ALPHA FOUNDATION FOR THE IMPROVEMENT OF MINE SAFETY AND HEALTH

Final Technical Report

Comprehensive atmospheric monitoring in underground coal mines: long-term critical trend analysis and tablet-based communication

Grant No. AFCTG20-103

University of Kentucky Research Foundation

Dr. Zacharias Agioutantis
Phone (859) 257-2953
Fax (859) 323-1962
zach.agioutantis@uky.edu

Period of Performance: September 1, 2020 – November 30, 2022

Acknowledgment/Disclaimer

"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."

Contents

List of Figures

List of Tables

1 Executive Summary

Methane gas management continues to be a challenge concerning underground coal mine safety 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. 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 by using Automated Atmospheric Monitoring Systems, which use sensors that collect massive amounts of data. These data, however, unless properly analyzed, cannot provide the information needed for informed decisions regarding safety and health in the workplace. Therefore, methodologies are required not only for data analysis but also to develop potential risk indicators that allow communication of the results to mine personnel, stakeholders, and decision-makers in real or near real-time more efficiently and straightforwardly.

Atmospheric data were collected from underground mines and analyzed using long-term trend analysis. The goal of this project was to investigate 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. Consequently, the project group developed a unique database that includes underground sensor data and external data for mining operations, such as barometric pressure, temperature, and humidity. 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 the methane gas concentration time series and the independent variables.

Autocorrelation means that each value of the time series (e.g., methane concentration) is related to the values of the same series at previous time instants, i.e., the series has a memory of its past values. This memory property is first quantified during the method estimation stage and then exploited to derive forecasts of methane concentration based on past values. Barometric pressure and production also have an impact on methane concentration. This is expressed by the cross-correlation function between the independent variables and methane gas concentration. Such associations were employed to develop univariate and multivariate forecasting models for methane gas emissions in underground coal mines. It was concluded that the univariate and multivariate methane gas forecasting methodologies developed in this project could reliably predict methane gas concentrations in underground coal mine operations that can lead to enhanced health and safety of mining personnel.

A web-based application that can process real time data was developed. At this time, the web application only utilizes methane emissions time-series data and barometric pressure data, and it does not utilize coal production data. The application can easily be installed and run on a webserver and provide real-time forecasts on methane emissions. Data input consists of real-time local methane emissions measurements and regional real-time barometric pressure data which are typically provided free from public weather stations. The web application can automatically read data from selected public weather stations, and, therefore, the end-user will only need to provide a direct feed of methane measurements at a particular location, e.g., at an exhaust shaft.

2 Problem Statement and Objectives

Monitoring atmospheric conditions in underground coal mines is an important task that helps mine operators control the ventilation systems more efficiently and, therefore, ensures a safe environment for all mine personnel. The layout of the sensors in each mine depends on regulatory requirements, mine geometry, the design of the ventilation system, the availability of power and communication lines to each sensor location, and other such parameters. Currently, several real-time monitoring techniques are available that allow mine operators to monitor all ventilation characteristics, including air velocity, pressure change, and gas concentration at various locations throughout a mine.

Atmospheric monitoring in underground coal mines is mandatory and should be designed and implemented according to existing regulations, i.e., Title 30 of the Code of Federal Regulations (CFR). However, data archiving is not required under current Atmospheric Monitoring System (AMS) regulations (CFR 30 §75.351). Instead, the regulation states that records must be kept regarding alert or alarm signals, AMS malfunctions, and seven-day alert and alarm signals tests. As a result, only a few massive ventilation datasets are available from underground mines in the U.S. This project utilized these data and extracted the valuable information that has been amassed.

Depending on the size and type of the underground mine, atmospheric monitoring should gather data that cover several components sufficient to characterize the underground atmospheric conditions. Such parameters include, but are not limited to, concentrations of various gases (CO, CO_2, CH_4, etc.), wet and dry bulb temperatures, relative humidity, barometric pressure, fan performance indicators, air velocity, and total air pressure loss. Appropriate monitoring becomes even more critical in the case of coal mines where high methane (CH_4) and carbon monoxide (CO) concentrations present immediate and definite hazards for mine personnel. Indeed, Methane gas is the most hazardous flammable gas found in underground coal operations worldwide. Explosions in underground coal mines because of 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.

Implementing real-time atmospheric monitoring in most underground coal mines has significantly enhanced the safety of the workforce. At the same time, atmospheric monitoring data are under-utilized. A generic trend analysis is usually available at most mines, but the literature suggests that such data are not used to establish short or long-term predictive models. Also, there is not enough evidence of software that allows for easy identification of emerging trends in atmospheric monitoring data.

Several researchers (Fauconnier, 1992; Lloyd and Cook, 2004) have reported that methane gas emissions are related to changes in barometric pressure; emissions increase as barometric pressure decreases. However, this relationship is not easily quantified, as high-resolution time series datasets of methane gas emissions are not readily available. This project tackles the

problem mentioned above by developing an Atmospheric Monitoring Analysis and Database Management (AMANDA) system explicitly designed to manage the massive data generated during AMS monitoring (Agioutantis et al., 2014; Agioutantis et al., 2015a,b). This data management system can manage high-resolution measurements (i.e., CH_4 measurements every 3s for multiple sensors) and provides archiving and management functions as well as simple statistical data analysis. Furthermore, the system is capable of managing data from different projects and supports multiple sensor types and multiple tags for a given project (mine).

After the mine and atmospheric data were collected, the data were 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. Their performance was evaluated using cross-validation metrics to determine the best model among different model families for each specific dataset.

The goal of this project is to provide near-real-time information to mine operations personnel with respect to atmospheric conditions at underground mines as well as 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 to develop robust and reliable methane gas forecast models for underground coal mine operations.

The specific objectives of this project are:

1. Conduct background research. 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.
2. Populate the database with atmospheric data from different mining operations. Gather mine data from the case studies, which mainly consist of methane gas and coal production rate, and weather data comprising barometric pressure from the nearest weather station to each case study.
3. Develop automated data reduction techniques aimed at the long-term analysis of atmospheric time series data and their relationships to meteorological and possibly other related time series data.
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.
5. Develop a web subsystem and web interface optimized for real-time or near real-time atmospheric data visualization and alerting for mine personnel.
6. Disseminate and publish the main findings.

3 Research Approach

Data were collected from underground mines and atmospheric data were retrieved from public weather stations. All data were utilized in developing a long-term trend analysis. The analysis mainly aimed to determine factors that indicate developing critical conditions. When these conditions are found in the development stage, they can be mitigated safely, timely, and effectively. Figure 1 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 project.

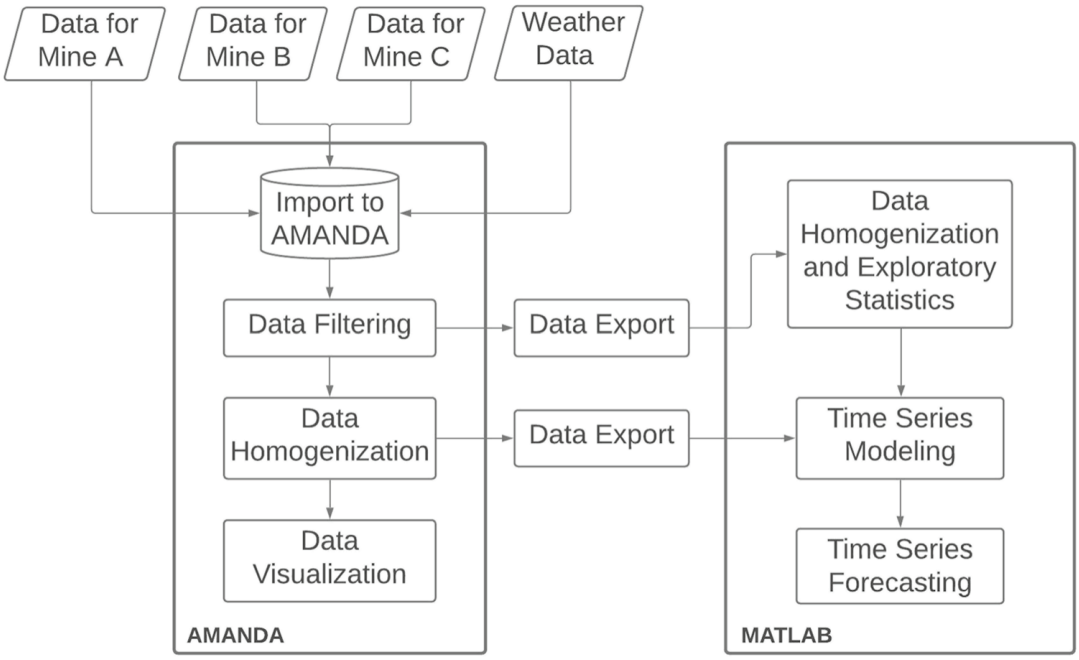


Figure 1: Flow diagram of data management

The first step consists of collecting the mine data and weather data. The mine data, consisting of methane gas time series, were collected from three case studies, which are referred to as Mine A, Mine B, and Mine C. The first case study (Mine A) uses an automated atmospheric monitoring system identified as a wireless multigas monitor (Figure 2) installed on the exhaust shafts. This device can simultaneously monitor four gases (i.e., CH_4, O_2, CO, CO_2). Furthermore, it has various advantages, such as remote operation through a Wi-Fi connection, reduced costs, and being user-friendly because it monitors several gases simultaneously. Moreover, no instruments or special skills are required to replace sensor modules. In addition, its firmware or computer software is updated automatically (Diaz et al., 2022b; AMR PEMCO, 2002).

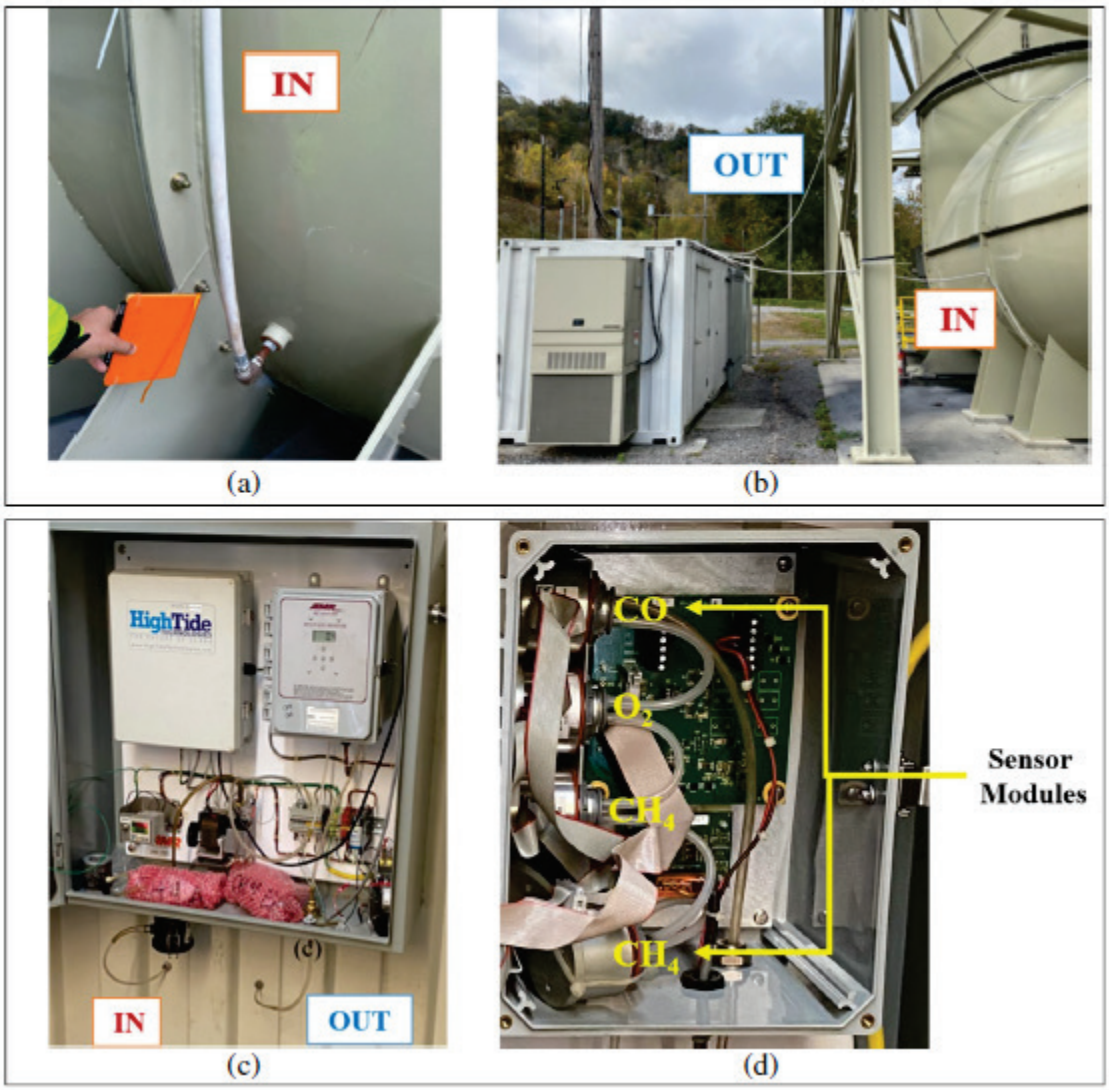


Figure 2: Monitoring system installation at Mine A; a multigas monitor is shown in (c) and (d)

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, 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 gathered manually. Methane gas is measured weekly at the exhaust shaft(s) using a manual process and appropriately recorded.

In addition to the methane gas data collected from the above-mentioned underground coal mines, weather 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).

After the data are collected, the weather and the mining monitoring data are stored in AMANDA (Atmospheric Monitoring Analysis and Database mAnagement), a custom relational database designed to manage AMS data. Details about the AMANDA system are given in Section 4.2.

More detailed information concerning the characteristics (e.g., source, frequency, and units) of the data collected are shown in Section 4.2.

The second step refers to data pre-processing. 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, abrupt changes, and outliers). The filtered data values are flagged as faulty values and are not replaced with zero or null values during this stage. Thus, the original data are preserved and the "cleaned" data can be used in any subsequent data calculations or analyses performed. Examples of data pre-processing are given in Appendix 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. Data homogenization was performed both using AMANDA and the MATLAB® programming environment. As implemented in AMANDA, data homogenization is a straightforward process that can develop 12-h, daily, or weekly averages for each 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 3 illustrates an example of two different data sets homogenized using interpolation. The blue circles represent a set of data on methane gas concentration taken from one of the case studies. In contrast, the white circles correspond to a set of barometric pressure data collected from WU. As illustrated in Figure 3-left, the data points within the same series and from the two different series do not share common and/or regular date/time stamps.

The fourth step of atmospheric data management includes all the processes run on either raw or homogenized data. These range from simple calculations of the Pearson correlation coefficient between two data streams (e.g., methane data vs. barometric pressure, methane data vs. coal production) to linear correlation relationships. Finally, the last step consists of using the different time series (methane gas, barometric pressure, and coal production rate) to run and train the univariate and multivariate forecasting models developed by this project. Results of exploratory data analysis are given in Appendix 2.

More details on data management are provided in Sections 4.1 and 4.2 below.

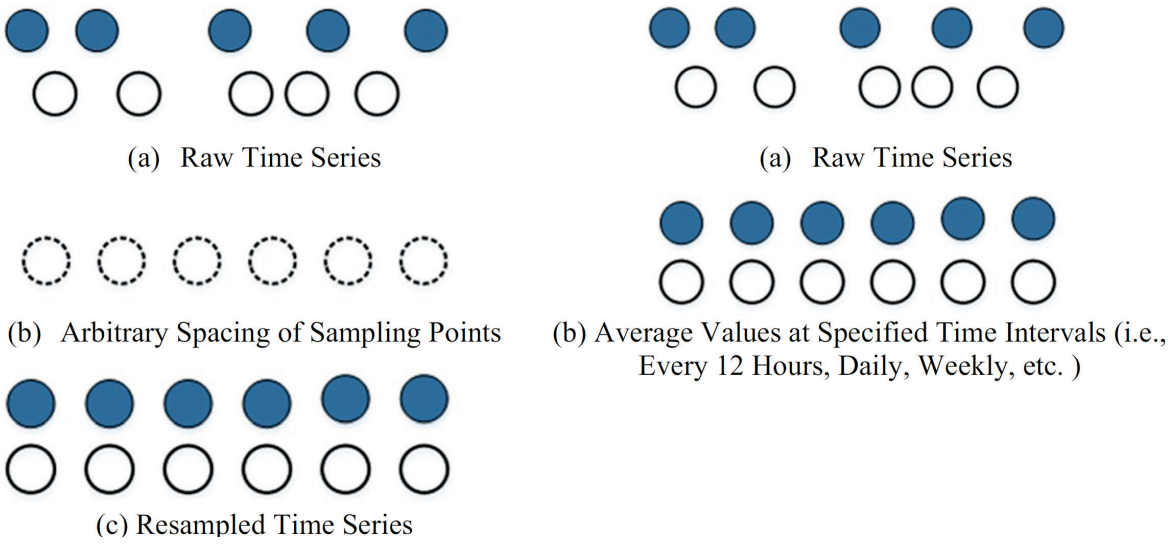


Figure 3: Representation of data homogenization: left: raw time series data with irregular sampling steps, which are converted to resampled data at an arbitrary spacing. Right: raw time series data with irregular spacing, which are converted to average values for specified time intervals

Homogenized data were then analyzed to determine trends. the analysis was mainly performed in the MATLAB environment and was later implemented under Python for the web based application.

A number of approaches were tested as detailed in sections 4.3 and 4.4 below.

Finally, an optimization routine was constructed so that a real-time monitoring system can select the most appropriate methodology to forecast methane concentrations based on the latest dataset. Details are given in section 4.5.

4 Research Findings and Accomplishments

This section outlines the main findings and accomplishments drawn from the current project based on the specific aims and research objectives listed in Section 2.

4.1 Background research and application development

The background research and application development performed in this project determined that methane prediction methods have been a topic of interest for the mining industry and academia for many decades (Airey, 1968; Curl, 1978; Tructin and Wasilewski, 1987; Dixon, 1992; Karacan et al., 2005; Luxbacher et al., 2009). There has been significant progress in monitoring and forecasting methane gas emissions in underground coal mining in recent years due to technological advances in different fields. However, the development of reliable methane gas prediction methods is still a challenge due to the multiple variables and sources that influence methane gas emissions into the underground mining environment (Agioutantis et al., 2015). Consequently, methane gas calculation and forecasting methods are still limited to information origin, and most of them remain empirical (Booth et al., 2016; Booth et al., 2017).

Methane gas forecast techniques can be classified into three categories based on the approach employed (Dixon and Longson, 1993; Borowski et al., 2009). The first category is the empirical approach based on data collected by observing a process or phenomenon for making decisions. Depending on the nature of the research, the data employed can be qualitative or quantitative (Patten, 2005). The second category is the numerical approach, which implements a numerical approximation or mathematical tools to solve physical models (Ramasamy, 1994). 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 for forecasting methane gas emissions and concentration (Brockwell and Davis, 2016).

The background research and review of time-series trends and correlations have highlighted that empirical and numerical approaches for predicting methane gas 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, and expensive, and data collection is challenging. Most importantly, they cannot be broadly implemented because they depend on the geographical and geological conditions for each particular case. 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. In contrast, statistical methane prediction methods (such as time series analysis) are less time-consuming and less 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.

Appendix 3 includes a complete literature review of coal mine methane forecasting by implementing empirical, numerical, or statistical methods for time series analysis.

4.2 Database development and population

Storage of mine and atmospheric data presents challenges as management and utilization of these data is critical to the operation. Optimizing the placement of monitoring systems to acquire the most critical information and transmitting the information to a centralized location with a sophisticated storage method only has value if the data can be understood and utilized by a decision maker.

Data management of an AMS should ensure that the integrity of historical data is maintained. Underground mines have a large range of data being generated. This presents a challenge for the AMS database system, especially with systems that have many sensors reporting a large quantity of data. This does not include other data that can further illuminate trends such as measures of load on mechanized cutting and hauling equipment, production data, etc.

In a paper published in 2014, Agioutantis et al. discussed a relational database application developed to study specific aspects of methane emissions in mines and possible correlations with other collected variables. This application is called Atmospheric Monitoring Analysis and Database mAnagement (or AMANDA) and is designed explicitly for AMS data with capabilities that are discussed in this section. Under this project, AMANDA was updated with new data management and import routines. The advantage of AMANDA was that it could combine data collected from automated mine systems with other data which could be collected outside mine systems.

This data management system has several subsystems, e.g., for data acquisition, data analysis, validation, and storage; visualization and reporting of the data; alarm generation; and tools for statistical evaluation and cross-correlation. Thus, the database system was designed with the following characteristics:

- Deployable on a 64-bit system to allow for large database files;
- Built on a relational database model;
- Implemented as a client/server system;
- Developed with multiple indexing of the data records to ensure a quick response to queries.

In addition, the data management application (AMANDA) allows the following:

- Data can be collected for multiple projects. A project is defined as a collection of sensor data; it can refer to one or multiple mines.
- Sensor types for each projects can be defined in a fully parametric fashion. Each parameter measured by PLC driven sensor is called a tag; multiple user-defined tags can be defined per project; a unique sensor type can be assigned per tag;
- Data files exported by a number of AMS implementations can be easily imported;

- Data files from online weather data systems can be easily imported. Data can include barometric pressure, temperature, etc.,
- Data statistics and identification of missing data can be generated;
- Weekly, daily, 12-h and 6-h averages can be generated for each data stream.

Figure 4 presents a simplified diagram that shows data flow from the sensors to a generic "Mine Database", i.e., a database implemented within a mining operation. AMANDA is external to this data flow and only reads data available through export functions or downloads available by the mine system. This way, AMANDA cannot directly or indirectly interfere with the installed data acquisition system.

More information about AMANDA is presented in Appendix 4.

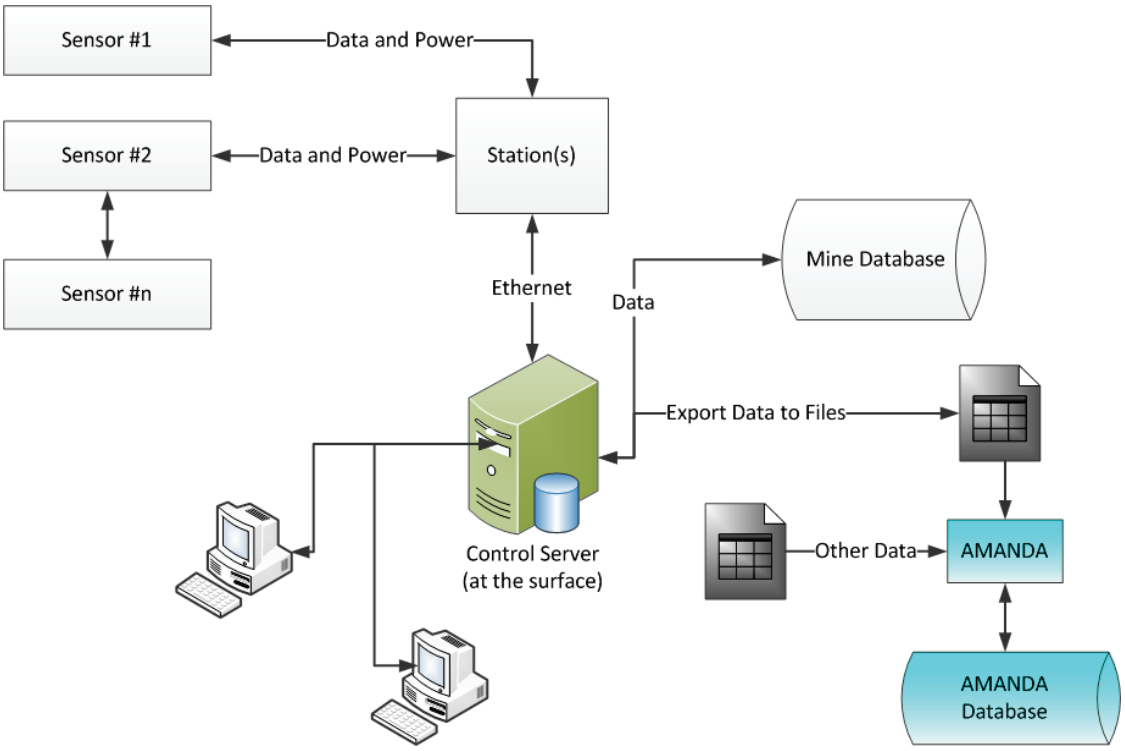


Figure 4: Data flow to the Mine Database and to the AMANDA Database

AMANDA was populated with over one year of continuous data, including measurements (CH_4, CO, and O_2) from Mine A sensors. Furthermore, data from the second mine operation, Mine B were imported. These data include recent measurements as well as historical data since 2012-2013. Moreover, data from the third case study, Mine C, which provides methane data collected at exhaust shafts for different panel districts, were imported. Finally, meteorological data (barometric pressure) by sensors available at the mine location as well as atmospheric data

(barometric pressure, temperature, humidity, etc.) from three weather stations were collected. Table 1 summarizes the stored data. Table 2 presents typical time series for weather data.

Table 1: Collected data

Source		Parameter	Frequency	Total number of records	Units
Mines	Mine A	Coal production	Daily	More than 2,500	Tons
		Methane concentration	Approx. Hourly	More than 250,000	%
		Oxygen concentration	Approx. Hourly	More than 250,000	%
		Carbon monoxide concentration	Approx. Hourly	More than 250,000	ppm
	Mine B	Barometric pressure	Every 10 s	3 mil/year	inWG
		Methane concentration	Every 10 s	3 mil/year	%
	Mine C	Methane concentration	Weekly	374	%
Nearest public weather station	Weather station for Mine A	Barometric pressure	Approx. Hourly	More than 60,000	InWG
	Weather station for Mine A	Barometric pressure	Approx. Hourly	More than 70,000	InWG
	Weather station for Mine A	Barometric pressure	Approx. Hourly	More than 78,000	InWG

Table 2: Typical time series values for weather data

Date	Time	Temp.	Dew Point	Hum.	Wind Speed	BP
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

4.3 Develop and validate long-term relationship(s) between meteorological parameters and methane emissions

The barometric pressure was the variable selected in order to investigate its potential association with methane gas concentration and emissions using correlation, cross-correlation, and autocorrelation techniques. The first step was to homogenize the time series data. This step was necessary because although the data collection frequency for both methane concentration and barometric pressure is approximately hourly, data points are not collected at the exact time within a time interval.

Data homogenization was performed using both 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 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.

In the final implementation of data homogenization, the homogenization process was performed excursively in AMANDA. Homogenized data were exported and used in MATLAB without any additional homogenization process. Data were exported in datasets with specific time periods (i.e., 180 or 360 days) so that all variables exported were complete (i.e., there were no missing values).

The data streams were then imported into the MATLAB environment in order to investigate the potential associations and 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.

The linear correlation between any 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, 2006; 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 on nonlinear dependence), the point-biserial correlation (used when one of the variables is dichotomous), Kendall rank correlation, and Pearson's (linear) correlation coefficient (Shumway and Stoffer, 2006). 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. Furthermore, the cross-covariance also can be 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, 2006). Such relations can be investigated for different time lags between these time series. This helps identify if the association between the two variables exhibits a time delay. The Pearson's (linear) correlation coefficient and the cross-covariance were implemented to assess the potential correlation between methane gas and barometric pressure.

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.

In addition, the potential autocorrelation of the methane gas concentration time series was also evaluated implementing two techniques; the autocorrelation function (ACF) plots and the variogram function. 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, 2006; Thomas, 2014). The 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 times (Shumway and Stoffer, 2006). Figure 5 presents an autocorrelation plot for non-stationary time series. The very slow decay of the auto-correlation function is a sign of the non-stationarity of the time series. In the stationary case, the ACF quickly decays from 1 to values near zero as the time lag increases.

A time series whose statistical properties change over time is called a non-stationary time series. Thus, a time series with a trend or seasonality is non-stationary in nature. This is because the presence of trend or seasonality will affect the mean, variance and other properties at any given point in time.

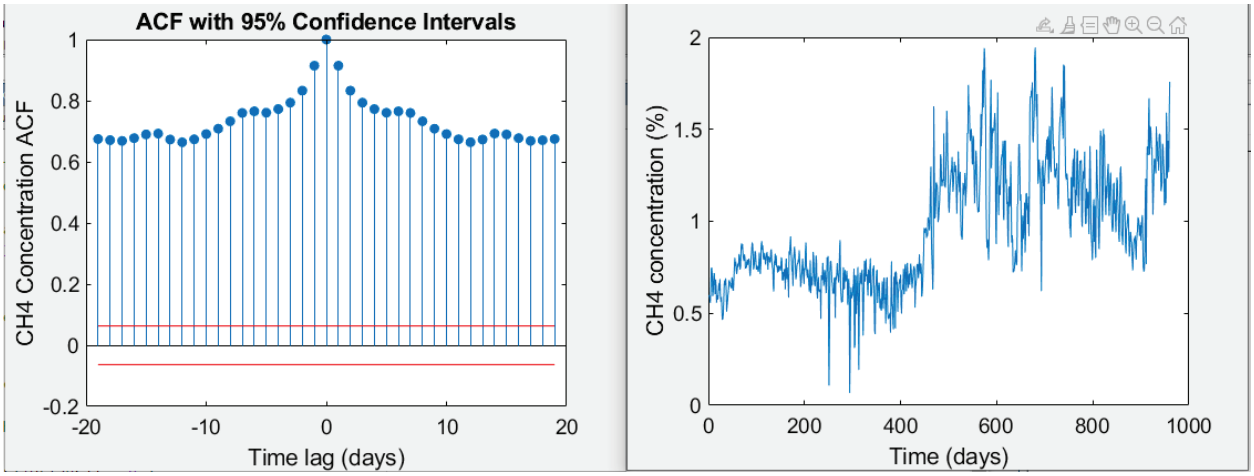


Figure 5: Methane gas autocorrelation plot (left) for a non-stationary time series (right).

Figure 6 (left) presents the autocorrelation plot for barometric pressure which is a stationary time series (right). The function plotted quickly decays from 1 to values near zero as the time lag increases.

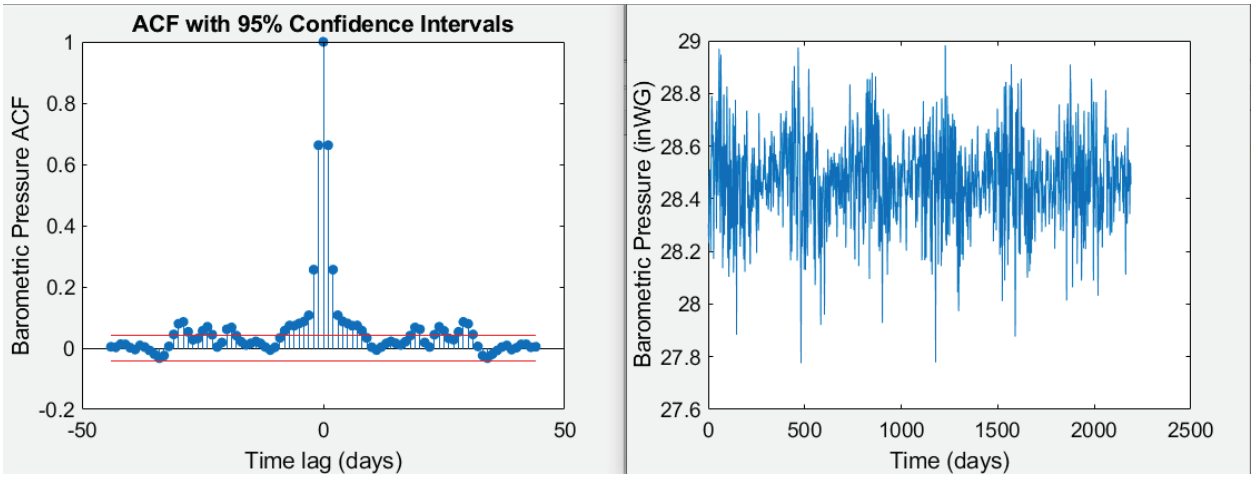


Figure 6: Barometric pressure autocorrelation plot (left) for a stationary time series (right).

Similarly, 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, 2006; 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+τ)-X(t), where τ is the time lag. If τ=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. See Figure 62 to Figure 66 for a number of variogram plots from two different mines.

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. Furthermore, another advantage of the variogram is that the differencing operation implied in calculating 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. 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).

It was determined that interpreting the autocorrelation of the methane gas concentration from Mines A and B using the ACF plot was complex due to the non-stationary nature of the time series (Figure 5). 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 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.

A series of graphs were generated when evaluating the potential correlation between methane gas concentration vs. barometric pressure and methane gas concentration vs. coal production using the Pearson correlation coefficient. Figure 7 presents daily averages for three variables for a specific time span. The graph was generated in AMANDA.

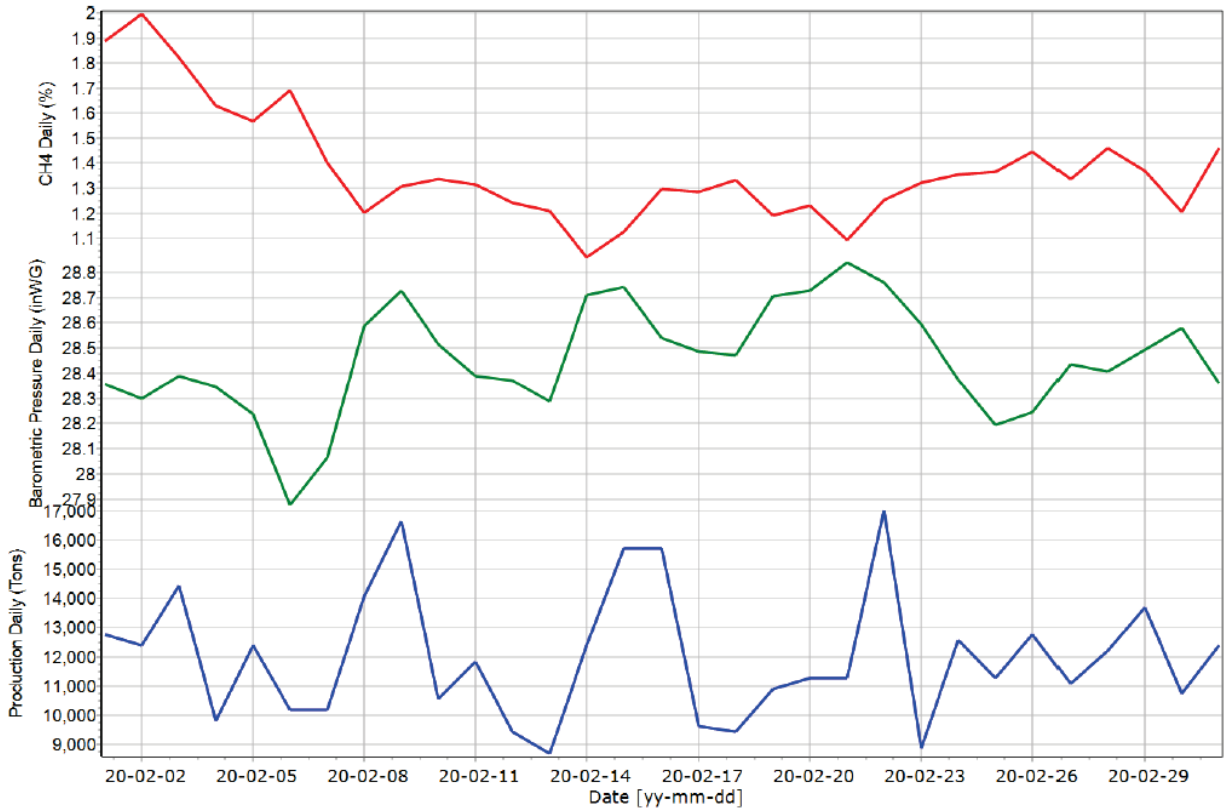


Figure 7: Daily average time series data of methane gas concentration (top plot; red), barometric pressure (middle plot; green), and coal production rate (bottom plot; blue)

4.4 Develop univariate and multivariate methane gas forecasting methodologies

4.4.1 Overview

Different univariate and multivariate forecasting approaches have been selected to validate the long-term relationship(s) between methane emissions and meteorological parameters and other independent variables such as coal production rate. Furthermore, methane gas concentrations were successfully forecasted using the time series data retrieved from the case studies. The three forecasting approaches used are discussed in this section. They involve the

- the univariate Autoregressive Integrated Moving Average (ARIMA(*p,d,q*)) model
- the Multivariate Vector Autoregressive, VAR(p), and
- the Autoregressive Integrated Moving Average with Exogenous input, ARIMAX(*p,d,q*), models.

All three models were implemented in the MATLAB® environment employing functions from the econometrics toolbox. Sample datasets from the three case studies covering one and six years were used to build the respective forecast models.

The main goal of time series analysis is to predict the future values of an observed variable as reliably as possible based on the available data. Forecasting models based on time series analysis are generally classified into three categories:

- Subjective Forecasting Models founded on judgment, opinion, or intuition,
- Univariate Forecasting Models which employ past values of a given time series to predict its future value; and
- Multivariate Forecasting Models

The models are based on values of one or more time-series to predict the value of a specified variable (Shumway and Stoffer, 2006; NIST, 2003; Hristopulos, 2020). 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 variable (Shumway and Stoffer). ARIMA models are flexible univariate stochastic methods that handle both stationary and non-stationary time series. Therefore, the ARIMA model can analyze methane gas time series.

