5-2024

GREY RELATIONAL ANALYSIS AND CANONICAL CORRELATION ANALYSIS OF AIR POLLUTION IN THREE KENTUCKY COUNTIES

Sarah Hartman

GREY RELATIONAL ANALYSIS AND CANONICAL CORRELATION ANALYSIS OF AIR POLLUTION IN THREE KENTUCKY COUNTIES

A Thesis submitted in partial fulfillment
of the requirements for the degree
Master of Science

Department of Mathematics
Western Kentucky University
Bowling Green, Kentucky

By
Sarah Hartman

May, 2024

GREY RELATIONAL ANALYSIS AND CANONICAL CORRELATION ANALYSIS OF AIR POLLUTION IN THREE KENTUCKY COUNTIES

Sarah Hartman

Date Recommended 4/5/2024

DocuSigned by:
Ngoc Nguyen
BD6F5C9E9564434...
Chair

DocuSigned by:
Melanie A. A
C0C3B268ED004E6...
Committee Member

DocuSigned by:
Lukun Zheng
8D4AF2BC8830416...
Committee Member

Committee Member

DocuSigned by:
Jennifer Hammonds
EBE3858E068F42D...
Interim Director of the Graduate School

ABSTRACT

GREY RELATIONAL ANALYSIS AND CANONICAL CORRELATION ANALYSIS OF AIR POLLUTION IN THREE KENTUCKY COUNTIES

Air pollution is a crucial factor that affects both the environment and public health. Various methods are available for assessing air quality and pollution levels, such as regression models, principal component analysis, and factor analysis tools. However, some of these methods present issues in multicollinearity and the nature of collected data. It is important to recognize that air pollution data is often uncertain, incomplete, and contains limited valid data points. Weather conditions and economic activities are also factors that can affect air pollution. With growing communities in Kentucky (KY), it is essential to address these factors as the state has unique economic imports and exports. Air pollution has a significant impact on both morbidity and mortality (Z. Song, Deng, & Ren, 2020). Therefore, monitoring and regulating air quality is necessary to mitigate these harmful effects. In this study, Grey Relational Analysis (GRA), Dynamic Grey Relational Analysis (DGRA), and Canonical Correlation Analysis (CCA) methodologies are used to analyze economic and meteorological factors and their relationship with three criteria pollutants: Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), and Ozone (O_3). Data is gathered from 2016-2019 from various sources and localized to 3 counties in Kentucky.

KEYWORDS: Grey Relational Analysis, Environmental Science, Canonical Correlation Analysis, Dynamic Grey Relational Analysis, Grey System Theory

This thesis is firstly dedicated to my Lord and Savior Jesus Christ who sat right next to me as I wrote this paper. It is because of his love, grace, and mercy through his death on the cross that I was able to write this thesis in peace, knowing no matter the outcome that I am safe in his arms. May this work glorify His name. Secondly, to Spencer's Coffee, where I (happily) spent a large amount of money on Spiced Apple Chai and many hours writing this thesis.

ACKNOWLEDGMENTS

Special thanks to my advisor, Dr. Ngoc Nguyen, for dedicating months of her time to my cause. I am indebted to her guidance throughout these last few months. Thank you for listening and responding with intention and patience, and for laughing at my cheesy jokes. Thank you to my committee members, who encouraged my idea from its beginnings and saw it through. Your time, patience, and guidance were incredibly valuable to me in all stages of this research project. Thanks to the math department, who prepared me to tackle this work. I appreciate all the smiles and encouraging waves in the hallway. Thanks to my family who challenged me to think critically about this project and the data, and for their support over the last year.

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1

Introduction, Background, Outline, and Data

1.1. Introduction

Recently, the intricate dynamics between environmental degradation, economic prosperity, and climatic patterns have garnered unprecedented attention. Among the myriad of environmental concerns, air pollution stands as a significant threat, posing multifaceted challenges to human health, ecosystem integrity, and socio-economic stability. Concurrently, the intricate interrelationships among air pollution, economic variables, weather patterns, and their collective impact on human civilization have become subjects of scholarly inquiry.

Various methods are available for assessing air quality and pollution levels, such as regression models, principal component analysis, and factor analysis tools. However, some of these methods present issues in multicollinearity and the nature of collected data. It is important to recognize that air pollution data is often uncertain, incomplete and contains limited valid data points. Weather conditions and economic activities are also factors that can affect air pollution. With growing communities in Kentucky (KY), it is essential to address these factors as the state has unique economic imports and exports. Air pollution has a significant impact on both morbidity and mortality (Z. Song et al., 2020). Therefore, monitoring and regulating air quality is necessary to mitigate these harmful effects.

This thesis comprehensively explores the intricate relationship between air pollution, economic parameters, and weather variables through Grey Relational

Analysis (GRA), Dynamic Grey Relational Analysis (DGRA), and Canonical Correlation Analysis (CCA). The rationale behind the selection of GRA, DGRA, and CCA lies in their adeptness in handling multi-dimensional datasets and discerning latent relationships within heterogeneous variables. By employing GRA, this study works to quantify the relative closeness among air pollution levels, economic indicators, and meteorological parameters, thereby bringing forward the underlying patterns of association. The adoption of DGRA further enhances this methodology through a unique addition to the methodology of GRA. Additionally, CCA is a robust multivariate statistical tool for elucidating the latent correlations between disparate sets of variables. The use of these tools within this thesis brings a full picture of the environmental health of three counties in Kentucky, while also providing a focused critique of the methods themselves. Further, this research contributes to the existing body of knowledge by offering insights into the differential impacts of various economic strata and climatic zones on air pollution across the state.

1.2. Background

Numerous studies have explored the impact of environmental factors on global public health. Not only has hot weather been studied in relation to respiratory health, but a literature review in this area of public health also reveals air pollution and hot weather exposure beyond certain thresholds have serious effects on respiratory health, with the elderly and young children being the most vulnerable groups. Additionally, there is a general acceptance of global warming increasing health effects of outdoor air pollution, and resulting in more heat waves (Grigorieva & Lukyanets, 2021). Temperature has been known as a catalyst for pollutants, and during heatwaves. Kalisa et al. (2018) found maximum temperature coinciding with the peak of O_3 and PM_{10}, which established that temperature's association with air pollution does not change according to rural and urban locations although air pollutants increased with increasing temperatures, particularly during heatwaves. Xu *et al.* (2021) globally analyzed the relationship between weather, air pollution, and SARS-CoV-2 (COVID-19) transmission, finding warmer temperatures and moderate outdoor ultraviolet exposure result in a slight reduction in the transmission of COVID-19. Still, changes in weather or air pollution alone are not enough to contain the spread of COVID-19 with other factors having greater effects. In light of recent global health events, COVID-19 transmissibility and factors affecting this phenomenon have been studied, specifically finding that temperature has a nonlinear negative relationship with COVID-19 transmissibility (Irfan, Ikram, Ahmad, Wu, & Hao, 2021). Meteorological data has also been studied in pollutant relationships. There have been multiple examples of chemical relationships to pollutants, such as humidity, as it promotes hygroscopic growth and continual chemical reactions of other pollutants in the atmosphere. Though this study does not focus on data collected within the COVID-19 era, it does provide a valuable example of environmental health's intersection with public health. A study conducted in Wuhan, China found "temperature, relative humidity, precipitation, and wind speed are negatively correlated with [air quality index] AQI, which atmospheric pressure is positively correlated with AQI for the entire study period" (Song et al., 2020, p.2) . Also mentioned in their paper is strong multicollinearity among meteorological elements, so principal components analysis is used to eliminate this. Abdul-Wahab et al. (2005) used principal component and multiple regression analysis in modeling ground-level ozone and factors affecting its concentration. It is found that while high temperature and high solar energy tended to increase the data time ozone concentrations, the pollutants NO_2 and SO_2 being emitted to the atmosphere are being depleted . Abdul-Wahab et al (2005) also mentioned significant amounts of multicollinearity among pollutant predictors, which makes multiple of their models inappropriate, as high levels of multicollinearity in the data can impact the accuracy of the model and increase variances. Multiple transformations are needed on the models, specifically logarithmic transformations. Though multiple regression analysis is a widely used methodology for showing relationships between predictor variables, it can face serious difficulties when independent variables are correlated with each other (Mcadams, Crawford, & Hadder, 2000). Additionally, the relationship between economic indicators and air pollution is noted within the literature in exposure science (Davis, Laden, Hart, Garshick, & Smith, 2010). Li et al. (2017) found all socio-economic indicators they considered in their study are related to multiple air pollutants (to some extent) . A study conducted in Switzerland revealed that economically oriented parameters correlate with high Ozone levels (Kuebler, Van Den Bergh, & Russell, 2001). Among these and other works of literature, there is evidence supporting the idea that "economic and regulatory activities have a significant impact on exposure to pollution" (Davis et al., 2010, pg. 18). Thus, there is empirical evidence of the effect of certain economic factors' impact upon air pollutant levels. These are explored using a variety of methods, including regression models and grey relational analysis. By integrating GRA, DGRA, and CCA into the analysis framework, researchers can overcome the limitations of traditional methods, such as multicollinearity or linear assumptions, and gain a deeper understanding of the complex interactions shaping air quality, economic development, and societal well-being.

1.3. Description of Thesis

In this study, Grey Relational Analysis, Dynamic Grey Relational Analysis, and Canonical Correlation Analysis methodology are used to analyze economic and meteorological factors and their relationship with three criteria pollutants: Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), and Ozone (O_3). Data are gathered from

2016-2019 from various sources and localized to 3 counties in Kentucky. Weather variables include average temperature, maximum relative humidity, precipitation, prevailing wind direction, and average wind speed. economic variables include mining, logging, and construction; manufacturing; and trade, transportation and utilities. These are further discussed in Section 1.4.

1.4. Data

1.4.1. Criteria Pollutants, Meteorological, and Economic Factors

Nitrogen dioxide is part of a group of highly reactive gasses and is used as an indicator for the larger group of nitrogen oxides in the atmosphere. The most likely source of this type of pollution is the burning of fuel from cars, trucks, buses, power plants, and off-road equipment. From a human health perspective, it can irritate airways and aggravate respiratory diseases. From an environmental perspective, it can form acid rain and make the air hazy and difficult to see through while also contributing to nutrient pollution in coastal waters (US EPA, 2016). Multiple monitoring sites are along the Ohio and Mississippi Rivers bordering Kentucky, and rivers are also susceptible to nutrient pollution.

Sulfur Dioxide is an indicator for a larger group of gaseous sulfur oxides. The largest source of SO_2 pollution is fossil fuel combustion at power plants and other industrial facilities. Exposure to high levels of SO_2 can affect adults and children who are sensitive to respiratory issues. High levels of this pollutant can lead to the "formation of other sulfur oxides and can react with other compounds in the atmosphere to form small particles, which contribute to Particulate matter pollution" (Environmental Protection Agency (EPA), 2019). Sulfur Dioxide also contributes to acid rain, which puts sensitive ecosystems in danger, and can reduce visibility in areas of the US, including national parks and wilderness areas.

Ozone pollution, also known as ground-level ozone, is a harmful air pollutant that can have negative impacts on human health and the environment. Ozone forms when nitrogen oxides and volatile organic compounds react in sunlight, and can be found in urban and rural areas. Long-term exposure to ozone can cause respiratory problems such as asthma, reduce lung function, and even premature death (Environmental Protection Agency (EPA), 2021a). In addition to its effects on human health, ozone pollution can also harm crops and other vegetation, reducing their productivity and negatively impacting ecosystems (NASA, 2022).

Meteorological factors such as temperature, humidity, wind speed and direction, and precipitation play an important role in the dispersion and transport of air pollutants in the atmosphere. For example, temperature and solar radiation are correlated with the formation and concentration of ground-level ozone (Environmental Protection Agency (EPA), 2021b), while wind speed and direction can determine the direction and distance that pollutants are transported from their sources. Precipitation can remove pollutants from the atmosphere through wet deposition, while dry deposition can occur when particles settle on surfaces due to changes in temperature and humidity. Understanding how meteorological factors are associated with the behavior of air pollutants is crucial for predicting and managing air quality and protecting public health.

The economic factors considered in this study include Mining, Logging and Construction; Manufacturing; and Trade, Transportation, and Utilities. Including economic variables in a study of air pollution and meteorological elements serves as a crucial endeavor for several reasons. Firstly, economic activities often serve as significant drivers of air pollution, with industries, transportation, and energy production contributing substantial emissions of pollutants such as NO_2, SO_2, and O_3. By incorporating these economic indicators, researchers can discern the extent to which economic development influences air quality and pollutant levels.

1.4.2. Specific Data Information

TABLE 1.4.1. Data Collection Locations

	Site 5	Site 67	Site 3002
Pollution	Daviess County	Jefferson County	Campbell County
Weather	Henderson County	Bullitt County	Campbell County
Economic	Daviess County	Jefferson County	Campbell County

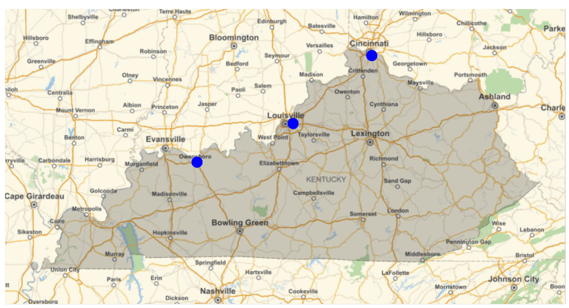

FIGURE 1.4.1. Map of Monitoring Sites

Table 1.4.1 describes Data Collection locations for each type of variable. Figure 1 provides a map of site locations. Daily values for each variable are collected during the summer months (June, July, and August) during 2016-2019, at three site locations in Kentucky, as defined by the US Environmental Protection Agency (US EPA, 2016). Sites are selected based on availability and completeness of data for the three months. In some cases, data from neighboring counties are utilized as some of the EPA monitoring sites are not included by the KY Mesonet or the US BEA. This is not anticipated to cause a large issue because of the proximity of each county. Sites use instruments that continually monitor air pollution levels. The instruments are kept in temperature-controlled shelters (US EPA, 2016). Pollutant data are obtained from the

US Environmental Protection Agency monitoring sites (US EPA, 2016). Pollutants are measured in parts per million (ppm). Because of sampling issues, there are some missing data for the period we investigated. An executive decision is made to delete those data from the analysis because of the nature of GRA, DGRA, and CCA. All weather variables are collected from the KY Mesonet, located at Western Kentucky University (Mahmood, Schargorodski, Foster, & Quillingan, 2019). All economic data are collected from the US Bureau of Economic Analysis (BEA) (Bureau of Economic Accounts (BEA), n.d.). Each variable is measured using all employees of that industry, in thousands (Bureau of Economic Accounts (BEA), n.d.).

Table 1.4.2 gives an abbreviation guide for use throughout the thesis, and includes not only weather and economic data, but also a few important abbreviations for future discussion (i.e., lower the better, higher the better, initial value processing, and mean value processing).

Table 1.4.2. Abbreviation Guide

AT (F)	Average Temperature
MRH (%)	Maximum Relative Humidity
P (in)	Precipitation
PWD	Prevailing Wind Direction
AWS (mph)	Average Wind Speed
MLC	Mining, Logging and Construction
M	Manufacturing
TTU	Trade, Transportation and Utilities
LTB	Lower the Better
HTB	Higher the Better
IVP	Initial Value Processing
MVP	Mean Value Processing

Tables 1.4.3, 1.4.4, and 1.4.5 present summary statistics for each collected data set for each site. The summary statistics feature Mean, Standard Deviation, and quantile information that includes minimum, maximum, and median values, among others.

Table 1.4.3. Summary Statistics Site 5

Variable	Mean	Std. Dev	Quantile				
			Min	25%	50%	75%	Max
NO_2 (ppm)	8.0377	4.0038	0	5	7	10	24
SO_2 (ppm)	4.4591	6.3088	0	1	2	5	42
O_3 (ppm)	0.0453	0.0104	0.014	0.038	0.045	0.05175	0.088
AT (F)	75.7245	4.1623	62	73.5	76.4	78.5	84
MRH (%)	97.5811	3.1107	76.4	96	98.9	100	100
P (in)	0.1315	0.3386	0	0	0	0.0675	2.37
PWD (Degrees)	178.2311	92.7530	0	112.5	202.5	247.5	337.5
AWS (mph)	4.0223	1.7805	0.7	2.8	3.6	5.1	10.6
MLC	2.7629	0.1293	2.5	2.7	2.8	2.9	2.9
M	8.4947	0.3551	7.9	8.5	8.6	8.8	8.9
TTU	10.2796	0.2804	10	10	10.2	10.7	10.7

Table 1.4.4. Summary Statistics Site 67

Variable	Mean	Std. Dev	Quantile				
			Min	25%	50%	75%	Max
NO_2 (ppm)	19.1288	6.9275	4.40	13.58	19.05	24.30	36.00
SO_2 (ppm)	2.4490	1.8927	0.20	1.20	1.90	3.10	13.00
O_3 (ppm)	0.0500	0.0119	0.02	0.04	0.05	0.06	0.09
AT (F)	75.2212	4.3812	60.70	71.18	76.05	78.30	83.90
MRH (%)	98.9481	2.3942	69.60	98.78	99.50	100.00	100.00
P (in)	0.1562	0.3904	0.00	0.00	0.00	0.08	2.60
PWD (Degrees)	173.9423	80.7063	0.00	135.00	191.25	225.00	337.50
AWS (mph)	3.0372	1.2623	1.10	2.10	2.90	3.70	7.30
MLC	30.1346	0.7912	28.90	29.20	30.40	30.80	31.10
M	82.7603	0.5242	81.90	82.30	82.70	83.20	83.60
TTU	147.5897	2.1909	143.90	145.10	148.90	149.70	150.30

The tables present a comprehensive overview of environmental, economic, and meteorological conditions across three distinct sites. Notably, the data span multiple scales, reflecting the varied nature of factors correlated with each site. Furthermore, an absence of outlying weather events during the data collection period underscores the reliability and consistency of the dataset. The absence of such anomalies ensures that any observed patterns or correlations are reflective of typical environmental and meteorological conditions experienced by the sites.

Table 1.4.5. Summary Statistics Site 3002

Variable	Mean	Std. Dev	Quantile				
			Min	25%	50%	75%	Max
NO_2 (ppm)	8.2828	4.9422	0	4	7	11.5	24
SO_2 (ppm)	1.3440	2.9516	0	0	1	2	42
O_3 (ppm)	0.0455	0.0111	0.017	0.037	0.046	0.053	0.075
AT (F)	74.3449	4.5063	59.2	71.65	74.8	77.7	83.4
MRH (%)	92.6475	6.9844	55.8	89.5	93.6	98.3	100
P (in)	0.1819	0.4656	0	0	0	0.1	3.4
PWD (Degrees)	176.8513	86.6420	0	135	2.6	47.1	337.5
AWS (mph)	3.3991	1.0812	1.6	2.6	3.2	4	7.1
MLC	48.6673	1.0859	46.7	47.1	49	49.5	49.6
M	118.0983	2.0849	115.3	115.6	117.3	119.5	121
TTU	215.1117	3.9249	209.1	210.7	214	218.7	219.3

The remaining part of this thesis is separated into chapters by analysis methods. Each chapter comprises a literature review, delineates the methodologies employed alongside their respective results, and discusses the outcomes and limitations of each methodology. The final chapter compares the three methods, the limitations of the study overall, and recommendations for future work.

