2022

DEVELOPMENT OF UNIVARIATE AND MULTIVARIATE FORECASTING MODELS FOR METHANE GAS EMISSIONS IN UNDERGROUND COAL MINES

Juan Diaz
University of Kentucky, jcdi246@uky.edu
Author ORCID Identifier:
https://orcid.org/ 0000-0002-1838-1927
Digital Object Identifier: https://doi.org/10.13023/etd.2022.050

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Juan Diaz, Student

Dr. Zach Aguiotantis, Major Professor

Dr. Jhon Silva Castro, Director of Graduate Studies

DEVELOPMENT OF UNIVARIATE AND MULTIVARIATE FORECASTING MODELS FOR METHANE GAS EMISSIONS IN UNDERGROUND COAL MINES

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Engineering at the University of Kentucky

By
Juan Carlos Díaz Martínez
Lexington, Kentucky
Director: Dr. Zacharias Agioutantis, Professor of Mining Engineering
Lexington, Kentucky
2022

https://orcid.org/0000-0002-1838-1927

ABSTRACT OF DISSERTATION

DEVELOPMENT OF UNIVARIATE AND MULTIVARIATE FORECASTING MODELS FOR METHANE GAS EMISSIONS IN UNDERGROUND COAL MINES

Methane gas management continues to be a challenge concerning underground coal mine safety and productivity worldwide despite the extraordinary effort of the mining industry, governmental agencies, and academia to develop new technologies to monitor and control methane gas emissions more efficiently. The risk of hazardous methane gas concentrations in underground environments cannot be underestimated. Statistical data for the last 100 years indicate that around 80% of the accidents and 90% of the fatalities in the underground coal mining industry in the US were related to methane gas explosions.

Modern underground mine operations monitor and evaluate atmospheric parameters such as barometric pressure, temperature, gas concentrations, and ventilation parameters (e.g., fan performance and airflow) by means of Automated Atmospheric Monitoring Systems, which use sensors that collect a massive amount of data implemented by mine operators to make decisions concerning mine safety and operate ventilation systems more effectively. In addition, however, some of these data can be statistically studied to develop forecast models to help improve the safety and health parameters of underground coal mining operations.

The research presented in this dissertation investigates potential correlations between methane gas concentrations and independent variables such as barometric pressure and coal production rate to build reliable forecasting models capable of predicting future concentrations of methane gas, mainly based on time series data collected by the Atmospheric Monitoring System of three active underground coal mining operations in the eastern US and weather data retrieved from public weather stations in the proximity of the case studies. The mine and weather data were stored and pre-processed using an Atmospheric Monitoring Analysis and Database Management system explicitly designed to manage Atmospheric Monitoring Systems data. Furthermore, various statistical techniques were implemented to assess the potential association (e.g., autocorrelation and cross-correlation) between methane gas concentration time series and the independent variables. Such associations were employed to develop univariate and multivariate forecasting models for methane gas emissions in underground coal mines. Finally, the optimal model is selected using the Akaike Information Criterion, and the results obtained from the different forecast approaches (univariate and multivariate) are compared using cross-validation metrics to determine the best model.

It was concluded that the ARIMA, VAR, and ARIMAX methane gas forecasting methodologies proposed in this research can accurately predict methane gas concentrations in underground coal mines operations. The methane gas forecasted from the models matched the validation data consistently, and their linear correlation was positive and strong in most cases. In addition, the 95% confidence interval consistently captured the forecast and validation data.

KEYWORDS: Underground coal mining industry, methane gas, time series analysis, univariate and multivariate forecasting.

Juan C. Díaz Martínez

04/26/2022

Date

DEVELOPMENT OF UNIVARIATE AND MULTIVARIATE FORECASTING MODELS FOR METHANE GAS EMISSIONS IN UNDERGROUND COAL MINES

By

Juan Carlos Díaz Martínez

Dr. Zach Agioutantis

Director of Dissertation

Dr. Jhon Silva-Castro

Director of Graduate Studies

04/26/2022

Date

DEDICATION

This work is entirely dedicated to my loved mother; Consuelo and family, and friends from Colombia; without whose constant support, this dissertation would be impossible. They always inspire and motivate me to keep working to fulfill my dream of becoming a Doctor of Philosophy. Furthermore, my thanks also go to my friends from the University of Kentucky for their constant help and encouragement.

Foremost, I would like to express my sincere gratitude to Professor Oscar Jaime Restrepo Baena from La Universidad Nacional de Colombia for always providing me with his guidance, support, and advice when I most needed to succeed in my career and professional life.

ACKNOWLEDGEMENTS

I would like to thank the following people, without whom I would not have been able to complete this research; first, my advisor, Dr. Zach Agioutantis, for allowing me to work on an exciting and valuable topic for the mining industry. Furthermore, for his permanent support and guidance during this research. Similarly, I would like to thank the other two research project members, Dr. Dionissios Hristopulos and Dr. Steven Schafrik, for their trust, assistance, and help throughout this journey. Finally, I wish to thank Dr. Thomas Novak and Dr. Kray Luxbacher, who serve as members of my Dissertation Committee. Each member has provided significant insights and support in developing and publishing this dissertation.

In addition, I would like to thank the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. for providing the funding to conduct this research.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS....iii

LIST OF TABLES....vii

LIST OF FIGURES....viii

1. INTRODUCTION....1

1.1 Objectives....2

1.2 Innovation of Dissertation....2

1.3 Publications....3

1.4 Structure of Dissertation....3

2. LITERATURE REVIEW ON TIME SERIES ANALYSIS FOR METHANE GAS FORECASTING....5

2.1 Abstract....5

2.2 Introduction....5

2.3 Methane Prediction Methods....6

2.3.1 Empirical Methane Prediction Methods....7

2.3.2 Numerical Methane Prediction Methods....8

2.3.3 Time Series Analysis for Methane Forecasting....9

2.4 Comparison and Discussion on Methane Prediction Methods....11

2.5 Case Study....12

2.6 Preliminary Results and Discussion....12

2.7 Summary and Conclusions....13

2.8 Acknowledgment....14

2.9 References....14

3. TIME SERIES DATA MANAGEMENT AND ANALYSIS....17

3.1 MANAGING AND UTILIZING BIG DATA IN ATMOSPHERIC MONITORING SYSTEMS FOR UNDERGROUND COAL MINES....17

3.1.1 Abstract....17

3.1.2 Introduction....18

3.1.3 Methane Gas Generation in Underground Coal Mines....19

3.1.4 Coal Mine Methane Forecasting Methods....20

3.1.5 Case Study....20

3.1.6 Results and Discussion....21

3.1.7 Summary and Conclusions....25

3.1.8 References....26

3.2 INVESTIGATING RELATIONSHIPS BETWEEN METHANE EMISSIONS AND ATMOSPHERIC DATA IN UNDERGROUND COAL MINES TO DEVELOP A FORECASTING MODEL 28
3.2.1 Abstract 28
3.2.2 Introduction 28
3.2.3 Case Study 29
3.2.4 Results and Discussion 35
3.2.5 Summary and Conclusions 38
3.2.6 Acknowledgment 39
3.2.7 References 39
4. UNIVARIATE FORECASTING APPROACH FOR METHANE GAS CONCENTRATIONS IN UNDERGROUND COAL MINES 40
4.1 Abstract 40
4.2 Introduction 41
4.3 Time Series Analysis for Methane Gas Forecasting 42
4.4 AMS in Underground Coal Mines 46
4.5 Atmospheric Data Management 48
4.5.1 Data Pre-Processing 50
4.5.2 Spike Detection and Analysis 52
4.6 Exploratory Analysis of Methane Time Series 54
4.6.1 Correlation of Methane Concentration with Barometric Pressure and Production 54
4.6.2 Estimation of Lagged Cross-Correlations 56
4.6.3 Autocorrelation of Methane Time Series 57
4.6.4 Variogram Estimation 58
4.7 Time Series Forecasting Models 60
4.7.1 ARIMA Model of Methane Concentration 61
4.7.2 ARIMA One-Step-Ahead Model of Methane Concentration 62
4.8 Discussion 66
4.9 Conclusions 68
4.10 Acknowledgments 68
4.11 Statements and Declarations 68
4.12 References 69
5. MULTIVARIATE FORECASTING APPROACH FOR METHANE GAS CONCENTRATIONS IN UNDERGROUND COAL MINES 74
5.1 Abstract 74
5.2 Introduction 75

5.3 Materials and Methods 77
5.3.1 Case Studies and Data 77
5.3.2 Univariate and Multivariate Forecasting Methodologies 79
5.3.3 Proposed Forecasting Approaches 79
5.3.3.1 ARIMA (p,d,q) model 79
5.3.3.2 Multivariate Vector Autoregressive model 80
5.3.3.3 ARIMAX (p,d,q) model 81
5.4 Results 81
5.4.1 VAR(p) One-Step-Ahead Model of Methane Concentration 82
5.4.2 ARIMAX One-Step-Ahead Model of Methane Concentration 85
5.5 Discussion 87
5.6 Conclusions 94
5.7 Acknowledgments 94
5.8 References 95
6. OVERALL DISCUSSION AND CONCLUSIONS 99
6.1 Overall Discussion 99
6.2 Overall Conclusions 100
6.3 Overall Recommendations 101
REFERENCES 103
VITA 110

LIST OF TABLES

Table 2.1 Description of collected data 12
Table 3.1.1 Description of collected data – Shafts B and D 20
Table 4.1 One-step-ahead forecast results summary 65
Table 5.1 Weather data available from Weather Underground 78
Table 5.2 Summary of methane gas forecasting results for the VAR(p) one-step-ahead model (using a vector of methane gas concentration and barometric pressure) 84
Table 5.3 Summary of methane gas forecasting results for the ARIMAX one-step-ahead model (using barometric pressure time series as an independent variable) 86
Table 5.4 Univariate and multivariate forecasting results summary.. 93

LIST OF FIGURES

Figure 2.1 Methane forecasting after Kaffanke (1980, cited in Dixon, 1992) 10
Figure 2.2 Methane multivariate forecasting after (Dixon and Longson, 1993) 11
Figure 2.3 Correlation of time series data for Shaft A 13
Figure 3.1.1 Schematic of data collection and pre-processing sequence 19
Figure 3.1.2 Data visualization in AMANDA: (a) Barometric pressure for a five-day interval and (b) methane gas concentration for a five-day interval 21
Figure 3.1.3 Superposition of BP from the mine and regional stations for a period of 30 days 22
Figure 3.1.4 Data visualization in AMANDA: (a) Barometric pressure, methane concentration, and production for 180 days (b) Barometric pressure and methane concentration for 180 days 23
Figure 3.1.5 (a) Raw data, (b) raw and interpolated data, (c) daily median data, (d) correlation of daily median data for a specific data segment 24
Figure 3.1.6 The plot of the cross-correlation function between the barometric pressure and methane gas concentration. The horizontal axis represents the time lag between the two variables, and the vertical axis represents the cross-correlation values 24
Figure 3.1.7 (a) Raw data, (b) raw and interpolated data, (c) daily median data, (d) correlation of daily median data for a specific data segment 25
Figure 3.2.1 Schematic representation of data collection, pre-processing, and processing 35
Figure 3.2.2 (a) Raw (before cleaning) CH_4 data from Mine A, (b) CH_4 data from Mine A – after data is cleaned up, (c) Raw CH_4 data from Mine B, (d) CH_4 data from Mine B – after data is cleaned up 37
Figure 3.2.3 (a) Time series of daily averages for barometric pressure (green) and CH_4 concentration (red) extending over a six-month period. The value of the Pearson correlation coefficient (R=-0.83) is also shown. (b) The same plots as in (b) for weekly averages; R=-0.84. 37
Figure 3.2.4 (a) Daily median values of gas concentration and barometric pressure for a specific time window (data segment 1). (b) Scatter plot and correlation of daily median gas concentration and barometric pressure values for data segment 1. (c) Daily median values of gas concentration and barometric pressure for a different time window (data segment 2). (d) Scatter plot and correlation of gas concentration and barometric pressure for data segment 2. Both data segments refer to Mine A 38
Figure 4.1 Fatalities and incidents in underground US coal mines since 1900; 80% of the incidents and 90% of the fatalities can be attributed to methane gas explosions (NIOSH, 2020) 42
Figure 4.2 Wireless Multi-Gas Monitor (AMR PEMCO, 2002) 47
Figure 4.3 Wireless Multi-Gas Monitor process 48
Figure 4.4 Flow diagram of data management 49
Figure 4.5 Data homogenization representation: Left: Raw time series data with irregular sampling steps converted to resampled data at an arbitrary spacing. Right: Raw time series data with irregular spacing are converted to average values for specified time intervals 50
Figure 4.6 Data pre-processing (a) Raw CH_4 data from Mine A (before filtering), (b) CH_4 data from Mine A (after data is cleaned up), (c) Inconsistent CH_4 data from Mine A, (d) Segment where no CH_4 data were collected from Mine A sensors 51
Figure 4.7 Data pre-processing, (a) Raw CH_4 data from Mine B (before cleaning), (b) CH_4 data from Mine B (after data is cleaned up), (c) Barometric pressure data (before cleaning), (d) Barometric pressure data (after cleaning) 52

Figure 4.8 Spike detection, (a) CH_4 spike identification, (b) Correlation between CH_4 spike and daily coal production, (c) CH_4 inverted spike identification, (d) Correlation between CH_4 inverted spike and daily coal production 53
Figure 4.9 Spike detection, (a) CH_4 spike identification, and (b) Correlation between CH_4 spike and daily coal production........ 54
Figure 4.10 Pearson correlation between BP and CH_4 emissions for Mine A, (a) Time series plot of average daily values for segment 1, (b) Scatterplot for segment 1, (c)Time series plot of average daily values for segment 2, (d) Scatterplot for segment 2........ 55
Figure 4.11 Pearson correlation between CH_4 emissions and coal production for Mine A, (a) Time series plot of average daily values, (b) Time series plot of average weekly values........ 56
Figure 4.12 Correlation investigation using the empirical cross-correlation function, (a) Cross-correlation between CH_4 and BP, (b) Correlation coefficient (R) between CH_4 and BP 57
Figure 4.13 Correlation investigation using cross-covariance, (a) Cross-covariance between CH_4 and BP, (b) Correlation coefficient (R) between CH_4 and BP........ 57
Figure 4.14 Demonstration of non-stationarity of methane time series, (a) Average daily values from Mine A, (b) Plot of average weekly values from Mine C........ 58
Figure 4.15 Variograms for methane time series, (a) Average daily values from Mine A – segment 1, (b) Average daily values from Mine A – segment 2, (c) Average weekly values from Mine C – segment 1, (c) Average weekly values from Mine C – segment 2 60
Figure 4.16 ARIMA one-step-ahead CH_4 concentration forecasts using a daily average time step;(a) Forecasting of segment 1,(b) Magnified view of the forecast in (a), (c) Forecasting of segment 2,(d) Magnified view of the forecast in (b) 64
Figure 4.17 ARIMA one-step ahead CH_4 concentration forecasting using a 12-hour average time step, (a) Forecasting of segment 1,(b) magnified view of the graph a,(c) Forecasting of segment 2,(d) Magnified view of the graph b 65
Figure 5.1 Time series data collected using daily average values: methane gas time series represented by the red line, coal production rate time series exemplified by the blue line, and barometric pressure time series symbolized by the green line........ 78
Figure 5.2 VAR(p) one-step-ahead CH_4 concentration forecasts using daily average values;(a) Forecasting for dataset 1, (b) Magnified view of the forecast in (a), (c) Forecasting for dataset 2, (d) Magnified view of the forecast in (b)........ 83
Figure 5.3 VAR(p) one-step-ahead CH_4 concentration forecasts using 12-hour average values;(a) Forecasting for dataset 3, (b) Magnified view of the forecast in (a), (c) Forecasting for dataset 4, (d) Magnified view of the forecast in (b)........ 84
Figure 5.4 ARIMAX one-step-ahead CH_4 concentration forecasts;(a) Forecast for dataset 1 using daily average values, (b) Forecast for dataset 2 using daily average values, (c) Forecast for dataset 3 using 12-hour average values, (d) Forecast for dataset 4 using 12-hour average values........ 86
Figure 5.5 Methane gas and barometric pressure time series included in dataset 6 89
Figure 5.6 Proposed methodology for selecting the best univariate or multivariate forecast model 89

1. INTRODUCTION

Variations in gas emission levels in underground coal mines are related to many factors, including coal production rates. Therefore, as underground coal mine production increases, ventilation systems must deal with an intensified load of pollutants (e.g., carbon monoxide, carbon dioxide, and methane gas). The importance of these pollutants cannot be underestimated. Methane gas has threatened underground coal mining safety and productivity for more than 100 years. Methane gas was nicknamed "*the miner's curse*" in the 19^{th} century after the first documented methane gas explosions in the United States (US) and France (Byrer et al., 2014; Flores, 1998).

Methane gas is the most hazardous flammable gas found in underground coal operations worldwide. Explosions in underground coal mines because of unsafe methane gas concentrations have been the leading cause of incidents and fatalities in the US mining industry. Since 1900, more than 11,000 underground coal mine workers have died in over 500 mining accidents (NIOSH, 2020). Recent examples due to methane gas explosions worldwide include the mining disaster at Soma Coal Mine, Turkey, in 2014, which caused more than 300 fatalities and has been considered the worst mining disaster so far in the 21^{st} century. Moreover, the latest mine accident at Listvyazhnaya coal mine, Russia, in November 2021, which left more than 52 fatalities, is a reminder of how dangerous (a) unmanaged concentration of methane gas and (b) the implementation of inadequate mine monitoring systems can be (Kozlov, 2021; Düzgün, 2014).

Real-time monitoring and evaluation of underground environmental parameters (e.g., gas concentration, barometric pressure, temperature, fan performance, and airflow) are essential for handling hazardous methane gas concentrations. Automated Atmospheric Monitoring Systems (AMS) used in underground coal mines operations typically collect and store a tremendous amount of data employed by operators to make decisions regarding mine safety and operate ventilation systems more efficiently. However, these data are generally under-utilized, whereas it is feasible that they can be analyzed and evaluated using different statistical techniques to develop forecasting models of future methane emission levels in underground environments (Agioutantis et al., 2014).

The research presented in this dissertation tackles the problem mentioned above by developing different methane gas forecasting models based on time series analysis. First, historical time series data from different sources (e.g., underground coal mines and weather stations) have been collected and imported into an Atmospheric Monitoring Analysis and Database mAnagement (AMANDA) system for data pre-processing and homogenization. After that, the data have been exported into the MATLAB® programming environment for further processing and statistical analysis. Finally, different methane gas concentration forecasting models were developed based on univariate and multivariate forecasting approaches, and their performance was evaluated using cross-validation metrics to determine the best forecast model among different model families for each specific dataset.

The general objective of this research is to safeguard and improve the safety and health conditions of mineworkers by identifying and quantifying techniques that provide meaningful correlations between methane gas concentrations and independent variables such as barometric pressure and coal production rate in order to develop robust and accurate methane gas forecast models for underground coal mine operations. In addition, this research will (a) contribute to the development of forecast methods for predicting methane gas concentrations and emissions in underground environments, especially in underground coal mines, that can lead to enhanced health and safety of mining personnel, and (b) identify research gaps in this field which should encourage new studies.

1.1 Objectives

This dissertation addresses the following objectives:

- ❖ Objective 1: Determine the state-of-the-art in methane emissions forecasting for underground coal mines. Study the literature about developing forecasting models to predict methane gas concentrations and emissions in underground coal mine operations, particularly the implementation of time series analysis.

- ❖ Objective 2: Collect mine and weather data. Gather mine data from the case studies, which, in most cases, consist of methane gas and coal production rate, and weather data comprised of barometric pressure from the nearest weather station to each case study.

- ❖ Objective 3: Pre-process and homogenize the mine and weather data collected to ensure data consistency and integrity. This objective includes detecting and filtering erroneous values and homogenizing the different time series data to guarantee that they share the same timestamp.

- ❖ Objective 4: Develop and validate long-term relationship(s) between methane emissions and independent variables. Estimate the potential autocorrelation of methane gas concentration time series and evaluate the possible cross-correlation between methane gas vs. barometric pressure and methane gas vs. coal production rate.

- ❖ Objective 5: Develop univariate and multivariate forecasting models to predict methane gas concentrations and compare their accuracy and reliability to determine the best forecasting model.

1.2 Innovation of Dissertation

Although the mining industry and academia have made a lot of efforts to manage methane gas concentrations in underground environments more efficiently and effectively, this is one of the first documented attempts that use methane gas, barometric pressure, and coal production rate time series data simultaneously to develop univariate and multivariate forecasting methodologies capable of predicting future concentrations of methane gas to improve the safety and health of underground coal mines.

The following general innovative elements of this research arise from this dissertation:

- ❖ The identification of appropriate filtering mechanisms for pre-processing the data collected by Automated Atmospheric Monitoring Systems installed in underground coal mines and data collected from external sources.
- ❖ The implementation of time series data to develop different univariate and multivariate forecasting techniques to predict future levels of methane gas in underground coal mines operations.
- ❖ The comparison of various univariate and multivariate methane gas forecasting approaches using cross-validation metrics and validation data to assess their accuracy and reliability.

1.3 Publications

The following publications are the result of the research presented in this dissertation:

- ❖ Diaz, JC., Agioutantis, Z., Schafrik, S., Hristopulos, DT., Luxbacher, K. (2022). Investigating relationships between methane emissions and atmospheric data in underground coal mines to develop a forecasting model. Society for Mining, Metallurgy, and Exploration (SME). Feb. 27 - Mar. 02, 2022, Salt Lake City, UT. Preprint 22-025.
- ❖ Diaz, JC., Agioutantis, Z., Schafrik, S., Hristopulos, DT. (2021). Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. Mater. Proc. 2021, 5, 78. https://doi.org/10.3390/materproc2021005078.
- ❖ Diaz JC., Agioutantis Z., Schafrik S., Hristopulos DT. (2021). Towards atmospheric monitoring data analysis in underground coal mines. In: Proceedings of the 18th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781003188476-51.
- ❖ Diaz, JC., Agioutantis, Z., Hristopulos, DT., Luxbacher, K., Schafrik, S. (2022). Time series modeling of methane gas in underground mines. Mining, Metallurgy, and Exploration Journal *(submitted)*.
- ❖ Diaz JC., Agioutantis Z., Hristopulos DT., Luxbacher K., Schafrik S. (2022). Forecasting of methane gas in underground coal mines: univariate vs. multivariate modeling approaches. Process Safety and Environmental Protection Journal (*submitted)*.

1.4 Structure of Dissertation

This dissertation is composed of several peer-reviewed journal and conference papers that have been published, have already been submitted, or are in preparation. Therefore, the chapters within this dissertation are structured as follows:

Chapter 2 presents a comprehensive literature review of different techniques used in previous research to forecast methane gas concentrations and emissions in underground coal mining environments. The literature review is presented through a paper published in the North American Mine Ventilation Symposium in June 2021.

Chapter 3 describes the main steps of data pre-processing and analysis employed in this research. Two manuscripts are included: an article published in the International Conference on Raw Materials and Circular Economy in September 2021 and a paper published and presented at the Society of Mining, Metallurgy, and Exploration symposium in February 2022.

Chapter 4 explains the statistical techniques implemented to investigate the potential association (correlation and autocorrelation) and validate the long-term relationship(s) between methane gas emissions and possible independent variables (e.g., barometric pressure and coal production rate) along with the preliminary results of the univariate methane gas forecasting methods proposed by this research. It is covered by a paper submitted in January 2022 to the Mining, Metallurgy, and Exploration Journal.

Chapter 5 presents the results obtained from the multivariate methane gas forecasting methods proposed by this research and compares their accuracy with the univariate approach presented in Chapter 4 through an article that will be submitted to the Process Safety and Environmental Protection Journal.

Finally, Chapter 6 presents a short discussion of the main findings as well as the conclusions and recommendations for the entire body of work.

2. LITERATURE REVIEW ON TIME SERIES ANALYSIS FOR METHANE GAS FORECASTING

The literature review of previous research for methane gas forecasting in underground coal mines is covered by the following peer-reviewed article published in the proceedings of the 2021 North American Mine Ventilation Symposium (NAMVS), held online from June 12-17, 2021.

TOWARDS ATMOSPHERIC MONITORING DATA ANALYSIS IN UNDERGROUND COAL MINES

Juan Diaz, University of Kentucky; Lexington, Kentucky; United States of America

Zach Agioutantis, University of Kentucky; Lexington, Kentucky; United States of America

Steven Schafrik, University of Kentucky; Lexington, Kentucky; United States of America

Dionissios Hristopulos, Technical University of Crete; Chania, Crete; Greece

Full Citation:

Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Towards atmospheric monitoring data analysis in underground coal mines. In: Proceedings of the 18th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781003188476-51.

2.1 Abstract

Atmospheric Monitoring Systems (AMS) used in underground coal mines typically collect and store a tremendous amount of data. Logged values, such as gas concentrations, ambient temperature, barometric pressure, humidity, air velocity, as well as a variety of fan related data are seldom integrated into a single system for analysis. Typically, the data are utilized in individual systems by operators in order to make decisions regarding the health and safety of the workforce as well as for managing ventilation systems more efficiently.

This research discusses different studies that investigate the dependence between gas concentration measurements from underground mines and atmospheric data by means statistical measures of association. The objective is to identify and quantify techniques that provide meaningful correlations between potentially harmful gas concentrations and meteorological variables, which will allow the development of a robust predictive model. Preliminary results from a case study in the eastern USA are presented.

2.2 Introduction

Variations in gas emission rates in underground coal mines are directly related to coal production rates. Consequently, as underground coal mine production increases, ventilation systems have to deal with an intensified load of pollutants such as carbon monoxide (CO), carbon dioxide (CO_2), hydrogen sulfide (H_2S), and methane gas (CH_4), among others. The importance of these pollutants cannot be underestimated. Primarily, methane gas has always been considered a serious threat to underground coal mining safety and productivity due to its high toxicity and explosivity risk. The mining disaster at Soma Coal Mine, Turkey, in 2014, which caused more than 300 fatalities, and has been considered the worst mining accident so far in the 21st century, is a reminder of how dangerous (a) the excessive concentration of methane gas and (b) the implementation of inadequate mine monitoring systems can be (Düzgün and Leveson, 2018).

Methane gas is the most frequently threatening flammable gas found in underground longwall coal mines (Chaulya and Prasad, 2016). It is highly explosive when its concentration exceeds a limit of 5% to 15%, known as the explosive range. In this scope, methane gas can be readily ignited by the presence of an ignition source such as an open flame, explosives, a spark from electrical equipment, or a cutting bit. Sometimes spontaneous heating of coal can trigger an ignition (Karacan et al., 2011; Kissell, 2006).

There are generally two methods for dealing with excessive methane emissions in underground coal mines. The first and most widely used is the implementation of mine ventilation systems. Their main objective is to dilute methane concentration to acceptable levels minimizing the risk of explosion. The second one is coal seam degasification techniques based on implementing borehole designs to remove the coal mine methane from the underground mining environment before, during, or after coal production (Karacan et al., 2011). This research focuses on underground coal mine ventilation and monitoring systems.

Real-time monitoring and evaluation of environmental underground mine parameters, such as barometric pressure, dry bulb and wet bulb temperature, humidity, air velocity, fan performance, gas concentration, and airflow, are essential for dealing with hazardous methane gas concentrations (Agioutantis et al., 2014). Atmospheric Monitoring Systems (AMS) used in underground coal mines typically collect and store a tremendous amount of data employed by operators to make decisions regarding mine safety and operate ventilation systems more efficiently and effectively. However, these data are generally under-utilized, whereas it is feasible that they can be used to create short or long-term predictive models of future levels of methane emission (Agioutantis et al., 2015; Dixon and Longson, 1993).

Two main points have been discussed so far. The first one is the high gas emission rates in underground coal mines due to the rise in coal production rates. The second one is the under-utilization of data collected by atmospheric monitoring systems. This research tackles this problem. This study investigates the dependence between gas concentration measurements from underground mines and atmospheric data by employing statistical measures of association. The objective is to predict the concentration of harmful gases using the data collected and stored by AMS.

2.3 Methane Prediction Methods

In recent years, there has been significant progress concerning monitoring and prediction of methane gas emissions in underground coal mining due to technological advances in different fields such as electronics, data transmission systems, atmospheric monitoring systems, the internet of things (IoT), artificial intelligence, and machine learning (Agioutantis et al., 2015; Byung and Asad, 2018). However, the development of accurate methane gas prediction methods still presents challenges primarily due to the multiple variables and sources involved in methane gas emission into the underground mining environment.

The most critical parameters that influence methane emission in underground coal mines can be classified into two major groups (Karacan, 2008). The first group includes parameters related to the geological characteristics of the coal deposits, such as gas content of mined coal seams (Boyer and Qingzhao, 1998), depth of the mined coalbed, reservoir properties of coalbed (Lunarzewski, 1998), coal rank, and strength of the overlying strata (Karacan et al., 2011). The second group includes the mining process parameters, also known as operational factors, which involve mining and coal productivity (Karacan, 2008).

Identifying and analyzing the parameters that influence methane emission in coal mining is essential for developing an accurate methane gas prediction method. These methods can be classified into three different categories based on the approach employed. The first category is the empirical approach based on data collected by observing a process or phenomenon. Depending on the nature of the research, the data employed can be qualitative or quantitative. The second category is the numerical approach, which implements a numerical approximation or mathematical tools to solve physical models. In this case, numerical methods are used to predict the emission and concentration of methane gas. Finally, the third category is the statistical approach, which is based on collecting and analyzing raw data using different mathematical techniques to find patterns and build a statistical model that forecasts methane gas emissions and concentration (Dixon and Longson, 1993).

2.3.1 Empirical Methane Prediction Methods

Dunmore (1981) developed an empirical methane prediction method for longwall coal mines used by the Mining Research and Development Establishment (MRDE) of the British Government. This study was based on Airey's theoretical treatments of gas emissions from coal seams (Airey, 1968). The MRDE methane prediction method is founded on the coal seams' initial gas content, the coal seams' thickness, the degree of emission expected, and the rate of coal production at the working face.

Creedy (1993) recommended a methane prediction model (Equation 2.1) by studying the methane emissions from basically three primary sources. The first one is methane releases from coal mines that implement drainage techniques, the second source is methane from non-drainage coal mines, and the third source is methane releases from coal storage. This study was based on annual historical data (collected from 1966 to 1988) of methane emissions to the atmosphere from deep mines in the United Kingdom.

$$E_D = LfP_w + \big((1.857D) + (D - U)\big) + RfP_t \tag{2.1}$$

where P_w = coal production from mines without methane drainage (tons/year), P_t = deep mine coal production (total tons/year), D = mass of methane drained from all mines (total volume/year), U = total methane used, L = methane gas released from mines without drainage of 6 m^3/ton, R = remaining CH_4 content arriving at the surface of 2 m^3/ton, and f = converting factor of volume to mass flow.

Kirchgessner et al. (1993) introduced a regression equation (Equation 2.2) that satisfactorily predicts the concentration of methane gas emissions from underground coal mines. This study's regression equation is based on the relationships between mine emissions, coalbed methane content, and coal production rate.

$$\mathrm{ME} = 1.08 \times 10^{-7}(CP \times MC) + 31.44 - 26.76 \times DV \quad (2.2)$$

where *ME* = total emissions of methane gas in a year, *CP* = annual coal production (tons); *MC* = total methane content of the unmined coal (m^3/ton); and *DV* = step function. In this equation DV =1 if (CP×MC) is less than 7.6×10^5, and DV = 0 if (CP×MC) is greater than or equal to 7.6×10^5.

Diamond et al. (1997) proposed a forecast model for methane emission based on the increase of longwall panel face width from 229 to 305 m. The predictions were based on emission trends established by continuous monitoring of methane emission rates on existing longwall panels.

Bustin and Clarkson (1998) attempted to characterize and calculate the effect of rank and mineral composition variability on the methane adsorption/desorption characteristics of coals to establish a methane predictive multivariant model based on different types of coals. However, the authors concluded that their attempts to develop this predictive methane emission model were unsuccessful. The multiple regression analysis of the model data set resulted in a large standard error.

2.3.2 Numerical Methane Prediction Methods

Numerical methane prediction methods are principally based on two considerations: Darcy's Law, which describes a fluid's flow through a porous medium, and forecasting techniques. The first researchers to consider numerical prediction techniques were Owili-Eger, Stefanko, and Ramani from the Pennsylvania State University (Dixon, 1992).

Owili-Eger et al. (1973) developed a mathematical model that can forecast methane emissions based on computational models and algorithms that approximate methane flow patterns and quantities through coal seams and the mine atmosphere. The applications of this mathematical model were demonstrated successfully. However, further research proved that this model only was able to work accurately for shallow mines (Dixon, 1992).

Sung et al. (1987) explored a two-dimensional finite-difference model to forecast methane gas concentration and emission rates into active coal mine working areas based on different methane drainage schemes. This numerical model applies a flow theory (Fick and Darcy's Laws) that describes the flow of methane gas in a heterogeneous and anisotropic coal seam, which models coal's microscopic and macroscopic pore structure.

Ediz and Edwards (1991) proposed a numerical model for longwall mines that simulates gas flow through a porous medium. This model employs a general time-dependent gas flow equation for anisotropic media based on Darcy's Law. An approximate solution to this equation was obtained numerically, employing the finite element method.

Tauziede and Pokryszka (1993) investigated a dynamic mathematical model to predict methane gas concentration and emission in longwall coal mines. This numerical model is based on a stratigraphy function and the gas content present in the mined formation. This model was based on experimental results.

Karacan et al. (2005) studied an advanced numerical model that can simulate the rock mass and the gas flow responses to longwall mining to predict methane flow and emissions. This research was based on two main phases. The first phase involved using Fast Lagrangian Analysis of Continua 2D (FLAC 2D) to simulate the rock's geomechanical characteristics. The second stage consisted of implementing the Computer Modeling Group's GEM software to simulate methane emissions and gas flow associated with underground longwall coal mining.

Luxbacher et al. (2009) established a numerical model that uses computer applications to explore the effects of porosity and permeability changes of the coal seam on methane emissions in an underground coal mine. In this research, the variations in porosity, permeability, and effective coal stress were included in the model. Furthermore, methane emissions, possible leakage from the stoppings, water inflow, and air requirements were analyzed by employing a coalbed methane reservoir technique.

2.3.3 Time Series Analysis for Methane Forecasting

A time series can be defined as a series of observations or data recorded at regular times. The observations can be captured over an entire interval at fixed time points or randomly sampled points. Operational monitoring, statistical research, event impact analysis, warning, anomaly detection, machine learning, and forecasting are some of the most popular time series applications (Shumway and Stoffer, 2006; Brillinger, 2001). As mentioned before, this research focuses on time series forecasting, where the main idea is to calculate and predict future values of the target variable based on past observations. A short literature review of time series analysis for forecasting and controlling atmospheric monitoring systems and methane concentration in underground coal mines is presented below.

Kaffanke (1980, cited in Dixon, 1992) proposed a medium-term prediction of the methane concentration method based on discrete multiple regression techniques. Coal daily production, total coal production, methane flow, gas content, and non-working days were the main variables selected to run the model for describing methane concentration and emission. Accurate methane forecasting was obtained by this study (Figure 2.1). The author recommended further research employing regression techniques, time series analysis, statistical methods, spectral analysis, and filter theory.

Tructin and Wasilewski (1987) presented a method that applies digital filtering algorithms and time series analysis to separate signals of different amplitude and duration that interfere with a mine ventilation system's operation in longwall coal mines, including monitoring methane gas emissions.

Tauziede and Pokryszka (1993) developed a dynamic statistical method for methane prediction. A general expression that predicts methane emission and concentration in the working faces was obtained after studying coal working faces using statistical analysis and time series techniques, as shown by Equation 2.3.

$$D_n = D_s[306\,A_n + 150\,A_{n-1} + 75\,A_{n-2} + 5{,}470] \qquad (2.3)$$

where D_n = expected volume of methane for week n (m^3), D_s = specific emission of methane (m^3) per meters of advance, A_n = planned advance per week n (m), A_{n-1} and A_{n-2} = real advances for weeks n-1 and n-2 (m).

Dixon (1992) described a model for predicting methane concentration based mainly on time series analysis. Univariate and multivariate time series models were developed using monitored data and the Box-Jenkins method of time series analysis. This technique's implementation identified the relationship between methane concentration and its explanatory variables such as coal production, barometric pressure, and airflow velocity. Finally, the author recommended time series analysis for application to mining process control and forecasting of methane concentration.

Dixon and Longson (1993) established a statistical method for predicting methane concentration in longwall coal mines based on time series analysis developed using data obtained from atmospheric monitoring systems. Figure 2.2 presents hourly forecasts based on this methodology. Methane drainage, barometric pressure, coal production, and air velocity were the main variables studied in this model. This research concluded that the coal production rate is a crucial variable for the prediction of methane gas.

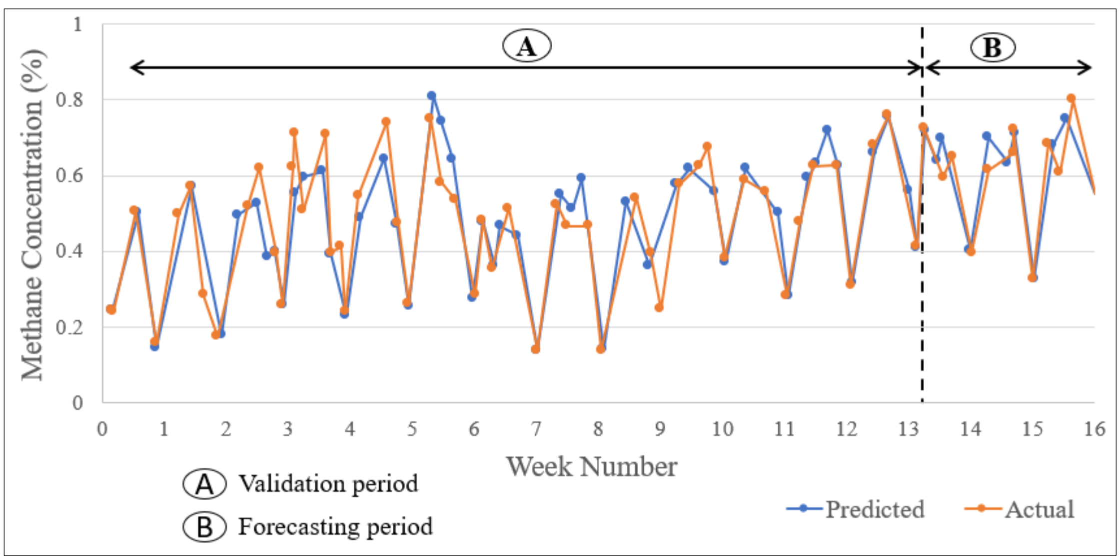

Figure 2.1 Methane forecasting after Kaffanke (1980, cited in Dixon, 1992)

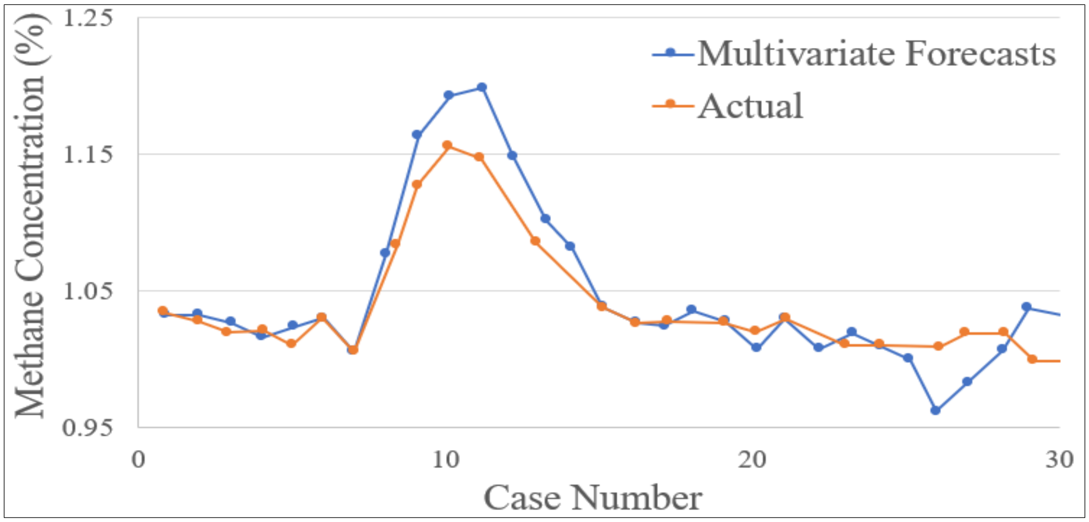

Figure 2.2 Methane multivariate forecasting after (Dixon and Longson, 1993)

Tominaga and Bandopadhyay (2002) designed a monitoring system to forecast spontaneous combustion in underground longwall coal mines. The main objective was to precisely predict the origin of excessive concentration of hazardous gasses in underground coal mine environments by implementing time series analysis and Fick's second law of diffusion.

