18] Price, G. (2007) Effects of Weather, Air Quality and Geographical Location on Asthma and COPD Exacerbations in the Localities of Worcester and Dudley. University of Worcester, Worcester.

[19] Son, J.-Y., Bell, M.L. and Lee, J.-T. (2014) The Impact of Heat, Cold, and Heat Waves on Hospital Admissions in Eight Cities in Korea. *International Journal of Biometeorology*, **58**, 1893-1903. http://dx.doi.org/10.1007/s00484-014-0791-y

[20] Anderson, F. (2014) Multivariate Geostatistical Model for Groundwater Constituents in Texas. *International Journal of Geosciences*, **5**, 1609-1617. http://dx.doi.org/10.4236/ijg.2014.513132

[21] Goovaerts, P. (2008) Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. *Mathematical Geosciences*, **40**, 101-128. http://dx.doi.org/10.1007/s11004-007-9129-1

[22] Schabenberger, O. and Gotway, C.A. (2004) Statistical Methods for Spatial Data Analysis. CRC Press, Boca Raton.

[23] Begley, C.E., Behan, P. and Seo, M. (2010) Who Uses Hospital Emergency Rooms?: Evidence from Houston/Harris County Texas. *Journal of Health Care for the Poor and Underserved*, **21**, 606-616. http://dx.doi.org/10.1353/hpu.0.0312

[24] Rencher, A.C. and Christensen, W.F. (2012) Methods of Multivariate Analysis. John Wiley & Sons, Hoboken. http://dx.doi.org/10.1002/9781118391686