CHAPTER 2

Grey Relational Analysis

2.1. Literature

Natural phenomena observations can present significant challenges for scientists as air quality and pollution data provide an incomplete perspective in regards to system variables, and sampling methods can be influenced by various environmental and economic factors. This concept is present in various systems, including the social, environmental, economic, and even human anatomical systems (Ng, 1994). Grey system theory, first proposed by Deng in 1983, aims to bridge the gap between social science and natural science (Julong, 1988). Grey information is either incomplete or undetermined, whereas black and white information is complete and lacking information, respectively. Grey system theory complements existing statistical methods such as fuzzy logic and rough set theory, and it helps establish a non-functional model for predicting variable levels and analyzing correlations. The idea of a 'grey' system stems from the color grey- black representing a complete lack of information, white representing complete information. Combined, we have an incomplete system, called 'grey'. Grey system theory deals with incomplete and uncertain information about systems, which is relevant to studying air pollution. This does not imply missing data but rather indicates the conditions of the system in place. A grey system has four defining characteristics: incomplete information about system parameters, incomplete structure, incomplete boundaries, and incomplete knowledge of system behaviors (Julong, 1988). These characteristics necessitate special analytical methods, such as grey relational analysis. According to Yamaguchi et al. (2015), grey relational analysis is one of grey system theory's most effective mathematical tools. It is particularly suited for air pollution analysis because it considers the similarities and differences between sets of data to rank each variable's impact on the system, resulting in a grey relational grade.

Applications of grey system theory are widespread and span various fields, including economics (C.-Y. Huang, Hsu, Chiou, & Chen, 2020), business (Wiecek-Janka, Mierzwiak, Nowak, Kujawinska, & Majchrzak, 2021), spatial planning practices (Gerus-Gościewska & Gościewski, 2022), environmental science (Pan, Sun, & Wang, 2011), agriculture, medicine, history, geography, industry, traffic, sports, and biological protection and water quality (J. Song, Guang, Li, & Xiang, 2016). Though many studies used principal components and multiple regression analysis (among other methods) to study this relationship, it is noted in the literature that there is a need for a method that considers the nature of data. Due to the influence of various environmental factors on air quality indicators, and the variability of primary air pollutants across different regions, there is a growing emphasis on examining the impact of pollutants on air quality from diverse viewpoints (Wang, Chen, Xie, & Wu, 2021). Grey relational analysis is a new approach to factor analysis and has been used in multiple different areas, specifically pollutants and air quality. For example, Wang et al. (2021) researched influence factors of air quality in the host area of the Olympic Games and found the three factors most closely related to the Beijing Air Quality Index are population, energy consumption, and motor vehicles. Pan used grey relational analysis to forecast city air quality and ran an impact factors analysis of various pollutants. They found industrial pollution and energy consumption had a great influence on pollutant concentrations (Pan et al., 2011). Zeydan and Pekkaya (2021) used grey relational analysis to study air quality monitoring stations and found air quality of stations located near industrial zones and thermal power plants is poorer than other monitoring sites. Zhan et al. (2018) looked at the correlation between air pollutant indicators and economic factors in Hong Kong. Ni (2013) found the main factors associated with air quality are industrial SO_2 emissions, public green space per capita, and total urban population. Li et al. (2018) looked at influence factors of air quality indexes in Beijing and other places for 5 years (2013-2017) and found rainfall, air temperature, and wind speed are associated with air quality index. Relative humidity and air pressure vary from city to city (Li, H & Wang, X., 2018).

2.2. Methodology

This section encompasses several key components, including an introduction to standardization and normalization techniques employed in data preprocessing. This discussion emphasizes the rationale behind these techniques and their application to ensure uniformity and comparability of the data across variables. Additionally, the section provides a detailed outline of the specific methodology utilized for Grey Relational Analysis.

2.2.1. Standardization and Normalization

Normalization and standardization are critical parts of data preparation, particularly in the context of GRA, where variables may exhibit diverse scales and units. This is further seen in Tables 1.4.3, 1.4.4, and 1.4.5 in Section 1.4. This process aims to standardize data to a common scale, facilitating meaningful comparisons and interpretations across variables. Within GRA, normalization establishes that each attribute contributes proportionally to the analysis, regardless of its initial magnitude. It involves transforming data values to conform to specific ranges, with considerations for variables where higher values indicate desirable outcomes, lower values denote preferable conditions, and moderate values are optimal.

2.2.1.1. Normalization and Standardization

Let y_k be the normalized value of the data point x. Let $j = 1, 2...n$, and $k = 0, 1, ..., m$, and x^* is the moderate value (Javed, Gunasekaran, & Mahmoudi, 2022).

When the data is higher the better, indicating higher values provide desirable outcomes, the following is used:

Type 'Higher the Better'

$$y_k(j) = \frac{x_k(j) - \min_j x_k(j)}{\max_j x_k(j) - \min_j x_k(j)}$$

In the context of this study, this normalization method is used for some meteorological variables because of how they interact with certain pollutants. See section 1.4.1 for more information.

When the data is lower the better, indicating lower values provide desirable outcomes, the following is used:

Type 'Lower the Better'

$$y_k(j) = \frac{\max_j x_k(j) - x_k(j)}{\max_j x_k(j) - \min_j x_k(j)}$$

In this study, the pollution data is normalized using this method, as pollution would ideally be as low as possible.

Type 'Moderate the Better'

$$y_k(j) = 1 - \frac{|x_k(j) - x^*|}{\max_j x_k(j) - \min_j x_k(j)}$$

(Javed et al., 2022)

This type of normalization is used when there is a pre-determined value the data should be at (Javed et al., 2022). This does not apply to any of our data, so it is not used.

Initial Value Processing (IVP)

According to Ng (1994) pg. 5, "the first value of each sequence to divide each succeeding value of the corresponding sequence" describes this pre-processing system.

$$y_k(j) = \frac{y_k(j)}{y_k(1)}$$

Mean Value Processing (MVP)

According to Ng (1994) pg. 5 , mean value processing works by finding the mean values of each primitive sequence, then those are used to divide each value of the corresponding sequences.

$$y_k(j) = \frac{y_k(j)}{y_{avg}}$$

Within this study, both IVP and MVP are used in separate analyses. This is to observe differences between grey relational orderings, if any. Table 2.2.1 represents the choices for normalization and standardization of variables by pollutant used within both GRA and DGRA analyses.

Table 2.2.1. Normalization and Standardization of Variables by Pollutant for GRA and DGRA

Variable	NO_2	SO_2	O_3
Pollutant	LTB	LTB	LTB
AT (F)	LTB	LTB	LTB
MRH (%)	LTB	LTB	HTB
P (in)	LTB	LTB	HTB
PWD	IVP/MVP	IVP/MVP	IVP/MVP
AWS (mph)	IVP/MVP	IVP/MVP	IVP/MVP
MLC	IVP/MVP	IVP/MVP	IVP/MVP
M	IVP/MVP	IVP/MVP	IVP/MVP
TTU	IVP/MVP	IVP/MVP	IVP/MVP

It is important to note that in various fields, including industry settings and environmental studies, the utilization of different standardization and normalization methods often hinges on the specific nature of the data and the objectives of the analysis. In industry settings, where data often adhere to well-defined standards, the choice of standardization or normalization method may be straightforward, guided by established protocols or industry best practices. For example, studies involving industry-related data, J.T. Huang & Liao (2003), Kuo, Yang & Huang (2008 and Muqeem et al. (2017) use higher the better and lower the better normalization methods when conducting GRA for each of their studies. However, in environmental data analysis, the literature presents a different normalization and standardization technique. For example, Shexia, Yaoqui, & Zhu (2018) used Initial Value Processing for water quality data when conducting GRA. Pan et al. (2011) made no mention of a normalization method used when conducting GRA with air quality index (AQI) data and economic data, and Tao (2015) used IVP or MVP for their study of both AQI and economic data. For studies using raw pollution data (as is used in this study), most often, the higher the better and lower the better techniques are used (Zeydan & Pekkaya, 2021). A study using exclusively economic data (C.-Y. Huang et al., 2020) used IVP and MVP techniques, and a general study of GRA (Ng, 1994) used IVP and MVP as well. Consequently, the selection of standardization or normalization techniques in environmental and economic studies often necessitates careful consideration of the characteristics of the data, the underlying assumptions of the analytical methods, and the research objectives. Thus, while different standardization and normalization methods may be employed across different contexts, in environmental data analysis, the choice seems to be driven by the unique complexities and nuances inherent in the data itself. That said, the choices in standardization and normalization made for this data are informed by the literature mentioned above and are outlined in Table 2.2.1. In all cases, pollutants are normalized as lower the better, while economic variables and some weather variables are normalized using either IVP or MVP. There is literature supporting the decision to characterize maximum relative humidity, average temperature, and precipitation to be normalized with higher and lower the better, from the perspective of wanting the pollutant's level to be lowered. For example, precipitation can remove pollutants from the atmosphere through wet deposition, while dry deposition can occur when particles settle on surfaces due to changes in temperature and humidity. Additionally, temperature and solar radiation can influence the formation and concentration of ground-level ozone (Environmental Protection Agency (EPA), 2021b). In choosing normalization and standardized methods for this study, because there is no consistency among normalization and standardization choices within the literature, decisions were made based on what the literature specified depending on the specific pollutant and type of factor being investigated. Palczewski & Salabun (2019) note that within Multi-criteria decision analysis, though normalization is required, the normalization choices do have an impact on final rankings, which can provide issues in critical areas that use Grey Relational Analysis, leading to incorrect results. To further investigate this, this study features both IVP and MVP results, along with the normalization methods presented in Table 2.2.1.

2.2.2. Grey Relational Analysis Methodology; Deng's Model of Grey Relational Analysis

Let us denote

$$X_0(k) = (x_0(1), x_0(2), ..., x_0(n)),$$

$$X_i(k) = (x_i(1), x_i(2), ..., x_i(n)),$$

where $k = 1, 2, 3..., n, i = 1, 2, 3, ...m$, and $\xi \in (0, 1)$. We can define

$$\alpha((x_0(k), x_i(k))) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \xi \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \xi \max_i \max_k |x_0(k) - x_i(k)|} \tag{2.1}$$

and

$$\gamma(X_0, X_i) = \frac{1}{n} \sum_{k=1}^{n} \alpha((x_0(k), x_i(k))) \tag{2.2}$$

X_0 is the reference sequence and X_i is the comparison sequence. In this study, the reference sequence is each pollutant level sequence, while the comparison sequence is each other variable. Now, α is the grey relational coefficient of each iteration of the comparison sequences, and $\gamma(X_0, X_i)$ is the grey relational degree between X_0 and X_i, which reflects the degree of closeness between the two comparing sequences. $\xi \in (0, 1)$ is the distinguishing coefficient, chosen to be .5 as indicated by literature. The purpose of the distinguishing coefficient is to weaken the effect of $\max_i \max_k |x_0(k) - x_i(k)|$ when it gets too big, which enlarges the difference significance of the relational coefficient (Ng, 1994). The closer the value of the grey relational degree (γ) is to 1 the higher the relational grade of the reference sequence; otherwise, it is lower (C.-Y. Huang et al., 2020). Once each grade is obtained, they are ordered from highest to lowest and called a grey relational order which indicates the most highly correlated variables in each set. This study utilizes Excel to carry out these computations.

2.3. Results

In this section, we present the results of our analysis focusing on air quality indicators at Site 67. For a full representation of the results, see Section 6.1. Given the scope of our study and to conserve space, we have chosen to concentrate solely on this site while acknowledging the broader dataset available in the discussion section. Our analysis encompassed both IVP and MVP for each pollutant examined.

TABLE 2.3.1. Site 67 NO_2 IVP/MVP GRA Results

Site 67 NO_2				
GRG γ with IVP		GRG γ with MVP		RANK
AT (F)	0.8039	AT (F)	0.8039	1
P (in)	0.7493	P (in)	0.7493	2
AWS (mph)	0.7389	AWS (mph)	0.7429	3
PWD	0.7384	MLC	0.7340	4
M	0.7319	TTU	0.7338	5
TTU	0.7228	M	0.7338	6
MRH (%)	0.7164	MRH (%)	0.7164	7
MLC	0.7159	PWD	0.6947	8

Table 2.3.1 reflects Site 67 NO_2 results. Initial value processing reveals average temperature (.8039), precipitation (.7493), and average wind speed (.7389) are the most highly correlated factors of NO_2 at Site 67. The least correlated factors are maximum relative humidity (.7164) and mining, logging and construction (.7159). Among economic variables, manufacturing (.7319) is the most correlated factor. Mean value processing reveals average temperature (.8039), precipitation (.7493), and average wind speed (.7429) are the most correlated factors of NO_2 at Site 67. The least correlated factors are maximum relative humidity (.7163) and mining, logging, and construction (.7159).

Table 2.3.2 reflects SO_2 results. Initial value processing reveals precipitation (.8836), manufacturing (.8787), and trade, transportation and utilities (.8665) are the most correlated factors of SO_2 at site 67. The least correlated factors are average temperature (.7388) and maximum relative humidity (.6207). Among weather variables, precipitation is the most correlated factor, while maximum relative humidity (.6207) is the least correlated. Among economic variables, manufacturing (.8835) is the most correlated factor. Mean value processing reveals precipitation (.8836), trade, transportation and utilities (.8814), and manufacturing (.8809) are the most correlated

Table 2.3.2. Site 67 SO_2 IVP/MVP GRA Results

Site 67 SO_2				
GRG γ with IVP		GRG γ with MVP		RANK
P (in)	0.8836	P (in)	0.8836	1
M	0.8787	TTU	0.8814	2
TTU	0.8665	M	0.8813	3
MLC	0.8572	MLC	0.8809	4
AWS (mph)	0.8020	AWS (mph)	0.8042	5
PWD	0.7912	PWD	0.7490	6
AT (F)	0.7388	AT (F)	0.7388	7
MRH (%)	0.6207	MRH (%)	0.6207	8

factors of SO_2 at site 67. The least correlated factors are average temperature (.7388) and maximum relative humidity (.6207). Among economic variables, trade, transportation and utilities (.8814) is the most correlated factor.

Table 2.3.3. Site 67 O_3 IVP/MVP GRA Results

Site 67 O_3				
GRG γ with IVP		GRG γ with MVP		RANK
AT (F)	0.8753	AT (F)	0.8753	1
PWD	0.7977	AWS (mph)	0.7985	2
MRH (%)	0.7968	MRH (%)	0.7968	3
AWS (mph)	0.7952	MLC	0.7916	4
M	0.7899	TTU	0.7915	5
P (in)	0.7857	M	0.7914	6
TTU	0.7826	P (in)	0.7857	7
MLC	0.7769	PWD	0.7600	8

Table 2.3.3 reflects O_3 results. Initial value processing reveals average temperature (.8753), prevailing wind direction (.7977), and maximum relative humidity (.7967) are the most correlated factors of O_3 at Site 67. The least correlated

factors are trade, transportation, and utilities (.7826) and mining, logging, and construction (.7769). Among economic variables, manufacturing (.7899) is the most correlated factor. Mean value processing reveals that average temperature (.8753), average wind speed (.7985), and maximum relative humidity (.7967) are the most correlated factors of O_3 at Site 67. The least correlated factors are precipitation (.7857) and prevailing wind direction (.7600). Among economic variables, mining, logging and construction (.7916) is the most correlated factor.

For O_3, both analyses highlight the significance of average temperature, average wind speed, and maximum relative humidity, with economic factors such as mining, logging, and construction also playing a notable role. Similarly, for SO_2, precipitation emerges as a consistently correlated factor in both IVP and MVP, alongside economic variables like trade, transportation, and utilities, as well as manufacturing. On the other hand, NO_2 results indicate the importance of average temperature, precipitation, and average wind speed, with manufacturing standing out as a key economic factor.

2.4. Discussion

In the discussion section, we comprehensively examine the findings from all sites, providing a thorough analysis of the factors correlated with air quality across different locations and discussing the method itself. While Site 67 is the sole focus of the results section for the sake of brevity, we discuss results from multiple sites.

Before discussing interpretations of rankings among each site and pollutant, it is important to discuss standardization methods and how those impact grey relational ordering. When comparing standardization methods, almost all rankings widely differ. Rankings are different among both weather and economic variables, as well as among variables of their kind (i.e., among just weather variables and just economic variables). In general, Site 5 exhibits the poorest performance in terms of disordered rankings. Site

67 and Site 3002 are less severe. Overall, for these sites, these rankings are consistent. Rankings are almost always the same among the individual groups of variables (for example, just weather and just economics) while among both types of variables, in some cases there is little variation in ordering, while in others there is much variation.

These results indicate that grey relational analysis is sensitive to preprocessing of data. Ideally, when conducting this analysis, the research team would opt for a single preprocessing method, and based on the results obtained, that method would prove to be adequate. MVP provides (generally) higher grey relational grades overall, and in some cases for Sites 67 and 3002, grades are the same. Additionally, GRA results in very similar grey relational grades among pollutants and standardization methods at each site. This may be because of the similarities between the normalization methods used and the similarity of the data used for each site, given the only levels that are altered in each round of computation are the pollutants themselves. Otherwise, the normalization methods only vary among NO_2 and SO_2 versus O_3 because Ozone interacts with some weather variables differently than other pollutants, so normalization and standardization are altered to reflect this. Even so, results indicate that GRA is still sensitive to the preprocessing of data.

For both IVP and MVP, results across Sites 67 and 3002 for nitrogen dioxide levels indicate the top three most correlated factors to be weather variables. Specifically, Site 5 has the opposite result- the top variables are primarily economic. Site 67 features consistently higher grey relational grades than the other two sites. Rankings are also different for Sites 67 and 3002 but are overall consistent among both standardization methods. While there is not a clear first ranking factor across all sites for NO_2, this could be due to location. Each site features different economic outputs and meteorological features.

For both IVP and MVP, results across all sites for sulfur dioxide indicate the top three most correlated factors to be economic variables or precipitation, and the dominating factor in two of three sites is manufacturing, followed by trade and transportation (despite IVP and MVP differences). Orderings are much more consistent with the results of this pollutant. It is also clear which factors are not as highly correlated with SO_2 levels; average temperature and maximum relative humidity consistently rank lowest. Because rankings resulting from GRA methodology are very similar among sites, we can conclude there is a clear conclusion for which factors have the most association with sulfur dioxide levels. In this case, manufacturing is the most correlated factor of SO_2 across all sites.

For IVP and MVP, results across all sites for Ozone levels indicate that weather variables are the top three most correlated factors. Grey relational grades are much lower at Site 67 than at Sites 5 and 3002. Ozone is the only pollutant that has consistent rankings among each site and standardization method, but rankings vary slightly between each site. This leads to the conclusion that Ozone is influenced primarily by weather variables, but those economic variables' rankings differ based on site location. The most correlated factors among weather variables are consistently average temperature and maximum relative humidity, which is consistent with the literature, though they do not rank higher than economic factors.

This discussion provides a comprehensive examination of findings from all sites. While Site 67 is the exclusive focus of the results section for brevity, results from multiple sites are discussed. When interpreting grey relational rankings, standardization methods and their impacts on grey relational ordering are important to consider. A notable disparity in rankings is observed across different standardization methods, with Site 5 demonstrating the most significant discrepancies. Moving forward, it is essential to acknowledge that these results suggest that grey relational analysis is sensitive to data preprocessing. Ideally, a singular preprocessing method would be chosen to ensure consistency in the analysis. Furthermore, the transition to discussing limitations is warranted, as the sensitivity of grey relational analysis to preprocessing methods may introduce inherent biases or limitations in the interpretation of results.