Zagorecki (2015) presented a numerical method based on the analysis of a data set in multivariate time series to predict the excessive concentration of methane gas at three locations at an underground mine. This method was based on statistical analysis, selection algorithms, correlation, cross-correlation techniques, and the machine learning random forest algorithm implemented in Waikato Environment for Knowledge Analysis (WEKA). This software includes a collection of mathematical models, algorithms, and visualization tools. It is mainly used for data analysis and forecast modeling.

2.4 Comparison and Discussion on Methane Prediction Methods

The literature review presented in this paper has highlighted that empirical and numerical approaches for predicting methane emission and concentration have been studied for a long time by many researchers. Some of these investigations have succeeded in developing models that can forecast methane emissions. Nevertheless, these techniques have some disadvantages that hinder their implementation. For example, empirical methods are time-consuming, expensive, and data collection is challenging. Most importantly, they cannot be broadly implemented because they depend on each particular case's geographical and geological conditions. The main disadvantage of numerical methods is the required amount of previous knowledge of the physical conditions and parameters that influence methane gas behavior in each particular case.

Statistical methane prediction methods (such as time series analysis) are less time-consuming and expensive. Unlike empirical and numerical approaches, statistical methods can be easily generalized and focus on the statistical interpretation of the results rather than on the process that affects methane emissions and concentration. For these reasons, the discussion presented below focuses on implementing statistical methods to forecast methane emissions and concentration in underground coal mines, emphasizing the study of time series analysis for methane forecasting based on monitored data.

2.5 Case Study

Data were collected from an active longwall mine in the eastern USA. Data included methane concentration measurements at the mine's exhaust shafts and daily production values. Data are available from four shafts (A, B, C, D). Barometric pressure data were retrieved from the nearest public weather station. Table 2.1 presents a summary of the available data as well as the measurement frequency for each parameter.

Table 2.1 Description of collected data

Parameter	***Frequency***	***Total number of points***	***Units***
Coal production	Daily	1,737	Tons
Methane concentration	Approx. Hourly	223,857	%
Barometric pressure	Approx. Hourly	57,270	InWG

2.6 Preliminary Results and Discussion

When analyzing time-series data, it is essential to ensure that data points from different series share common date/time stamps. As shown in Table 2.1, the data collection frequency for both methane concentration and barometric pressure is approximately hourly, although data points are not collected at the exact same time within the hour. Hence, the first step was to homogenize the time series data. This was done by applying interpolation on one of the data sets so that timestamps matched those of the second data set. Interpolation was implemented using the interp1 MATLAB® interpolation function.

As methane concentration data span about five years, a smaller set was selected that corresponds to 160 days. Daily data were aggregated from hourly values for both methane concentration and barometric pressure obtained from the nearest weather station (at a distance of about 50 miles) and then plotted together, as shown in Figure 2.3.

The blue line indicates barometric pressure (InWG) (left axis), while the red line represents methane gas concentration (%) collected from shaft A (right axis).

Different correlations and trends can be observed in Figure 2.3. For example, a negative correlation between barometric pressure and methane gas rate can be identified between around days 40, 70, 155, 175, where barometric pressure decreases while methane gas concentrations increase. This relationship is similar to that reported by Agioutantis et al. (2014). On the contrary, barometric pressure increases while methane gas increases between days 90-110. Hence, at different times during these 160 days, time interval different trends are evident.

The trends presented in Figure 2.3 can be attributed to different factors, such as the data collection process, the aggregation process, the potential presence of outliers and extreme values in the dataset (due to calibration practices, power outage, etc.) or the influence of other mine-related parameters such as the ventilation controls, coal production levels, etc.

These contributing elements need to be identified and evaluated in order to construct a model that can accurately predict gas concentration.

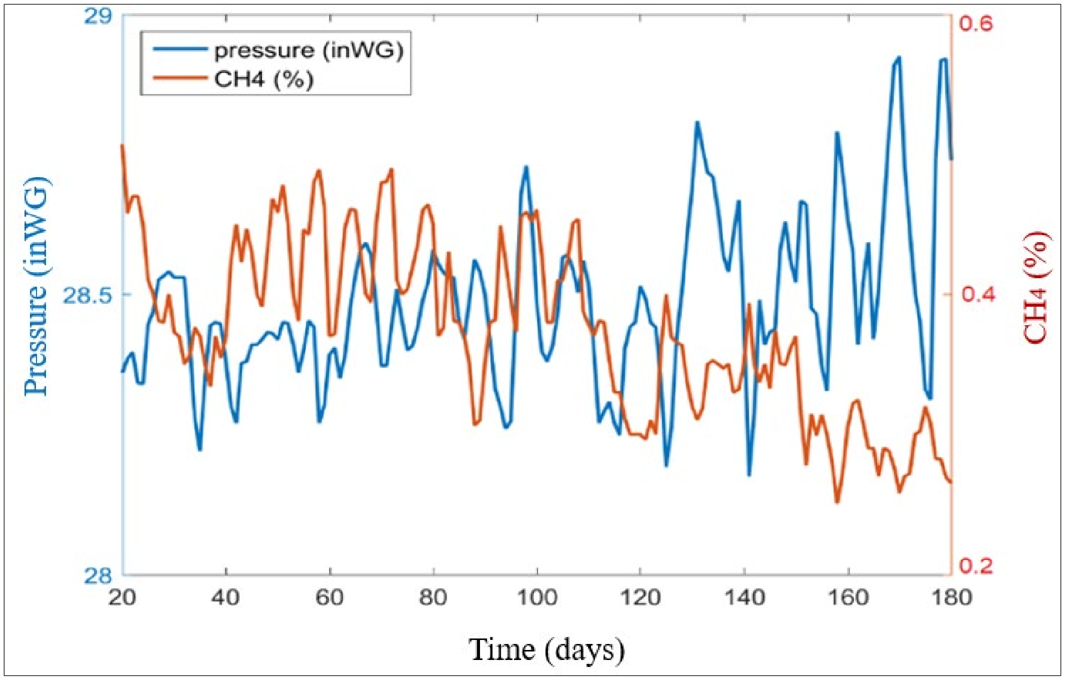

Figure 2.3 Correlation of time series data for Shaft A

2.7 Summary and Conclusions

This paper presents a literature review of the main methods used for methane gas forecasting in underground coal mines. Moreover, it identifies the advantages of implementing statistical methods for the prediction of methane emissions and concentrations. In particular, the flexibility and forecasting ability of time series analysis for methane forecasting based on monitored data should be emphasized. A preliminary analysis of data from a mine in the eastern US is presented. Forecasting mine gas concentrations is a complex problem, and the existing methods rely on simplifying assumptions that may be sufficient under certain conditions and mine settings but not general enough to allow accurate and reliable forecasting. The goal of this research is to further investigate the dependence between gas concentration and other factors and to ultimately develop improved forecasting methods based on time-series analysis.

The next stages of this research are focused on (a) identifying the factors and variables that have an impact on methane concentration, such as daily production and atmospheric pressure, (b) identifying and removing outliers due to sensor calibration and malfunction, (c) estimating cross-correlations to quantify the interdependence of variables (d) investigating different techniques for data reduction, aggregation, and homogenization, and (e) determining optimal time-series predictive models.

2.8 Acknowledgment

This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION). The views, opinions, and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its directors, and staff.

2.9 References

Agioutantis Z, Luxbacher K, Karmis M, Schafrik S (2014) Development of an atmospheric data management system for underground coal mines. J. South. Afr. Inst. Min. Metall, 114:1059–1063.

Agioutantis Z, Luxbacher K, Dougherty M, Karmis M (2015) Develop and implement a data management platform for atmospheric data in underground coal mines. In Proceedings of the 15th North American Mine Ventilation Symposium, 20–25 June 2015, Blacksburg, VA, USA.

Airey EM (1968) Gas emission from broken coal. An experimental and theoretical investigation. Int. J. Rock Mech. Min. Sci. Geomech, 5:475–494.

Boyer CM, Qingzhao B (1998) Methodology of coalbed methane resource assessment. Int. J. Coal Geol, 35:349–368.

Brillinger DR (2001) Time Series: Data analysis and theory. Elsevier, Philadelphia, USA.

Bustin RM, Clarkson CR (1998) Geological controls on coalbed methane reservoir capacity and gas content. Int. J. Coal Geol, 38:3–26.

ByungWang J, Rana K (2018) An internet of things system for underground mine air quality pollutant prediction based on azure machine learning. Sensors. https://doi:10.3390/s18040930.

Creedy DP (1993) Methane emissions from coal related sources in Britain: development of a methodology. Chemosphere, 26:419–439.

Diamond WP, Garcia F, Aul G, Ray R (1997) Analysis and prediction of longwall methane emissions: a case study in the Pocahontas No.3 coalbed, V.A. National Institute of Occupational Safety and Health (NIOSH). https://search.cdc.gov/search. Accessed March 15, 2021.

Dixon DW (1992) A statistical analysis of monitored data for methane prediction. Dissertation, University of Nottingham. Nottingham, United Kingdom.

Dixon DW, Longson I (1993) Statistical method for methane prediction and improved environmental control. In: Proceedings of the 6th North American Mine Ventilation Symposium, pp. 94–98. Salt Lake City, Utah, United States.

Dunmore R (1982) Predicting methane in real-time for longwall faces. In Proceedings of the 1st North American Mine Ventilation Symposium. 29–31 March, Alabama, USA.

Düzgün HS, Leveson N (2018) Analysis of soma mine disaster using causal analysis based on systems theory (CAST). Saf. Sci. https://doi.org/10.1016/j.ssci.2018.07.028.

Ediz IG, Edwards JS (1991) Numerical simulation of time-dependent methane flow. Mining Science and Technology, 12:1–15.

Karacan CÖ (2008) Modeling and prediction of ventilation methane emissions of U.S. longwall mines using supervised artificial neural networks. Int. J. Coal Geol, 73:371–387.

Karacan CÖ, Diamond WP, Esterhuizen GS, Schatzel SJ (2005) Numerical analysis of the impact of longwall panel width on methane emissions and performance of gob gas ventholes. National Institute for Occupational Safety and Health (NIOSH). https://search.cdc.gov/search. Accessed October 01, 2021.

Karacan CÖ, Ruiz F, Cote M, Phipps S (2011) Coal mine methane: A review of capture and utilization practices with benefits to mining safety and greenhouse gas reduction. Int. J. Coal Geol, 86:121–156.

Kirchgessner DA, Piccot SD, Winkler JD (1993) The estimate of global methane emissions from coal mines. Chemosphere, 26:453-472.

Kissell FN (2006) Handbook for Methane Control in Mining; National Institute for Occupational Safety and Health (NIOSH), Cincinnati, OH, USA.

Lunarzewski LW (1998) Gas emission prediction and recovery in underground coal mines. Int. J. Coal Geol, 35:117–145.

Luxbacher KD, Erdogan SS, Karacan CÖ (2009) Modeling methane emissions and ventilation needs by examining mining-induced permeability changes and related damage to ventilation controls. In Proceedings of the 43rd U.S. Rock Mechanics Symposium and 4th U.S.-Canada Rock Mechanics Symposium. June 28 - July 01, Ashville, USA.

Owili-Eger A, Stefanko S, Ramani RV (1973) Simulation of quantity and quality control in mine ventilation. Special Research Report SR-95. The Pennsylvania State University. Pennsylvania, United States.

Shumway RH, Stoffer DS (2006) Time series analysis and its applications. With R Examples. Springer, New York City, USA.

Sung W, Ertekin T, Ramani RV (1987) Evaluation of the impact of the degasification process on the inflow of methane gas into coal mines: A numerical exercise. In: Proceedings of the Society for Mining, Metallurgy & Exploration (SME), pp. 328–333. The United States of America.

Tauziede C, Pokryszka Z (1993) Dynamic prediction of CH_4 emission in longwalls. In Proceedings of the Conférence Internationale des Instituts de Recherches sur la Sécurité dans les Mines Pretoria, South Africa.

Tominaga Y, Bandopadhyay S (2002) Monitoring of spontaneous combustion based on time series data. In Proceedings of the 9th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781439833742.ch48.

Trutwin W, Wasilewski S (1987) Application of digital filter and time series analysis in ventilation control systems. In: Proceedings of the 5th North American Mine Ventilation Symposium (NAMVS), pp. 534–540. Morgantown, West Virginia, United States.

Zagorecki A (2015) Application of sensor fusion and data mining for prediction of methane concentration in coal mines. Mining-Informatics, Automation, and Electrical Engineering ed. 53:33–38.

3. TIME SERIES DATA MANAGEMENT AND ANALYSIS

The following two manuscripts describe and discuss data collection, storage, pre-processing, and processing steps implemented in this research. The first article is a peer-reviewed published in the International Conference on Raw Materials and Circular Economy in September 2021, and the second paper was published and presented at the Society of Mining, Metallurgy, and Exploration symposium in February 2022.

3.1 MANAGING AND UTILIZING BIG DATA IN ATMOSPHERIC MONITORING SYSTEMS FOR UNDERGROUND COAL MINES

Juan Diaz, University of Kentucky; Lexington, Kentucky; United States of America

Zach Agioutantis, University of Kentucky; Lexington, Kentucky; United States of America

Steven Schafrik, University of Kentucky; Lexington, Kentucky; United States of America

Dionissios Hristopulos, Technical University of Crete; Chania, Crete; Greece

Full Citation:

Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. Mater. Proc. https://doi.org/10.3390/materproc2021005078.

3.1.1 Abstract

Underground coal mining Atmospheric Monitoring Systems (AMS) have been implemented for real-time or near real-time monitoring and evaluation of the mine atmosphere and related parameters such as gas concentration (e.g., CH_4, CO, O_2), fan performance (e.g., power, speed), barometric pressure, ambient temperature, humidity, etc. Depending on the sampling frequency, AMS can collect and manage a tremendous amount of data, which mine operators typically consult for everyday operations as well as long-term planning and more effective management of ventilation systems. The raw data collected by AMS need considerable pre-processing and filtering before they can be used for analysis. This paper discusses different challenges related to filtering raw AMS data in order to identify and remove values due to sensor breakdowns, and sensor calibration periods, transient values due to operational considerations, etc., as well as to homogenize time series for different variables. The statistical challenges involve the removal of faulty values and outliers (due to systematic problems) and transient effects, gap-filling (by means of interpolation methods), and homogenization (setting a common time reference and time step) of the respective time series. The objective is to derive representative and synchronous time series values that can subsequently be used to estimate summary statistics of AMS and to infer correlations or nonlinear dependence between different data streams. Identification and modeling of statistical dependencies can be further exploited to develop predictive equations based on time series models.

Keywords: big data; atmospheric monitoring systems; time series analysis

3.1.2 Introduction

In recent years, the monitoring of atmospheric conditions in underground coal mines has considerably improved due to technological advances in different areas such as machine learning, the internet of things (IoT), electronics, and data transmission, among others. As a result, Atmospheric Monitoring Systems (AMSs) used in underground coal mines operations can collect and store big datasets; large mine operations can generate more than 100 GB of data annually (Agioutantis et al., 2014). Mine operators typically utilize these data to manage everyday operations and for more effective and efficient management of ventilation systems as well as for long-term planning (Agioutantis et al., 2015; Diaz et al., 2021). In addition, the raw data regarding mine atmosphere and related parameters collected by AMSs can be used to develop forecasting models for toxic and explosive gases in underground coal mines based on time series models. However, prior to conducting any analysis, the collected data need to be pre-processed (i.e., cleaned, filtered, and homogenized).

Real-world databases are generally inconsistent or incomplete. As a result, information may be missing, and/or existing data may not be accurate, which directly affects the outcomes of forecasting models, including time series models (Verma, 2021). Data pre-processing is a crucial technique that deals with such issues. Data pre-processing, also known as data cleaning, can be defined as a set of operations to detect and remove erroneous values and outliers, determine missing values, smooth noisy data, and adjust time series inconsistencies (Baur et al., 2015). In other words, data cleaning is the process of transforming raw data collected and stored from different data collection systems (DCSs), in this particular case, atmospheric monitoring systems in underground coal mines, into datasets that can be used for planning, modeling, visualization, and decision-making (Buttrey and Lyn, 2017; Ranjan et al., 2021).

There are many challenges concerning datasets collected and stored by AMSs in underground coal mines, such as data gaps due to sensor malfunction or calibration. In addition, identifying and removing faulty values including outliers (due to systematic problems) is another critical matter that compromises measurement accuracy. Moreover, outlier detection helps identify malfunctioning sensors or unusual events (Hongzhi et al., 2019). Finally, data homogenization is challenging when dealing with time series data, as it is crucial to guarantee that data points from different series share common date/time stamps. Figure 3.1.1 schematically presents a typical sequential procedure of data collection and pre-preprocessing (Agioutantis et al., 2014).

The research presented in this paper discusses all the challenges related to AMSs data collection and storage in the process of developing representative and synchronous time series values that can be used to estimate summary statistics of AMSs and identify correlations (linear or nonlinear) between different data streams with the ultimate goal of developing forecasting models based on time series analysis.

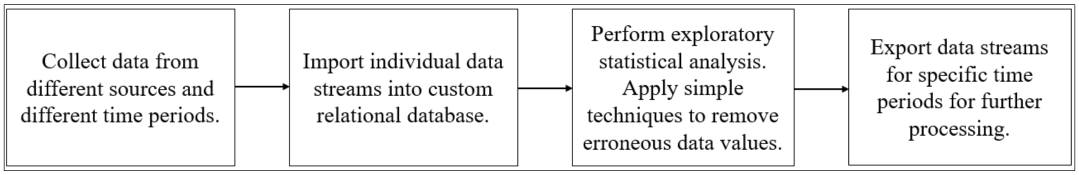

Figure 3.1.1 Schematic of data collection and pre-processing sequence

The research presented in this paper discusses all the challenges related to AMS data collection and storage in the process of developing representative and synchro-nous time series values that can be used to estimate summary statistics of AMS and identify correlations (linear or nonlinear) between different data streams with the ultimate goal of developing forecasting models based on time series analysis.

3.1.3 Methane Gas Generation in Underground Coal Mines

Coalbed methane (CBM), also known as coal mine methane (CMM), is methane gas present in underground coal seams. It is produced due to the geological process of coalification, i.e., the decomposition of organic matter into coal (US EPA, 2021). CBM is mainly composed of methane gas (CH_4), carbon dioxide (CO_2), nitrogen (N_2), butanes (C_4H_{10}), propane (C_3H_8), and ethane (C_2H_6) (Thakur et al., 2015), with methane being the principal component as it covers approximately 80–90% of the total gas volume (Mahdevari, 2019). CMM refers mainly to methane released during coal mining activities when the coal seam is fractured. Thus, CMM and CBM can be defined as subsects of the methane gas found in coal seams. However, CMM refers exclusively to the methane gas present and released from mined coal seams, while CBM reflects the methane gas present in unminable coal seams (US EPA, 2021; Thakur et al., 2015).

The amount of CMM generated at a specific underground coal mine operation depends basically on three main parameters; operational variables such as the mining method and productivity of the coal mine, the gassiness of the coal seam, and its geological conditions (e.g., coal rank, coal seam fractures, and coal seam depth) (Mahdevari, 2019; Irving and Tailakov, 1996). Coal extraction releases more methane than was initially confined within the mined coal seam itself due to the fractures developed in the surrounding strata and the pressure drop caused during mining, which draws additional gas from the adjacent strata. That is particularly true in longwall mining which can reach high production rates, e.g., 20,000 and 30,000 tons of coal daily (Mahdevari, 2019; Karacan et al., 2011; Bise, 2013).

The large amount of methane released in underground coal mines is an essential concern for mine ventilation management in order to ensure worker health and safety as CH_4 is highly explosive for concentrations ranging from 5% to 15% (Kissell, 2006). Excessive methane gas concentration accounts for more than 80% and almost 90% of the accidents and fatalities in the underground coal mining industry in the United States, respectively (NIOSH, 2020). Dealing with coal mine methane involves understanding and identifying the critical parameters that influence methane emission and hazardous accumulations such as coal production, atmospheric pressure, humidity, air velocity, and fan performance (Agioutantis et al., 2014; 2015; Diaz et al., 2021).

3.1.4 Coal Mine Methane Forecasting Methods

Methane gas forecast techniques can be categorized mainly into three groups depending on the scientific method implemented. The first group consists of methods that are based on empirical data and expertise to forecast methane gas emissions. The second group includes the numerical forecasting methods; it comprises methodologies that solve physical models represented mathematically by a set of equations. Finally, the third approach is based on collecting, pre-processing, processing, and analyzing raw data using statistical techniques, also known as the statistical approach (Diaz et al., 2021; Kissel, 2016; Dixon, 1992).

Empirical and numerical methane forecasting methods are expensive, time-consuming, and limited to a specific location and mine. In addition, they are directly affected by the combination of natural and technical factors that influence methane gas behavior in each particular case, such as geological conditions of the surrounding rock and mined coal seams (e.g., fractures, amount of methane concentrated and emitted), technical specifications of the mining process (e.g., mining method, advance rate, mine depth, ventilation system).

In contrast, unlike empirical and numerical methods, statistical forecasting techniques (such as time series models) are less expensive and time-consuming. Most importantly, they can be easily generalized because they are based on analyzing and interpreting historical data rather than the physical processes and the relevant factors (mine operation, geological setting) that influence methane emissions (Diaz et al., 2021).

3.1.5 Case Study

The atmospheric monitoring data were collected from an active underground longwall coal mine located in the eastern US, renamed Mine A, due to confidentiality reasons. The dataset consists of methane concentration measurements collected by sensors at a number of exhaust shafts as well as daily production values (tons/day) that could be attributed to each shaft. Data records are available for several years. The work presented in this study is based on data from two shafts only (B and D). In addition, a python routine was developed to download meteorological data (i.e., barometric pressure, temperature, humidity, among others) from a public weather station in proximity to Mine A (Weather Underground (WU)) that provides real-time weather information online. Barometric pressure (BP) data were also provided by sensors available at the mine location. Table 3.1.1 summarizes the data used for this research and the measurement frequency for each parameter.

Table 3.1.1 Description of collected data – Shafts B and D

Source	***Parameter***	***Approx. Frequency***	***Approx. data points***	***Units***
Shaft B	Coal production	Daily	1,800	Tons
	Methane concentration	Hourly	50,000	%
Shaft D	Coal production	Daily	1,800	Tons
	Methane concentration	Hourly	45,000	%
Nearest public weather station	Barometric pressure, temperature, etc.	Hourly	81,000	InWG
Mine weather station	Barometric pressure	Every 10 s	3 mil/year	InWG

Mine A and WU data were populated into a custom relational database known as AMANDA, which stands for Atmospheric Monitoring Analysis and Database mAnagement. AMANDA has been designed explicitly for AMSs data. It has many subsystems such as data analysis, validation and storage, data reporting, and visualization. For example, the two plots in Figure 3.1.2 illustrate a visualization of BP for a five-day interval and the corresponding variation of CH_4 emissions for the same five-day interval. In addition, the AMANDA system can run several basic statistics on the imported data. These statistics allow the user to check for negative values or any obvious outlier values in the data streams. AMANDA can accommodate multiple projects as well as multiple data streams per project (Agioutantis et al., 2014).

Once data were populated in the database, a number of tools were employed to disable (not delete) evident erroneous values (such as negative methane concentration measurements or methane measurements that correspond to either faulty sensors or calibration periods). The "cleaned" data streams were then exported for specific time periods.

These data streams were then imported into the MATLAB environment for further data pre-processing. Finally, different MATLAB commands (e.g., interp1 performing interpolation) were used to transform the methane gas and BP data streams into a common time stamp.

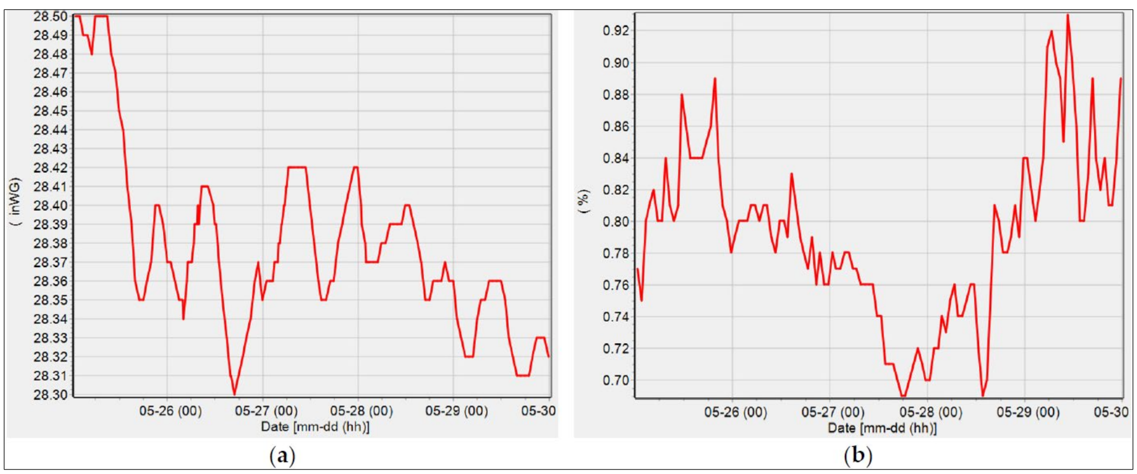

Figure 3.1.2 Data visualization in AMANDA: (a) Barometric pressure for a five-day interval and (b) methane gas concentration for a five-day interval

3.1.6 Results and Discussion

The BP values collected at the public weather station were compared to the values collected at the mine (Figure 3.1.3). The correlation between the two data streams was very high. For example, for the 30-day interval shown in Figure 3.1.3, the correlation was R=0.99. Thus, it was decided to use the regional BP instead of BP mine data since the former data stream was of a higher quality (i.e., fewer missing points).

Figure 3.1.4a presents a superposition of methane gas concentration, coal production, and BP for 180 days for shaft B. Figure 3.1.4b presents a similar plot but only for methane gas concentration and BP for a different 180-day interval for the same shaft. The latter figure indicates that there are still methane emissions even without production due to already mined-out areas and/or exposed pillar ribs of all the development entries and crosscuts.

Initially, the analysis concentrated on time periods where production was zero to analyze the effect of BP on methane emissions and develop a baseline. Then, several such time periods were identified, and the respective data streams were analyzed separately. Figure 3.1.5 includes four plots (a to d) that correspond to a segment of 250 days of data collected from shaft B. Figure 3.1.5a shows a plot of raw data for methane gas concentration and BP. Figure 3.1.5b illustrates the application of the interpolation function so that the two data streams acquire a common reference timestamp. Each of the two plots in Figure 3.1.5b shows the raw data and the interpolated data. Figure 3.1.5c shows the daily median (of the data shown in Figure 3.1.5a) as a representative value for each data stream. The scatter plot in Figure 3.1.5d exhibits a strong correlation between methane gas concentration and BP with a linear correlation coefficient R=-0.77. The negative sign of the correlation indicates that when BP drops, methane concentration tends to increase.

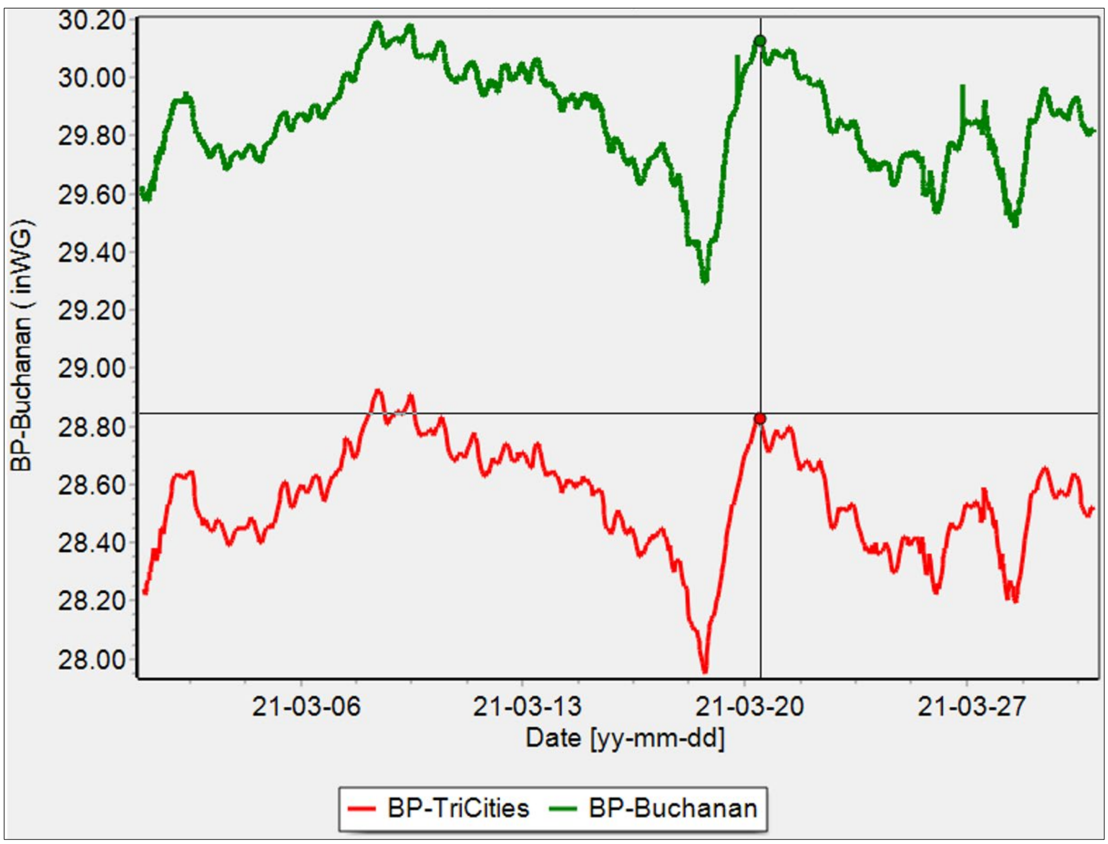

Figure 3.1.3 Superposition of BP from the mine and regional stations for a period of 30 days

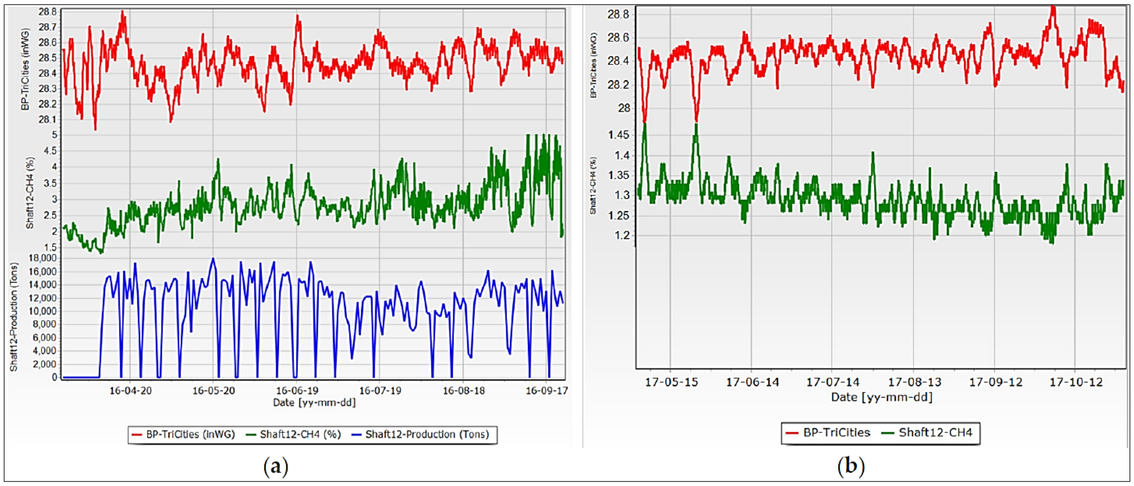

Figure 3.1.4 Data visualization in AMANDA: (a) Barometric pressure, methane concentration, and production for 180 days (b) Barometric pressure and methane concentration for 180 days

Figure 3.1.6 shows the plot of the cross-correlation function between the barometric pressure and the methane gas concentration for this segment for various time lags (shown along the horizontal axis). The cross-correlation at zero lag is ~ -0.77, in agreement with the value shown in the scatter plot of Figure 3.1.5d. Note that negative values of the cross-correlation between the two variables persist even for lags of several days. The presence of such correlations signals that the barometric pressure could be used to forecast the methane gas concentration (provided that the existence of such correlations is systematically observed between the two variables).

Figure 3.1.7 includes four plots (a-d) corresponding to a segment of 180 days of collected data from Shaft B. This is a different time segment than the one analyzed in Figure 3.1.5. Figure 3.1.7a shows a plot of the raw data for methane gas concentration and BP. Figure 3.1.7b illustrates the application of the interpolation function so that the two data streams acquire a common reference timestamp. Finally, figure 3.1.7c shows a daily median (of the data shown in Figure 3.1.7a) as a representative value for each data stream. Note that in this case, the scatter plot in Figure 3.1.7d presents a poor correlation between methane gas concentration and BP with R=0.24.

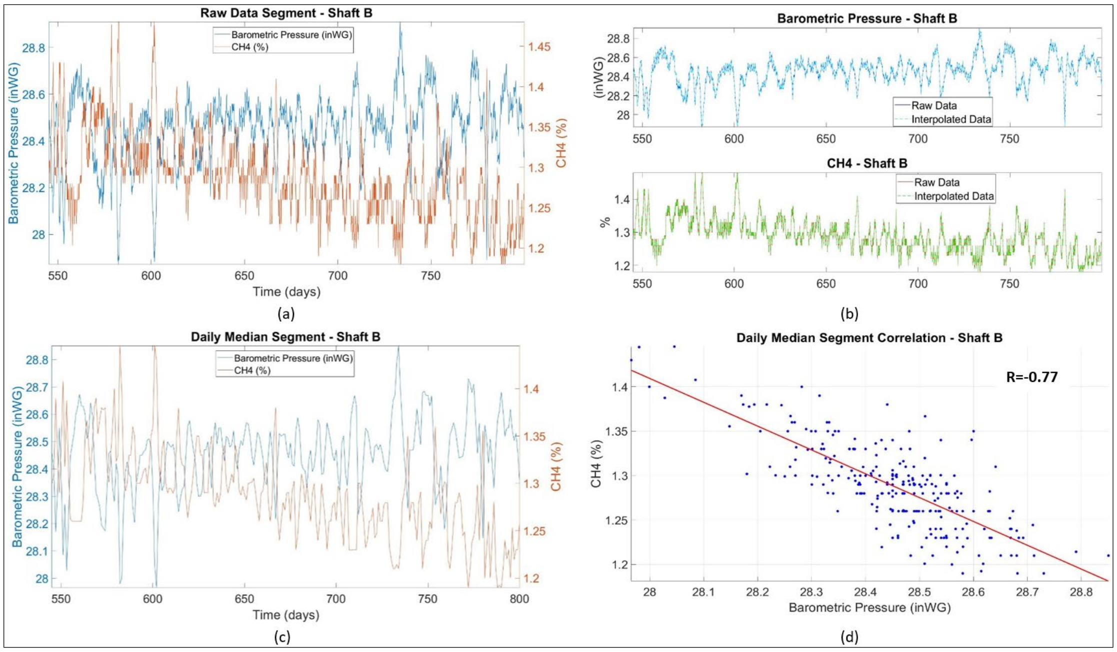

Figure 3.1.5 (a) Raw data, (b) raw and interpolated data, (c) daily median data, (d) correlation of daily median data for a specific data segment

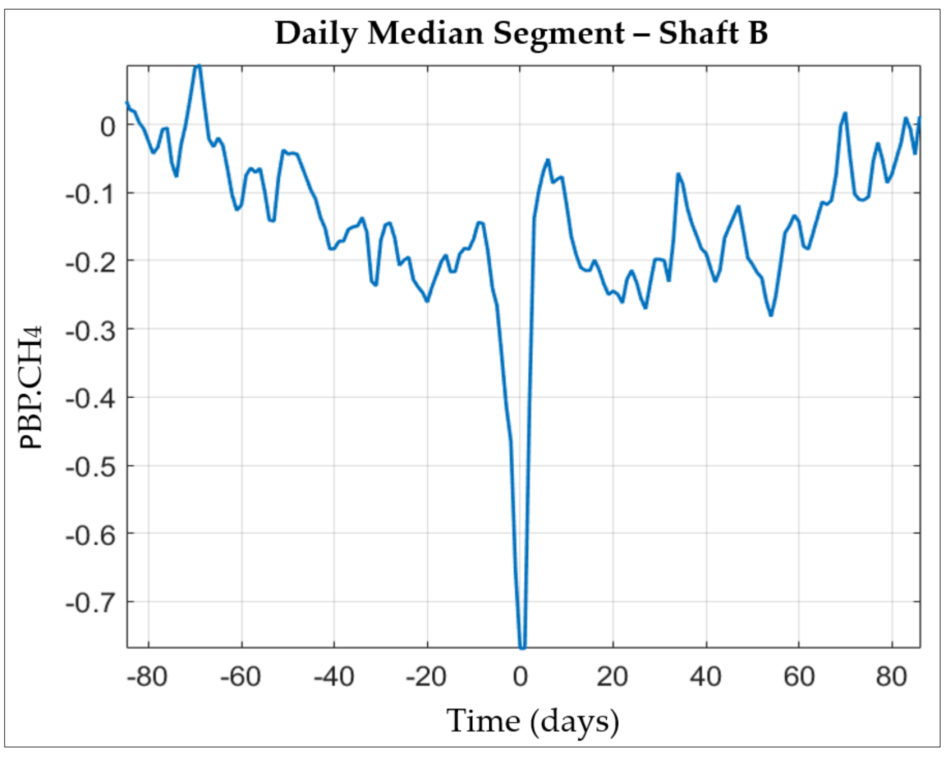

Figure 3.1.6 The plot of the cross-correlation function between the barometric pressure and methane gas concentration. The horizontal axis represents the time lag between the two variables, and the vertical axis represents the cross-correlation values

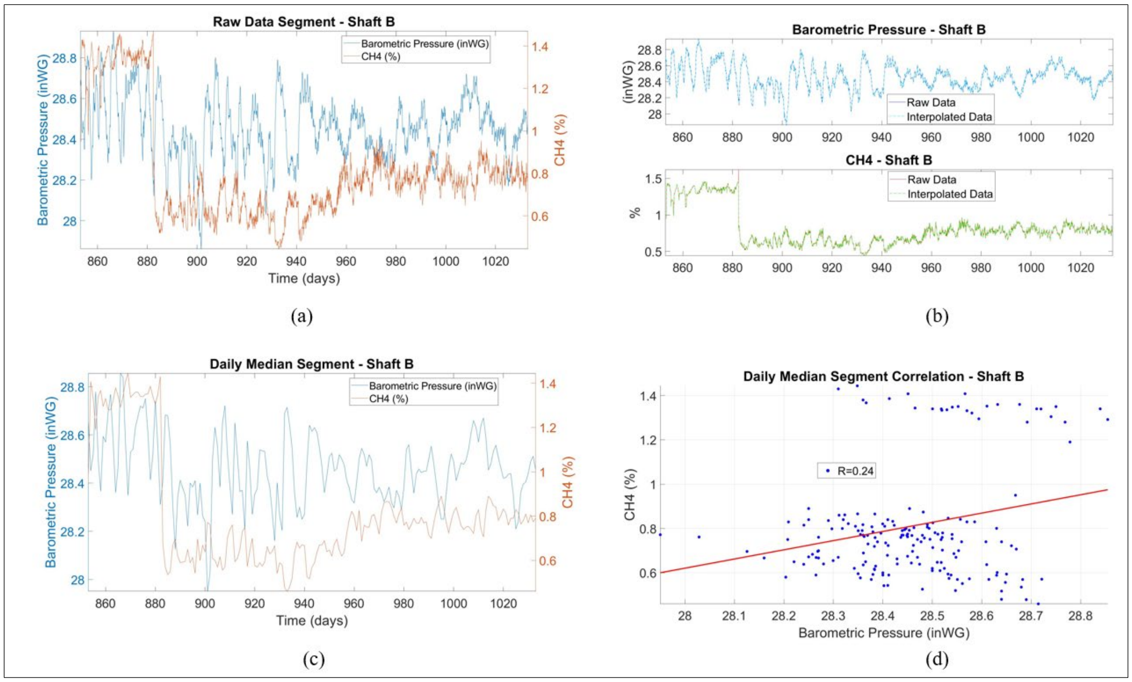

Figure 3.1.7 (a) Raw data, (b) raw and interpolated data, (c) daily median data, (d) correlation of daily median data for a specific data segment

3.1.7 Summary and Conclusions

This paper discusses challenges related to Atmospheric Monitoring Systems data collection and storage facing the development of representative and synchronous time series models. The latter could be used to estimate summary statistics of AMSs and identify correlations (linear or nonlinear) between different data streams with the ultimate goal of developing methane gas forecasting models based on time series analysis. The development and deployment of such tools can help to improve mine health and safety.