4.4.2 ARIMA Models

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 to stationary.

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 the three case studies used to construct ARIMA(*p,d,q*), the VAR(*p*) and the ARIMAX(*p,d,q*) models in the MATLAB environment. 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 “training set” and is used to train the model (i.e., to estimate the optimal parameters); the other is called “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 univariate and multivariate models 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).

In the case of the multivariate models it was found that the forecast performance of the traditional ARIMA(*p,d,q*) model 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). Therefore, it was concluded that the regular ARIMA model does not provide reliable forecasts of methane gas concentration over a time horizon involving many future steps. Consequently, a different forecasting technique based on ARIMA models had to be explored. The ARIMA(*p,d,q*) one-step-ahead model was chosen to predict the concentration of methane gas.

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 reliable one-step-ahead forecasts, the following cross-validation methodology was 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.

In general, in time series modeling, there are two criteria that can be used to determine the optimum model. The Bayesian information criterion (BIC) that measures the trade-off between model fit and complexity of the model and the Akaike Information Criterion (AIC). A lower AIC or BIC value indicates a better fit (Mohammed et al, 2015). The Bayesian Information Criterion (BIC) is more useful in selecting a correct model while the AIC is more appropriate in finding the best model for predicting future observations (Chakrabarti and Gosh, 2011).

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 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 3. 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. (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.

$$\left[\hat{x}(t_p) - 1.96\sqrt{\mathrm{MSE}(t_p)}, \hat{x}(t_p) + 1.96\sqrt{\mathrm{MSE}(t_p)}\right] \quad \text{Eq. (1)}$$

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. 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.

It was determined that the ARIMA one-step-ahead model provides reliable 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 more extended methane time series (six years) than shorter ones (1 year) to train the ARIMA one-step-ahead model. It should be noted that potential users of this methodology do not need six-year datasets to achieve a good correlation and reliable forecasts. However, it should be noted, that an increase in the length of the dataset will potential increase the quality/reliability of the forecast.

The methane time series collected with the 12-hour average time step provides a more reliable 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.

A series of graphs were generated by employing the ARIMA (*p,d,q*) one-step-ahead approach to forecasting methane gas concentrations using a daily and 12-hour average time stamp. Diagrams that show correlation data are included in Appendix 5.

Statistical approaches, particularly time series analysis, have recently been used to forecast methane gas in underground coal mines. Such approaches take advantage of inherent correlations in the time series. Correlations reflect the "memory" of the process and allow the formulation of future probabilistic predictions based on information from the past, such as the ARIMA model presented and discussed previously (Diaz et al., 2021a,b; Dixon, 1992; Dixon and Longson, 1993). However, other research has demonstrated that methane gas concentration and emissions are correlated with auxiliary variables (covariates) such as barometric pressure and coal production rate (Hemp, 1994; Xu et al., 2014; Wasilewski, 2014; Lolon, 2017; Yuan and Smith, 2010; Diaz et al., 2022a,b). For example, a strong negative correlation between methane gas and barometric pressure has been found in many cases: methane gas concentration increases when barometric pressure decreases and vice versa. On the other hand, the correlation between methane gas and coal production rate is generally positive: methane gas concentration increases when coal production rate increases and vice versa (Diaz et al., 2021a,b, 2022a,b).

4.4.3 VAR(p) models

The VAR(p) model is a flexible and easy to numerically implement approach for analyzing multivariate time series (Johansen, 1995; Lütkepohl, 2005). 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. (2). More information about the VAR(p) model is given in (Johansen, 1995; Lütkepohl, 2005; Kirchgässner and Wolters, 2007).

The equations for the VAR(1) model of order *p = 1* for two-time series, denoted by *xt* and *yt* are presented in Eq. 2:

$$x_t = \emptyset_{1,0} + \emptyset_{1,1}\, x_{t-1} + a_{1,1}\, y_{t-1} + u_t\,, \qquad \text{Eq. (2a)}$$

$$x_t = \emptyset_{2,0} + \emptyset_{2,1}\, y_{t-1} + b_{1,1}\, x_{t-1} + v_t\,. \qquad \text{Eq. (2b)}$$

In Eq. (2), *xt, yt* correspond to CH_4 concentration and barometric pressure, respectively; $\emptyset_{1,0}, \emptyset_{1,1}, \emptyset_{2,0}, \emptyset_{2,1}$ and $a_{1,1}, b_{1,1}$ are time-independent model parameters, while ut, vt are independent Gaussian white noise processes that represent the innovation terms for each series. The optimal *p value* is obtained by means of

$$p^* = \arg min \quad AIC[VAR(p)],$$

$$p \in \{1,2,\ldots.,pmax\}$$

where AIC[VAR(p)] is the AIC value for the VAR(p) model and *pmax* is the maximum autoregressive order considered.

The ARIMAX(*p,d,q*) model incorporates dependence on one or more explanatory variables (Andrews et al., 2013; Kravchuk, 2017). The general ARIMAX equation in the case of one explanatory variable is:

$$\Phi(B)[\nabla^d X_t] = \beta y_t + \theta(B)\epsilon_t \qquad \text{Eq. (3)}$$

Where y_t is the explanatory variable, β is a constant coefficient, and $\Phi(B)$, $\theta(B)$ are defined in Eq. (2).

4.4.4 ARIMAX Models (with 1 exogenous variable)

ARIMAX 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 potential autocorrelations to enhance the accuracy of the forecasts (Hyndman and Athanasopoulos, 2021; Wang, 2020). ARIMAX is used in order to take advantage of the negative correlation observed between CH_4 concentrations and barometric pressure. The ARIMAX(*p,d,q*) model is applied herein with d = q = 0 in order to compare its performance with the VAR(p) model. The optimal ARIMAX model has an autoregressive order $p*$ which is established based on AIC.

The same general observations (trends) that apply to the VAR(*p*) and ARIMA models also hold for the ARIMAX model as well, albeit there are some differences in the validation measures between different methods. This outcome is not surprising, since all three methods exploit to a large extent the auto-correlations of the methane concentration series. VAR(*p*) is a vector model that incorporates the atmospheric pressure and assumes an autoregressive formulation. ARIMA, on the other hand, is a univariate method that accounts only for the methane concentration. However, its advantage is that it allows more flexibility in modeling the autocorrelations than the VAR(*p*) model. The ARIMAX model combines the ARIMA autocorrelations with an exogenous input. In this case, the latter is the atmospheric pressure. Analysis results resonate with the common experience in time series analysis, namely that no single model is optimal for all datasets.

It was also concluded that the ability of the univariate ARIMA, multivariate VAR, and ARIMAX models to predict future concentrations of CH_4 is influenced by data irregularities such as discontinuities, faulty values, and abrupt changes (Figure 8) as well as the averaging time step (e.g., 12-hour or daily average values), and the presence of independent variables (e.g., coal production rate) not accounted for the models with one exogenous (auxiliary) variable. Due to such factors, a single model is not able to deliver consistently the best results for all datasets. Consequently, it was required to develop a methodology for selecting the best forecast model based on cross-validation analysis. This methodology is discussed at the end of section 4.4.3.

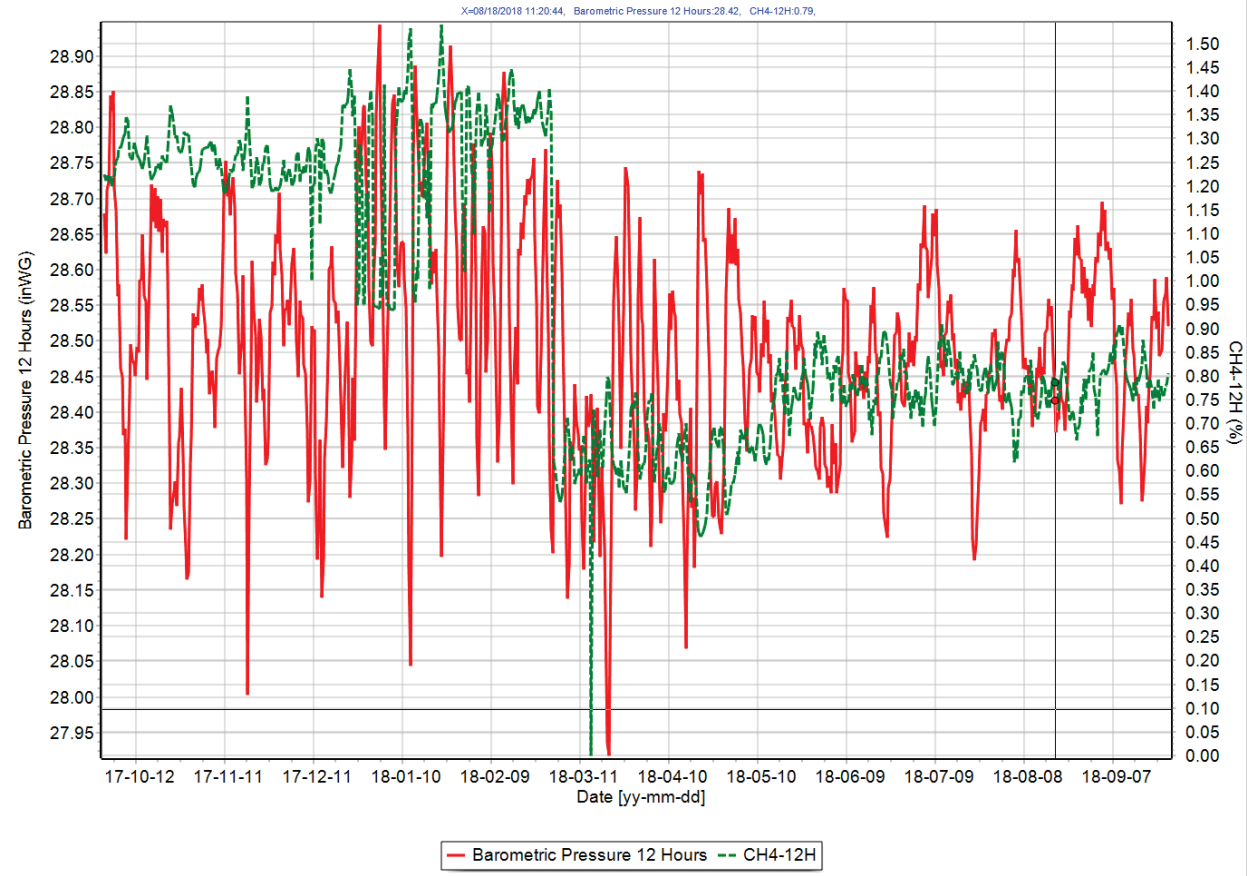


Figure 8: Methane gas concentration (red line) and barometric pressure time series (red line) for 360 days. The correlation coefficient is very poor (close to zero)

4.4.5 ARIMAX Models (with 2 exogenous variables)

A few models were developed that utilized two exogenous variables with the ARIMAX(*p,d,q*) framework. Coal production data that corresponded to long methane time series segments were not always available. In addition, even when utilizing daily averages, there were missing data for coal production, i.e., when the mine(s) stopped producing over weekends or holidays. Figure 9 presents the forecast for a 30-day dataset using the ARIMAX method and only utilizing barometric pressure as the exogenous (external or independent variable). The correlation coefficient for this prediction is R=0.86.

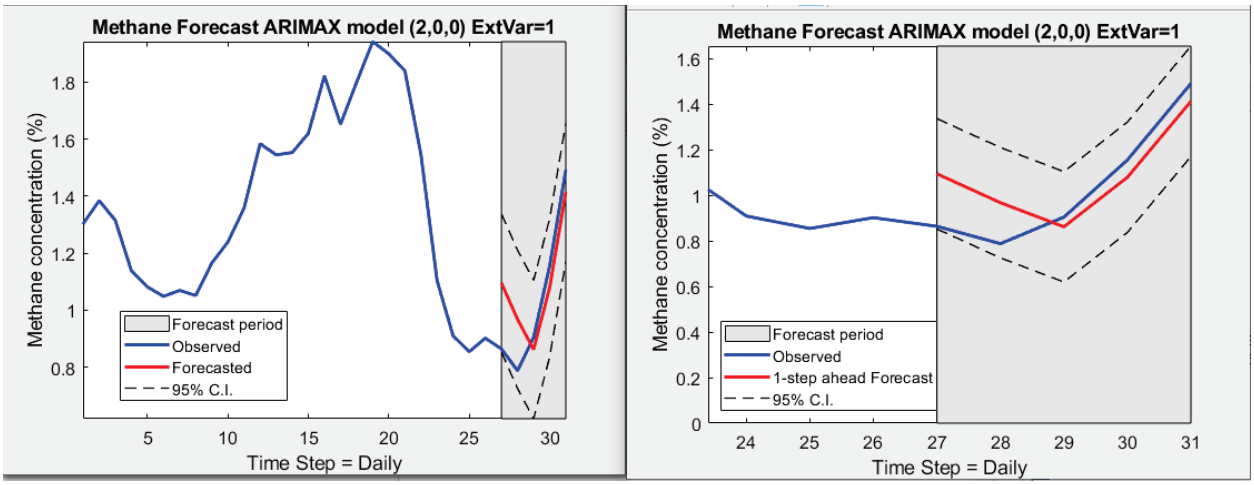


Figure 9: Forecast using the ARIMAX(p,d,q) method and barometric pressure for a 30-day timeseries

Figure 10 presents the forecast for the same methane timeseries using two exogenous variables, i.e. barometric pressure and coal production. In this case, the correlation coefficient is higher (R=0.93).

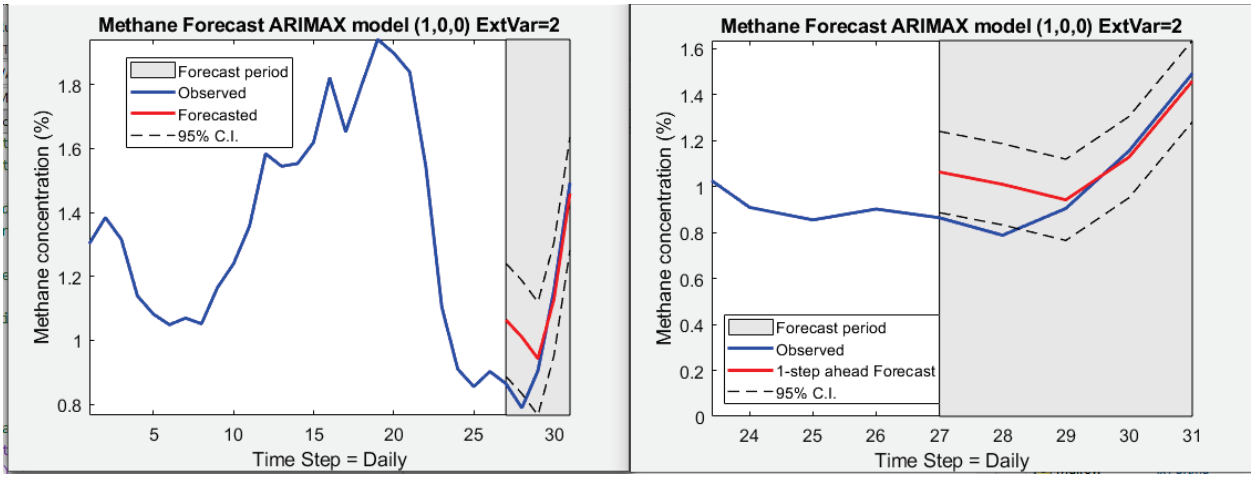


Figure 10: Forecast using the ARIMAX(p,d,q) method, barometric pressure and coal production for a 30-day timeseries

4.4.6 Results

This section presents the results of CH_4 concentration forecasting using VAR(*p*) and ARIMAX(*p,d,q*) for four datasets (from 1 to 4). Each dataset comprises two time series: (i) the CH_4 concentration time series (dependent variable) and (ii) the barometric pressure time series (explanatory variable). The methane data were retrieved from Mine A and the barometric pressure from the nearest weather station. The first two datasets (1 and 2) represent daily average values and consist of one year and six years of data, respectively. Datasets 3 and 4 are derived from the same time series as datasets 1 and 2, respectively, but represent 12-hour average values. Note that twelve datasets were analyzed in total with both models (six datasets use daily average values and six used 12-hour averages). For reasons of brevity, this section

focuses on representative results from datasets 1-4. A summary of the results obtained with all 12 datasets is shown in Table 3).

It should be noted that although coal production has an impact on methane concentration, the collection and utilization of coal production data presented a number of challenges:

a) Coal production data were not always readily available.
b) In many cases, coal production data did not have a clear relationship to the methane gas monitoring station.
c) Coal production may be zero (or vey low) on weekends and holidays.

Nonetheless, since coal production has an impact on methane concentration, the impact of coal production is indirectly included to some extent in the auto-correlation of the methane concentration. Using either the auto-correlation or cross-correlation with barometric pressure good forecasts of methane concentration can be obtained. These can be further improved if consistent data on coal production become available.

Table 3: Summary of univariate and multivariate one-step-ahead cross validation measures. L: Length of dataset (in number of days). Ns: Sample size (number of points). δt: Time step. No.: Number of dataset (1-12).

No.	Data Segment Features					Univariate					Multivariate									
	δt	Training		Validation		ARIMA(p,d,q)					VAR(p)					ARIMAX(p,d,q)				
		L	N_s	L	N_s	R	(p,d,q)	RMSE	ME	MAE	R	p	RMSE	ME	MAE	R	$p_{\max}$	RMSE	ME	MAE
1	Daily	365	365	18	18	0.89	(4,1,4)	0.54	−0.19	0.44	0.89	2	0.50	−0.12	0.44	0.87	28	0.58	−0.20	0.49
2	Daily	2,200	2,200	109	109	0.65	(3,1,4)	0.04	0.00	0.03	0.66	13	0.05	−0.01	0.04	0.65	23	0.05	−0.01	0.03
3	12 hr	365	730	18	37	0.90	(4,1,4)	0.47	−0.11	0.35	0.91	8	0.46	−0.06	0.32	0.91	8	0.46	−0.06	0.33
4	12 hr	2,200	4,380	109	219	0.71	(4,1,4)	0.05	0.00	0.04	0.66	30	0.05	−0.01	0.05	0.68	29	0.05	−0.01	0.04
5	Daily	365	365	18	18	0.54	(2,0,2)	0.05	0.00	0.04	0.34	3	0.06	−0.01	0.05	0.53	3	0.06	0.00	0.04
6	12 hr	365	730	18	37	0.33	(1,1,1)	0.08	0.00	0.06	0.35	11	0.08	−0.01	0.06	0.34	10	0.09	−0.01	0.06
7	Daily	365	365	18	18	0.33	(1,0,2)	0.05	0.00	0.04	0.57	3	0.04	0.00	0.03	0.41	3	0.05	0.00	0.03
8	12 hr	365	730	18	37	0.72	(3,1,4)	0.04	0.01	0.03	0.78	8	0.03	0.00	0.02	0.72	8	0.04	0.00	0.03
9	Daily	365	365	18	18	0.65	(3,0,4)	0.02	0.00	0.01	0.50	5	0.02	0.02	0.00	0.68	16	0.02	0.00	0.01
10	12 hr	365	730	18	37	0.52	(2,1,1)	0.02	0.00	0.02	0.48	12	0.03	0.00	0.02	0.54	17	0.02	0.00	0.02
11	Daily	365	365	18	18	0.79	(1,1,2)	0.20	0.08	0.16	0.79	3	0.19	0.04	0.15	0.83	25	0.18	0.07	0.14
12	12 hr	365	730	18	37	0.85	(1,0,1)	0.17	0.01	0.13	0.83	9	0.18	0.01	0.14	0.84	15	0.18	0.02	0.13

Figure 9 displays the forecasting results, employing the multivariate VAR(p) one-step-ahead approach, for the daily-averaged datasets (1 and 2). Figure 9a shows the training and validation data as well as the forecasts for dataset 1. The optimal autoregressive order is p = 2 based on the lowest AIC. The magnified segment in Figure 9b demonstrates that the forecast (continuous line) is close to the true values (dash-dot line) in the validation period. The CV correlation coefficient (henceforth, R) is calculated at R = 0.89, implying a strong correlation between the validation data and the forecasts. Moreover, the true values are captured by the 95% prediction interval (the interval's bounds are given by the dashed lines). It is concluded that the VAR(2) model provides a reliable forecast for dataset 1.

Figure 9c displays the forecast for the longer dataset 2. In this case the optimal model is VAR(13). Figure 9d demonstrates 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 sharp fluctuations), as shown in Figure 9c. Nevertheless, the VAR(13) model reliably follows the fluctuations of the validation data, which lie inside the 95% prediction interval.

Furthermore, the cross-validation measures (except for R) are significantly better than those for dataset 1, as presented in Table 4.

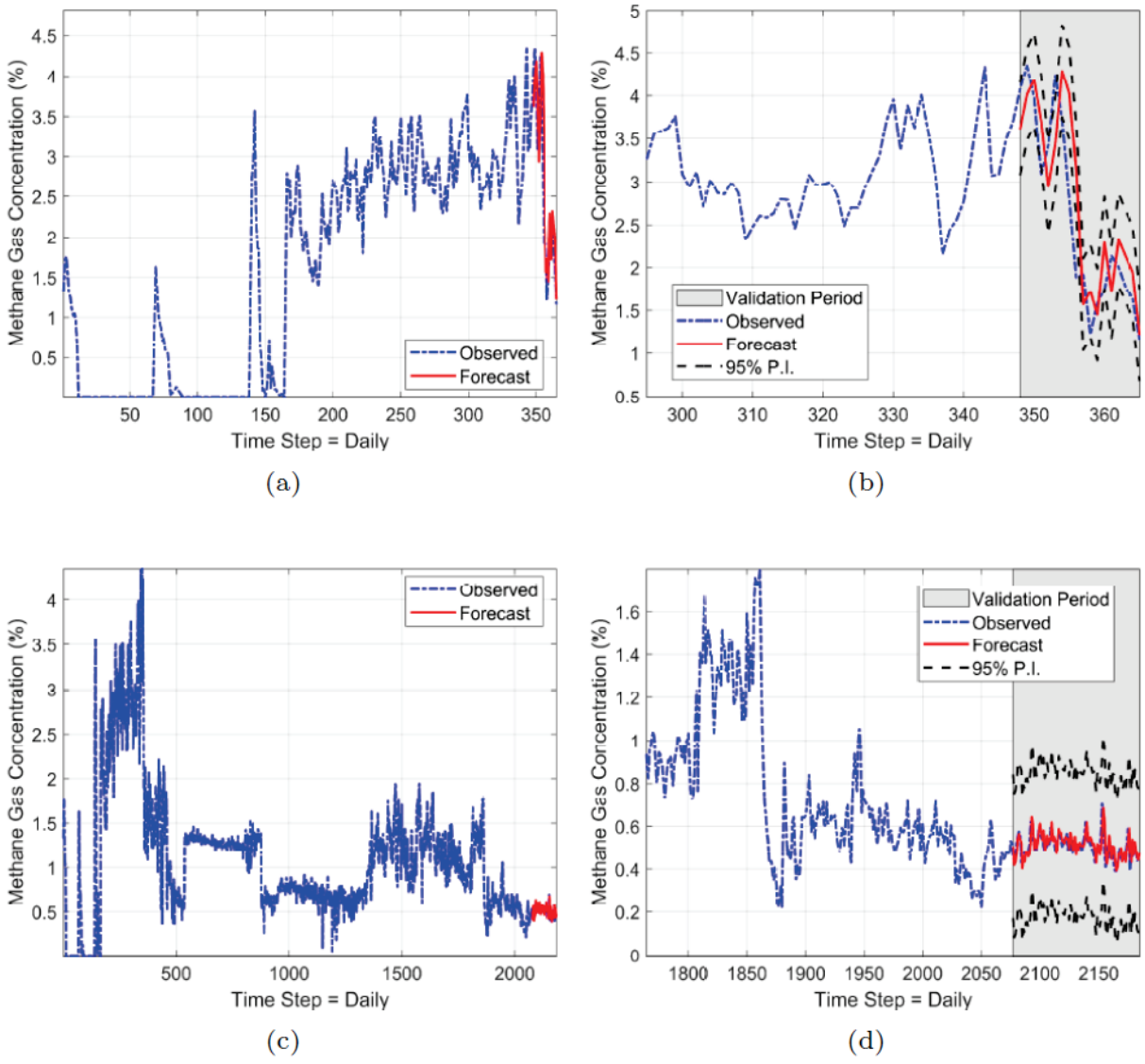


Figure 11: VAR(p) one-step-ahead CH4 concentration forecasts based on daily average values: (a) Forecast for dataset 1; (b) Magnified view of the forecast in (a); (c) Forecast for dataset 2; (d) Magnified view of the forecast in (b).

Table 4: Summary of CH4 concentration validation measures for the VAR(p) one-step-ahead model (using a bivariate vector of methane gas concentration and barometric pressure). p∗: Optimal autoregressive order of VAR(p) model based on AIC.

	Time Step	Training Data Length (days)	Training Data Sample Size	Validation Data Length (days)	Validation Data Sample Size	p^*	R	RMSE	ME	MAE
Dataset 1	24-hour	365	365	18	18	2	0.89	0.50	−0.11	0.44
Dataset 2	24-hour	2,200	2,200	109	109	13	0.66	0.05	−0.01	0.04
Dataset 3	12-hour	365	730	18	37	8	0.91	0.46	−0.06	0.32
Dataset 4	12-hour	2,200	4,380	109	219	30	0.66	0.05	−0.01	0.04

Figure 10 shows the VAR(p) forecast for datasets 3 and 4 (these use 12-hour average values). Dataset 3 and the respective forecasts are shown in Figure 10a. VAR(8) is the optimal model in this case. Figure 10b reveals that the forecasts closely follow the validation data; the CV correlation is R = 0.91, higher than for the daily averages (see Figure 9a and Figure 9b). Figure 10c presents the forecast for the longer dataset 4. In this case the optimal model is VAR(30). Visual inspection of Figure 10d indicates that the forecast and the validation have similar variations. The CV correlation is R = 0.66, i.e., the same value as for the daily-averaged time series shown in Figure 9c. The CV measures obtained for dataset 4 are significantly better than those achieved in dataset 3, except for R, as shown in Table 5. This is due to the fact that the CH_4 variations in the validation segment of the dataset 4 are considerably smoother than those in the respective segment of dataset 3. In the latter, the concentration drops from ≈ 5% to ≈ 0.5%, while in the former case all concentration values are > 0.4% and < 0.8%.

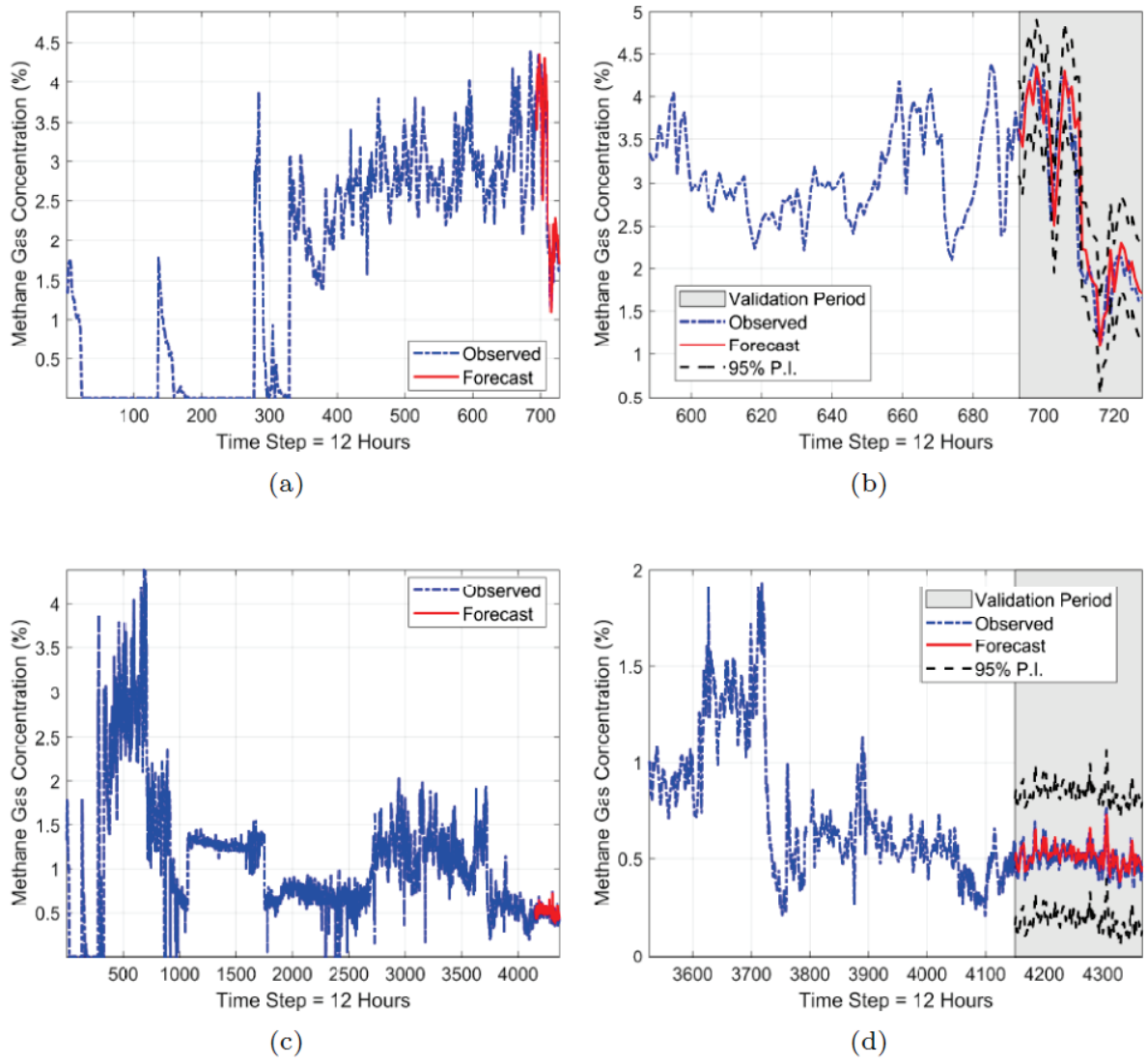


Figure 12: VAR(p) one-step-ahead CH4 concentration forecasts based on 12-hour averages: (a) Forecast for dataset 3; (b) Magnified view of the forecast in (a); (c) Forecast for dataset 4; (d) Magnified view of the forecast in (b).

Table 5: Summary of CH4 concentration validation measures for the ARIMAX one-step-ahead model (using barometric pressure time series as an independent variable). $p*$: Optimal autoregressive order of ARIMAX model based on AIC.

	Time Step	Training Data		Validation Data		p^*	R	RMSE	ME	MAE
		Length (days)	Sample Size	Length (days)	Sample Size					
Dataset 1	24-hour	365	365	18	18	28	0.87	0.58	−0.20	0.49
Dataset 2	24-hour	2,200	2,200	109	109	23	0.65	0.05	−0.01	0.03
Dataset 3	12-hour	365	730	18	37	8	0.91	0.46	−0.07	0.33
Dataset 4	12-hour	2,200	4,380	109	219	29	0.68	0.05	−0.01	0.04

Figure 11 comprises four plots that represent CH_4 forecasts obtained using the ARIMAX(*p,d,q*) model. The plot segments with the white background contain the training data, the black dashed lines mark the boundaries of the 95% prediction intervals, and the segment in the shaded background contains the validation data (dash-dot line, blue line) and the forecast (continuous line, red line). Figure 11a shows the ARIMAX forecast for dataset 1. The optimum model is achieved with p=28. Visual inspection of the plot demonstrates that the forecast (continuous line, red line) is close to the true validation data (dashdot line, blue line), and the latter lie within the 95% prediction interval. The CV correlation is R = 0.87 (Table 6) showing strong correlation between the validation data and the forecasts. Figure 11b displays the forecast for dataset 2. The optimal model is obtained for p = 23. The CV correlation is R = 0.65.

The forecast consistently follows the CH_4 validation peaks, and the validation data are contained in the 95% prediction interval. Figure 11c and Figure 11d present the forecast for the datasets 3 and 4 (12-hour averages), respectively. The best model for the data in Figure 11c has p = 8. The correlation coefficient between the forecasts and the validation data is R = 0.91, exceeding the value obtained for the daily-average samples (Figure 11a). For dataset 4, shown in Figure 11d, the optimal model is obtained for p=29. The correlation coefficient is R = 0.68, which is slightly higher than for the daily-averaged time series shown in Figure 11b.

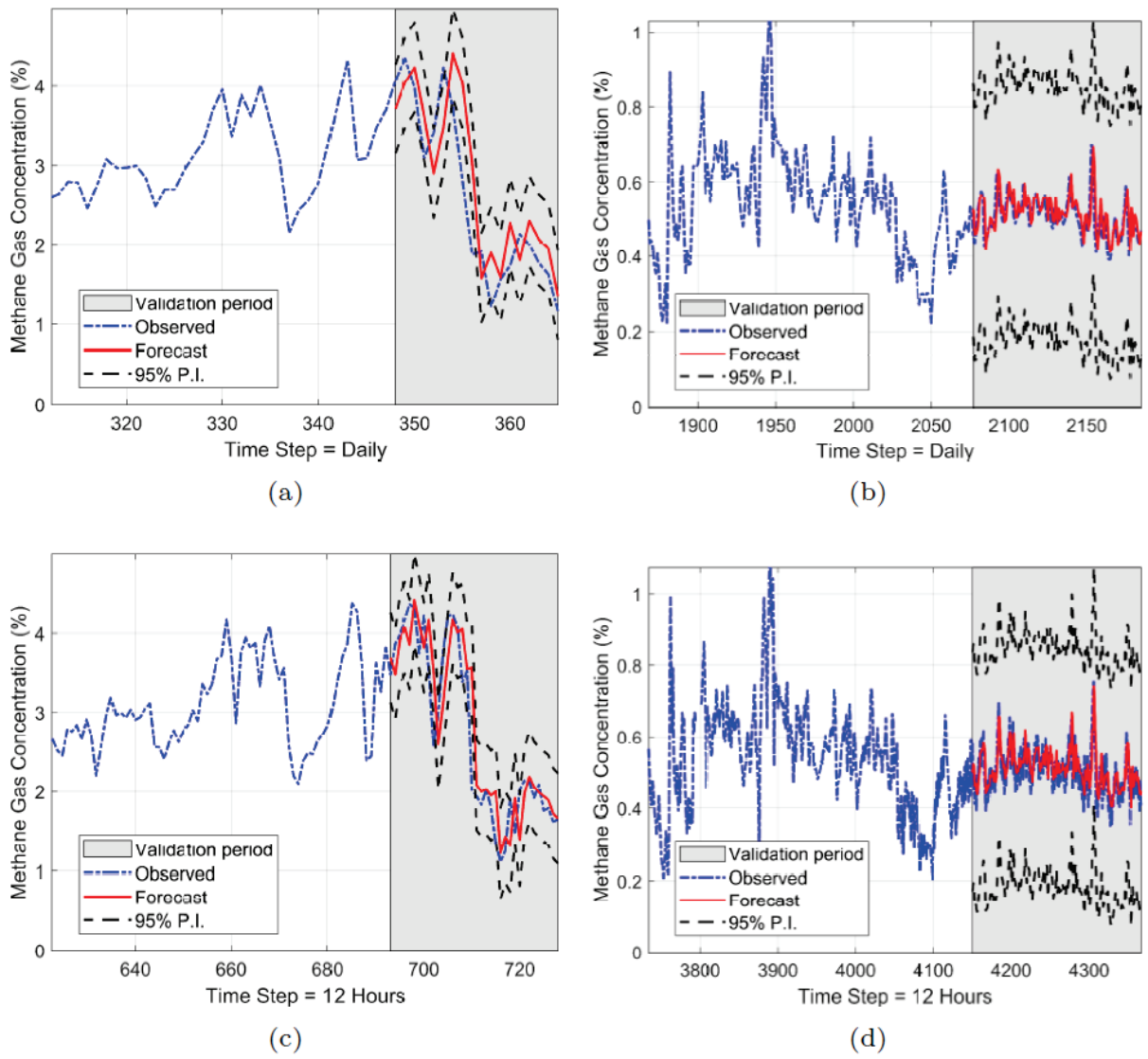


Figure 13: ARIMAX one-step-ahead CH4 concentration forecasts: (a) Forecast for dataset 1 using daily average values; (b) Forecast for dataset 2 using daily average values; (c) Forecast for dataset 3 using 12-hour average values; (d) Forecast for dataset 4 using 12-hour average values.

Table 6: Summary of CH4 concentration validation measures for the ARIMAX one-step-ahead model (using barometric pressure time series as an independent variable). p∗: Optimal autoregressive order of ARIMAX model based on AIC.

	Time Step	Training Data		Validation Data		p^*	R	RMSE	ME	MAE
		Length (days)	Sample Size	Length (days)	Sample Size					
Dataset 1	24-hour	365	365	18	18	28	0.87	0.58	−0.20	0.49
Dataset 2	24-hour	2,200	2,200	109	109	23	0.65	0.05	−0.01	0.03
Dataset 3	12-hour	365	730	18	37	8	0.91	0.46	−0.07	0.33
Dataset 4	12-hour	2,200	4,380	109	219	29	0.68	0.05	−0.01	0.04

4.4.7 Discussion

The multivariate time series models, VAR(p) and ARIMAX(*p,d,q*), are used to forecast methane gas concentrations. The results obtained from these multivariate approaches are compared in Table 3 with those obtained from the univariate ARIMA(*p,d,q*) model, as discussed in Diaz et al, (2022b). All forecasting methods use the same training and validation data, and their predictive performance is assessed using the same cross-validation measures.