2.5. Limitations

The Grey Relational Analysis methodology presents a drawback due to the necessity for data preprocessing, as highlighted in the literature. When discussing the nature of GRA and pollutant type, J. T. Huang & Liao (2003, pg. 1711) say

"When the range of the sequence is too large or the standard value is too enormous, it will cause the influence of some factors to be neglected. In addition, in the sequence, if the factors' goals and directions are different, the relational analysis might also produce incorrect results. Therefore, preprocessing of all the data is necessary." The language used in this quote indicates the specific criteria for determining the magnitude of data ranges and factor goals remain ambiguous. While preprocessing is deemed necessary, there is a lack of consensus on which normalization processes to employ. Unlike in industry settings, where standardization or normalization methods are straightforwardly chosen based on established protocols and uniform distributions, the choice in environmental studies lacks convention, leading to potential variability in results and interpretations. Additionally, Javed et al., (2022) indicate that within the field of multi-criterion decision-making, it is "not unusual to obtain different rankings using different normalization techniques" (pg. 6). This is further expanded upon within Chapters 3 and 5. Additionally, the distinguishing coefficient ξ is blindly chosen to be .5, with little to no contextual support for this decision. This proves to be an important aspect in future analysis, specifically when looking at Dynamic Grey Relational Analysis methodology and results.

2.6. Conclusion

In conclusion, the findings from our analysis (using both IVP and MVP), shed light on the complex dynamics associated with air quality across different sites in Kentucky. The examination of Site 67, in particular, reveals insights into the significant factors affecting pollutant levels, with weather variables often playing a prominent role. However, economic factors also emerge as correlated, albeit with variations across sites. The sensitivity of grey relational analysis to preprocessing methods underscores the importance of careful data preparation to ensure reliable results. While MVP generally yields higher grey relational grades, both IVP and MVP highlight consistent trends in pollutant influences, albeit with variations in rankings among sites. Notably, for nitrogen dioxide, weather variables predominate as most associated with NO_2 levels across Sites 67 and 3002, contrasting with Site 5, where economic factors hold greater sway. Conversely, sulfur dioxide levels are predominantly influenced by economic variables or precipitation across all sites, with manufacturing consistently emerging as the most correlated factor. Finally, Ozone levels demonstrate a notable consistency in rankings among sites and standardization methods, indicating a primary association with weather variables, particularly average temperature and maximum relative humidity. These findings underscore the complex interplay of environmental and economic factors in shaping air quality, highlighting the need for tailored strategies to mitigate pollution and safeguard public health. However, it is crucial to acknowledge the limitations inherent in our methodology, including the potential biases introduced by data preprocessing and the constraints of our analysis focusing on a select number of sites. Future research endeavors should strive to address these limitations.

CHAPTER 3

Dynamic Grey Relational Analysis

3.1. Literature

Scientists face many challenges when observing natural phenomena, particularly in assessing air quality and pollution data, which may be incomplete due to various environmental and economic factors associated with sampling methods. Grey system theory, introduced by Deng in 1983, aims to bridge the gap between social science and natural science by addressing incomplete or undetermined 'grey' information, contrasting with complete 'black' and lacking 'white' information. This theory complements existing statistical methods like fuzzy logic and rough set theory, providing a non-functional model for predicting variable levels and analyzing correlations. Applications of grey system theory are extensive, spanning fields such as economics, business, spatial planning, environmental science, agriculture, medicine, history, geography, industry, traffic, sports, and biological protection, as well as water quality. Despite the use of principal components and multiple regression analysis in studying air quality relationships, there is a recognized need for methods that account for the nature of the data. Grey relational analysis emerges as a novel approach, applied across various areas and pollutants to forecast air quality and investigate associated with factors. Studies employing this method reveal insights into the impact of industrial pollution, energy consumption, population density, motor vehicles, economic factors, and environmental conditions on air quality indices in different regions. An area of grey system theory and a derivative of GRA, called Dynamic Grey Relational Analysis (DGRA) can also be used for this study to assess air pollution data, rather than traditional methods such as regression or principal components analysis, which may have issues in handling incomplete data, as indicated by previous literature. As discussed in Section 2.5, GRA has shortcomings such as the value of its distinguishing coefficient; it is consistently taken to be .5 (Ouali, 2022). "Also, why after almost four decades, still no one identified this inconsistency could be attributed to the fact that the scholars mostly assume $\xi = 0.5$" (Javed et al. 2022, pg. 5) . Recent literature has argued a dynamic method is needed to compare the effects of sequential data movement (C.-Y. Huang et al., 2020). DGRA dynamically changes the distinguishing coefficient ξ and calculates it optimally (Javed et al., 2022). This methodology is thought to be more precise than typical GRA models (Bai, Jin, Wang, Wang, & Xu, 2020). As noted by Huang et al. (2020) pg. 3, "The literature related to grey relation has employed static analysis in data studies, without considering leading or lagging time" . Thus, DGRA can be used to compare the effects of sequential data movement (C.-Y. Huang et al., 2020). DGRA is an innovative take on grey relational analysis and addresses issues with Deng's model. Some studies have noted discrepancies between traditional GRA and DGRA in final grades (C.-Y. Huang et al., 2020). Ongoing research aims to determine which yields the most accurate results, with this study contributing to the discussion. There is limited literature about the reliability of DGRA, as DGRA is a new development in the field of Multi-Criteria Decision Making (MCDM) and Grey System Theory.

3.2. Methodology

This section encompasses several key components, including standardization and normalization techniques employed in data preprocessing. This discussion emphasizes the rationale behind these techniques and their application to ensure uniformity and comparability of the data across variables. Additionally, the section provides a detailed outline of the specific methodology utilized for Dynamic Grey Relational Analysis.

3.2.1. Standardization and Normalization

Normalization and standardization are a critical part of data preparation, particularly in the context of Dynamic Grey Relational Analysis, where variables may exhibit diverse scales and units. This process aims to standardize data to a common scale, facilitating meaningful comparisons and interpretations across variables. Within DGRA, normalization ensures that each attribute contributes proportionally to the analysis, regardless of its initial magnitude. It involves transforming data values to conform to specific ranges, with considerations for variables where higher values indicate desirable outcomes, lower values denote preferable conditions, and moderate values are optimal. Normalization and standardization processes are the same as those presented in Section 2.2. See Table 2.2.1 for normalization and standardization choices.

Dynamic Grey Relational Analysis offers a refined approach to address the challenges present in Grey Relational Analysis, particularly concerning standardization and normalization procedures. Unlike GRA, where normalization is often a mandatory step to ensure comparability among variables with differing scales, DGRA presents a flexible modeling framework wherein normalization is not obligatory (Ouali, 2022; Javed et al., 2022). Javed et al. (2022) give the option not to normalize data using higher the better, lower the better, and moderate the better (what is referred to as the Ideal Alternative Function), but only when the research team "lack consensus on the choice of normalization technique." Briefly mentioned priorly, normalization of data is needed when the removal of the effects of different measurement units needs to be reduced (Bai et al., 2020). In this project, normalization is needed because of the vastly different measurement types. Interestingly, (C.-Y. Huang et al., 2020) suggests the use of various standardization methods, though this study only examines economic variables when performing DGRA. (Ouali, 2022) does not mention which normalization methods are used for their industry-type data when performing DGRA. Nonetheless, there is no consistency among normalization and standardization choices within the literature, and the choices made for these projects are dependent upon the specific pollutant. Palczerski & Salabun (2019) note that within Multi-Criteria Decision Analysis, though normalization is required, the normalization choices do have an impact on final rankings, which can provide issues in critical areas that use Grey Relational Analysis, leading to incorrect results.

Thus, because of the ambiguity in normalization and standardization procedures, to stay consistent in pre-processing techniques, the choices made for DGRA methodology and these data are informed by the literature mentioned in Section 2.2 as outlined in Table 2.2.1, and are the same as GRA choices.

3.2.2. Dynamic Grey Relational Analysis Methodology

First, we establish a reference sequence and a comparative sequence

$$X_0(k) = x_0(1), x_0(2), ..., x_0(n),$$

$$X_i(k) = x_i(1), x_i(2), ..., x_i(n),$$

where $k = 1, 2, 3..., n, i = 1, 2, 3, ...m,$ $X_0(k)$ represents a reference sequence and $X_i(k)$ is a comparative sequence (Javed et al., 2022). After normalization, the Dynamic Grey Relational Grade is

$$\gamma_{0k} = \frac{1}{n}\sum_{j=1}^{n} \alpha_{0k}(j), \tag{3.1}$$

where α_{0k} (Grey Relational Coefficient) is

$$\alpha_{0k}(j) = \frac{\Delta_{\min} + \xi(j)_{0k}\Delta_{\max}}{|\Delta_{0k}(j)| + \xi(j)_{0k}\Delta_{\max}}, \xi(j) \in (0,1], k = 1, 2, ..., m, \tag{3.2}$$

where

$$|\Delta_{0k}(j)| = |x_0(j) - x_k(j)| \tag{3.3}$$

$$\Delta_{\min} = \min_k \min_j |x_0(j) - x_k(j)| \tag{3.4}$$

$$\Delta_{\max} = \max_k \max_j |x_0(j) - x_k(j)| \tag{3.5}$$

$$\xi(j) = \{\xi(1), \xi(2)...\xi(n)\}, \xi(j) \in (0,1]. \tag{3.6}$$

$$\phi(j) = \frac{\frac{1}{m}\sum_{k=1}^{m} |x_0(j) - x_k(j)|}{\Delta_{\max}}. \tag{3.7}$$

$\xi(j)$ is the vector of dynamic distinguishing coefficients, not a fixed value as in the GRA methodology. To determine these, we estimate an h-multiplier. $h \in [1,2]$ is a unique continuous multiplier that defines the relative position of each coefficient in the set $\alpha(j)$. To determine the dynamic distinguishing coefficients, a one-variable linear programming method of maximizing (Javed et al., 2022) is used:

$$\text{Maximize } \xi(j) = h(\phi(1) + \phi(2) + ... + \phi(n)) \text{ s.t. } h \in [1,2] \text{ and } h\phi(j) \leq 1. \tag{3.8}$$

Furthermore, to properly calculate the distinguishing coefficients, the weight of each criterion ω must be calculated. Javed et. al (2022) note the use of the Ordinal Priority Approach (OPA) to Multi-criterion decision making to calculate the weights of the criteria based on preference data (Ataei, Mahmoudi, Feylizadeh, & Li, 2020), though because there is no preference for rankings of each pollutant, the mathematics behind this simplifies to the weight being $\frac{1}{n}$, where n is the number of variables included in the study, as indicated within the methodology section. Typically, multiple experts would use data to develop a ranking as to which criteria are most and least important, then estimate those scores and input that information into computer software to calculate each weight. This coupled with data type classification in the normalization process allows for dynamic calculation of ξ, though again, in this study, weights are calculated as $\frac{1}{n}$.

All computations for this calculation are conducted in Matlab using Javed et al. (2022) Matlab code developed for replication of their study. No changes are made to the code, though we test multiple algorithmic methods for the linear programming method (linprog) used with Matlab, and results do not differ when utilizing different algorithms, such as the Dual-Simplex Algorithm, the Interior-Point-Legacy Algorithm, and the Interior-Point Algorithm. The global grey relational variators are $h\phi(j)$. The calculation method is carried out using the $\xi(j)$ from Matlab and Excel for the rest of the computation. γ is then calculated. Now, α is the grey relational coefficient of the comparison sequences which reflects the degree of closeness between the two comparing sequences. The closer the value of the grey relational grade is to 1 the higher the relational grade of the reference sequence; otherwise, it is lower (C.-Y. Huang et al., 2020). Once each grade is obtained, they are ordered from highest to lowest and called a grey relational order which indicates the most correlated factors affecting various variables.

3.3. Results

The results section presents findings regarding both Initial Value Processing and Mean Value Processing for Site 67 across various pollutants. Due to space constraints, only results for Site 67 are detailed in this section, with comprehensive data for other sites available in the appendix. Analysis of Site 67 reveals notable patterns and trends for each pollutant examined.

Table 3.3.1 reflects NO_2 results. Initial value processing reveals average temperature (γ=.8857, ξ=.9965), maximum relative humidity (γ=.8301, ξ=1), and

TABLE 3.3.1. Site 67 NO_2 IVP/MVP DGRA Results

Site 67 NO_2						
ξ	GRG (γ) with IVP		ξ	GRG (γ) with MVP		RANK
0.9965	AT (F)	0.8857	0.9965	AT (F)	0.8857	1
1.0000	MRH (%)	0.8301	1.0000	MRH (%)	0.8301	2
0.4536	AWS (mph)	0.7218	0.4484	AWS (mph)	0.7239	3
0.2594	PWD	0.6119	0.3017	PWD	0.5911	4
0.0184	P (in)	0.1362	0.0184	P (in)	0.1362	5
0.0176	MLC	0.0995	0.0169	MLC	0.1113	6
0.0129	TTU	0.0805	0.0126	TTU	0.0871	7
0.0101	M	0.0719	0.0100	M	0.0733	8

average wind speed (γ=.7218, ξ=.4536), are the most correlated factors of NO_2 at Site 67. The least correlated factors are trade, transportation, and utilities (γ= .0805, ξ=.0129), and manufacturing (γ= .0719, ξ=.0101). Among economic variables, mining, logging, and transportation is the most correlated. Standardization methods present the same orderings with very similar grey relational grades. Mean value processing reveals average temperature (γ=.8857, ξ=.9965), maximum relative humidity (γ=.8301, ξ=1), and average wind speed (γ=.7239, ξ=.4484) are the most correlated factors of NO_2 at Site 67. The least correlated factors are trade, transportation and utilities (γ= .0871, ξ=.0126), and manufacturing (γ= .0733, ξ=.0100). Among economic variables, mining, logging and transportation is the most correlated.

Table 3.3.2 reflects SO_2 results. Initial value processing reveals average temperature (γ=.8451, ξ=.9965), maximum relative humidity (γ=.7638, ξ=1), and average wind speed (γ=.7885, ξ=.4484), are the most correlated factors of NO_2 at Site 67. The least correlated factors are trade, transportation, and utilities (γ= .1998, ξ=.0129), and manufacturing (γ= .1994, ξ=.0101). Among economic variables, mining, logging, and transportation are the most correlated. Standardization methods present the same orderings with very similar grey relational grades. Mean value processing

TABLE 3.3.2. Site 67 SO_2 IVP/MVP DGRA Results

Site 67 SO_2						
ξ	GRG (γ) with IVP		ξ	GRG (γ) with MVP		RANK
0.9965	AT (F)	0.8451	0.9965	AT (F)	0.8451	1
0.4536	AWS (mph)	0.7879	0.4484	AWS (mph)	0.7885	2
1.0000	MRH (%)	0.7638	1.0000	MRH (%)	0.7638	3
0.2594	PWD	0.6778	0.3017	PWD	0.6525	4
0.0184	P (in)	0.3173	0.0184	P (in)	0.3173	5
0.0176	MLC	0.2321	0.0169	MLC	0.2995	6
0.0129	TTU	0.1998	0.0126	TTU	0.2486	7
0.0101	M	0.1994	0.0100	M	0.2070	8

reveals average temperature (γ=.8451, ξ=.9965), average wind speed (γ=.7884, ξ=.4484), and maximum relative humidity (γ=.7638, ξ=1) are the most correlated factors of NO_2 at Site 67. The least correlated factors are trade, transportation, and utilities (γ= .2485, ξ=.0126) and manufacturing (γ= .2070, ξ=.01). Among economic variables, mining, logging, and transportation is the most correlated.

TABLE 3.3.3. Site 67 O_3 IVP/MVP DGRA Results

Site 67 O_3						
ξ	GRG (γ) with IVP		ξ	GRG (γ) with MVP		RANK
1.0000	AT (F)	0.9023	1.0000	AT (F)	0.9023	1
0.4552	AWS (mph)	0.7132	0.4500	AWS (mph)	0.7154	2
0.4083	MRH (%)	0.6963	0.4083	MRH (%)	0.6963	3
0.2968	P (in)	0.6026	0.2968	P (in)	0.6026	4
0.2603	PWD	0.6015	0.3027	PWD	0.5808	5
0.0177	MLC	0.0869	0.0170	MLC	0.0930	6
0.0130	TTU	0.0681	0.0127	TTU	0.0720	7
0.0101	M	0.0580	0.0101	M	0.0584	8

Table 3.3.3 reflects O_3 results. Initial value processing reveals average temperature (γ=.9023, ξ=1), maximum relative humidity (γ=.6963, ξ=.4083), and average wind speed (γ=.7132, ξ=.4552) are the most correlated factors of NO_2 at Site 67. The least correlated factors are trade, transportation, and utilities (γ= .0681, ξ=.0130) and manufacturing (γ= .5800, ξ=.0101). Among economic variables, mining, logging, and transportation is the most correlated. Standardization methods result in the same orderings with the same grey relational grades. Mean value processing reveals average temperature (γ=.9023, ξ=1), maximum relative humidity (γ=.6963, ξ=.4083), and average wind speed (γ=.7154, ξ=.4500), are the most correlated factors of NO_2 at Site 67. The least correlated factors are trade, transportation, and utilities (γ= .0720, ξ=.0127), manufacturing (γ= .0584, ξ=.0101). Among economic variables, mining, logging, and transportation is the most correlated.

In summary, the results from both MVP and IVP provide insights into the factors associated with air quality at Site 67. For ozone, average temperature, maximum relative humidity, and average wind speed emerge as the most correlated factors, with economic variables such as mining, logging, and construction with significant association. Notably, standardization methods yield consistent orderings and grey relational grades across both MVP and IVP analyses for O_3. Similarly, for sulfur dioxide, weather variables like average temperature and average wind speed are correlated, along with economic factors such as mining, logging, and construction. Standardization methods again produce uniform orderings and grey relational grades for SO_2. Finally, nitrogen dioxide levels are predominantly influenced by weather variables across both MVP and IVP analyses, with economic variables also playing a role. The consistency in orderings and grey relational grades across standardization methods emphasizes the reliability of the findings. Moving forward, these results provide a foundation for discussing the implications of the identified factors on air quality and informing future research directions.

3.4. Discussion

In the discussion section, we comprehensively examine the findings from all sites, providing a thorough analysis of the factors associated with air quality across different locations and discussing the method itself. While Site 67 is the sole focus of the results section for the sake of brevity, we now discuss results from multiple sites.

Before discussing interpretations of rankings among each site and pollutant, it is important to discuss standardization methods and how those impact dynamic grey relational ordering. Overall, grey relational orderings resulting from DGRA are consistent among IVP and MVP, while Site 5 has the most variation in grey relational orderings across all pollutants. Sites 67 and 3002 present the same orderings with very similar grey relational grades.

Despite the variation in ordering of Site 5, the dynamic grey relational analysis seems to be only very slightly sensitive to the preprocessing of data, specifically in the grey relational grades themselves. IVP processing generally provides higher grey relational grades than MVP. The method employed results in consistent grey relational grades for pollutants, notably NO_2 and SO_2, across different sites, possibly due to uniform normalization methods and data consistency, except for adjustments made for O_3 due to its distinct nature. In this sense, DGRA is still sensitive to preprocessing of data, but on a much lower level than GRA.

For IVP and MVP, results across all sites for nitrogen dioxide levels indicate the top three most correlated factors are weather variables. In contrast, economic variables fall toward the end of the ranking. The prevailing factors are average wind speed, average temperature, and maximum relative humidity. Grades are similar across sites, though Site 5 has the most dissimilar ranking of factors. There is consistent variation in ranks 3-5 across each site. For example, maximum relative humidity ranks second at Site 67, while it ranks fourth at both Site 5 and Site 67. Though the top three factors are weather variables, there is slight variation among which variables rank higher across each site, lending a conclusion to be drawn about the reflection of site differences in the results.