Preliminary results and analysis of the time series data collected over five years from two shafts (B and D) of an active longwall mine in the eastern US are presented. A notable negative correlation between barometric pressure and methane gas concentration has been identified. The presence of such correlations suggests that the barometric pressure could be used to forecast the methane gas concentration in underground coal mines. Furthermore, this research has shown that methane gas concentration does not depend exclusively on barometric pressure. Therefore, more advanced multivariate analysis methods should also be explored to determine potential factors and variables that are correlated (with negative or positive coefficients) with methane concentration, such as daily production and changes in mine operations.

The focus of future work will be to (a) collect data and develop import routines from the atmospheric monitoring system and meteorological stations as well as filter and homogenize such data from a second case study (Mine B), (b) develop and validate long-term relationship(s) between meteorological parameters and methane gas concentration (c) investigate quantitative measures of statistical dependence between methane concentration and mine operations.

Author Contributions: Supervision, Z.A.; Writing-original draft, J.D.; Writing-review and editing, Z.A., D.H, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION). The views, opinions, and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its directors, and staff.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable as data is proprietary.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

3.1.8 References

Agioutantis Z, Luxbacher K, Karmis M, Schafrik S (2014) Development of an atmospheric data management system for underground coal mines. J. South. Afr. Inst. Min. Metall, vol 114:1059–1063.

Agioutantis Z, Luxbacher K, Dougherty M, Karmis M (2015) Develop and implement a data management platform for atmospheric data in underground coal mines. In Proceedings of the 15th North American Mine Ventilation Symposium, 20–25 June 2015, Blacksburg, VA, USA.

Baur J, Moreno-Villanueva M, Kötter T, Sindlinger T, Bürkle A, Berthold N, Junk M (2015) MARK-AGE data management: Cleaning, exploration, and visualization of data. Mech. Ageing Dev. https://doi.org/10.1016/j.mad.2015.05.007.

Bise CJ (2013) Modern American coal mining methods and applications. (ed) Christopher JB Ed.; Society for Mining, Metallurgy & Exploration (SME), Englewood, CO, USA.

Buttrey SE, Lyn RW (2017) A Data Scientist's Guide to Acquiring, Cleaning, and Managing Data in R. JohnWiley & Sons, Incorporated. Hoboken, NJ, USA.

Byrer C, Havryluk I, Uhrin D (2014) Coalbed methane: a miner's curse and a valuable resource. In: Pramod T, Schatzel S, Aminian K (ed) Coalbed methane: from prospect to pipeline, 1st ed. Elsevier, Amsterdam, pp 1–6.

Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Towards atmospheric monitoring data analysis in underground coal mines. In: Proceedings of the 18th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781003188476-51.

Dixon DW (1992) A statistical analysis of monitored data for methane prediction. Dissertation, University of Nottingham. Nottingham, United Kingdom.

Hongzhi W, Bah M, Hammad M (2019) Progress in outlier detection techniques: a survey. Inst. Electr. Electron. Eng. (IEEE), 7:107964–108000.

Irving W, Tailakov O (1996) CH_4 Emissions: coal mining and handling. Good practice guidance and uncertainty management in national greenhouse gas inventories. https://www.ipcc-nggip.iges.or.jp/public/gp/english/. Accessed November 11, 2021.

Karacan CÖ, Ruiz F, Cote M, Phipps S (2011) Coal mine methane: A review of capture and utilization practices with benefits to mining safety and greenhouse gas reduction. Int. J. Coal Geol, 86:121–156.

Kissell FN (2006) Handbook for Methane Control in Mining; National Institute for Occupational Safety and Health (NIOSH), Cincinnati, OH, USA.

Mahdevari S (2019) Coal Mine Methane: control, utilization, and abatement. In Advances in productive, safe, and responsible coal mining. Elsevier. Amsterdam, The Netherlands.

Ranjan K, Gaurav B, Rajanarayan P, Debashisha J (2021) Review of preprocessing methods for univariate volatile time series in power system applications. Electr. Power Syst. Res, https://doi.org/10.1016/j.epsr.2020.106885.

The National Institute for Occupational Safety and Health (NIOSH) (2020) NIOSH data and statistics. https://www.cdc.gov/niosh/injury/data.html. Accessed March 28, 2021.

US EPA. Coalbed Methane Outreach Program. The United States Environmental Protection Agency (US EPA) (2020). https://www.epa.gov/cmop. Accessed November 11, 2021.

Verma U (2019) Data cleaning and pre-processing. Analytics Vidhya. https://medium.com/analytics-vidhya/data-cleaning-and-preprocessing-a4b751f4066f. Accessed November 02, 2021.

3.2 INVESTIGATING RELATIONSHIPS BETWEEN METHANE EMISSIONS AND ATMOSPHERIC DATA IN UNDERGROUND COAL MINES TO DEVELOP A FORECASTING MODEL

Juan Diaz, University of Kentucky; Lexington, Kentucky; United States of America

Zach Agioutantis, University of Kentucky; Lexington, Kentucky; United States of America

Steven Schafrik, University of Kentucky; Lexington, Kentucky; United States of America

Dionissios Hristopulos, Technical University of Crete; Chania, Crete; Greece

Kray Luxbacher, Virginia Polytechnic Institute and State University; Blacksburg, Virginia; United States of America

Full Citation:

Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT, Luxbacher K (2022) Investigating relationships between methane emissions and atmospheric data in underground coal mines to develop a forecasting model. SME Annual Meeting, Feb. 27 - Mar. 02, 2022, Salt Lake City, UT. Preprint 22-025.

3.2.1 Abstract

Big data is generated from both surface and underground mining operations. Such data contain a wealth of information concerning safety and health in the workplace and production parameters. This paper will discuss the progress towards developing an accurate forecasting model for methane gas concentration based on the analysis of data collected from Atmospheric Monitoring Systems in underground coal mines employing time series models. Several procedures need to be applied to raw data, such as data cleaning and filtering for outlier removal, data homogenization and segmentation, and exploratory statistical analysis. The data analyzed were retrieved from two coal operations in the Eastern US. A negative correlation between barometric pressure and methane gas concentration was established, at least for certain data segments. Such correlations raise the possibility that barometric pressure data can predict variations of gas concentration in the mine. The datasets will be further investigated to establish the robustness of barometric pressure and gas concentration correlation and explore the dependence of gas concentration on other factors related to mine design and operations.

Keywords: Big data, methane forecasting, and Atmospheric Monitoring Systems.

3.2.2 Introduction

The atmosphere of underground coal mines typically contains toxic and flammable gases, such as carbon dioxide, carbon monoxide, sulfur dioxide, methane, and hydrogen sulfide. Some of these gases are emitted from cracks in the coal seams due to either natural phenomena or human activity (e.g., methane gas), and others are produced due to spontaneous combustion or coal fire (e.g., carbon monoxide) (Chaulya and Prasad 2016; Byrer et al., 2014). Consequently, the study and monitoring of atmospheric variables in underground coal mine operations are crucial to protecting the health and safety of the miners. Furthermore, such advances can help mine operators to manage more effectively most of the parameters directly related to the ventilation systems (Agioutantis et al., 2014; 2015).

Monitoring atmospheric conditions in underground coal mines has been an essential concern for researchers and the mining industry for a long time. The use of flame safety lamps for detecting methane gas or canaries as warning systems to alert miners when the atmosphere becomes dangerous due to excessive accumulation of toxic and explosive gases are excellent examples (Barry 2013; Taylor and Karacan 2010). However, the technological advances in different fields such as transmission systems, electronics, and computer science have modernized mine monitoring methods leading to accurate and reliable monitoring. Indeed, real-time monitoring techniques are accessible and straightforward to operate; it is easy for mine operators to monitor all ventilation parameters in underground coal mines, such as temperature, barometric pressure (BP), airflow, air velocity, fan performances, and gas concentration (e.g., methane gas, carbon monoxide, sulfur dioxide, and oxygen) (Griffin et al., 2011).

Automated Atmospheric Monitoring Systems (AMS) installed in underground coal mines can collect and store an immense amount of data. Sophisticated mining operations include hundreds or even thousands of sensors that continuously collect data from the mining environment and mining equipment. For example, one large operation could contain more than 700 sensors for monitoring mining operations (e.g., drilling, blasting, hauling, and loading) and processing plants. These sensors could generate almost 300 MB of daily and 100 GB of yearly data (Agioutantis et al., 2014; Diaz et al., 2021).

In most cases, raw data need to be processed before they can be used to generate meaningful input for operators. This may include the following: (a) cleaning and filtering (including outlier identification and removal where appropriate), (b) data homogenization (data timestamps should be homogenized across different datasets), (c) segmentation aiming to determine stationary segments without discontinuities due to calibration or changes in mining operations and (d) exploratory statistical analysis (Diaz et al., 2021).

The research presented in this paper discusses the progress toward developing an accurate forecasting model for methane gas concentration and emissions based on data processing and statistical analysis employing a time series methodology. Raw data collected from Atmospheric Monitoring Systems in underground coal mines in the Eastern US were used.

3.2.3 Case Study

The mine data (i.e., methane gas, oxygen, and carbon monoxide) collected and implemented for this research were retrieved from two coal mines located in the US, henceforth referred to as Mine A and Mine B. In addition, the atmospheric data (i.e., temperature, humidity, precipitation, and barometric pressure) were collected from two public weather stations in the vicinity of Mine A and Mine B, respectively. Table 3.2.1 shows a summary of the data that was used in this research.

Methane data from Mines A and B were provided by the mine personnel. Furthermore, a python routine was developed to download meteorological data rapidly and efficiently from the two closest public weather stations to each mine. Then, atmospheric and mine data for both case studies were populated into a custom relational database identified as AMANDA (Atmospheric Monitoring Analysis and Database mAnagement), designed explicitly for AMS data management. Preliminary processing, including data cleaning and calculation of correlation coefficients between data streams, is accomplished in AMANDA; however, data can also be exported for further processing in different programming environments.

AMANDA employs several tools and functions to identify and filter outliers, missing data, and zero data values due to calibration or defective sensors. Additionally, the Pearson Correlation Coefficient between two different time series can be determined, provided that the time series are sampled at the same time instants. Thus, for example, the correlation between methane gas and barometric pressure and/or coal production can be established on a daily and weekly basis.

The "filtered or cleaned" data streams can be exported for segments with specific time duration (e.g., five days, one month, and six months). These pre-processed data streams were then imported into the MATLAB programming environment for further data processing. In this stage, several commands (e.g., interp1, corrcoef, and scatter) were utilized to transform the barometric pressure and CH_4 data streams to a common time stamp in order to determine and visualize potential correlations. Figure 3.2.1 presents a flowchart that illustrates the data collection, storage, pre-processing, and processing steps implemented in this research.

Table 3.2.1 Summary of collected data sets– Mines A and B

Source	***Parameter***	***Approx. Frequency***	***Approx. data points***	***Units***
Mine A (Shaft B)	Coal production	Daily	2,100	Tons
	CH_4 concentration	Hourly	52,000	%
Mine B	CH_4 concentration	10 sec	3 mil/year	%
Nearest public weather station	BP-Mine A	Hourly	62,000	InWG
	BP-Mine B	Hourly	90,000	InWG

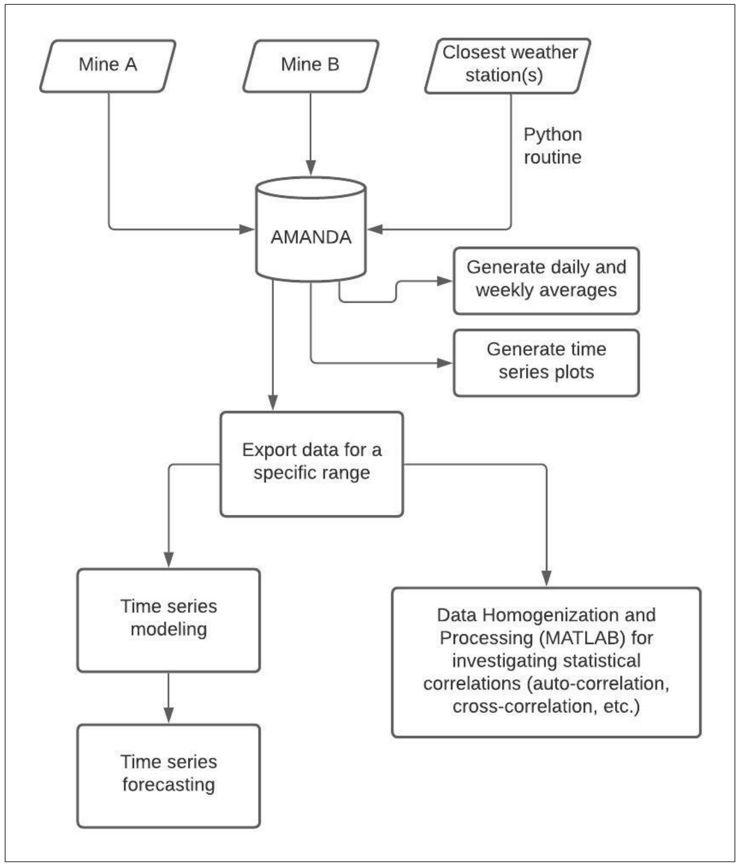

Figure 3.2.1 Schematic representation of data collection, pre-processing, and processing

3.2.4 Results and Discussion

Figure 3.2.2 includes four plots (a to d) which illustrate the techniques implemented for filtering and cleaning the gas concentration data collected from mines A and B. Figure 3.2.2a shows raw methane gas time series from Mine A for a ten-day period with a time step of about hourly. The data include sudden drops in methane gas concentration (CH_4 reaches 0%) due to sensor malfunction. Figure 3.2.2b displays the same time segment from Mine A after the zero values have been filtered out. Figure 3.2.2c presents a methane gas time series from Mine B for a one-day time window with a time step of about 10 seconds. Again, the data contain a few negative values due to sensor malfunction. Finally, Figure 3.2.2d shows the same methane gas data from Mine B after removing negative values. In general, filtering of methane gas data involves removing three classes of values: (a) negative values as methane concentrations can never be negative, (b) zero values at bleeder shafts where methane concentration values, even if small, are always greater than zero, and (c) values that are recorded during sensor calibration (typically 2% or 2.5%) that appear as single value spikes in the time series. In addition, barometric pressure data and gas concentration are typically recorded with different time steps. Hence, an interpolation procedure is applied to both time series (using the interp1 function in MATLAB) in order to generate time series with a common time axis. The method of shape-preserving piecewise cubic interpolation (option 'pchip' in interp1) was selected. The interpolated series are in excellent agreement with the original data for both variables.

Figure 3.2.3 establishes the presence of negative correlations between the barometric pressure and the methane gas concentration measurements. Figure 3.2.3a presents two time series plots covering a period of 180 days. The time series involve daily averages of methane concentrations and barometric pressure related to Mine A. Visual inspection of Figure 3.2.3a reveals a negative correlation between methane gas (red line) and barometric pressure (green line). The visual perception is confirmed by calculating the Pearson coefficient, which takes the value R=-0.83. The negative sign signifies that when BP drops, methane concentration tends to increase. Figure 3.2.3b shows the time series plots for weekly averages of the same data. Each day of the week is assigned a standard number (1-53). Thus, each week is represented by a single day irrespective of the number of data points in that week. The weekly data also display negative correlations with the Pearson correlation coefficient calculated at R=-0.84. Note that when data are filtered out (e.g., Figure 3.2.2), the respective values are not replaced with zero or null values; hence the corresponding time points are not in any way involved in the daily or weekly averages.

Next, the correlations between barometric pressure and gas concentration are investigated using median daily values. Calculating the median tends to reduce random fluctuations, which helps in identifying statistical patterns. Figure 3.2.4 includes four plots (a to d) which involve two time series plots (a and c), and two scatter plots (b and d) for two different data segments both from Mine A. Plots a and b correspond to a segment of 250 days, while plots c and d correspond to a different segment of 180 days of collected data, Figure 3.2.4a shows the time series of the daily medians which are taken as representative values for each data stream. The scatter plot between barometric pressure and gas concentration for the same data segment shown in Figure 3.2.4b exhibits a strong negative correlation between methane gas concentration and barometric pressure with a linear Pearson correlation coefficient R=-0.77. The negative sign of the correlation suggests that when BP drops, methane concentration tends to increase. Figure 3.2.4c shows the time series of the daily median values of the gas concentration and the barometric pressure for the second data segment. In this case, the scatter plot in Figure 3.2.4d presents a weak, positive correlation between methane gas concentration and BP with R=0.24.

The weak correlation in Figure 3.2.4d is most likely the result of the unexpected drop of methane concentration around day 880, as shown in Figure 3.2.4c. This is also evident in the presence of two "clusters" of gas concentration values (a cluster of high values before the day of the drop and a cluster of low values later) shown in Figure 3.2.4d. The weak positive correlation is due to the fact that the regression analysis attempts to fit a straight line to these two different clusters). The sudden drop in the gas concentration, in this case, seems unrelated to changes in the barometric pressure. Most likely, its cause involves other variables directly affecting the correlation between CH_4 and BP, such as coal production and changes in mine operations (e.g., development of a new panel, coal recovery in more than one panel simultaneously, and downtimes). This dataset underscores the importance of incorporating data that represent operational changes into a successful AMS.

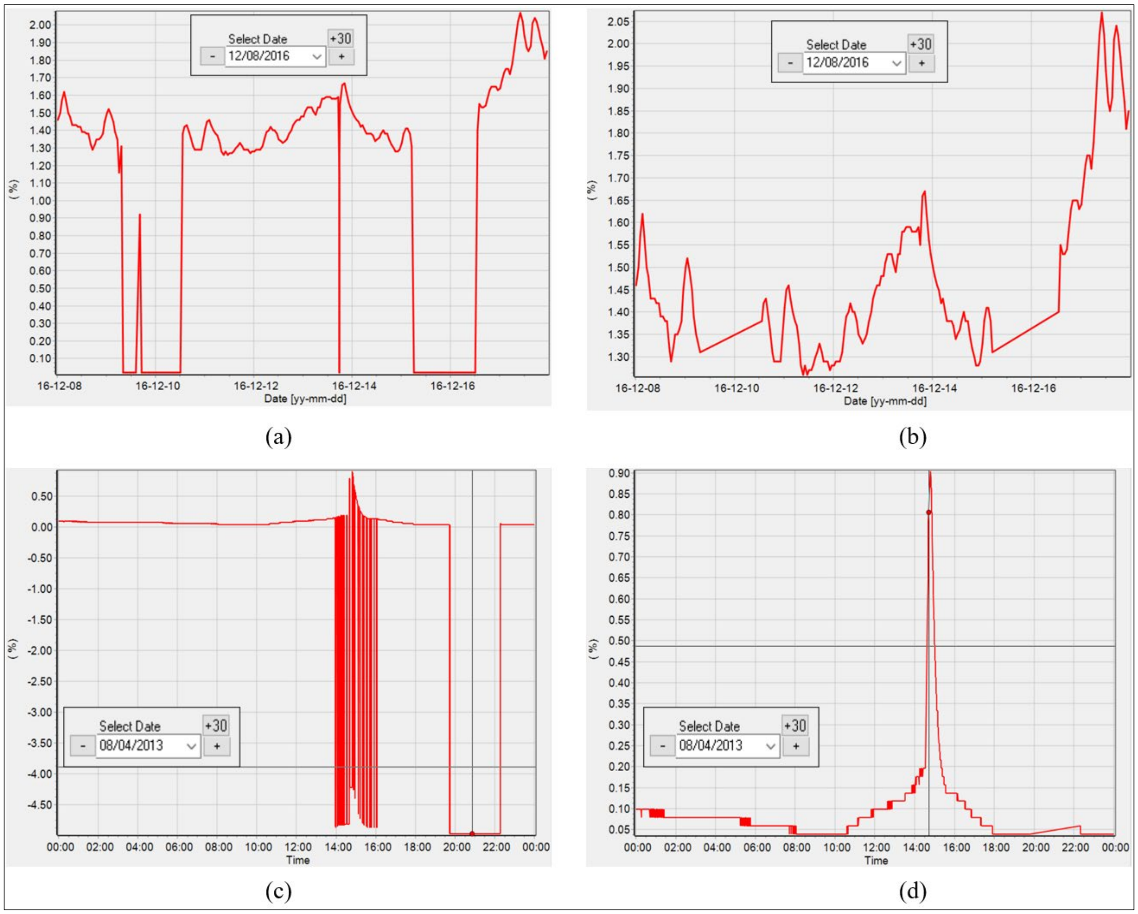

Figure 3.2.2 (a) Raw (before cleaning) CH_4 data from Mine A, (b) CH_4 data from Mine A – after data is cleaned up, (c) Raw CH_4 data from Mine B, (d) CH_4 data from Mine B – after data is cleaned up

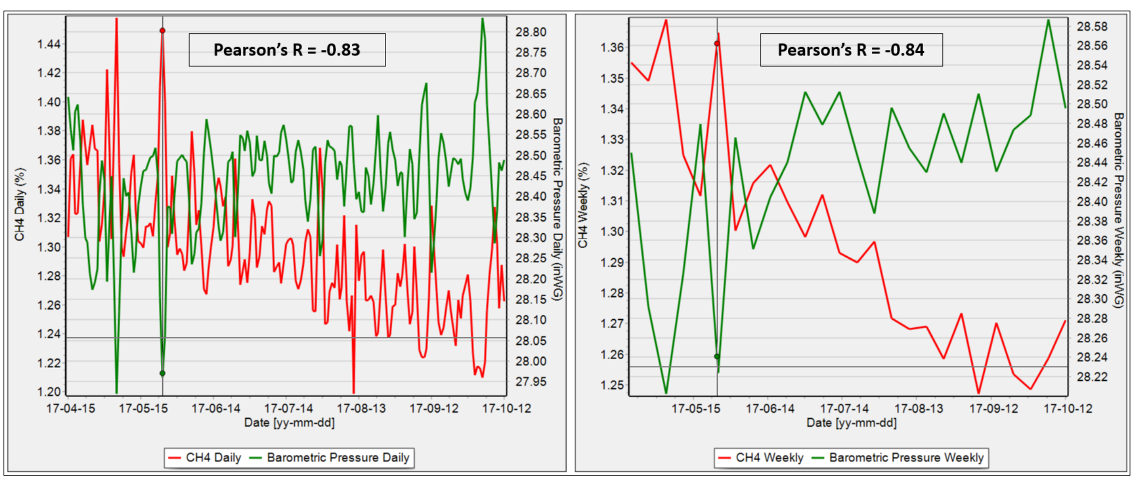

Figure 3.2.3 (a) Time series of daily averages for barometric pressure (green) and CH_4 concentration (red) extending over a six-month period. The value of the Pearson correlation coefficient (R=-0.83) is also shown. (b) The same plots as in (b) for weekly averages; R=-0.84

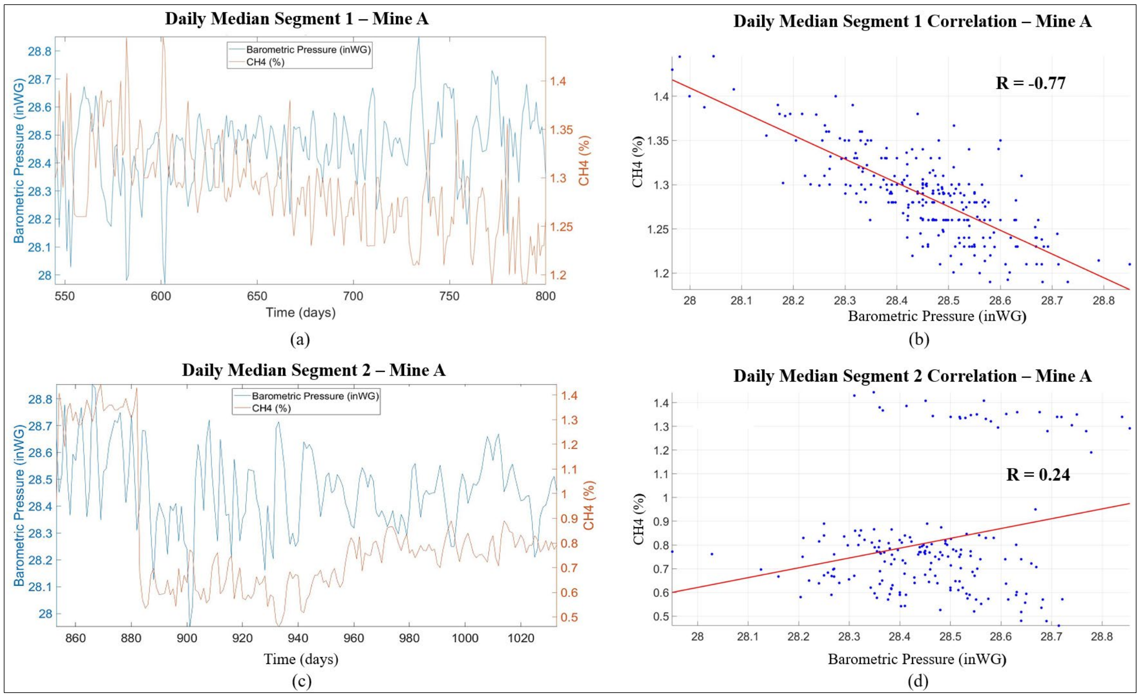

Figure 3.2.4 (a) Daily median values of gas concentration and barometric pressure for a specific time window (data segment 1). (b) Scatter plot and correlation of daily median gas concentration and barometric pressure values for data segment 1. (c) Daily median values of gas concentration and barometric pressure for a different time window (data segment 2). (d) Scatter plot and correlation of gas concentration and barometric pressure for data segment 2. Both data segments refer to Mine A

3.2.5 Summary and Conclusions

This paper discusses progress towards developing an accurate forecasting model for methane gas concentration in coal mines. The research analyzes data collected from Atmospheric Monitoring Systems in two underground coal mines in the US by employing time series models. In addition, a relational database management system (AMANDA) is presented, along with some of the statistical techniques implemented for data pre-processing and exploratory analysis. Finally, preliminary results of the potential correlation between methane gas and barometric pressure are presented and discussed.

This research has identified a significant negative correlation between methane gas and barometric pressure, at least for certain segments of the time series. Establishing such correlations increases the possibility that barometric pressure data can be used to predict variations in methane gas concentration in underground coal mines. However, more sophisticated statistical techniques for data analysis (e.g., time series analysis) need to be applied in order to identify variables potentially influencing methane gas and barometric pressure correlation; such variables may include changes in mine operation and coal production.

Future work will focus on (a) studying the correlation between methane gas and barometric pressure for the second case study (Mine B), (b) developing and validating long-term relationship(s) between meteorological parameters and methane gas concentration for both cases (Mines A and B), and (c) the identification and training of an accurate time series model which will allow forecasting methane gas concentration.

3.2.6 Acknowledgment

This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION), contract number AFCTG20-103. The views, opinions, and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its directors, and staff.

3.2.7 References

Agioutantis Z, Luxbacher K, Karmis M, Schafrik S (2014) Development of an atmospheric data management system for underground coal mines. J. South. Afr. Inst. Min. Metall, 114:1059–1063.

Agioutantis Z, Luxbacher K, Dougherty M, Karmis M (2015) Develop and implement a data management platform for atmospheric data in underground coal mines. In Proceedings of the 15th North American Mine Ventilation Symposium, 20–25 June 2015, Blacksburg, VA, USA.

Barry M, (2013) The canaries in the Coal Mine. European Respiratory Journal. https://doi.org/10.1183/09031936.00137913.

Byrer C, Havryluk I, Uhrin D (2014) Coalbed methane: a miner's curse and a valuable resource. In: Pramod T, Schatzel S, Aminian K (ed) Coalbed methane: from prospect to pipeline, 1st ed. Elsevier, Amsterdam, pp 1–6.

Chaulya SK, Prasad GM (2016) Sensing and monitoring technologies for mines and hazardous areas, 1st (ed). Amsterdam, Netherlands.

Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Towards atmospheric monitoring data analysis in underground coal mines. In: Proceedings of the 18th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781003188476-51.

Diaz J, Agioutantis Z, Hristopulos DT, Schafrik S (2021) Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. Mater. Proc. 2021. https://doi.org/10.3390/materproc2021005078.

Griffin R, Jong EC, Luxbacher KD, Westman EC (2011) A review of atmospheric monitoring systems in underground coal mines: implications for explosion prevention. In: Proceedings for the Society for Mining, Metallurgy, and Exploration (SME), Preprint 11-127, pp. 683–686. Denver, Colorado, United States.

Taylor C, Karacan CÖ (2010) Historical development of technologies for controlling methane in underground coal mines. In: Extracting the Science: A century of mining research. Brune JF (ed), Littleton, CO, Pittsburgh, United States.

4. UNIVARIATE FORECASTING APPROACH FOR METHANE GAS CONCENTRATIONS IN UNDERGROUND COAL MINES

The study of the potential association (correlation and autocorrelation) between methane gas emissions and possible independent variables (e.g., barometric pressure and coal production rate) along with the preliminary results of the univariate methane gas forecasting methods developed and proposed in this research is covered by the following peer-reviewed article submitted in January 2022 to the Mining, Metallurgy, and Exploration Journal.

TIME SERIES MODELING OF METHANE GAS IN UNDERGROUND MINES

Juan Diaz, University of Kentucky; Lexington, Kentucky; United States of America

Zach Agioutantis, University of Kentucky; Lexington, Kentucky; United States of America

Steven Schafrik, University of Kentucky; Lexington, Kentucky; United States of America

Dionissios Hristopulos, Technical University of Crete; Chania, Crete; Greece

Kray Luxbacher, Virginia Polytechnic Institute and State University; Blacksburg, Virginia; United States of America

Full Citation:

Diaz J, Agioutantis Z, Hristopulos DT, Schafrik S, Luxbacher K (2022) Time series modeling of methane gas in underground mines. Mining, Metallurgy, and Exploration Journal *(Submitted)*.

4.1 Abstract

Methane gas is emitted during both underground and surface coal mining. Underground coal mines need to monitor methane gas emissions to ensure adequate ventilation is provided to guarantee that methane concentrations remain low under different production and environmental conditions. Prediction of methane concentrations in underground mines can also contribute towards the successful management of methane gas emissions. The main objective of this research is to develop and validate long-term relationship(s) between methane gas emissions and meteorological parameters (e.g., barometric pressure, temperature, and humidity) and other potential variables (e.g., coal production) to build an accurate forecasting model of methane gas emissions and concentrations based on time series analysis. Methane time series data were retrieved from Atmospheric Monitoring Systems (AMS) of three underground coal mines in the United States. The AMS data were stored and pre-processed using an Atmospheric Monitoring Analysis and Database mAnagement system. Furthermore, different statistical techniques such as cross-correlation, autocorrelation, cross-covariance, and variograms were implemented to investigate the potential association between the dependent variable (methane gas) and independent variables (meteorological parameters and coal production). The ARIMA one-step-ahead model provides accurate forecasts that match the direction (increase/decrease) of the validation data.

Keywords: Methane gas concentration, Atmospheric monitoring Systems, Time series modeling, ARIMA forecasting.

4.2 Introduction

For more than a century, coal mine methane has been considered a significant threat to mining safety and productivity. That is especially true for underground coal mining, which is regarded as a hazardous economic activity due to numerous incidents that have killed thousands of coal miners worldwide. Explosions in underground mining due to methane gas have been the leading cause of incidents and fatalities in the mining industry. Methane gas is an explosive gas frequently found in underground coal mines, especially longwall mining (Chaulya and Prasad, 2016). Indeed, methane gas was nicknamed "*the miner's curse*" in the 19th century after the first documented mine explosions in the United States and France, where more than 1,400 miners lost their lives (Byrer et al., 2014; Flores, 1998). Since 1900 approximately 12,000 underground coal mine workers have been killed in over 500 mining accidents only in the United States (see Figure 4.1). Around 80% of the accidents and 90% of the fatalities were related to methane gas explosions. The Upper Big Branch Mine-South disaster at Montcoal, West Virginia, in 2010, was the last underground coal mining disaster due to a methane explosion in the US. This tragedy took the lives of 29 mine workers (Kowalski-Trakofler et al., 2009; NIOSH, 2021).

In the last decades, the number of disasters and fatalities has considerably decreased mainly due to the remarkable effort made by the mining industry, academia, and governmental agencies to develop new technologies and techniques to monitor and more efficiently control methane gas in underground coal mining operations (Airey, 1968; Curl, 1978). However, methane gas still presents challenges to underground coal mines globally. For example, the mining disaster at Listvyazhnaya coal mine, Russia, in November 2021, caused more than 52 fatalities (Kozlov, 2021). Other examples include the methane explosion in an underground coal mine in Colombia in August 2021, where 12 miners died (La República, 2021), and the disaster at Soma Coal Mine, Turkey, in 2014, which killed 300 mineworkers and is considered the worst mining disaster so far in this century (Düzgün and Leveson, 2018). These disasters are reminders of the need for developing reliable methane forecasting methods to improve mine safety and health in underground coal mine operations.

This paper aims to develop and validate long-term relationships between methane gas emissions and meteorological parameters and other potential variables such as coal production to build a reliable forecasting model of methane gas concentration. Time series data have been collected from Automated Atmospheric Monitoring Systems (AMS) installed at underground coal mines in the United States. The data have been filtered, cleaned, and subsequently analyzed using various techniques, including autocorrelation, cross-correlation, and variograms. Finally, preliminary results of the methane gas concentration forecasting model employing Autoregressive Integrated Moving Average (ARIMA) techniques are presented and discussed. This forecasting model is also expected to be applicable in other underground mining and industrial operations.

Section 4.3 presents a literature review on time series that identifies the most relevant previous studies that have attempted to develop a reliable methane gas forecasting model for underground coal mine operations based on time series analysis. Furthermore, Section 4.4 briefly describes the atmospheric monitoring system implemented by each case study. Moreover, Section 4.5 explains the main steps used in this research for Atmospheric Data Management, including data collection, store, pre-processing, and processing. Likewise, Section 4.6 presents the statistical techniques implemented to investigate the potential association (correlation and autocorrelation) and validate long-term relationship(s) between methane gas emissions and meteorological parameters and other variables such as coal production. Additionally, Section 4.7 displays the preliminary methane gas concentration forecasting results obtained by employing ARIMA models. Moreover, Section 4.8 discusses the main findings of the research. Finally, Section 4.9 presents the conclusions and the future work of the research discussed in this study.

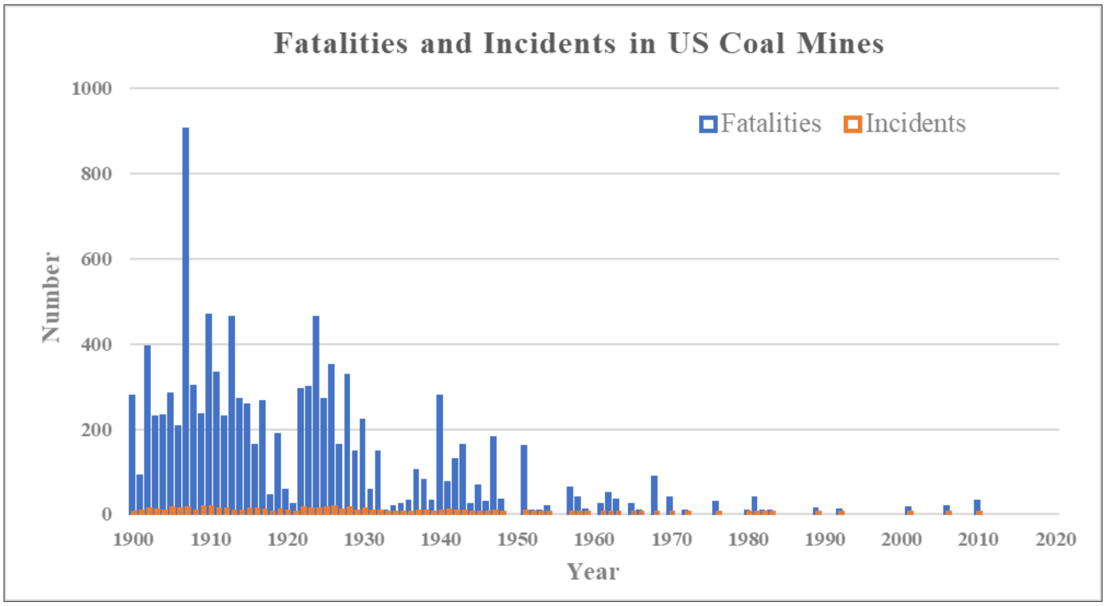

Figure 4.1 Fatalities and incidents in underground US coal mines since 1900; 80% of the incidents and 90% of the fatalities can be attributed to methane gas explosions (NIOSH, 2020)

4.3 Time Series Analysis for Methane Gas Forecasting

The study and analysis of time series data started a long time ago when the statistician Undy Yule presented his work named "*On a Method of Investigating Periodicities in Disturbed Series, with Spatial Reference to Wolfer's Sunspot Number*" in 1927 (Yule, 1927; Tsay, 2000). At that time, statisticians were the only ones permitted to analyze and deduce theories and hypotheses from data. Consequently, they used to develop and implement complex methods characterized by numerous assumptions regarding the nature of data (Nielsen, 2019). However, in 1970, the field of time series analysis and forecasting was opened to a much larger public when the statisticians George Box and Gwilym Jenkins published their book "*Time Series Analysis: Forecasting and Control.*" This book contained straightforward methods to work with time series data and its analysis (Box et al., 2015).