All three forecasting models can reliably predict methane gas concentrations (details are given in Table 3). In most cases, the CH_4 concentration forecasts by the VAR(p), ARIMAX(*p,d,q*), and ARIMA(*p,d,q*) models match the direction (increasing/decreasing trend) of the validation data. Moreover, the observed CH_4 levels are captured by the 95% prediction intervals (this is illustrated in Figure 9 to Figure 11). In addition, the linear correlation between the forecasts and the validation data is strong and positive, indicating that both the forecasts and the validation data tend to increase and decrease in harmony. The other cross-validation statistics that measure the distance between the validation data and the forecasts take in general small values which are similar between different methods (as evidenced in Table 3). In general terms, the performance of the three methods is deemed as quite satisfactory given the quality of the data. All three methods can be used to provide reliable statistical forecasts of methane concentration given that the prediction intervals capture the true values in the validation tests.

A number of observations pertaining to the performance of the time series forecasting methods are discussed below.

- Datasets 7 and 8 also represent daily and 12-hour averages of the same records. All three models yield superior results for dataset 8 (i.e., for the 12-hour averages) in terms of the CV measures (Table 3).
- Similar behavior (i.e., better performance of the models with the 12-hour averaged data) is evident for all datasets presented herein, except for dataset 9. Daily averages tend to reduce fluctuations more than 12-hour averages. Hence, the models are fitted to smoother time series in this case. On the other hand, the one-step-ahead forecast for 12-hour averaging is closer in time to the last training point than the daily average. Thus, the dependence of the predicted value on the past is stronger than in the case of daily averages.
- Ideally, data averages over shorter (than 12-hour) time windows are expected to lead to more reliable predictions. This is supported by tests that were run on partial data segments that allowed calculating 6-hour averages. Training respective time series models, however, requires data streams of higher quality that contain fewer data gaps.

In some cases, the performance of the forecasting models is beset by irregularities in the data such as abrupt changes in the CH_4 time series. As a result, the correlation between the forecasts and the validation data is weak.

- The CV analysis of dataset 6 reveals weak correlation between forecasts and validation data; the values of R achieved by ARIMA, VAR, and ARIMAX were 0.33, 0.35, and 0.34, respectively, Table 3. This can be attributed to abrupt changes in the average CH_4 concentration, most likely due to sensor failure or/and calibration, or the potential influence of an unaccounted auxiliary variable(s) (e.g., coal production rate) which affects methane gas emissions. Figure 8 shows the CH_4 concentration (continuous lines, red line) and barometric pressure (dashed line, green line) for dataset 6. Visual inspection demonstrates that indeed the CH_4 time series changes abruptly with the concentration suddenly dropping from 1.4% to 0.60%. Furthermore, for this particular dataset the correlation between CH_4 and the barometric pressure time series is extremely low.
- Dataset 5 is based on the same data as dataset 6 but uses daily average values instead of 12-hour average values. The univariate ARIMA and the multivariate ARIMAX models show better performance than in dataset 6. A higher CV correlation is achieved, i.e., R = 0.54 (ARIMA) and R = 0.53 (ARIMAX). On the other hand, the performance of the VAR(p) model did not improve; the CV correlation is slightly lower (R = 0.34) than for dataset 6, as shown in Table 3. These results are interpreted as follows: Abrupt changes in CH_4 concentration are reduced by taking daily averages, leading to an overall better performance for dataset 5. However, the daily averaging does not eliminate "short-term variations" (fluctuations with correlations that are cut off after a few steps) which are better captured by the moving average terms in ARIMA and ARIMAX models.

The above results are attributed to the fact that abrupt changes in CH_4 concentration are reduced by taking daily averages, leading to an overall better performance for dataset 5 than the 12-hour-average of dataset 6. However, the daily averaging does not eliminate "short-term variations" (fluctuations with correlations that are cut off after a few steps) which are better captured by the moving average terms in ARIMA and ARIMAX models but are absent in the VAR(*p*) model.

A flow diagram that illustrates such a methodology is presented in Figure 12. (i) The ARIMA(*p,d,q*), VAR(p) and ARIMAX(*p,d,q*) models are applied to the target dataset. (ii) The optimal parametrization for each model is determined using the lowest AIC value. (iii) The best model among the ARIMA, VAR, and ARIMAX is selected based on a specified cross-validation measure (e.g., R or RMSE). (iv) The optimal model is used to forecast the CH_4 concentrations. (v) If the dataset is updated with new records for CH_4 concentration and barometric pressure, the algorithm is rerun to determine the optimal model. This last step implies continuous model updating in light of incoming data.

This flowchart can be used as a guideline for future studies, which will aim to further explore the use of time series models in the presence of higher-frequency, more complete records, and better quality datasets that will also incorporate consistently recorded estimates of production activity.

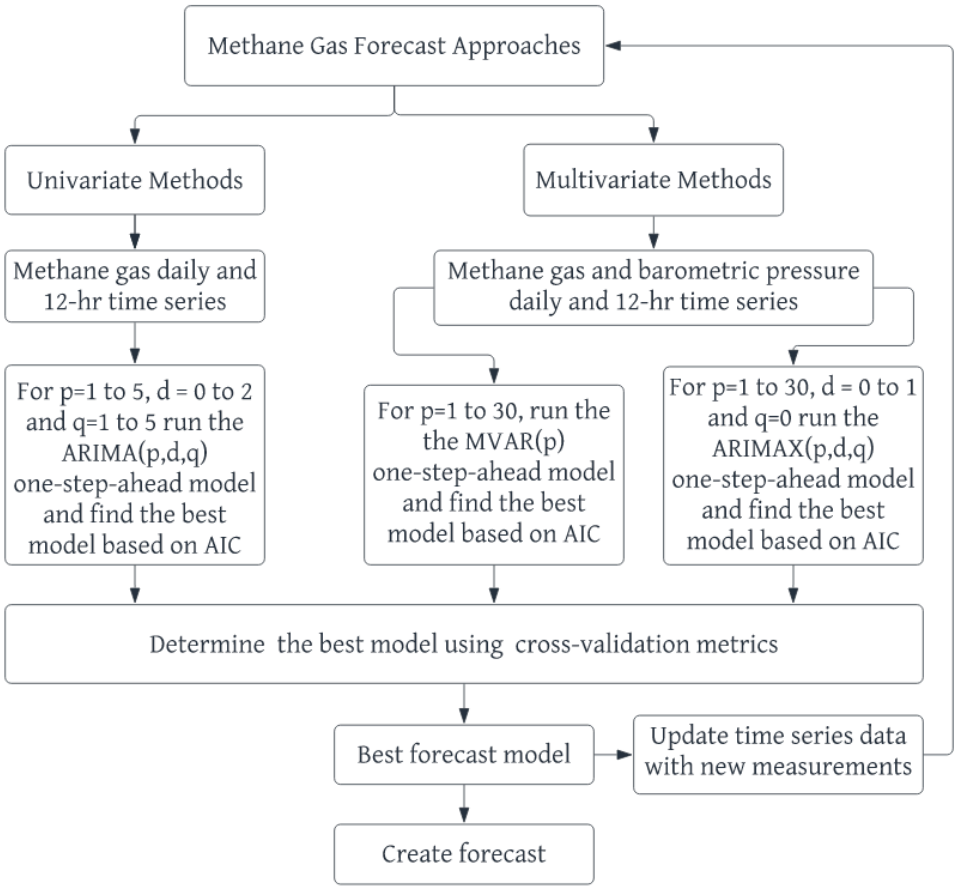


Figure 14: Flowchart of the methodology developed for selecting the optimal (univariate or multivariate) time series forecast model.

4.4.8 Discussion in the context of previous studies

Results for CH_4 forecasting are discussed below in the context of previous studies. Wang (2020) used several time series methodologies to forecast CH_4 based on sensor data from an underground coal mine. The RMSE obtained by means of ARIMA and VAR models were 5.4×10^{-3} and 4.5×10^{-3}, respectively. It is not possible to directly compare our results with those obtained by Wang (2020). First, statistics of the gas concentration series are not shown nor are graphs of the time series provided. Second, the partitioning of the dataset into training and validation sets is not described. Third, the (p d,q) orders of the optimal ARIMA models and the selection method used are not presented. Fourth, the dataset used in Wang (2020) contains significantly more information, since it involves measurements for three different gases from 15 monitoring sensors with a sampling step of six seconds for a total of about six million time points. Finally, the forecasting horizon (seconds, minutes, hours or days) over which the validation measures are evaluated is not specified.

A different study (Karacan, 2008) proposed a predictive approach based on Artificial Neural Networks for longwall mines. It was found that the linear correlation between the forecasts and

the validation data was around R = 0.93 for all datasets. The RMSE and R found in these investigations are similar to those obtained herein from six-year-long datasets using 12-hour averages (see Table 3).

In order to establish reliable and reliable forecasting methods for CH_4 concentration in mines several steps need to be taken by the research community. (i) Good quality gas CH_4 concentrations data are essential, as is reducing the frequency of erroneous records due to sensor malfunctions and recording gaps. (ii) The correlations between CH_4 concentration, atmospheric pressure, and coal production rate, including potential confounding factors, need to be better understood. (iii) With respect to modeling efforts, data-driven methods (whether based on statistical time series analysis or machine learning tools) have an advantage over methods based on computational fluid dynamics, since the latter demand significant computational resources and information (e.g., values of diffusion coefficients, initial and boundary conditions) which is at best only partially known. More research is needed to establish the scope, accuracy, and reliability of data-driven forecasting methods.

Given that different data-driven 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 methods. Therefore, it is proposed that the following critical elements of the data analysis be 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.
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:

a) The reporting should include the number, type (e.g., concentration, atmospheric pressure, etc.), and units of the time series used in the forecasting model.
b) The length of the time series and the sampling step (e.g., hour, day) should be specified.
c) Any pre-processing steps used to filter, smoothen, or coarse-grain (downsample) the data, or to remove outliers should be described.
d) Graphs of the time series studied should be presented, as they can provide valuable visual aids for the reader.
e) The results of exploratory statistical analysis should be listed, including the mean, median, standard deviation, skewness, and kurtosis coefficients of the data.
f) An analysis of the probability distribution(s) that the different time series follow should be included (using suitable probability plots if needed), and deviations from the normal distribution should be reported.
g) Two-point autocorrelations should be investigated by means of the autocorrelation function (ACF), the partial autocorrelation function (PACF), and the variogram function.

h) Trends, periodic behavior, and non-stationarities (if present) should be identified and discussed.
i) The visual inspection of time series graphs should 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 to allow reproducibility of the results. 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 orders used in the optimal model search should also be reported, 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 explanatory variables (e.g., atmospheric pressure and production rate). In addition, if a nonlinear transformation (e.g., Box-Cox) is used, 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. Hence, 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 used to avoid overfitting), the training method, the hyperparameters involved in the training process as well as their optimal values.

With regard to the third point above, 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. The validation should be performed with values that are not included in the training set. Statistical measures of forecasting performance such as the mean absolute error, the root mean square error, and the correlation coefficient between validation and forecast values should be reported. 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 CH_4 may vary between different mines or even different sections of the same mine. If data transformations are used, the performance measures should be reported in the original domain (e.g., if the logarithmic transform has been applied, the RMSE should be calculated and reported for the concentration, not for its logarithm). If the method allows for uncertainty estimation (e.g., ARIMA time series methods and Gaussian process regression), measures of uncertainty quantification should also be reported. One such measure involves the prediction intervals. In addition, proper scoring rules can be implemented for uncertainty quantification as described in (Gneiting and Raftery, 2007; Bessac and Naveau, 2021). Finally, it is helpful to supplement forecasting performance analysis with residual diagnostic testing to check if 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 have a long history, while the latter have gained momentum in the last decade. Even though a natural tendency is to prefer more modern approaches over older methods, it is prudent 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, 2014), such as autoregressive heteroskedastic (ARCH) models and their generalized (GARCH) versions, as well as regime-switching models such as Markov switching AR and Self-Exciting Threshold Autoregressive (SETAR) models. Based on the literature review, such models could be better suited for handling irregular (i.e., non-Gaussian, non-stationary) data but have not yet been applied to CH_4 concentration forecasting. Finally, the machine learning method of Gaussian process regression (Rasmussen and Williams, 2006; Agou et al, 2022) and geostatistical analysis (De Iaco et al, 2022), also provide flexible forecasting frameworks that deserve further investigation.

4.5 Development of a web subsystem and web interface optimized for real-time or near real-time atmospheric data visualization for mine personnel

A web subsystem was developed based on results presented above. The web subsystem features a database backend where all data are stored. The user interface allows data management as well as running the forecasting algorithm. User credentials need to be specified when accessing the application.

The web application utilizes methane emissions time-series data and barometric pressure data, however it does not utilize coal production data as coal production data do not typically consist of a continuous time series.

Different locations can be parametrically specified on the web app and data can be uploaded for each location as shown in Figure 13 and Figure 14. In this context, the term location refers to a mine location where data can be collected. A forecast can then be created on the collected data. Data are collected from two sources:

a) From the mine using a manual import (import of an excel file).
b) From the nearest weather station. The weather station can be specified using its initials and the application will draw data automatically from the station.

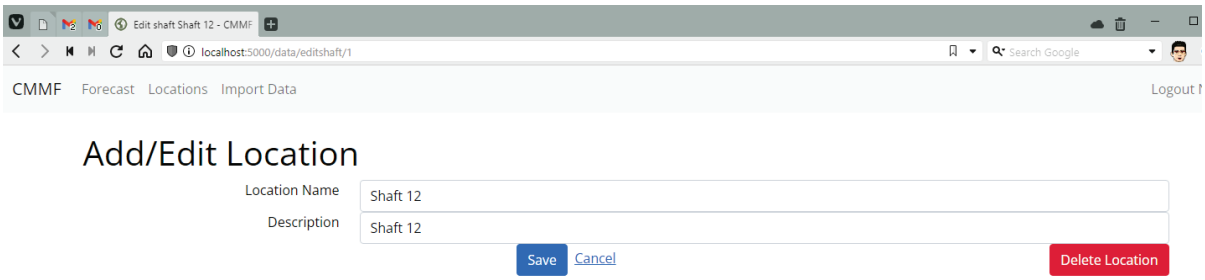


Figure 15: Creating a location for data processing.

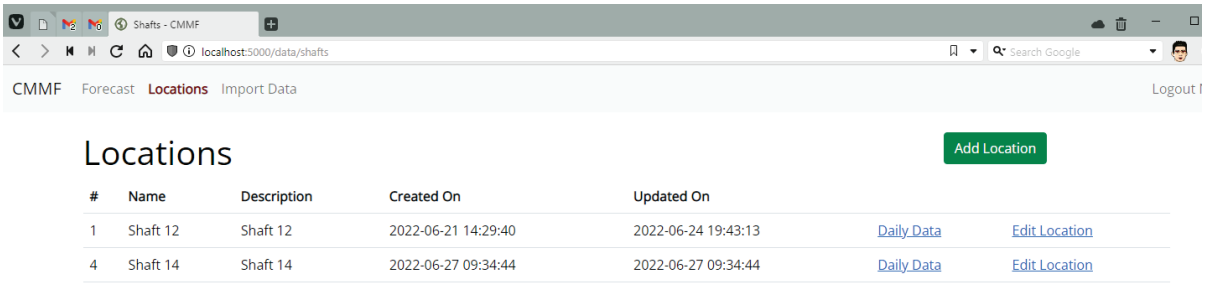


Figure 16: Summary of locations and data available.

Data can be imported to each location and also displayed as shown in Figure 15 and Figure 16. Data can be automatically retrieved from public weather stations as shown in Figure 17.

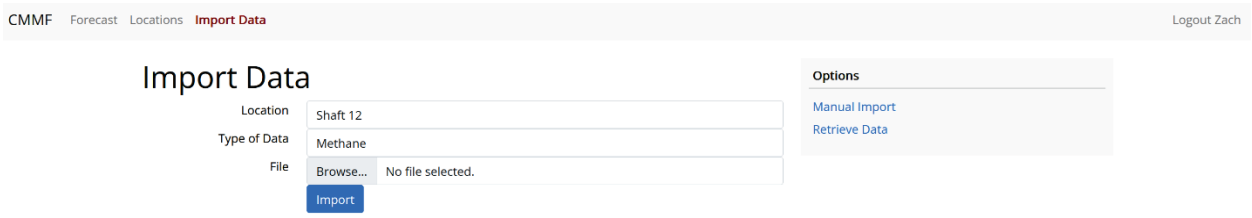
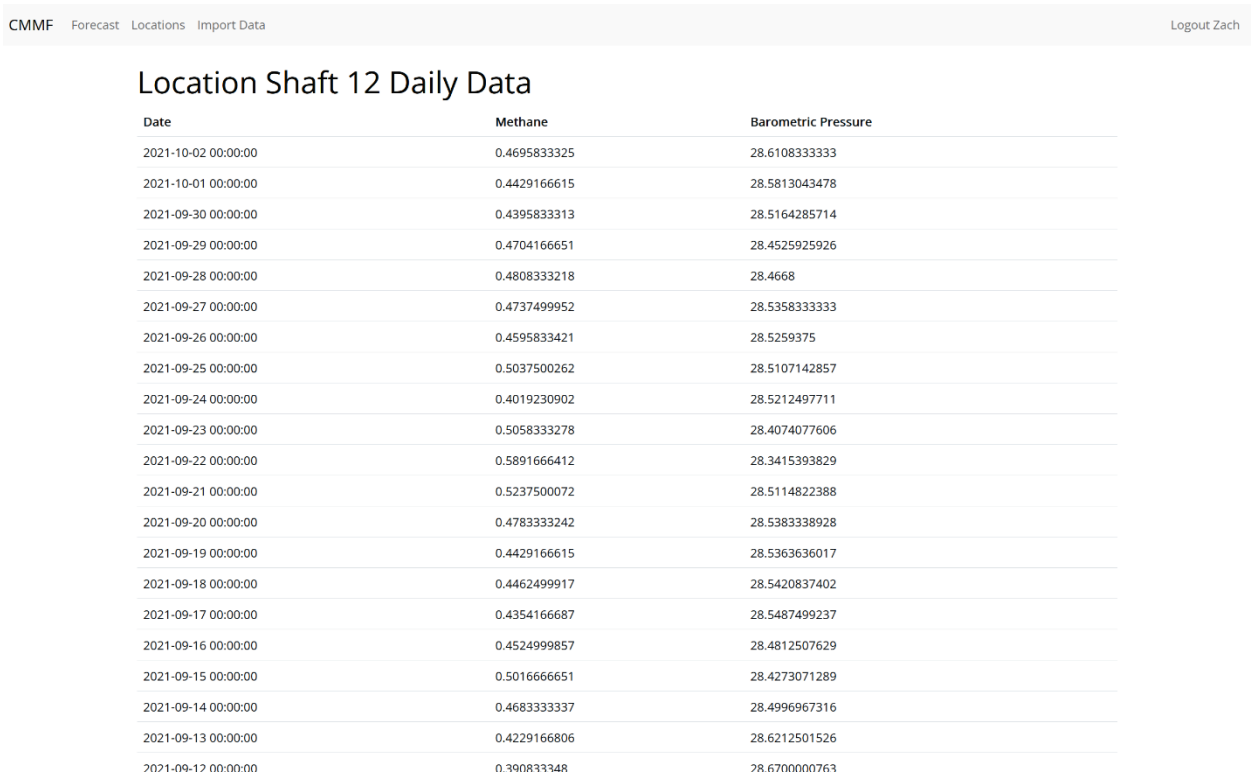
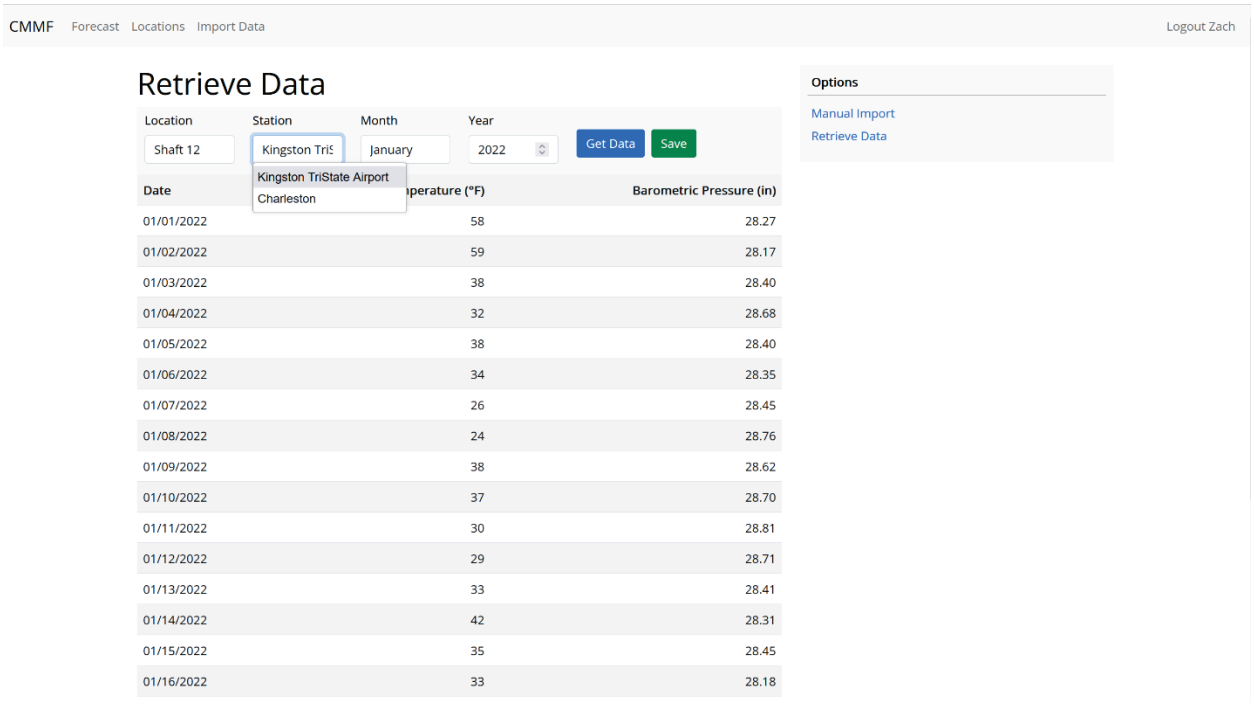


Figure 17: Selecting a data file to import for a specific location



Date	Methane	Barometric Pressure
2021-10-02 00:00:00	0.4695833325	28.6108333333
2021-10-01 00:00:00	0.4429166615	28.5813043478
2021-09-30 00:00:00	0.4395833313	28.5164285714
2021-09-29 00:00:00	0.4704166651	28.4525925926
2021-09-28 00:00:00	0.4808333218	28.4668
2021-09-27 00:00:00	0.4737499952	28.5358333333
2021-09-26 00:00:00	0.4595833421	28.5259375
2021-09-25 00:00:00	0.5037500262	28.5107142857
2021-09-24 00:00:00	0.4019230902	28.5212497711
2021-09-23 00:00:00	0.5058333278	28.4074077606
2021-09-22 00:00:00	0.5891666412	28.3415393829
2021-09-21 00:00:00	0.5237500072	28.5114822388
2021-09-20 00:00:00	0.4783333242	28.5383338928
2021-09-19 00:00:00	0.4429166615	28.5363636017
2021-09-18 00:00:00	0.4462499917	28.5420837402
2021-09-17 00:00:00	0.4354166687	28.5487499237
2021-09-16 00:00:00	0.4524999857	28.4812507629
2021-09-15 00:00:00	0.5016666651	28.4273071289
2021-09-14 00:00:00	0.4683333337	28.4996967316
2021-09-13 00:00:00	0.4229166806	28.6212501526
2021-09-12 00:00:00	0.390833348	28.6700000763

Figure 18: Listing of daily data for a specific location



Date	perature (°F)	Barometric Pressure (in)
01/01/2022	58	28.27
01/02/2022	59	28.17
01/03/2022	38	28.40
01/04/2022	32	28.68
01/05/2022	38	28.40
01/06/2022	34	28.35
01/07/2022	26	28.45
01/08/2022	24	28.76
01/09/2022	38	28.62
01/10/2022	37	28.70
01/11/2022	30	28.81
01/12/2022	29	28.71
01/13/2022	33	28.41
01/14/2022	42	28.31
01/15/2022	35	28.45
01/16/2022	33	28.18

Figure 19: Data retrieved from nearby weather station through the Weather Underground API

In addition, the application is programmed to collect data directly from Weather Underground weather stations through the API (Application Program Interface) provided by that platform.

The user can then generate a forecast model either as a univariate time series (methane gas only) or as a multivariate time series (methane gas and barometric pressure). The user can also specify the date range for the forecast model and also the percentage of the time series that will be used for training and validation. Figure 18 presents the user interface for running the forecast model with two variables, i.e., as a multivariate analysis.

The forecasted period is shown in red at the end of the time series, while the observed values are shown in black. The values used for model calibration are shown in gray. The user can also enable or disable visualization of each time series by clicking on (selecting or unselecting) the respective legend icons (color boxes) at the bottom of the chart. The user can zoom into a specific area of the chart using the mouse. The view can be reset by clicking on the blue "Reset Zoom" button. Figure 19 presents a detailed view of the forecast area. The gray shaded region corresponds to the 95% confidence interval.

Figure 20 presents the user interface for running the forecast model with only one variable (methane gas). Figure 21 presents a detailed view of the forecast area. Figure 22 to Figure 25 present information similar to that in Figure 18 to Figure 21 which corresponds to a different time period.

When the user clicks on the refresh button the program will run the optimization algorithm presented in Figure 12. This typically takes a few seconds – the larger the dataset the longer it takes. It should also be noted that the univariate analysis is faster than the multivariate analysis. The application will display the best model based on the input parameters. Model information is shown at the bottom of the screen below the chart (Figure 18).

It should also be noted that the p-value shown in Figure 18 is not wrong. The p-value is really small and shows as 0 with 4 decimal places displayed (i.e., as "0.0000"). The user can re-run the model with a different set of input parameters and/or a different location and get a fresh forecast. Forecast values may change pretty much the same way as a weather forecast may change with time.

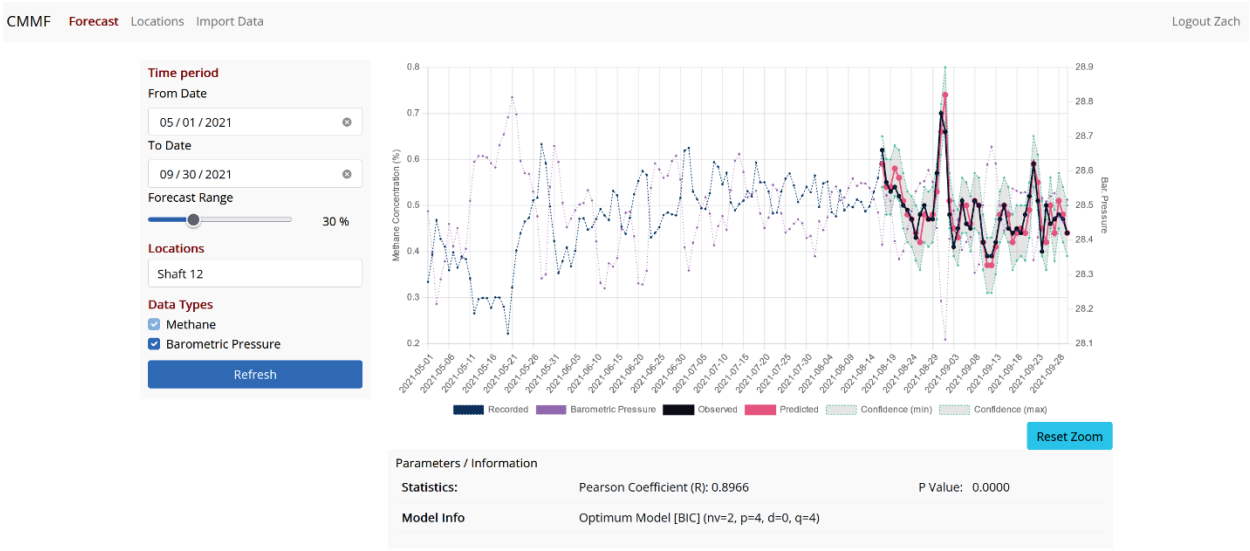


Figure 20: Forecast for time series 1 based on both CH4 and BP values

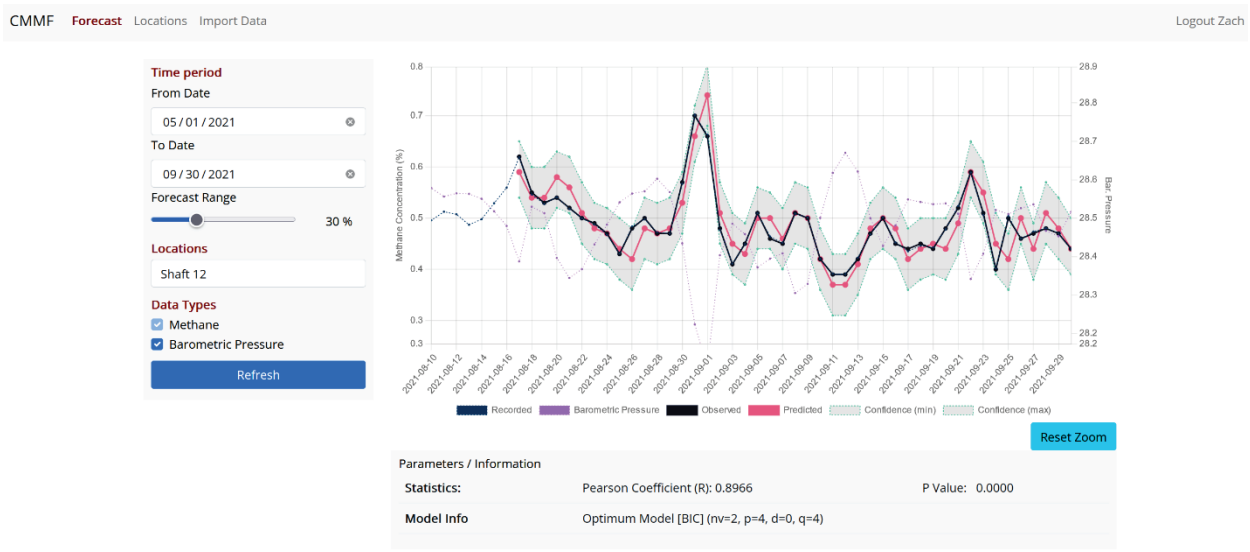


Figure 21: Forecast for time series 1 based on both CH4 and BP values – Detail view

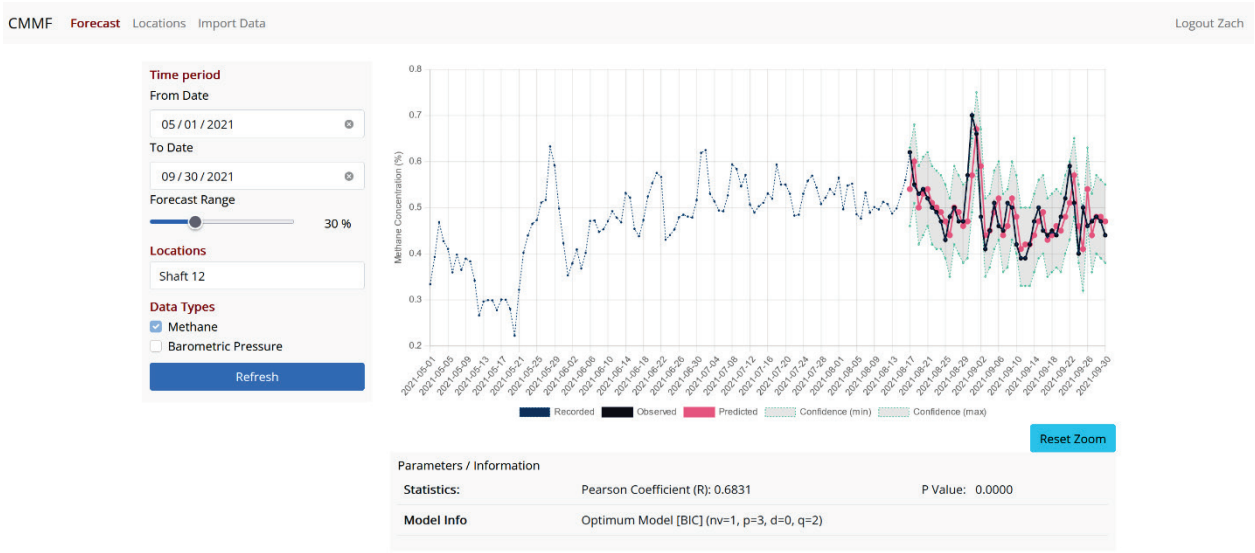


Figure 22: Forecast for time series 1 based on CH4

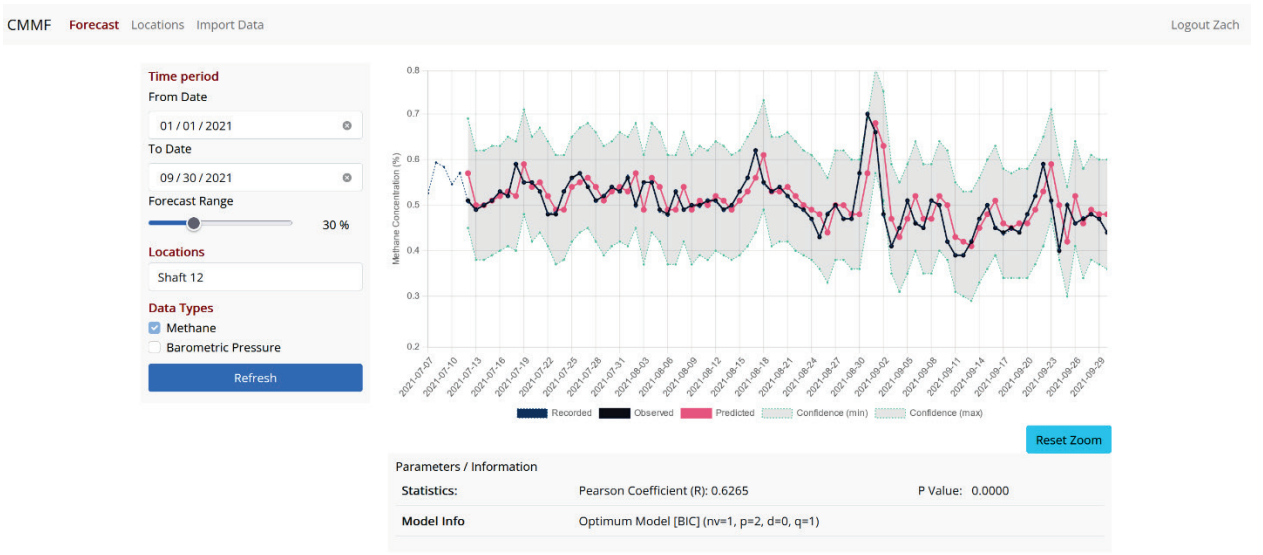


Figure 23: Forecast for time series 1 based on CH4 – Detailed view

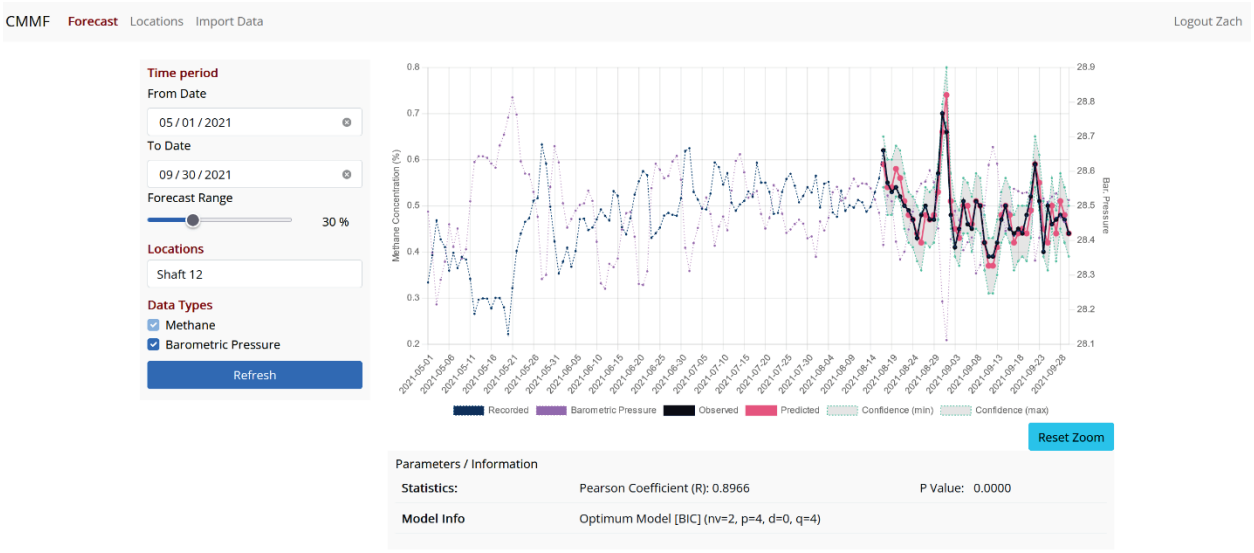


Figure 24: Forecast for time series 2 based on CH4 and BP

Figure 25: Forecast for time series 2 based on CH4 and BP – Detailed view

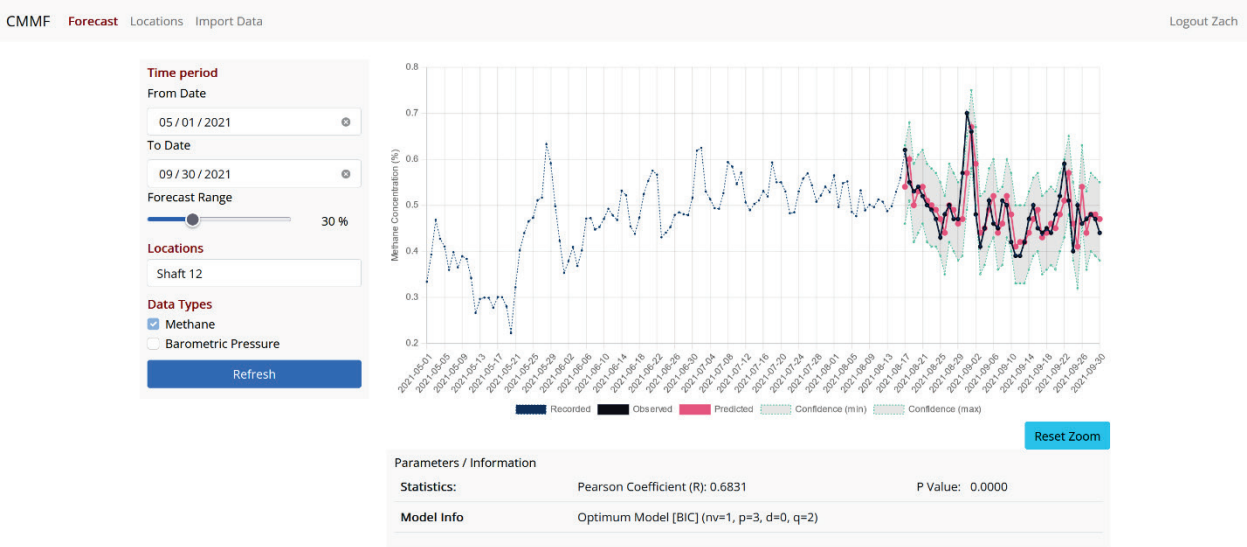


Figure 26: Forecast for time series 2 based on CH4

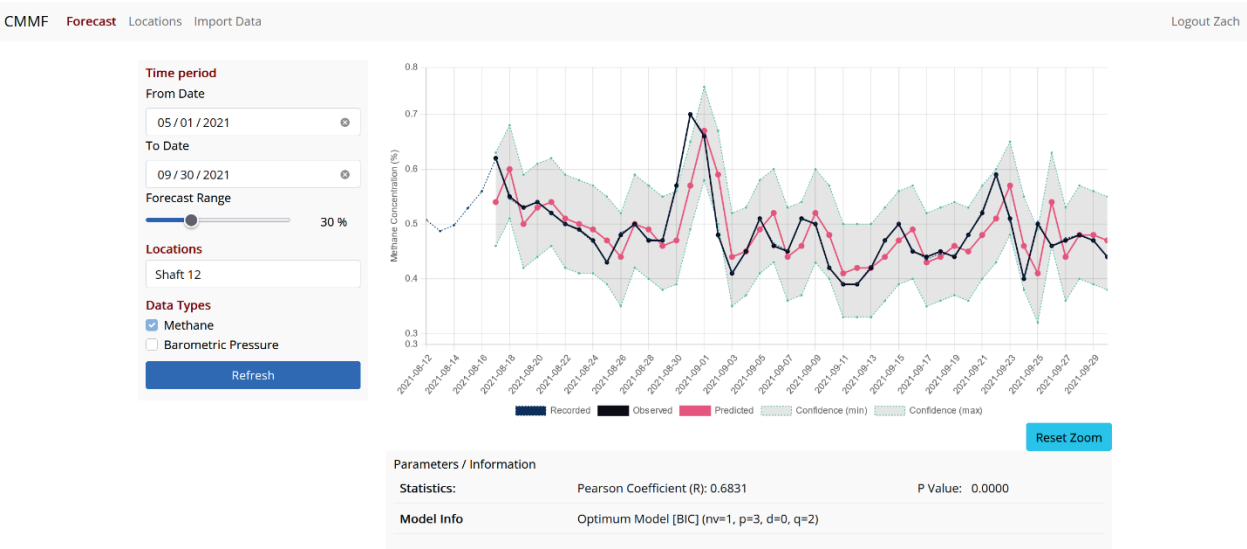


Figure 27: Forecast for time series 2 based on CH4 – detailed view

A help button provides basic information on data needed and how to run the application (Figure 28).