For IVP and MVP, results across all sites for sulfur dioxide levels indicate the top three most correlated factors are weather variables. In contrast, economic variables fall towards the end of the ranking. Average wind speed, average temperature, and maximum relative humidity are the prevailing factors. Grades are similar across each site, though site 5 has the most dissimilar ranking of factors. There is consistent variation in ranks 3-5 across each site. For example, precipitation ranks third at Site 3002 while at Sites 67 and 5, it ranks in the fifth position. Though the top three factors are weather variables, there is slight variation among which variables rank higher across each site.

For IVP and MVP, results across all sites for ozone levels indicate the top three most correlated factors to be weather variables. In contrast, economic variables fall toward the end of the ranking order. The prevailing factors are average wind speed, prevailing wind direction, and maximum relative humidity. There is consistent variation in ranks 3-5 across each site. For example, the prevailing wind direction ranks third at Site 5, while it ranks fifth and fourth at Site 67 and Site 3002, respectively. Though the top three factors are weather variables, there is slight variation among which variables rank higher across each site, lending a conclusion to be drawn about the reflection of site differences and pollutant changes in the results.

3.5. Limitations

Each criterion must be classified as a lower the better, higher the better, or moderate the better type. This indicates how data is to be normalized, though this classification (as indicated by literature) is not necessarily clear when analyzing multiple types of data sources. Similarly to GRA, a decision is made regarding the normalization methods based on several literature sources because of the diverse types of variables investigated in this study. Javed et. al (2022), pg. 7 also mentioned the need of consideration of "considering the uncertainty in the input data, it is suggested that instead of taking the face value of [the Dynamic Grey Relational Grade], uncertainty quantification should be done through the Grey Relational Standard Deviation." This is not considered for this study, as there is no comparable method to utilize for both GRA and CCA.

3.6. Conclusion

In conclusion, the findings from our analysis (using both IVP and MVP), shed light on the complex dynamics associated with air quality across different sites in Kentucky. While Site 67 receives primary focus in the results section for the sake of brevity, our subsequent discussion encompasses results from all sites, offering a broader perspective on air quality dynamics. The sensitivity of dynamic grey relational analysis to preprocessing methods underscores the importance of careful data preparation to ensure reliable results. Overall, our analysis reveals consistent orderings resulting from Dynamic Grey Relational Analysis among both IVP and MVP results, with Site 5 exhibiting the most variation in grey relational orderings across pollutants. However, Sites 67 and 3002 present uniform orderings with very similar grey relational grades. Despite slight variations in rankings, our findings suggest that DGRA is only minimally sensitive to data preprocessing, underscoring the reliability of the method. Moving forward, the consistency in grey relational grades across pollutants at each site suggests the robustness of DGRA in assessing air quality dynamics. Subsequent analysis reveals weather variables to be associated with NO_2, SO_2, and O_3 levels, with economic variables exhibiting lesser association.

CHAPTER 4

Canonical Correlation Analysis

4.1. Literature

Canonical Correlation Analysis (CCA) stands as a powerful multivariate statistical technique used to explore the relationships between sets of variables. Unlike univariate analyses that focus on single variables, CCA aims to uncover the associations between two sets of variables simultaneously, revealing underlying patterns of correlation between them. A goal of CCA is to "establish the maximum correlation among sets of variables" (Statheropoulos, Vassiliadis, & Pappa, 1998, pg. 1088). This methodology is particularly useful when dealing with complex datasets comprising multiple interrelated variables, as it allows researchers to identify latent relationships and extract meaningful information from high-dimensional data. A similar method employed with data of this type is Principal Components Analysis (PCA) and multiple regression. CCA differs from PCA in that PCA seeks to maximize variance within each set of variables independently, whereas CCA maximizes the correlation between the sets (Statheropoulos, Vassiliadis, & Pappa, 1998). Similarly, CCA differs from regression analysis in that it does not assume a dependent-independent variable relationship but rather explores the overall association between two sets of variables. Abdul-Wahab et al. (2005) used principal component and multiple regression analysis in modeling ground-level ozone and factors affecting its concentration. They found that while high temperature and high solar energy tended to increase the data time ozone concentrations, the pollutants NO_2 and SO_2 being emitted to the atmosphere are being depleted. Abdul-Wahab et al. (2005) also mentioned significant amounts of multicollinearity among pollutant predictors, which makes multiple of their models inappropriate. Multiple transformations are needed on the models, specifically logarithmic transformations. Though multiple regression analysis is a widely used methodology for showing relationships between predictor variables and the response variables, it can face serious difficulties when independent variables are correlated with each other (Mcadams et al., 2000). Other studies have mentioned using artificial neural networks in combination with stepwise regression analysis which provides a better analysis of air quality detectors themselves (Liu, Zhao, Jin, Shen, & Li, 2021) and the use of Pearson correlation coefficient methods, specifically in its limitations–it can only reflect the "linear relationship between two variables, and the direction of the correlation has limitations" (Wang et al. 2021, pg. 3). Zyromski et al. (2014) used a bivariate correlation matrix to investigate the relationship between meteorological factors and air pollutants and found that wind speed is the most correlated factor. (Shihab, 2022). Similarly, Habeebullah et al. (2015)used canonical correlation analysis to achieve the same goal and concluded there is a contribution from meteorological factors to pollutant emission (Habeebullah, Munir, Awad, Seroji, & Mohammed, 2015; Shihab, 2022). The versatility of CCA makes it applicable across various domains, including environmental sciences. In the context of air pollution and meteorological data analysis, CCA offers a comprehensive approach to discern the intricate interactions between air quality parameters, meteorological variables, and other relevant factors. By uncovering the underlying relationships between these sets of variables, CCA can provide valuable insights into the complex dynamics of air pollution and weather patterns, facilitating informed decision-making and policy interventions aimed at mitigating environmental risks and promoting public health. CCA has primarily been used in education, psychology, marketing, and sociology (Laessig & Duckett, 1979).

Multiple studies have focused on air pollution and meteorological data. Statheropoulos et al. (1998) explore the application of PCA and CCA in assessing variability in air quality data. The study focuses on the analysis of air quality parameters, such as sulfur dioxide, nitrogen dioxide, and particulate matter (PM10), collected from monitoring stations in Thessaloniki, Greece. The results of the PCA found that air pollution data is related to gasoline combustion, oil combustion, and ozone interactions, while the most prominent component in meteorological data is dry conditions and high-speed western winds. Canonical correlation analysis reveals a relationship between pollution and high humidity along with low wind speed (Statheropoulos et al., 1998). Additionally, Binaku & Schmeling (2017) investigated air pollutants concerning meteorological factors during summer months using CCA. The analysis reveals low wind speed influenced air pollutant variables as well as wind direction. Borowiak, Zbierska, & Jusik (2011) aimed to study correlations between tropospheric ozone concentration, wind speed, temperature, solar radiation, and leaf injury using CCA. They found CCA may be a useful tool for bioindication studies. Canonical Correlation Analysis has been used in many settings, and its application to this project may prove useful in discerning complex relationships among pollution, meteorology, and economic factors.

4.2. Methodology

4.2.1. CCA methodology

Let

$$X = (X_1, ..., X_p),\ Y = (Y_1, ..., Y_q)$$

denote random vectors. X and Y are based on the number of variables in each set. Now define a set of linear combinations named U and V, where U corresponds to X, and V corresponds to the second set of variables, Y (*Lesson 13: Canonical Correlation Analysis*, n.d.).

$$U_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p$$

$$U_2 = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p$$

$$\vdots$$

$$U_p = a_{p1}X_1 + a_{p2}X_2 + ... + a_{pp}X_p$$

$$V_1 = a_{11}Y_1 + a_{12}Y_2 + ... + a_{1q}Y_q$$

$$V_2 = a_{21}Y_1 + a_{22}Y_2 + ... + a_{2q}Y_q$$

$$\vdots$$

$$V_q = a_{q1}Y_1 + a_{q2}Y_2 + ... + a_{qq}Y_q.$$

So we can define (U_i, V_i) as the i^{th} canonical variate pair. Then, computing the variance of U_i and V_j,

$$\text{var}(U_i) = \sum_{k=1}^{p}\sum_{l=1}^{q} b_{ik}b_{il}\text{cov}(X_k, X_l)$$

$$\text{var}(V_j) = \sum_{k=1}^{p}\sum_{l=1}^{q} b_{jk}b_{jl}\text{cov}(Y_k, Y_l),$$

and calculating the correlation between U_i and V_j we have

$$\frac{\text{cov}(U_i, V_j)}{\sqrt{\text{var}(U_i)\text{var}(V_j)}}$$

Then the canonical correlation for the i^{th} canonical variate pair is the correlation between U_i and V_i

$$\rho_i = \frac{\text{cov}(U_i, V_i)}{\sqrt{\text{var}(U_i)\text{var}(V_i)}}$$

(*Lesson 13: Canonical Correlation Analysis*, n.d.).

4.2.2. Interpretation of Results

Assumptions in canonical correlation analysis encompass various aspects such as multivariate normality and homogeneity of variance. Multivariate normality is affirmed based on the ample data available, ensuring that the data points follow a normal distribution across multiple variables (Sherry & Henson, 2005). A canonical function is a mathematical construct comprising standardized coefficients derived from two linear equations representing the observed predictor and criterion variables. These coefficients are optimized to maximize the canonical correlation, thereby capturing the underlying relationship between the two sets of variables. Each canonical function corresponds to a distinct linear combination of the observed variables. Similar to principal component analysis where each component captures a different aspect of variance in the data, each canonical function in canonical correlation analysis encapsulates a specific pattern of association between the predictor and criterion variables (Samuels, n.d.). Moreover, the number of canonical functions is determined by the smaller of the two variable sets. This ensures that all relevant information from both sets is accounted for in the analysis. Each canonical function serves as a basis for constructing synthetic variables, which are linear combinations of the observed variables. These synthetic variables are used to compute the canonical correlation coefficient and assess the strength of the relationship between the predictor and criterion variable sets.

The analyzed data comprise observed predictor variables and criterion variables. Weather and economic variables are the predictor variables within this study, while pollutant variables are the criterion variables. Synthetic variables, employed as placeholders for dimension reduction, are used in determining the canonical correlation coefficient. These synthetic variables represent shared variance between predictor and criterion variable sets, akin to the effect size represented by R_c^2 in multiple regression analysis (Sherry & Henson, 2005). Canonical correlation coefficients (R_c) denote the Pearson correlation between two synthetic variables derived from canonical functions, exhibiting a positive range from 0 to 1. These coefficients, analogous to multiple R in regression, determine the relationship between the predictor and criterion variable sets. Squared canonical correlation (R_c^2) represents the variance accounts for effect size shared by each synthetic variable (Sherry & Henson, 2005). This measure is interpreted similarly to the R^2 in multiple regression. Standardized canonical function coefficients are utilized in linear equations to combine observed variables, maximizing canonical correlation. They are interpreted similarly to beta weights in regression. Structure coefficients (r_s) represent the relationship between the observed and synthetic variables, akin to structure coefficients in factor analysis or multiple regression. They help to define the structure of the synthetic variable (Sherry & Henson, 2005). The squared canonical structure coefficients specify how much variance a predictor variable shares with a synthetic variable (Sherry & Henson, 2005). The canonical commonality coefficient (h^2) reflects the proportion of variance explained by the complete canonical solution, gauging the utility of observed variables in the analysis. This is found by summing the squared canonical structure coefficient.

4.3. Results

We now present a comprehensive analysis encompassing all sites and pollutants, providing valuable insights into the complex relationships between environmental variables and air quality. Each table includes canonical coefficients, structure coefficients, squared canonical structure coefficients, and the commonality coefficient, offering an examination of the interrelationships among variables. More detailed results are available in the appendix.

4.3.1. Site 5

A canonical correlation analysis is conducted using the three *economic variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., economic and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .0315, .0212, and .001 for each successive function.

Collectively, only the first dimension of the model across all functions is statistically significant using Wilk's λ = .947 criterion, F(9,759.477) = 1.905, p = .0482. Because Wilk's λ represents the variance unexplains by the model, $1 - \lambda$ yields the full model effect size in an R^2 metric. Thus, for the set of one canonical function, the R^2 type effect size is 0.0529, which indicates that the full model explains a small portion, only about 5.2%, of the variance shared between the variable sets.

The dimension reduction analysis allows for testing of the hierarchical arrangement of functions for statistical significance (*Canonical Correlation Analysis in Detail*, 2018). As noted, the full model (Functions 1 to 3) is statistically significant. Function 2 to 3 do not explain a statistically significant amount of shared variance between the variable sets, $F(4, 626) = 1.759$, $p = .135$. Additionally, Function 3 (which is the only function that is tested in isolation) does not explain a statistically significant amount of shared variance between the variable sets, $F(1, 314) = .300$, $p = .5843$. Given the (R_c^2) effects for each function, none of the functions could be considered noteworthy in the context of this study ($3.15\%, 2.12\%$, and .09% of the shared variance, respectively).

TABLE 4.3.1. Canonical Solution for Weather Predicting Pollutant Levels for Functions 1, 2, and 3 at Site 5

	Function 1			Function 2			Function 3			
Variable	*Coef*	r_s	$r_s^2(\%)$	*Coef*	r_s	$r_s^2(\%)$	*Coef*	r_s	$r_s^2(\%)$	$h^2(\%)$
AT	0.2125	0.1042	1.0850	0.5784	0.1783	3.1787	-0.4071	-0.1191	1.4193	5.6829
MRH	0.3444	0.0287	0.0823	0.0413	-0.0214	0.0457	0.5348	0.1756	3.0849	3.2129
P	0.3679	0.2559	6.5502	0.0207	0.0410	0.1684	0.5416	0.1412	1.9929	8.7115
PWD	0.2362	0.2209	4.8813	-0.8712	-0.2569	6.5998	-0.2498	-0.0939	0.8822	12.3633
AWS	0.8194	0.4316	18.6241	0.2914	0.0184	0.0338	-0.1306	-0.0793	0.6281	19.2860
O_3	-0.4375	-0.3603	12.9842	0.1805	0.0014	0.0002	-0.9555	-0.1859	3.4559	16.4403
NO_2	-0.4710	-0.4174	17.4248	-0.9225	-0.1881	3.5388	0.4284	0.0638	0.4067	21.3703
SO_2	-0.4722	-0.3764	14.1655	0.8503	0.1863	3.4724	0.4396	0.1087	1.1806	18.8185

A canonical correlation analysis is conducted using the *five weather variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., weather and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .283, .110, and .064 for each successive function. Collectively, the full model across all functions is statistically significant using the Wilk's λ = .597 criterion, F(15, 856.175) = 11.722, p < .001. For the set of one canonical function, the R^2 type effect size is 0.403, which indicates that the full model explains a substantial portion, about 40.3%, of the variance shared between the variable sets. As noted, the full model (Functions 1 to 3) is statistically significant. Functions 2 to 3 and 3 to 3 are also statistically significant, $F(8, 622)$ = 7.448, p < .001, and $F(3, 312)$ = 7.095, p < .001, respectively. Given the (R_c^2) effects for each function, only the first function could be considered noteworthy in the context of this study (28.29% of the shared variance). The last two functions only explain 11.04% and 6.39%, respectively, of the remaining variance in the variable sets after the extraction of the prior functions.

Table 4.3.1 presents the standardized canonical function coefficients and structure coefficients for Function 1. The squared structure coefficients are also given as well as the commonalities (h^2) across the three functions for each variable. Looking at the standardized coefficients for Function 1, one sees that the relevant criterion variables are primarily average wind speed, precipitation, and prevailing wind direction. This conclusion is supported by both the standardized and squared structure coefficients. Additionally, maximum relative humidity has a modest standardized coefficient with a small squared structure coefficient. Furthermore, all these variables' structure coefficients in Function 1 are positive, indicating they are positively related with respect to the first function. Regarding the predictor variable set in function 1, NO_2, and SO_2 are the primary contributors to the synthetic predictor variable set. Both NO_2, and SO_2's structure coefficients are negative, indicating that they are negatively associated with each of the weather variables.

TABLE 4.3.2. Canonical Solution for Weather and Economic Variables Predicting Pollutant Levels for Functions 1, 2, and 3 at Site 5

	Function 1			Function 2			Function 3			
Variable	*Coef*	r_s	$r_s^2(\%)$	*Coef*	r_s	$r_s^2(\%)$	*Coef*	r_s	$r_s^2(\%)$	$h^2(\%)$
AT	0.1227	0.0755	0.5703	0.5372	0.2016	4.0650	0.3182	0.1024	1.0476	5.6829
MRH	0.5068	0.0488	0.2378	-0.8599	-0.1595	2.5436	0.4657	0.0657	0.4316	3.2129
P	0.3636	0.2650	7.0222	-0.1357	-0.0698	0.4875	0.1582	0.1096	1.2019	8.7115
PWD	0.2543	0.2308	5.3256	-0.1091	-0.0094	0.0088	-0.8033	-0.2651	7.0289	12.3633
AWS	0.8235	0.4177	17.4445	0.1355	0.1352	1.8291	0.3267	-0.0111	0.0124	19.2860
MLC	-0.1554	-0.0920	0.8468	-0.0417	-0.0426	0.1815	0.4150	0.1307	1.7093	2.7376
M	-0.2574	0.0088	0.0078	-0.2032	0.1022	1.0436	-0.1902	-0.0407	0.1657	1.2171
TTU	0.0578	-0.0034	0.0012	-1.1266	-0.1357	1.8416	-0.0484	0.0123	0.0151	1.8579
O_3	-0.5496	-0.4081	16.6552	0.8603	0.2260	5.1076	-0.3078	-0.1277	1.6318	23.3946
NO_2	-0.3432	-0.3829	14.6616	-0.8701	-0.2411	5.8152	-0.6177	-0.1392	1.9365	22.4133
SO_2	-0.4935	-0.3831	14.6757	-0.0468	-0.0840	0.7053	0.9453	0.2390	5.7143	21.0952

A canonical correlation analysis is conducted using the eight *weather and economic variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., economic, weather and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .297, .167, and .123 for each successive function.

Collectively, the full model across all functions is statistically significant using Wilk's λ = .514 criterion, F(24, 890.995) = 9.584, p < .001. For the set of one canonical function, the R^2 type effect size is 0.486, which indicates that the full model explains a substantial portion, about 48.6%, of the variance shared between the variable sets.

As noted, the full model (Functions 1 to 3) is statistically significant. Functions 2 to 3 and 3 to 3 are also statistically significant, $F(14, 616) = 7.478$, $p < .001$, and $F(6, 309) = 7.247$, $p < .001$, respectively. Given the (R_c^2) effects for each function, only the first function could be considered noteworthy in the context of this study (29.7% of the shared variance). The last two functions only explain 16.7% and 12.3%, respectively, of the remaining variance in the variable sets after extracting the prior functions.

Table 4.3.2 presents the standardized canonical function coefficients and structure coefficients for Functions 1, 2, and 3. The squared structure coefficients are also given as well as the commonalities (h^2) across the three functions for each variable. Looking at the coefficients for Function 1 across all weather and economic variables, one sees that the relevant criterion variables are primarily average wind speed, precipitation, and prevailing wind direction. These results are supported by both the standardized coefficients and squared structure coefficients. Additionally, maximum relative humidity has a modest standardized coefficient with a small squared structure coefficient. This could be due to multicollinearity. Furthermore, all the weather variables' structure coefficients in Function 1 are positive, indicating they are positively related with respect to the first synthesis function.