In recent years, time series analysis has become one of the most essential and predominant statistical techniques in experimental science and data analysis, mainly due to advances in statistics and computer science such as machine learning and the internet of things (IoT), which have considerably improved data processing, collection, storage, and visualization (Nielsen, 2019; Brockwell and Davis, 2016). Mathematically speaking, a time series can be defined as a doubly infinite sequence (a series a_n where n varies from $-\infty$ to $+\infty$) of multiple random variables. In other words, a time series is a set of observations X_t, where each observation is recorded at a specific time t. The most common series are in hourly, daily, weekly, monthly, and annual frequencies. Time series analysis refers to a collection of tools for analyzing massive datasets that consist of records arranged in chronological order. The main objective of the analysis is to determine the statistical properties inherent in the data and construct a model (or models) that allows filling gaps in time series and forecasting their values in the future (Box et al., 2015; Brockwell and Davis, 2016).

The practice of time series analysis in the past was limited to a few disciplines and applications due to the complexity of the data collection process. For example, medicine, meteorology, economics, and astronomy were the first disciplines to use time series methodologies (Cryer and Chan, 2008). One of the earliest documented works that used time series data was the work done by John Graunt in 1662. He published a book named "*Natural and Political Observations Made Upon the Bills of Mortality*." This book was based on the death records kept in London districts since the early 1500s. In his book, Graunt presented the first "*life tables,*" known today as actuarial tables. John Graunt's research was one of the first works based on time series data and analysis to address medical concerns (Harkness, 2020). However, nowadays, the number of areas in which time series are studied and implemented is massive (Astudillo et al., 2020).

The mining and minerals industry is one field that has benefited from the rapid development of time series analysis. It is used, among other purposes, (a) to forecast mineral commodity prices (Astudillo et al., 2020; Tapia et al., 2018; Olayiwola, 2016), (b) to approximate the future mineral supply and demand (Watari et al., 2020; Rosienkiewicz et al., 2017; Renner and Wellmer, 2020), (c) to estimate production in real or near real-time (providing operators and engineers with insight into current operating conditions) (Rodriguez, 2020) and to optimize blasting and drilling operations for mineral recovery (Gupta et al., 2020; Bilal et al., 2013). Additionally, time series analysis can be used to ensure the health and safety conditions of the workforce, particularly in underground coal mines operations.

The study and implementation of forecasting methodologies to predict the concentration and emissions of hazardous gases in underground coal mines have been a topic of interest for academia, the mining industry, and governmental agencies for many decades (Bilal et al., 2013; Dixon, 1992; Dixon and Longson, 1993; Trutwin and Wasilewski, 1987; Karacan et al., 2005; Griffin et al., 2011; Diaz et al., 2021). As a result, different approaches have been implemented to tackle this problem. For example, the empirical approach for methane gas prediction was the first methodology employed (Owili-Eger et al., 1973). The first documented empirical research dates back to the 1960s, when researchers attempted to build a methane gas forecasting method to analyze time series data collected manually from underground coal mines for a short period (Dixon, 1992; Dixon and Longson, 1993; Trutwin and Wasilewski, 1987). Then, with the advances in computer technology and mathematical sciences, the numerical approach, mainly the Computational Fluid Dynamics (CFD) methodology, has been explored as a potential solution to predict methane emissions (Sung et al., 1987). However, precise explosive and flammable forecast methods are still challenging due to the numerous in-situ characteristics of each mine operation (e.g., production rate and mining parameters, geological characteristics, topography features) that affect methane gas emissions into the underground mining environment (Diaz et al., 2021; Booth et al., 2017).

Furthermore, the statistical approach, principally the implementation of time series analysis for hazardous gas forecasting in underground coal mines, has been studied in recent years mainly due to its advantages over other methods. For example, time series methodology is considered less time-consuming and expensive. In addition, the data collection process is rapid and reliable due to the identification of outliers, patterns, and missing data. Also, it allows data cleaning and validation (Booth, 2017; Shumway and Stoffer, 2017). This section presents the most relevant previous studies that have attempted to develop a methane gas forecasting model for underground coal mine operations based on time series analysis.

Kaffanke (cited in Dixon, 1992) developed a methodology for predicting concentration and emissions of methane gas for segments between one day to one month in length by using discrete multiple linear regression, which is a statistical technique implemented to calculate the degree of association between two or more independent variables (explanatory variables) and a dependent variable (response variable) by fitting a linear equation to experimental data (Shumway and Stoffer, 2017). The most relevant variables included in the model were: daily coal output (tons/day), accumulated output (total tons), previous day methane gas flow (m^3/s), and idle day methane flow (m^3/s) or the number of no working days and released gas content (m^3/t). Methane emissions and concentration forecasts were done by implementing seven equations for each day of the week. The author made two important conclusions: first, accurate methane emission prediction can be achieved by implementing statistical methods, in this particular case, discrete multiple linear regression. Secondly, further research was recommended by implementing different statistical techniques for methane prediction, such as linear filter theory, spectral analysis, and time series analysis.

Trutwin and Wasilewski (1987) established an approach for modeling airflow in underground coal mine operations by studying their atmospheric data and implementing time series analysis and digital filtering techniques, which is a method that uses mathematical tools to identify and modify specific aspects of a signal such as noise due to mechanical vibration and electrical interference. In this particular case, the low-pass first-order recursive filter was employed. The primary objective was to identify and separate signals of different amplitudes and durations that disturb and hinder monitoring parameters such as methane gas emissions, oxygen concentration, and air velocity in underground coal mines. The authors concluded that random disturbances influencing ventilation systems in underground coal mining could be classified into three different groups based on their frequencies; disturbances with a low frequency below $1.5x10^{-5}\ Hz$, disturbances with an intermediate frequency between $1.5x10^{-5}\ Hz$ and $2.77x10^{-4}\ Hz$, and the last group is disturbances with a high frequency above $2.77x10^{-4}\ Hz$.

Dixon (1992) recommended a methane concentration and emissions forecasting method using atmospheric data collected manually from longwall coal mines based on the Box-Jenkins time series analysis technique. This research implemented the Autoregressive Integrated Moving Average model for describing stationary and non-stationary time series. The implementation of this technique identified the relationship between the dependent variable, methane emissions, and its independent variables, such as coal production, barometric pressure, and airflow velocity. In addition, the author highlighted that this model could be constructed without previous knowledge of the series itself, and the model was trained and built using past values of the same series. Finally, the author recommended further research to forecast methane concentrations in underground coal mining operations by implementing time series analysis methods.

Dixon and Longson (1993) proposed a statistical method for short-term forecasting methane gas emissions based on time series techniques developed by analyzing data collected and stored from non-automated AMS. The model considers the most important sources of methane gas, such as the methane released from the working coal seam, the stratigraphy above and below the working seam, and the degree of methane emissions from the adjacent seams and strata. Furthermore, the potential correlation between methane gas and barometric pressure, coal production, and air velocity were studied. However, the association between methane gas and barometric pressure could not be measured because the BP data could not be collected. Nevertheless, the correlation between methane gas and coal production was assessed. It was concluded that the method selected for coal recovery (mining method) and its production rate is a crucial independent variable that affects methane gas concentration and, consequently, methane gas forecasting.

Tominaga and Bandopadhyay (2002) introduced a technique to monitor and forecast spontaneous combustion in underground coal mine operations based on time series data collected by carbon monoxide (CO) sensors in a longwall mine located in Hokkaido, Japan. This research made use of the second law of diffusion of Fick to simulate a mathematical model able to identify time series data characteristics, such as the concentration-time curve. The researchers concluded that in a forced ventilation system, the location of the carbon monoxide source could be determined precisely when the distance from the origin of CO emission has a linear correlation with the shape characteristics of the concentration-time curve.

Shu-gang et al. (2008) developed a method to predict methane gas emissions originated at the faces of longwall coal mines based on time series analysis and the Least Square Support Vector Multi-classifier (LS-SVM), which is a collection of techniques implemented to explore time series data and identifies patterns implemented mainly for regression analysis. The authors concluded that such a method could be implemented for methane gas concentration and emissions forecasting and a warning system at the longwall face to alert operators when the mine atmosphere becomes dangerous.

Shengrui et al. (2011) recommended a model of gas emissions forecast based on the study of time series methane data and Chaos Theory, which is a state of apparently random disorders (chaos), anomalies, and irregularities controlled for basic laws. The Chaos theory implies that predictions directly depend on initial conditions (Jorgensen and Fath, 2019). This short-time prediction model was built using data from an underground coal mine in

China collected during five days, at a time interval of five minutes. It was concluded that methane gas forecasting in underground coal operations could be achieved by studying time series data and the Chaos theory methodology. According to this research, the model showed good results under unstable conditions (e.g., rockburst, coal burst, and gas outburst). However, they suggested further investigation to obtain more accurate calculations.

Zagorecki (2015) proposed a forecasting method based on sensor fusion and data mining techniques to predict methane gas emissions by studying time series data from an underground coal mine. The main objective of this research was to predict in a short time (3 to 6 minutes) methane outbreaks (exceedance of CH_4 threshold levels) at three specific locations in an underground coal mine located in Poland using multivariate time series data, approximately 57,000 records were collected for a period of three months. Temperature, humidity, and barometric pressure were some of the atmospheric variables studied for building the methane forecasting model. Furthermore, the most important data mining technique implemented was the machine learning random forest algorithm implemented in Waikato Environment for Knowledge Analysis (WEKA). This system implements a set of algorithms, mathematical tools, and data visualization packages for data analysis and forecasting. The author concluded that the model is an acceptable solution for methane gas forecasting in underground coal mines under stable conditions.

4.4 AMS in Underground Coal Mines

Atmospheric Monitoring Systems are characteristically implemented in underground coal mines in order to monitor parameters such as concentrations of toxic and explosive gases in exhaust air as well as weather-related parameters. Processing AMS data ensures that hazardous gases do not exceed regulatory standards and that engineering controls are effective. Consequently, reducing high-risk mining environments. Rudimentary methods of monitoring the mining atmosphere can be traced back to the implementation of warm-blooded animals such as canaries to alert miners when the atmosphere becomes harmful. Nowadays, more sophisticated and accurate AMS are employed to monitor underground coal mines more efficiently (Taylor and Karacan, 2010; Goodman et al., 2008). This section briefly describes the AMS employed in each case study (Mines A, B, and C) analyzed in this research.

The first case study (Mine A) uses an automated atmospheric monitoring system identified as Wireless Multi-Gas Monitor (Figure 4.2) installed on the exhaust shafts. The Wireless Multi-Gas Monitor is an excellent example of the modern technologies available to monitor the atmosphere of underground coal mines. This device can simultaneously monitor up to four gases (e.g., CH_4, O_2, CO, CO_2, NO, H_2, and SO_2). Furthermore, it has a variety of advantages such as remote operation through a Wi-Fi connection, reduced costs, and being user-friendly because it monitors several gases simultaneously, and no instruments or special skills are required to replace sensor modules. In addition, its firmware or computer software is updated automatically (AMR PEMCO, 2002).

Figure 4.3 contains four plots (a to d) that illustrate the operation of the Wireless Multi-Gas Monitor utilized in Mine A. First, a line is installed in the exhaust shaft to carry in and out to the monitoring station the returned air from the mine atmosphere, as shown in Figure

4.3a and 4.3b. Then, the return air (polluted air) is taken inside the Wireless Multi-Gas Monitor box, where the concentration of the gases is assessed, as illustrated in Figure 4.3c. In this particular case, Mine A, the concentration of three gases is monitored using four sensor modules: one for carbon monoxide (CO), another one for oxygen (O_2), and two for methane gas, as shown in Figure 4.3d. Finally, the information collected and stored by the Wireless Multi-Gas Monitor is transferred to an online system used to monitor gas emissions at several exhaust shafts where the data can be directly downloaded.

The second case study (Mine B) employs an automated AMS that collects gas concentration data (with a sampling rate of about 10 seconds) from different sensors throughout the mine, in addition to the standard sensors that collect data from fans, and conveyor belts, etc. Collected data are electronically transmitted to a central monitoring system on the surface for further processing. Data for the third case study (Mine C) are collected manually. Methane gas is measured weekly at the exhaust shaft(s) using a manual process and appropriately recorded.

Figure 4.2 Wireless Multi-Gas Monitor (AMR PEMCO, 2002)

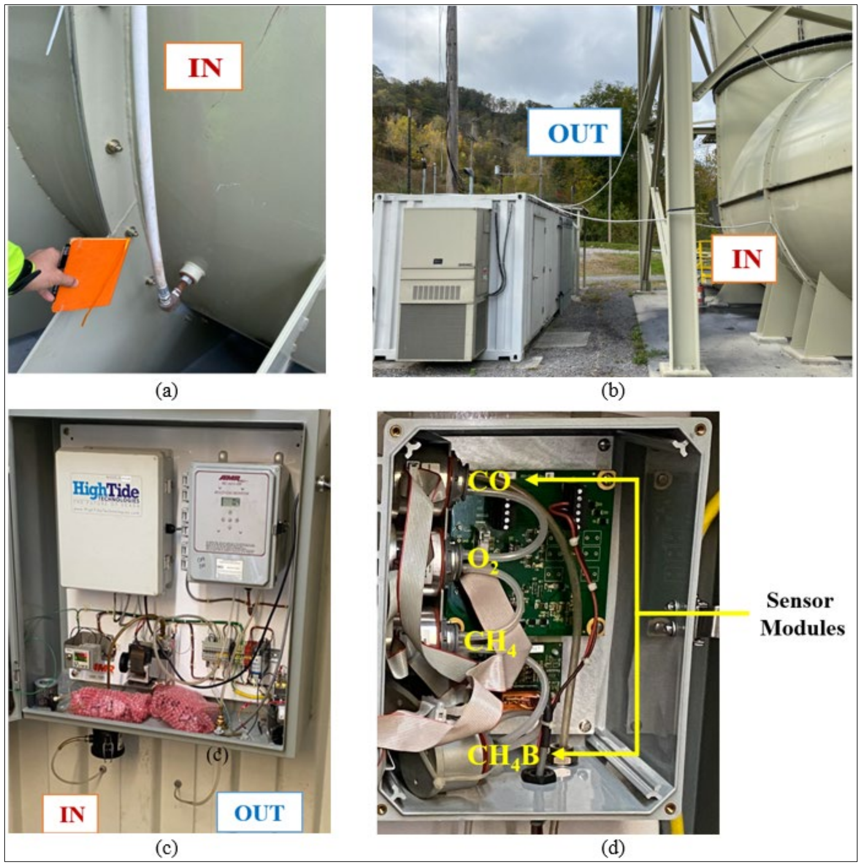

Figure 4.3 Wireless Multi-Gas Monitor process

4.5 Atmospheric Data Management

In addition to the gas emission data (e.g., CH_4, CO, and O_2) collected from three underground coal mines, meteorological data (e.g., barometric pressure, temperature, humidity, precipitation, and wind speed) were retrieved from the nearest weather station to each mine. The data were automatically downloaded from a commercial weather service that provides real-time meteorological conditions information over the internet called Weather Underground Commercial Company (WU).

Figure 4.4 shows a flow diagram describing the main steps of the data management process (e.g., data collection, storage, pre-processing, and processing) implemented in this research. The first step consists of collecting the data: (a) the atmospheric data are retrieved from the closest weather stations to each mine, and (b) gas concentration data are recovered from the mine monitoring systems in the form of excel files. Then, the atmospheric and the mining monitoring data are stored into AMANDA (Atmospheric Monitoring Analysis and Database mAnagement), a custom relational database designed to manage AMS data. More detailed information concerning the characteristics (e.g., source, frequency, and units) of the data collected can be found in Diaz et al. (2021).

The second step refers to data pre-processing. It is already established that real-world data must be pre-processed for appropriate time series analysis (Grimberg et al., 2021; Verma, 2019; Baur et al., 2015). In this case, data pre-processing is performed using AMANDA;

this stage includes data cleaning and filtering (e.g., identifying missing values, zero values, erroneous data, spikes, and outliers). The filtered data values are flagged as erroneous values and are not replaced with zero or null values during this stage. Thus, they are excluded from any subsequent data calculations or analyses performed. More details are given in section 4.5.1.

The third step is data homogenization, which is crucial when analyzing time series data because it guarantees that data points from different series share a common date/time stamp (Diaz et al., 2021a). Data homogenization was performed both using AMANDA and the MATLAB® programming environment. Data homogenization, as implemented in AMANDA, is a straightforward process that can develop 12-hour, daily, or weekly averages for each and every data stream and manage these generated time series as separate data streams. Also, the data streams can be exported for further processing in MATLAB. Data homogenization in the MATLAB environment utilized the interp1 command that interpolates between existing data points in given pairs of time series to determine new points with a common time stamp. The new points are used in subsequent processing. Figure 4.5 illustrates an example where two different data sets are homogenized using interpolation. The blue circles represent a set of data of methane gas concentration taken from one of the case studies, while the white circles correspond to a set of barometric pressure data collected from WU. As illustrated in Figure 4.5a, the data points within the same series and from the two different series do not share common and/or regular date/time stamps.

The final step of atmospheric data management includes all the processes run either on the raw data or the homogenized data. These range from simple calculations of Pearson Correlation Coefficient between two data streams (e.g., methane data vs. barometric pressure, methane data vs. coal production) as well as the linear correlation relationships.

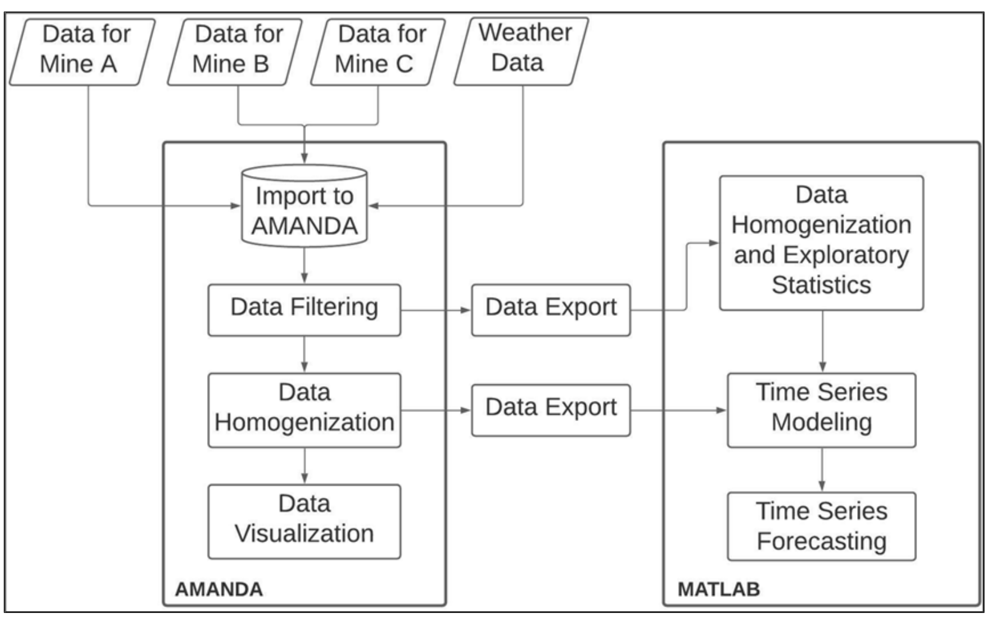

Figure 4.4 Flow diagram of data management

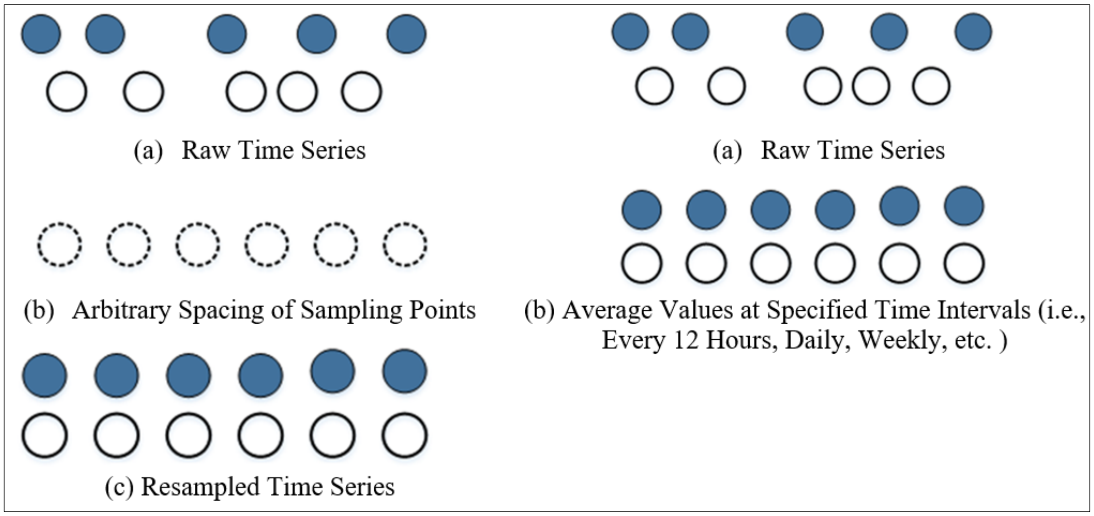

Figure 4.5 Data homogenization representation: Left: Raw time series data with irregular sampling steps converted to resampled data at an arbitrary spacing. Right: Raw time series data with irregular spacing are converted to average values for specified time intervals

4.5.1 Data Pre-Processing

The majority of the AMS are affected by the complex nature of underground coal mining environments (e.g., humidity, presence of dust, power, and calibration issues). As a result, the data streams provided by AMS frequently present inconsistencies such as calibration spikes, erroneous and/or negative values, and gaps mainly due to sensor malfunction. Therefore, collected data will need to be pre-processed before being used for analysis. As a result, inspecting, identifying, and deleting (or filtering out) anomalous records in the database is vital to ensure data reliability and integrity. In addition, the presence of correlation between data streams should always be investigated on pre-processed datasets to achieve a better time series forecasting model (Griffin, 2013; Zhou et al., 2017).

Figure 4.6 includes four plots (a to d) that illustrate some of the techniques implemented in this research using AMANDA to filter the methane gas data collected, in this case, from Mine A. Figure 4.6a shows raw methane data time series for a period of fifteen days with an estimated hourly time step. The data include sudden drops in methane gas concentration (CH_4 reaches 0%) due to sensor malfunction. Figure 4.6b exhibits the same time segment from Mine A after the zero values have been filtered out. Figure 4.6c presents a segment of methane data for twenty days where the methane gas concentration does not change; it keeps the exact concentration (CH_4=2.83%) for the whole period due to improper sensor calibration. Finally, Figure 4.6d shows a segment of thirty days with methane gas missing data (note that missing data is different from zero data), most likely due to sensor failure. It should be noted that problematic data points are not deleted; they are flagged so that they are not included in the analysis, but they can always be restored if needed.

Figure 4.7 comprises four plots (a to d); the first two plots (a and b) correspond to Mine B while the last two plots (c and d) belong to the closest weather station from Mine B. Figure 4.7a shows raw methane data time series from Mine B for one day sampled using a time step around ten seconds. The data include negative values for methane gas concentration (CH_4 reaches –5.0%) due to sensor malfunction; methane concentrations can never be negative. Figure 4.7b exhibits the same time segment from Mine B after the methane gas negative values have been filtered out. Figure 4.7c describes a segment of barometric pressure for six months with an approximate hourly time step. It is shown that BP reaches zero values (BP = 0.0 InWG) two times, which is a mistake (barometric pressure cannot be zero) due to sensor failure. Figure 4.7d exhibits the same time segment shown in Figure 4.7c after the zero barometric pressure values have been filtered out.

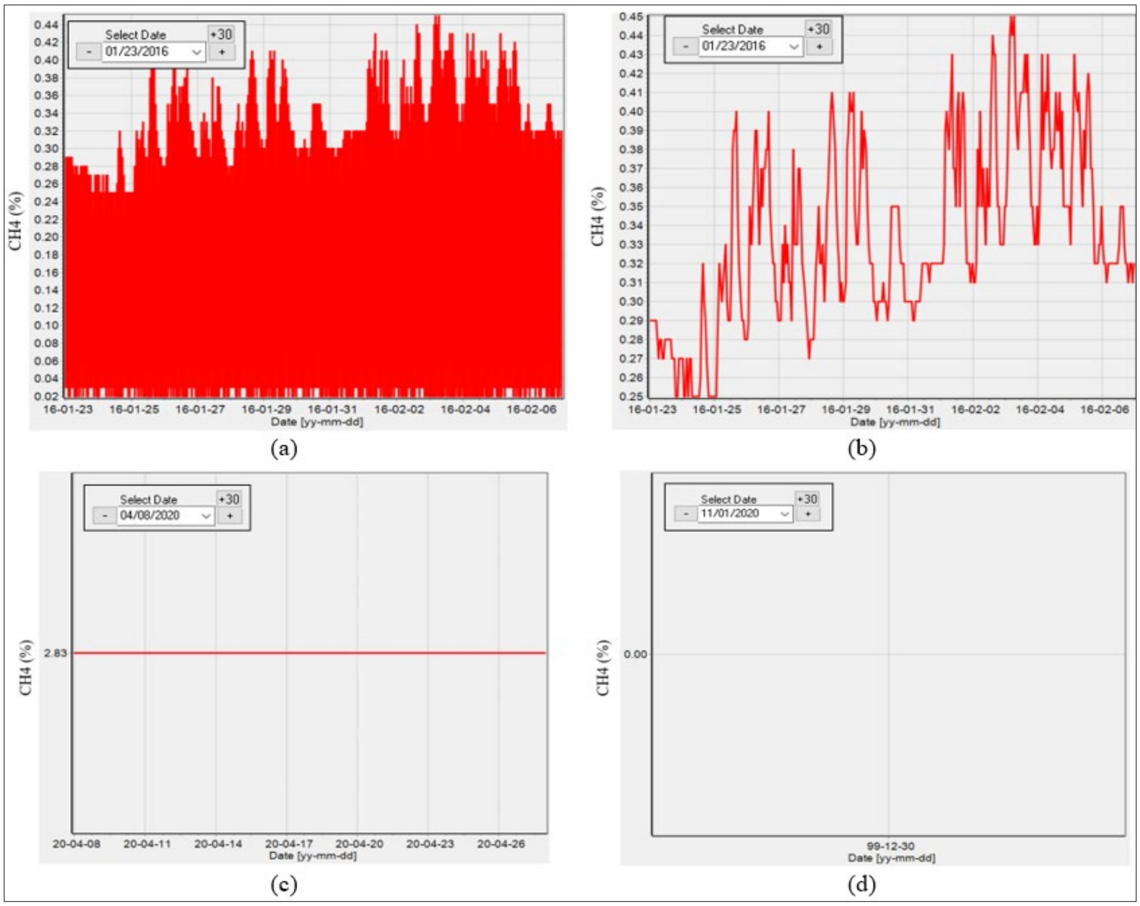

Figure 4.6 Data pre-processing (a) Raw CH_4 data from Mine A (before filtering), (b) CH_4 data from Mine A (after data is cleaned up), (c) Inconsistent CH_4 data from Mine A, (d) Segment where no CH_4 data were collected from Mine A sensors

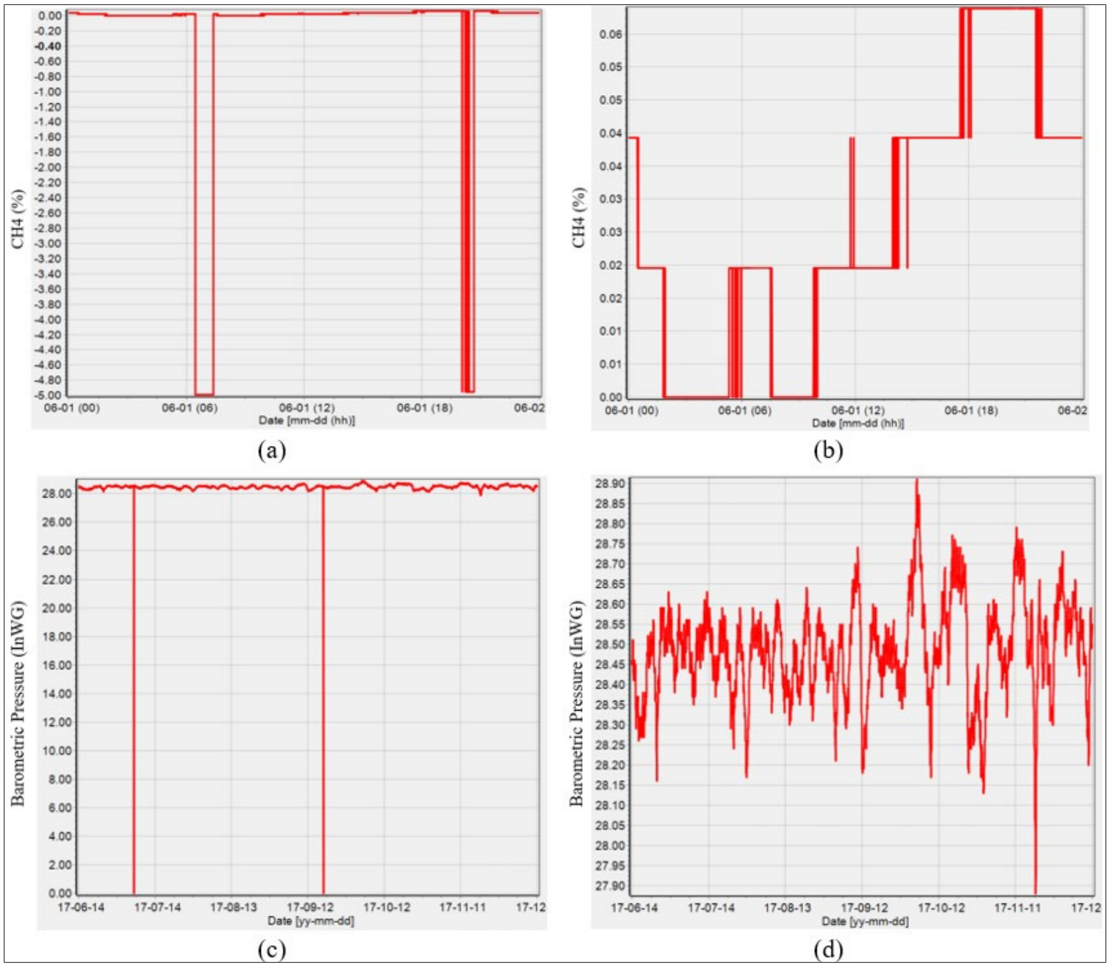

Figure 4.7 Data pre-processing, (a) Raw CH_4 data from Mine B (before cleaning), (b) CH_4 data from Mine B (after data is cleaned up), (c) Barometric pressure data (before cleaning), (d) Barometric pressure data (after cleaning)

4.5.2 Spike Detection and Analysis

Identifying and analyzing peaks (spikes and inverted spikes) in a time series is essential to ensure the accuracy of a forecasting model. If spikes are not detected during data pre-processing, they will compromise further data processing, directly impacting outcomes (Palshikar, 2009; Broquet et al., 2018; Goin and Ahern, 2019). Spike analysis includes different steps such as identification of the length of spikes (e.g., seconds, minutes, hours, and days), the magnitude of spikes, the similarities between spikes (e.g., time or frequency of occurrence, duration) as well as whether spikes occur in multiple time series (Vlachos et al., 2004). This section identifies and analyses the spikes and inverted spikes found during the pre-processing phase of AMS data streams collected from Mine A utilizing AMANDA.

Figure 4.8 consists of four plots (a to d) corresponding to Mine A for a five-day window. The red line represents methane gas concentrations (%) sampled using an hourly average time step, and the green line denotes daily coal production (tons/day). Figure 4.8a shows a methane gas spike, the concentration of methane gas increases unexpectedly from approximately 0.60% to more than 1.60%. Figure 4.8b illustrates the potential correlation between methane gas spike and daily coal production; the spike on the methane gas time series occurred when there was no coal production, as shown in Figure 4.8b. On the other hand, Figure 4.8c outlines an inverted methane gas spike; the concentration of methane gas decreases suddenly from around 1.60% to 0.80%. Finally, Figure 4.8d shows the correlation between the inverted methane gas spike and daily coal production; the CH_4 inverted spike occurred when there was no coal production, as shown in Figure 4.8b. The unexpected methane gas peaks (spike and inverted spike) presented in Figure 4.8 are mainly due to sensor calibration. This information was confirmed with the mine personnel from Mine A.

Figure 4.9 includes plots a and b that correspond to a period of ninety days for Mine A. Again, the red line represents methane gas concentrations (%) sampled using a daily average time step, and the green line denotes daily coal production (tons/day). Figure 4.9a shows a methane gas spike, the concentration of methane gas increases unexpectedly from about 0.20% to more than 3.0%. Figure 4.9b illustrates the correlation between methane gas and daily coal production. In this case, unlike in Figure 4.8, the spike in the methane time series occurs almost at the same time when coal production increases. This behavior can be explained due to other variables directly affecting the emissions and concentrations of methane gas, such a significant increase in coal production is most likely due to coal recovery in more than one panel simultaneously.

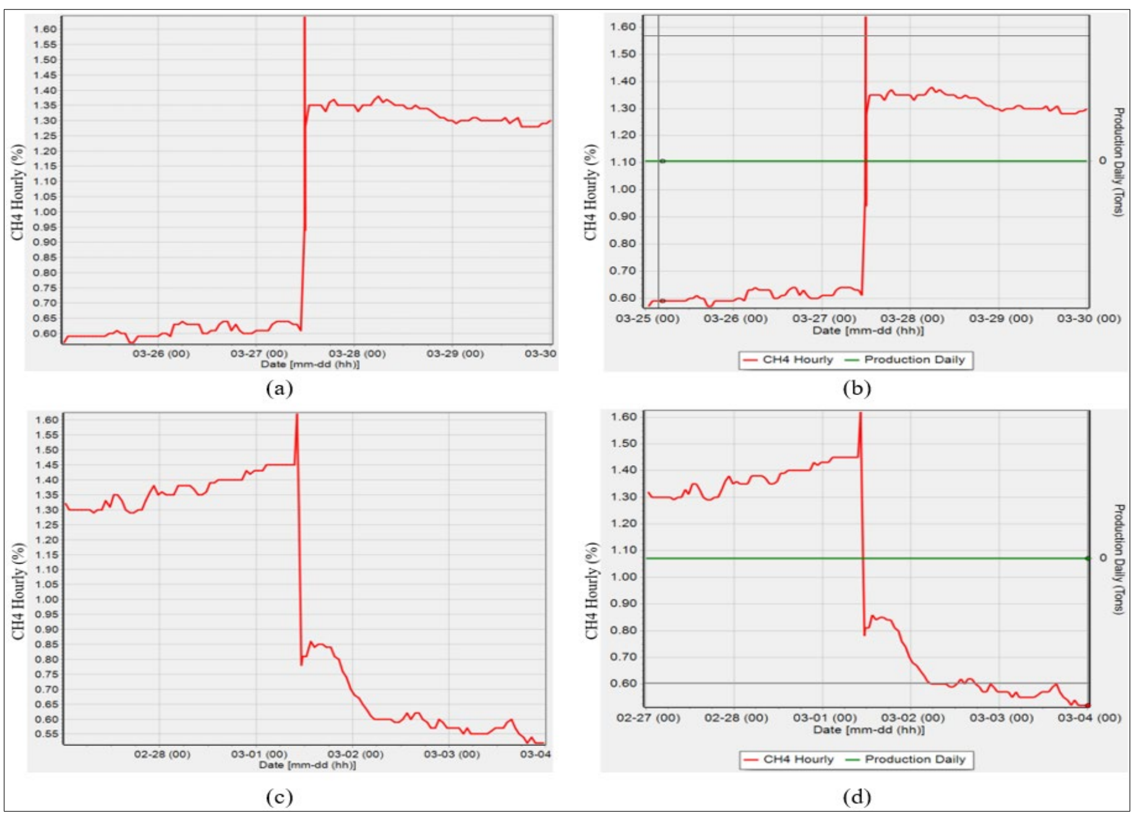

Figure 4.8 Spike detection, (a) CH_4 spike identification, (b) Correlation between CH_4 spike and daily coal production, (c) CH_4 inverted spike identification, (d) Correlation between CH_4 inverted spike and daily coal production

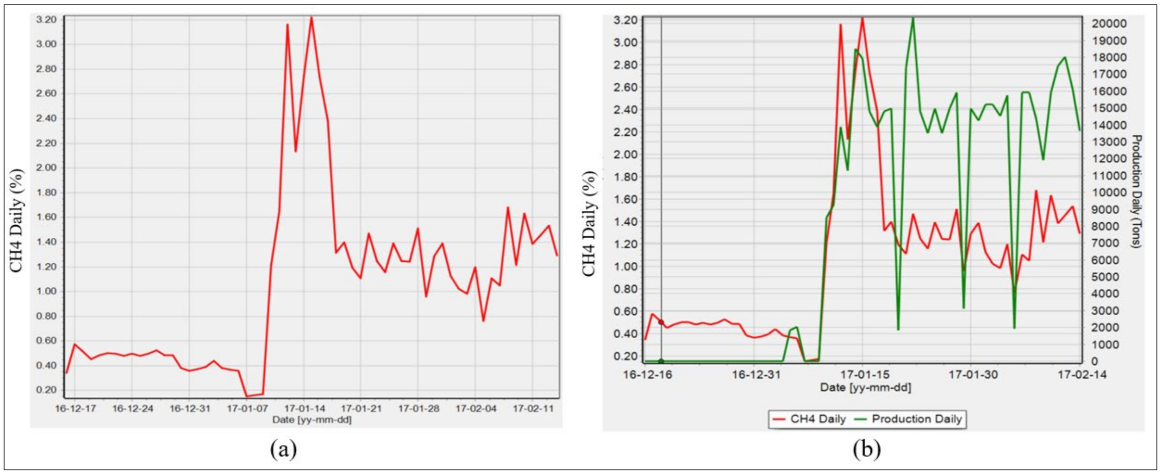

Figure 4.9 Spike detection, (a) CH_4 spike identification, and (b) Correlation between CH_4 spike and daily coal production

4.6 Exploratory Analysis of Methane Time Series

This section presents the statistical techniques (e.g., cross-correlation, autocorrelation, cross-covariance, scatter plots, and variograms) implemented to investigate the potential association (correlation and autocorrelation) and validate long-term relationship(s) between methane gas emissions and meteorological parameters and other variables such as coal production.