In conclusion, this application can easily be installed and run on a webserver and provide real-time forecasts on methane emissions. Data input consists of real-time local methane emissions measurements and regional real-time barometric pressure data which are typically provided free from public weather stations. The web application can automatically read data from

selected public weather stations, and, therefore, the end-user will only need to provide a direct feed of methane measurements at a particular location, e.g., at an exhaust shaft.

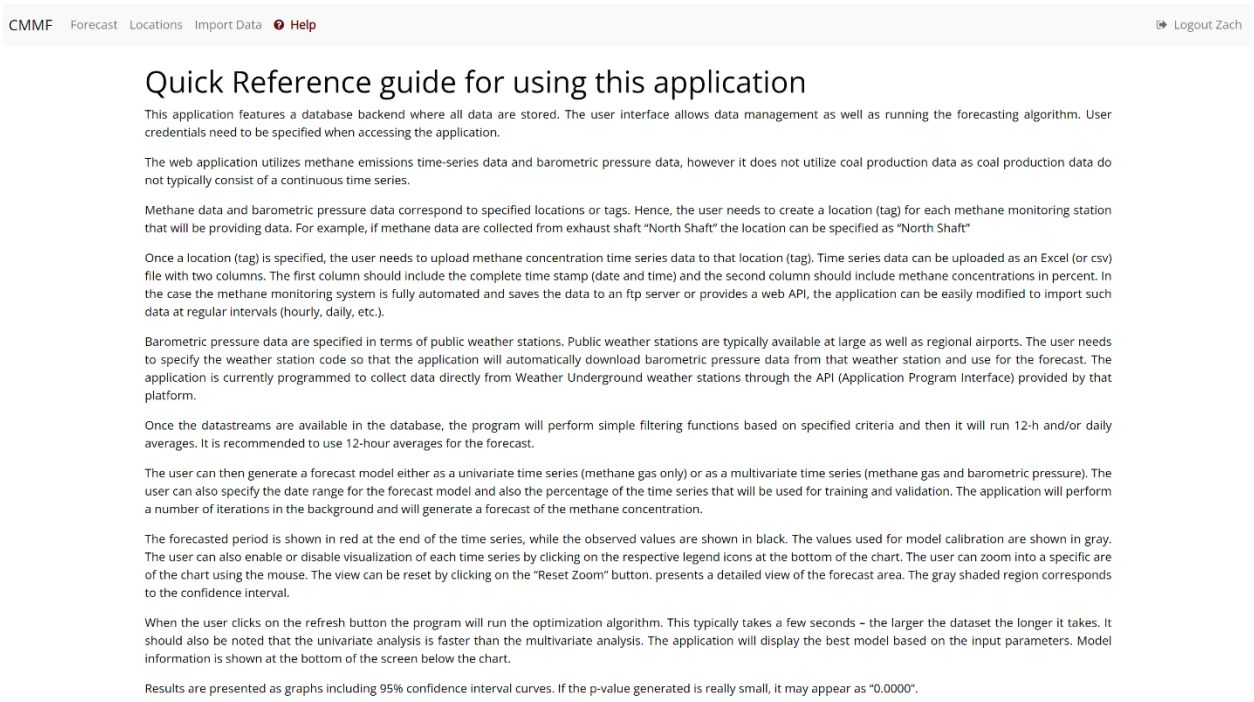


Figure 28: Help screen of the methane forecast application.

5 Publication Record and Dissemination Efforts

This section presents a complete record of publications and presentations which were generated from this project.

5.1 Technical Papers

1. Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Towards atmospheric monitoring data analysis in underground coal mines. Proceedings, 18th North American Mine Ventilation Symposium (NAMVS), June 12-16, 2021, https://doi.org/10.1201/9781003188476-51
2. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Managing and utilizing big data in atmospheric monitoring systems for underground coal mines, Proceedings, International Conference on Raw Materials and Circular Economy (RawMat), September 2021, Athens, Greece, Mater. Proc. https://doi.org/10.3390/materproc2021005078
3. 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
4. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2022) Forecasting of methane gas concentrations in underground coal mines, Proceedings, International Conference on Green Intelligent Mining for Thick Coal Seam - 40th Anniversary of Longwall Top-Coal Caving Mining in China, July 22-24, 2022. Beijing, China
5. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT, Luxbacher K (2022) Time series modeling of methane gas in underground mines. Mining, Metallurgy & Exploration Journal. https://doi.org/10.1007/s42461-022-00654-5
6. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2022) Methane Gas Concentration Prediction in Underground Coal Mines – As a Sustainable Mining Approach Proceedings, 10th International Conference of Sustainable Development in the Minerals Industry (SDIMI), September 15 – 17, 2022. Windhoek, Namibia (*extended abstract*)
7. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2023) Underground coal methane gas forecasting using multivariate time series with one and two auxiliary variables, Proceedings, 19th North American Mine Ventilation Symposium, Rapid City, SD, June 2023 (*under review*)
8. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT, Luxbacher K (2022) Forecasting of Methane Gas in Underground Coal Mines: Univariate versus Multivariate Time Series Modeling. Stochastic Environmental Research and Risk Assessment Journal, https://doi.org/10.1007/s00477-023-02382-8
9. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT, Luxbacher K (2022) Time series modeling of methane gas in underground mines. *Mining Engineering, Feb 2023*

5.2 Technical Presentations (in Person and Virtual) and Poster Presentations

1. Diaz J, (2021) *Mineria desde Casa*, a series of Spanish conferences about different topics of Mining Engineering organized by La Universidad Nacional de Colombia, April 15, 2021 *(virtual presentation)*.
2. Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Towards atmospheric monitoring data analysis in underground coal mines. 18th North American Mine Ventilation Symposium (NAMVS), June 12-16, 2021, https://doi.org/10.1201/9781003188476-51 *(virtual presentation)*
3. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2021) Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. International Conference on Raw Materials and Circular Economy (RawMat), September 2021, Athens, Greece, Mater. Proc. https://doi.org/10.3390/materproc2021005078 *(virtual presentation)*
4. 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 (*in person presentation*)
5. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2022) Forecasting of methane gas concentrations in underground coal mines. International Conference on Green Intelligent Mining for Thick Coal Seam - 40th Anniversary of Longwall Top-Coal Caving Mining in China, July 22-24, 2022. Beijing, China *(virtual presentation)*
6. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2022) Methane Gas Concentration Prediction in Underground Coal Mines – As a Sustainable Mining Approach. International Conference of Sustainable Development in the Minerals Industry (SDIMI), September 15 – 17, 2022. Windhoek, Namibia (*in person presentation*)
7. Diaz JC, Agioutantis Z, Hristopulos DT, Schafrik S, Implementing Time Series Analysis for Methane Gas Forecasting in Underground Coal Mines, SME Annual Meeting, Feb 27 – March 2, 2022, Salt Lake City, UT (*poster presentation in the poster competition*)
8. Diaz, J., D.T. Hristopulos, Z. Agioutantis and S. Schafrik, Time series forecasting of methane gas in underground coal mines, SME Annual Meeting, Feb 26 – March 1, 2023, Denver, CO (*accepted abstract, planned in person presentation*)
9. Diaz JC, Agioutantis Z, Schafrik S, Hristopulos DT (2023) Underground coal methane gas forecasting using multivariate time series with one and two auxiliary variables, Proceedings, 19th North American Mine Ventilation Symposium, Rapid City, SD, June 2023 (*planned in person presentation*)

5.3 Dissertations

Diaz J. (2022), Development of Univariate and Multivariate Forecasting Models for Methane Gas Emissions In Underground Coal Mines, PhD Dissertation, University of Kentucky

6 Conclusions and Impact Assessment

There are a number of forecasting models that assume data that is constantly monitored is simply time-dependent, which use statistical methodologies to forecast the next set of values in the time series. These approaches are not time-consuming and do not need much training data because they only account for the past and next values of the time series. The research presented here used a more complicated and reliable methodology, considering the input factors to the time-series data under study. This was done with several years of mine data and weather data that have been under continuous collection for several years.

Data inconsistencies (e.g., missing data, erroneous values, and abrupt changes) in the methane gas time series collected were filtered out after pre-processing. In addition, it was determined that using daily or 12-hour averages was more representative of trends than using hourly averages. This has a direct impact to the outcomes of statistical tests (e.g., variogram function, Pearson linear correlation, and scatterplots) and, consequently, the reliability and performance of the different forecasting approaches.

Three main associations were identified between the time series (e.g., methane gas, barometric pressure, and coal production rate):

- There is autocorrelation in the methane gas time series. Autocorrelation becomes better for larger datasets.
- 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.
- 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. However, production data are not always consistent. Production may stop over weekends and holidays and producing faces may not be easily correlated to methane gas monitoring stations. Nonetheless, since production has an impact on methane concentration, the impact of production is indirectly included to some extent in the auto-correlation of the methane concentration. Using either the auto-correlation or cross-correlation with barometric pressure good forecasts of methane concentration can be obtained. These can be further improved if consistent data on production become available.

The autocorrelation[1] in the methane gas time series is the reason that simplified statistical models that use the methane gas series alone perform in a reliable manner over different time windows. A univariate forecasting model, the ARIMA(*p,d,q*) one-step-ahead model, was

[1] Autocorrelation means that each value of the time series (e.g., methane concentration) is related to the values of the same series at previous time instants. In other words, the series has a memory of its past values. This memory property is first quantified during the method estimation stage and then exploited to derive forecasts of methane concentration based on past values. Barometric pressure and production also have an impact on methane concentration. This is expressed by the cross-correlation function between the independent variables and methane gas concentration.

developed based on the autocorrelation of the methane time series. The forecasting reliability 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 reliably when the time training data is sufficiently cleaned. 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.

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.

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 developed by averaging data every 12 hours contained more information (number of records) than the ones with daily average values.

The performance and reliability 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.

The univariate and multivariate methane gas forecasting models proposed in this project 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.

Finally, an algorithm capable of assessing the results of both multivariate and univariate models and selecting the best model among them for a given dataset was developed. See Figure 12 in Chapter 4, for the algorithm description. The algorithm can reliably predict methane gas concentration for the upcoming days. As the new data are populated in the database, the prediction becomes better (as is the case with weather prediction systems).

A web-based application that can process real time data was developed. At this time, the web application only utilizes methane emissions time-series data and barometric pressure data, and it does not utilize coal production data. The application can easily be installed and run on a webserver and provide real-time forecasts on methane emissions. Data input consists of real-time local methane emissions measurements and regional real-time barometric pressure data which are typically provided free from public weather stations. The web application can automatically read data from selected public weather stations, and, therefore, the end-user will only need to provide a direct feed of methane measurements at a particular location, e.g., at an exhaust shaft.

7 Recommendations for Future Work

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 can also perform data pre-processing and develop weekly, daily, 12-hour and 6-hour averages. 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 fully automate the pre-processing data stage to make it more efficient.

Early in the data collection process, it was identified that methane sensor measurement that measured methane through the same airstream had differences. Also, methane measurements drifted with time and had to be corrected after calibration. The reliability and performance of methane sensors under actual operating conditions needs to be examined in detail.

The statistical data analysis and the development of the different forecast models were carried out by implementing the MATLAB programming environment, and where subsequently programmed in Python for the web interface. Results between MATLAB and Python routines are slightly different as the underlying routines may use different processes for time series forecasting. Performance and results of black-box routines need to be compared to ensure that they perform similarly under all conditions.

This research investigated the univariate and multivariate forecasting approaches to develop three models (ARIMA, VAR(p), and ARIMAX) capable of reliably predicting methane gas concentrations in underground coal mines based on time series data. It is recommended to explore utilization of more sophisticated and complex forecasting methodologies such as Artificial Neural Networks, Complex Seasonality, Prophet model, and bootstrapping to compare their reliability and performance with the forecast models proposed.

Two multivariate forecasting approaches (VAR(p) and ARIMAX) were developed to forecast concentrations of methane gas in underground coal mines based on the negative cross-correlation between methane gas and one (barometric pressure) independent variable. Production was also used as a second independent variable when available. As mine production does not always occur every calendar day, production time series have gaps. It is recommended that a model is developed that can negotiate production gaps and utilize daily production as a second independent variable.

Continue the collection of atmospheric data to build a better database of methane concentration data and respective production data, where available.

8 References

1. Agioutantis Z., K. Luxbacher, H. Dougherty and M. Karmis, (2015a), Development and Implementation of a Data Management Platform for Atmospheric Data in Underground Coal Mines, Proceedings, 15th North American Ventilation Symposium, June 20-24, 2015, Blacksburg VA, pp. 135-140.
2. Agioutantis, Z. and S. Papaterpos, (2015b), A real-time event driven data management application for equipment monitoring in continuous surface mining operations, Proceedings, AIMS Conference, May 27-28, 2015, Aachen, Germany, pp. 19-30.
3. Agioutantis, Z., K. Luxbacher, M. Karmis, S. Schafrik, (2014), Development of an atmospheric data management system for underground coal mines, The Journal of The Southern African Institute of Mining and Metallurgy, Vol. 114, (December 2014), pp. 1059-1063.
4. Agou VD, Pavlides A, Hristopulos DT (2022) Spatial modeling of precipitation based on data-driven warping of Gaussian processes. Entropy 24(3):321. https://doi.org/10.3390/e24030321
5. Airey, E. M. (1968). Gas emission from broken coal. An experimental and theoretical investigation. International Journal of Rock Mechanics and Mining Sciences and Geomechanics, 5, 475–494.
6. AMR PEMCO (2002) Multi-Gas Monitor. https://www.amrpemco.com/mc-6410-multi-gas-monitor/, Accessed December 29, 2022.
7. Andrews BH, Dean MD, Swain R, et al (2013) The National Institute for Occupational Safety and Health. Data and Statistics. Tech. rep., Society of Actuaries, URL https://www.soa.org/globalassets/assets/files/research/projects/research-2013-arima-arimax-ben-appl-rates.pdf
8. Bessac J, Naveau P (2021) Forecast score distributions with imperfect observations. Advances in Statistical Climatology, Meteorology and Oceanography 7(2):53–71. https://doi.org/10.5194/ascmo-7-53-2021
9. Booth, P., Brown, H., Nemcik, J., and Ting, R. (2017). The spatial context in the calculation of gas emissions for underground coal mines. International Journal of Mining Science and Technology, 27(5), 787–794.
10. Booth, P., Nemcik, J., and Ren, T. (2016). A Critical Review and New Approach for Determination of Transient Gas Emission Behaviour in Underground Coal Mines. Proceedings, 16th Coal Operators' Conference, Mining Engineering, University of Wollongong.
11. Borowski, M., Szlazak, N., and Obracaj, D. (2009). Methane Hazard Predictions in Underground Coal Mining. Deep Mining Challenges, International Mining Forum 2009, 18–21 February, Krakow, Poland, 9–22.
12. Brockwell, P. J., and Davis, R. A. (2016). Introduction to Time Series and Forecasting, 3rd Ed, Springer Texts in Statistics, Springer.
13. Boyer, C. M., Qingzhao, B. (1998). Methodology of coalbed methane resource assessment. International Journal of Coal Geology, 35: 349–368
14. Box GE, Jenkins GM, Reinsel GC, et al (2015) Time Series Analysis: Forecasting and Control, 5th edn. John Wiley & Sons, Hoboken, NJ, USA

15. Brillinger, D.R., (2001). Time Series: Data analysis and theory. Society for Industrial and Applied Mathematics.
16. ByungWan J, Rana A (2018). An internet of things system for underground mine air quality pollutant prediction based on azure machine learning. Sensors. https://doi:10.3390/s18040930.
17. Bustin, R.M., Clarkson, C.R. (1998). Geological controls on coalbed methane reservoir capacity and gas content. International Journal of Coal Geology, 38: 3–26.
18. Chakrabarti A, Ghosh, J.K. (2011). AIC, BIC and Recent Advances in Model Selection, In Editor(s): Prasanta S. Bandyopadhyay, Malcolm R. Forster, Handbook of the Philosophy of Science, Philosophy of Statistics, North-Holland, Volume 7, 2011, pp. 583-605, ISSN 18789846, https://doi.org/10.1016/B978-0-444-51862-0.50018-6.
19. Creedy, D.P. (1993). Methane emissions from coal related sources in Britain: development of a methodology. Chemosphere, 26: 419–439
20. Curl, S.J. (1978). Methane prediction in coal mines. International Energy Agency Coal Research.
21. De Iaco S, Hristopulos DT, Lin G (2022) Special issue: Geostatistics and machine learning. Mathematical Geosciences 54:459–465. https://doi.org/10.1007/s11004-022-09998-6
22. Diamond, W.P., 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., Proceedings of the 6th International Mine Ventilation Congress, Ramani (ed), May 17-22, 1997, Pittsburgh, Pennsylvania.
23. Diaz J, Agioutantis Z, Hristopulos DT, Schafrik S (2021a) Managing and utilizing big data in atmospheric monitoring systems for underground coal mines. Materials Proceedings 5(1):78. https://doi.org/10.3390/materproc2021005078.
24. Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2021b) Towards atmospheric monitoring data analysis in underground coal mines. Proceedings, 18th North American Mine Ventilation Symposium (NAMVS 2021), Tukkaraja P (ed), June 21-23. CRC Press, Boca Raton, FL, pp. 498–506, https://doi.org/10.1201/9781003188476-51.
25. Diaz J, Agioutantis Z, Schafrik S, Hristopulos DT (2022a) Forecasting of methane gas concentrations in underground coal mines. Proceedings, International Conference on Green Intelligent Mining for Thick Coal Seam - 40th Anniversary of Longwall Top-Coal Caving Mining in China, July 22-24, 2022. Beijing, China.
26. Diaz, J., Agioutantis Z, Hristopulos DT, Schafrik S, Luxbacher K, (2022b), Time Series Modeling of Methane Gas in Underground Mines, Mining, Metallurgy & Exploration, 2022, https://dx.doi.org/10.1007/s42461-022-00654-5
27. Dixon, D. (1992). A statistical analysis of monitored data for methane prediction. PhD Dissertation, University of Nottingham.
28. Dixon, D., and Longson, I. (1993). Statistical method for methane prediction and improved environmental control. Proceedings, 6th US Mine Ventilation Symposium, June 21-23, Salt Lake City, UT, pp. 94–98.
29. Dunmore, R. (1982). Predicting methane in real-time for longwall faces. Proceedings of the 1st North American Mine Ventilation Symposium. Alabama, 29 31 March 1982.
30. Ediz, I. G. & Edwards, J. S. 1991. Numerical simulation of time-dependent methane flow. Mining Science and Technology, 12: 1–15.

31. Enders W (2014) Applied Econometric Time Series, 4th edn. John Wiley & Sons, Hoboken, NJ, USA
32. Fauconnier, C.J. (1992), Fluctuations in barometric pressure as a contributory factor to gas explosions in South African mines. Journal of the South African Institute of Mining and Metallurgy, vol. 92, no. 5. pp. 131–147.
33. Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102(477):359–378. https://doi.org/10.1198/016214506000001437
34. 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.
35. Hemp R (1994) The effect of changes in barometric pressure on mines in the highveld of South Africa. Journal of the Southern African Institute of Mining and Metallurgy vol. 94, no. 6, pp. 133–146.
36. Hristopulos, D.T. (2020). Random Fields for Spatial Data Modeling. Springer, Dordrecht, the Netherlands.
37. Hyndman RJ, Athanasopoulos G (2021) Forecasting: principles and practice. https://otexts.com/fpp3/. Accessed December 29, 2022
38. Jensen, B., Gillies, A. D. S., Anderson, J. M., and Jones, N. (1992). Review of methane emission and prediction research in longwall coal mines. Proceedings, The Australian Institute of Mining and Metallurgy, 1, 11–17.
39. Johansen S (1995) The vector autoregressive model. In: Likelihood-Based inference in cointegrated vector autoregressive models. Oxford University Press, England.
40. Karacan, C.Ö., Esterhuizen, G. S., Schatzel, S. J., and Diamond, W. P. (2005). Numerical analysis of the impact of longwall panel width on methane emissions and performance of gob gas ventholes. International Journal of Coal Geology, 71(2–3), 225–245.
41. Karacan C.Ö. (2008) Modeling and prediction of ventilation methane emissions of U.S. longwall mines using supervised artificial neural networks. International Journal of Coal Geology, Volume 73, Issues 3–4, 1 February 2008, Pages 371-387 https://doi.org/10.1016/j.coal.2007.09.003
42. Karacan, C. Ö., 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. International Journal of Coal Geology, 86: 121–156
43. Kirchgässner G, Wolters J (2007) Introduction to Modern Time Series Analysis. Springer, Berlin, Germany
44. Kravchuk K (2017) Forecasting: ARIMAX model exercises (Part-5). https://www.r-bloggers.com/2017/05/forecasting-arimax-model-exercises-part-5/. Accessed December 29, 2022.
45. Krog, R (2006), Predicting methane emissions from longer longwall faces by analysis of emission contributors, Proceedings, 11th US/North American Mine Ventilation Symposium, May, https://doi.org/10.1201/9781439833391.ch54
46. Lloyd, P.J.D. and A. Cook, (2004), Methane release from South African coal mines, Journal of the South African Institute of Mining and Metallurgy vol. 105, no. 8. pp. 483–490.

47. Lolon SA (2017) Computational modeling of barometric pressure fluctuation effects on explosive methane-air mixtures in a longwall mine gob. PhD dissertation, Colorado School of Mines, Colorado, USA, URL https://mountainscholar.org/handle/11124/170983.
48. Lütkepohl, H (2005) New Introduction to Multiple Time Series Analysis. Springer Science & Business Media, Berlin, Germany.
49. Lunarzewski, L. W. 1998. Gas emission prediction and recovery in underground coal mines. International Journal of Coal Geology, 35: 117–145.
50. Luxbacher, K. D., Erdogan, S. S., and Karacan, C. Ö. (2009). Modeling methane emissions and ventilation needs by examination of mining induced permeability changes and related damage to ventilation controls. Proceedings, 43rd US Rock Mechanics Symposium and 4th U.S.-Canada Rock Mechanics Symposium, June, Asheville, NC, paper ARMA-09-146
51. Mohammed, E.A., Naugler C, Far B.H. (2015), Emerging Business Intelligence Framework for a Clinical Laboratory Through Big Data Analytics, Chapter 32, in Editor(s): Quoc Nam Tran, Hamid Arabnia, Emerging Trends in Computer Science and Applied Computing, Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology, Morgan Kaufmann, pp. 577-602, ISBN 9780128025086, https://doi.org/10.1016/B978-0-12-802508-6.00032-6.
52. National Institute of Standards and Technology (NIST). (2003). Engineering Statistics Handbook. https://www.itl.nist.gov/div898/handbook/, Accessed December 29, 2022.
53. 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.
54. Patten, M. L. (2005). Proposing empirical research: A guide to the fundamentals. In Glendale, CA: Pyrczak Pub.
55. Ramasamy, A. M. S. (1994). A Numerical Method for Forecasting. Center for Future Studies, Pondicherry University.
56. Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, URL https://www.GaussianProcess.org/gpml
57. 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.
58. Shumway, R. H., and Stoffer, D. S. (2006). Time Series Analysis and its Applications with R Examples. Time Series Analysis and Its Applications (Second Ed). Springer.
59. Sung W, Ertekin T, Ramani RV (1987) Evaluation of the impact of the degasification process on inflow of methane gas into coal mines: A numerical exercise. Proceedings, 3rd U.S. Mine Ventilation Symposium, University Park, Pennsylvania, pp. 328-333 (October, 1987).
60. 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. Petroria, 1993
61. Thomas, S. (2014). Basic Statistics, 1st (ed). Alpha Science International Ltd. Oxford, United Kingdon.

62. 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/10.1201/9781439833 742.ch48
63. Tructin, W., and Wasilewski, S. (1987). Application of digital filter and time series analysis in ventilation control systems, Proceedings, 3rd North American Mine Ventilation Symposium, 12-14 October, Pennsylvania.
64. 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
65. Wasilewski S (2014) Influence of barometric pressure changes on ventilation conditions in deep mines. Archives of Mining Sciences 59(3):621-639. https://doi.org/10.2478/amsc-2014-0044
66. Xu L, Lin X, Amen J, et al (2014) Impact of changes in barometric pressure on landfill methane emission. Global Biogeochemical Cycles 28(7):679-695. https://doi.org/10.1002/2013GB004571
67. Yuan L, Smith AC (2010) Modeling the effect of barometric pressure changes on spontaneous heating in bleederless longwall panels. Transactions of the Society for Mining, Metallurgy, and Exploration, 2010, Vol. 328, No. 1, pp. 485-492.

Appendix 1: Data Pre-Processing (Cleaning and Filtering)

This section contains information about the techniques implemented for filtering and cleaning the data collected from Mine A, Mine B and the Weather Underground website. In addition, some examples about finding missing values and removing outliers due to sensor malfunction are presented.

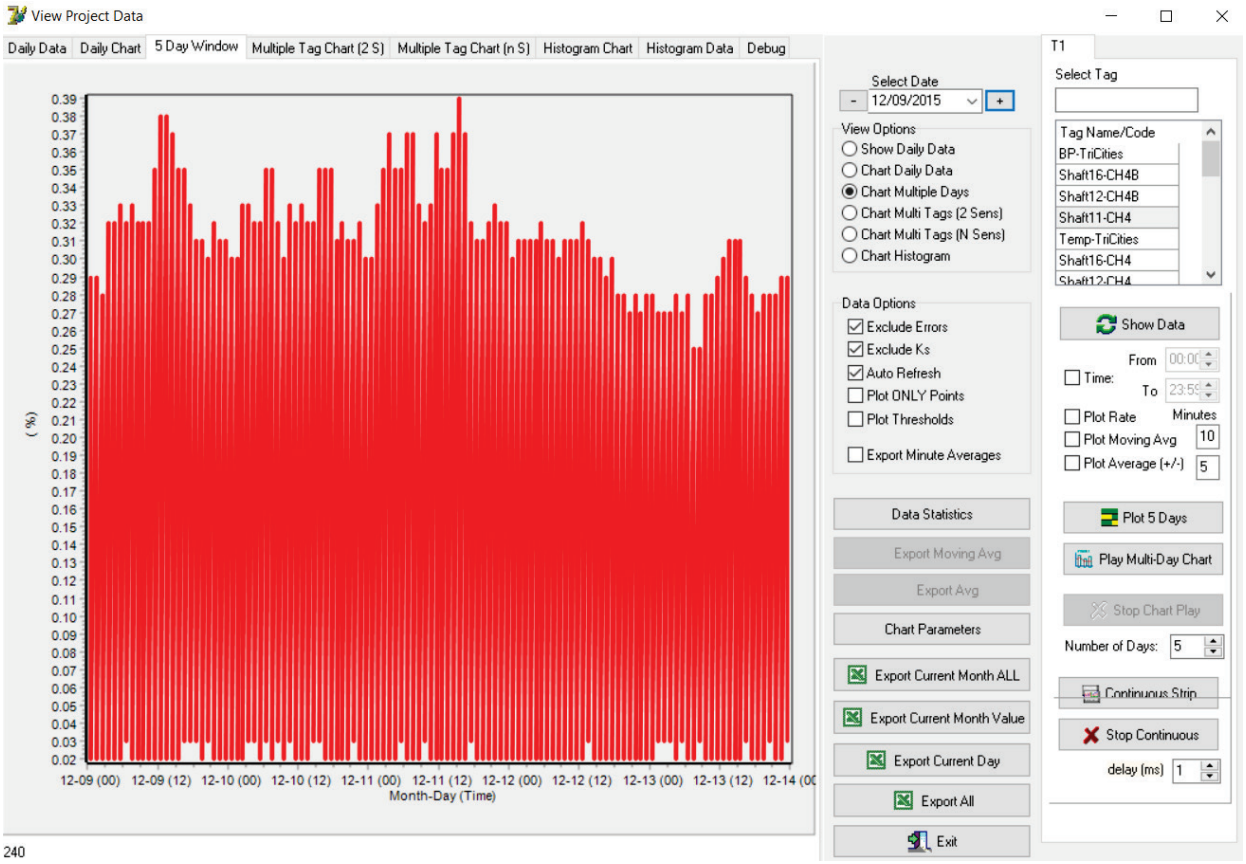


Figure 29: Segment of methane data from Shaft A, Mine A – Before data is cleaned up (AMANDA interface)

Figure 26 shows a segment of methane gas data from Shaft A of Mine A that corresponds to five days period of time. The data includes a significant number of zero values due to sensor malfunction (this information was verified by a discussion with mine personnel).

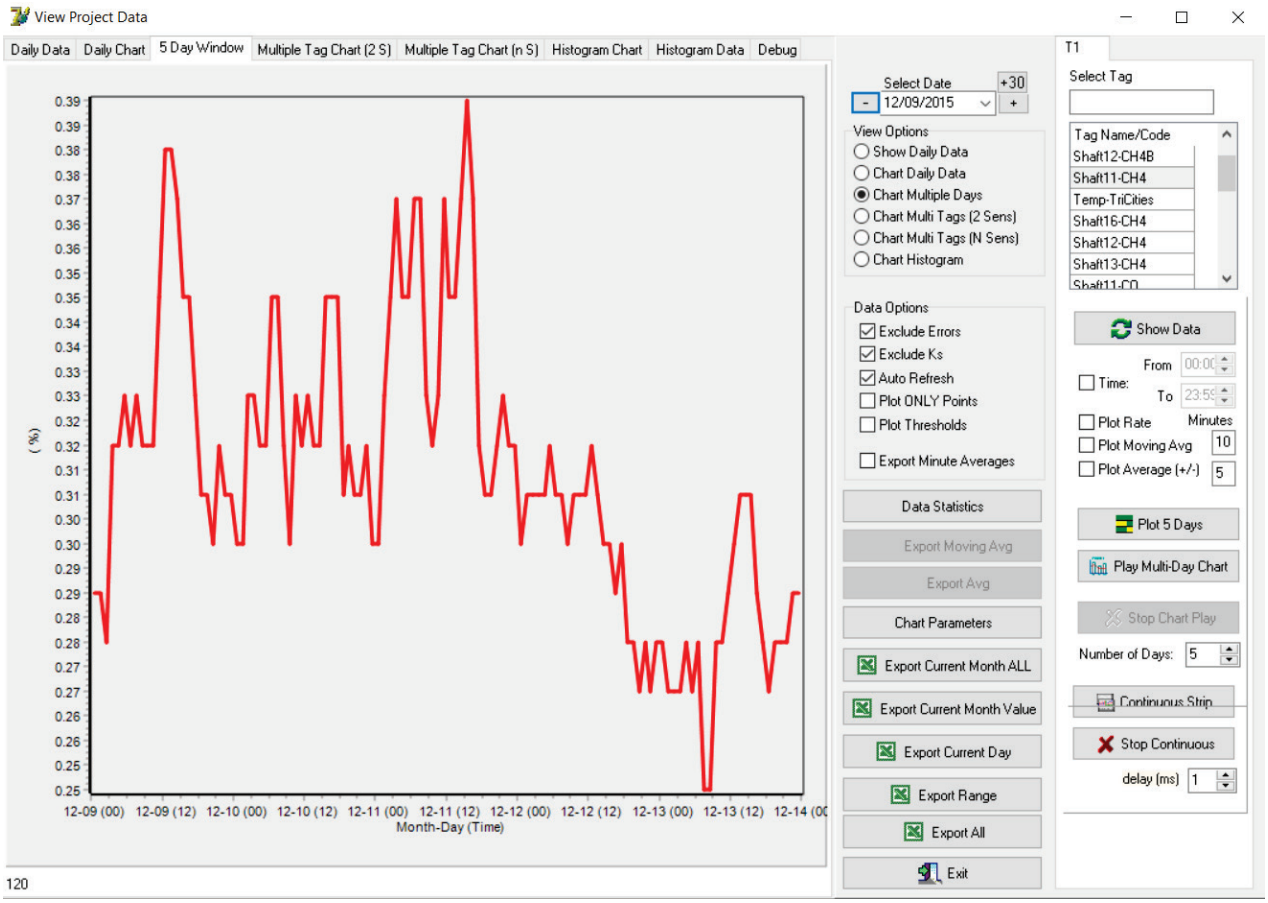


Figure 30: Segment of methane data from Shaft A, Mine A – After data is cleaned up (AMANDA Interface)

Figure 27 shows a segment of methane gas data from Shaft A of Mine A that corresponds to a five-day period. This is the same period presented in Figure 26. However, the data does not include any zero values as the data has been filtered and cleaned up, as shown in Figure 27.