Regarding the predictor variable set in function 1, O_3 and SO_2 are the primary contributors to the synthetic predictor variable, followed by NO_2. Because the structure coefficient (for Function 1) for each is negative, they are negatively related to each of the weather variables and trade, transportation, and utilities, and positively related to mining, logging, and construction as well as manufacturing. Because all three Functions present significant findings, we can consider h^2. Table 4.3.2 indicates average wind speed, prevailing wind direction, and precipitation are the most useful variables for the entire analysis.

4.3.2. Site 67

A canonical correlation analysis is conducted using the three *economic variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., economic and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .0739, .0278, and .0004 for each successive function.

Collectively, only the first dimension of the model across all functions is statistically significant using Wilk's λ = .900 criterion, F(9,744.874) = 3.66, p < .001. for the set of one canonical function, the R^2 type effect size is 0.1, which indicates that the full model explains a small portion, about 10%, of the variance shared between the variable sets.

As noted, the full model (Functions 1 to 3) is statistically significant. Functions 2 to 3 do not explain a statistically significant amount of shared variance between the variable sets, $F(4, 614)$ = 2.211, p = .0664. Additionally, Function 3 (which is the only function that is tested in isolation) does not explain a statistically significant amount of shared variance between the variable sets, $F(1, 308)$ = .129, p = .7197. Given the (R_c^2) effects for each function, none of the functions could be considered noteworthy in the context of this study (7.39%, 2.78%, and .04% of the shared variance, respectively).

TABLE 4.3.3. Canonical Solution for Weather Predicting Pollutant Levels for Functions 1, 2, and 3 at Site 67

	Function 1			Function 2			Function 3			
Variable	*Coef*	r_s	$r_s^2(\%)$	*Coef*	r_s	$r_s^2(\%)$	*Coef*	r_s	$r_s^2(\%)$	$h^2(\%)$
AT	0.1066	0.0887	0.7861	-0.9594	-0.4090	16.7310	-0.1214	-0.0681	0.4634	17.9806
MRH	0.0708	-0.1408	1.9815	0.3043	0.1243	1.5445	-0.7194	-0.1215	1.4768	5.0028
P	0.3814	0.4043	16.3453	0.1713	0.0894	0.7990	0.5359	0.0813	0.6612	17.8055
PWD	-0.2224	-0.0939	0.8826	0.1833	0.0108	0.0117	-0.5180	-0.1783	3.1780	4.0723
AWS	0.8016	0.5419	29.3644	0.1401	0.0039	0.0015	-0.6566	-0.0684	0.4672	29.8331
O_3	-0.5363	-0.4965	24.6503	-0.7041	-0.2051	4.2078	0.6780	0.0980	0.9611	29.8193
NO_2	-0.6459	-0.5338	28.4906	0.7754	0.1614	2.6060	-0.4841	-0.0883	0.7792	31.8758
SO_2	0.0208	-0.0995	0.9907	-0.6458	-0.2764	7.6405	-0.7859	-0.2129	4.5318	13.1630

A canonical correlation analysis is conducted using the five *weather variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., weather and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .371, .201, and .076 for each successive function. Collectively, the full model across all functions is statistically significant using Wilk's $\lambda = .464$ criterion, F(15, 839.612) = 17.928, $p < .001$. For the set of one canonical function, the r^2 type effect size is 0.535, which indicates that the full model explains a substantial portion, about 53.5%, of the variance shared between the variable sets. As noted, the full model (Functions 1 to 3) is statistically significant. Functions 2 to 3 and 3 to 3 are also statistically significant, $F(8, 610) = 12.503$, $p < .001$, and $F(3, 306) = 8.443$, $p < .001$, respectively. Given the (R_c^2) effects for each function, only the first two functions could be considered noteworthy in the context of this study (37.1% and 20.08% of the shared variance, respectively). The last function explains 7.65% of the remaining variance in the variable sets after extracting the prior functions.

Table 4.3.3 presents the standardized canonical function coefficients and structure coefficients for Functions 1 and 2. The squared structure coefficients are also given as well as the commonalities (h^2) across the three functions for each variable. Looking at the standardized coefficients for Function 1, one sees that the relevant criterion variables are primarily average wind speed and precipitation. This conclusion is supported by both the standardized and squared structure coefficients. Furthermore, both average wind speed and precipitation structure coefficients are positive, indicating positive associations with the first synthesis function.

Regarding the predictor variable set in Function 1, O_3 and NO_2 are the primary contributors to the synthetic predictor variable. Because the structure coefficient for each is negative, they are negatively related to the primary weather variables; average wind speed, and precipitation.

Looking at the standardized coefficients for Function 2, one sees that the relevant criterion variable is average temperature. This conclusion is supported by both the standardized and squared structure coefficients.

Regarding the predictor variable set in Function 2, all three pollutants are contributors to the synthetic predictor variable. Because the structure coefficients for each are negative for O_3 and SO_2 and are positive for NO_2, O_3, and SO_2 are positively correlated to the primary weather factor average temperature in function 2 while NO_2 is negatively correlated to average temperature.

TABLE 4.3.4. Canonical Solution for Weather and Economic Variables Predicting Pollutant Levels for Functions 1, 2, and 3 at Site 67

	Function 1			Function 2			Function 3			
Variable	*Coef*	r_s	$r_s^2(\%)$	*Coef*	r_s	$r_s^2(\%)$	*Coef*	r_s	$r_s^2(\%)$	$h^2(\%)$
AT	0.1001	0.0662	0.4387	0.9016	0.4181	17.4791	-0.1759	-0.0251	0.0628	17.9806
MRH	0.0765	-0.1379	1.9027	-0.2161	-0.1012	1.0234	0.5957	0.1441	2.0766	5.0028
P	0.3753	0.4104	16.8463	-0.1825	-0.0831	0.6904	-0.3345	-0.0518	0.2688	17.8055
PWD	-0.2318	-0.0985	0.9701	-0.1128	0.0246	0.0604	0.4780	0.1744	3.0419	4.0723
AWS	0.7824	0.5392	29.0707	0.0060	0.0414	0.1715	0.5500	0.0769	0.5908	29.8331
MLC	-0.2231	-0.0601	0.3608	0.3143	0.1261	1.5894	-0.0328	0.1572	2.4702	4.4205
M	-0.1689	-0.0639	0.4088	0.1746	0.1673	2.7995	0.2296	0.1601	2.5636	5.7719
TTU	0.2722	-0.0149	0.0223	-0.0707	0.0983	0.9665	0.3783	0.1960	3.8429	4.8317
O_3	-0.5504	-0.5081	25.8169	0.5022	0.1516	2.2990	-0.8294	-0.1571	2.4678	30.5837
NO_2	-0.6206	-0.5329	28.4005	-0.6803	-0.1545	2.3875	0.6362	0.1280	1.6373	32.4252
SO_2	-0.0343	-0.1331	1.7708	0.8067	0.3620	13.1069	0.6189	0.2057	4.2311	19.1089

A canonical correlation analysis is conducted using the eight *weather and economic variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., economic, weather, and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .378, .224, and .115 for each successive function.

Collectively, the full model across all functions is statistically significant using Wilk's λ = .427 criterion, F(24, 873.593) = 12.41, p < .001. For the set of one canonical function, the R^2 type effect size is 0.573, which indicates that the full model explains a substantial portion, about 57.3%, of the variance shared between the variable sets. As noted, the full model (Functions 1 to 3) is statistically significant. Functions 2 to 3 and

3 to 3 are also statistically significant, $F(14, 604) = 8.917$, $p < .001$, and $F(6, 303) = 6.557$, $p < .001$, respectively. Given the (R_c^2) effects for each function, only the first two functions could be considered noteworthy in the context of this study (37.8% and 22.4% of the shared variance, respectively). The last function only explains 11.5% of the remaining variance in the variable sets after extracting the prior functions.

Table 4.3.4 presents the standardized canonical function coefficients and structure coefficients for Functions 1, 2, and 3. The squared structure coefficients are also given as well as the commonalities (h^2) across the three functions for each variable. Looking at the coefficients for Function 1, one sees that the relevant criterion variables are primarily average wind speed and precipitation. These results are supported by both the standardized coefficient and squared structure coefficients. Furthermore, both variables' structure coefficients are positive indicating they are positively related to the first function. Economic variables except for trade, transportation, and utilities have negative standardized coefficients. However, all economic variables' structure coefficients are very small, indicating negligible association with the first function. Regarding the predictor variable set in Function 1, O_3 and NO_2 are the primary contributors to the synthetic predictor variable. Because the structure coefficient for each is negative, they are negatively related to the two primary weather variables.

For Function 2, the average temperature is the most relevant criterion variable, followed by mining, logging, and construction, though mining, logging, and construction has a relatively small squared structure coefficient, comparatively.

Regarding the predictor variable set in Function 2, SO_2 is the primary contributor to the synthetic predictor variable, followed by NO_2. Because the structure coefficient is positive and negative respectively, they are positively and inversely related to the primary weather variable average temperature.

4.3.3. Site 3002

A canonical correlation analysis is conducted using the *three economic variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., economic and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .0370, .0248, and .0053 for each successive function. Collectively, only the first and second dimensions of the model across all functions are statistically significant using Wilk's λ = .934 criterion, F(9,820.32) = 2.587, p = .0061. For the set of one canonical function, the R^2 type effect size is 0.065, which indicates that the full model explains a small portion, about 6.5%, of the variance shared between the variable sets. Function 2 to 3 explain a statistically significant amount of shared variance between the variable sets, $F(4, 676)$ = 2.689, p = .0357. As noted, the full model (Functions 1 to 3) is statistically significant, as well as Functions 2 to 3. Function 3 (which is the only function that is tested in isolation) does not explain a statistically significant amount of shared variance between the variable sets, $F(1, 339)$ = 1.81, p = .179. Given the (R_c^2) effects for each function, none of the functions could be considered noteworthy in the context of this study ($3.7\%, 2.48\%$, and $.53\%$ of the shared variance, respectively).

TABLE 4.3.5. Canonical Solution for Weather Predicting Pollutant Levels for Functions 1 and 2 at Site 3002

	Function 1			Function 2			
Variable	$Coef$	r_s	$r_s^2(\%)$	$Coef$	r_s	$r_s^2(\%)$	$h^2(\%)$
AT	-0.0943	-0.1222	1.4944	0.9346	0.2442	5.9627	7.4572
MRH	0.6966	0.3739	13.9787	-0.2269	-0.0927	0.8596	14.8383
P	0.3379	0.3153	9.9407	-0.0184	0.0005	0.0000	9.9407
PWD	-0.1373	-0.0870	0.7566	-0.0676	0.0042	0.0018	0.7584
AWS	0.6157	0.3047	9.2824	0.6531	0.1430	2.0450	11.3274
O_3	-0.9138	-0.5675	32.2075	0.4599	0.0513	0.2634	32.4709
NO_2	-0.1579	-0.2741	7.5139	-1.0636	-0.2991	8.9479	16.4617
SO_2	-0.0936	-0.1727	2.9824	0.1768	0.0035	0.0012	2.9836

A canonical correlation analysis is conducted using the *five weather variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., weather and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .334, .117, and .0115 for each successive function. Collectively, the full model across all functions is statistically significant using the Wilk's $\lambda = .581$ criterion, F(15, 925.19) = 13.407, $p < .001$. For the set of one canonical function, the R^2 type effect size is 0.419, which indicates that the full model explains a substantial portion, about 41.9%, of the variance shared between the variable sets. As noted, the full model (Functions 1 to 3) is statistically significant. Functions 2 to 3 are also statistically significant with $F(8, 672) = 12.503$, $p < .001$. Given the (R_c^2) effects for each function, only the first function could be considered noteworthy in the context of this study (33.4% of shared variance). The last two functions account for 11.7% and 1.15%, respectively, of the remaining variance in the variable sets after the extraction of the prior functions. Table 4.3.5 presents the standardized canonical function coefficients and structure coefficients for Functions 1 and 2. The squared structure coefficients are also given as well as the commonalities (h^2) across the three functions for each variable. Looking at the coefficients for Function 1, one sees that the relevant criterion variables are primarily maximum relative humidity, average wind speed, and precipitation. This conclusion is supported by both the standardized and squared structure coefficients. Furthermore, all these variables' structure coefficients are positive indicating they are positively related to the first function. Regarding the predictor variable set in Function 1, O_3 is the primary contributor to the synthetic predictor variable. Because the structure coefficient for O_3 is negative, it is negatively related to the primary weather variables.

A canonical correlation analysis is conducted using the eight *weather and economic variables* as predictors of the three pollutant variables to evaluate the multivariate shared relationship between the two variable sets (i.e., economic, weather

TABLE 4.3.6. Canonical Solution for Weather and Economic Variables Predicting Pollutant Levels for Functions 1, 2 and 3 at Site 3002

	Function 1			Function 2			
Variable	$Coef$	r_s	$r_s^2(\%)$	$Coef$	r_s	$r_s^2(\%)$	$h^2(\%)$
AT	-0.1756	-0.1167	1.3609	0.8490	0.2461	6.0576	7.4184
MRH	0.8066	0.3713	13.7897	-0.1882	-0.0999	0.9974	14.7871
P	0.2776	0.3167	10.0282	-0.0627	-0.0078	0.0061	10.0343
PWD	-0.1091	-0.0833	0.6939	-0.0992	0.0019	0.0003	0.6943
AWS	0.6047	0.3076	9.4595	0.5667	0.1366	1.8673	11.3268
MLC	-0.7883	0.0702	0.4930	-0.6217	-0.0865	0.7483	1.2414
M	-0.3556	0.0828	0.6856	0.8919	-0.0029	0.0008	0.6864
TTU	0.8720	0.1045	1.0929	-0.2294	-0.0295	0.0870	1.1799
O_3	-0.9140	-0.5984	35.8067	0.4905	0.0634	0.4023	36.2091
NO_2	-0.1822	-0.2977	8.8636	-1.0574	-0.3155	9.9535	18.8171
SO_2	-0.0512	-0.1597	2.5510	0.1333	-0.0095	0.0091	2.5600

and pollutant levels). The analysis yields three functions with squared canonical correlations (R_c^2) of .371, .132, and .023 for each successive function. Collectively, the full model across all functions is statistically significant using the Wilk's λ = .533 criterion, F(24, 963.502) = 9.717, $p < .001$. For the set of one canonical function, the R^2 type effect size is 0.467, which indicates that the full model explains a substantial portion, about 46.7%, of the variance shared between the variable sets. As noted, the full model (Functions 1 to 3) is statistically significant. Functions 2 to 3 and 3 to 3 are also statistically significant, $F(14, 666) = 4.080$, $p < .001$, and $F(6, 334) = 1.289$, $p < .001$, respectively. Given the (R_c^2) effects for each function, only the first function could be considered noteworthy in the context of this study (37.1% of the shared variance). The last two functions only explain 13.2% and 2.3%, respectively, of the remaining variance in the variable sets after extracting the prior functions. Table 4.3.6 presents the standardized canonical function coefficients and structure coefficients for Functions 1 and 2. The squared structure coefficients are also given as well as the commonalities (h^2) across the three functions for each variable. Looking at the coefficients for Function 1, one sees that the relevant criterion variables are primarily maximum relative humidity, precipitation, and average wind speed with relatively higher standardized and structure coefficients. Additionally, mining, logging, and construction has a high standardized coefficient with a small squared structure coefficient, indicating negligible association with the first synthesis function. Furthermore, all three weather variables' structure coefficients are positive, indicating they are positively related to the first function. Regarding the predictor variable set in Function 1, O_3 is the primary contributor to the synthetic predictor variable. Because the structure coefficient of O_3 is negative, it is negatively correlated to the three primary weather variables.

4.4. Discussion

This model indicates that the most correlated factors among weather variables at Site 5 are primarily average wind speed, precipitation, and prevailing wind direction. Each factor has a high standardized coefficient with a sizeable structure coefficient, indicating that not only are the variables important to the model, but they also influence the model significantly. This is also the case amongst both economic and weather variables. However, maximum relative humidity has a high standardized coefficient with a very small structure coefficient. This could be due to multicollinearity amongst the data. Among only economic variables, only one dimension of the model is statistically significant and does not explain enough of the variation within the model to consider formal interpretation.

These models indicate that at Site 67, the most correlated factors among weather variables are primarily average wind speed and precipitation. Prevailing wind direction features a moderate standardized coefficient with a low structure coefficient, indicating a small influence on the model itself. This could be due to multicollinearity. Among just weather variables, average wind speed and precipitation are inversely correlated with both ozone and nitrogen dioxide, while positively correlated with sulfur dioxide. Among both economic and weather variables, average wind speed and precipitation are the most correlated factors, as they both feature large standardized coefficients with large structure coefficients. In the second Function, the average temperature has a high standardized coefficient as well as a large structure coefficient. Among both weather variables and economic variables, precipitation and average wind speed are all inversely correlated with all pollutant levels.

The models for Site 3002 indicate that the most correlated variables among just weather variables and among both economic and weather variables are maximum relative humidity, average wind speed, and precipitation. Each feature has similar structure coefficients and high standardized coefficients. Trade, transportation, and utilities as well as mining, logging, and construction feature high standardized coefficients with small structure coefficients, indicating they have much less influence on the model than the most correlated variables.

These results are similar to results found in the extant literature, which primarily indicate average wind speed to be a top contributor to pollutant levels. For more information regarding this, see the literature section of this chapter. Therefore, we can conclude that using CCA on this data set is an effective methodology for studying air pollution and meteorological factors. Additionally, there is some association with economic factors, which is logically to be expected, though this is not fully indicated within other analyses of the same data (DGRA and GRA). This is further discussed in Chapter 5.

4.5. Conclusion

In conclusion, Canonical Correlation Analysis serves as a valuable tool for exploring relationships between sets of variables, providing insights into air quality dynamics across different sites. However, it is essential to acknowledge the limitations of CCA, including its assumption of linear relationships, sensitivity to outliers and multicollinearity, and challenges in interpretation, particularly with low canonical correlations. Despite these constraints, our analysis reveals significant findings regarding the correlated factors affecting air quality at various sites. At Site 5, average wind speed, precipitation, and prevailing wind direction emerge as primary influencers, aligning with existing literature. Similarly, at Site 67, these weather variables, along with average temperature and maximum relative humidity, play pivotal roles. Site 3002 exhibits similar patterns, with maximum relative humidity, average wind speed, and precipitation being associated with air quality significantly. The incorporation of economic factors further enriched our understanding, highlighting the complex interplay between meteorological and socioeconomic variables. Although CCA offers valuable insights, its limitations necessitate careful consideration and complementary methods. Nonetheless, our findings underscore the effectiveness of CCA in studying air pollution and meteorological factors, providing a robust framework for future research in environmental management and public health.

4.6. Limitations

While Canonical Correlation Analysis offers valuable insights into the relationships between sets of variables and has been widely used in various fields, it does have several limitations. Firstly, CCA assumes linear relationships between variables, which may not always accurately capture the complex non-linear interactions present in real-world data. Additionally, CCA requires a relatively large sample size to produce reliable results, and small sample sizes may lead to unstable estimates. Moreover, CCA is sensitive to outliers, correlated observations, and multicollinearity between variables, which can distort results and reduce the robustness of the analysis.

Furthermore, CCA does not provide information on causality or the directionality of relationships between variables, limiting its ability to infer causal mechanisms. Lastly, interpretation of CCA results can be challenging, particularly when dealing with a large number of variables or when correlations between canonical variates are low. This study features some low canonical correlations. Overall, while CCA is a powerful multivariate analysis technique, researchers should be cautious of its limitations and consider complementary methods to enhance the interpretability of their findings.