4.6.1 Correlation of Methane Concentration with Barometric Pressure and Production

Correlation is a statistical measure used to quantify (i) the strength of association between two variables and (ii) the direction of their relationship. The linear correlation between two variables varies between +1 and -1. A value of ±1 suggests a perfect positive/negative correlation between the variables. As the correlation coefficient value tends to 0, the association between the two variables becomes weaker. Furthermore, the direction of the relationship between the variables is indicated by the sign of the coefficient; a positive sign (+) indicates that the variables are directly proportional (when one variable increases, the other variable also increases, and vice versa), and a negative sign (–) signifies an inverse proportional relationship (when one variable increases the other variable decreases and vice versa) (Shumway and Stoffer, 2017; Thomas, 2014). Different measures of correlation exist in the literature, including the following: Spearman and Kendall rank correlation coefficients (used to measure ordinal association and applied in cases of nonlinear dependence), the point-biserial correlation (used when one of the variables is dichotomous), Kendall rank correlation, and Pearson's (linear) correlation coefficient (Schober and Schwarte, 2018). The Pearson correlation (R) was chosen based on exploratory data analysis to investigate the relationship of methane gas concentration with barometric pressure and coal production.

Figure 4.10 consists of four plots (a to d) corresponding to Mine A. Plots a and c were obtained using AMANDA, while plots b and d were generated employing MATLAB. Figure 4.10a presents two time series: the red line represents methane gas concentrations (%), and the green line denotes barometric pressure (InWG), both sampled on a daily average basis for 180 days. Figure 4.10a shows a strong negative correlation between methane gas and barometric pressure with the Pearson correlation coefficient calculated at R=–0.77. The negative sign indicates that methane concentration increases when barometric pressure decreases and vice versa. Figure 4.10b presents a scatter plot using the same data streams in Figure 4.10a, illustrating the strong negative correlation between barometric pressure and methane gas emissions.

Figure 4.10c shows two time series, methane gas concentrations and barometric pressure; both sampled on a daily average basis for a different interval of 180 days. These plots reveal a lack of correlation between methane gas (red line) and barometric pressure (green line), with the Pearson correlation coefficient estimated at R=0.00. This is further supported by the scatter plot in Figure 4.10d, which shows a lack of systematic relation between gas concentration and barometric pressure. This behavior is due to a sudden drop in the methane gas time series from values exceeding 4% to around 1%, as shown in Figure 4.10c. Such discontinuities can directly affect the outcomes of any method used for data analysis, as discussed in Section 5.5.2. They also suggest that the gas concentration depends not only on the barometric pressure but also on factors related to mine operations.

Figure 4.11 consists of two plots (a and b) corresponding to Mine A. The red line represents methane gas concentrations (%) sampled with a daily average step, and the green line denotes coal production (tons/day) sampled daily. Both time series in the plot (a) correspond to an interval of 30 days. Figure 4.11a illustrates a strong positive correlation between methane gas and coal production, with the Pearson correlation coefficient calculated at R=0.81. The positive sign denotes that methane concentration increases when coal production increases. Furthermore, Figure 4.11b shows a high positive correlation between methane emissions and coal production. In this case, a weekly average time step is used over 12 weeks, and the Pearson correlation coefficient is calculated at R=0.84. This analysis supports the hypothesis that gas concentration is not unilaterally related to barometric pressure. It is more likely that the correlation with barometric pressure is stronger during nearly constant production activity periods.

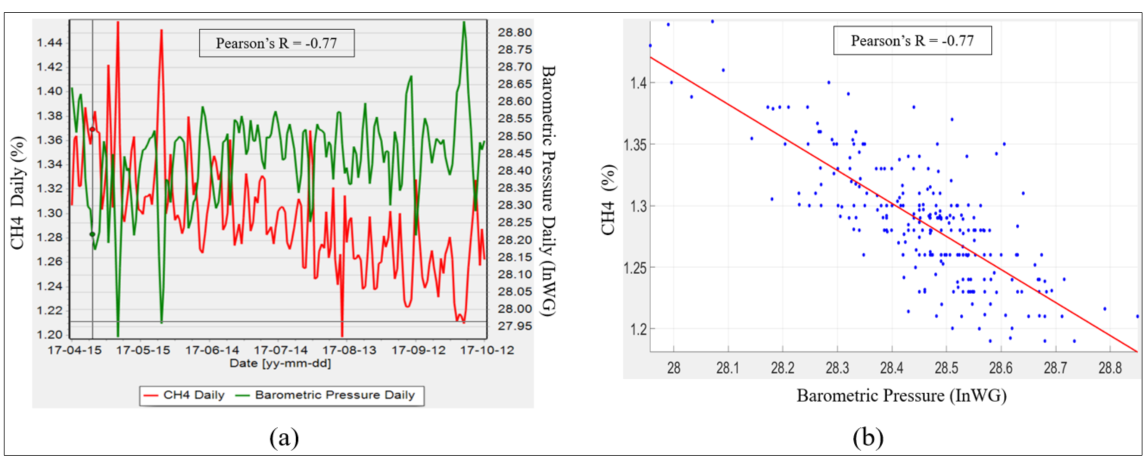

Figure 4.10 Pearson correlation between BP and CH_4 emissions for Mine A, (a) Time series plot of average daily values for segment 1, (b) Scatterplot for segment 1, (c)Time series plot of average daily values for segment 2, (d) Scatterplot for segment 2

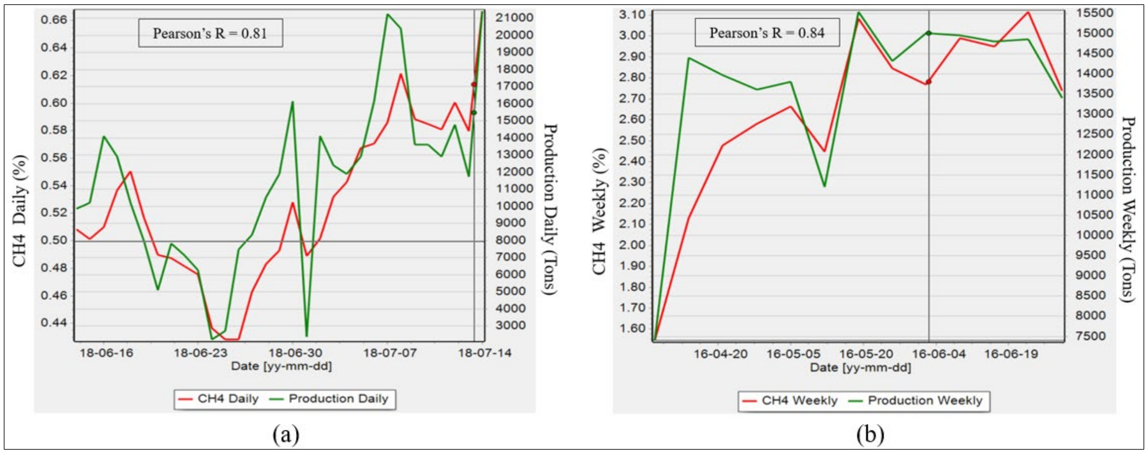

Figure 4.11 Pearson correlation between CH_4 emissions and coal production for Mine A, (a) Time series plot of average daily values, (b) Time series plot of average weekly values

4.6.2 Estimation of Lagged Cross-Correlations

In statistics, the cross-covariance is used to investigate the relation between two time series allowing for time offsets between the two series. The cross-covariance can take positive or negative values; a positive value indicates that the variables tend to move in the same direction, and a negative value signifies that the variables move in opposite directions (Shumway and Stoffer, 2017; Smith, 2021). Such relations can be investigated for different time lags between the two time series. This helps identify if the association between the two variables exhibits a time delay. In this section, the cross-correlation is estimated to investigate the possibility of time-lagged correlations between methane gas emissions and barometric pressure.

Figures 4.12 and 4.13 consist of plots a and b, corresponding to different datasets from Mine A. Both datasets were collected during an interval of 180 days and sampled using a daily average time step. Figure 4.12a presents the cross-correlation function between barometric pressure and methane gas for several time lags (displayed along the horizontal axis). It shows that the highest (negative) correlation between these two variables occurs at zero lag with a value of –0.85. The negative sign indicates that the variables tend to move in opposite directions. Figure 4.12b presents a scatter plot using the same data streams shown in Figure 4.12a. It demonstrates that barometric pressure and methane gas emissions are highly correlated for this specific data segment, with the correlation coefficient calculated at $R=-0.84$.

Figure 4.13a shows the cross-covariance function between barometric pressure and methane gas concentration. Visual examination indicates no significant cross-correlation between the two series for this particular data segment. The scatterplot in Figure 4.13b with the correlation coefficient calculated at $R=0.03$ supports this information.

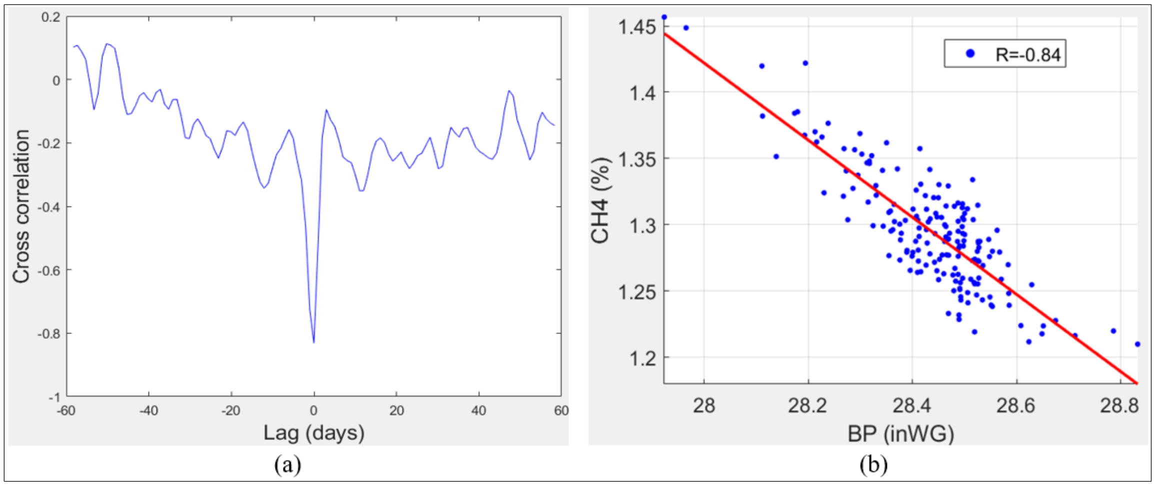

Figure 4.12 Correlation investigation using the empirical cross-correlation function, (a) Cross-correlation between CH_4 and BP, (b) Correlation coefficient (R) between CH_4 and BP

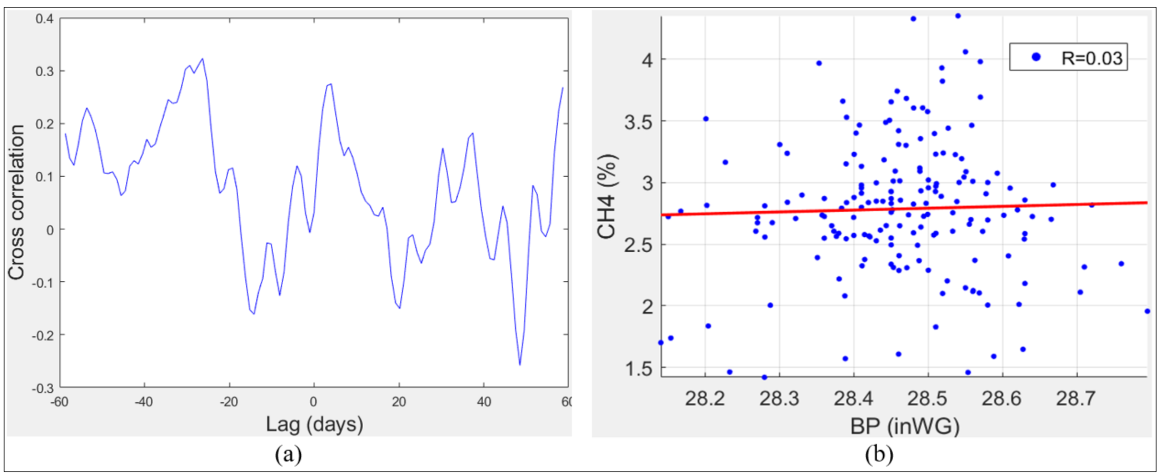

Figure 4.13 Correlation investigation using cross-covariance, (a) Cross-covariance between CH_4 and BP, (b) Correlation coefficient (R) between CH_4 and BP

4.6.3 Autocorrelation of Methane Time Series

Autocorrelation, also known as serial correlation, measures the degree of correlation between a time series and a lagged version of itself. In other words, autocorrelation measures the association between the present value of a variable and its past values. The autocorrelation is technically similar to the correlation between two different time series. However, the autocorrelation uses the same time series twice in its original and lagged forms (Shumway and Stoffer, 2017; Smith, 2021). Autocorrelation function (ACF) plots are among the most popular tools for investigating temporal dependence in stationary time series. A time series is called stationary if; its statistical properties (e.g., mean, median, variance, and autocorrelation) do not change over time. In other words, stationary time series do not have trends or periodic fluctuations (seasonality), and the statistical features (e.g., variance and characteristic time) of fluctuations are invariant in time (NIST, 2003).

Figure 4.14 consists of two plots (a and b) corresponding to Mines A and C, respectively. The blue line in each plot represents measurements of methane gas concentration (%). The time series in Figure 4.14a spans an interval of more than 300 days and is sampled using a daily average time step, while the plot in Figure 4.14b shows methane concentration on a weekly average basis for approximately 120 weeks. Visual inspection of Figures 4.14a and b show downward trends and non-repeating large spikes. Consequently, the methane time series shown in Figures 4.14a and b can be assumed non-stationary. This behavior is typical of the methane gas time series over different time windows analyzed in this research from different mines (Mines A, B, and C), even though certain stationary periods can also be found in the data. As a result, the ACF is not recommended for assessing the autocorrelations of the methane time series. In the case of non-stationary series, the ACF estimated from the data declines very slowly with time due to the non-stationarities. However, the main problem is that the true ACF for non-stationary processes does not depend only on the time lag but also on the time of observation. Therefore, a different statistical measure should be used, which can capture two-point correlations as a function purely of the time lag. The variogram function, used in the next section, provides such a measure.

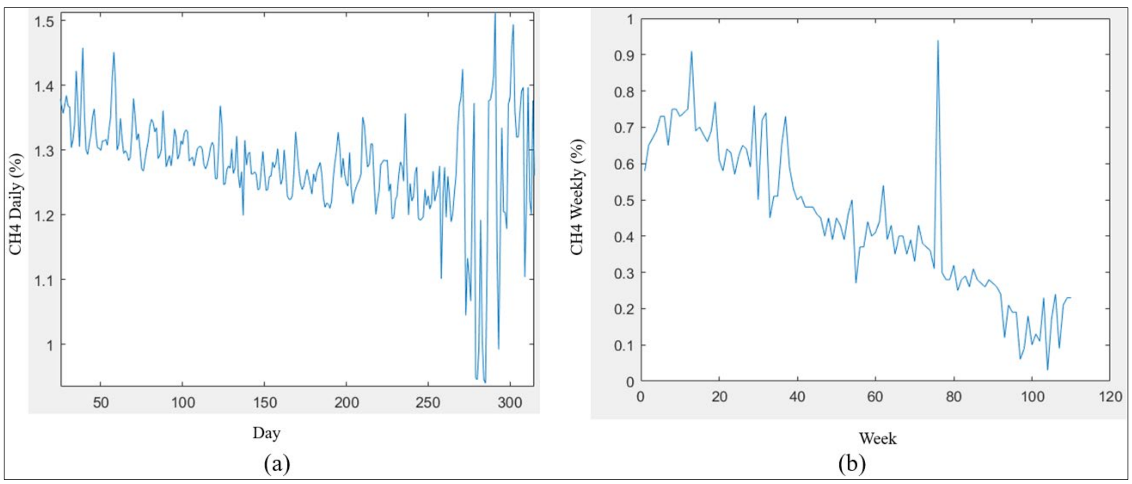

Figure 4.14 Demonstration of non-stationarity of methane time series, (a) Average daily values from Mine A, (b) Plot of average weekly values from Mine C

4.6.4 Variogram Estimation

The variogram function can be used to estimate the variability (degree of similarity or dissimilarity) of time series values at a particular time lag (Shumway and Stoffer, 2017; NIST, 2003; Hristopulos, 2020). More precisely, for a time series denoted by $X(t)$, the variogram is given by the semi-variance of the increment time series $X(t+\tau) - X(t)$ Where τ is the time lag. If $\tau = 0$, the value of the variogram is zero for all t since the increment vanishes. As τ increases, so does the value of the variogram function. If the process is stationary, the variogram attains a plateau (sill). The sill is reached after a characteristic time lag which determines the range of the temporal correlations. If the time series does not have autocorrelations, the variogram jumps from zero to the sill value discontinuously. However, if the process is non-stationary, the variogram continues to increase without bound. Nevertheless, the variogram remains a function of purely the temporal lag for non-stationary processes that have stationary increments. This is a substantial advantage compared to the autocorrelation function.

Another advantage of the variogram is that the differencing operation implied in the calculation of the time series increments tends to eliminate potential short-range increasing or decreasing trends (stochastic trends as they are called). This property is also used in ARIMA time series modeling (see Section 4.7). The variogram function was initially used in studies of fluid turbulence to account for the non-stationarity of fluid velocity in turbulent flows and in geostatistical studies to capture the correlations of non-stationary spatial patterns (Hristopulos, 2020). This section analyzes empirical variograms estimated from mines A and C to investigate potential autocorrelations of the methane gas time series.

Figure 4.15 consists of four variogram plots (a to d); the first two plots (a and b) correspond to Mine A and the last two (c and d) to Mine C. The methane data were collected on a daily and weekly average basis from mines A and C, respectively. The horizontal axis in these plots represents the temporal distance (in days or weeks) between pairs of points, and the vertical axis represents the calculated semi-variance of the methane gas concentration at the respective time lag. All four plots show variogram functions that start at zero and rise, albeit at different rates. For example, in Figure 4.15a, it seems that the sill is reached after ~60 days, while in Figure 4.15b, the variogram function seems to reach a plateau after ~5 days. Both plots use a maximum time lag of 100 days and provide evidence of autocorrelations in the methane concentration series; however, the characteristic time is different for each time series. This is not surprising since the production conditions are not the same for the two time segments investigated.

Figure 4.15c and 4.15d show two variogram functions from Mine C with a weekly average time step over a considerably longer time span of 140 weeks. These plots reveal that the variogram functions seem to stabilize around week 10. However, after lag 10, they start increasing, a tendency that is maintained over the entire range of time lags studied. This behavior is a clear mark of non-stationarity. Hence, the lessons learned from the variogram analysis are that (i) the methane concentration series exhibit autocorrelations, implying that a stochastic predictive model can be constructed, and (ii) the time series may exhibit non-stationarity, thus requiring the use of suitable time series models that allow for the presence of non-stationarities.

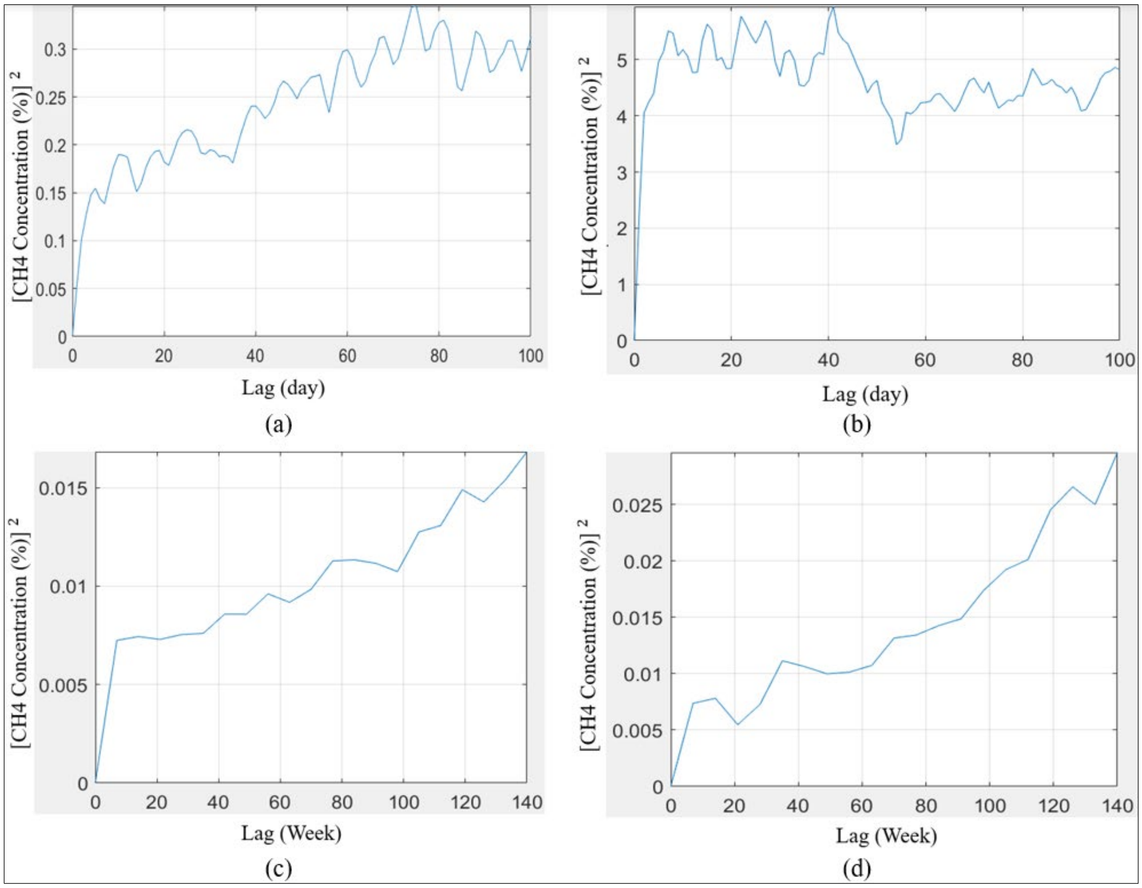

Figure 4.15 Variograms for methane time series, (a) Average daily values from Mine A – segment 1, (b) Average daily values from Mine A – segment 2, (c) Average weekly values from Mine C – segment 1, (c) Average weekly values from Mine C – segment 2

4.7 Time Series Forecasting Models

The main goal of time series analysis is to predict the future values of an observed variable as accurately as possible based on the available data. Forecasting models based on time series analysis are generally classified into three categories; (1) Subjective Forecasting Models founded on judgment, opinion, or intuition, (2) Univariate Forecasting Models which employ past values of a given time series to predict its future value; and (3) Multivariate Forecasting Models based on values of one or more time series to predict the value of a specified variable (Box et al., 2015; Shumway and Stoffer, 2017). Some forecasting techniques are straightforward and computationally efficient (e.g., Mean, Naïve, Seasonal Naïve, and Drift methods); others are more advanced and complex (e.g., Complex Seasonality, Prophet model, and Bootstrapping and Bagging) but offer more flexibility and improved accuracy. The selection of a forecasting method depends on different considerations, such as the accessibility of the historical data, the accuracy of the model, the forecasting context, and the associated time and cost (Hyndman and Athanasopoulos, 2021). The research presented in this article focuses on Autoregressive Integrated Moving Average (ARIMA) models (Box et al., 2015; Shumway and Stoffer, 2017). These are flexible, univariate stochastic models which can handle both stationary and non-stationary time series. Therefore, the ARIMA model can be used to analyze methane gas time series.

4.7.1 ARIMA Model of Methane Concentration

This section focuses on a specific forecasting method based on the ARIMA (p,d,q) model for non-stationary time series. ARIMA is a univariate forecasting technique based on the concept that the past values of a time series can be employed to predict its future values (Box et al., 2015). ARIMA models comprise a linear superposition of time series values at earlier times and a respective superposition of stochastic innovation terms. The innovation terms represent realizations of Gaussian white noise and are responsible for introducing randomness in the model. In addition, ARIMA models are built using time series differences (increments); this procedure helps remove non-stationarities. In general, three integer-valued parameters characterize the ARIMA model: first, the order of the autoregressive (AR) term (p), which indicates the number of lags (past values) that are used as predictors in the model; secondly, the order of the moving average (MA) term (q), which signifies the number of innovation terms included and finally, the order of differencing (d) which is necessary to render the time series stationary. Depending on the complexity of the non-stationarities in the time series, more than one differencing (d) operation may be required. Therefore, the value of d is the minimum differencing order needed for transforming a non-stationary time series into stationary (Box et al., 2015; Shumway and Stoffer, 2017). For more information about the ARIMA models, refer to Shumway and Stoffer. (2017) or Box et al. (2015).

Sample datasets with different lengths (e.g., one month, six months, and one year) and time steps (e.g., 12 hours, daily, and weekly) from Mines A, B, and C were used to construct ARIMA (p,d,q) models in the MATLAB environment (using functions in the econometrics toolbox). There are various ways to assess the performance of a given time series model. Models can be compared concerning measures of fit to the data, such as the Akaike and Bayesian information criteria. Alternatively, they can be compared based on their predictive performance using the approach of cross-validation (CV). The latter evaluates how well the model forecasts compare with reality. There are different approaches for implementing CV. They all partition in some way the dataset into two disjoint sets: the first is called the “training set” and is used to train the model (i.e., to estimate the optimal parameters); the other is called the “validation set” and is used to provide the ground truth against which the model forecasts will be compared. In this research, the partitioning used consists of 95% of the data points in the training set and 5% in the validation set. The predictive accuracy of the model was assessed by employing statistical cross-validation metrics such as the Mean Error (ME), the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), and the Pearson correlation coefficient (R). It was found that the forecast performance of ARIMA (p,d,q) models for methane gas concentration was poor in most of the cases studied; for example, in most cases, the correlation between the validation and the forecast values was low ($R < 0.5$). Consequently, it was concluded that the ARIMA model does not provide accurate forecasts of methane gas concentration over a time horizon involving many future steps.

4.7.2 ARIMA One-Step-Ahead Model of Methane Concentration

In practical situations, it is not often required to forecast the time series for many times ahead. Instead, it suffices to forecast the time series for the next time step (i.e., 12-hour intervals or day). Therefore, in order to evaluate the ability of ARIMA models to provide accurate one-step-ahead forecasts, the following cross-validation methodology is used: (i) The ARIMA model coefficients are estimated using the data in the training set, (ii) the model is used to predict the next value of the time series, implementing a continuously updated dataset: the latter at first involves the point in the training set (e.g., up to time index t; once the forecast at t+1 is generated, the training set is augmented to include the true value of the time series at t+1; using the updated dataset the forecast at t+2 is generated, and so on). Finally, (iii) the one-step-ahead forecasts are compared with the true values in the validation set through CV metrics as described above (Shumway and Stoffer, 2017; Hyndman and Athanasopoulos, 2021).

The optimal ARIMA model was determined using the following algorithm in the MATLAB environment. For each dataset examined, all values of p and q between 1 and 4 and all values of d between 0 and 4 were evaluated. This leads to eighty different ARIMA models estimated using the MATLAB function estimate. In some instances, combinations that correspond to values of d=4 do not produce valid estimates; such models are disregarded. The optimal model is the one that achieves the lowest Akaike Information Criterion (AIC). Then, the optimal model was used to derive one-step-ahead forecasts of the methane concentration, obtained through the MATLAB function forecast. The forecasts are compared with the true values in the validation set by means of the CV measures (e.g., ME, MAE, RMSE, and R), as shown in Table 4.1. The function forecast also estimates the mean square error (MSE) of the prediction. The MSE is then used to generate 95% prediction intervals given by Eq. (4.1). Prediction intervals are essential for two reasons: (i) they allow an assessment of the precision of the forecast, and (ii) if the forecasts deviate from the true values, it permits identifying if the true values are at least contained within the prediction intervals.

$$\left[\hat{x}(t_p) - 1.96\sqrt{\mathrm{MSE}(t_p)}, \hat{x}(t_p) + 1.96\sqrt{\mathrm{MSE}(t_p)}\right] \tag{4.1}$$

where $\hat{x}(t_p)$ is the optimal ARIMA prediction, MSE is the mean square error of the prediction, and 1.96 is a value used to obtain the 95% prediction intervals.

Given the irregular variations of the methane time series, four different nonlinear transformations (logarithm, square root, inverse, and inverse square root) were applied to the methane time series. These transformations were used to stabilize the variance and mitigate potential heteroscedasticity effects (i.e., the dependence of the local variance on the local mean). Furthermore, the time series analysis described above was applied to each resulting (transformed) time series. Moreover, at the end of each calculation, the forecasts of the transformed data need to be inverted to the original domain, which is a straightforward step by invoking the conservation of the probability of random variables under nonlinear monotone transformations (Hristopulos, 2020). However, analyzing the forecasts based on these transformations, it was found that they only marginally improved the CV metrics of the untransformed time series in the best cases. Consequently, the following ARIMA modeling focuses on the untransformed data. Examples of the one-step-ahead forecast of methane gas concentration obtained using the optimal ARIMA model are presented below.

Figures 4.16 and 4.17 consist of four plots each (a to d). Graphs (a) and (c) are the forecasts obtained using the ARIMA (p,d,q) one-step-ahead model, while plots (b) and (d) are magnified versions of the forecasts shown in graphs (a) and (c) respectively. The methane time series data used to generate these figures were retrieved from Mine A. The gray line represents the training data (methane gas time series), the upper and lower black dashed lines signify the boundaries of the 95% Confidence Interval (C.I.), the blue line indicates the time series in the validation set, while the red line represents the forecast. The validation and forecast periods contain five percent (5%) of points used in the training time series.

Figure 4.16a shows the forecast obtained for a time series of methane gas concentrations spanning 360 days using a daily average time step. The ARIMA (4,1,4) model was selected based on the lowest AIC value. As a result, the autoregressive order of the AR(p) term is four (p=4), the order of the differencing (d) is one (d=1), and the order of the MA(q) term is four (q=4). Figure 4.16b demonstrates that the one-step-ahead forecast (red line) is quite close to the true value during the validation period (blue line); the correlation coefficient calculated was R=0.89, implying a strong correlation between the validation data and the forecasts. Therefore, the ARIMA (4,1,4) model provides an accurate forecast. Moreover, the observed values lie within the 95% prediction interval.

Figure 4.16c displays the one-step-ahead forecast for a time series of methane gas concentrations spanning more than 2100 days (six years) with a daily average time step. In this case, the ARIMA (3,1,4) model is selected (the autoregressive order is 3, the differencing order is 1, and the moving average order is 4). Figure 4.16d demonstrates a strong correlation between the validation data and the forecast, with the correlation coefficient evaluated at R=0.65. This value is satisfactory considering that the training data are far from ideal (e.g., they include zero values and some sharp fluctuations), as shown in Figure 4.16c. Nonetheless, the ARIMA (3,1,4) one-step-ahead tracks the peaks of the validation data reasonably accurately, while the validation data lie within the 95% prediction interval. Furthermore, the cross-validations metrics, including the root mean square error obtained for the one-step-ahead forecast presented in Figure 4.16c (RMSE=0.05), are significantly lower than in Figure 4.16a (RMSE=0.54), as shown in Table 4.1, which means that the second forecast is closer to the true values.

Figure 4.17 shows the one-step-ahead forecast for the gas concentration time series for the same period presented in Figure 4.16 but using a twelve-hour average time step. The ARIMA (4,1,4) is again the best model. Figure 4.17b reveals that the forecasts closely follow the validation data; the correlation coefficient calculated was R=0.90, higher than the correlation achieved with the daily average samples (cf. Figure 4.16). Furthermore, Figure 4.17c presents the one-step-ahead forecast for a time series spanning more than 2100 days (six years), corresponding to more than 4300 data points. Again, the ARIMA (4,1,4) remains the optimal model. Visual inspection of Figure 4.17d indicates that the forecast (red line) and the validation data (blue line) are notably similar. The correlation coefficient was calculated at R=0.71, significantly higher than the respective R for the time series shown in Figure 4.16c. In addition, Table 4.1 shows that the CV metrics presented in Figure 4.17c are significantly lower than in Figure 4.17a, which means that the second forecast provides a higher approximation to the true values.

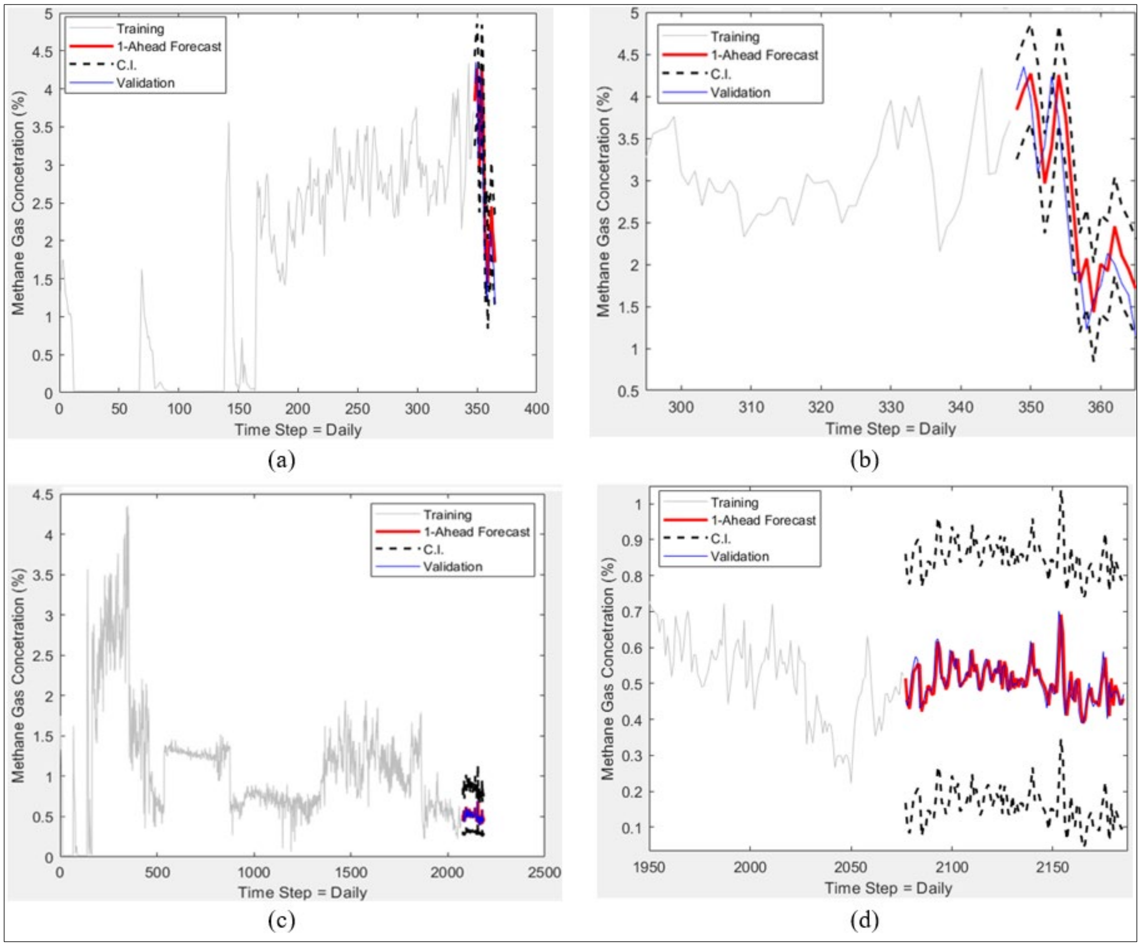

Figure 4.16 ARIMA one-step-ahead CH_4 concentration forecasts using a daily average time step;(a) Forecasting of segment 1,(b) Magnified view of the forecast in (a), (c) Forecasting of segment 2,(d) Magnified view of the forecast in (b)

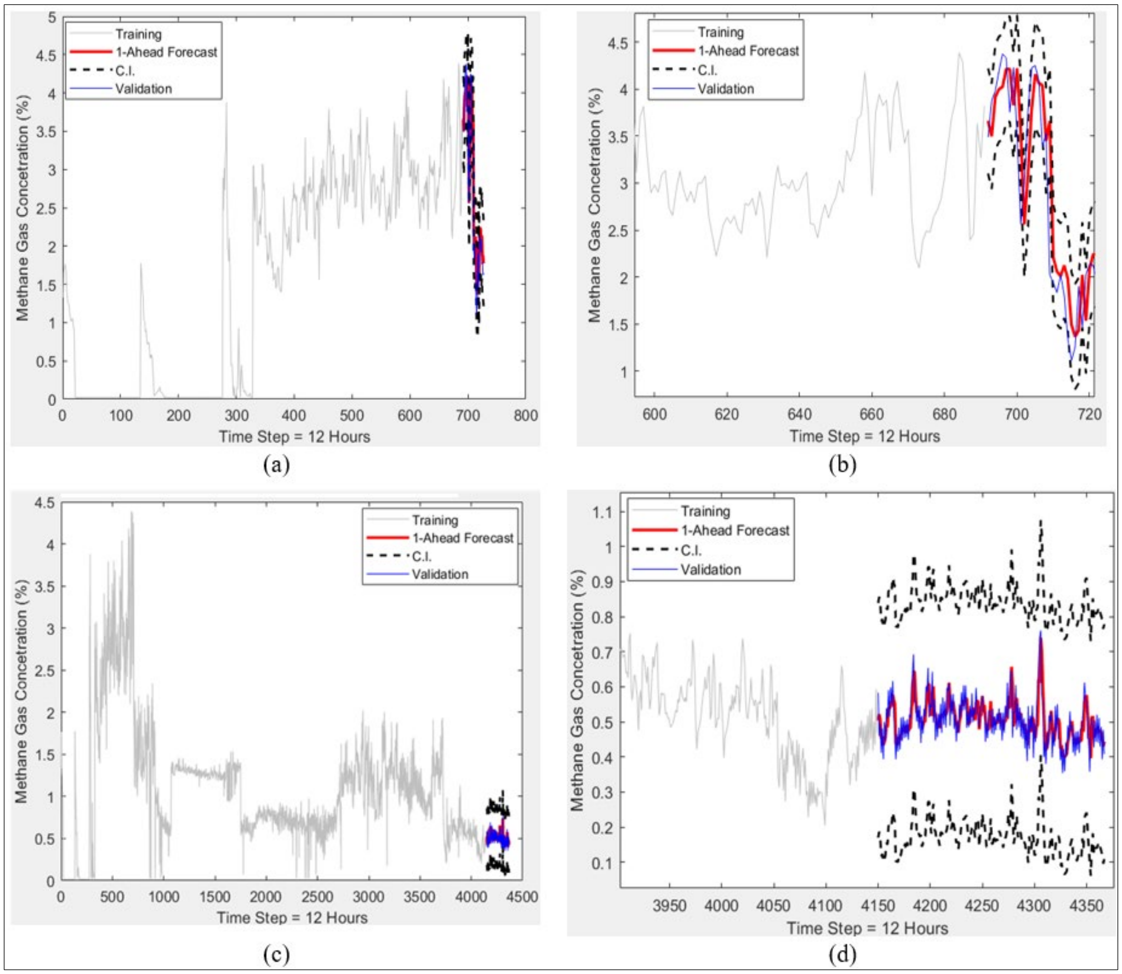

Figure 4.17 ARIMA one-step ahead CH_4 concentration forecasting using a 12-hour average time step, (a) Forecasting of segment 1,(b) magnified view of the graph a,(c) Forecasting of segment 2,(d) Magnified view of the graph b

Table 4.1 One-step-ahead forecast results summary

	Time Step	Training Data Length (days)	Training Data Sample Size	Validation Data Length (days)	Validation Data Sample Size	Best ARIMA (p,d,q) Model	Correlation Coefficient (R)	RMSE	ME	MAE
Forecast 1	Daily average	365	365	18	18	(4,1,4)	0.89	0.54	−0.19	0.44
Forecast 2	Daily average	2,200	2,200	109	109	(3,1,4)	0.65	0.05	0.00	0.03
Forecast 3	12 Hours average	365	730	18	37	(4,1,4)	0.90	0.47	−0.11	0.35
Forecast 4	12 Hours average	2,200	4,380	109	219	(4,1,4)	0.71	0.05	0.00	0.04

4.8 Discussion

The research presented in this paper has found that methane gas forecasting methods based on time series analysis are less costly and time-consuming than empirical and numerical forecasting methods basically because the data collection process is more rapid and reliable.