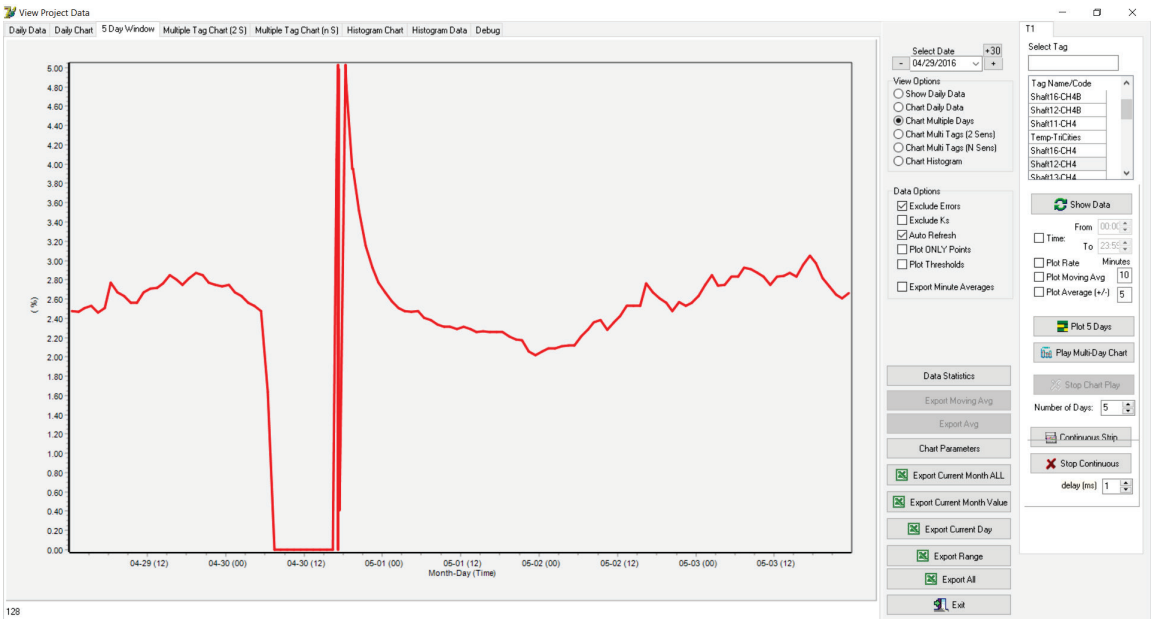


Figure 31: Segment of methane data from Shaft B, Mine A – Before data is cleaned up (AMANDA Interface)

Figure 28 shows a segment of methane gas data from Shaft B of Mine A that corresponds to a five-day period. The data includes some spikes or outliers due to sensor malfunction (this information was verified by a discussion with mine personnel).

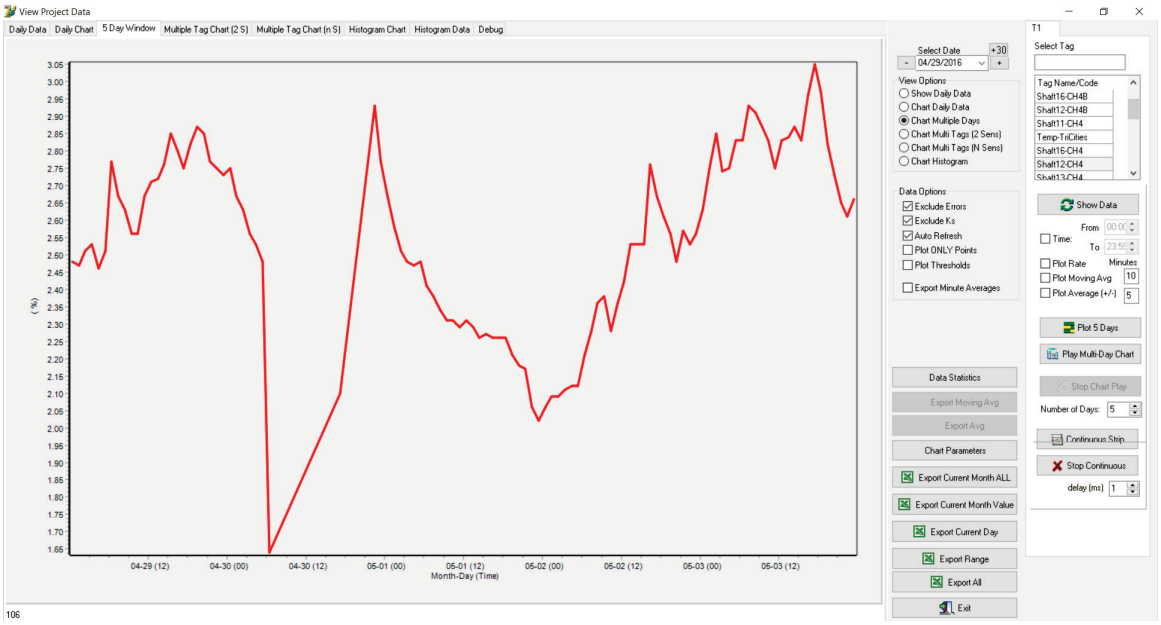


Figure 32. Segment of methane data from Shaft B, Mine A – After data is cleaned up (AMANDA Interface)

Figure 29 shows a segment of methane gas data from Shaft B of Mine A that corresponds to a five-day period. This is the same period presented in Figure 28. However, the data does not show any spikes or outlier values as the data has been filtered and cleaned up.

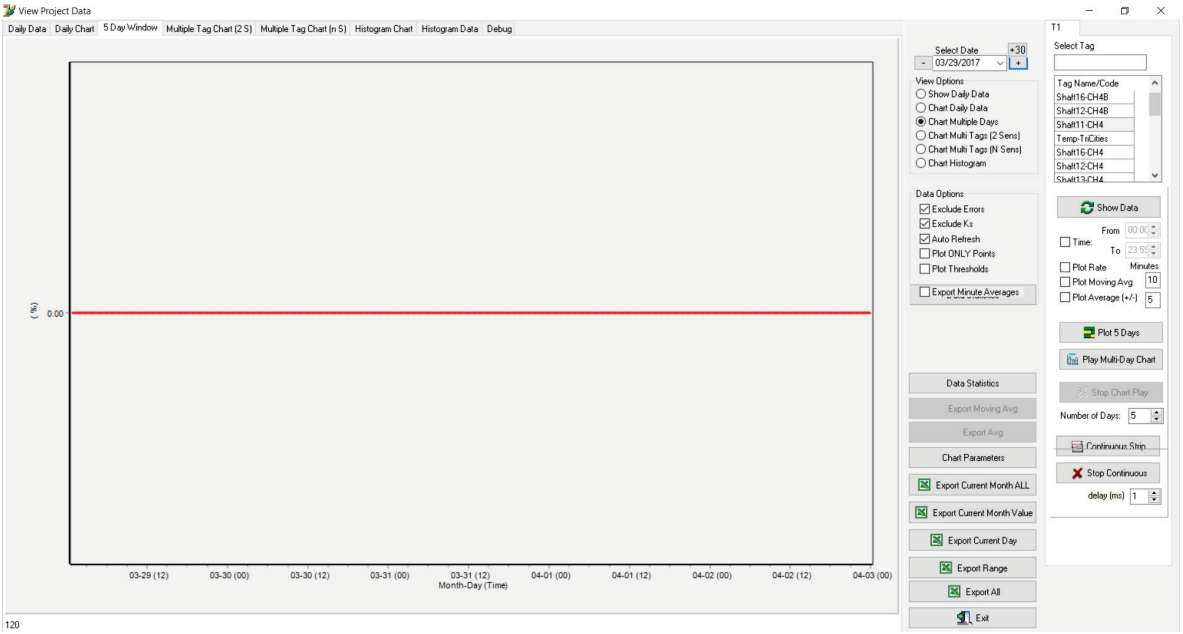


Figure 33: Identification of zero values data – Shaft A, Mine A (AMANDA Interface)

Figure 30 presents a segment of data from Shaft A of Mine A for a period of one year, where the mine sensors record a methane gas concentration equal to zero percent (0%).

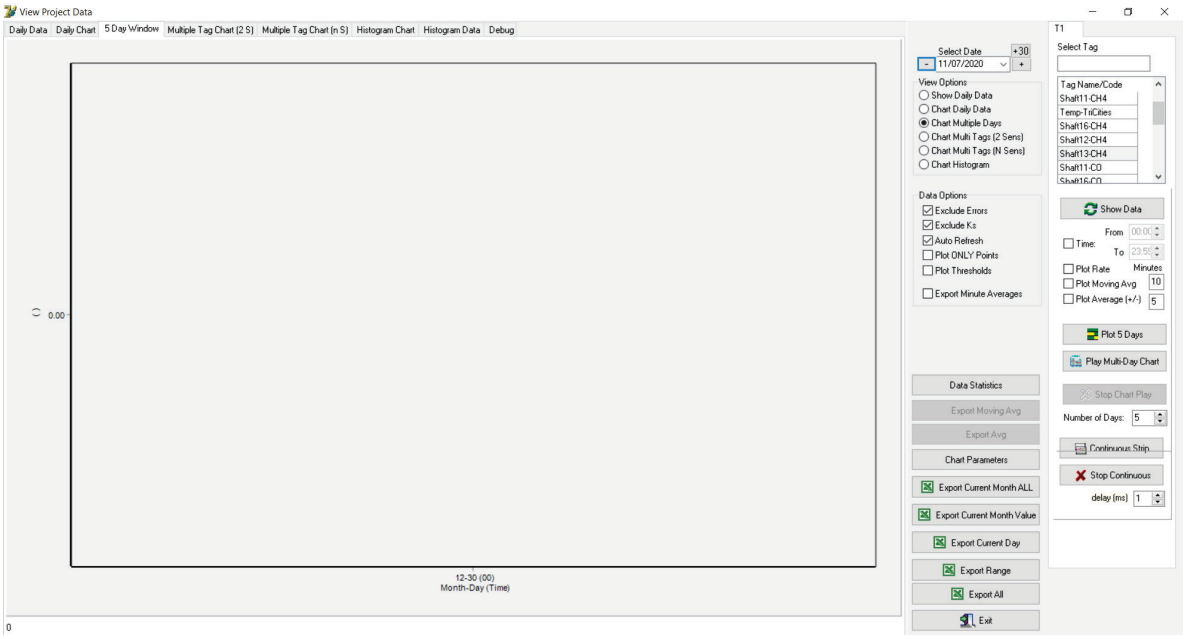


Figure 34: Identification of no data – Shaft C, Mine A (AMANDA Interface)

Figure 31 presents a segment of data from Shaft C of Mine A for a period of five days where there is no data recorded.

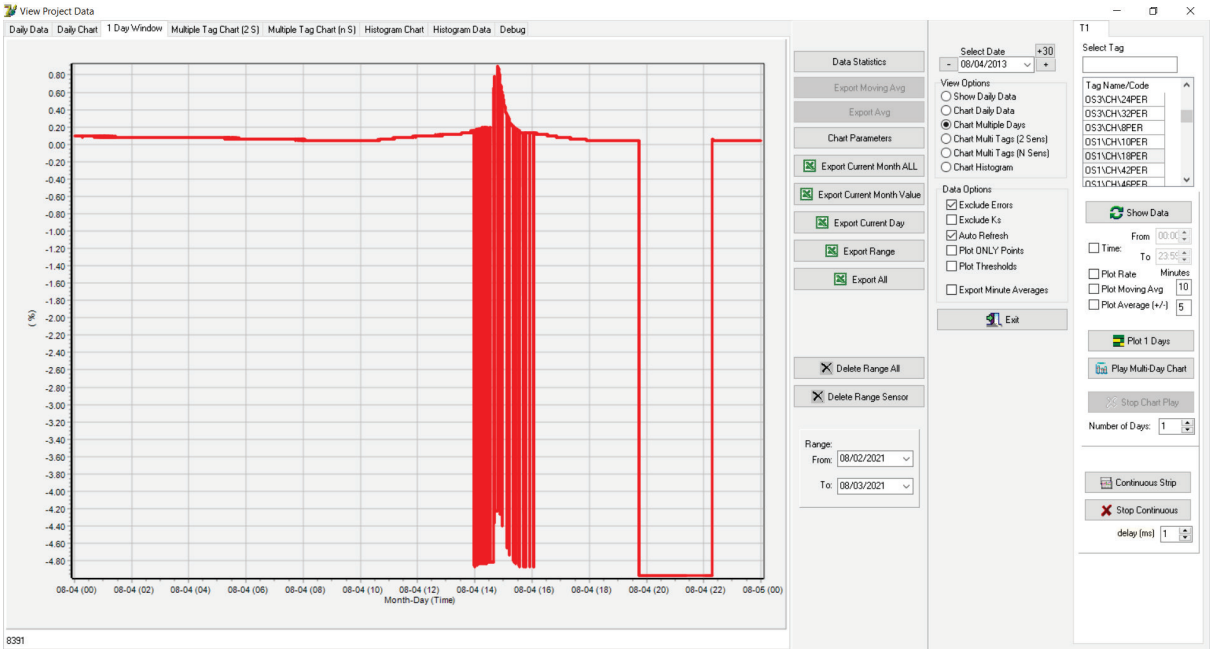


Figure 35: Methane data from Mine B– Before data is cleaned up

Figure 32 shows methane gas data from Mine B that corresponds to one day. The data includes some negative values (outliers) due to sensor malfunction.

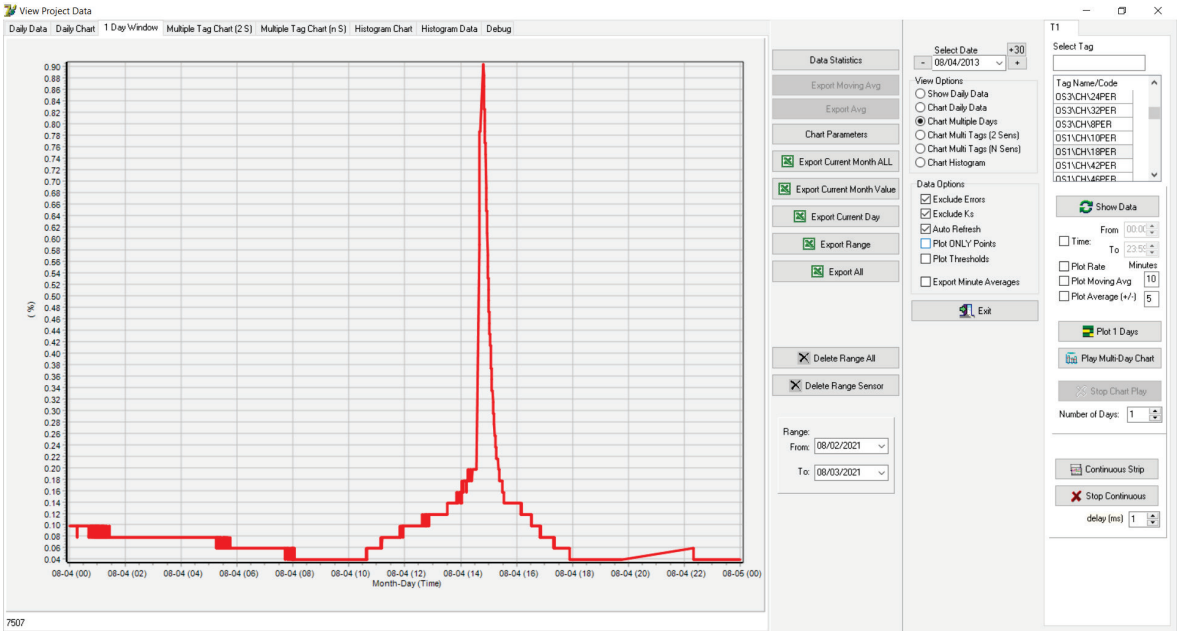


Figure 36: Methane data from Mine B– After data is cleaned up (AMANDA Interface)

Figure 33 shows a segment of methane gas data from Mine B that corresponds to one day. This is the same day presented in Figure 32. However, the data has no negative values as the data has been filtered and cleaned up.

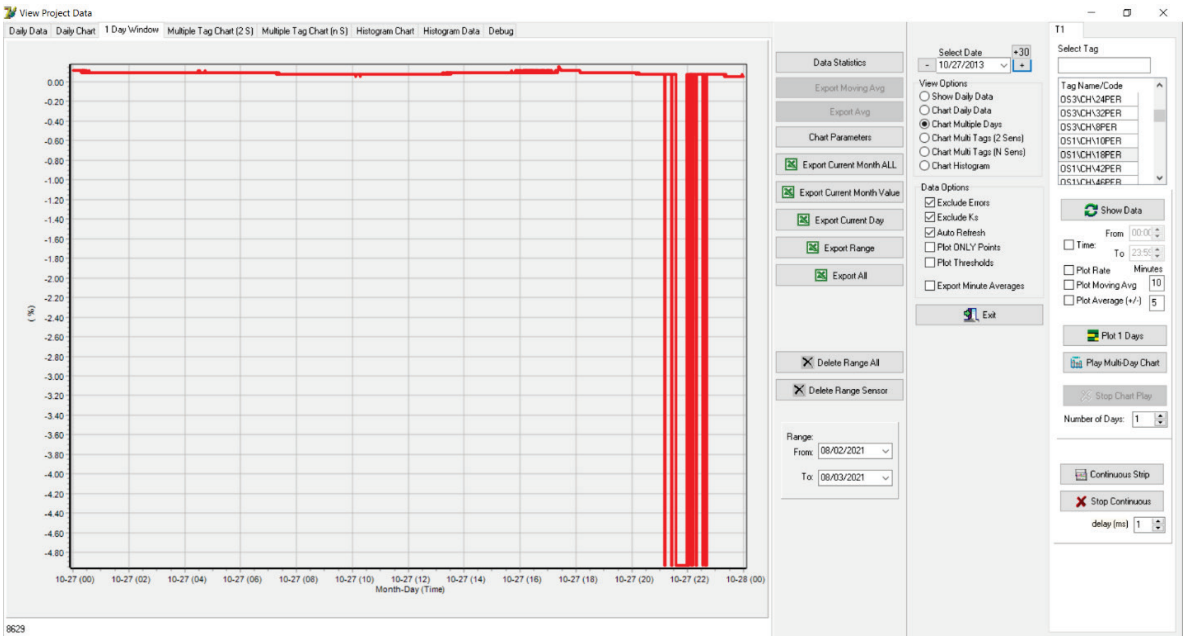


Figure 37: Methane data from Mine B– Before data is cleaned up (AMANDA Interface)

Figure 34 presents methane gas data from Mine B that corresponds to one day. The data presents some negative values due to sensor malfunction.

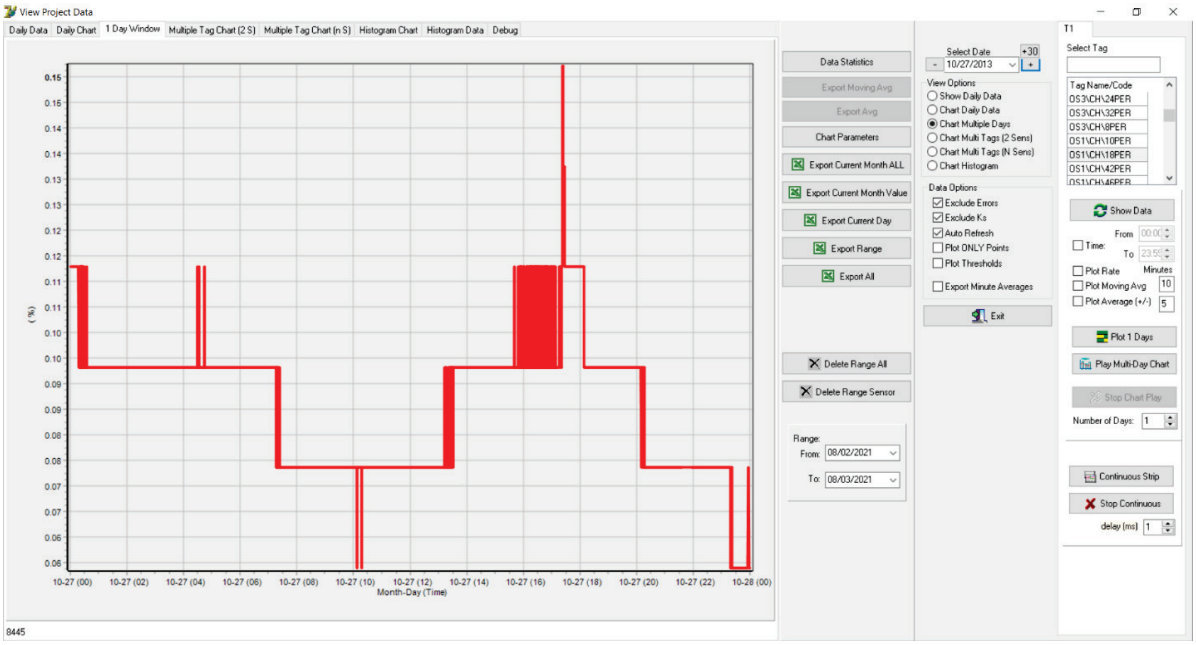


Figure 38: Methane data from Mine B– After data is cleaned up (AMANDA Interface)

Figure 35 shows a segment of methane gas data from Mine B that corresponds to one day. This is the same day presented in Figure 34. However, the data has no negative values or outliers as the data has been filtered and cleaned up.

Appendix 2: Exploratory Data Analysis

The following presents the preliminary results of the interpolation of the raw data collected over five years from sensors at Shaft A of Mine A and barometric pressure from closest public weather station. The following figures were obtained employing exploratory data analysis, interpolation, correlation, and cross-correlation techniques.

Figure 36 shows the interpolation of raw data from the methane gas concentration collected from shaft A, Mine A and barometric pressure from the closest public weather station. The methane concentration and barometric pressure data were collected over five years. The blue line indicates barometric pressure (inWG) (left axis), while the red line represents methane gas concentration (%) (right axis).

Spikes in the methane data correspond to calibration measurements. Also, there are time intervals where no data were recorded.

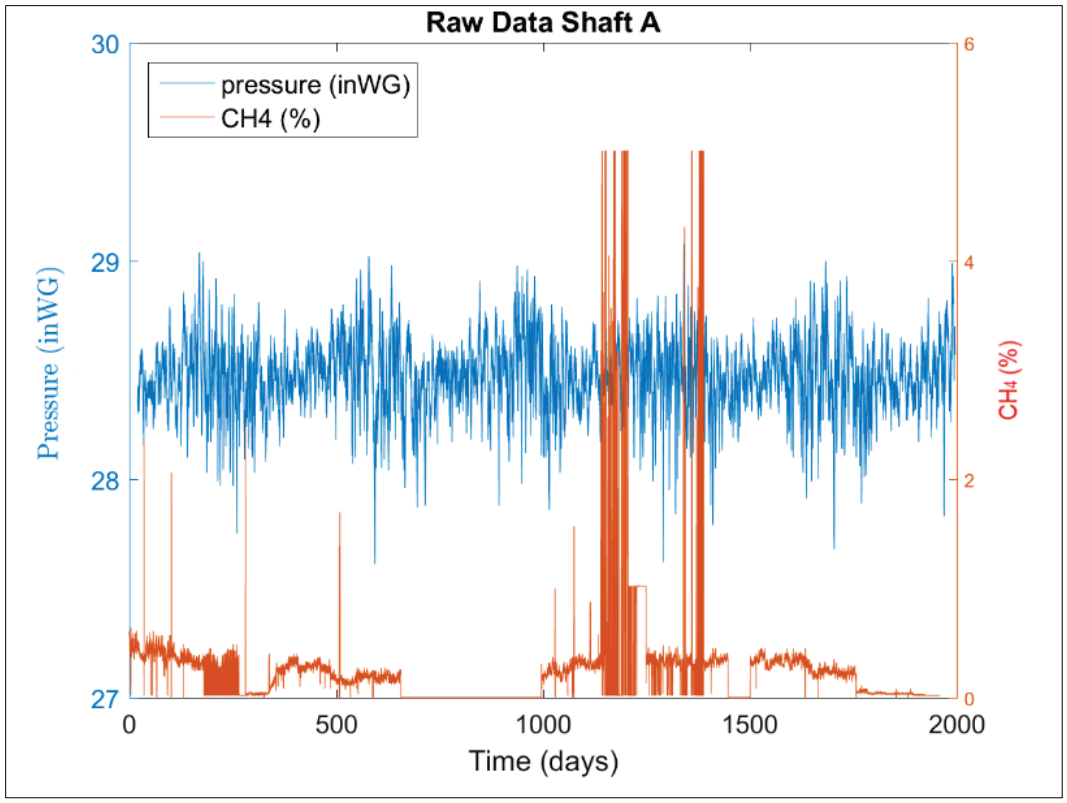


Figure 39: Interpolation of raw data from shaft A and atmospheric pressure.

Figure 37 shows the interpolation of raw data from methane gas concentration collected from shaft A, Mine A and barometric pressure from the closest public weather station. Methane concentration and barometric pressure data are plotted for a segment of 180 days. The blue line in the top plot indicates interpolated barometric pressure data (inWG), while the green line in the bottom plot represents interpolated data of methane gas concentration (%) to correspond to the timestamps of the barometric pressure data.

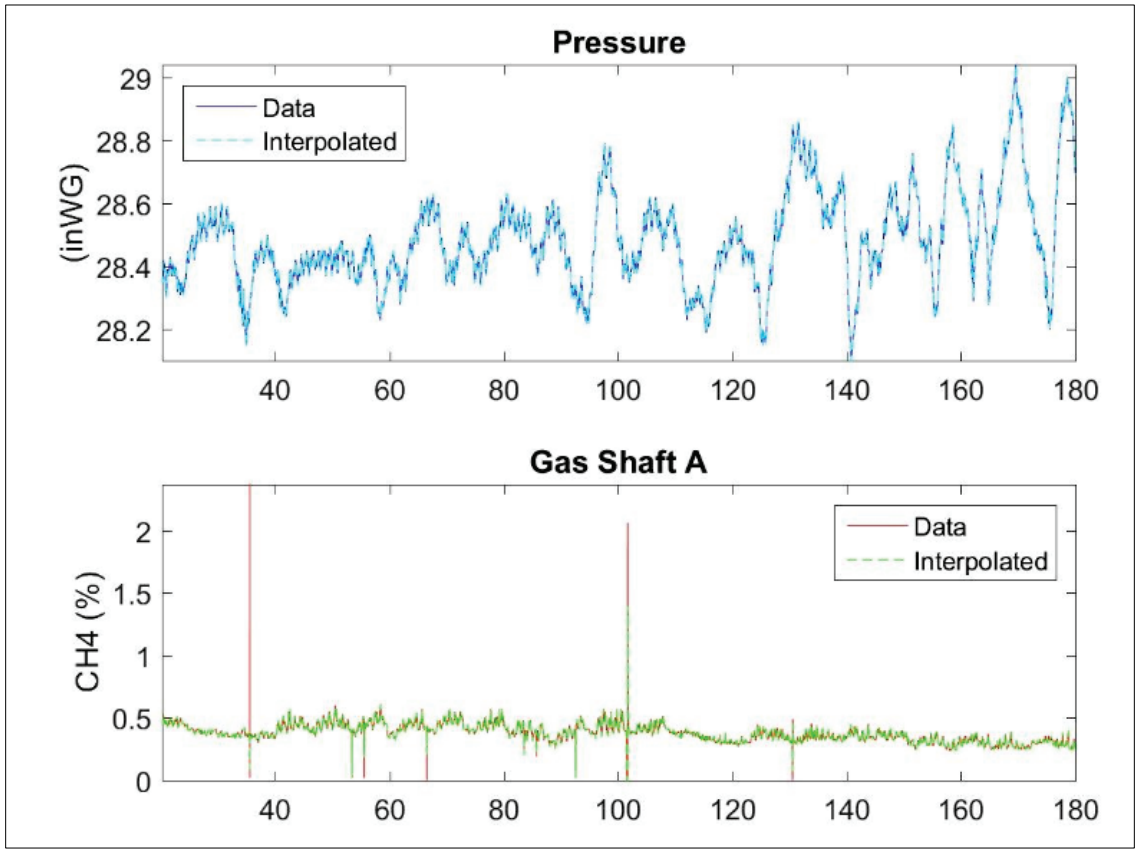


Figure 40: Interpolation segment of raw data from shaft A and barometric pressure.

Figure 38 shows a preliminary correlation between methane gas concentration from shaft A of Mine A and barometric pressure from the closest public weather station for a subset of 160 days. Daily data were aggregated from hourly values for both methane concentration and barometric pressure and then plotted together, as shown in Figure 38.

The blue line indicates barometric pressure (inWG) (left axis), while the red line represents methane gas concentration (%) (right axis). Different correlations and trends can be observed in Figure 38. For example, an inversely proportional relationship between barometric pressure and methane gas rate can be identified between around days 40, 70, 155, and 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 38 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.

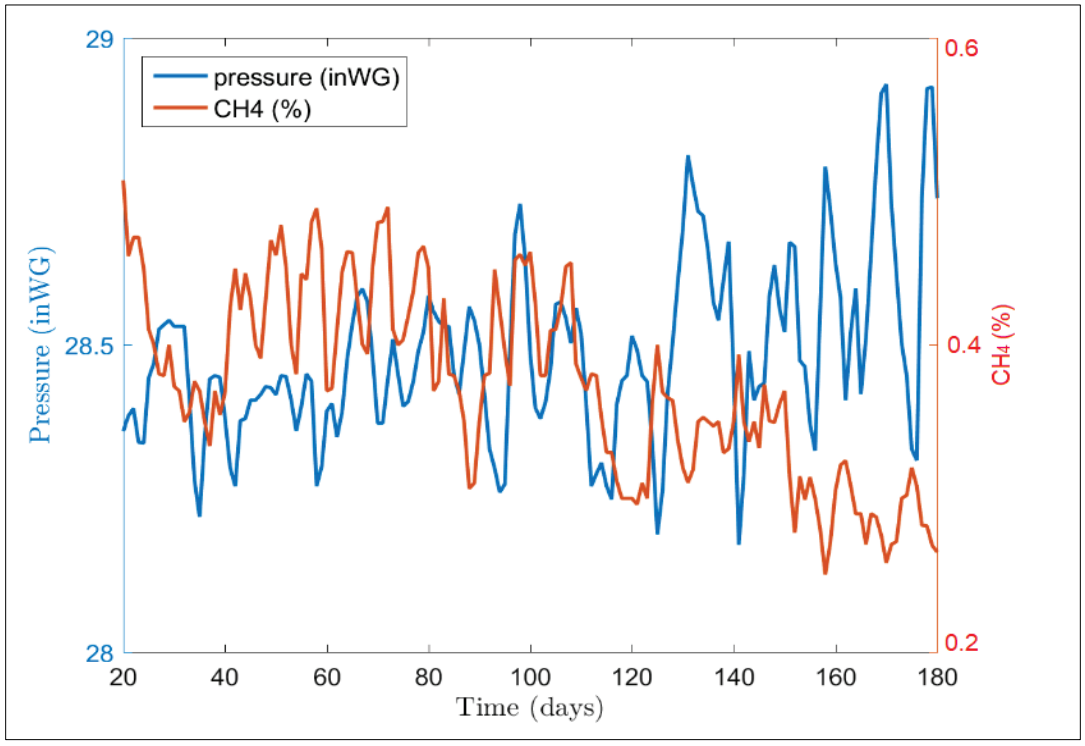


Figure 41: Correlation of time series data for Shaft A.

Figure 39 shows the correlation between methane gas concentration from shaft A, Mine A (X-axis) and barometric pressure from the closest public weather station (Y-axis). This dataset corresponds to 160 days of collected data. A negative correlation between methane gas and barometric pressure is illustrated in Figure 39. The presence of outliers can be attributed to different factors, such as calibration practices, data collection processes, or external elements.

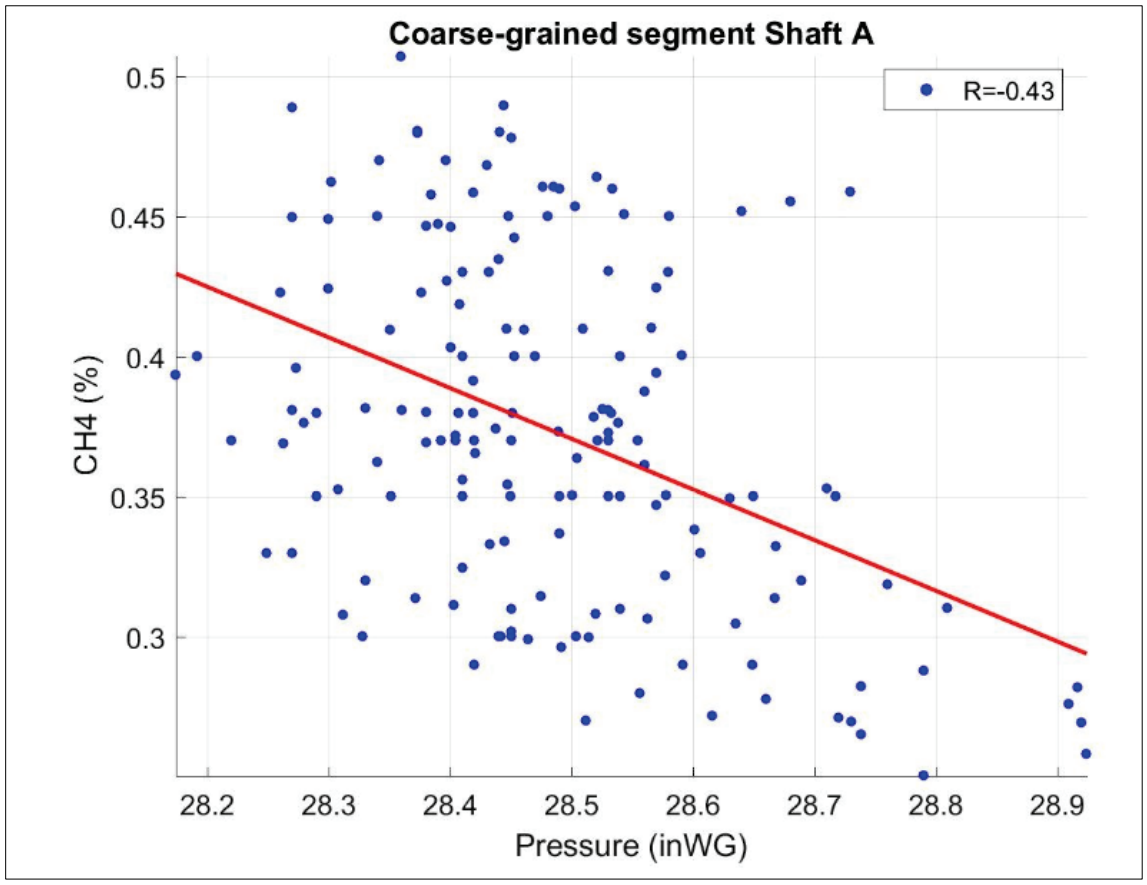


Figure 42: Scatter plot between methane gas concentration for shaft A and barometric pressure.

Figure 40 illustrates the correlation between methane gas concentration from Shaft A of Mine A (X-axis) and barometric pressure from the closest public weather station (Y-axis). This dataset corresponds to over 2,000 days of collected data. A negative correlation between methane gas and barometric pressure is illustrated in Figure 40, i.e., when barometric pressure increases, methane gas concentration decreases. The presence of outliers can be attributed to different factors, such as calibration practices, data collection processes, or external elements.

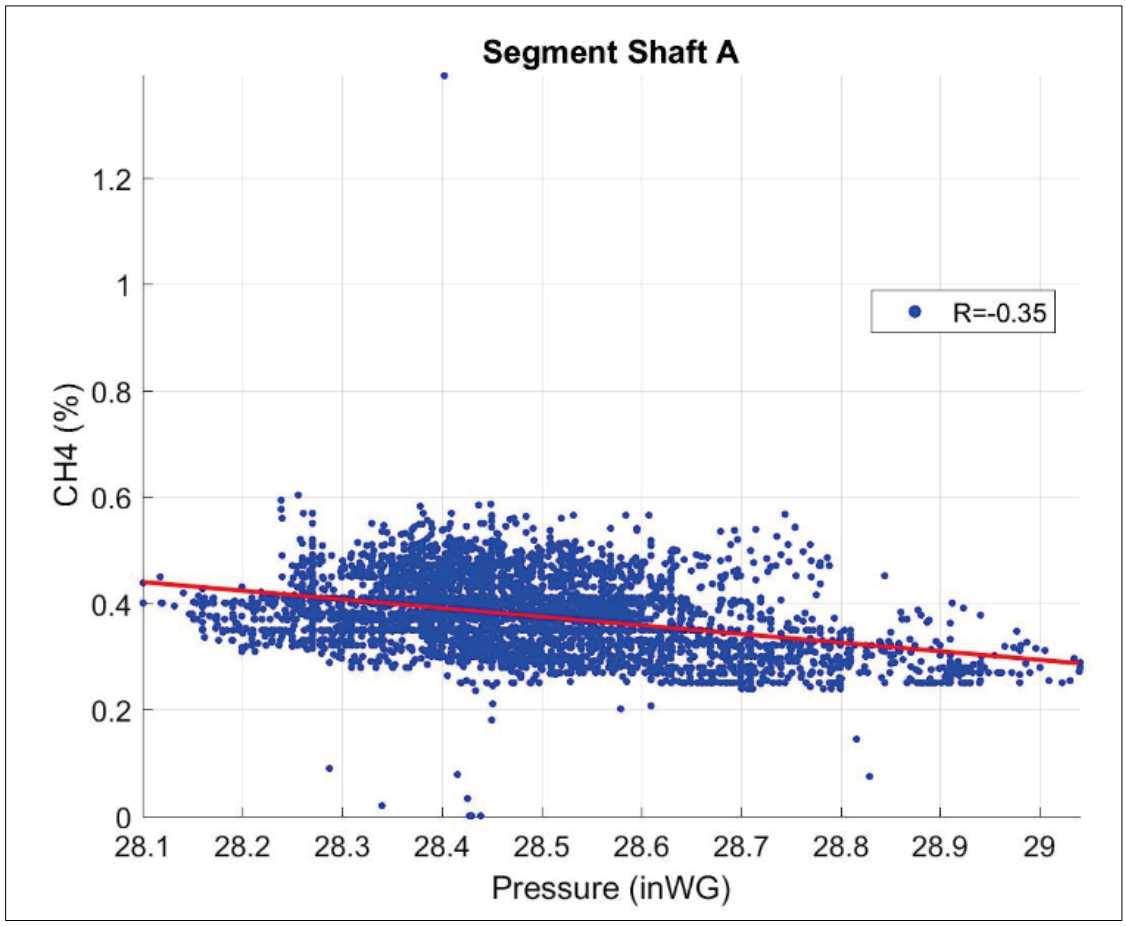


Figure 43: Scatter plot of a segment between methane gas concentration for shaft A and barometric pressure.