CHAPTER 5

Comparison and Discussion

5.1. Discussion and Limitations

In this comprehensive discussion section, we discuss the similarities and differences between three prominent multivariate analysis techniques: Dynamic Grey Relational Analysis, Canonical Correlation Analysis, and Grey Relational Analysis. These methodologies offer distinct approaches to understanding the complex relationships within datasets, each with its strengths and limitations. We also explore how DGRA and GRA share similarities in their utilization of grey relational coefficients to measure the similarity between variables, yet differ in their treatment of the distinguishing coefficient, with DGRA dynamically adjusting this parameter based on the data, while GRA maintains a static value. Conversely, CCA stands apart as a functional model that explicitly models linear relationships between sets of variables. Furthermore, we address the nuances of normalization and standardization methods employed in our study, which play a crucial role in preprocessing the data for analysis. Additionally, the limitations inherent in our study are examined, including assumptions made in the analysis, potential biases, and constraints in data availability and quality. By elucidating these methodological considerations and limitations, we aim to provide a comprehensive understanding of the analytical approaches employed and the implications of our findings. Dynamic Grey Relational Analysis and Grey Relational Analysis share similarities in their fundamental approach to analyzing relationships within datasets. Both methods utilize grey relational coefficients to measure the similarity between variables, allowing for the identification of correlated factors. However, they differ in their implementation: DGRA incorporates a dynamic calculation of the distinguishing coefficient, adapting it based on the data, while GRA employs a static value of 0.5 for this coefficient. This dynamic aspect of DGRA enables it to better capture the variability within the dataset, potentially leading to more accurate results. Moreover, both DGRA and GRA are non-functional models, meaning they do not explicitly model the functional relationship between variables but rather assess their relational patterns. In contrast, Canonical Correlation Analysis is a functional model that explicitly models the linear relationships between sets of variables. Unlike DGRA and GRA, CCA does not require preprocessing other than addressing missing data, offering a more straightforward approach to analyzing multivariate relationships. Additionally, it is important to mention the difference in the interpretative nature of both CCA and DGRA. CCA gives a researcher much more information into the dynamic workings of a system (while considering all dependent variables), while DGRA ranks factors using a grade between zero and one. Overall, while DGRA and GRA provide valuable insights into the relational patterns within datasets, CCA offers a more structured and comprehensive framework for understanding the functional relationships between variables.

Grey relational analysis methodology reveals inconsistencies in relational ranking across standardization methods, suggesting the method is sensitive to pre-processing. Results indicate the most correlated variables for NO_2 are average temperature, precipitation, average wind speed, and maximum relative humidity. Specifically at Site 5, mining, logging, and construction as well as trade, transportation, and utilities are among the most correlated. However, overwhelmingly at Sites 67 and 3002, weather variables dominate. Conversely, for SO_2, economic variables are consistently highly ranked among each site, with precipitation being the only other variable mentioned in the top half of most correlated variables. For O_3, results are somewhat similar to NO_2 in that average temperature consistently ranks first across all sites and standardization methods. Consistencies in every other grey relational order are practically non-existent at each site and standardization method. Generally, economic variables rank low across all sites for both ozone and nitrogen dioxide, but results are overall not consistent.

Dynamic grey relational analysis methodology reveals consistencies in relational ranking across standardization methods, suggesting it is not as sensitive to these processes as grey relational methodology is. That said, for all three pollutants, results indicate the most correlated factors of all pollutants to be weather variables, primarily average wind speed (Sites 5 and 3002 across all pollutants), average temperature (Site 67 across all pollutants), followed widely by maximum relative humidity and prevailing wind direction. Economic variables rank low, with very small grey relational grades, indicating their association with pollutant levels to be small. Canonical Correlation Analysis methodology reveals multiple correlated factors that differ across each site. At Site 5, the most correlated factors among both economic and weather variables are average wind speed, precipitation, and wind direction. At Site 67, the most correlated factors among both economic and weather variables are average wind speed and precipitation followed by average temperature and mining, logging, and construction. Thirdly, at Site 3002, the most correlated variables among both weather and economic variables are trade, transportation and utilities, maximum relative humidity, and mining, logging, and construction.

DGRA and GRA methodologies present wildly different results. Not only do rankings amongst variables nearly flip between the two methods, but grades also widely differ. Additionally, standardization methods used in the pre-processing of data provide differences in ranking amongst both methods, though more widely in GRA results. Firstly, the comparison of DGRA and GRA results gives insight into the possible limitations of GRA. Not only are rankings inconsistent, but GRA results show economic variables as highly ranked in comparison to weather variables. DGRA has the opposite result. Comparing both of these to CCA results, DGRA has much more similar results to CCA than GRA has, which nearly completely discredits GRA methodology in and of itself because CCA is a widely used and trusted multivariate statistical method. This is additionally supported by the low grey relational grades resulting from DGRA results for economic variables as well as low standardized coefficients in CCA. While rankings exhibit discrepancies, it is challenging to interpret the results of Canonical Correlation Analysis in terms of a definitive "ranking." However, weather variables are still generally considered more correlated than economic variables. This is not the outcome of using Grey Relational Analysis. There are several reasons this could be. Firstly, a drawback of the GRA methodology is the necessity for preprocessing of data, and the nature of the data that is used. The Grey Relational Analysis (GRA) methodology presents a drawback due to the necessity for data preprocessing, as highlighted in the literature. When discussing the nature of GRA and pollutant type, J.T. Huang & Liao (2003, pg. 1711) say "When the range of the sequence is too large or the standard value is too enormous, it will cause the influence of some factors to be neglected. In addition, in the sequence, if the factors' goals and directions are different, the relational analysis might also produce incorrect results. Therefore, preprocessing of all the data is necessary." This could explain the lack of consistency in rankings resulting from GRA. However, there is a bit of nuance in the language used by Huang and Liao. For example, what exactly constitutes "too large" or "too enormous" in terms of the standard value, and what is meant by the "goals and directions" of the factors? This conclusion is not supported by mathematical justification, which leads to the question of to what extent this affects the reliability of the method. Though there have been improvements made to GRA methodology (through DGRA and separate approaches to data [see (Javed et al., 2022; Dai, Liu, & Hu, 2014; Yamaguchi, Li, & Nagai, n.d.))], the method itself is inherently flawed for studies of this type, where multiple types of data are investigated. That said, the literature also indicates that this is why we use normalization and standardization techniques to remedy this issue. However, there is no convention of which normalization processes to use, as discussed previously. In industry settings, standardization or normalization methods are often straightforwardly chosen based on established protocols or best practices due to well-defined data standards and uniform distributions. Studies such as (J. T. Huang & Liao, 2003; Kuo, Yang, & Huang, 2008; Muqeem, Sherwani, Ahmad, & Khan, 2017) utilized normalization methods like "higher the better" or "lower the better" in their Grey Relational Analysis research involving industry-related data. However, in environmental data analysis, the literature reveals a more varied approach. For instance, Shexia et al. (2018) applied Initial Value Processing to water quality data, while Pan et al., (2011) and Tao et al. (2015) didn't specify normalization methods for air quality index (AQI) and economic data in their GRA studies. Studies like Zeydan et al. (2021) predominantly used "higher the better" and "lower the better" techniques for raw pollution data. Similarly, Huang (2020) and Ng (1994) employed IVP and MVP techniques in economic data analysis. Therefore, the selection of standardization or normalization techniques in environmental and economic studies requires careful consideration of data characteristics, analytical assumptions, and research objectives. The choice of methods in environmental data analysis appears to be influenced by the unique complexities inherent in the data itself.

Conclusively, the preprocessing of data is an issue that is highlighted in both DGRA and GRA results. Results for GRA indicate widely different rankings, with very few sites and pollutants providing similar rankings across pre-processing methods. Though this is seen in a few places in DGRA results, orderings are much more consistent, and grey relational grades calculated are also much more consistent. This finding is corroborated by Javed et al., (2022), which indicates that within the field of multi-criterion decision-making, it is "not unusual to obtain different rankings using different normalization techniques" (pg. 6). Additionally, Javed et al. (2022) briefly discussed when presenting the methodology of DGRA, and proposed the Ideal

Alternative Function to solve this problem. However, this is still a normalization technique that relies on extracting an ideal alternative from a decision matrix (Javed et al., 2022). This is not utilized in this study because normalization is comprised of literature-based decisions, especially since the behavior of pollutants is taken into consideration when the pre-processing of data occurs. Additionally, it is important to mention the difference in the interpretative nature of both CCA and DGRA. CCA gives a researcher much more information into the dynamic workings of a system (while considering all dependent variables), while DGRA ranks factors using a grade between zero and one. Thus, when considering these results, it is imperative to consider the bigger picture when looking for consistencies and inconsistencies between the results of each methodology. However, multiple regression could be used to further discuss the results of GRA and DGRA. Multiple Linear Regression (MLR) is used to determine the relationship between a quantitative dependent variable and two or more independent variables as a way to validate GRA and DGRA results. Site 67 is included within this discussion to stay consistent with results in other chapters. Table 6.4.1 and 6.4.2 include a comparison of MLR results to GRA and DGRA results. Each model is significant with p-value $< .001$, and the models chose, on average, 5 predictors. The predictors chosen along with the estimated standardized coefficients (ESC) align with DGRA results more so than GRA results, though (with MLR, based on the ESC) the rankings themselves are not the same. These results are presented Tables 6.4.1 and 6.4.2 in the Appendix. A similar pattern is observed for other sites.

Though we know through literature and logic that economic variables may impact air pollution data, the results in this study do not indicate such a conclusion. Further, CCA provides a result that is corroborated by literature (See Chapter 4 literature review and discussion). Given the robust usefulness of the method and multivariate statistical basis, the method proved to be reliable within the context of this study. Looking at other methodologies' results, Canonical Correlation Analysis leads us to a different conclusion than both grey relational analysis and dynamic grey relational analysis, though the conclusion is more similar to DGRA results. This is similar to the result of using MLR, as discussed previously. In any case, economic variables are low-ranks in the DGRA results and do not account for many variations- in both grey relational grades and the CCA model. Similarly, multiple regression chose similar factors to DGRA, rather than GRA. Thus, there is reason to recommend the use of DGRA over GRA because of its similarity in results. However, caution is still recommended because of the sensitive nature of DGRA to preprocessing methods and a lack of formal methodology to accompany the choice of preprocessing techniques. Extraneous factors to consider include the fact that though the sites are located in suburban locations, the locations themselves are in a rural state where these economic variables may not be as prolific as if the data were taken from a larger, active city with both a larger workforce and a higher population.

5.2. Future work

While this study provides valuable insights into the interplay between air pollution, economic variables, and weather patterns using GRA and DGRA, several limitations and avenues for future research warrant consideration. Firstly, the study would benefit from incorporating a broader range of pollutant data, economic factors, and weather variables to comprehensively understand their interactions. An example of this may be investigating more criteria pollutants such as $PM_{2.5}$, Pb, and CO as well as expanding types of factors considered. To further the economic portion of this study, factors that could be included are Propane, Electricity Net generation, geothermal energy, and CO_2 emissions across the nation. Additionally, integrating a public health aspect into the study could offer valuable insights into the health implications of air pollution and inform targeted interventions. An example of this may be asthma diagnostics in vulnerable populations. A downfall of this, however, is the availability of data. Moreover, future work should clarify the extent of the limitations extending into method validity. Namely, what constitutes "too large" or "too enormous" (J. T. Huang & Liao, 2003) within a data set, and exactly how normalization and standardization affect grey relational orderings, especially considering the methods' use in sensitive topics, such as healthcare (Ouali, 2022; Javed et al., 2022). Further research is also needed to explore the suitability of different normalization and standardization methods for various types of data; what is best used? How do we corroborate results if differing methods of normalization present converse orderings? Assessing alternative methods alongside GRA and DGRA could also shed light on the most suitable approach for analyzing complex relationships among environmental, economic, and health variables. Overall, addressing these limitations and pursuing future research avenues will enhance the effectiveness and applicability of GRA and DGRA in analyzing the multifaceted dynamics of air pollution and its impacts on society.

5.3. Conclusion

In conclusion, this thesis has delved into the intricate relationships among air pollution, economic factors, weather variables, and their impact on civilization through the lens of Grey Relational Analysis, Dynamic Grey Relational Analysis, and Canonical Correlation Analysis. By employing these advanced analytical techniques, we have gained valuable insights into the complex interplay between these factors, shedding light on their nuanced relationships and implications for societal well-being. The Grey Relational Analysis methodology uncovered inconsistencies in relational rankings across standardization methods, indicating sensitivity to preprocessing. Notably, correlated variables for nitrogen dioxide included average temperature, precipitation, and economic activities at specific sites. Conversely, for sulfur dioxide, economic variables consistently rank high, with precipitation being the only other significant factor. Ozone results mirrored NO_2 to some extent, with average temperature consistently correlated but other rankings varied greatly. On the other hand, Dynamic Grey Relational Analysis (DGRA) shows consistent relational rankings across standardization methods, suggesting lesser sensitivity to preprocessing. Weather variables, particularly average wind speed, and temperature emerge as the most correlated factors for all pollutants, while economic variables have minimal impact. Lastly, Canonical Correlation Analysis reveals site-specific correlated factors, with weather variables like wind speed and precipitation consistently prominent across sites, alongside varying economic factors. While GRA demonstrated sensitivity to preprocessing, DGRA provides more stable rankings, and CCA highlights site-specific correlated factors. However, it is essential to acknowledge the limitations of this study, including the need for more comprehensive pollutant, economic, and weather data, as well as the necessity to integrate a public health perspective. Future research endeavors should also address the challenges posed by specific conditions required by GRA, explore alternative normalization and standardization methods, and assess the effectiveness of various analytical approaches. Moreover, based on our findings, we recommend the adoption of Dynamic Grey Relational Analysis over GRA due to its flexibility and ability to accommodate diverse datasets without extensive preprocessing, and the closeness of results to CCA, given the multivariate nature of CCA and the unique perspective CCA gives. Overall, this study contributes to our understanding of the complex dynamics shaping air quality, economic development, and societal well-being, paving the way for informed decision-making and proactive interventions toward a healthy future.

References

Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005, October). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software*, *20*(10), 1263–1271. doi: 10.1016/j.envsoft.2004.09.001

Ataei, Y., Mahmoudi, A., Feylizadeh, M. R., & Li, D.-F. (2020). Ordinal priority approach (opa) in multiple attribute decision-making. *Applied Soft Computing*, *86*, 105893. doi: 10.1016/j.asoc.2019.105893

Bai, Y.-t., Jin, X.-b., Wang, X.-y., Wang, X.-k., & Xu, J.-p. (2020, January). Dynamic Correlation Analysis Method of Air Pollutants in Spatio-Temporal Analysis. *IJERPH*, *17*(1), 360. doi: 10.3390/ijerph17010360

Binaku, K., & Schmeling, M. (2017, December). Multivariate statistical analyses of air pollutants and meteorology in Chicago during summers 2010-2012. *Air Qual Atmos Health*, *10*(10), 1227–1236. doi: 10.1007/s11869-017-0507-7

Borowiak, K., Zbierska, J., & Jusik, S. (2011). CANONICAL CORRESPONDENCE ANALYSIS (CCA) AS A TOOL FOR INTERPRETATION OF BIOINDICATOR PLANT RESPONSE TO AMBIENT AIR POLLUTION. *Fresenius Environmental Bulletin*, *20*(9).

Bureau of Economic Accounts (BEA). (n.d.). *Regional Economic Accounts: Download.* https://apps.bea.gov/regional/downloadzip.cfm.

Canonical Correlation Analysis in Detail. (2018, July).

Dai, J., Liu, X., & Hu, F. (2014). Research and Application for Grey Relational Analysis in Multigranularity Based on Normality Grey Number. *The Scientific World Journal*, *2014*, 1–10. doi: 10.1155/2014/312645

Davis, M. E., Laden, F., Hart, J. E., Garshick, E., & Smith, T. J. (2010, May). Economic Activity and Trends in Ambient Air Pollution. *Environ Health Perspect*, *118*(5), 614–619. doi: 10.1289/ehp.0901145

Environmental Protection Agency (EPA). (2019, June). *Sulfur Dioxide (SO2) Pollution.* https://www.epa.gov/so2-pollution.

Environmental Protection Agency (EPA). (2021a). *Ground-level ozone pollution.* https://www.epa.gov/ground-level-ozone-pollution.

Environmental Protection Agency (EPA). (2021b). *Meteorology and air quality.* https://www.epa.gov/air-research/meteorology-and-air-quality.

Gerus-Gościewska, M., & Gościewski, D. (2022, January). Grey Relational Analysis (GRA) as an Effective Method of Research into Social Preferences in Urban Space Planning. *Land*, *11*(1), 102. doi: 10.3390/land11010102

Grigorieva, E., & Lukyanets, A. (2021, June). Combined Effect of Hot Weather and Outdoor Air Pollution on Respiratory Health: Literature Review. *Atmosphere*, *12*(6), 790. doi: 10.3390/atmos12060790

Habeebullah, T. M., Munir, S., Awad, E. A., Abdel Hameed A. A. Morsy, Seroji, A. R., & Mohammed, A. M. F. (2015). The Interaction between Air Quality and Meteorological Factors in an Arid Environment of Makkah, Saudi Arabia. *IJESD*, *6*(8), 576–580. doi: 10.7763/IJESD.2015.V6.660

Huang, C.-Y., Hsu, C.-C., Chiou, M.-L., & Chen, C.-I. (2020, October). The main factors affecting Taiwan's economic growth rate via dynamic grey relational analysis. *PLoS ONE*, *15*(10), e0240065. doi: 10.1371/journal.pone.0240065

Huang, J. T., & Liao, Y. S. (2003, January). Optimization of machining parameters of Wire-EDM based on Grey relational and statistical analyses. *International Journal of Production Research*, *41*(8), 1707–1720. doi: 10.1080/1352816031000074973

Irfan, M., Ikram, M., Ahmad, M., Wu, H., & Hao, Y. (2021, November). Does temperature matter for COVID-19 transmissibility? Evidence across Pakistani provinces. *Environ Sci Pollut Res*, *28*(42), 59705–59719. doi: 10.1007/s11356-021-14875-6

Javed, S. A., Gunasekaran, A., & Mahmoudi, A. (2022, November). DGRA: Multi-sourcing and supplier classification through Dynamic Grey Relational Analysis method. *Computers & Industrial Engineering*, *173*, 108674. doi: 10.1016/j.cie.2022.108674

Julong, D. (1988). *Introduction to Grey System Theory.*

Kalisa, E., Fadlallah, S., Amani, M., Nahayo, L., & Habiyaremye, G. (2018, November). Temperature and air pollution relationship during heatwaves in Birmingham, UK. *Sustainable Cities and Society*, *43*, 111–120. doi: 10.1016/j.scs.2018.08.033

Kuebler, J., Van Den Bergh, H., & Russell, A. G. (2001, January). Long-term trends of primary and secondary pollutant concentrations in Switzerland and their response to emission controls and economic changes. *Atmospheric Environment*, *35*(8), 1351–1363. doi: 10.1016/S1352-2310(00)00401-5

Kuo, Y., Yang, T., & Huang, G.-W. (2008, August). The use of grey relational analysis in solving multiple attribute decision-making problems. *Computers & Industrial Engineering*, *55*(1), 80–93. doi: 10.1016/j.cie.2007.12.002

Laessig, R. E., & Duckett, E. J. (1979, April). Canonical correlation analysis: Potential for environmental health planning. *Am J Public Health*, *69*(4), 353–359. doi: 10.2105/AJPH.69.4.353

Lesson 13: Canonical Correlation Analysis. (n.d.).