The atmospheric monitoring system employed in each case study (Mines A, B, and C) was described. Mine A uses an automated AMS known as Wireless Multi-Gas Monitor (see Figure 4.2) installed on the exhaust shafts. This device can simultaneously monitor up to four gases. It provides remote operation through a Wi-Fi connection; no instruments or special skills are required to replace sensor modules, and the computer software is updated automatically. Mine B employs an automated AMS that collects gas concentration data from different sensors throughout the mine. These data are electronically transmitted to a central monitoring system on the surface for further processing. Finally, methane gas data for Mine C are calculated weekly at the exhaust shaft(s) using a manual process.

The main steps of the data management process used in this research were explained. First, methane gas concentrations were collected from the AMS of three underground coal mines, and meteorological data were retrieved from the weather stations nearest to each mine using the WU website. Then, the atmospheric and methane gas data were stored into AMANDA, which is also used for data pre-processing. The next step is data homogenization which is performed by means of AMANDA and MATLAB®. The final step includes several statistical procedures run either on the raw or the homogenized data to assess the potential autocorrelation and cross-correlation of the studied variables such as methane gas, coal production, and atmospheric pressure parameters. Data homogenization is an essential step when evaluating time series data. It ensures that the records collected from methane gas and barometric pressure time series have a common date/time stamp, which is required to evaluate the potential correlation between the variables.

The data pre-processing processes, including data cleaning and filtering, are also discussed herein. The examples provided in Section 4.5.2 demonstrate that historical data like the methane gas and barometric pressure time series such as the one analyzed in this research regularly present inconsistencies, such as erroneous and/or negative values and gaps (Figures 4.6 and 4.7) and calibration spikes (Figure 4.8). However, some methane gas concentration peaks (spikes and inverted spikes) were found to be caused by changes in independent variables that directly affect the emissions and concentrations of methane gas, like for example, a substantial increase in coal production, most likely due to coal recovery in more than one panel simultaneously (as shown in Figure 4.9). Consequently, inspecting, identifying, and filtering out anomalous records in the database is essential to guarantee data consistency and integrity.

Statistical dependence measures such as cross-correlation, autocorrelation, cross-covariance, and variograms were implemented to investigate potential associations and validate long-term relationship(s) between the dependent variable (e.g., methane gas emissions) and the independent variables (e.g., meteorological parameters and coal production rate) for the different case studies. In addition, the Pearson correlation coefficient was selected to investigate the relationship between the independent and dependent variables. It was determined that methane gas and coal production rate exhibit a strong positive correlation: when coal production rates increase, methane gas concentration increases for most cases. In contrast, the correlation between methane gas concentration and barometric pressure is significant but negative: methane gas decreases when barometric pressure increases and vice versa. Nevertheless, it was found that for some data segments, the correlation between these two variables (CH_4 vs. BP) was weak; in some cases, the correlation coefficient was zero (R=0.00), which can be explained due to the presence of inconsistent records in the methane time series such as spikes and inverted spikes most likely due to sensor calibration, sensor failure or independent variable(s) (e.g., coal production rate) directly affecting methane gas emissions and barometric pressure correlation.

The cross-correlation function between methane gas concentration and barometric pressure shows that the highest association occurs at lag zero, and the negative sign of the cross-correlation indicates that methane gas concentration and barometric pressure have an inverse relationship. Again, for some data segments, no significant cross-correlation was detected. It was complex to interpret the autocorrelation of the methane gas concentration from Mines A and C using the ACF plot due to the non-stationary nature of the time series. Instead, the variogram function, which is more suitable for non-stationary data, was assessed, and it revealed both short-range correlations and long-range stochastic trends on time scales that vary between datasets.

The research presented in this paper proposed a methane gas concentration forecasting method based on the ARIMA (p,d,q) one-step-ahead model. Methane gas time series from Mine A spanning different lengths (e.g., one year and six years) and using different time steps (e.g., daily and every 12 hours) were employed to estimate the optimal (among different choices of p, d, q values) ARIMA model. The optimal model was obtained by running suitable MATLAB® code. In all of the cases evaluated and presented in Section 4.7.2, the ARIMA one-step-ahead model provides accurate forecasts that match the direction (increase/decrease) of the validation data. In addition, the correlation between the forecasts and the data in the validation period was strong and positive. Moreover, the observed values of methane gas concentration were always captured by the 95% prediction interval.

It was also established that the forecasting model is improved (a higher correlation between the forecast and the validation data is achieved) by using longer methane time series (six years) than shorter ones (1 year) to train the ARIMA model (see Table 4.1). Furthermore, the methane time series collected with the 12-hour average time step provides a more accurate forecast than the daily average methane time series. It can be explained since the methane time series that uses a 12-hour average time step contains more information, and the one-step-ahead forecast refers to a time instant that is closer to the training data than in the case of the daily average step.

4.9 Conclusions

The research presented in this paper investigated the correlation between methane gas emissions, barometric pressure, and other variables (e.g., coal production) with the ultimate goal of designing a forecasting model based on time series analysis that can help prevent accidents and fatalities due to methane gas explosions in underground coal operations.

Data pre-processing and homogenization are fundamental steps for ensuring data consistency and integrity of the time series that will be fed to the forecasting model. For example, it has been demonstrated that methane gas data collected and stored by atmospheric monitoring systems in underground coal mines contain inconsistencies that compromise further data processing steps, directly impacting the outcomes of statistical tests and the accuracy of forecasting methods.

This research identified (i) a significant negative correlation between methane gas and barometric pressure and (ii) the presence of autocorrelations in the methane gas time series. The latter was used to build a methane gas forecasting method based on the ARIMA (p,d,q) model. As a result, accurate one-step-ahead predictions of methane gas concentration for different case scenarios were obtained, where the one-step implies either 12 or 24 hours after the last recorded time. Furthermore, the accuracy of the predictions was established using cross-validation analysis. Therefore, the ARIMA one-step-ahead methane gas forecasting method presented in this paper, coupled with atmospheric monitoring system, provides a solution that could improve the health and safety conditions in underground coal mines and different underground operations. However, the performance of the forecasting model needs to be further assessed with different datasets and conditions. Also, at this point, the data pre-processing and homogenization require expert human intervention and judgment.

The focus of future work will be to (a) assess the performance of the forecasting model by employing an extended set of statistical cross-validation measures to conduct residual analysis, (b) characterize the uncertainty of the forecasting model by calculating the statistical performance of the prediction intervals and (c) determine the ARIMA model's precision and accuracy for different periods and operating conditions.

4.10 Acknowledgments

This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION), contract number AFCTG20-103. The views, opinions, and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its Directors, and staff.

4.11 Statements and Declarations

Author Contributions: Supervision, Z.A.; Writing—original draft, J.D.; Writing—review & editing, Z.A., D.H, S.S, KL. All authors have read and agreed to the published version of the manuscript.

Funding: This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION), contract number AFCTG20-103. The

views, opinions, and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its Directors, and staff.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable as data is proprietary.

4.12 References

Airey EM (1968) Gas emission from broken coal. An experimental and theoretical investigation. Int. J. Rock Mech. Min. Sci. Geomech., vol. 5:475–494.

AMR PEMCO (2002) Multi-Gas Monitor. https://www.amrpemco.com/mc-6410-multi-gas-monitor/. Accessed October 10, 2021.

Astudillo G, Carrasco R, Fernández-Campusano C, Chacon M (2020) Copper price prediction using support vector regression technique Appl. Sci. https://doi.org/10.3390/app10196648.

Baur J, Moreno-Villanueva M, Kötter T, Sindlinger T, Bürkle A, Berthold N, Junk M (2015) MARK-AGE data management: Cleaning, exploration and visualization of data. Mech. Ageing Dev. https://doi.org/10.1016/j.mad.2015.05.007.

Bilal E, Arnaout A, Fruhwirth RK, Thonhauser G (2013) A Statistical feature-based approach for operations recognition in drilling time series. Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 5:454–461.

Booth P, Brown H, Nemcik J, Ting R (2017) Spatial context in the calculation of gas emissions for underground coal mines. Int. J. Min. Sci. Technol. https://doi.org/10.1016/j.ijmst.2017.07.007.

Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis forecasting and control, 5th (ed). John Wiley & Sons. New Jersey, United States.

Brockwell PJ, Davis RA (2016) Introduction to time series and forecasting, 3rd (ed). Springer. Switzerland.

Broquet G, Pison I, Abbaris A et al. (2018) Identification of spikes associated with local sources in continuous time series of atmospheric CO, CO_2, and CH_4. Atmos. Meas. Tech. https://doi.org/10.5194/amt-11-1599-2018.

Byrer C, Havryluk I, Uhrin D (2014) Coalbed methane: a miner's curse and a valuable resource. In: Pramod T, Schatzel S, Aminian K (ed) Coalbed methane: from prospect to pipeline, 1st ed. Elsevier, Amsterdam, pp 1–6.

Chaulya SK, Prasad GM (2016) Sensing and monitoring technologies for mines and hazardous areas, 1st (ed). Amsterdam, Netherlands.

Cryer JD, Chan KS (2008) time series analysis with applications in R, 2nd (ed). Springer. New York, United States.

Curl SJ (1978) Methane prediction in coal mines (IEA Coal Research Report), London, United Kingdom.

Diaz J, Agioutantis Z, Hristopulos DT, Schafrik S (2021) Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. Mater. Proc. 2021. https://doi.org/10.3390/materproc2021005078.

Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021a) Towards atmospheric monitoring data analysis in underground coal mines. In: Proceedings of the 18th North American Mine Ventilation Symposium (NAMVS). https://doi.org/10.1201/9781003188476-51.

Dixon DW (1992) A statistical analysis of monitored data for methane prediction. Dissertation, University of Nottingham. Nottingham, United Kingdom.

Dixon DW, Longson I (1993) Statistical method for methane prediction and improved environmental control. In: Proceedings of the 6th North American Mine Ventilation Symposium (NAMVS), pp. 94–98. Salt Lake City, Utah, United States.

Düzgün HS, Leveson N (2018) Analysis of soma mine disaster using causal analysis based on systems theory (CAST). Saf. Sci. https://doi.org/10.1016/j.ssci.2018.07.028.

Flores RM (1998) Coalbed methane: From hazard to resource. Int. J. Coal Geol. https://doi.org/10.1016/S0166-5162(97)00043-8.

Goin DE, Ahern J (2019) Identification of spikes in time series. Epidemiol. Method. https://doi.org/10.1515/em-2018-0005.

Goodman GVR, Karacan CÖ, Schatzel SJ, Krog RB, Taylor CD, Thimons ED (2008) NIOSH Research for monitoring and controlling methane at US, underground coal mining operations. In: Proceedings of the 21st World Mining Congress. Krakow, Poland.

Griffin R (2013) Utilization and implementation of atmospheric monitoring systems in United States underground coal mines and application of risk assessment. Dissertation, The Virginia Polytechnic Institute, and State University. West Virginia, United States.

Griffin R, Jong EC, Luxbacher KD, Westman EC (2011) A review of atmospheric monitoring systems in underground coal mines: implications for explosion prevention. In: Proceedings for the Society for Mining, Metallurgy, and Exploration (SME), Preprint 11-127, pp. 683–686. Denver, Colorado, United States.

Grimberg F, Asprion PM, Schneider B, Miho E, Babrak L, Habbabeh A (2021) the real-world data challenges radar: a review on the challenges and risks regarding the use of real-world data. Digit. Biomarkers, https://doi.org/10.1159/000516178.

Gupta N, Tran D, Devegowda V, Jayaran C, Rai C, Sondergeld C, Karami H (2020) Looking ahead of the bit using surface drilling and petrophysical data: machine-learning-based real-time geosteering involve field. SPE J. https://doi.org/10.2118/199882-PA.

Harkness T (2020) John Graunt at 400. Fighting disease with Numbers. J. Royal Statistical Society. https://doi.org/10.1111/1740-9713.01421.

Hristopulos DT (2020) Random fields for spatial data modeling: A primer for scientists and engineers. Springer Nature, Dordrecht.

Hyndman RJ, Athanasopoulos G (2021) Forecasting: principles and practice. https://otexts.com/fpp3/. Accessed January 10, 2022.

Jorgensen S, Fath B (2019) Chaos. Encycl. Ecol. https://doi.org/10.1016/B978-008045405-4.00148-8.

Karacan CÖ, Diamond WP, Esterhuizen GS, Schatzel SJ (2005) Numerical analysis of the impact of longwall panel width on methane emissions and performance of gob gas ventholes. National Institute for Occupational Safety and Health (NIOSH). https://search.cdc.gov/search. Accessed October 01, 2021.

Kowalski-Trakofler DW, Alexander KM, McWilliams LJ, Reissman DB (2009) Underground coal mining disasters and fatalities-Unites States, 1900-200. MMWR vol. 57:1379–1382. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5751a3.htm. Accessed November 02, 2021.

Kozlov P (2021) Russian coal mine: dozens killed in Siberia accident, BBC News. https://www.bbc.com/news/world-europe-59421319. Accessed December 01, 2021.

La República (2021) Doce muertos dejó una explosión en mina de carbón en Boyacá, a 190 km de Bogotá. Diario La República. https://www.larepublica.co/economia/doce-muertos-dejo-una-explosion-en-mina-de-carbon-en-boyaca-a-190-km-de-bogota-3222133. Accessed December 01, 2021.

National Institute of Standards and Technology (NIST) (2003) Engineering Statistics Handbook. https://www.itl.nist.gov/div898/handbook/. Accessed December 07, 2021.

Nielsen A (2019) Practical time series analysis. Prediction with statistics & machine learning preview edition, 1st (ed). O'Reilly Media Inc. San Francisco, United States.

Olayiwola A (2016) Forecasting copper spot prices: A knowledge-discovery approach. Dissertation, The University of Manchester, United Kingdom.

Owili-Eger A, Stefanko S, Ramani RV (1973) Simulation of quantity and quality control in mine ventilation. Special Research Report SR-95. The Pennsylvania State University. Pennsylvania, United States.

Palshikar GK (2009) Simple algorithms for peak detection in time series. In: Proceedings of 1st International Conference on Advanced Data Analysis (IIMA), July 2009. Ahmedabad, Gujarat, India.

Renner S, Wellmer FW (2020) Volatility drivers on the metal market and exposure of producing countries. Miner. Econ. https://doi.org/10.1007/s13563-019-00200-8.

Rodríguez SP (2020) Modeling and forecasting complex patterns of mineral production. Gold mining in Canada. Cogent Eng. https://doi.org/10.1080/23311916.2020.1724849.

Rosienkiewicz M, Chlebus E, Detyna J (2017) A hybrid spares demand forecasting method dedicated to the mining industry. Appl. Math. Model https://doi.org/10.1016/j.apm.2017.04.027.

Schober P, Schwarte LA (2018) Correlation coefficients: Appropriate use and interpretation. Anesth. Analg. https://doi.org/10.1213/ANE.0000000000002864.

Shengrui Z, Baisheng N, Shuiwen L, Hui W, Qian L, Hailong L (2011) Model of gas concentration forecast based on chaos theory. Procedia Eng. https://doi.org/10.1016/j.proeng.2011.11.2160.

Shu-gang C, Yan-bao L, Yan-Ping W (2008) A forecasting and forewarning model for methane hazard in working face of coal mine based on LS-SVM. J. China Univ. Min. Technol. https://doi.org/10.1016/S1006-1266(08)60037-1.

Shumway RH, Stoffer DS (2017) Time series analysis and its applications, 4th (ed). Springer International Publishing. Cham, Switzerland.

Smith T (2021) Autocorrelation.
ttps://www.investopedia.com/terms/a/autocorrelation.asp. Accessed December 04, 2021.

Sung W, Ertekin T, Ramani RV (1987) Evaluation of the impact of the degasification process on the inflow of methane gas into coal mines: A numerical exercise. In: Proceedings of the Society for Mining, Metallurgy & Exploration (SME), pp. 328–333. The United States of America.

Tapia Cortez CA, Saydam S, Coulton J, Sammut C (2018) Alternative techniques for forecasting mineral commodity prices. Int. J. Min. Sci. Technol. https://doi.org/10.1016/j.ijmst.2017.09.001.

Taylor C, Karacan CÖ (2010) Historical development of technologies for controlling methane in underground coal mines. In: Extracting the Science: A century of mining research. Brune JF (ed), Littleton, CO, Pittsburgh, United States.

The National Institute for Occupational Safety and Health (NIOSH) (2020) NIOSH data and statistics. https://www.cdc.gov/niosh/injury/data.html. Accessed November 02, 2021.

Thomas S (2014) Basic Statistics, 1st (ed). Alpha Science International Ltd. Oxford, United Kingdon.

Tominaga Y, Bandopadhyay S (2002) Monitoring of spontaneous combustion based on time series data. In Proceedings of the 9th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781439833742.ch48.

Trutwin W, Wasilewski S (1987) Application of digital filter and time series analysis in ventilation control systems. In: Proceedings of the 5th North American Mine Ventilation Symposium (NAMVS), pp. 534–540. Morgantown, West Virginia, United States.

Tsay RS (2000) Time series and forecasting: Brief history and future research. J. Am. Stat. Assoc. https://doi.org/10.1080/01621459.2000.10474241.

Verma U (2019) Data cleaning and pre-processing. Analytics Vidhya. https://medium.com/analytics-vidhya/data-cleaning-and-preprocessing-a4b751f4066f. Accessed November 02, 2021.

Vlachos M, Meek C, Vagena Z, Gunopulos D (2004) Identifying similarities, periodicities, and bursts for online search queries, In Proceedings of the 2004 ACM SIGMOD international conference on Management of Data – SIGMOD. https://doi.org/10.1145/1007568.1007586.

Watari T, Nansai K, Nakajima K (2020) Major metals demand, supply, and environmental impacts to 2100: A critical review. Resour. Conserv. Recycl. https://doi.org/10.1016/j.resconrec.2020.105107.

Yule UG (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfer's Sunspot Numbers. Philos. Trans. R. Soc. London, vol. 226:167–298.

Zagorecki A (2015) Application of sensor fusion and data mining for prediction of methane concentration in coal mines. Mining-Informatics, Automation, and Electrical Engineering ed. vol. 53, nr 4, pp. 33–38.

Zhou C, Damiano N, Whisner B, Reyes M (2017) Industrial Internet of Things (IIoT) applications in underground coal mines. Min. Eng. https://doi.org/10.19150/me.7919.

5. MULTIVARIATE FORECASTING APPROACH FOR METHANE GAS CONCENTRATIONS IN UNDERGROUND COAL MINES

The multivariate methane gas forecasting approaches developed and proposed by this research and their comparison with the univariate forecasting approach presented in Chapter 4 is covered by the following peer-reviewed article that will be submitted to Process Safety and Environmental Protection Journal.

FORECASTING OF METHANE GAS IN UNDERGROUND COAL MINES: UNIVARIATE VS. MULTIVARIATE MODELING APPROACHES

Juan Diaz, University of Kentucky; Lexington, Kentucky; United States of America

Zach Agioutantis, University of Kentucky; Lexington, Kentucky; United States of America

Steven Schafrik, University of Kentucky; Lexington, Kentucky; United States of America

Dionissios Hristopulos, Technical University of Crete; Chania, Crete; Greece

Kray Luxbacher, Virginia Polytechnic Institute and State University; Blacksburg, Virginia; United States of America

Full Citation:

Diaz J, Agioutantis Z, Hristopulos DT, Luxbacher K, Schafrik S (2022) forecasting of methane gas in underground coal mines: univariate vs. multivariate modeling approaches. Process Safety and Environmental Protection Journal. *(Submitted).*

5.1 Abstract

Coal mining operations provide the coal required to satisfy more than 36% of the electricity demand worldwide, making coal the most abundant fuel source used for electricity generation. As coal is mined, methane gas is released, which can constitute a significant threat to the health and safety of underground coal miners. Therefore, methane gas concentrations need to be monitored and managed to prevent hazards. This paper introduces a new framework to forecasting methane gas emissions in underground coal mine operations by analyzing time series data using univariate and multivariate forecasting techniques. The methane time series data used during this research have been collected from the Atmospheric Monitoring System of three active underground coal mines in the eastern US. Methane gas forecasting is initially assessed based on a univariate time series model (ARIMA). Then, multivariate time series models (VAR and ARIMAX) are investigated, including barometric pressure in the forecasts. The optimal model for each family (ARIMA, VAR, ARIMAX) is selected based on the Akaike Information Criterion. Finally, the results obtained from the different forecast approaches (univariate and multivariate) are compared using cross-validation metrics to determine the best overall model. It is concluded that the ARIMA, VAR, and ARIMAX methane gas forecasting methodologies proposed in this research can successfully predict methane gas concentrations in underground coal mines in most cases. Keywords: univariate and multivariate forecasting, time series analysis, methane gas, underground coal mines.

5.2 Introduction

Modern life is unthinkable without electrical energy. It provides domestic and industrial heat and power to equipment and technology employed in factories, hospitals, buildings, and homes. Coal is the most abundant fuel source used for electricity generation as it produces more than 36% of global energy. Coal mines are responsible for supplying approximately 25% of the electricity in the United States (SME, 2021; WCA, 2022). In addition, metallurgical coal is a necessary component in steel manufacturing. Furthermore, coal and coal by-products are implemented in numerous areas to produce goods such as carbon fiber for the automotive and aviation industries, activated carbon for purified air systems and water treatment plants, synthetic petroleum-based fuels, and medicine (Wei et al., 2015). However, as coal is mined, methane gas (CH_4) trapped between coal particles is emitted (Flores, 1998; Kissell, 2006). Methane gas concentrations ranging between 5% and 15% can create explosive mixtures in mine atmospheres that, when ignited, can result in explosions with catastrophic consequences (Kissell, 2006; NIOSH, 2020). Therefore, underground coal mines need to monitor and manage methane concentrations to guarantee the health and safety conditions of their personnel (Agioutantis et al., 2014; Diaz et al., 2021). The prediction of methane gas concentration would tremendously help that effort.

Different methane gas concentration forecast models have been discussed in the literature; most are based on empirical, numerical, and statistical approaches (Diaz et al., 2021). For example, Airey (1968) proposed an empirical methodology to quantify methane gas emissions from coal excavated during underground coal mining operations and the factors that influence their generation. Furthermore, Dunmore (1982) developed an experimental method to forecast methane gas emissions in underground longwall mines in the United Kingdom, based on Airey's theoretical analysis of gas emissions from coal seams (Dixon, 1992). This research focused on the geological characteristics of coal seams (e.g., thickness, depth, and gas content) and mine operation parameters such as coal production rate and extraction method. The author concluded that the accuracy of the proposed model is significantly affected by the particular geological conditions of each underground coal mine operation.

Kirchgessner et al. (1993) recommended a method to estimate methane gas emissions from underground coal mining operations. The authors developed a regression equation that reasonably predicts methane emissions based on three main variables: coal production rate, mine emissions, and coalbed methane concentration. Furthermore, it was identified that some characteristics of the mine, such as the depth, pressure, and humidity, are the main parameters directly influencing coalbed methane gas concentration. In a different work, Schatze et al. (2008) proposed a methodology for predicting methane emissions when longwall panel dimensions increase, especially face lengths. This research was conducted in a longwall mine operating at the Pittsburgh seam in Pennsylvania (US). Airflow and methane concentration was measured using methane sensors along the longwall face. The longwall face was divided into three equal segments. The authors assumed that the frequency of production delays and the mine advance rate were equivalent, methane gas emissions in the longwall panel were continuous when the panel was mined, and all potential sources of methane gas (e.g., methane gas in the coal seam) changed at a constant rate with respect to increasing face length. It was concluded that monitoring mine data (methane gas time series) could be used to predict methane concentrations when the coal panel dimensions change.

A different approach to forecasting methane gas emissions has been proposed by several researchers based on computational fluid dynamics. For example, Owili-Eger et al. (1973) developed a mathematical model for predicting methane gas emissions in the coal seams and the mining atmosphere based on the physics of gas flow through a coal seam which was solved numerically. The authors used a modified gas diffusion flow model through porous media. It assumed constant flow of gases, minimum temperature changes over the medium, the dependence of directional permeabilities only on pressure and location, and methane gas flow along the coal seam except for production and injection wells. The applications of this mathematical model were demonstrated through a hypothetical example. However, further research proved that the model is imprecise when assessing methane gas concentration for deep underground coal operations (Dixon, 1992). Guo et al. (2008) recommended a three-dimensional numerical model to forecast mine gas emission as well as rock mass deformation, water inflow, and mine stability in an underground longwall coal mine. This numerical model used a combined 3D mechanical deformation and dueled porosity multiphase flow finite element (COSFLOW). The authors indicated that COSFLOW has unique characteristics because it incorporates Cosserat's continuum theory, which allows a compelling description of mechanical deformation in weak layered rock and stress changes. It was concluded that the model could predict methane gas emissions and concentrations in the longwall panel reasonably accurately. However, transient fluctuations were found in the measurements, most likely generated by local variations in geology or gas content or independent variables not incorporated into the model.

The main disadvantage of the forecasting approaches mentioned above (empirical and numerical methane forecast models) is the difficulty in their implementation due to the significant number of empirically established parameters influencing methane gas concentration; as a result, misspecification of the parameters can lead to inaccurate prediction models (Diaz et al., 2021; Booth et al., 2017; 2016). Scientists have investigated statistical approaches, particularly time series analysis, to forecast methane gas in underground coal mines in recent years. Statistical approaches take advantage of inherent correlations in time series, which play the role of memory. Thus, if the correlations can be modeled, past information can be used to formulate future probabilistic predictions. In particular, the Autoregressive Integrated Moving Average (ARIMA) models can forecast methane gas concentration by modeling the autocorrelation of methane concentration time series (Diaz et al., 2021; Dixon, 1992; Dixon and Longson, 1993). Other research has demonstrated that methane gas concentration and emissions are directly influenced by other independent variables such as barometric pressure and coal production rate (Hemp, 1994; Xu et al., 2014; Lolon, 2017; Wasilewski, 2014; Yuan et al., 2007). For example, it has been determined that in most cases, methane gas has a strong negative correlation with barometric pressure: methane gas concentration increases when barometric pressure decreases and vice versa. In contrast, the correlation between methane gas and coal production rate generally is positive: methane gas concentration increases when coal production rate increases and vice versa (Diaz et al., 2021; 2022).

The research presented in this paper addresses the lack of methodologies that can be used for accurately forecasting methane gas concentrations in underground environments. This investigation is one of the first documented attempts to collect, process, and analyze historical mine data (e.g., methane gas concentrations and coal production rate) from automated Atmospheric Monitoring Systems of underground coal mines and weather data from public weather stations to develop different methane gas concentration forecasting models based on univariate and multivariate time series approaches and evaluate their performance using cross-validation metrics. Cross-validation allows determining the best forecast model among different model families for each specific dataset. The present work aims to enable a better understanding concerning (a) forecast methods for predicting methane gas concentrations and emissions in underground environments, especially in underground coal mines, leading to improved safety and health conditions of the workforce, and (b) identify research gaps in this field which should encourage new studies.

5.3 Materials and Methods

5.3.1 Case Studies and Data

The data used to conduct the research presented in this paper can be classified into two main categories. The first category includes mine data, which comprises two major time series, methane gas concentration and coal production rate, while the second category consists of barometric pressure time series data, as illustrated in Figure 5.1. The mine data are retrieved from three case studies; three active underground coal mines in the eastern US were renamed Mines A, B, and C due to confidentiality. The weather data were collected from the closest weather station of each mine using a public weather service known as Weather Underground (WU). Table 5.1 shows a sample of the weather data available on the WU website. In general, the database includes over six years of contiguous mine data from Mine A, non-contiguous data spanning almost nine years from Mine B, and seven years of data from Mine C. In addition, the barometric pressure data from three weather stations for the periods described above have been collected.

After the mining and weather data are collected from the different sources (underground mines and weather stations), they are imported and stored into an Atmospheric Monitoring Analysis and Database mAnagement (AMANDA) system, a custom relational database explicitly created to manage data from atmospheric monitoring systems. After that, the different time series data (e.g., methane gas, barometric pressure, and coal production) are pre-processed; this step identifies and filters erroneous values such as missing or zero records, outliers, and peaks. The next step is to bring all-time series data into a common geospatial framework (data points from different time series share a common date/time stamp) to ensure data uniformity and reliability; this stage is called data homogenization. As a result, two time series families are generated from the data collected; (i) daily average values and (ii) 12-hour average values.

Furthermore, once the data are pre-processed and homogenized, the data are exported into the MATLAB® programming environment for further processing and statistical analysis. This stage mainly evaluates the potential autocorrelation of the time series data (methane gas) and cross-correlation between different datasets (methane gas vs. barometric pressure and methane gas vs. coal production) using several statistical techniques such as variogram, cross-covariance, and autocorrelation function. The final steps include time series modeling and time series forecasting. For more detailed information about (i) the characteristics of the data and (ii) the data management process and analysis (data collection, store, pre-processing, and processing) utilized in this research, refer to Diaz et al. (2021; 2021a; 2022).

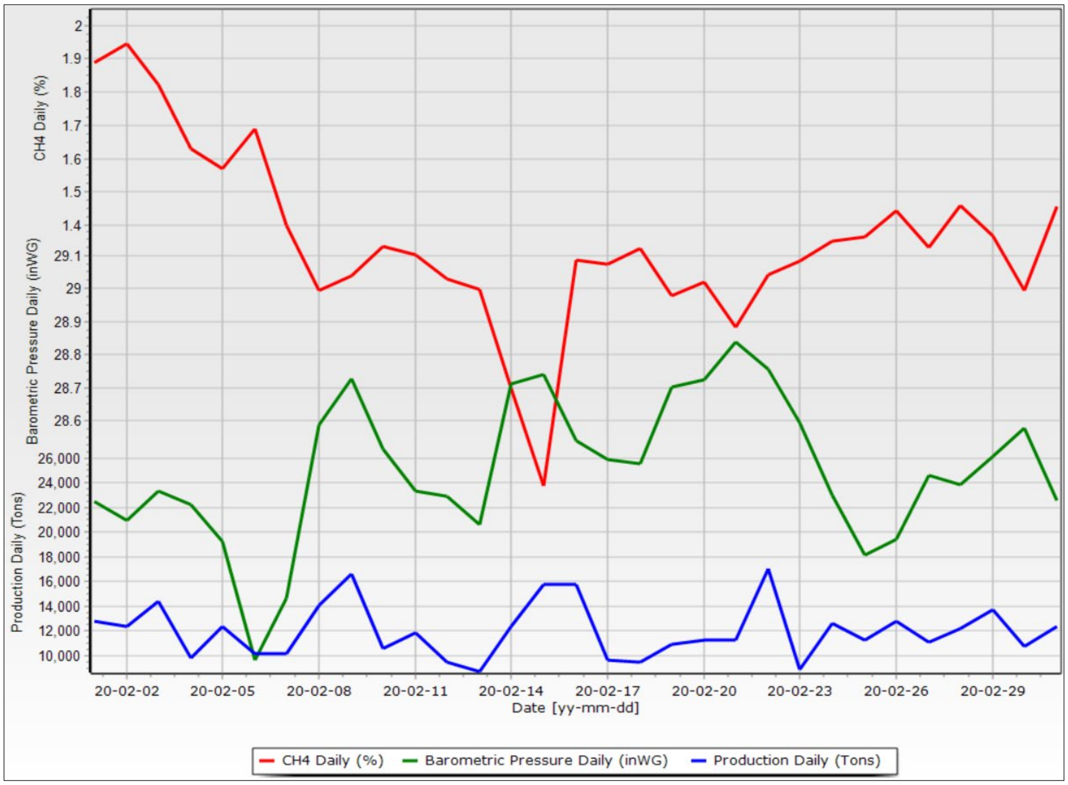

Figure 5.1 Time series data collected using daily average values: methane gas time series represented by the red line, coal production rate time series exemplified by the blue line, and barometric pressure time series symbolized by the green line

Table 5.1 Weather data available from Weather Underground

Date	Time	Temperature	Dew Point	Humidity	Wind Speed	Pressure
02/03/2022	7:13 AM	57 °F	55 °F	93 %	3 mph	28.18 in
02/03/2022	7:26 AM	57 °F	55 °F	93 %	0 mph	28.18 in
02/03/2022	7:53 AM	57 °F	56 °F	96 %	0 mph	28.19 in
02/03/2022	8:04 AM	57 °F	56 °F	96 %	0 mph	28.20 in
02/03/2022	8:20 AM	58 °F	56 °F	93 %	3 mph	28.21 in
02/03/2022	8:46 AM	58 °F	56 °F	93 %	5 mph	28.22 in
02/03/2022	8:53 AM	57 °F	56 °F	93 %	0 mph	28.22 in
02/03/2022	8:55 AM	57 °F	56 °F	96 %	0 mph	28.22 in
02/03/2022	9:33 AM	59 °F	57 °F	96 %	0 mph	28.23 in
02/03/2022	9:41 AM	59 °F	57 °F	93 %	3 mph	28.24 in

5.3.2 Univariate and Multivariate Forecasting Methodologies

Time series can be defined as a collection of data records compiled at regular time periods (e.g., hourly, daily, monthly, and annually). The principal objective of time series analysis is to determine trends and/or patterns in the time series data to predict its future value(s) (Shumway and Stoffer, 2017). Time series forecasting techniques are frequently classified into three major approaches: (i) subjective forecasting, (ii) univariate, and (iii) multivariate forecasting; the last two approaches are the most popular and, consequently, the focus of the research presented in this paper. A univariate time series is a sequence of records with a single time-dependent variable. The univariate forecasting techniques use lagged time series values to forecast its future values; an example of a univariate forecasting model is the Autoregressive Integrated Moving Average Model (Box et al., 2015). In contrast, a multivariate time series consists of more than one time-dependent variable. Multivariate forecasting methods employ pass values of the time series assessed, but the lagged values of other variables (time series) correlated are also considered (Brockwell and Davis, 2016; Chatfield, 2004). Univariate forecasting methods are less complex than multivariate methods mainly because there are fewer parameters to evaluate. However, most of the time, multivariate forecasting techniques offer a more comprehensive understanding of the model and higher accuracy (Brockwell and Davis, 2016). Both methodologies have been used to forecast methane gas concentrations in this research and are presented in the following section.

5.3.3 Proposed Forecasting Approaches

Three forecasting approaches (i.e., the univariate model ARIMA(p,d,q) and the multivariate models VAR(p) and ARIMAX(p,d,q)) are discussed in this section. The performance of these models was assessed with the method of cross-validation (CV), which evaluates how well the predicted values compare with the true values. There are different methodologies for implementing cross-validation. In this study, the datasets were divided into two disjoint sets: the first is the "training set," which is used to train the models and consists of 95% of the records in the time series. The second is the "validation set," which provides the ground truth against which the model forecasts will be compared and comprises 5% of the total records in the time series. The performance and accuracy of the models are assessed with statistical cross-validation metrics such as the Linear Correlation Coefficient (R), the Mean Absolute Error (MAE), the Mean Error (ME), and the Root Mean Squared Error (RMSE). Then, the models are applied in the one-step-ahead forecast mode: the optimal model is determined based on the training set using model selection criteria, specifically, the Akaike Information Criterium (AIC). The optimal model is then employed to forecast methane concentration one step ahead, and the forecast is compared to the true value. Later, the dataset is augmented to include methane concentration values collected at the current time, and a new forecast is generated for the next time instant. Finally, the RMSE is used to generate 95% confidence intervals for the forecasts.

5.3.3.1 ARIMA (p,d,q) model

ARIMA(p,d,q) is a univariate forecasting methodology that forecasts a time series based on its past values and past values of the innovations. In general, ARIMA models involve three integer-valued parameters: (i) the order of the autoregressive (AR) term (p), which specifies the number of lags used as predictors in the model, (ii) the order of the moving average (MA) term (q), which signifies the number of past innovation terms included, and (iii) the order of differencing (d) which is used to render a time series stationary; for particular datasets, more than one differencing operation may be required (Box et al., 2015; Chatfield, 2004).

A detailed description of developing the ARIMA(p,d,q) method to forecast methane gas concentrations is discussed by Diaz et al. (2022). Finally, the ARIMA(p,d,q) model results are compared and evaluated against the multivariate models presented and discussed below.

5.3.3.2 Multivariate Vector Autoregressive model

The Multivariate Vector Autoregressive (VAR) model is one of the most popular and straightforward techniques for analyzing multivariate time series developed by the econometrician and macroeconomist Christopher A. Sims in 1980 (Johansen, 1995). A VAR model of order p (VAR(p)) comprises n coupled variables (time series). Each variable depends on its p past values as well as on the past values of all other variables up to order p. A mathematical representation of the VAR(p) model for two variables and p=1 is given in Eq. (5.1). The main advantages of the VAR(p) model are that it provides a logical and accurate approach to data description, structural inference, and forecasting (Kirchgässner and Wolters, 2007). For more information about the VAR(p) model, refer to Johansen (1995) or Kirchgässner and Wolters (2007).

$MVAR(1)\ model - (y_t, x_t), using\ one\ lag\ (p = 1)$

$$y_t = a_1 + b_{11}y_{t-1} + b_{12}x_{t-1} + u_t$$

$$x_t = a_2 + b_{21}y_{t-1} + b_{22}x_{t-1} + v_t \qquad \text{Eq. (5.2)}$$

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

where $y_t\ and\ x_t$ are two stationary time series, a_1, a_2, b_1, b_2 are constant model parameters and u_t , v_t are independent white noise processes that represent the innovation terms. One time series represents methane gas concentration and the other atmospheric pressure for this study.

The VAR(p) model was developed in the MATLAB® environment employing functions from the econometrics toolbox. Sample datasets from the three case studies (Mines A, B, and C) covering periods of one and six years with two different time average values (e.g., 12 hours and daily) were used to build respective VAR(p) models.

Several statistical tests were implemented to evaluate the normality, heteroscedasticity, and stationarity properties of the time series data. These statistical properties can provide valuable indicators for the performance of the VAR(p) model since they are inherent in the model. First, the assumption of normality is evaluated using the Lilliefors test, which is an improvement of the Kolmogorov-Smirnov (K-S) test. Lilliefors test is more accurate when the population mean and standard deviation are unknown and need to be estimated from the sample data, as in this research (Seemon, 2014; Ghasemi and Zahediasl, 2012).

The presence of heteroscedasticity (i.e., variations of the variance) was investigated using the Autoregressive Conditional Heteroscedasticity test. Heteroscedasticity can occur due to outliers in the dataset or the omission of relevant variables in the model. Since the VAR(p) model assumes homoscedastic variance of the innovation terms establishing homoscedasticity is essential for validating the model and predicting confidence intervals (C.I.) (Williams, 2020).

The stationarity of the time series was tested using two statistical techniques: the Augmented Dickey-Fuller test (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. A time series is stationary if its statistical features such as median, variance, and mean are not affected by time. Stationary time series have no trends, seasonality, or fluctuations (NIST, 2003). The ADF and KPSS tests investigate the hypothesis that a time series is stationary around a deterministic trend (Prabhakaran, 2019).