As a result of the above analysis, it was decided to use both daily averages and 12-hour averages to develop homogenized datasets.

The following presents the second phase of data pre-processing using the experience gained by previous work. the pre-processing of raw data collected over six years from the sensors at Mine A and barometric pressure from the nearest public weather station to assess the potential correlation between methane gas and barometric pressure and coal production using the Pearson correlation coefficient (R). The following figures correspond to data segments with different lengths and time steps (e.g., daily and weekly) obtained employing AMANDA.

In Figure 41, the red line represents methane gas concentrations (%), and the green line denotes barometric pressure; both time series sampled with a daily average step corresponding to an interval of 30 days. Figure 41 shows a strong negative correlation between methane gas and barometric pressure with the Pearson correlation coefficient calculated at R=-0.92. The negative sign indicates that methane concentration increases when barometric pressure decreases and vice versa.

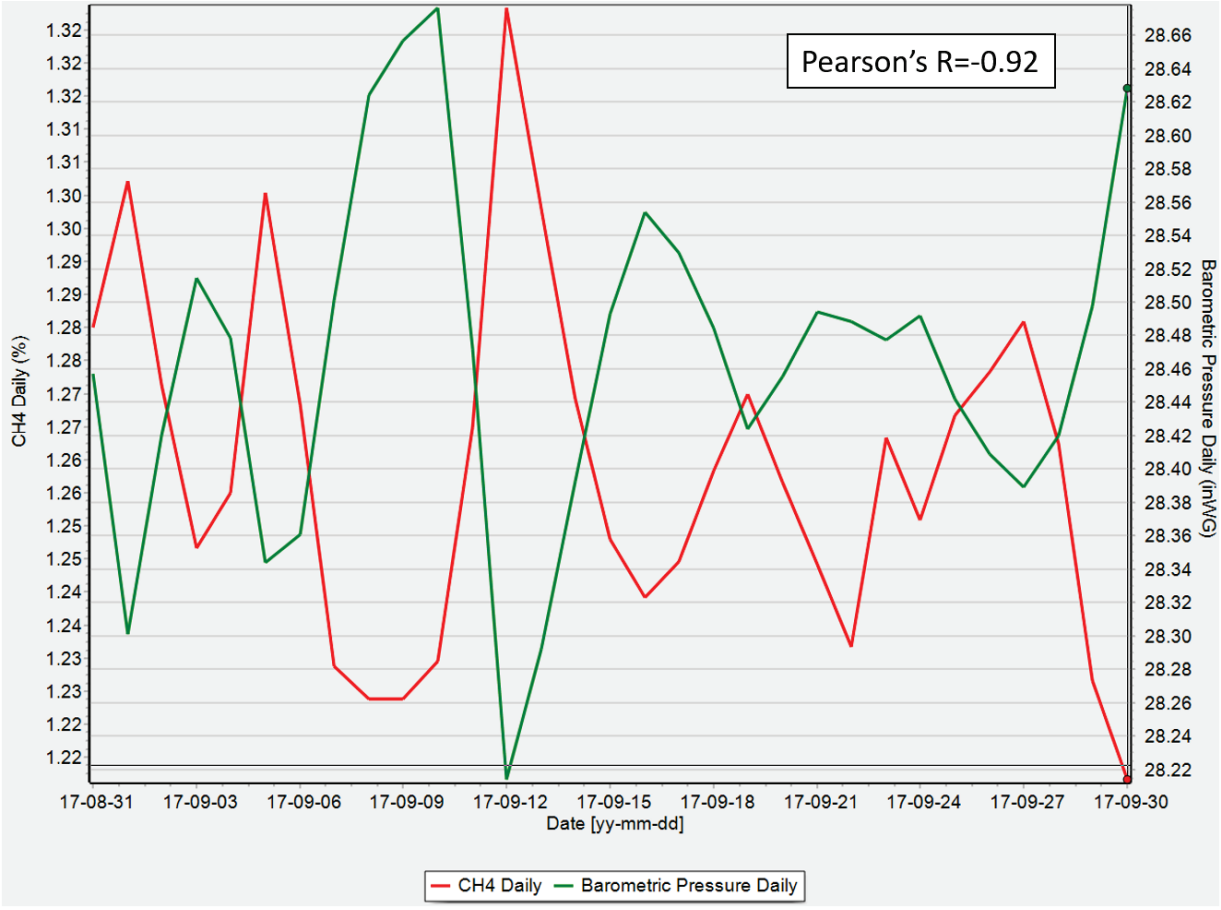


Figure 44: Correlation of CH4 vs. BP for 30 days - segment 1 – with a daily average step

In Figure 42, the red line represents methane gas concentrations (%), and the green line denotes barometric pressure; both time series sampled with a daily average step corresponding to an interval of 30 days. Figure 42 indicates a strong negative correlation between methane gas and barometric pressure with the Pearson correlation coefficient calculated at R=-0.92. The negative sign indicates that methane concentration increases when barometric pressure decreases and vice versa.

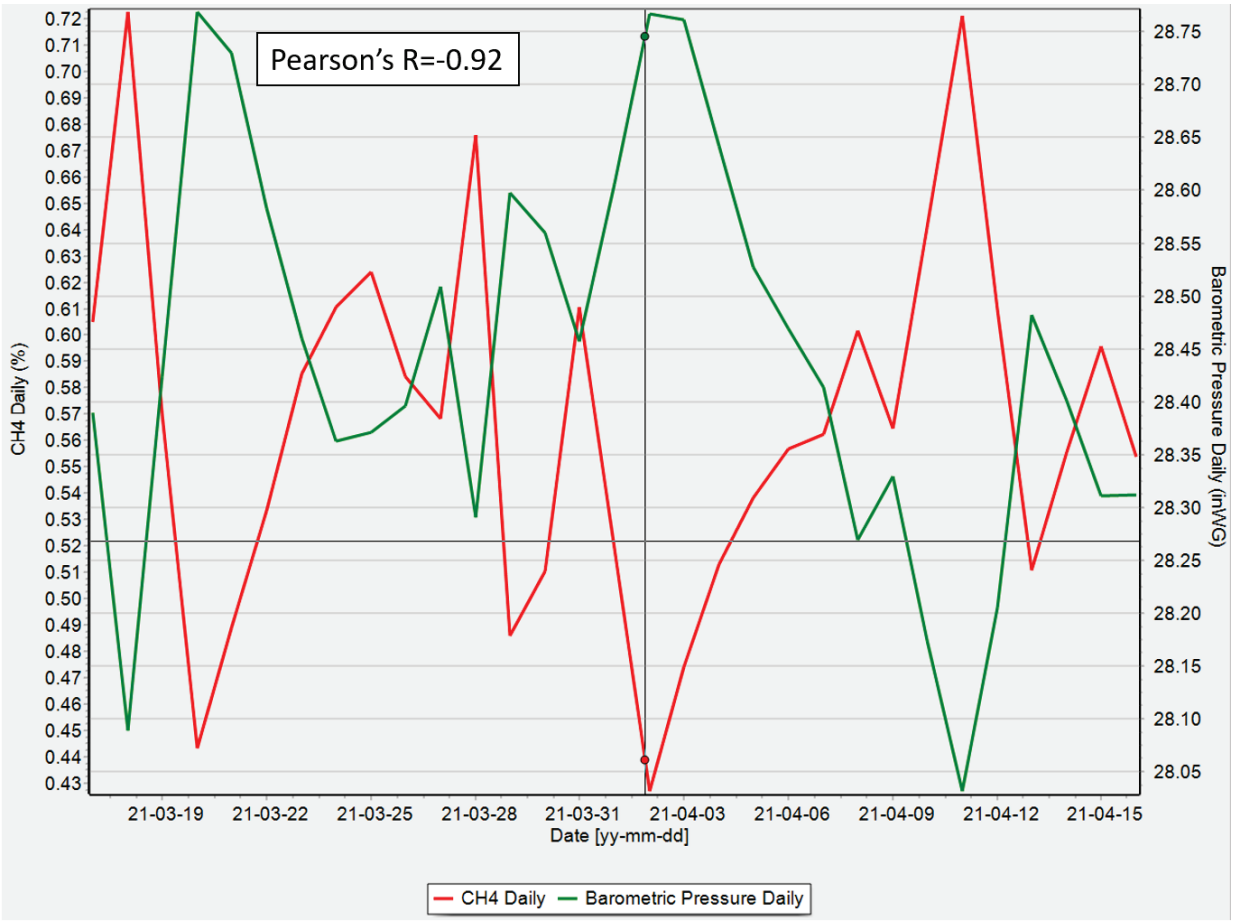


Figure 45: Correlation of CH4 vs. BP for 30 days - segment 2 – with a daily average step

In Figure 43, methane gas and barometric pressure time series are sampled with a daily average step and correspond to an interval of 120 days (4 months). Figure 43 shows a strong negative correlation between methane gas and barometric pressure with the Pearson correlation coefficient calculated at R=-0.81. The negative sign indicates that methane gas and barometric pressure have an inverse correlation (when one increases, the other decreases and vice versa).

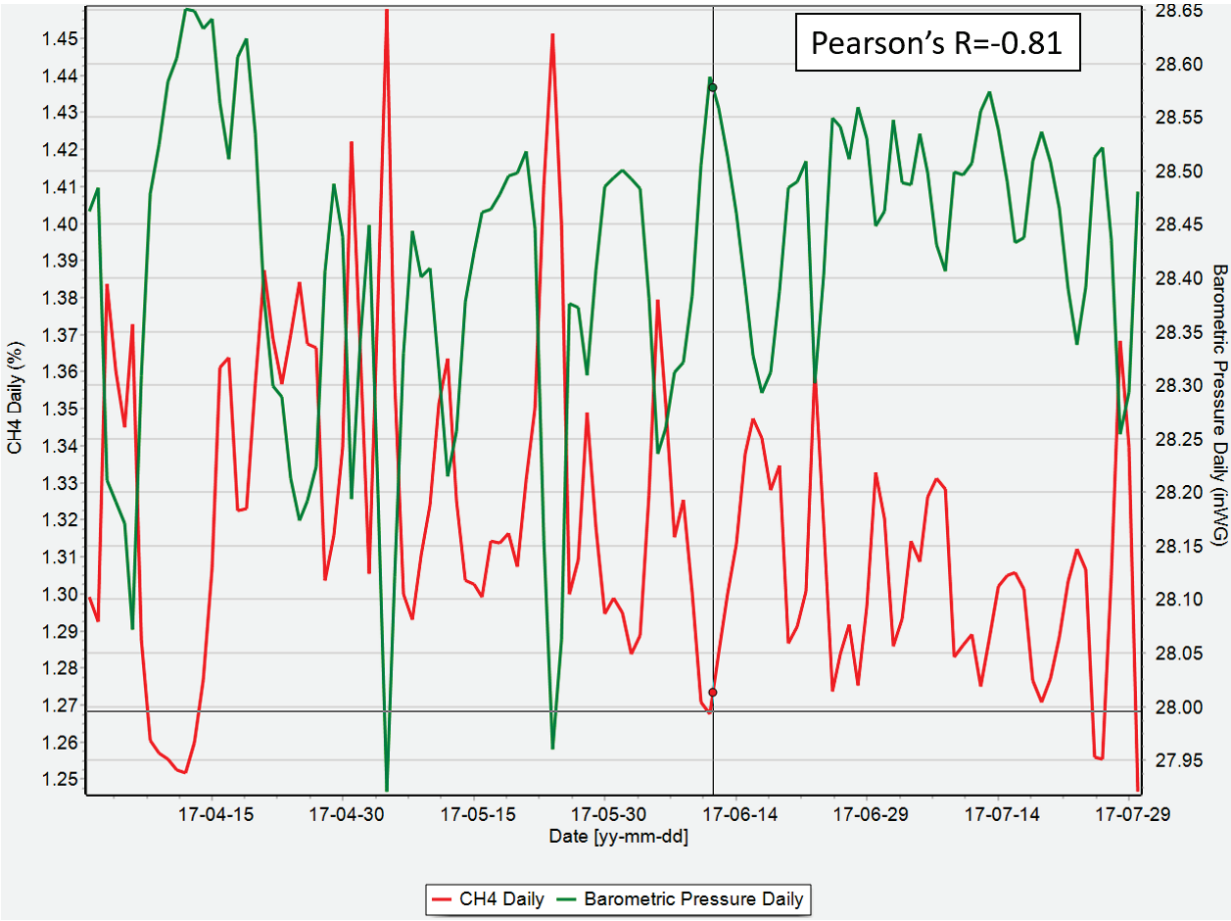


Figure 46: Correlation of CH4 vs. BP for 120 days - segment 3 – with a daily average step

In Figure 44, methane gas and barometric pressure time series are sampled with a daily average step and correspond to an interval of 120 days (4 months). Figure 44 shows a strong negative correlation between methane gas and barometric pressure with the Pearson correlation coefficient calculated at R=-0.78. The negative sign indicates that methane gas and barometric pressure have an inverse correlation (when one increases, the other decreases and vice versa).

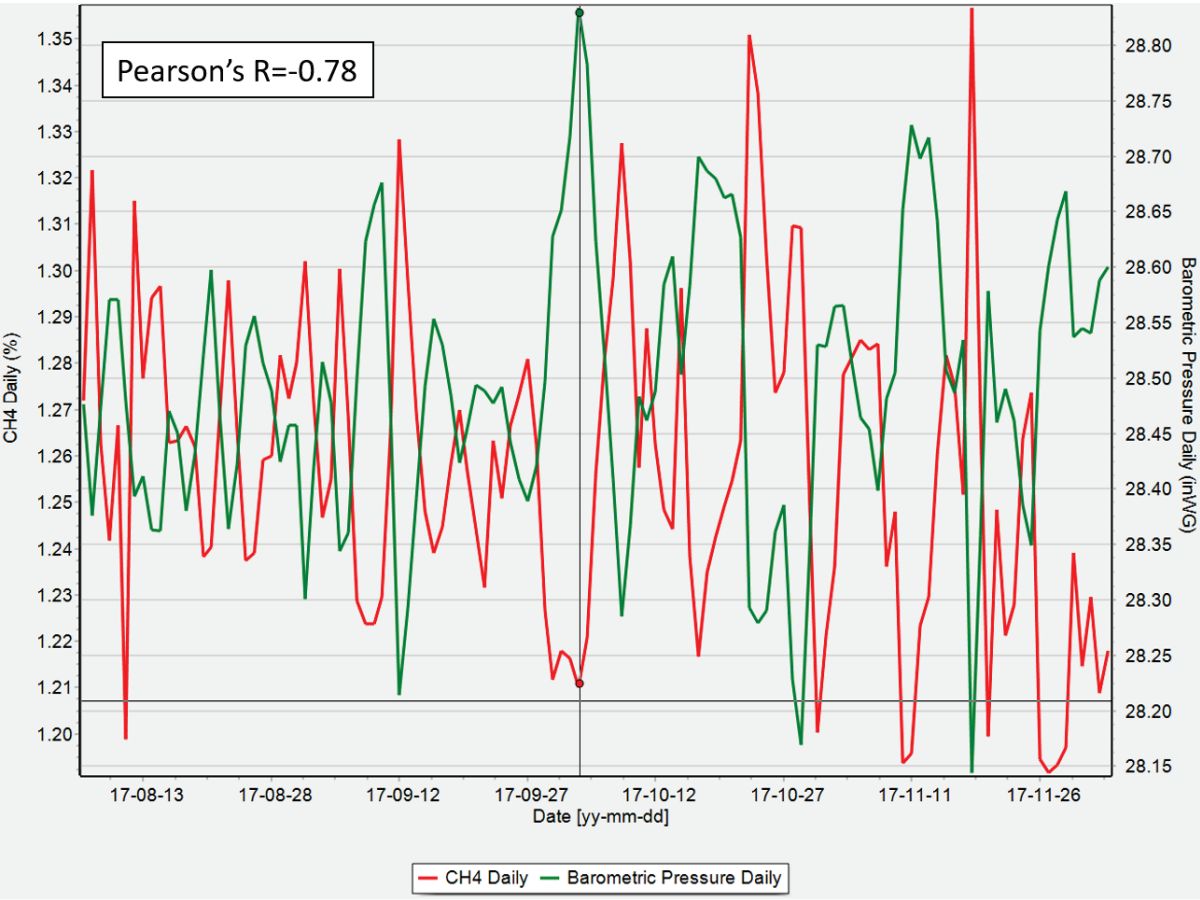


Figure 47: Correlation of CH4 vs. BP for 120 days - segment 4 – with a daily average step

Figure 45 shows two time series, methane gas and barometric pressure; both are sampled using a weekly average step and correspond to an interval of 8 weeks (2 months). Figure 45 reveals a strong negative correlation between methane gas and barometric pressure with the Pearson correlation coefficient calculated at R=-0.93. The negative sign indicates that methane gas and barometric pressure have an inverse correlation.

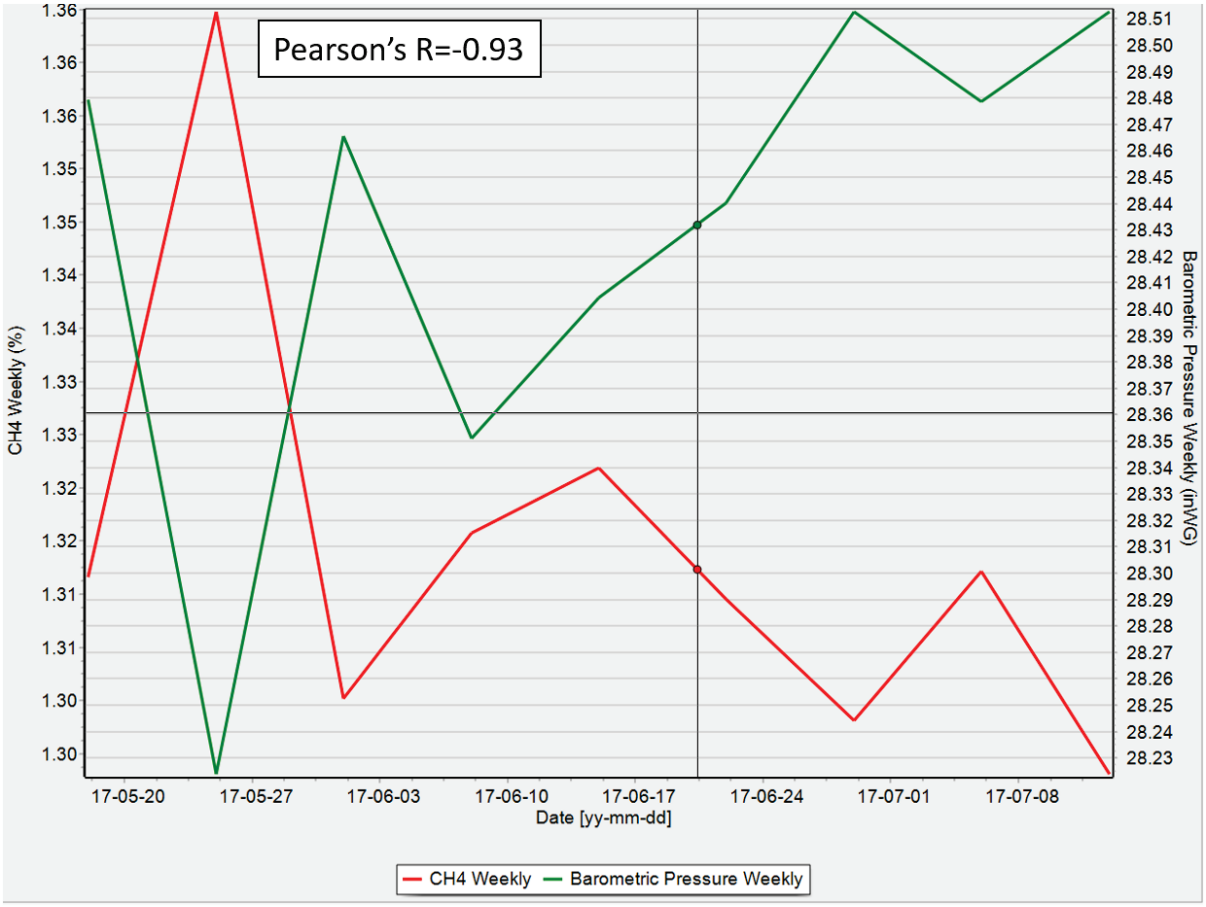


Figure 48: Correlation of CH4 vs. BP for 8 weeks - segment 5 – with a weekly average step

Figure 46 shows two time series, methane gas, and barometric pressure; both are sampled using a weekly average step and correspond to an interval of 26 weeks (6 months). Figure 46 reveals a strong negative correlation between methane gas and barometric pressure with the Pearson correlation coefficient calculated at R=-0.72. The negative sign indicates that methane gas and barometric pressure have an inverse correlation.

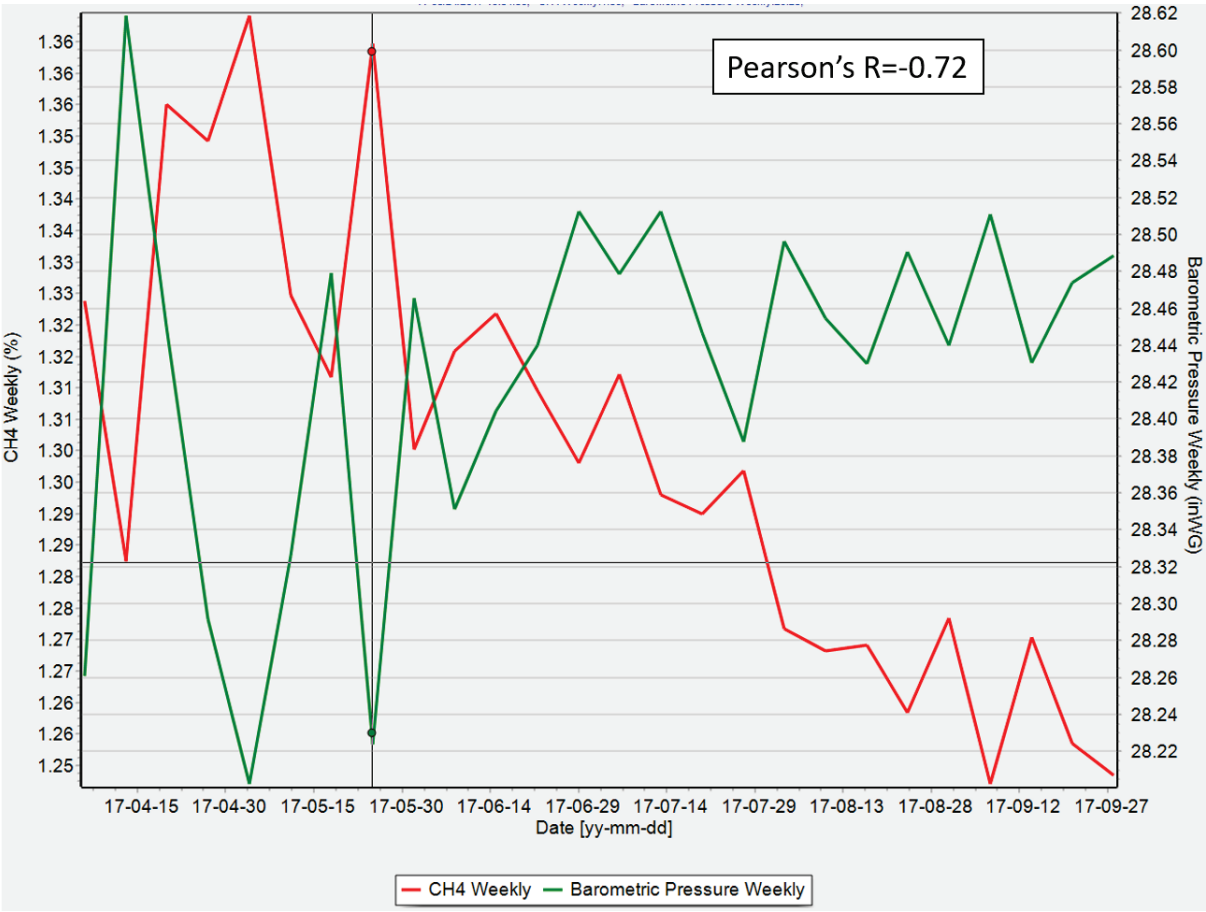


Figure 49: Correlation of CH4 vs. BP for 26 weeks - segment 6 – with a weekly average step

Figure 47 shows two time series; methane gas concentrations and barometric pressure, both sampled on a daily average basis for a different interval of 30 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 behavior is most likely to a sudden increase of the methane gas time series from values about 0.20% to more than 2.80%, as shown in Figure 47. Such discontinuities can directly affect the outcomes of any method used for data analysis. They also suggest that the gas concentration depends not only on the barometric pressure but also on factors related to mine operations.

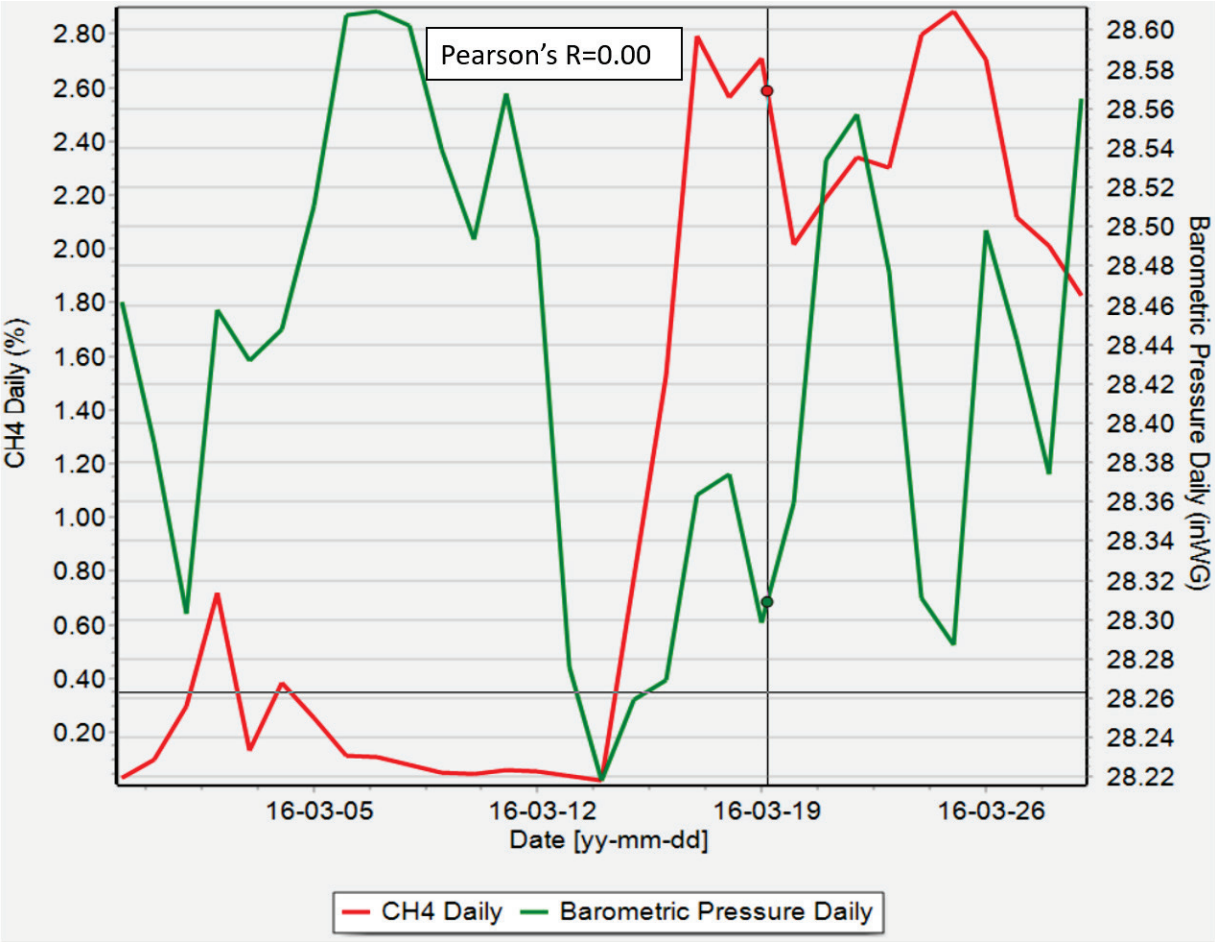


Figure 50: Correlation of CH4 vs. BP for 30 days - segment 7 – with a daily average step

Figure 48 presents the methane gas concentrations and barometric pressure time series, both sampled on a daily average basis for a different interval of 180 days. Again, these plots reveal a lack of correlation between methane gas and barometric pressure, with the Pearson correlation coefficient estimated at R=0.00. This odd behavior could be explained by the presence of discontinuities that affect the correlation between these two variables.

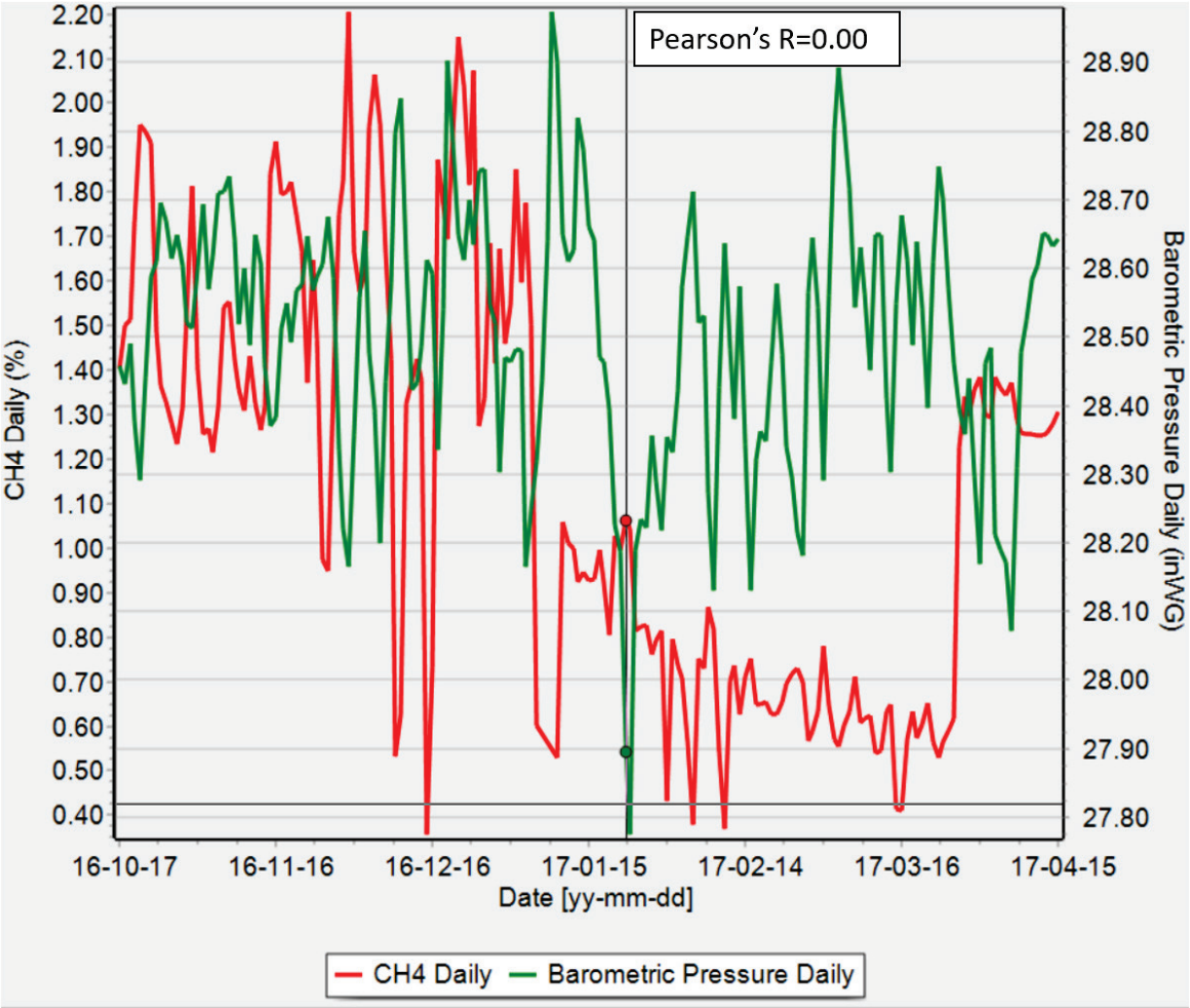


Figure 51: Correlation of CH4 vs. BP for 180 days - segment 8 – with a daily average step

Figure 49 shows two time series, methane gas concentrations, and barometric pressure; both sampled on a weekly average basis for a different interval of 180 days (26 weeks). These plots reveal a positive correlation between methane gas (red line) and barometric pressure (green line), with the Pearson correlation coefficient estimated at R=0.37. This uncommon behavior is most likely due to independent variables directly affecting the correlation between these two variables (CH_4 vs. BP), such as a significant increase in coal production due to coal recovery on more than one panel simultaneously.

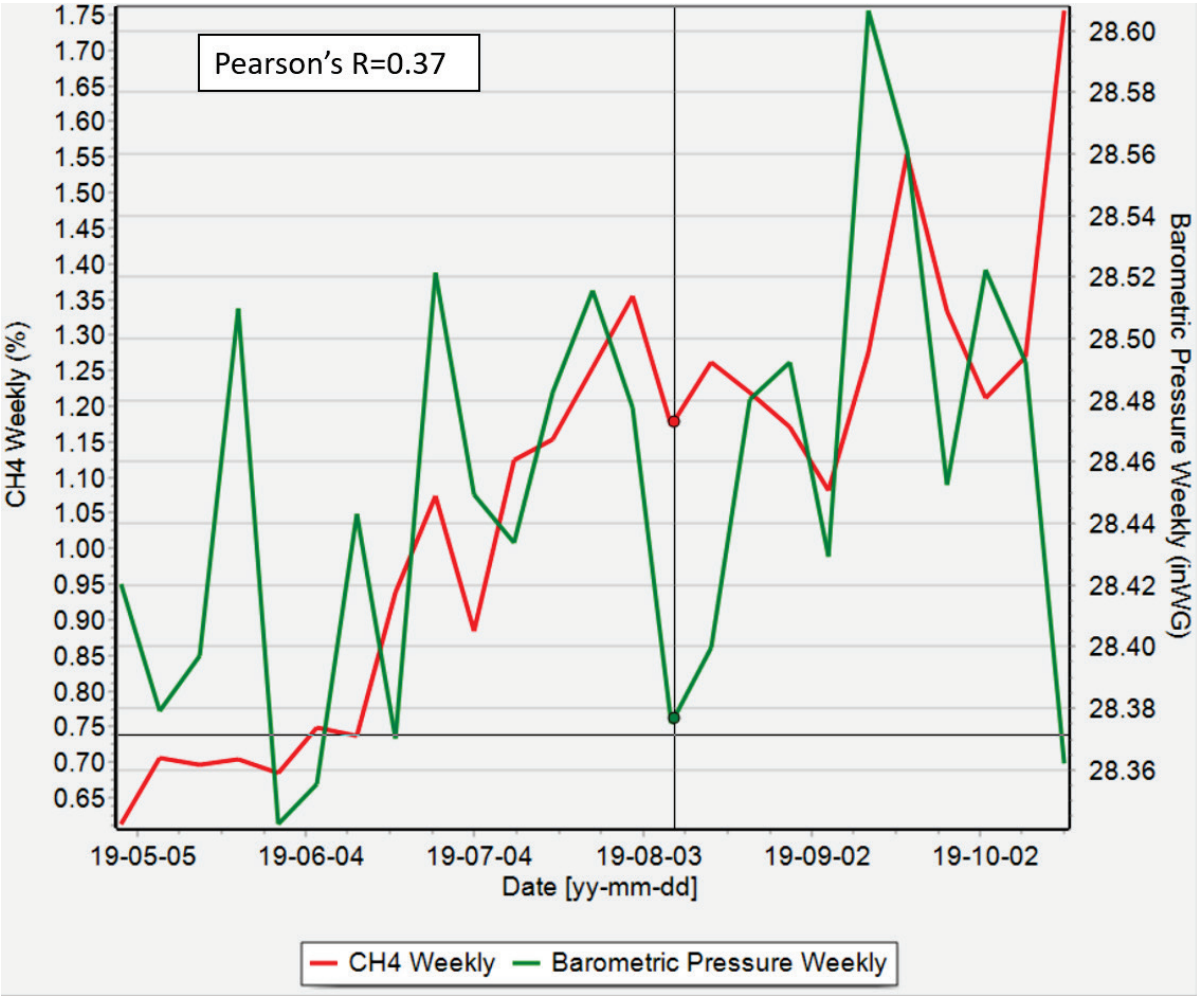


Figure 52: Correlation of CH4 vs. BP for 26 weeks - segment 9 – with a weekly average step

Figure 50 presents two time series, methane gas concentration and coal production 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 correspond to an interval of 30 days. Figure 50 illustrates a strong positive correlation between methane gas and coal production, with the Pearson correlation coefficient calculated at R=0.68. The positive sign denotes that methane concentration increases when coal production increases. This analysis supports the hypothesis that gas concentration is not related to barometric pressure unilaterally. It is more likely that the correlation with barometric pressure is stronger during nearly constant production activity periods.