Li, X., Zheng, W., Yin, L., Yin, Z., Song, L., & Tian, X. (2017, August). Influence of Social-economic Activities on Air Pollutants in Beijing, China. *Open Geosciences*, *9*(1). doi: 10.1515/geo-2017-0026

Li, H, & Wang, X. (2018). Analysis of Air Quality and Influencing Factors in Beijing-Tianjin-Hebei Region From 2013 to 2017. In *Adm. Tech. Environ. Monit.* (2nd ed., pp. 21–25.).

Liu, B., Zhao, Q., Jin, Y., Shen, J., & Li, C. (2021, February). Application of combined model of stepwise regression analysis and artificial neural network in data calibration of miniature air quality detector. *Sci Rep*, *11*(1), 3247. doi: 10.1038/s41598-021-82871-4

Mahmood, R., Schargorodski, M., Foster, S., & Quillingan, A. (2019). A technical overview of the kentucky mesonet. *Journal of Atmospheric and Oceanic Technology*, *36*(9). doi: 10.1175/JTECH-D-18-0198.1

Mcadams, H. T., Crawford, R. W., & Hadder, G. R. (2000, June). A Vector Approach to Regression Analysis and Its Application to Heavy-Duty Diesel Emissions. In *CEC/SAE Spring Fuels & Lubricants Meeting & Exposition* (p. 2000-01-1961). doi: 10.4271/2000-01-1961

Muqeem, M., Sherwani, A. F., Ahmad, M., & Khan, Z. A. (2017, November). Taguchi-based combined grey relational and principal component analyses for multi-response optimization of diesel engines. *GS*, *7*(3), 408–425. doi: 10.1108/GS-05-2017-0013

NASA. (2022). *Ozone Pollution.* https://climate.nasa.gov/health/pollution/ozone-pollution/. Retrieved from `https://airquality.gsfc.nasa.gov/surface-level-ozone`

Ng, D. K. W. (1994). Grey System and Grey Relational Model. *ACM SIGICE Bulletin*, *20*(2).

Ouali, M. (2022, December). Evaluation of Chinese Cloth Suppliers using Dynamic Grey Relational Analysis. *IJGS*, *2*(2), 34–46. doi: 10.52812/ijgs.62

Palczewski, K., & Sałabun, W. (2019). Influence of various normalization methods in PROMETHEE II: An empirical study on the selection of the airport location. *Procedia Computer Science*, *159*, 2051–2060. doi: 10.1016/j.procs.2019.09.378

Pan, L., Sun, B., & Wang, W. (2011). City Air Quality Forecasting and Impact Factors Analysis Based on Grey Model. *Procedia Engineering*, *12*, 74–79. doi:

10.1016/j.proeng.2011.05.013

Samuels, V. (n.d.). *CANONICAL CORRELATION and PRINCIPAL COMPONENT ANALYSES* (Unpublished doctoral dissertation). The University of Texas at El Paso.

Sherry, A., & Henson, R. K. (2005, February). Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer. *Journal of Personality Assessment*, *84*(1), 37–48. doi: 10.1207/s15327752jpa8401$_0$9

Shexia, Z., Yaoqui, K., & Zhu, R. (2018). Application of grey relational analysis in water quality evaluation. *The Journal of Grey System*, *58*(8). doi: 10.16511/j.cnki.qhdxxb.2018.26.031

Shihab, A. (2022, December). Investigating the Relationship between AirPollutants and Meteorology: A CanonicalCorrelation Analysis. *Pol. J. Environ. Stud.*, *31*(6), 5841–5849. doi: 10.15244/pjoes/151908

Song, J., Guang, W., Li, L., & Xiang, R. (2016, April). Assessment of Air Quality Status in Wuhan, China. *Atmosphere*, *7*(4), 56. doi: 10.3390/atmos7040056

Song, Z., Deng, Q., & Ren, Z. (2020, January). Correlation and principal component regression analysis for studying air quality and meteorological elements in Wuhan, China. *Environ Prog Sustainable Energy*, *39*(1), 13278. doi: 10.1002/ep.13278

Statheropoulos, M., Vassiliadis, N., & Pappa, A. (1998, March). Principal component and canonical correlation analysis for examining air pollution and meteorological data. *Atmospheric Environment*, *32*(6), 1087–1095. doi: 10.1016/S1352-2310(97)00377-4

Tao, A. (2015). Grey relational theory analysis of factors affecting air quality in Shanghai City. In *Proceedings of the 2015 6th International Conference on Manufacturing Science and Engineering.* Qingdao, China: Atlantis Press. doi: 10.2991/icmse-15.2015.107

US EPA. (2016, September). *Basic Information about NO2.* https://www.epa.gov/no2-pollution/basic-information-about-no2.

US EPA, OAR. (2016, August). *Monitor Values Report* [Data and Tools]. https://www.epa.gov/outdoor-air-quality-data/monitor-values-report.

Wang, W., Chen, Y., Xie, Y., & Wu, L. (2021, November). Research on Influence Factors of Air Quality in the Host Area of the 24th Winter Olympic Games Based on Grey Relational Analysis. *Front. Environ. Sci.*, *9*, 757533. doi: 10.3389/fenvs.2021.757533

Wiecek-Janka, E., Mierzwiak, R., Nowak, M., Kujawinska, A., & Majchrzak, J. (2021, March). Application of Grey Systems Theory in the Analysis of Data Obtained from Family Businesses. *ERSJ*, *XXIV*(Special Issue 1), 494–510. doi: 10.35808/ersj/2057

Xu, R., Rahmandad, H., Gupta, M., DiGennaro, C., Ghaffarzadegan, N., Amini, H., & Jalali, M. S. (2021, October). Weather, air pollution, and SARS-CoV-2 transmission: A global analysis. *The Lancet Planetary Health*, *5*(10), e671-e680. doi: 10.1016/S2542-5196(21)00202-3

Yamaguchi, D., Li, G.-D., & Nagai, M. (n.d.). Evaluation on the Effectiveness of Grey Relational Analysis Models.

Zeydan, Ö., & Pekkaya, M. (2021, May). Evaluating air quality monitoring stations in Turkey by using multi criteria decision making. *Atmospheric Pollution Research*, *12*(5), 101046. doi: 10.1016/j.apr.2021.03.009

A: GRA Results

Table 5.3.1. SO_2 GRA Complete Results

SO_2 GRA RESULTS								
Site 5			Site 67			Site 3002		
GRG (γ) with IVP		RANK	GRG (γ) with IVP		RANK	GRG (γ) with IVP		RANK
TTU	0.8926	1	P (in)	0.8836	1	M	0.9392	1
MLC	0.8836	2	M	0.8787	2	TTU	0.9338	2
P (in)	0.8594	3	TTU	0.8665	3	P (in)	0.9290	3
M	0.8023	4	MLC	0.8572	4	MLC	0.9197	4
AWS (mph)	0.7227	5	AWS (mph)	0.8020	5	AWS (mph)	0.7573	5
PWD	0.7189	6	PWD	0.7912	6	PWD	0.6357	6
AT (F)	0.5721	7	AT (F)	0.7388	7	AT (F)	0.5813	7
MRH (%)	0.4625	8	MRH (%)	0.6207	8	MRH (%)	0.5017	8
GRG (γ) with MVP		RANK	GRG (γ) with MVP		RANK	GRG (γ) with MVP		RANK
SO2	0.0000		SO2	1.0000		SO2	1.0000	
TTU	0.8985	1	P (in)	0.8836	1	M	0.9407	1
M	0.8976	2	TTU	0.8814	2	TTU	0.9393	2
MLC	0.8916	3	M	0.8813	3	MLC	0.9353	3
P (in)	0.8844	4	MLC	0.8809	4	P (in)	0.9059	4
AWS (mph)	0.7539	5	AWS (mph)	0.8042	5	AWS (mph)	0.7196	5
PWD	0.7136	6	PWD	0.7490	6	PWD	0.6446	6
AT (F)	0.6316	7	AT (F)	0.7388	7	AT (F)	0.4922	7
MRH (%)	0.5266	8	MRH (%)	0.6207	8	MRH (%)	0.4110	8

Table 5.3.2. O_3 GRA Complete Results

O_3 GRA RESULTS								
Site 5			Site 67			Site 3002		
GRG (γ) with IVP		RANK	GRG (γ) with IVP		RANK	GRG (γ) with IVP		RANK
AT (F)	0.7884	1	AT (F)	0.8753	1	AT (F)	0.8374	1
AWS (mph)	0.7480	2	PWD	0.7977	2	MRH (%)	0.7932	2
PWD	0.7407	3	MRH (%)	0.7968	3	P (in)	0.7190	3
MRH (%)	0.7396	4	AWS (mph)	0.7952	4	M	0.7053	4
MLC	0.7088	5	M	0.7899	5	TTU	0.7031	5
TTU	0.7038	6	P (in)	0.7857	6	MLC	0.6975	6
M	0.6450	7	TTU	0.7826	7	AWS (mph)	0.6599	7
P (in)	0.6329	8	MLC	0.7769	8	PWD	0.6198	8
GRG (γ) with MVP		RANK	GRG (γ) with MVP		RANK	GRG (γ) with MVP		RANK
AT (F)	0.8136	1	AT (F)	0.8753	1	AT (F)	0.8007	1
MRH (%)	0.7696	2	AWS (mph)	0.7985	2	MRH (%)	0.7492	2
AWS (mph)	0.7319	3	MRH (%)	0.7968	3	AWS (mph)	0.6734	3
M	0.7180	4	MLC	0.7916	4	P (in)	0.6644	4
MLC	0.7175	5	TTU	0.7915	5	MLC	0.6603	5
TTU	0.7170	6	M	0.7914	6	M	0.6602	6
PWD	0.6754	7	P (in)	0.7857	7	TTU	0.6602	7
P (in)	0.6706	8	PWD	0.7600	8	PWD	0.6506	8

TABLE 5.3.3. NO_2 GRA Complete Results

NO_2 GRA RESULTS								
Site 5			Site 67			Site 3002		
GRG (γ) with IVP		RANK	GRG (γ) with IVP		RANK	GRG (γ) with IVP		RANK
AWS (mph)	0.7639	1	AT (F)	0.8039	1	AT (F)	0.7706	1
MLC	0.7533	2	P (in)	0.7493	2	MRH (%)	0.7000	2
TTU	0.7466	3	AWS (mph)	0.7389	3	P (in)	0.7883	3
PWD	0.7391	4	PWD	0.7384	4	PWD	0.6430	4
P (in)	0.7317	5	M	0.7319	5	AWS (mph)	0.7128	5
AT (F)	0.7142	6	TTU	0.7228	6	MLC	0.7577	6
M	0.6741	7	MRH (%)	0.7164	7	M	0.7670	7
MRH (%)	0.5987	8	MLC	0.7159	8	TTU	0.7644	8
GRG (γ) with MVP		RANK	GRG (γ) with MVP		RANK	GRG (γ) with MVP		RANK
P (in)	0.7645	1	AT (F)	0.8039	1	P (in)	0.7404	1
MLC	0.7554	2	P (in)	0.7493	2	AWS (mph)	0.7319	2
AWS (mph)	0.7552	3	AWS (mph)	0.7429	3	MLC	0.7295	3
TTU	0.7549	4	MLC	0.7340	4	TTU	0.7292	4
M	0.7548	5	TTU	0.7338	5	M	0.7291	5
AT (F)	0.7481	6	M	0.7338	6	AT (F)	0.7201	6
PWD	0.6805	7	MRH (%)	0.7164	7	PWD	0.6854	7
MRH (%)	0.6407	8	PWD	0.6947	8	MRH (%)	0.6405	8

B: DGRA Results

Table 5.3.4. NO_2 DGRA Complete Results

NO_2 DGRA RESULTS											
Site 5				Site 67				Site 3002			
ξ	GRG (γ) with IVP		RANK	ξ	GRG (γ) with IVP		RANK	ξ	GRG (γ) with IVP		RANK
1.0000	AWS (mph)	0.8564	1	0.9965	AT (F)	0.8857	1	1.0000	AWS (mph)	0.8233	1
0.4511	PWD	0.7211	2	1.0000	MRH (%)	0.8301	2	0.4916	AT (F)	0.7678	2
0.4375	AT (F)	0.6885	3	0.4536	AWS (mph)	0.7218	3	0.9357	PWD	0.7583	3
0.6276	MRH (%)	0.6498	4	0.2594	PWD	0.6119	4	0.6531	MRH (%)	0.7502	4
0.0386	P (in)	0.2144	5	0.0184	P (in)	0.1362	5	0.0441	P (in)	0.3005	5
0.0332	MLC	0.2099	6	0.0176	MLC	0.0995	6	0.0213	M	0.1542	6
0.0275	TTU	0.1744	7	0.0129	TTU	0.0805	7	0.0170	TTU	0.1264	7
0.0359	M	0.1407	8	0.0101	M	0.0719	8	0.0172	MLC	0.1213	8
ξ	GRG (γ) with MVP		RANK	ξ	GRG (γ) with MVP		RANK	ξ	GRG (γ) with MVP		RANK
1.0000	AWS (mph)	0.8499	1	0.9965	AT (F)	0.8857	1	1.0000	AWS (mph)	0.8349	1
0.3833	AT (F)	0.6989	2	1.0000	MRH (%)	0.8301	2	0.8367	PWD	0.7730	2
0.5494	PWD	0.6986	3	0.4484	AWS (mph)	0.7239	3	0.5800	AT (F)	0.7455	3
0.5498	MRH (%)	0.6615	4	0.3017	PWD	0.5911	4	0.7705	MRH (%)	0.7263	4
0.0338	P (in)	0.2222	5	0.0184	P (in)	0.1362	5	0.0520	P (in)	0.2772	5
0.0305	MLC	0.1869	6	0.0169	MLC	0.1113	6	0.0246	M	0.1494	6
0.0292	M	0.1783	7	0.0126	TTU	0.0871	7	0.0195	MLC	0.1259	7
0.0251	TTU	0.1620	8	0.0100	M	0.0733	8	0.0195	TTU	0.1249	8

Table 5.3.5. SO_2 DGRA Complete Results

SO_2 DGRA RESULTS											
Site 5				Site 67				Site 3002			
ξ	GRG (γ) with IVP		RANK	ξ	GRG (γ) with IVP		RANK	ξ	GRG (γ) with IVP		RANK
1.0000	AWS (mph)	0.8302	1	0.9965	AT (F)	0.8451	1	0.9934	AWS (mph)	0.8523	1
0.4511	PWD	0.7005	2	0.4536	AWS (mph)	0.7879	2	0.9511	PWD	0.7513	2
0.4375	AT (F)	0.5412	3	1.0000	MRH (%)	0.7638	3	0.0451	P (in)	0.6860	3
0.6276	MRH (%)	0.5177	4	0.2594	PWD	0.6778	4	0.5010	AT (F)	0.5816	4
0.0386	P (in)	0.4693	5	0.0184	P (in)	0.3173	5	0.6650	MRH (%)	0.5714	5
0.0275	TTU	0.4413	6	0.0176	MLC	0.2321	6	0.0216	M	0.4869	6
0.0332	MLC	0.4412	7	0.0129	TTU	0.1998	7	0.0172	TTU	0.4114	7
0.0359	M	0.2753	8	0.0101	M	0.1994	8	0.0175	MLC	0.3595	8
ξ	GRG (γ) with MVP		RANK	ξ	GRG (γ) with MVP		RANK	ξ	GRG (γ) with MVP		RANK
1.0000	AWS (mph)	0.8514	1	0.9965	AT (F)	0.8451	1	1.0000	AWS (mph)	0.8245	1
0.5494	PWD	0.7303	2	0.4484	AWS (mph)	0.7885	2	0.8598	PWD	0.7371	2
0.3833	AT (F)	0.5716	3	1.0000	MRH (%)	0.7638	3	0.0533	P (in)	0.6499	3
0.5498	MRH (%)	0.5496	4	0.3017	PWD	0.6525	4	0.0250	M	0.5358	4
0.0338	P (in)	0.4916	5	0.0184	P (in)	0.3173	5	0.5930	AT (F)	0.5354	5
0.0305	MLC	0.4576	6	0.0169	MLC	0.2995	6	0.7872	MRH (%)	0.5257	6
0.0292	M	0.4576	7	0.0126	TTU	0.2486	7	0.0198	TTU	0.4793	7
0.0251	TTU	0.4438	8	0.0100	M	0.2070	8	0.0199	MLC	0.4551	8

Table 5.3.6. O_3 DGRA Complete Results

O_3 DGRA RESULTS											
Site 5				Site 67				Site 3002			
ξ	GRG (γ) with IVP		RANK	ξ	GRG (γ) with IVP		RANK	ξ	GRG (γ) with IVP		RANK
1.0000	AWS (mph)	0.8458	1	1.0000	AT (F)	0.9023	1	0.4916	AT (F)	0.8262	1
0.4375	AT (F)	0.7672	2	0.4552	AWS (mph)	0.7132	2	1.0000	AWS (mph)	0.7880	2
0.4511	PWD	0.7224	3	0.4083	MRH (%)	0.6963	3	0.7418	P (in)	0.7845	3
0.6616	P (in)	0.6937	4	0.2968	P (in)	0.6026	4	0.9357	PWD	0.7381	4
0.0726	MRH (%)	0.3268	5	0.2603	PWD	0.6015	5	0.1328	MRH (%)	0.5318	5
0.0332	MLC	0.1529	6	0.0177	MLC	0.0869	6	0.0213	M	0.1128	6
0.0275	TTU	0.1289	7	0.0130	TTU	0.0681	7	0.0170	TTU	0.0922	7
0.0359	M	0.1229	8	0.0101	M	0.0580	8	0.0172	MLC	0.0887	8
ξ	GRG (γ) with MVP		RANK	ξ	GRG (γ) with MVP		RANK	ξ	GRG (γ) with MVP		RANK
1.0000	AWS (mph)	0.8349	1	1.0000	AT (F)	0.9023	1	0.5800	AT (F)	0.8158	1
0.3833	AT (F)	0.7735	2	0.4500	AWS (mph)	0.7154	2	1.0000	AWS (mph)	0.7935	2
0.5796	P (in)	0.7017	3	0.4083	MRH (%)	0.6963	3	0.8752	P (in)	0.7694	3
0.5494	PWD	0.6939	4	0.2968	P (in)	0.6026	4	0.8367	PWD	0.7428	4
0.0636	MRH (%)	0.3343	5	0.3027	PWD	0.5808	5	0.1567	MRH (%)	0.5057	5
0.0305	MLC	0.1451	6	0.0170	MLC	0.0930	6	0.0246	M	0.1028	6
0.0292	M	0.1448	7	0.0127	TTU	0.0720	7	0.0195	MLC	0.0845	7
0.0251	TTU	0.1220	8	0.0101	M	0.0584	8	0.0195	TTU	0.0843	8

C: CCA Results

5.3.1. Site 5

5.3.1.1. Economic Results

TABLE 5.3.7. Canonical Solution for Economic Variables Predicting Pollutant Levels for Function 1 at Site 5

	Function 1			
Variable	*Coef*	r_s	$r_s^2(\%)$	$h^2(\%)$
MLC	0.9414	0.1401	1.9624	1.9624
M	-0.9200	-0.0074	0.0055	0.0055
TTU	-1.3723	-0.0282	0.0795	0.0795
O_3	0.2827	0.0441	0.1946	6.1830
NO_2	-0.5815	-0.0244	0.0595	1.8898
SO_2	1.0023	0.1504	2.2623	71.8743

TABLE 5.3.8. Canonical Correlations

Canonical Correlations		
Root No.	Canonical Correlation	Squared Correlation
1	0.1774	0.0315
2	0.1455	0.0212
3	0.0309	0.0010