5.3.3.3 ARIMAX (p,d,q) model

A second multivariate forecasting approach was also utilized based on the Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) model, an extended version of the traditional ARIMA model. The conventional univariate ARIMA (p,d,q) model permits forecasting the future values of a time series based on its past values. In contrast, ARIMAX(p,d,q) model also incorporates past values of one or more exogenous dependent variables (Andrews et al., 2013; Kravchuk, 2017). The ARIMAX model is similar to a multivariate regression forecast model; the main differences are (i) that ARIMAX incorporates autoregressive and moving average terms, and (ii) it utilizes the potential autocorrelation to enhance the accuracy of the forecasts (Kravchuk, 2017; Hyndman and Athanasopoulos, 2021). The ARIMAX (p,d,q) model is applied herein with d=q=0 to compare its performance with the VAR(p) model. The optimal ARIMAX model has an autoregressive order p established using the AIC.

5.4 Results

This section presents the results of the methane gas concentration forecasting using the multivariate VAR(p) and ARIMAX(p,d,q) techniques for four datasets (from 1 to 4). Each dataset comprises two time series; (i) the methane gas time series (dependent variable) and (ii) the barometric pressure time series (independent variable). The methane data used were retrieved from Mine A and the barometric pressure data from the nearest weather station of the mine. Furthermore, the first two datasets (1 and 2) implement daily average values and consist of one year and six years of data, respectively. On the other hand, the last two datasets (3 and 4) use the same data as datasets 1 and 2, correspondingly, but employ 12-hour average values. Though this section focuses on the results of four datasets, 12 datasets were employed in total to feed both models; six of the datasets use daily average values, and the other 12-hour average values, as presented in Table 5.4.

5.4.1 VAR(p) One-Step-Ahead Model of Methane Concentration

Figures 5.2 and 5.3 consist of four plots each (a to d). Graphs (a) and (c) are the forecasts obtained using the optimal VAR(p) models, while plots (b) and (d) are magnified versions of the forecasts shown in graphs (a) and (c), respectively. The gray line represents the training data (methane gas time series), the blue line represents the validation set, the red line denotes the forecast, and the black dashed lines indicate the boundaries of the 95% confidence interval for the forecasts. The validation and forecast periods contain five percent (5%) of the methane time series.

Figure 5.2a shows the forecast obtained for dataset 1, which contains the time series of methane gas concentrations spanning more than 360 days using daily average values. The optimal autoregressive order (p) term is two (p=2) obtained with the lowest AIC. Figure 5.2b demonstrates that the forecast (red line) is relatively close to the true value during the validation period (blue line); the correlation coefficient was calculated at R=0.89, implying a strong correlation between the validation data and the forecast. Moreover, the observed values lie within the 95% confidence interval. Therefore, the proposed VAR(2) model provides an accurate forecast for this particular dataset.

Figure 5.2c displays the forecast for dataset 2. It includes the time series of methane gas concentrations spanning more than 2,100 days (six years) with a daily average value. In this case, the optimal model obtained was VAR(13). Figure 5.2d demonstrates a significant correlation between the validation data and the forecast (R=0.66), considering that the training data are far from ideal (e.g., they include zero values and some sharp fluctuations), as shown in Figure 5.2c. Nevertheless, the VAR(13) model follows the sharp fluctuations of the validation data accurately, while the validation data lie within the 95% prediction interval. Additionally, the cross-validation metrics achieved are significantly better than those obtained in dataset 1, except for the value of R, as presented in Table 5.2.

Figure 5.3 shows the VAR(p) forecast for datasets 3 and 4 using 12-hour average values. Figure 5.3a shows that the VAR(8) is the best model according to the lowest AIC. Furthermore, Figure 5.3b reveals that the forecasts closely follow the validation data; the correlation coefficient calculated was R=0.91, higher than the correlation achieved with the daily average samples (cf. Figures 5.2a and 5.2b). Figure 5.3c presents the forecast for a time series spanning more than 2,100 days. In this case, the optimal model was VAR(30), which corresponds to the lowest AIC. Visual inspection of Figure 5.3d indicates that the forecast (red line) and the validation data (blue line) are notably similar. The correlation coefficient was calculated at R=0.66, which is the same value for the time series presented in Figures 5.2c. Additionally, the cross-validation metrics obtained for dataset 4 are significantly better than those achieved in dataset 3, except for the value of R, as indicated in Table 5.2.

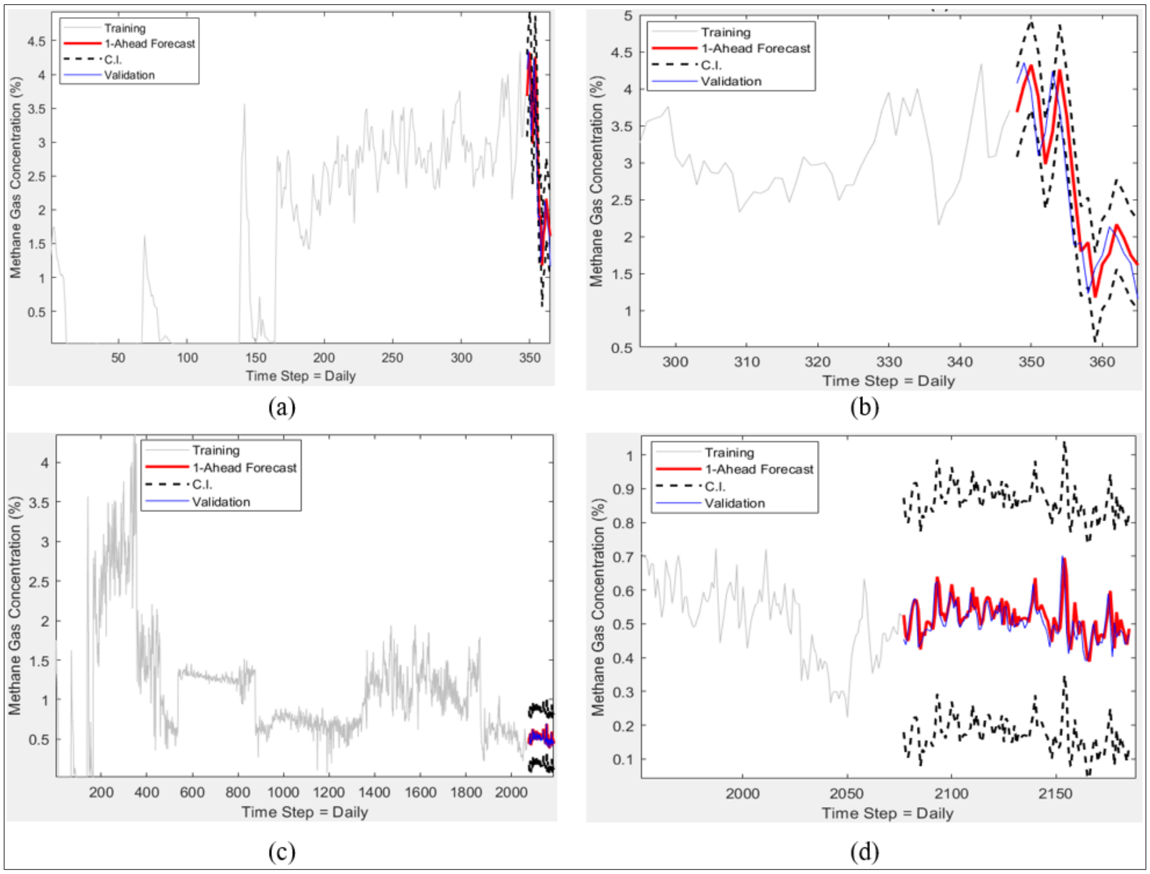

Figure 5.2 VAR(p) one-step-ahead CH_4 concentration forecasts using daily average values;(a) Forecasting for dataset 1, (b) Magnified view of the forecast in (a), (c) Forecasting for dataset 2, (d) Magnified view of the forecast in (b)

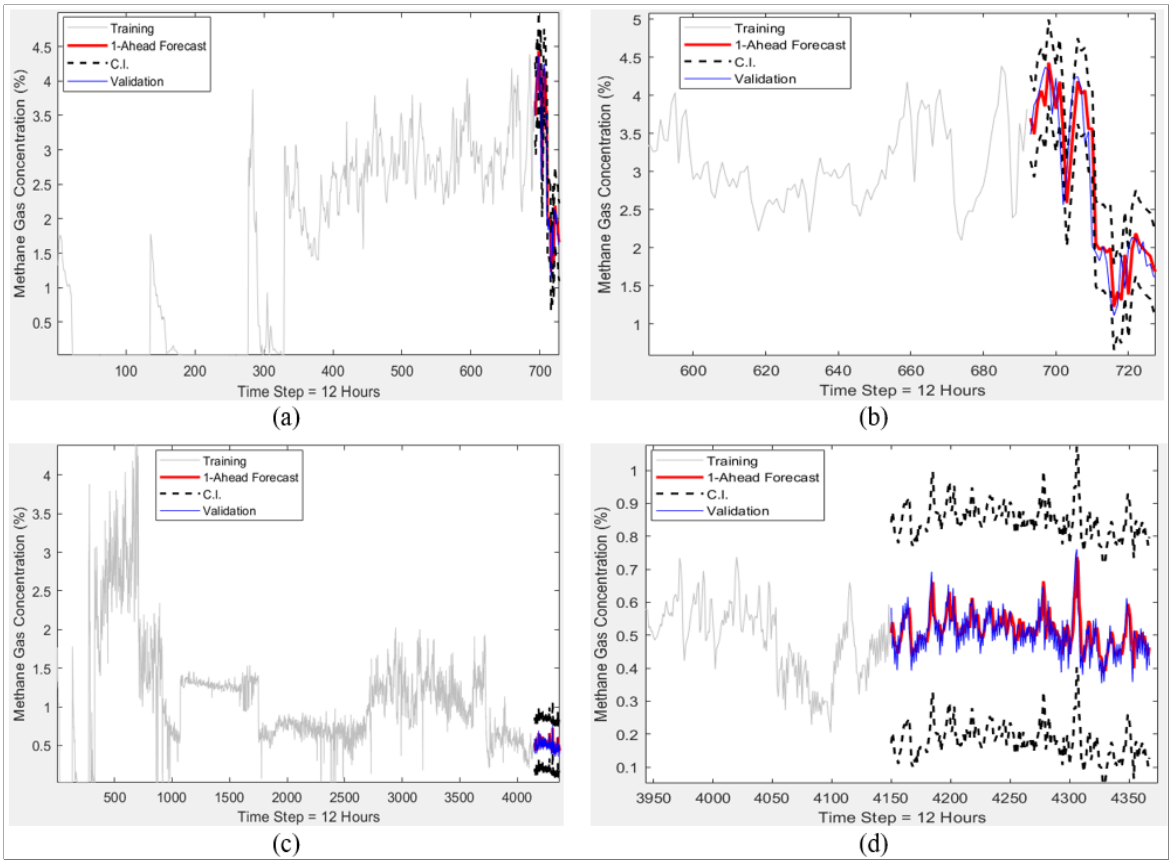

Figure 5.3 VAR(p) one-step-ahead CH_4 concentration forecasts using 12-hour average values;(a) Forecasting for dataset 3, (b) Magnified view of the forecast in (a), (c) Forecasting for dataset 4, (d) Magnified view of the forecast in (b)

Table 5.2 Summary of methane gas forecasting results for the VAR(p) one-step-ahead model (using a vector of methane gas concentration and barometric pressure)

	Time Step	Training Data Length (days)	Training Data Sample Size	Validation Data Length (days)	Validation Data Sample Size	Optimal order of VAR(p) model	Correlation Coefficient (R)	RMSE	ME	MAE
Dataset 1	Daily average	365	365	18	18	2	0.89	0.50	−0.11	0.44
Dataset 2	Daily average	2,200	2,200	109	109	13	0.66	0.05	−0.01	0.04
Dataset 3	12 Hour average	365	730	18	37	8	0.91	0.46	−0.06	0.32
Dataset 4	12 Hour average	2,200	4,380	109	219	30	0.66	0.05	−0.01	0.04

5.4.2 ARIMAX One-Step-Ahead Model of Methane Concentration

Figure 5.4 consists of four plots (a to d) representing the methane gas forecasts obtained using the ARIMAX (p,d,q) model. The white background contains the data used to train the model, the upper and lower black dashed lines signify the boundaries of the 95% confidence interval, and the gray background contains the validation data (blue line) and the forecast (red line).

Figure 5.4a shows the forecast obtained for dataset 1. The optimum model was achieved using an autoregressive term equal to twenty-eight (p=28), based on the lowest AIC. Furthermore, visual inspection demonstrates that the forecast (red line) is close to the true value during the validation period (blue line), and the observed values lie within the 95% confidence level. Moreover, the correlation coefficient calculated was R=0.87, as shown in Table 5.3. Therefore, it implies a strong correlation between the validation data and the forecast. Consequently, the proposed model provides a reliable forecast for this particular data segment.

Figure 5.4b displays the forecast for dataset 2. In this case, the optimal model was obtained when the autoregressive term equals twenty-three (p=23). Again, visual examination shows that the forecast and validation data have a strong correlation, with the correlation coefficient computed at R=0.65. The forecast model consistently follows the peaks in the validation data accurately, and the 95% confidence interval continuously contains the forecast.

Figures 5.4c and d present the forecast for datasets 3 and 4, respectively. The time series employed have the same lengths (360 days and 2,100 days) as datasets 1 and 2 but utilize 12-hour average values. In Figure 5.4c, the best model according to the lowest AIC is when the autoregressive term is equal to eight (p=8). Furthermore, Figure 5.4c reveals that the forecasts closely follow the validation data. In fact, the correlation coefficient calculated was R=0.91, higher than the correlation achieved with the daily average samples (Figure 5.4a).

Moreover, in Figure 5.4d, the best model was obtained when the autoregressive term equals twenty-nine (p=29). It can be observed that the forecast and the validation data are remarkably similar. The correlation coefficient was calculated at R=0.68, which is slightly higher than for the time series shown in Figure 5.4b.

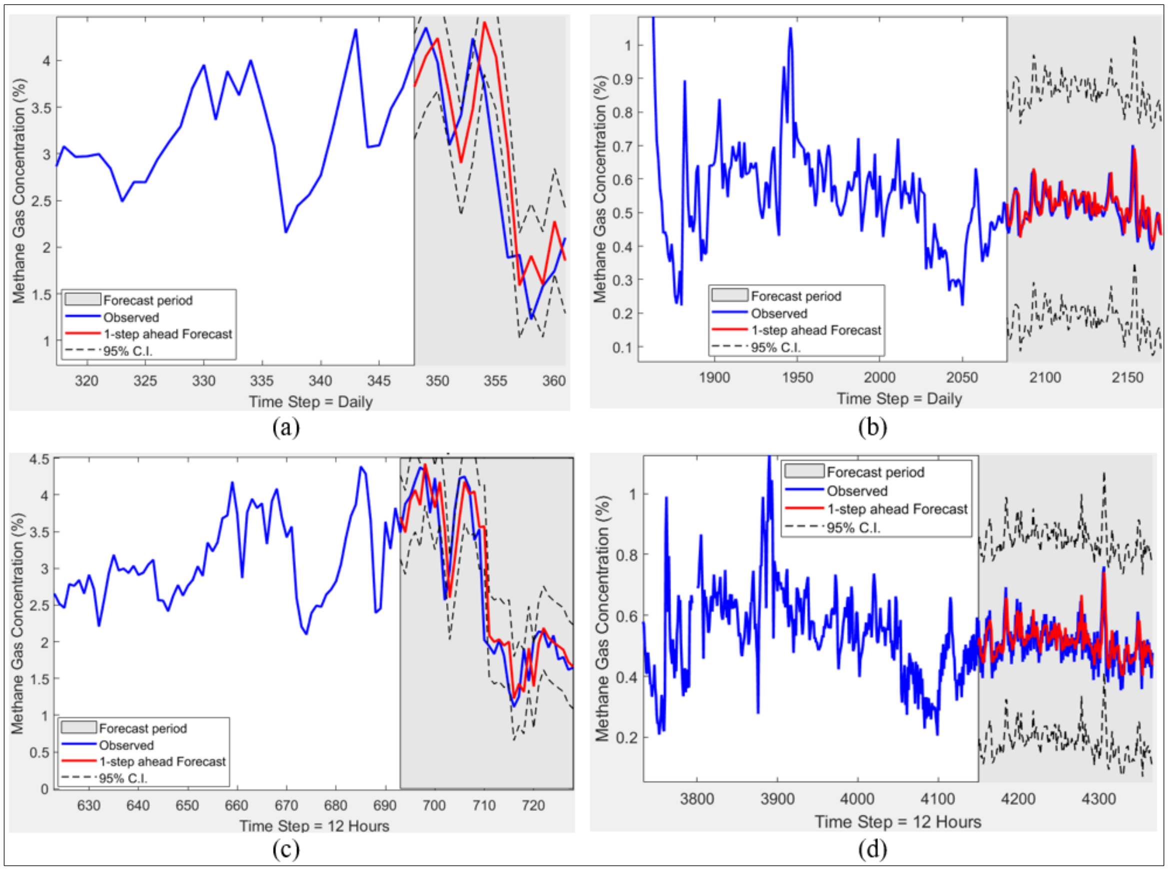

Figure 5.4 ARIMAX one-step-ahead CH_4 concentration forecasts;(a) Forecast for dataset 1 using daily average values, (b) Forecast for dataset 2 using daily average values, (c) Forecast for dataset 3 using 12-hour average values, (d) Forecast for dataset 4 using 12-hour average values

Table 5.3 Summary of methane gas forecasting results for the ARIMAX one-step-ahead model (using barometric pressure time series as an independent variable)

		Training Data		**Validation Data**		**ARIMAX**	**Correlation**			
	Time Step	**Length (days)**	**Sample Size**	**Length (days)**	**Sample Size**	**(Optimal p)**	**Coefficient (R)**	**RMSE**	**ME**	**MAE**
Dataset 1	Daily average	365	365	18	18	28	0.87	0.58	−0.20	0.49
Dataset 2	Daily average	2,200	2,200	109	109	23	0.65	0.05	−0.01	0.03
Dataset 3	12 Hour average	365	730	18	37	8	0.91	0.46	−0.07	0.33
Dataset 4	12 Hour average	2,200	4,380	109	219	29	0.68	0.05	−0.01	0.04

5.5 Discussion

Two multivariate forecasting approaches have been presented to predict methane gas concentrations. First, this paper developed and presented the multivariate methodology that includes the VAR(p) and ARIMAX(p,d,q) models. The results from these approaches are compared with those obtained by means of the univariate approach, the ARIMA(p,d,q) model, as explained in previous work by Diaz et al. [26]. All forecasting methodologies used the same training and validation datasets, and their performance and accuracy were assessed using cross-validation metrics.

This investigation identified that the three forecasting models could accurately predict methane gas concentrations. For example, in most cases, the concentrations of methane gas forecasted by the VAR(p), ARIMAX(p,d,q), and ARIMA (p,d,q) models match the direction (increase/decrease) of the validation data. Moreover, the observed values of methane gas levels were captured by the 95% confidence intervals, as illustrated in Figures 5.2, 5.3, and 5.4. In addition, the linear correlation between the forecasts and the validation data was strong and positive, and the value of the cross-validation metrics was similar, as shown in Table 5.4.

However, in some cases, the performance and accuracy of the three forecasting models were hindered by complex datasets containing inconsistencies such as abrupt changes in the methane gas time series. As a result, the linear correlation between the forecast and the validation data was weak. For example:

- The results from the analysis of dataset 6 show that the linear correlation between the forecast and validation data was weak ($R<0.5$) for all three forecasting methods; the correlation achieved by the ARIMA, VAR, and ARIMAX models was 0.33, 0.35, and 0.34, respectively, as indicated in Table 5.4. This can be attributed to an abrupt change in the average concentration of the methane time series, most likely due to sensor failure or/and calibration, or the potential influence of an independent variable(s) (e.g., coal production rate) directly affecting methane gas emissions and barometric pressure correlation. Indeed, Figure 5.5 shows the methane gas (red line) and barometric pressure (green line) time series with 12-hour average values implemented for dataset 6. Visual inspection demonstrates that the methane time series contains an abrupt change; the methane gas concentration decreases unexpectedly from 1.4% to 0.60%. Furthermore, the linear correlation between methane gas and barometric pressure time series was calculated at $R=0.16$.
- Dataset 5 includes the same data as dataset 6 but with daily average values instead of the 12-hour average values. As a result, the univariate ARIMA and multivariate ARIMAX models show a better performance than in dataset 6. In addition, a higher correlation between the forecasts and validation set was achieved, i.e., $R=0.54$ and $R=0.53$, respectively. On the other hand, the performance of the VAR model did not improve; the correlation between the forecasts and validation data was slightly lower ($R=0.34$) than the one achieved in dataset 6, as indicated in Table 5.4.
- Similarly, datasets 7 and 8 were based on the same records averaged differently; in dataset 7, the time series data represent daily average values, while the time series in dataset 8 uses 12-hour average values. As a result, the ARIMA, VAR, and ARIMAX models in dataset 8 yield superior results in terms of linear correlation

between the forecast and validation data as well as for the rest of the cross-validation metrics compared to those for dataset 7 (Table 5.4). Furthermore, a similar pattern is evident for all the datasets presented in this paper, except dataset 9. Therefore, the better performance and accuracy achieved by the forecast models that employ 12-hour average values can be explained due to these methane time series containing two times more information than the methane time series with daily average values.

The analysis of the results obtained from the different forecast approaches indicates that the ability of the ARIMA, MVAR, and ARIMAX models to predict future concentrations of methane gas may be influenced by several factors such as irregularities in the dataset (e.g., unusual variations and faulty values), the time scale used for averaging (e.g., 12-hour or daily average values) and the presence of additional independent variables (e.g., coal production rate) not accounted by the model. Therefore, all these factors determine the best forecast model able to deliver consistently superior results in all datasets. Consequently, it is required to develop a methodology for selecting the best (univariate or multivariate) forecast model based on cross-validation analysis. Figure 6 presents a flow diagram illustrating such a methodology. First, the univariate ARIMA(p,d,q) and multivariate VAR(p) and ARIMAX(p,d,q) forecasting models are applied to the same dataset. Second, the optimal parametrization for each model is determined based on the lowest AIC value. Third, the best model among the ARIMA, VAR, and ARIMAX forecast methods is selected based on a specified cross-validation measure (e.g., linear correlation or RMSE). Fourth, the optimal model is used to forecast the methane gas concentrations. Finally, as the dataset is updated with new values for methane gas concentration and barometric pressure, the algorithm is rerun to determine the best methane gas forecasting model for the updated dataset. This last step implies a continuous model updating in light of incoming data.

The results obtained from the methane gas forecasts proposed by this research line up with previous studies. For example, Wang (2020) developed several methodologies to forecast methane gas using time series from the sensors of an underground coal mine. The RMSE obtained for the ARIMA and VAR models proposed by the author were 5.4E-3 and 4.5E-3, respectively. However, it is not possible to directly compare our results with those obtained by Wang. First, statistics of the gas concentration series that were used in that study are not shown and no graphical plots of the time series for the time series are provided. Second, there are no details regarding the partitioning of the data into training and validation sets. Third, the authors give no information regarding the orders of the optimal ARIMA models used and how they were determined. Fourth, the dataset used in that study contained significantly more information, since it involved measurements for three different gases from 15 monitoring sensors with a sampling step of 6 seconds for a total of about 6 million time points. Finally, the forecasting horizon (seconds, minutes, hours or days) over which the validation measures are evaluated is not specified.

In a different study, Karacan (2007) proposed a more sophisticated approach to predicting concentrations of methane gas in longwall mines using Artificial Neural Networks. The author fed the model using ten datasets. It was found that the linear correlation between the forecast and the validation data was around R=0.93 for all datasets. Consequently, the RMSE and linear correlation found in these previous investigations are similar to the values obtained in the forecast of the datasets having six years length and 12-hours average values in this research (see Table 5.4).

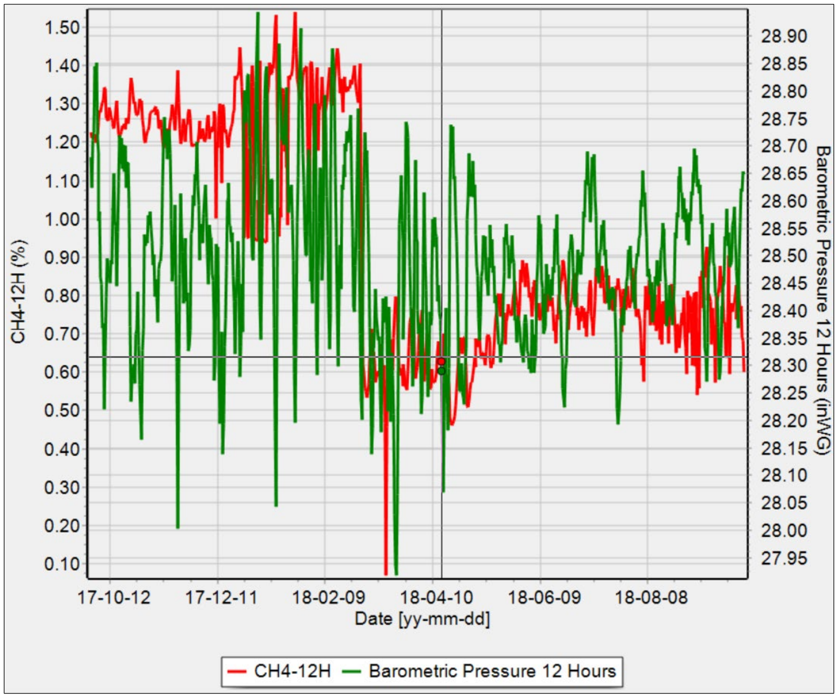

Figure 5.5 Methane gas and barometric pressure time series included in dataset 6

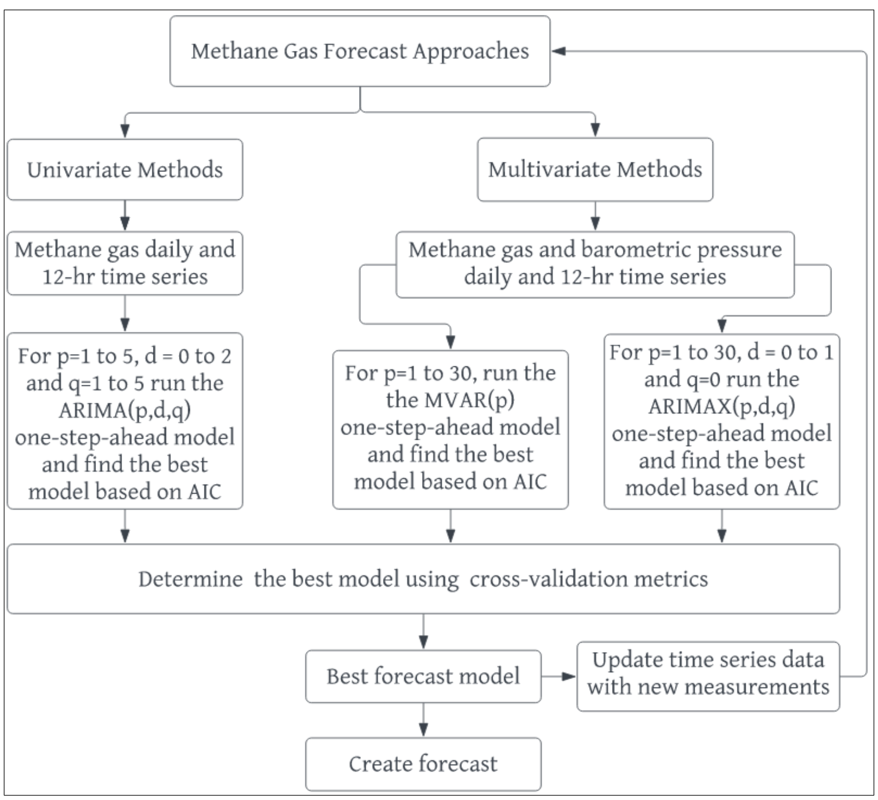

Figure 5.6 Proposed methodology for selecting the best univariate or multivariate forecast model

In order to develop improved methods of forecasting methane gas concentrations in mines, several steps need to be taken by the research community. First, it is essential to obtain good quality gas concentrations measurements and limit the number of erroneous records due to sensor malfunction or recording gaps. Second, the correlations between atmospheric pressure, coal production rate, and gas concentration, including potential confounding variables, need to be better understood. Third, with respect to modeling efforts, statistical methods (whether based on classical time series analysis or more modern machine learning tools) have an advantage over methods that are based on computational fluid dynamics since the latter demands significant computational resources as well as information (e.g., values of diffusion coefficients, initial and boundary conditions) which are usually not fully known. However, more research is needed to establish the scope, accuracy, and reliability of statistical forecasting methods.

Given that a number of different statistical methods can be applied to methane gas forecasting, it is essential to agree on a minimal set of reporting principles that will allow performance comparison between different methods. Therefore, it is proposed that the following critical elements of the data analysis be thoroughly reported: (1) adequate statistical characterization of the data and the pre-processing protocol, (2) complete specification of the statistical forecasting model, including the values of all the model parameters and the methods used to estimate their values and (3) explicit description of the training and cross-validation practice and presentation of statistical performance measures.

Concerning the first point above, the following is recommended: (i) the reporting should include the number, nature (e.g., concentration, atmospheric pressure, etc.), and units of the time series used in the forecasting analysis, (ii) the length of the time series and the sampling step (e.g., hour, day) should be specified and (iii) any pre-processing steps used to filter, smoothen, or coarse-grain (down-sample) the data or remove outliers should be described. Furthermore, graphical plots of conventional time series are recommended, as they can provide valuable visual aids for the readers. Moreover, the results of exploratory statistical analysis should be listed, including the mean, median, standard deviation, skewness, and kurtosis coefficients of the data. Likewise, an analysis of the probability distribution(s) followed by the different time series should be included (using suitable probability plots if needed), and deviations from the normal distribution should be modeled. This analysis should be followed by estimating two-point correlations employing the autocorrelation function (ACF), partial autocorrelation function (PACF), and variogram function plots for the time series. Trends, periodic behavior, periodicities, and non-stationarities (if present) should be identified and discussed. In addition to the visual analysis of graphs, these efforts can be supported by statistical tests that investigate the normality, stationarity, and heteroscedasticity of the data.

Regarding the second point, the statistical model should be adequately specified for the results to be reproducible by other researchers. For example, in the case of ARIMA models, it is necessary to report the orders (p,d,q) of the autoregressive component (p), of the differencing operator (d), and the moving average component (q). The maximum investigated orders should also be declared, as well as the statistical criterion used for model selection (e.g., AIC, Bayesian Information Criterion, or cross-validation). Similar considerations apply to vector autoregressive (VAR) and ARIMAX models with exogenous variables (e.g., atmospheric pressure and production). In addition, if a nonlinear transformation (e.g., Box-Cox) was applied to the data, the functional form and pertinent parameters should be given. It should also be clarified if standardization (Z-score normalization) has been applied to the data (mainly when using multivariate methods). In the case of machine learning methods (e.g., artificial neural networks), the results can be highly dependent on several decisions related to the structure and training of the network. Again, all relevant details should be presented, including the network's architecture (e.g., number and type of layers, number of nodes per layer, selection of activation function and the regularization approach to avoid overfitting), the training method used, and the values of the different hyperparameters involved in the training process.

Finally, for point (3), it is equally important to specify how the model was trained (i.e., what percentage of the data was used for training) and which protocol was used to conduct the validation. For example, one pertinent issue is whether a one-step-head or a k-step-ahead (where $k>1$) forecasting protocol is used. It may sound redundant, but it should be stressed that the validation should be performed with values not included in the training set. The statistical measures of forecasting performance must include the mean absolute deviation, the root mean square error, and the correlation coefficient between validation and forecast values. Relative measures of performance (i.e., concerning the average value of the data) are also helpful (e.g., relative root mean square error) since the average gas concentration may vary between different mines or even different sections of the same mine. If the data have been transformed for processing, the performance measures should be reported in the original domain (e.g., the RMSE should be calculated and reported for the concentration and not for its logarithm, in case the logarithmic transform has been applied). If the method allows for uncertainty estimation (e.g., the ARIMA-based time series methods and Gaussian process regression), measures of uncertainty quantification should also be reported. One such measure involves confidence intervals for the forecast. In addition, proper scoring rules can be implemented for uncertainty quantification as described in Gneiting and Raftery (2007) and Bessac and Naveau (2021). Finally, it is helpful to supplement the analysis of forecasting performance with model diagnostic testing to investigate whether the forecasting model is consistent with the underlying assumptions (Box et al., 2015).

The statistical (stochastic) and machine learning forecasting approaches are data-driven methods. The former methods have a long history, while the latter has gained momentum in the last decade. Even though a natural tendency is to prefer more modern approaches than older methods, it is recommended to analyze the merits of both approaches. The statistical methods, for example, are inherently capable of estimating forecast uncertainty, and they provide interpretable results. On the other hand, machine learning approaches do not depend on parametric assumptions regarding the probability distribution of the data. Therefore, a fair comparison of the two approaches requires adherence to a set of reporting principles as described above. In addition, the computational resources (e.g., CPU memory usage, computational time, and scaling of resources with size) should be parts of such comparisons. It should also be mentioned that the classical time series approach involves several nonlinear generalizations (Enders, 2008), such as autoregressive heteroskedastic (ARCH) models and their generalized (GARCH) versions, regime-switching models such as Markov switching AR and Self-Exciting Threshold Autoregressive (SETAR) models. Based on the literature review, such models that sound better suited for handling irregular (i.e., non-Gaussian, non-stationary) data have not yet been applied to methane gas concentration forecasting. Finally, the machine learning method of Gaussian process regression (Agou et al., 2022) and geostatistical analysis, which shares many features with the former (De Iaco et al., 2022), also provide flexible forecasting frameworks that deserve further investigation.

Table 5.4 Univariate and multivariate forecasting results summary

		Data Segment Features				**Univariate Forecasting Approach**					**Multivariate Forecasting Approach**									
	Time Step	**Training Data**		**Validation Data**		**ARIMA(p,d,q) One-Step-Ahead**					**VAR(p,d,q) One-Step-Ahead**					**ARIMAX(p,d,q) One-Step-Ahead**				
		Length (days)	Sample Size	Length (days)	Sample Size	COR	Best Model (p,d,q)	RMSE	ME	MAE	COR	Best Model VAR(p)	RMSE	ME	MAE	COR	MAX (P)	RMSE	ME	MAE
Dataset 1	Daily average	365	365	18	18	0.89	(4,1,4)	0.54	-0.19	0.44	0.89	VAR(2)	0.50	-0.12	0.44	0.87	28	0.58	-0.20	0.49
Dataset 2	Daily average	2,200	2,200	109	109	0.65	(3,1,4)	0.04	0.00	0.03	0.66	VAR(13)	0.05	-0.01	0.04	0.65	23	0.05	-0.01	0.03
Dataset 3	12 hr average	365	730	18	37	0.90	(4,1,4)	0.47	-0.11	0.35	0.91	VAR(8)	0.46	-0.06	0.32	0.91	8	0.46	-0.06	0.33
Dataset 4	12 hr average	2,200	4,380	109	219	0.71	(4,1,4)	0.05	0.00	0.04	0.66	VAR(30)	0.05	-0.01	0.05	0.68	29	0.05	-0.01	0.04
Dataset 5	Daily average	365	365	18	18	0.54	(2,0,2)	0.05	0.00	0.04	0.34	VAR(3)	0.06	-0.01	0.05	0.53	3	0.06	0.00	0.04
Dataset 6	12 hr average	365	730	18	37	0.33	(1,1,1)	0.08	0.00	0.06	0.35	VAR(11)	0.08	-0.01	0.06	0.34	10	0.09	-0.01	0.06
Dataset 7	Daily average	365	365	18	18	0.33	(1,0,2)	0.05	0.00	0.04	0.57	VAR(3)	0.04	0.00	0.03	0.41	3	0.05	0.00	0.03
Dataset 8	12 hr average	365	730	18	37	0.72	(3,1,4)	0.04	0.01	0.03	0.78	VAR(8)	0.03	0.00	0.02	0.72	8	0.04	0.00	0.03
Dataset 9	Daily average	365	365	18	18	0.65	(3,0,4)	0.02	0.00	0.01	0.50	VAR(5)	0.02	0.02	0.00	0.68	16	0.02	0.00	0.01
Dataset 10	12 hr average	365	730	18	37	0.52	(2,1,1)	0.02	0.00	0.02	0.48	VAR(12)	0.03	0.00	0.02	0.54	17	0.02	0.00	0.02
Dataset 11	Daily average	365	365	18	18	0.79	(1,1,2)	0.20	0.08	0.16	0.79	VAR(3)	0.19	0.04	0.15	0.83	25	0.18	0.07	0.14
Dataset 12	12 hr average	365	730	18	37	0.85	(1,0,1)	0.17	0.01	0.13	0.83	VAR(9)	0.18	0.01	0.14	0.84	15	0.18	0.02	0.13

5.6 Conclusions

This work was motivated by the need to develop an accurate and reliable model for predicting methane concentration in underground mine environments. Therefore, an extensive study on univariate and multivariate forecasting models that can predict methane gas concentrations in underground coal mines based on time series analysis has been conducted. Three forecasting approaches (ARIMA, VAR, and ARIMAX) are presented and compared using cross-validation metrics to establish the best methane gas forecasting model. ARIMA is based exclusively on the autocorrelations of the methane concentration series, while ARIMAX and VAR also consider the variation of barometric pressure and its cross-correlations with the methane gas concentration. A total of 12 datasets were employed utilizing daily and 12-hour average values. As a result, the following conclusions can be drawn:

- The performance of the forecast methods is directly influenced by the presence of irregularities in the dataset and potential independent variables (e.g., coal production rates) that are not included in the models.
- The data pre-processing steps, in most cases, require human intervention and assessment.
- None of the models proposed can uniformly outperform the other forecasting approaches for all datasets.
- In 10 of 12 datasets, the 12-hour average time series gave better forecasts than the daily average time series.
- An algorithm is proposed to assess the results of both multivariate and univariate models and select the best model for the given dataset. Furthermore, the algorithm should be run continuously to update the best model based on available new data.
- Besides being one of the first documented attempts to use time series data from AMS, the univariate and multivariate methane gas forecast models proposed in this research also achieved excellent results. They offer a potential solution to fill the gap of reliable methodologies capable of forecasting methane gas concentrations to improve the safety and health conditions of the workforce in the underground coal mining industry.

The focus of future work will be to (a) develop multivariate forecasting methods that will include coal production rate as a second exogenous variable, (b) compare the performance of forecasting models with one (barometric pressure) against two (barometric pressure and coal production rate) exogenous variables and (c) collect more information from different case studies to establish if the different forecast models proposed in this research can be broadly implemented. It will help improve the safety and health conditions of underground coal mines worldwide.

5.7 Acknowledgments

This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION), contract number AFCTG20-103. The views, opinions, and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its Directors, and staff.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

5.8 References

Agioutantis Z, Luxbacher K, Karmis M, Schafrik S (2014) Development of an atmospheric data management system for underground coal mines. J. South. Afr. Inst. Min. Metall, vol 114: 1059–1063.

Agou V.D, Andrew P, Dionissios T.H. (2022) Spatial Modeling of Precipitation Based on Data-Driven Warping of Gaussian Processes" Entropy 24, no. 3: 321. https://doi.org/10.3390/e24030321.

Airey EM (1968) Gas emission from broken coal. An experimental and theoretical investigation. Int. J. Rock Mech. Min. Sci. Geomech, 5:475–494.

Andrews BH, Dean MD, Swain R, Cole C (2013) Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. Society of Actuaries. The University of Southern Maine. Portland, USA.