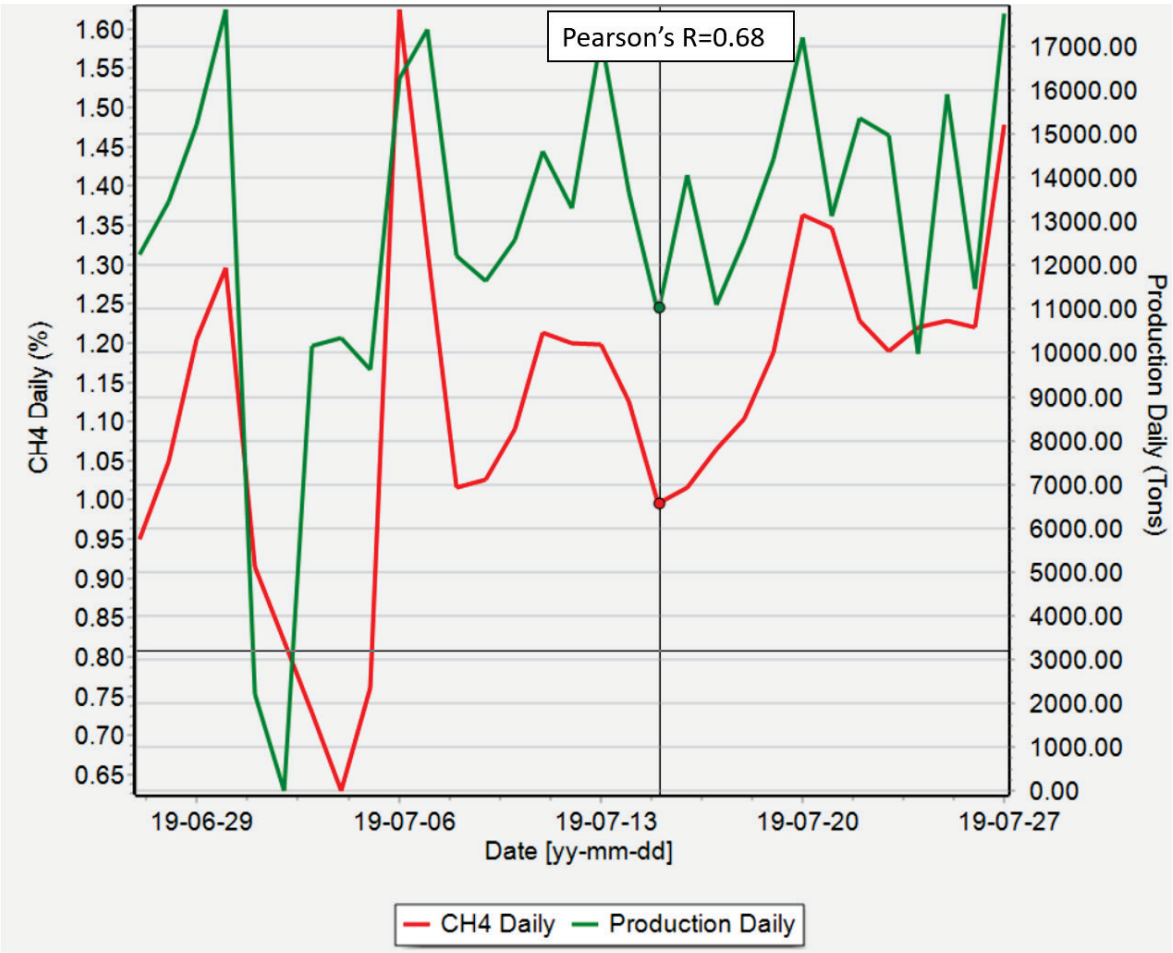


Figure 53: Correlation of CH4 vs. Coal Production for 30 days - segment 1 – with a daily average step

Figure 51 presents two time series, methane gas concentration and coal production 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 correspond to an interval of 60 days. Figure 51 illustrates a strong positive correlation between methane gas and coal production, with the Pearson correlation coefficient calculated at R=0.76. The positive sign denotes that methane concentration increases when coal production increases.

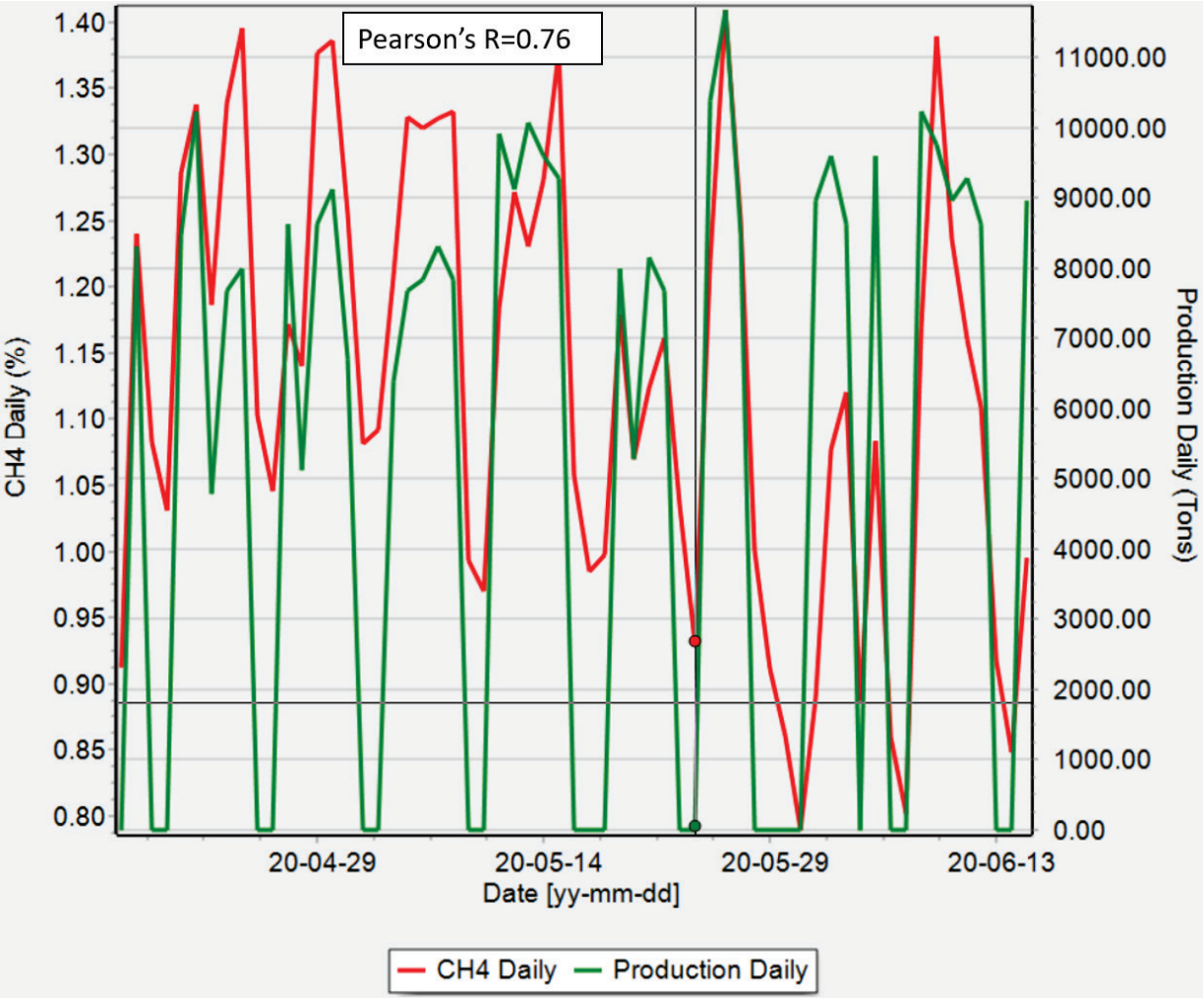


Figure 54: Correlation of CH4 vs. Coal Production for 30 days - segment 2 – with a daily average step

Figure 52 consists of methane gas concentration and coal production time series corresponding to Mine A. The red line represents methane gas concentrations (%), and the green line denotes coal production (tons/day); both time series are sampled using a weekly average and correspond to an interval of 60 days. Figure 52 illustrates a strong positive correlation between methane gas and coal production, with the Pearson correlation coefficient calculated at R=0.87. The positive sign denotes that methane concentration increases when coal production increases.

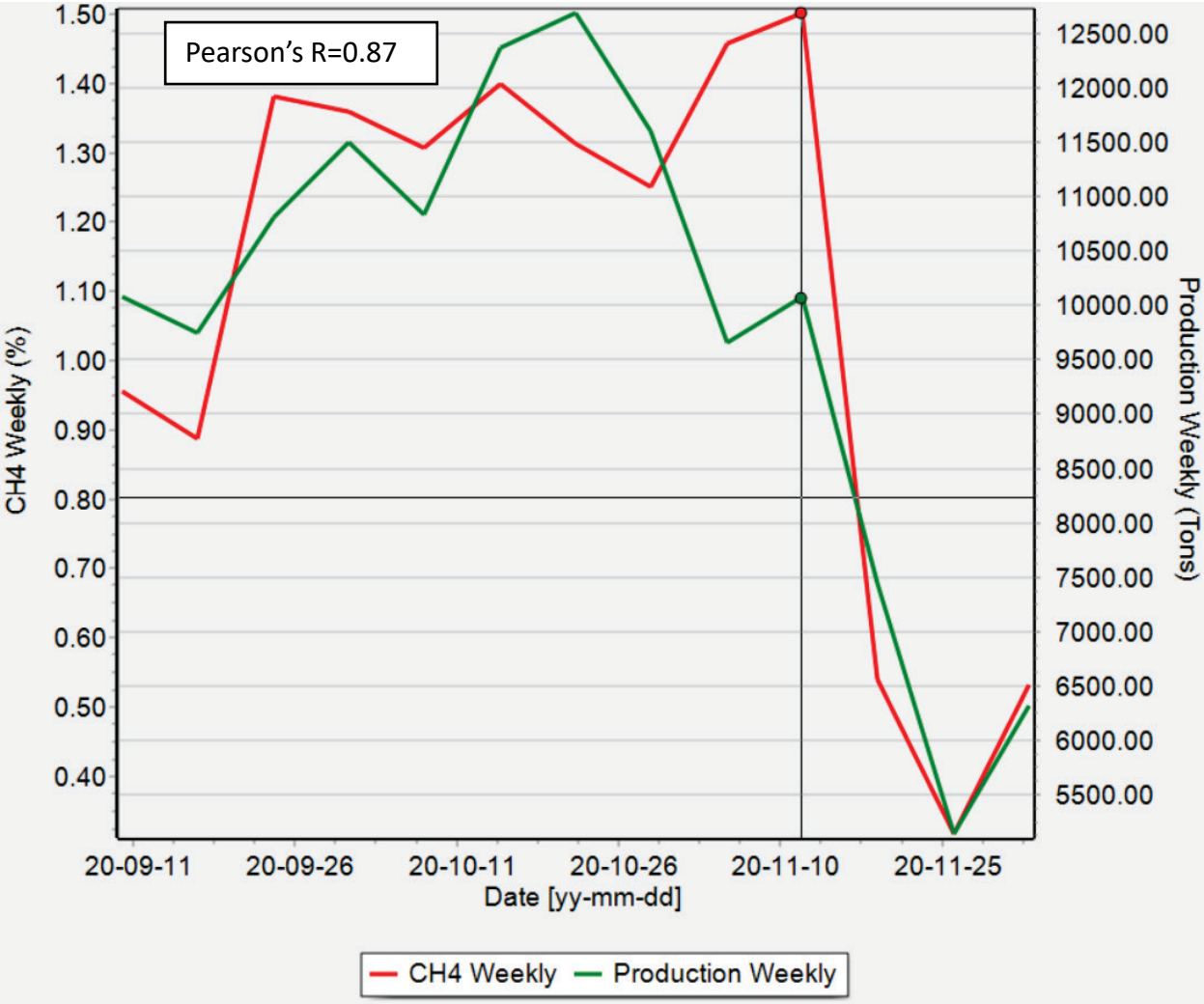


Figure 55: Correlation of CH4 vs. Coal Production for 8 weeks, segment 3 – with a weekly average step

.

This section presents several methane gas time series from Mine A, Mine B and Mine C using different lengths (e.g., one month, six months, one year, etc.) and time steps obtained by implementing AMANDA for data visualization. The main objective of this section is to show the non-stationary nature of the methane gas time series investigated in this research.

Figure 53 presents a methane gas concentration time series from Mine A, which spans an interval of 180 days and is sampled using a daily average time step. Visual inspection shows upward trends and non-repeating large spikes. Consequently, the methane gas time series shown in Figure 53 can be assumed non-stationary. This behavior is typical of the methane gas time series over different time windows analyzed for Mine A. However, specific 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.

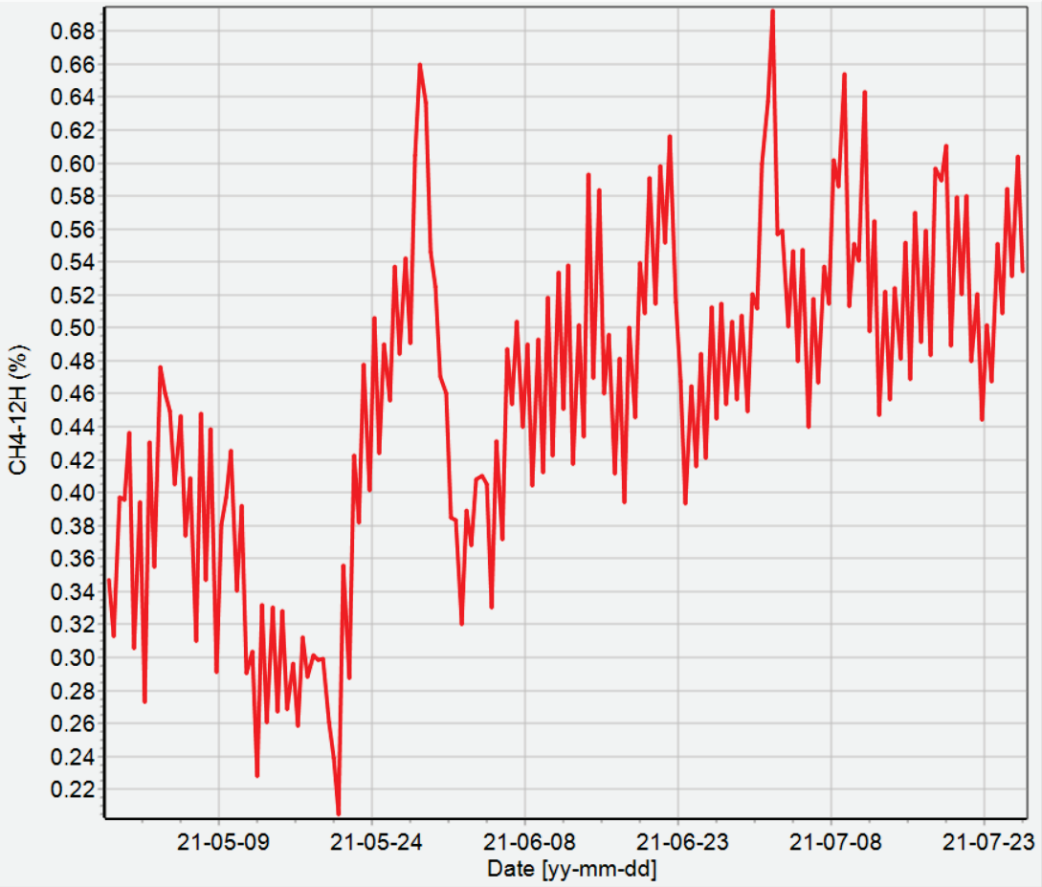


Figure 56. Demonstration of non-stationarity of methane time series - average daily values from Mine A

Figure 54 shows a methane gas concentration time series from Mine A, which spans an interval of 90 days and is sampled using a 12-hour average time step. Visual inspection shows upward trends and non-seasonality behavior. Consequently, the methane gas time series shown in Figure 54 can be assumed non-stationary. As a result, the ACF is not recommended for assessing the autocorrelations of the methane time series.

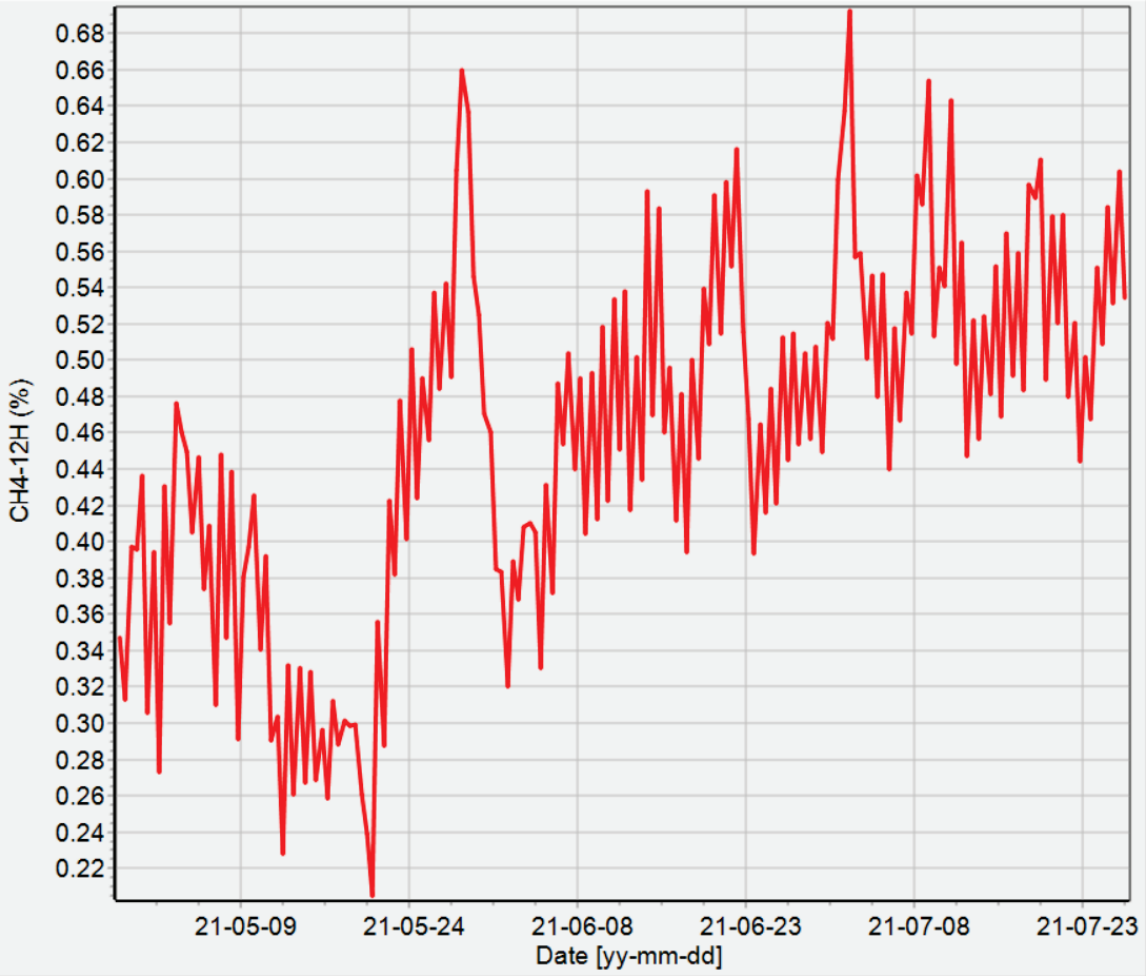


Figure 57. Demonstration of non-stationarity of methane time series - average 12-Hour data values from Mine A

Figure 55 reveals a methane gas concentration time series from Mine B, which spans an interval of 10 days and is sampled using an approximately 10 seconds time step. Visual inspection shows unexpected upward and downward trends and non-seasonality behavior for the methane gas concentration. Therefore, the methane gas time series shown in Figure 55 can be presumed non-stationary. This behavior is typical of the methane gas time series over different time windows analyzed for Peerless mine. As a result, the ACF is not recommended for assessing the autocorrelations of the methane time series. However, specific stationary periods can also be found in the data.

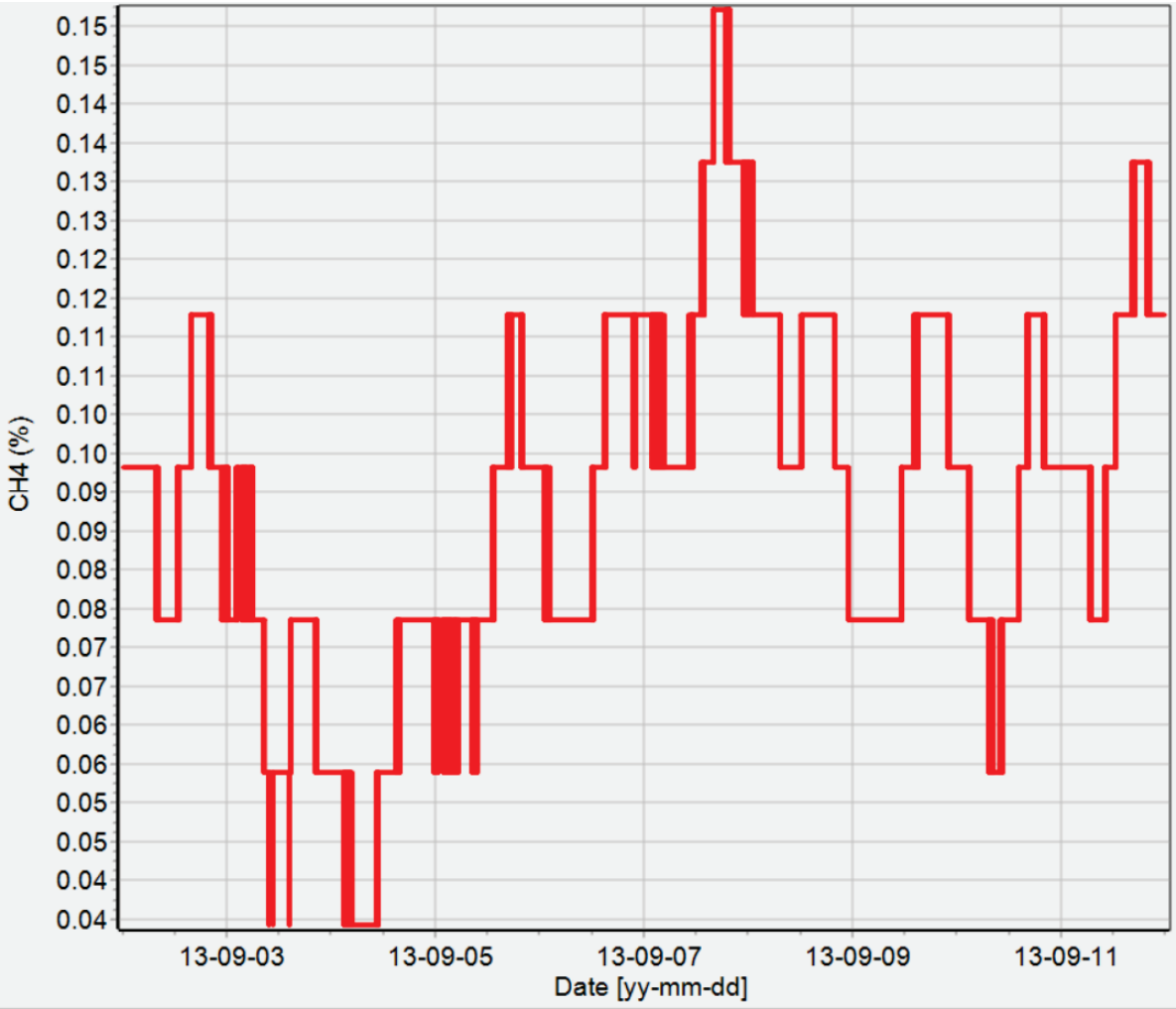


Figure 58. Demonstration of non-stationarity of methane time series – data from Mine B

Figure 56 shows a methane gas concentration time series from Mine C, which spans an interval of 90 days (12 weeks) and is sampled using an average weekly time step. It shows an upward trend and non-seasonality behavior for the methane gas concentration time series. Consequently, the methane gas time series shown in Figure 56 can be assumed non-stationary. This behavior is characteristic of the methane gas time series over different time windows analyzed Mine C. As a result, the ACF is not recommended for assessing the autocorrelations of the methane time series. Nevertheless, particular stationary periods can also be found in the data.

Figure 59. Demonstration of non-stationarity of methane time series – data from Mine C

Figure 57 shows a methane gas concentration time series from Mine C, which spans an interval of 180 days (24 weeks) and is sampled using an average weekly time step. It shows a downward trend and non-seasonality behavior for the methane gas concentration time series. Consequently, the methane gas time series shown in Figure 57 can be assumed non-stationary. As a result, the ACF is not recommended for assessing the autocorrelations of the methane time series.

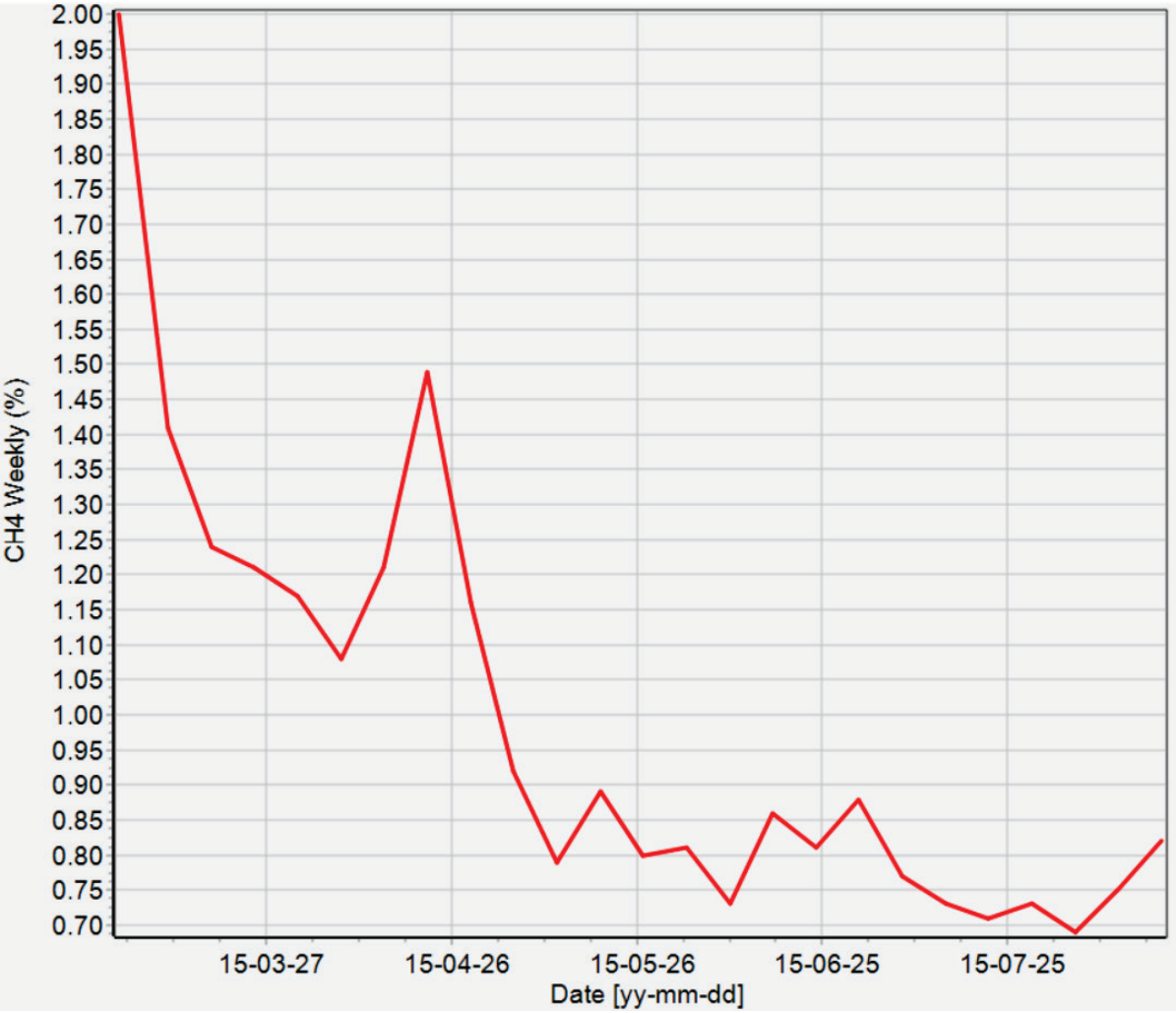


Figure 60. Demonstration of non-stationarity of methane time series – data from Mine C

The following presents the processing of raw data collected over six years from Mine A sensors and barometric pressure from the weather station to assess the potential correlation between methane gas and barometric pressure and coal production implementing a cross-correlation function. The following figures correspond to data segments with different lengths (e.g., one month and six months) and time steps (e.g., 12-hours and daily) obtained employing the MATLAB environment.

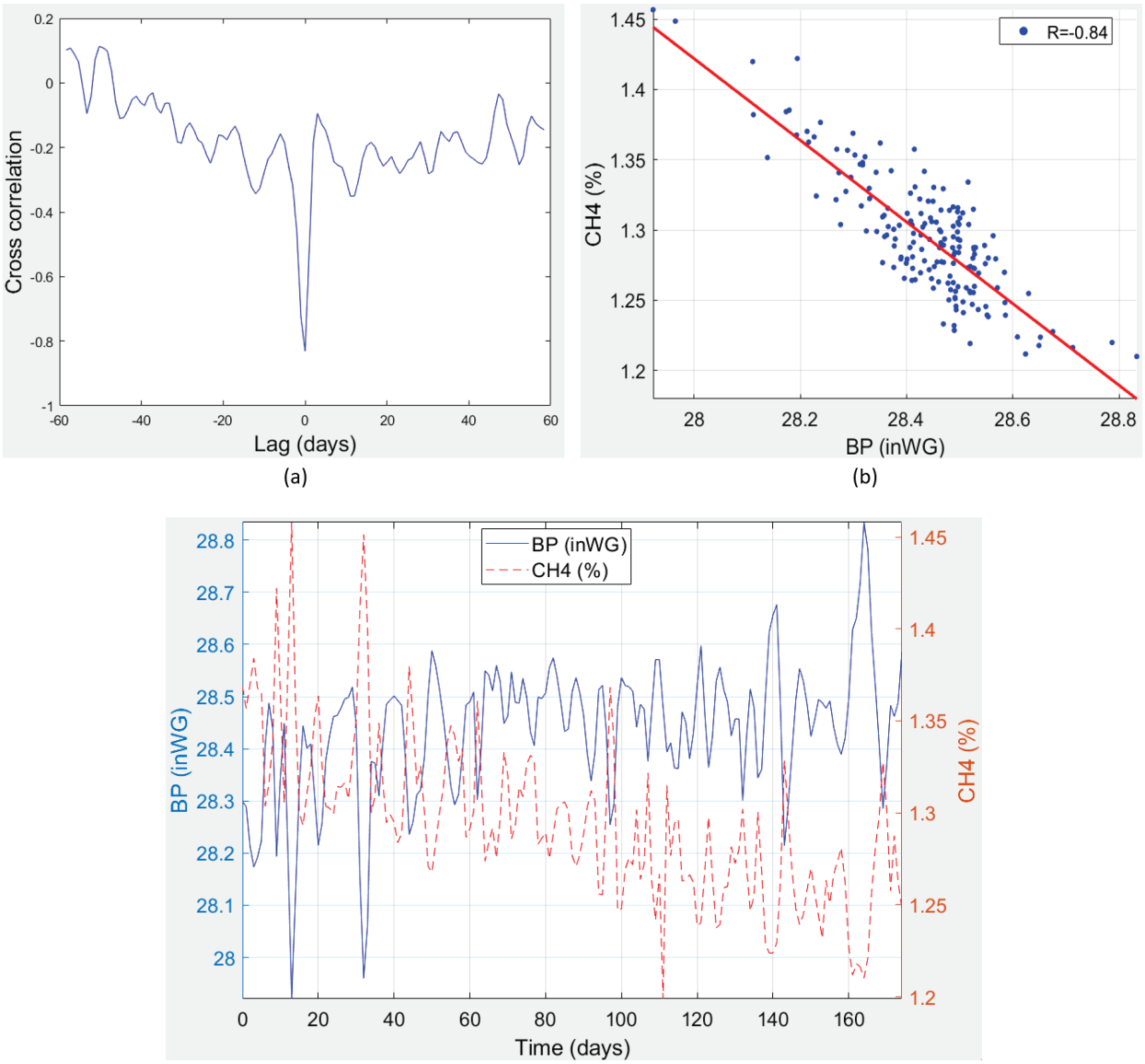


Figure 61: Correlation investigation using the empirical cross-correlation function, (a) Cross-correlation between CH4 and BP, (b) Correlation coefficient (R) between CH4 and BP, (c) methane gas concentration and barometric pressure time series after data cleaning – Segment 1

Figure 58 includes three plots (a, b, and c) corresponding to the Mine A. Figure 58a 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 58b describes a scatter plot using the same data streams shown in Figure 58a. 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. Finally, Figure 58c shows the methane gas concentration ad barometric pressure implemented to perform the cross-correlation analysis. Both datasets were collected during an interval of 180 days and sampled using a daily average time step.

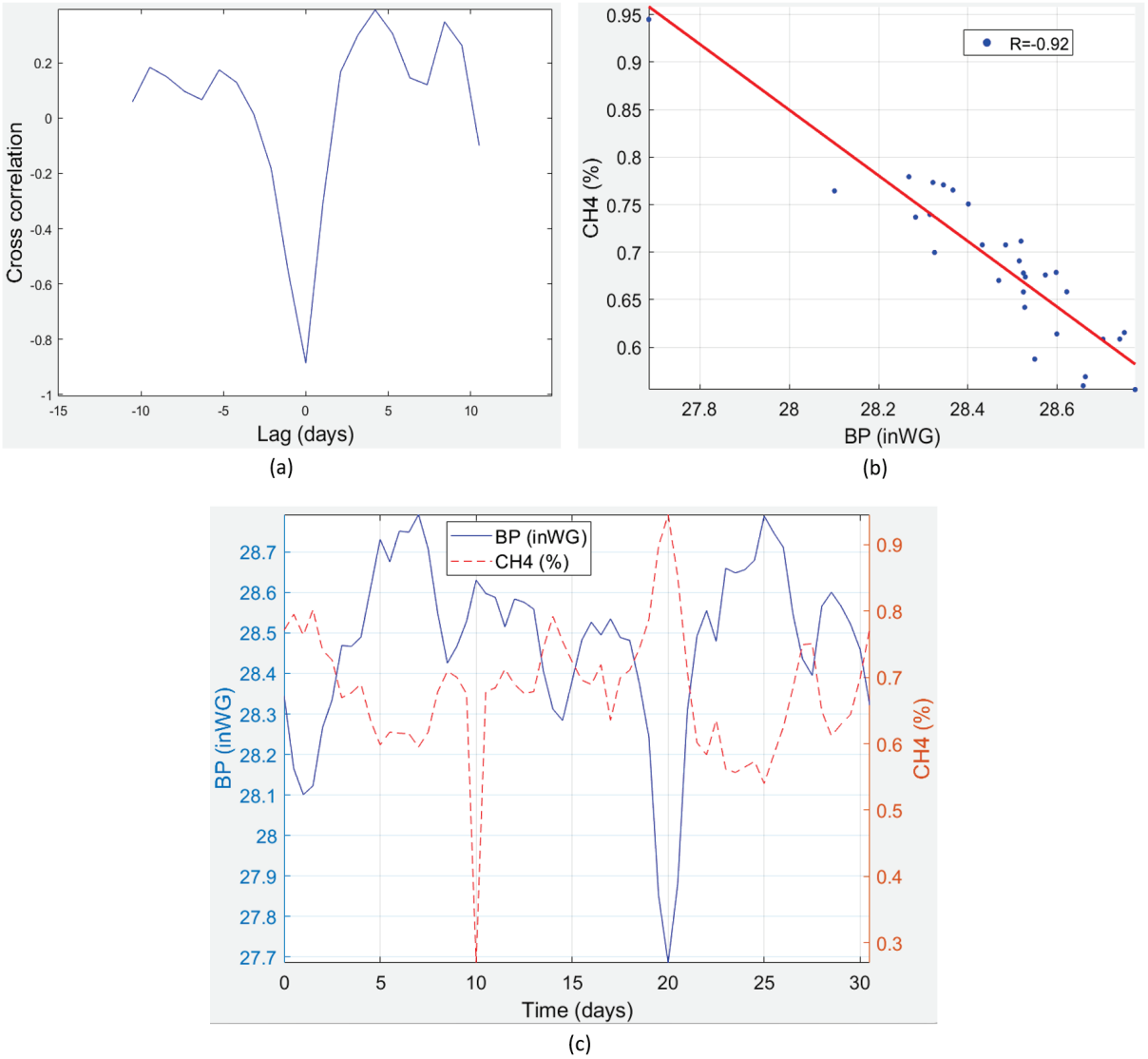


Figure 62: Correlation investigation using the empirical cross-correlation function, (a) Cross-correlation between CH4 and BP, (b) Correlation coefficient (R) between CH4 and BP, (c) methane gas concentration and barometric pressure time series after data cleaning – Segment 2

Figure 59 includes three plots (a, b, and c) corresponding to the Mine A. Figure 59a 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 about −0.9. The negative sign indicates that the variables tend to move in opposite directions. Furthermore, Figure 59b describes a scatter plot using the same data streams shown in Figure 59a. It demonstrates that barometric pressure and methane gas emissions are highly correlated for this specific data segment, with the correlation coefficient calculated at R=−0.92. Finally, Figure 59c shows the methane gas concentration ad barometric pressure implemented to perform the cross-correlation analysis. Both datasets were collected during an interval of 31 days and sampled using a 12-hour average time step.