TABLE 5.3.9. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.9471	1.9054	9.0000	759.4766	0.0482
2 to 3	0.9779	1.7594	4.0000	626.0000	0.1354
3 to 3	0.9990	0.3000	1.0000	314.0000	0.5843

TABLE 5.3.10. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
MLC	0.9414	-0.3473	0.4454
M	-0.9200	-0.7983	2.3271
TTU	-1.3723	-1.5151	1.5990

Table 5.3.11. Correlations Between Dependent and Canonical Variables

	Function No.		
Variable	1	2	3
MLC	0.1401	-0.0880	0.0032
M	-0.0074	0.1081	0.0206
TTU	-0.0282	-0.1328	-0.0116

Table 5.3.12. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	0.2827	0.7740	0.6767
NO_2	-0.5815	-0.7672	0.5741
SO_2	1.0023	-0.3514	-0.1054

Table 5.3.13. Correlations Between Covariates and Canonical Variables

	Canonical Variable		
Covariate	1	2	3
O_3	0.04411	0.06584	0.02646
NO_2	-0.02439	-0.09066	0.02379
SO_2	0.15041	-0.07116	0.00634

5.3.1.2. Weather Results

Table 5.3.14. Canonical Correlations

Root No.	Canonical Correlation	Squared Correlation
1	0.5320	0.2830
2	0.3322	0.1104
3	0.2527	0.0639

Table 5.3.15. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.5971	11.7219	15	856.1749	0.0000
2 to 3	0.8328	7.4481	8	622.0000	0.0000
3 to 3	0.9361	7.0945	3	312.0000	0.0001

Table 5.3.16. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
AT	0.2125	0.5784	-0.4071
MRH	0.3444	0.0413	0.5348
P	0.3679	0.0207	0.5416
PWD	0.2362	-0.8712	-0.2498
AWS	0.8194	0.2914	-0.1306

Table 5.3.17. Correlations Between Dependent and Canonical Variables (Canonical Loadings)

	Function No.		
Variable	1	2	3
AT	0.1042	0.1783	-0.1191
MRH	0.0287	-0.0214	0.1756
P	0.2559	0.0410	0.1412
PWD	0.2209	-0.2569	-0.0939
AWS	0.4316	0.0184	-0.0793

Table 5.3.18. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	-0.4375	0.1805	-0.9555
NO_2	-0.4710	-0.9225	0.4284
SO_2	-0.4722	0.8503	0.4396

Table 5.3.19. Correlations Between Covariates and Canonical Variables (Canonical Loadings)

	Canonical Variable		
Covariate	1	2	3
O_3	-0.3603	0.0014	-0.1859
NO_2	-0.4174	-0.1881	0.0638
SO_2	-0.3764	0.1863	0.1087

5.3.1.3. Weather and Economic Results

Table 5.3.20. Canonical Correlations

Root No.	Canonical Correlation	Squared Correlation
1	0.5448	0.2968
2	0.4082	0.1666
3	0.3512	0.1234

Table 5.3.21. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.5137	9.5837	24	890.9946	0.0000E+00
2 to 3	0.7306	7.4775	14	616	1.3545E-14
3 to 3	0.8766	7.2473	6	309	2.9706E-07

Table 5.3.22. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
AT	0.1227	0.5372	0.3182
MRH	0.5068	-0.8599	0.4657
P	0.3636	-0.1357	0.1582
PWD	0.2543	-0.1091	-0.8033
AWS	0.8235	0.1355	0.3267
MLC	-0.1554	-0.0417	0.4150
M	-0.2574	-0.2032	-0.1902
TTU	0.0578	-1.1266	-0.0484

TABLE 5.3.23. Correlations Between Dependent and Canonical Variables (Canonical Loadings)

	Function No.		
Variable	1	2	3
AT	0.0755	0.2016	0.1024
MRH	0.0488	-0.1595	0.0657
P	0.2650	-0.0698	0.1096
PWD	0.2308	-0.0094	-0.2651
AWS	0.4177	0.1352	-0.0111
MLC	-0.0920	-0.0426	0.1307
M	0.0088	0.1022	-0.0407
TTU	-0.0034	-0.1357	0.0123

TABLE 5.3.24. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	-0.5496	0.8603	-0.3078
NO_2	-0.3432	-0.8701	-0.6177
SO_2	-0.4935	-0.0468	0.9453

TABLE 5.3.25. Correlations Between Covariates and Canonical Variables (Canonical Loadings)

	Canonical Variable		
Covariate	1	2	3
O_3	-0.4081	0.2260	-0.1277
NO_2	-0.3829	-0.2411	-0.1392
SO_2	-0.3831	-0.0840	0.2390

5.3.2. Site 67

5.3.2.1. Economic Results

TABLE 5.3.26. Canonical Solution for Economic Variables Predicting Pollutant Levels for Functions 1, 2, and 3 at Site 67

	Function 1			
Variable	*Coef*	r_s	$r_s^2(\%)$	$h^2(\%)$
MLC	-0.9484	-0.2037	4.1493	4.1493
M	-0.9005	-0.2371	5.6227	5.6227
TTU	0.7183	-0.1879	3.5291	3.5291
O_3	-0.1140	-0.0791	0.6256	0.6256
NO_2	-0.0943	-0.0823	0.6781	0.6781
SO_2	-0.9528	-0.2676	7.1630	7.1630

TABLE 5.3.27. Canonical Correlations

Root No.	Canonical Correlation	Squared Correlation
1	0.2718	0.0739
2	0.1667	0.0278
3	0.0205	0.0004

TABLE 5.3.28. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.9000	3.6608	9	744.8742	0.0002
2 to 3	0.9718	2.2111	4	614.0000	0.0664
3 to 3	0.9996	0.1291	1	308.0000	0.7197

TABLE 5.3.29. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
MLC	-0.9484	1.3666	-1.1978
M	-0.9005	0.6298	0.8222
TTU	0.7183	-2.2768	0.2611

TABLE 5.3.30. Correlations Between Dependent and Canonical Variables

	Function No.		
Variable	1	2	3
MLC	-0.2037	-0.0507	-0.0120
M	-0.2371	-0.0376	0.0089
TTU	-0.1879	-0.1140	-0.0048

Table 5.3.31. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	-0.1140	1.1042	0.1038
NO_2	-0.0943	-0.3894	-1.0451
SO_2	-0.9528	-0.2065	0.2909

Table 5.3.32. Correlations Between Covariates and Canonical Variables

	Canonical Variable		
Covariate	1	2	3
O_3	-0.0791	0.1508	-0.0064
NO_2	-0.0823	0.0097	-0.0195
SO_2	-0.2676	-0.0190	0.0027

5.3.2.2. Weather Results

Table 5.3.33. Canonical Correlations

Root No.	Canonical Correlation	Squared Correlation
1	0.6090	0.3708
2	0.4481	0.2008
3	0.2765	0.0764

Table 5.3.34. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.4644	17.9281	15	839.6115	0.0000E+00
2 to 3	0.7381	12.5029	8	610.0000	1.1102E-16
3 to 3	0.9236	8.4432	3	306.0000	2.0855E-05

Table 5.3.35. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
AT	0.1066	-0.9594	-0.1214
MRH	0.0708	0.3043	-0.7194
P	0.3814	0.1713	0.5359
PWD	-0.2224	0.1833	-0.5180
AWS	0.8016	0.1401	-0.6566

Table 5.3.36. Correlations Between Dependent and Canonical Variables (Canonical Loadings)

	Function No.		
Variable	1	2	3
AT	0.0887	-0.4090	-0.0681
MRH	-0.1408	0.1243	-0.1215
P	0.4043	0.0894	0.0813
PWD	-0.0939	0.0108	-0.1783
AWS	0.5419	0.0039	-0.0684

Table 5.3.37. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	-0.5363	-0.7041	0.6780
NO_2	-0.6459	0.7754	-0.4841
SO_2	0.0208	-0.6458	-0.7859

Table 5.3.38. Correlations Between Covariates and Canonical Variables (Canonical Loadings)

	Canonical Variable		
Covariate	1	2	3
O_3	-0.4965	-0.2051	0.0980
NO_2	-0.5338	0.1614	-0.0883
SO_2	-0.0995	-0.2764	-0.2129

5.3.2.3. Weather and Economic Results

Table 5.3.39. Canonical Correlations

Root No.	Canonical Correlation	Squared Correlation
1	0.6150	0.3782
2	0.4733	0.2240
3	0.3390	0.1149

Table 5.3.40. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.4271	12.4099	24	873.5928	0.0000E+00
2 to 3	0.6868	8.9167	14	604.0000	0.0000E+00
3 to 3	0.8851	6.5574	6	303.0000	1.5860E-06

Table 5.3.41. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
AT	0.1001	0.9016	-0.1759
MRH	0.0765	-0.2161	0.5957
P	0.3753	-0.1825	-0.3345
PWD	-0.2318	-0.1128	0.4780
AWS	0.7824	0.0060	0.5500
MLC	-0.2231	0.3143	-0.0328
M	-0.1689	0.1746	0.2296
TTU	0.2722	-0.0707	0.3783

Table 5.3.42. Correlations Between Dependent and Canonical Variables (Canonical Loadings)

	Function No.		
Variable	1	2	3
AT	0.0662	0.4181	-0.0251
MRH	-0.1379	-0.1012	0.1441
P	0.4104	-0.0831	-0.0518
PWD	-0.0985	0.0246	0.1744
AWS	0.5392	0.0414	0.0769
MLC	-0.0601	0.1261	0.1572
M	-0.0639	0.1673	0.1601
TTU	-0.0149	0.0983	0.1960

Table 5.3.43. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	-0.5504	0.5022	-0.8294
NO_2	-0.6206	-0.6803	0.6362
SO_2	-0.0343	0.8067	0.6189

Table 5.3.44. Correlations Between Covariates and Canonical Variables (Canonical Loadings)

	Canonical Variable		
Covariate	1	2	3
O_3	-0.5081	0.1516	-0.1571
NO_2	-0.5329	-0.1545	0.1280
SO_2	-0.1331	0.3620	0.2057

5.3.3. Site 3002

5.3.3.1. Economic Results

Table 5.3.45. Canonical Solution for Economic Variables Predicting Pollutant Levels for Functions 1, 2, and 3 at Site 3002

	Function 1			Function 2			
Variable	$Coef$	r_s	$r_s^2(\%)$	$Coef$	r_s	$r_s^2(\%)$	$h^2(\%)$
MLC	0.0663	0.0663	0.4401	0.4172	-0.1277	1.6308	2.0709
M	-0.0176	-0.0176	0.0309	1.2376	-0.1229	1.5094	1.5402
TTU	0.0024	0.0024	0.0006	-2.4977	-0.1452	2.1080	2.1086
O_3	-0.2310	0.0173	0.0298	0.7337	0.1270	1.6134	1.6432
NO_2	1.0692	0.1855	3.4426	-0.1199	0.0410	0.1681	3.6107
SO_2	-0.1284	0.0157	0.0246	0.6114	0.1129	1.2755	1.3001

Table 5.3.46. Canonical Correlations

Root No.	Canonical Correlation	Squared Correlation
1	0.1924	0.0370
2	0.1573	0.0248
3	0.0730	0.0053

TABLE 5.3.47. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.9341	2.5872	9	820.3201	0.0061
2 to 3	0.9700	2.5893	4	676.0000	0.0358
3 to 3	0.9947	1.8176	1	339.0000	0.1785

TABLE 5.3.48. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
MLC	2.6297	0.4172	0.9160
M	-1.1955	1.2376	3.0056
TTU	-1.3014	-2.4977	-3.3489

TABLE 5.3.49. Correlations Between Dependent and Canonical Variables

	Function No.		
Variable	1	2	3
MLC	0.0663	-0.1277	0.0344
M	-0.0176	-0.1229	0.0451
TTU	0.0024	-0.1452	0.0281

TABLE 5.3.50. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	-0.2310	0.7337	0.7345
NO_2	1.0692	-0.1199	0.0016
SO_2	-0.1284	0.6114	-0.8267

TABLE 5.3.51. Correlations Between Covariates and Canonical Variables

	Canonical Variable		
Covariate	1	2	3
O_3	0.0173	0.1270	0.0426
NO_2	0.1855	0.0410	0.0031
SO_2	0.0157	0.1129	-0.0505

5.3.3.2. Weather Results

Table 5.3.52. Canonical Correlations

Root No.	Canonical Correlation	Squared Correlation
1	0.5780	0.3341
2	0.3424	0.1172
3	0.1077	0.0116

Table 5.3.53. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.5810	13.4071	15	925.189	0.000E+00
2 to 3	0.8725	5.9258	8	672.000	2.364E-07
3 to 3	0.9884	1.3182	3	337.000	2.683E-01

Table 5.3.54. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
AT	-0.0943	0.9346	0.0492
MRH	0.6966	-0.2269	0.1935
P	0.3379	-0.0184	-0.5281
PWD	-0.1373	-0.0676	-0.9060
AWS	0.6157	0.6531	0.0701

Table 5.3.55. Correlations Between Dependent and Canonical Variables (Canonical Loadings)

	Function No.		
Variable	1	2	3
AT	-0.1222	0.2442	-0.0049
MRH	0.3739	-0.0927	0.0045
P	0.3153	0.0005	-0.0442
PWD	-0.0870	0.0042	-0.0926
AWS	0.3047	0.1430	-0.0023

Table 5.3.56. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	-0.9138	0.4599	0.2910
NO_2	-0.1579	-1.0636	0.0380
SO_2	-0.0936	0.1768	-1.0167

Table 5.3.57. Correlations Between Covariates and Canonical Variables (Canonical Loadings)

	Canonical Variable		
Covariate	1	2	3
O_3	-0.5675	0.0513	0.0125
NO_2	-0.2741	-0.2991	-0.0117
SO_2	-0.1727	0.0035	-0.1028

5.3.3.3. Weather and Economic Results

Table 5.3.58. Canonical Correlations

Root No.	Canonical Correlation	Squared Correlation
1	0.6093	0.3713
2	0.3635	0.1321
3	0.1504	0.0226

Table 5.3.59. Dimension Reduction Analysis- Wilks' Lambda, using F-Approximation (Rao's F)

Roots	Wilks λ	F	Hypothesis DF	Error DF	Significance of F
1 to 3	0.5333	9.7168	24	963.5022	0.0000E+00
2 to 3	0.8483	4.0801	14	666.0000	7.7090E-07
3 to 3	0.9774	1.2892	6	334.0000	0.2615

Table 5.3.60. Standardized Canonical Coefficients for Dependent Variables

	Function No.		
Variable	1	2	3
AT	-0.1756	0.8490	0.0213
MRH	0.8066	-0.1882	-0.1167
P	0.2776	-0.0627	-0.2052
PWD	-0.1091	-0.0992	-0.6993
AWS	0.6047	0.5667	-0.0161
MLC	-0.7883	-0.6217	0.2206
M	-0.3556	0.8919	0.7782
TTU	0.8720	-0.2294	-0.2145

Table 5.3.61. Correlations Between Dependent and Canonical Variables (Canonical Loadings)

	Function No.		
Variable	1	2	3
AT	-0.1167	0.2461	-0.0203
MRH	0.3713	-0.0999	0.0231
P	0.3167	-0.0078	-0.0319
PWD	-0.0833	0.0019	-0.0960
AWS	0.3076	0.1366	0.0033
MLC	0.0702	-0.0865	0.0974
M	0.0828	-0.0029	0.1028
TTU	0.1045	-0.0295	0.1004

Table 5.3.62. Standardized Canonical Coefficients for Covariates

	Canonical Variable		
Covariate	1	2	3
O_3	-0.9140	0.4905	0.2350
NO_2	-0.1822	-1.0574	0.0783
SO_2	-0.0512	0.1333	-1.0263

TABLE 5.3.63. Correlations Between Covariates and Canonical Variables (Canonical Loadings)

	Canonical Variable		
Covariate	1	2	3
O_3	-0.5984	0.0634	0.0108
NO_2	-0.2977	-0.3155	-0.0133
SO_2	-0.1597	-0.0095	-0.1451

D: General Results

Table 5.3.64. Stepwise Regression, DGRA and GRA Result Comparison

Site 67 Stepwise Regression Results								
NO_2			SO_2			O_3		
Predictors	p-value	ESC	Predictors	p-value	ESC	Predictors	p-value	ESC
AWS (mph)	2.7900E-09	-0.3475	AT (F)	7.1100E-07	0.2769	AWS (mph)	1.73E-15	-0.4830
AT (F)	2.2500E-05	-0.2152	M	7.3524E-02	0.1329	P (in)	0.00181	-0.1658
P (in)	4.7100E-05	-0.2146	P (in)	1.2830E-03	-0.1848	MRH (%)	0.00251	-0.1652
PWD	5.0300E-05	0.2021	PWD	1.0148E-01	0.0883	AT (F)	0.02578	0.1135
M	1.7300E-01	0.0924	TTU	9.0063E-01	0.0160			
Site 67 GRA results								
NO_2			SO_2			O_3		
GRG (g) with IVP		RANK	GRG (g) with IVP		RANK	GRG (g) with IVP		RANK
AT (F)	0.8039	1	P (in)	0.8836	1	AT (F)	0.8753	1
P (in)	0.7493	2	M	0.8787	2	PWD	0.7977	2
AWS (mph)	0.7389	3	TTU	0.8665	3	MRH (%)	0.7968	3
PWD	0.7384	4	MLC	0.8572	4	AWS (mph)	0.7952	4
M	0.7319	5	AWS (mph)	0.8020	5	M	0.7899	5
TTU	0.7228	6	PWD	0.7912	6	P (in)	0.7857	6
MRH (%)	0.7164	7	AT (F)	0.7388	7	TTU	0.7826	7
MLC	0.7159	8	MRH (%)	0.6207	8	MLC	0.7769	8

Table 5.3.65. Continuation of Stepwise Regression, DGRA and GRA Result Comparison

Site 67 DGRA results											
NO_2				SO_2				O_3			
ξ	GRG (g) with IVP		RANK	ξ	GRG (g) with IVP		RANK	ξ	GRG (g) with IVP		RANK
0.9965	AT (F)	0.8857	1	0.9965	AT (F)	0.8451	1	1.0000	AT (F)	0.9023	1
1.0000	MRH (%)	0.8301	2	0.4536	AWS (mph)	0.7879	2	0.4552	AWS (mph)	0.7132	2
0.4536	AWS (mph)	0.7218	3	1.0000	MRH (%)	0.7638	3	0.4083	MRH (%)	0.6963	3
0.2594	PWD	0.6119	4	0.2594	PWD	0.6778	4	0.2968	P (in)	0.6026	4
0.0184	P (in)	0.1362	5	0.0184	P (in)	0.3173	5	0.2603	PWD	0.6015	5
0.0176	MLC	0.0995	6	0.0176	MLC	0.2321	6	0.0177	MLC	0.0869	6
0.0129	TTU	0.0805	7	0.0129	TTU	0.1998	7	0.0130	TTU	0.0681	7
0.0101	M	0.0719	8	0.0101	M	0.1994	8	0.0101	M	0.0580	8

Copyright Permission

Name: Hartman, Sarah

Email (to receive future readership statistics): hartmansarah02@gmail.com

Type of document: ['Thesis']

Title: Grey Relational Analysis and Canonical Correlation Analysis of Air Pollution in Three Kentucky Counties

Keywords (3-5 keywords not included in the title that uniquely describe content): Environmental Science, Grey System Theory, Weather, Economics

Committee Chair: Ngoc Nguyen

Additional Committee Members: Lukun Zheng, Melanie Autin

Select 3-5 TopSCHOLAR® disciplines for indexing your research topic in TopSCHOLAR®: Life Sciences Physical Sciences and Mathematics Medicine and Health Sciences