Bessac J, Naveau P (2021) Forecast score distributions with imperfect observations. Advances in Statistical Climatology, Meteorology, and Oceanography, 7(2), pp.53-71.

Booth P, Brown H, Nemcik J, Ting R (2017) Spatial context in the calculation of gas emissions for underground coal mines. Int. J. Min. Sci. Technol. https://doi.org/10.1016/j.ijmst.2017.07.007.

Booth P, Nemcik J, Ren T (2016) A critical review and new approach for determination of transient gas emission behavior in underground coal mines. In Proceedings of the 16th Coal Operators' Conference, Mining Engineering, University of Wollongong, Naj Aziz and Bob Kininmonth (eds.) 10-12 February 2016, 367-379, Wollongong, Australia.

Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis forecasting and control, 5th (ed). John Wiley & Sons. New Jersey, United States.

Brockwell PJ, Davis RA (2016) Introduction to time series and forecasting, 3rd ed. Springer. Switzerland.

Byrer C, Havryluk I, Uhrin D (2014) Coalbed methane: a miner's curse and a valuable resource. In: Pramod T, Schatzel S, Aminian K (ed) Coalbed methane: from prospect to pipeline, 1st ed. Elsevier, Amsterdam, pp 1–6.

Chatfield C (2004) The analysis of time series - An introduction, Six ed. CHAPMAN & HALL/CRC, New York, USA.

De Iaco S, Hristopulos, D.T., Lin, G. (2022) Special Issue: Geostatistics and Machine Learning. Math Geosci 54, 459–465. https://doi.org/10.1007/s11004-022-09998-6.

Diaz, JC., Agioutantis, Z., Hristopulos, DT., Luxbacher, K., Schafrik, S. (2022). Time series modeling of methane gas in underground mines. Mining, Metallurgy, and Exploration Journal (submitted).

Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Towards atmospheric monitoring data analysis in underground coal mines. In: Proceedings of the 18th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781003188476-51.

Diaz J, Agioutantis Z, Hristopulos DT, Schafrik S (2021a) Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. Mater. Proc. 2021. https://doi.org/10.3390/materproc2021005078.

Dixon DW (1992) A statistical analysis of monitored data for methane prediction. Dissertation, University of Nottingham. Nottingham, United Kingdom.

Dixon DW, Longson I (1993) Statistical method for methane prediction and improved environmental control. In: Proceedings of the 6th North American Mine Ventilation Symposium, pp. 94–98. Salt Lake City, Utah, United States.

Dunmore R (1982) Predicting methane in real-time for longwall faces. In Proceedings of the 1st North American Mine Ventilation Symposium. 29–31 March, Alabama, USA.

Enders W (2008). Applied Econometric Time Series. John Wiley & Sons.

Flores RM (1998) Coalbed methane: From hazard to resource. Int. J. Coal Geol. https://doi.org/10.1016/S0166-5162(97)00043-8.

Ghasemi A, Zahediasl S (2012) Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. https://doi.org/10.5812/ijem.3505.

Gneiting T, Raftery A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. 102 (447): 359–378.

Guo H, Adhikary DP, Craig MS (2008) Simulation of mine water inflow and gas emission during longwall mining. Rock Mechanics and Rock Engineering. https://doi.org/10.1007/s00603-008-0168-9.

Hemp R (1994) The effect of changes in barometric pressure on mines in the highveld of South Africa. The Journal of the South African Institute of Mining and Metallurgy, 51:133–146

Hyndman RJ, Athanasopoulos G (2021) Forecasting: principles and practice. https://otexts.com/fpp3/. Accessed January 10, 2022.

Johansen S (1995) The vector autoregressive model. In: Likelihood-Based inference in cointegrated vector autoregressive models. Oxford University Press, England.

Karacan CÖ (2007) Modeling and prediction of ventilation methane emissions of U.S. longwall mines using supervised artificial neural networks. International Journal of Coal Geology. https://doi.org/10.1016/j.coal.2007.09.003.

Kirchgässner G, Wolters J (2007) Introduction to Modern Time Series Analysis. Springer, Berlin, Germany.

Kirchgessner DA, Piccot SD, Winkler JD (1993) The estimate of global methane emissions from coal mines. Chemosphere, 26: 453-472.

Kissell FN (2006) Handbook for Methane Control in Mining; National Institute for Occupational Safety and Health (NIOSH), Cincinnati, OH, USA.

Kravchuk K (2017) Forecasting: ARIMAX model exercises (Part-5). https://www.r-bloggers.com/2017/05/forecasting-arimax-model-exercises-part-5/. Accessed March 16, 2022.

Lolon SA (2017) Computational Modeling of barometric pressure fluctuation. Dissertation, Colorado School of Mines, Colorado, USA.

National Institute of Standards and Technology (NIST) (2003) Engineering Statistics Handbook. https://www.itl.nist.gov/div898/handbook/. Accessed March 07, 2021.

Owili-Eger A, Stefanko S, Ramani RV (1973) Simulation of quantity and quality control in mine ventilation. Special Research Report SR-95. The Pennsylvania State University. Pennsylvania, United States.

Prabhakaran S (2019) Augmented Dickey-Fuller Test (ADF Test) – Must Read Guide. https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/. Accessed March 07, 2021.

Schatzel SJ, Karacan CO, Krog RB, Esterhuizen GS, Goodman GVR (2008) Guidelines for the prediction and control of methane emissions on longwalls. The National Institute for Occupational Safety and Health (NIOSH). IC 9502 Information Circular/2008, Pittsburgh, PA, USA.

Seemon T (2014) Basic statistics. Alpha Science International Ltd, Oxford, England.

Shumway RH, Stoffer DS (2017) Time series analysis and its applications, 4th ed. Springer International Publishing. Cham, Switzerland.

Society for Mining, Metallurgy & Exploration (SME) (2021) Coal's importance to the world. https://www.smenet.org/What-We-Do/Technical-Briefings/Coal-s-Importance-in-the-US-and-Global-Energy-Supp. Accessed February 02, 2022.

The National Institute for Occupational Safety and Health (NIOSH) (2020) NIOSH data and statistics. https://www.cdc.gov/niosh/injury/data.html. Accessed February 02, 2022.

Wang W (2020) Mp-Matt: a time series prediction method with Mine Gas Sensor Data. J. Phys.: Conf. Ser. https://doi.org/10.1088/1742-6596/1544/1/012161.

Wasilewski S (2014) Influence of barometric pressure changes on ventilation conditions in deep mines. Archives of Mining Sciences. https://doi.org/10.2478/amsc-2014-0044.

Wei X, Zong Z, Fan X, Li Z (2015) Coal-based products and their uses. Coal Production and Processing Technology. https://doi: 10.1201/b19352-24.

Williams R (2020) Heteroskedasticity. The University of Notre Dame, Indiana, USA. https://www3.nd.edu/%7Erwilliam/. Accessed March 07, 2022.

World Coal Association (2022) Other uses of coal. https://www.worldcoal.org/coal-facts/other-uses-of-coal/. Accessed February 02, 2022.

Xu L, Lin X, Amen J, Welding K, McDermitt D (2014) Impact of changes in barometric pressure on landfill methane emission. Global Biogeochemical Cycles. https://doi.org/10.1002/2013GB004571.

Yuan L, Smith AC (2007) Modeling the effect of barometric pressure changes on spontaneous heating in bleeder's longwall panels. In Proceedings of the Society for Mining, Metallurgy & Exploration. Preprint 07-101. 25-28 February 2007, Denver, Colorado, USA.

6. OVERALL DISCUSSION AND CONCLUSIONS

6.1 Overall Discussion

The study and implementation of forecasting methodologies to predict the concentration and emissions of hazardous gases in underground coal mines, in particular methane gas, have been a topic of interest for academia, the mining industry, and governmental agencies for a long time ago. Consequently, several methodologies have been explored to tackle this problem. For example, the empirical and numerical forecasting approaches. Some have been successfully implemented, capable of predicting methane emissions in particular cases. However, methane concentration forecast methods based on empirical and numerical methodologies are still challenging due to the numerous in-situ characteristics of each mine operation (e.g., production rate and mining parameters, geological characteristics, topography features) that affect methane gas emissions in the underground mining environment.

Nevertheless, the recent technological advances, such as the development and implementation of automated atmospheric monitoring systems to monitor the atmosphere of underground coal mines operations, have facilitated data collection and storage, encouraging the study of statistical approaches to forecasting methane gas concentrations and emissions. These have proved to be more cost-effective and straightforward than empirical and numerical forecasting approaches.

Improving underground coal mine worker safety is the principal reason for developing an accurate methane gas concentration forecasting model(s) based on time series analysis. As coal mines continue to operate at deeper horizons and with higher coal production rates, it becomes even more essential to monitor and manage methane gas concentrations more efficiently and effectively in order to avoid or at least decrease the risk of explosions due to hazardous concentrations of methane gas.

Towards the same direction, the current dissertation addresses the need to develop reliable methane gas concentration forecast models for underground coal mines to safeguard the health and safety of the workforce. More precisely, to study the potential correlation between methane gas emissions in underground coal mines with potential independent variables such as barometric pressure and coal production rate to predict methane concentrations based on implementing univariate and multivariate forecasting techniques using methane gas time series data collected by automated atmospheric monitoring systems and weather data time series retrieved from public weather stations.

The dissemination efforts resulting from this research, which are included above, offer the mining industry and academia an excellent opportunity to understand the main steps involved in developing different forecasting methods for the prediction of methane gas concentrations and emissions in underground environments, especially in underground coal mines as well as the identification of potential research gaps in this field, which should encourage new studies.

6.2 Overall Conclusions

Based upon the discussions and conclusions of the previous chapters of this dissertation, the following list outlines the overall conclusions drawn from the current dissertation:

1. Statistical methane forecasting methodologies can be easily generalized, are less time-consuming, and are less expensive than empirical and numerical methane gas forecasting approaches. The research presented in this dissertation was accomplished by using mine and weather data that were already accessible, saving a substantial amount of time and expenses to obtain physical data (e.g., coal seam gas content and ore body geological characteristics) for each case study.
2. The data inconsistencies (e.g., missing data, erroneous values, and abrupt changes) in the methane gas time series collected and stored by atmospheric monitoring systems in underground coal mines compromise data processing and analysis. This has a direct impact to the outcomes of statistical tests (e.g., variogram function, Pearson linear correlation, and scatterplots) and, consequently, the accuracy and performance of the different forecasting approaches.
3. Three main associations were identified between the time series (e.g., methane gas, barometric pressure, and coal production rate). First, there is autocorrelation in the methane gas time series. Second, there is a strong positive correlation between the methane gas time series and the coal production rate time series; methane gas concentration increases when coal production rates increase and vice versa. Third, there is a significant negative correlation between the methane gas time series and the barometric pressure time series; methane gas concentration decreases when barometric pressure increases and vice versa.
4. A univariate forecasting model, the ARIMA(p,d,q) one-step-ahead model, was developed based on the autocorrelation of the methane time series. The forecasting accuracy and performance were assessed using statistical cross-validation metrics (e.g., Mean Error, the Mean Absolute Error, the Root Mean Squared Error, and the Pearson correlation coefficient). As a result, it was concluded that the ARIMA model can predict methane gas concentrations accurately. For instance, the concentrations of methane gas forecasted match the direction of the validation data; the model was able to forecast directional changes (increase/decrease) in methane concentrations. Moreover, the linear correlation between the forecast and the validation data was strong and positive, and the 95% confidence interval consistently captured the forecast and the validation data.
5. The negative correlation identified between methane gas and barometric pressure time series was employed to develop two multivariate forecasting models capable of effectively predicting future levels of methane gas: the Vector Autoregressive (VAR(p)) and the Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) one-step-ahead models. The performance of each forecast methodology was assessed using validation data and cross-validation metrics. As a result, the concentrations of methane gas forecasted by the VAR(p) and ARIMAX(p,d,q) models match the direction of the validation data. Furthermore, the forecasted methane gas concentration values were trapped consistently by the 95% confidence bound, and the linear correlation between the forecasts and the

validation data was strong and positive. Finally, the value of the cross-validation metrics was similar for both methods.

6. In most cases, the datasets composed of 12-hour average values time series yield better results than datasets comprised of daily average values time series. For example, the linear correlation between the forecast and the validation data was higher, and the cross-validation metrics (e.g., RMSE, ME, and MAE) were lower using 12-hour time series. This can be explained as the time series measured every 12-hour containing more information (number of records) than the ones with daily average values.
7. The performance and accuracy of the three forecast models (ARIMA, VAR, and ARIMAX) were compared using cross-validation metrics to establish the best methane gas forecasting model for underground coal mining operations. It has been concluded that none of the models can uniformly outperform the other forecasting approaches in all datasets. Nonetheless, an algorithm capable of assessing the results of both multivariate and univariate models and selecting the best model among them for the given dataset was developed.
8. The univariate and multivariate methane gas forecasting models proposed in this research offer an exceptional solution to fill the gap of reliable methodologies capable of forecasting methane gas concentrations to improve the safety and health conditions of the workforce in underground coal mining and other underground environments.

6.3 Overall Recommendations

In addition to the conclusions obtained from this body of work, further investigations and improvements of the developed methane gas forecasting models would provide a better assessment of the forecasts and improve their performance. The following actions are recommended to achieve this goal:

1. The atmospheric and the mining monitoring data are stored using an Atmospheric Monitoring Analysis and Database mAnagement (AMANDA) system, a custom relational database designed to manage atmospheric monitoring data. AMANDA is also implemented for data pre-processing. However, this crucial stage is not fully automated; it requires expert human intervention and judgment, making it time-consuming and laborious. Therefore, it is recommended to mechanize the pre-processing data stage to make it more efficient.
2. The statistical data analysis and the development of the different forecast models were carried out by implementing the MATLAB programming environment, which offered excellent results. Nevertheless, nowadays, statistical science and computer applications have developed state-of-the-art open-source programming languages such as Python and R that may be offered additional advantages for statistical analysis and consequently improve the development of methane gas forecasting methods in terms of model accuracy and performance.
3. This research has investigated the univariate and multivariate forecasting approaches to develop three models (ARIMA, VAR(p), and ARIMAX) capable of accurately predicting methane gas concentrations in underground coal mines based on time series data. However, it is recommended to explore implementing more

sophisticated and complex forecasting methodologies such as Artificial Neural Networks, Complex Seasonality, Prophet model, and bootstrapping to compare their accuracy and performance with the forecast models proposed.

4. Two multivariate forecasting approaches (VAR(p) and ARIMAX) were developed to predict future concentrations of methane gas in underground coal mines based on the negative cross-correlation between methane gas and one (barometric pressure) independent variable. It is suggested to include coal production rate as a second exogenous variable and compare the performance of forecasting models with one (barometric pressure) exogenous variable against forecasting models with two (barometric pressure and coal production rate) exogenous variables.
5. Even though methane gas data were retrieved from three case studies (Mines A, B, and C) and, one of the advantages of statistical techniques is that they can be employed in any circumstance or case; this research suggests that more information needs to be collected from different underground coal mines to assess the accuracy and performance of the methane gas forecasting models proposed using different datasets. It will help establish if the different forecast models developed in this research can be broadly implemented, which will help to improve the safety and health conditions of underground environments worldwide, especially underground coal mines operations.
6. Given that different forecasting approaches can be employed to predict methane concentrations in underground coal mines, it is essential to agree on a minimal set of reporting principles that allow performance comparison between different forecasting methods. Therefore, it is proposed that the following critical elements of the data analysis be carefully described: (1) the statistical characterization of the data and the pre-processing protocol, (2) a complete specification of the statistical forecasting model(s), including the values of all the model parameters and the methods used to estimate their values and (3) explicit description of the training and cross-validation practice and presentation of statistical performance measures.

Ultimately, methane gas forecasting approaches can significantly enhance health and safety in underground coal mines and other industrial operations, but further research is essential for maximum benefit to stakeholders.

REFERENCES

Agioutantis Z, Luxbacher K, Dougherty M, Karmis M (2015) Develop and implement a data management platform for atmospheric data in underground coal mines. In Proceedings of the 15th North American Mine Ventilation Symposium, 20–25 June 2015, Blacksburg, VA, USA.

Agioutantis Z, Luxbacher K, Karmis M, Schafrik S (2014) Development of an atmospheric data management system for underground coal mines. J. South. Afr. Inst. Min. Metall, 114:1059–1063.

Agou V.D, Andrew P, Dionissios T.H. (2022) Spatial Modeling of Precipitation Based on Data-Driven Warping of Gaussian Processes" Entropy 24, no. 3: 321. https://doi.org/10.3390/e24030321.

Airey EM (1968) Gas emission from broken coal. An experimental and theoretical investigation. Int. J. Rock Mech. Min. Sci. Geomech, 5:475–494.

AMR PEMCO (2002) Multi-Gas Monitor. https://www.amrpemco.com/mc-6410-multi-gas-monitor/. Accessed October 10, 2021.

Andrews BH, Dean MD, Swain R, Cole C (2013) Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. Society of Actuaries. The University of Southern Maine. Portland, USA.

Astudillo G, Carrasco R, Fernández-Campusano C, Chacon M (2020) Copper price prediction using support vector regression technique Appl. Sci. https://doi.org/10.3390/app10196648.

Barry M, (2013) The canaries in the Coal Mine. European Respiratory Journal. https://doi.org/10.1183/09031936.00137913.

Baur J, Moreno-Villanueva M, Kötter T, Sindlinger T, Bürkle A, Berthold N, Junk M (2015) MARK-AGE data management: Cleaning, exploration and visualization of data. Mech. Ageing Dev. https://doi.org/10.1016/j.mad.2015.05.007.

Bessac J, Naveau P (2021) Forecast score distributions with imperfect observations. Advances in Statistical Climatology, Meteorology, and Oceanography, 7(2), pp.53-71.

Bise CJ (2013) Modern American coal mining methods and applications. (ed) Christopher JB Ed.; Society for Mining, Metallurgy & Exploration (SME), Englewood, CO, USA.

Booth P, Brown H, Nemcik J, Ting R (2017) Spatial context in the calculation of gas emissions for underground coal mines. Int. J. Min. Sci. Technol. https://doi.org/10.1016/j.ijmst.2017.07.007.

Booth P, Nemcik J, Ren T (2016) A critical review and new approach for determination of transient gas emission behavior in underground coal mines. In Proceedings of the 16th Coal Operators' Conference, Mining Engineering, University of Wollongong, Naj Aziz and Bob Kininmonth (eds.) 10-12 February 2016, 367-379, Wollongong, Australia.

Borowski M, Szlazak N, Obracaj D (2009) Methane hazard predictions in underground coal mining. In Proceedings for Deep Mining Challenges: International Mining Forum. 18–21 February, Kraków, Poland.

Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis forecasting and control, 5th (ed). John Wiley & Sons. New Jersey, United States.

Boyer CM, Qingzhao B (1998) Methodology of coalbed methane resource assessment. Int. J. Coal Geol, 35:349–368.

Brillinger DR (2001) Time Series: Data analysis and theory. Elsevier, Philadelphia, USA.

Brockwell PJ, Davis RA (2016) Introduction to time series and forecasting, 3rd ed. Springer. Switzerland.

Broquet G, Pison I, Abbaris A et al. (2018) Identification of spikes associated with local sources in continuous time series of atmospheric CO, CO2, and CH4. Atmos. Meas. Tech. https://doi.org/10.5194/amt-11-1599-2018.

Bustin R M, Clarkson CR (1998) Geological controls on coalbed methane reservoir capacity and gas content. Int. J. Coal Geol, 38:3–26.

Buttrey SE, Lyn RW (2017) A Data Scientist's Guide to Acquiring, Cleaning, and Managing Data in R. JohnWiley & Sons, Incorporated. Hoboken, NJ, USA.

Byrer C, Havryluk I, Uhrin D (2014) Coalbed methane: a miner's curse and a valuable resource. In: Pramod T, Schatzel S, Aminian K (ed) Coalbed methane: from prospect to pipeline, 1st ed. Elsevier, Amsterdam, pp 1–6.

ByungWang J, Rana K (2018) An internet of things system for underground mine air quality pollutant prediction based on azure machine learning. Sensors. https://doi:10.3390/s18040930

Cao S, Liu Y, Wang Y (2008) A forecasting and forewarning model for methane hazard in working face of coal mine based on LS-SVM. J. China Univ. Min. Technol. https://doi.org/10.1016/S1006-1266(08)60037-1.

Chaulya SK, Prasad GM (2016) Sensing and monitoring technologies for mines and hazardous areas, 1st (ed). Amsterdam, Netherlands.

Creedy DP (1993) Methane emissions from coal related sources in Britain: development of a methodology. Chemosphere, 26:419–439.

Cryer JD, Chan KS (2008) time series analysis with applications in R, 2nd ed. Springer. New York, United States.

Curl SJ (1978) Methane prediction in coal mines (IEA Coal Research Report), London, United Kingdom.

De Iaco S, Hristopulos, D.T., Lin, G. (2022) Special Issue: Geostatistics and Machine Learning. Math Geosci 54, 459–465. https://doi.org/10.1007/s11004-022-09998-6.

Diamond WP, Garcia F, Aul G, Ray R (1997) Analysis and prediction of longwall methane emissions: a case study in the Pocahontas No.3 coalbed, V.A. National Institute of Occupational Safety and Health (NIOSH). https://search.cdc.gov/search. Accessed March 15, 2021.

Diaz, JC., Agioutantis, Z., Hristopulos, DT., Luxbacher, K., Schafrik, S. (2022). Time series modeling of methane gas in underground mines. Mining, Metallurgy, and Exploration Journal (submitted).

Diaz J, Agioutantis Z, Hristopulos DT, Schafrik S (2021) Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. Mater. Proc. 2021. https://doi.org/10.3390/materproc2021005078.

Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021a) Towards atmospheric monitoring data analysis in underground coal mines. In: Proceedings of the 18th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781003188476-51.

Dixon DW (1992) A statistical analysis of monitored data for methane prediction. Dissertation, University of Nottingham. Nottingham, United Kingdom.

Dixon DW, Longson I (1993) Statistical method for methane prediction and improved environmental control. In: Proceedings of the 6th North American Mine Ventilation Symposium, pp. 94–98. Salt Lake City, Utah, United States.

Dunmore R (1982) Predicting methane in real-time for longwall faces. In Proceedings of the 1st North American Mine Ventilation Symposium. 29–31 March, Alabama, USA.

Düzgün HS, Leveson N (2018) Analysis of soma mine disaster using causal analysis based on systems theory (CAST). Saf. Sci. https://doi.org/10.1016/j.ssci.2018.07.028.

Ediz IG, Edwards JS (1991) Numerical simulation of time-dependent methane flow. Mining Science and Technology, 12:1–15.

Enders W (2008). Applied Econometric Time Series. John Wiley & Sons.

Esmael B, Arnaout A, Fruhwirth RK, Thonhauser G (2013) A Statistical feature-based approach for operations recognition in drilling time series. Int. J. Comput. Inf. Syst. Ind. Manag. Appl, 5:454–461.

Flores RM (1998) Coalbed methane: From hazard to resource. Int. J. Coal Geol. https://doi.org/10.1016/S0166-5162(97)00043-8.

Ghasemi A, Zahediasl S (2012) Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. https://doi.org/10.5812/ijem.3505.

Goin DE, Ahern J (2019) Identification of spikes in time series. Epidemiol. Method. https://doi.org/10.1515/em-2018-0005.

Goodman GVR, Karacan CÖ, Schatzel SJ, Krog RB, Taylor CD, Thimons ED (2008) NIOSH Research for monitoring and controlling methane at US, underground coal mining operations. In: Proceedings of the 21st World Mining Congress. Krakow, Poland.

Gneiting T, Raftery A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. 102 (447): 359–378.

Griffin R (2013) Utilization and implementation of atmospheric monitoring systems in United States underground coal mines and application of risk assessment. Dissertation, The Virginia Polytechnic Institute, and State University. West Virginia, United States.

Griffin R, Jong EC, Luxbacher KD, Westman EC (2011) A review of atmospheric monitoring systems in underground coal mines: implications for explosion prevention. In: Proceedings for the Society for Mining, Metallurgy, and Exploration (SME), Preprint 11-127, pp. 683–686. Denver, Colorado, United States.

Grimberg F, Asprion PM, Schneider B, Miho E, Babrak L, Habbabeh A (2021) the real-world data challenges radar: a review on the challenges and risks regarding the use of real-world data. Digit. Biomarkers, https://doi.org/10.1159/000516178.

Guo H, Adhikary DP, Craig MS (2008) Simulation of mine water inflow and gas emission during longwall mining. Rock Mechanics and Rock Engineering. https://doi.org/10.1007/s00603-008-0168-9.

Gupta N, Tran D, Devegowda V, Jayaran C, Rai C, Sondergeld C, Karami H (2020) Looking ahead of the bit using surface drilling and petrophysical data: machine-learning-based real-time geosteering involve field. SPE J. https://doi.org/10.2118/199882-PA.

Harkness T (2020) John Graunt at 400. Fighting disease with Numbers. J. Royal Statistical Society. https://doi.org/10.1111/1740-9713.01421.

Hemp R (1994) The effect of changes in barometric pressure on mines in the highveld of South Africa. The Journal of the South African Institute of Mining and Metallurgy, 51:133–146

Hongzhi W, Bah M, Hammad M (2019) Progress in outlier detection techniques: a survey. Inst. Electr. Electron. Eng. (IEEE), 7:107964–108000.

Hristopulos DT (2020) Random fields for spatial data modeling: A primer for scientists and engineers. Springer Nature, Dordrecht.

Hyndman RJ, Athanasopoulos G (2021) Forecasting: principles and practice. https://otexts.com/fpp3/. Accessed January 10, 2022.

Irving W, Tailakov O (1996) CH4 Emissions: coal mining and handling. Good practice guidance and uncertainty management in national greenhouse gas inventories. https://www.ipcc-nggip.iges.or.jp/public/gp/english/. Accessed November 11, 2021.

Johansen S (1995) The vector autoregressive model. In: Likelihood-Based inference in cointegrated vector autoregressive models. Oxford University Press, England.

Jorgensen S, Fath B (2019) Chaos. Encycl. Ecol. https://doi.org/10.1016/B978-008045405-4.00148-8.

Karacan CÖ (2008) Modeling and prediction of ventilation methane emissions of U.S. longwall mines using supervised artificial neural networks. Int. J. Coal Geol, 73:371–387.

Karacan CÖ (2007) Modeling and prediction of ventilation methane emissions of U.S. longwall mines using supervised artificial neural networks. International Journal of Coal Geology. https://doi.org/10.1016/j.coal.2007.09.003.

Karacan CÖ, Diamond WP, Esterhuizen GS, Schatzel SJ (2005) Numerical analysis of the impact of longwall panel width on methane emissions and performance of gob gas ventholes. National Institute for Occupational Safety and Health (NIOSH). https://search.cdc.gov/search. Accessed October 01, 2021.

Karacan CÖ, Ruiz F, Cote M, Phipps S (2011) Coal mine methane: A review of capture and utilization practices with benefits to mining safety and greenhouse gas reduction. Int. J. Coal Geol, 86: 121–156.

Kirchgässner G, Wolters J (2007) Introduction to Modern Time Series Analysis. Springer, Berlin, Germany.

Kirchgessner DA, Piccot SD, Winkler JD (1993) The estimate of global methane emissions from coal mines. Chemosphere, 26: 453-472.

Kissell FN (2006) Handbook for Methane Control in Mining; National Institute for Occupational Safety and Health (NIOSH), Cincinnati, OH, USA.

Kowalski-Trakofler DW, Alexander KM, McWilliams LJ, Reissman DB (2009) Underground coal mining disasters and fatalities-Unites States, 1900-200. MMWR. 57:1379–1382. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5751a3.htm. Accessed November 02, 2021.

Kozlov P (2021) Russian coal mine: dozens killed in Siberia accident, BBC News. https://www.bbc.com/news/world-europe-59421319. Accessed December 01, 2021.

Kravchuk K (2017) Forecasting: ARIMAX model exercises (Part-5). https://www.r-bloggers.com/2017/05/forecasting-arimax-model-exercises-part-5/. Accessed March 16, 2022.

La República (2021) Doce muertos dejó una explosión en mina de carbón en Boyacá, a 190 km de Bogotá. Diario La República. https://www.larepublica.co/economia/doce-muertos-dejo-una-explosion-en-mina-de-carbon-en-boyaca-a-190-km-de-bogota-3222133. Accessed December 01, 2021.

Lolon SA (2017) Computational Modeling of barometric pressure fluctuation. Dissertation, Colorado School of Mines, Colorado, USA.

Lunarzewski LW (1998) Gas emission prediction and recovery in underground coal mines. Int. J. Coal Geol, 35:117–145.

Luxbacher KD, Erdogan SS, Karacan CÖ (2009) Modeling methane emissions and ventilation needs by examining mining-induced permeability changes and related damage to ventilation controls. In Proceedings of the 43rd U.S. Rock Mechanics Symposium and 4th U.S.-Canada Rock Mechanics Symposium. June 28 - July 01, Ashville, USA.

Mahdevari S (2019) Coal Mine Methane: control, utilization, and abatement. In Advances in productive, safe, and responsible coal mining. Elsevier. Amsterdam, The Netherlands.

National Institute of Standards and Technology (NIST) (2003) Engineering Statistics Handbook. https://www.itl.nist.gov/div898/handbook/. Accessed December 07, 2021.

Nielsen A (2019) Practical time series analysis. Prediction with statistics & machine learning preview edition, 1st (ed). O'Reilly Media Inc. San Francisco, United States.

Olayiwola A (2016) Forecasting copper spot prices: A knowledge-discovery approach. Dissertation, The University of Manchester, United Kingdom.

Owili-Eger A, Stefanko S, Ramani RV (1973) Simulation of quantity and quality control in mine ventilation. Special Research Report SR-95. The Pennsylvania State University. Pennsylvania, United States.

Palshikar GK (2009) Simple algorithms for peak detection in time series. In: Proceedings of 1st International Conference on Advanced Data Analysis (IIMA), July 2009. Ahmedabad, Gujarat, India.

Prabhakaran S (2019) Augmented Dickey-Fuller Test (ADF Test) – Must Read Guide. https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/.Accessed March 07, 2021.

Ranjan K, Gaurav B, Rajanarayan P, Debashisha J (2021) Review of preprocessing methods for univariate volatile time series in power system applications. Electr. Power Syst. Res, https://doi.org/10.1016/j.epsr.2020.106885.

Renner S, Wellmer FW (2020) Volatility drivers on the metal market and exposure of producing countries. Miner. Econ. https://doi.org/10.1007/s13563-019-00200-8.

Rodríguez SP (2020) Modeling and forecasting complex patterns of mineral production. Gold mining in Canada. Cogent Eng. https://doi.org/10.1080/23311916.2020.1724849.

Rosienkiewicz M, Chlebus E, Detyna J (2017) A hybrid spares demand forecasting method dedicated to the mining industry. Appl. Math. Model https://doi.org/10.1016/j.apm.2017.04.027.

Schober P, Schwarte LA (2018) Correlation coefficients: Appropriate use and interpretation. Anesth. Analg. https://doi.org/10.1213/ANE.0000000000002864.

Seemon T (2014) Basic statistics. Alpha Science International Ltd, Oxford, England.

Shengrui Z, Baisheng N, Shuiwen L, Hui W, Qian L, Hailong L (2011) Model of gas concentration forecast based on chaos theory. Procedia Eng. https://doi.org/10.1016/j.proeng.2011.11.2160.

Shumway RH, Stoffer DS (2006) Time series analysis and its applications. With R Examples. Springer, New York City, USA.

Shumway RH, Stoffer DS (2017) Time series analysis and its applications, 4th ed. Springer International Publishing. Cham, Switzerland.

Smith T (2021) Autocorrelation. https://www.investopedia.com/terms/a/autocorrelation.asp. Accessed December 04, 2021.

Society for Mining, Metallurgy & Exploration (SME) (2021) Coal's importance to the world. https://www.smenet.org/What-We-Do/Technical-Briefings/Coal-s-Importance-in-the-US-and-Global-Energy-Supp. Accessed February 02, 2022.

Sung W, Ertekin T, Ramani RV (1987) Evaluation of the impact of the degasification process on the inflow of methane gas into coal mines: A numerical exercise. In: Proceedings of the Society for Mining, Metallurgy & Exploration (SME), pp. 328–333. The United States of America.

Tapia Cortez CA, Saydam S, Coulton J, Sammut C (2018) Alternative techniques for forecasting mineral commodity prices. Int. J. Min. Sci. Technol. https://doi.org/10.1016/j.ijmst.2017.09.001.

Tauziede C, Pokryszka Z (1993) Dynamic prediction of CH4 emission in longwalls. In Proceedings of the Conférence Internationale des Instituts de Recherches sur la Sécurité dans les Mines Pretoria, South Africa.

Taylor C, Karacan CÖ (2010) Historical development of technologies for controlling methane in underground coal mines. In: Extracting the Science: A century of mining research. Brune JF (ed), Littleton, CO, Pittsburgh, United States.

The National Institute for Occupational Safety and Health (NIOSH) (2020) NIOSH data and statistics. https://www.cdc.gov/niosh/injury/data.html. Accessed November 02, 2021.

Thomas S (2014) Basic Statistics, 1st (ed). Alpha Science International Ltd. Oxford, United Kingdon.

Tominaga Y, Bandopadhyay S (2002) Monitoring of spontaneous combustion based on time series data. In Proceedings of the 9th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781439833742.ch48.

Trutwin W, Wasilewski S (1987) Application of digital filter and time series analysis in ventilation control systems. In: Proceedings of the 5th North American Mine Ventilation Symposium (NAMVS), pp. 534–540. Morgantown, West Virginia, United States.

Tsay RS (2000) Time series and forecasting: Brief history and future research. J. Am. Stat. Assoc. https://doi.org/10.1080/01621459.2000.10474241.

US EPA. Coalbed Methane Outreach Program. The United States Environmental Protection Agency (US EPA) (2020). https://www.epa.gov/cmop. Accessed November 11, 2021.

Verma U (2019) Data cleaning and pre-processing. Analytics Vidhya. https://medium.com/analytics-vidhya/data-cleaning-and-preprocessing-a4b751f4066f. Accessed November 02, 2021.

Vlachos M, Meek C, Vagena Z, Gunopulos D (2004) Identifying similarities, periodicities, and bursts for online search queries, In Proceedings of the 2004 ACM SIGMOD international conference on Management of Data – SIGMOD. https://doi.org/10.1145/1007568.1007586.

Wang W (2020) Mp-Matt: a time series prediction method with Mine Gas Sensor Data. J. Phys.: Conf. Ser. https://doi.org/10.1088/1742-6596/1544/1/012161.

Wasilewski S (2014) Influence of barometric pressure changes on ventilation conditions in deep mines. Archives of Mining Sciences. https://doi.org/10.2478/amsc-2014-0044.

Watari T, Nansai K, Nakajima K (2020) Major metals demand, supply, and environmental impacts to 2100: A critical review. Resour. Conserv. Recycl. https://doi.org/10.1016/j.resconrec.2020.105107.

Wei X, Zong Z, Fan X, Li Z (2015) Coal-based products and their uses. Coal Production and Processing Technology. https://doi: 10.1201/b19352-24.

Williams R (2020) Heteroskedasticity. The University of Notre Dame, Indiana, USA. https://www3.nd.edu/%7Erwilliam/. Accessed March 07, 2022.

World Coal Association (2022) Other uses of coal. https://www.worldcoal.org/coal-facts/other-uses-of-coal/. Accessed February 02, 2022.

Xu L, Lin X, Amen J, Welding K, McDermitt D (2014) Impact of changes in barometric pressure on landfill methane emission. Global Biogeochemical Cycles. https://doi.org/10.1002/2013GB004571.

Yuan L, Smith AC (2007) Modeling the effect of barometric pressure changes on spontaneous heating in bleeder's longwall panels. In Proceedings of the Society for Mining, Metallurgy & Exploration. Preprint 07-101. 25-28 February 2007, Denver, Colorado, USA.

Yule UG (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfer's Sunspot Numbers. Philos. Trans. R. Soc. London, 226:167–298.

Zagorecki A (2015) Application of sensor fusion and data mining for prediction of methane concentration in coal mines. Mining-Informatics, Automation, and Electrical Engineering ed. 53:33–38.

Zhou C, Damiano N, Whisner B, Reyes M (2017) Industrial Internet of Things (IIoT) applications in underground coal mines. Min. Eng. https://doi.org/10.19150/me.7919.

VITA

Juan Carlos Díaz Martínez

Education

- Master of Science in Advanced Mineral Resources Development at Technische Universität Bergakademie Freiberg, Germany and Montanuniversität Leoben, Austria, June 2019.
- Bachelor of Science in Mining and Metallurgical Engineering at Universidad Nacional de Colombia, Medellin, Colombia, August 2012.

Professional Positions

- Graduate Research Assistant, Department of Mining Engineering, University of Kentucky, US, January 2020-May 2022.
- Graduate Research Assistant, Department of Mining Engineering, Montanuniversität Leoben, Austria, October 2018-March 2019.
- Project Coordinator, Universidad Nacional de Colombia, Colombia, August 2015-December 2015.
- Health and Safety Supervisor, DISCURES S.A.S, Colombia, February 2014-August 2015.
- Junior Mining Engineer, Consortium HGC, Colombia, April 2013-December 2013.

Scholastic and Professional Honors

- Adolf Feizlmyr Stipendien, 2019.
- Colfuturo Scholarship, 2017.

Professional Publications

- Diaz, JC., Agioutantis, Z., Hristopulos, DT., Luxbacher, K., Schafrik, S. (2022). Time series modeling of methane gas in underground mines. Mining, Metallurgy, and Exploration Journal *(submitted)*.

- Diaz, JC., Agioutantis, Z., Hristopulos, DT., Luxbacher, K., Schafrik, S. (2022). Forecasting of methane gas in underground coal mines: univariate vs. multivariate modeling approaches. Process Safety and Environmental Protection Journal *(submitted)*.

- Díaz, JC., Agioutantis, Z., Schafrik, S., Hristopulos, DT., and Luxbacher, K. (2022). Investigating relationships between methane emissions and atmospheric data in underground coal mines to develop a forecasting model. Society for Mining, Metallurgy, and Exploration (SME). Feb. 27 - Mar. 02, 2022, Salt Lake City, UT. Preprint 22-025.
- Diaz, JC., Agioutantis, Z., Schafrik, S., Hristopulos, DT. (2021). Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. Mater. Proc. 2021, 5, 78. https://doi.org/10.3390/materproc2021005078.
- Diaz JC., Agioutantis Z., Schafrik S., Hristopulos DT. (2021). Towards atmospheric monitoring data analysis in underground coal mines. In: Proceedings of the 18th North American Mine Ventilation Symposium. https://doi.org/10.1201/9781003188476-51.
- Restrepo, OJ., and Díaz, JC. (2016). Artisanal miners and their social and environmental situation in Colombia. Society for Mining, Metallurgy, and Exploration (SME). Feb. 21 - 24, 2016, Phoenix, AZ. Preprint 16_014
- Bustamante, N., Danoucaras., N., Mclntyre, N., Díaz, JC., and Restrepo, OJ. (2016). Review of improving the water management for the informal gold mining in Colombia. Revista Facultad de Ingeniería, Universidad de Antioquia. https://doi.org/10.17533/udea.redin.n79a16.
- Jiménez, J., Guarín, M., and Díaz, JC. (2012). Análisis y diseño de la operación de perforación y voladuras en minería de superficie empleando el enfoque de la programación estructurada. Boletín Ciencias de la Tierra.