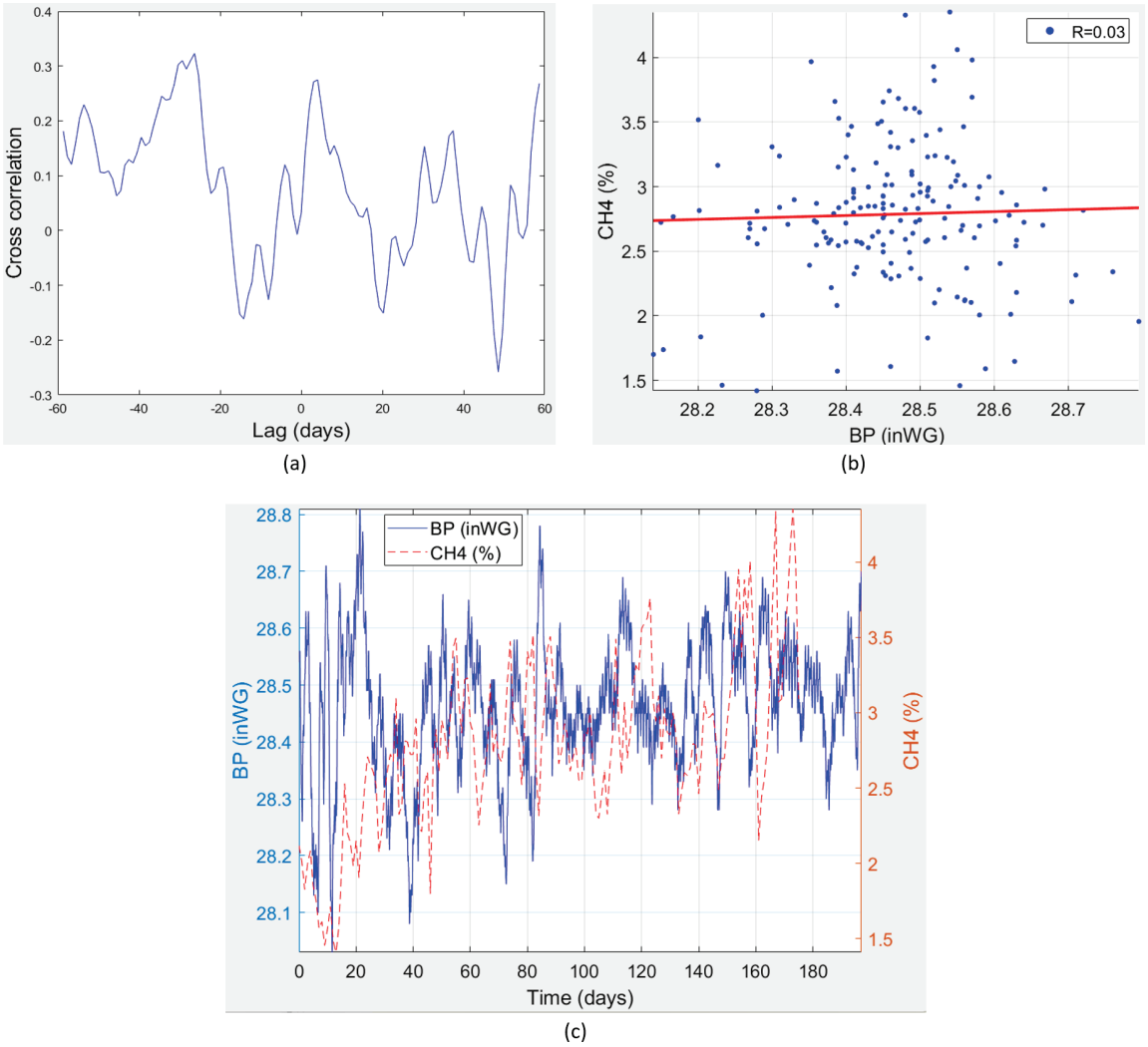


Figure 63: Correlation investigation using the empirical cross-correlation function, (a) Cross-correlation between CH4 and BP, (b) Correlation coefficient (R) between CH4 and BP, (c) methane gas concentration and barometric pressure time series after data cleaning – Segment 3

Figure 60 includes three plots (a, b, and c) corresponding to the Mine A. Figure 60a 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 data segment. The scatterplot in Figure 60b with the correlation coefficient calculated at R=0.03

supports this information. Finally, Figure 60c shows the methane gas concentration ad barometric pressure implemented to determine the cross-correlation function. Both datasets were collected during an interval of 180 days and sampled using a daily average time step.

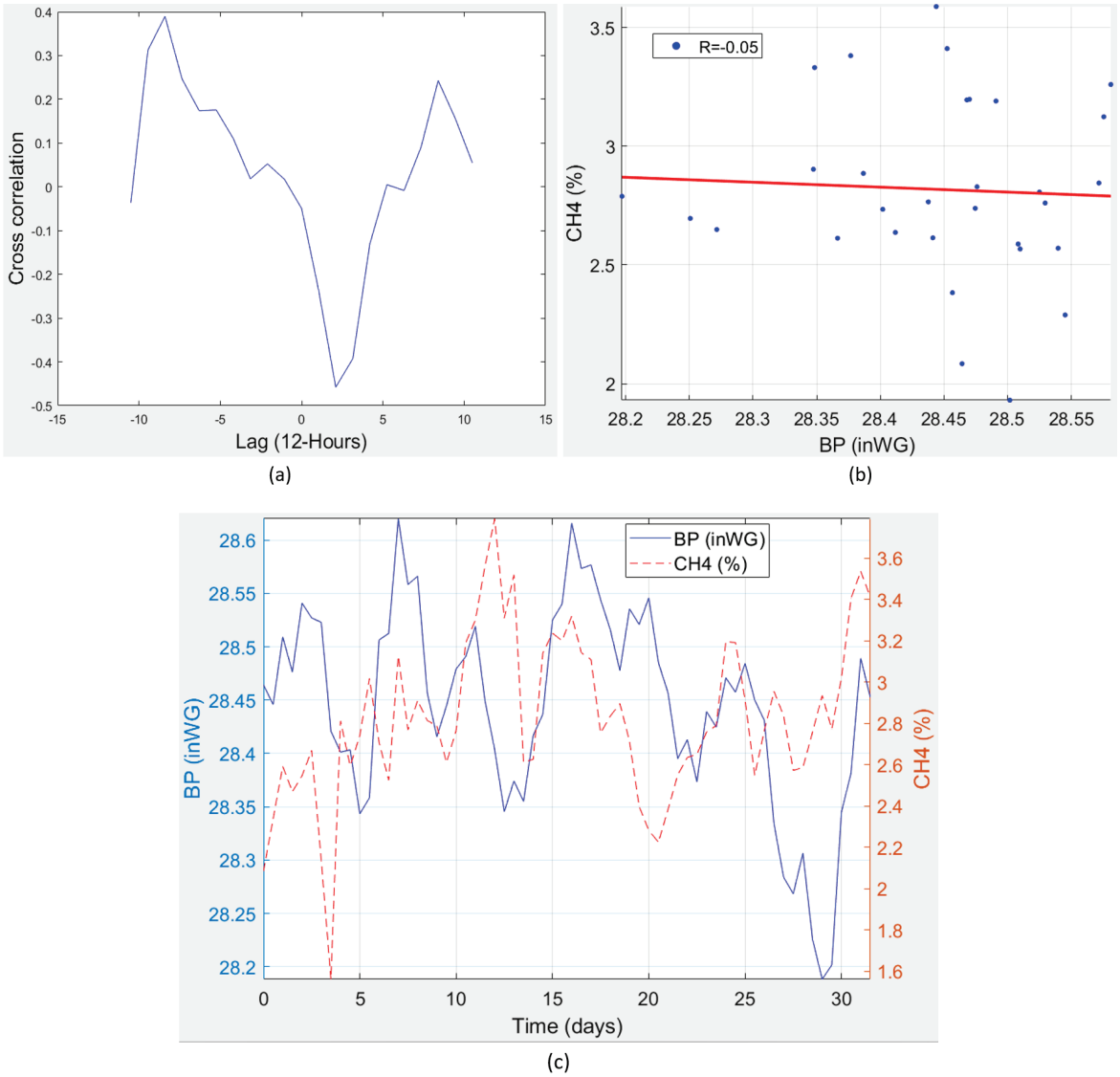


Figure 64: Correlation investigation using the empirical cross-correlation function, (a) Cross-correlation between CH4 and BP, (b) Correlation coefficient (R) between CH4 and BP, (c) methane gas concentration and barometric pressure time series after data cleaning – Segment 4

Figure 61 includes three plots (a, b, and c) corresponding to the Mine A. Figure 61a 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 data segment. The scatterplot in Figure 61b with the correlation coefficient calculated at R=-0.05 strengthens this information. Finally, Figure 61c shows the methane gas concentration ad barometric pressure implemented to determine the cross-correlation function. Both datasets were collected during an interval of 31 days and sampled using a 12-hour average time step.

The following figures present the empirical variograms calculated using the data from Mine A and Mine C as part of the exploratory data analysis, in order to investigate potential autocorrelation trends of the methane gas time series using the variogram function. The graphs were generated by custom code run in the MATLAB environment.

Figure 62 consists of a variogram plot for timeseries data from Mine A. The methane data were collected on a daily average basis. The horizontal axis represents the temporal distance (in days) between pairs of points, and the vertical axis exemplifies the calculated semi-variance of the methane gas concentration at the respective time lag. Figure 62 shows variogram functions that start at zero and rise, albeit at different rates. For example, in Figure 62, the sill is reached after ~60 days. This plot uses a maximum time lag of 100 days and provides evidence of autocorrelations in the methane concentration series; however, the characteristic time is different for each time series, which is not surprising since the production conditions are not the same for the segments investigated.

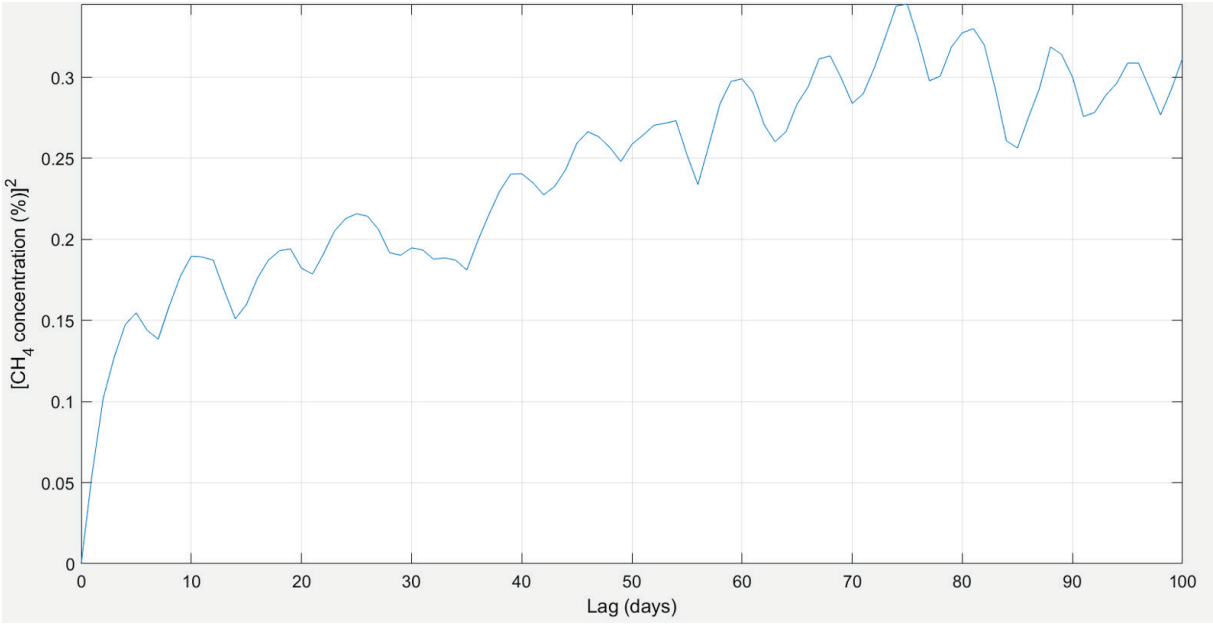


Figure 65. Variogram for methane time series from Mine A – Segment 1

Figure 63 consists of a variogram plot from a different data segment from Mine A. The methane data were collected on a daily average basis. The horizontal axis represents the temporal distance (in days) between pairs of points, and the vertical axis embodies the calculated semi-variance of the methane gas concentration at the respective time lag. Figure 62 shows variogram functions that start at zero and rise, albeit at different rates. For example, the sill is reached after ~60 days. This plot uses a maximum time lag of 100 days and provides evidence of autocorrelations in the methane concentration series.

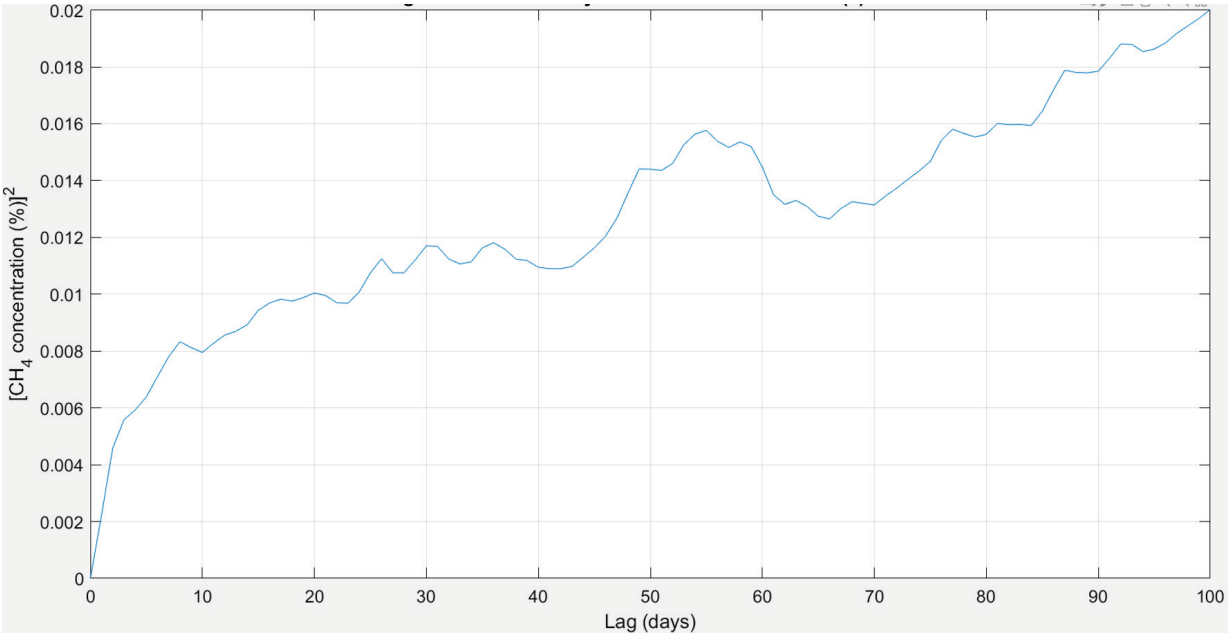


Figure 66. Variogram for methane time series from Mine A- Segment 2

Figure 64 consists of a variogram plot from a different data segment from Mine A. The methane data were collected on a daily average basis. The horizontal axis represents the temporal distance (in days) between pairs of points, and the vertical axis embodies the calculated semi-variance of the methane gas concentration at the respective time lag. Visual inspection shows that the sill is reached after ~5 days. This plot utilizes a maximum time lag of 100 days and offers evidence of autocorrelations in the methane concentration series. Again, however, the characteristic time is different for each time series. This is not surprising since the production conditions differ for every segment examined.

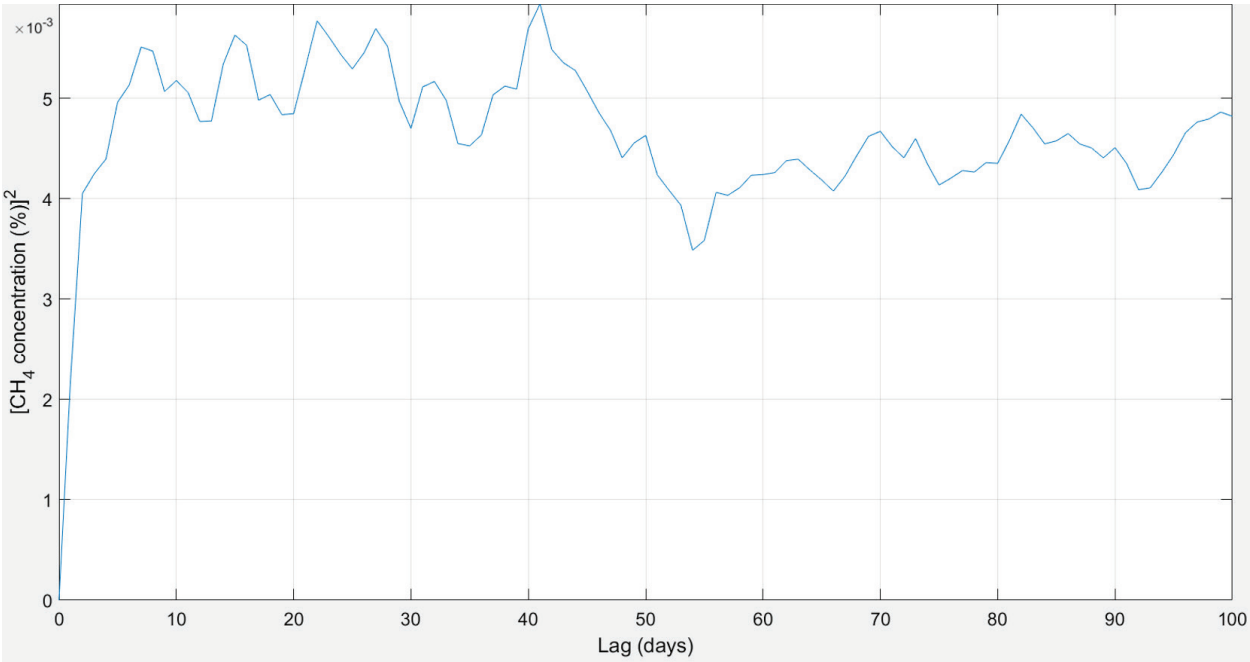


Figure 67. Variogram for methane time series from Mine A - Segment 3

Figure 65 consists of a variogram plot from Mine C. The methane data were collected weekly over a considerably more extended period of 140 weeks. The horizontal axis represents the temporal distance (in weeks) between pairs of points, and the vertical axis embodies the calculated semi-variance of the methane gas concentration at the respective time lag. Figure 65 reveals that the variogram function seems to stabilize around week 10. However, after lag 10, they start increasing, a tendency maintained over the entire range of time lags studied. This behavior is a clear mark of non-stationarity.

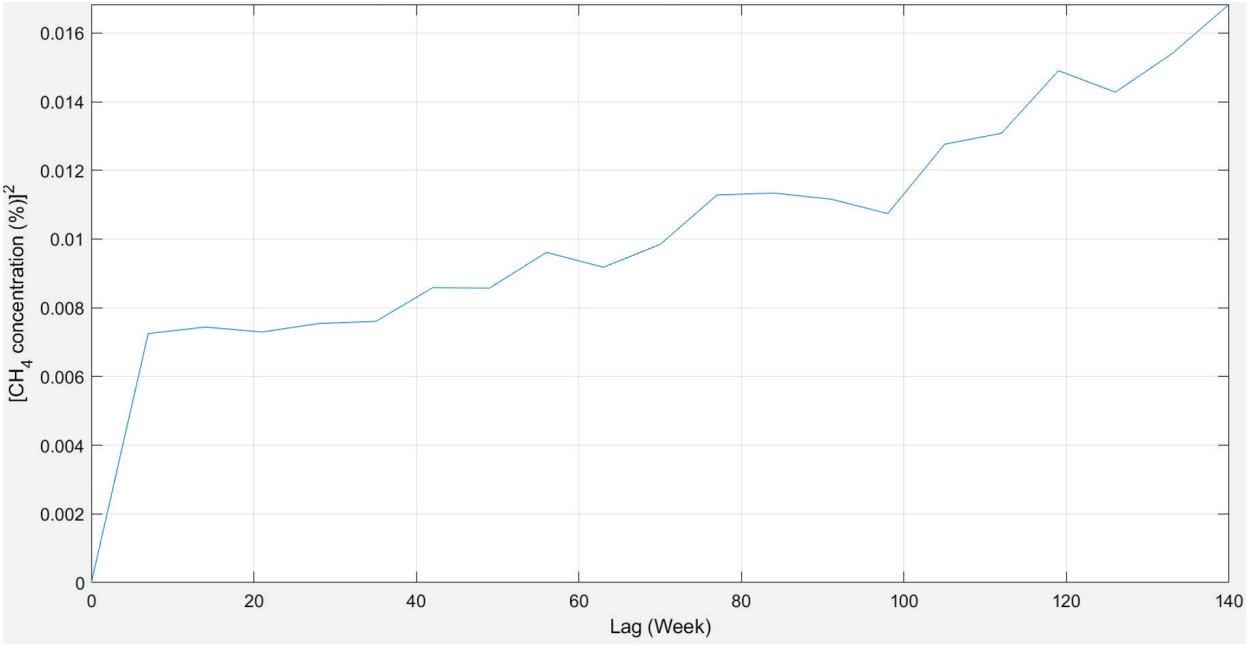


Figure 68. Variogram for methane time series from Mine C - Segment 1

Figure 66 consists of a variogram plot from a different data segment from Mine C. The methane data were collected weekly over a period of 140 weeks. The horizontal axis represents the temporal distance (in weeks) between pairs of points, and the vertical axis embodies the calculated semi-variance of the methane gas concentration at the respective time lag. Figure 66 shows that the variogram function stabilizes around week 10. However, after lag 10, it increases (similarly to Figure 65); this tendency is maintained over the entire range of time lags studied. This behavior is a clear mark of non-stationarity.

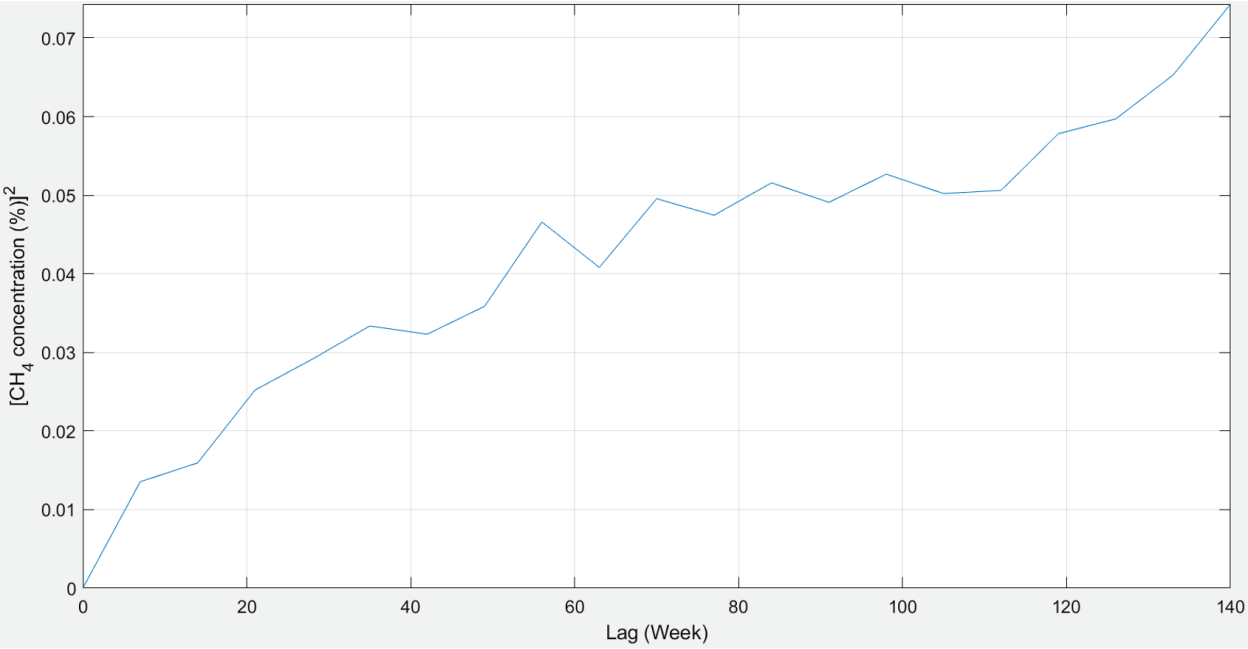


Figure 69. Variogram for methane time series from Mine C - Segment 2

Appendix 3: A review of Coal Mine Methane Prediction Methods

This Appendix discusses the state of the art of coal mine methane forecasting by implementing empirical, numerical, or statistical methods. In particular on the implementation of time series analysis. A brief description of the main finding of each publication and its advantages and disadvantages is presented.

Coal mine methane prediction methods have been a topic of interest for the mining industry and academy for many decades (Airey, 1968; Curl, 1978; Tructin and Wasilewski, 1987; Dixon, 1992; Karacan et al., 2005; Luxbacher et al., 2009). There is significant progress in monitoring and forecasting methane gas emissions in underground coal mining in recent years due to technological advances in different fields. However, the development of reliable methane gas prediction methods is still a challenge due to the multiple variables and sources that influence methane gas emissions into the underground mining environment (Agioutantis et al., 2015; Byungwan and Rana, 2018). Consequently, methane gas calculation and forecasting methods are still limited to information origin, and most of them remain empirical (Booth et al., 2016; Booth et al., 2017).

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 a reliable methane gas prediction method.

Methane gas forecast techniques can be classified into three categories based on the approach employed (Dixon and Longson, 1993; Borowski et al., 2009). The first category is the empirical approach based on data collected by observing a process or phenomenon for making decisions. Depending on the nature of the research, the data employed can be qualitative or quantitative (Patten, 2005). The second category is the numerical approach, which implements a numerical approximation or mathematical tools to solve physical models (Ramasamy, 1994). In this case, numerical methods are used to predict the emission and concentration of methane gas. Finally, the third category is the statistical approach, based on collecting and analyzing raw data using different mathematical techniques to find patterns and build a statistical model for forecasting methane gas emissions and concentration (Brockwell and Davis, 2016).

Empirical methane prediction methods

Empirical methane prediction techniques are designed principally for longwall mining operations and are based mainly on four parameters. The first parameter is the stratigraphy above and below the working seam. The second parameter is the methane releases in the worked coal seam. The zone of gas emissions and concentration in the roof and floor strata is the third parameter. Finally, the last factor is the degree of gas emission from adjacent seams and strata (Curl, 1978).

Airey (1968) conducted research using coal samples of the Deep Soft seam from the Sherwood Colliery, North Nottinghamshire, United Kingdom (UK), for establishing an empirical calculation as illustrated in Equation 1 that quantifies the emissions of gas methane from broken coal during underground coal mining operations and identifies the factors that influence the methane rates release. Equation 1 relates methane gas emissions with coal particle sizes and time. The author determined that the empirical equation allows reliable results when the methane gas flows through a cracked solid. However, imprecise results were obtained when the methane flows through a homogeneous solid.

$$V(t) = A\left[1 - e^{\left(-\left(\frac{t}{t_0}\right)^{\frac{1}{3}}\right)}\right] \quad (1)$$

where *A* = coal sample initial gas content, *t* = time after the start of desorption, and *to* = a time constant.

Curl (1978) presented a review of the available literature about predicting methane gas emissions into underground coal mining. The Technical Information Service of the International Energy Agency (IEA) of Coal Research funded this report. The primary research objective was to discuss the characteristics, advantages, and disadvantages of each empirical methane predicting method implemented in different countries, such as the United Kingdom – MRDE (Mining Research and Development Stablishment), Russia - the Skoczynski Institute, France - CERCHAR (Centre d'Etudes et de Recherches des Charbonnages de France), Belgium - INIEX (Institute National des Industries Extractives), Germany - STVB and WBK (Steinkohlenbergbauverein and Westfalische Berggeverkschaftskasse), Poland and the US. The author concluded that methane prediction techniques are based on three basic approaches. The first approach is modeling the zone of methane emission and the degree of gas release. This methodology is used for Russia, Germany, and France. The second approach, which is implemented in Poland, focuses on predicting the drop in methane gas pressure in the strata bordering mine workings. Finally, the third approach employed in the US and UK is based on mathematical modeling of methane flow. The author also concluded that each method has its advantages and disadvantages, and the selection of the most suitable methane prediction technique depends on each case's geological characteristics.

Dunmore (1982) developed an empirical method known as MRDE for predicting gas emission from longwall mining, as illustrated in Figure 67. This model was implemented by the Mining Research and Development Establishment (MRDE) of the British Government to forecast methane emissions in British underground coal mines. This study was based on Airey's theoretical treatments of gas emission from coal seams (Airey, 1968). The MRDE is focused on the geological characteristics of the coal seams, such as the initial gas content in the coal seam, the coal seam's thickness, the degree of emission expected, and the coal production rate the working face. The MRDE differs from other methods because calculating the degree of gas emission is a time-dependent function obtained by implementing equation 1 developed by Airey (Dixon, 1992; Jensen et al., 1992). The author concluded that the accuracy of the MRDE model is significantly affected by the particular geological conditions of each case.

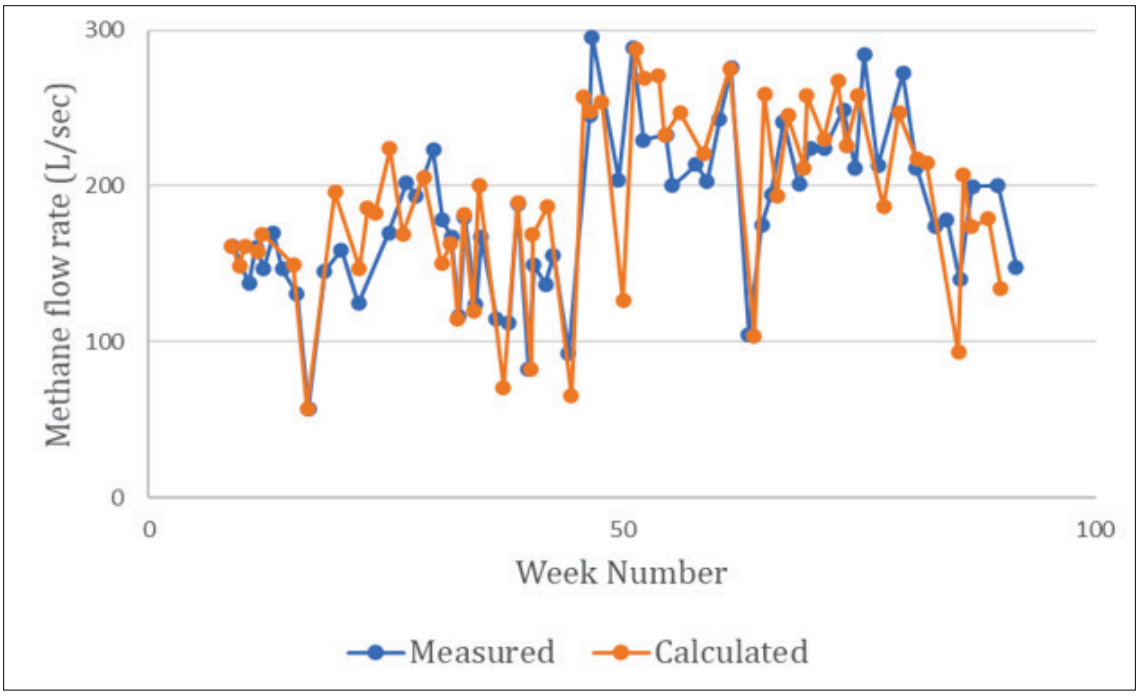


Figure 70: Comparison of prediction with measurement over 93 weeks after Dunmore (1982)

Creedy (1993) recommended an empirical model to predict methane gas emissions from coal-related sources to the atmosphere (Equation 2). This model studies the methane gas 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.

$$E_D = LfP_W + [(1.857) + (D - U)] + RfP_t \qquad (2)$$

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 releases from mines without drainage of 6 m^3/ton, *R* = is the average gas content of coal stored at the surface and is assumed to be 2 m^3/ton, and *f* = is a factor for converting volume flow to mass flow.

Kirchgässner et al. (1993) proposed a method to estimate global methane gas emissions from underground coal mining operations. In this research, the author presented a regression equation (Equation (3)) that satisfactorily predicts methane emissions based on three main variables. The first variable refers to mine emissions. The second variable corresponds to coalbed methane content, and the third parameter is coal production rate. This research identified that the deep, pressure, and moisture are the main essential parameters that directly affect methane gas emissions and concentration.

$$\mathrm{ME} = 1.08 \times 10^{-7}(CP \times MC) + 31.44 - 26.76 \times DV \qquad (3)$$

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) established a methane prediction method for longwall coal mining based on research conducted at two adjacent mines operating in the Pocahontas No.3 coalbed in Virginia. The model was based on emission trends established by continuous monitoring of methane emission rates on existing longwall panels. The author's research principal objective was to forecast the concentration of methane gas in two longwall panels when the panels' width increases from 209m to 305m. This investigation predicted that methane concentration would increase from 8.0 m^3/min to 8.6 m^3/min.

Bustin and Clarkson (1998) attempted to establish a multivariate methane gas emission predictive model based on the properties of the different types of coal, such as rank, physical, chemical, and mineralogical composition. However, the authors determined 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. The authors concluded that there are no reliable methane adsorption capacity trends based on coal rank or composition.

Krog et al. (2006) developed a model for forecasting methane gas emissions from longer longwall faces by analyzing emission contributors in the Pittsburgh Coalbed in southwestern Pennsylvania. This model was based on four parameters. The first parameter was the methane releases from the broken coal by the cutter or shearer. The second factor considered was the methane produced from the coal on the face conveyor. The third component was the methane gas emitted from the coal on the face conveyor. Finally, the last parameter considered was the background gas emanated from the coal face and the adjoining ribs. The authors concluded that transport factors and coal production considerably influence methane gas emissions and concentrations on the longwall face.

Schatzel et al. (2008) proposed a methodology for predicting methane emission rates when longwall face lengths increase. This method is a variation of a technique recommended by the National Institute for Occupational Safety and Health (NIOSH). This technique was first developed by Diamond et al. (1997). This research was conducted in a longwall mine operating in the Pittsburgh Coalbed in southwestern Pennsylvania. Airflow and methane concentration was measured using methane sensors located along the longwall face. The following assumptions were made to perform this study. First, the authors assumed that the frequency of delays and the mine advance rate were equivalent. Second, methane gas emission in the longwall was presumed to be continuous in each segment. Third, all potential sources of methane gas emissions and concentration were supposed to change at a constant rate to increasing face length. The authors concluded that this methodology could only be applied at the Pittsburgh Coalbed.

Numerical methane prediction methods

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

Owili-Eger et al. (1973) developed a mathematical model able to forecast methane gas flow patterns and emissions in the coal seams and into the mining atmosphere based on the study of the physics of gas flow through a coal seam by using a computer program to solve the mathematical model proposed The numerical technique proposed by the authors exemplifies a modified gas diffusion system for flow through porous media. In this research, the following assumptions were made. First, the flow of gases is assumed to be constant. The second assumption made was that temperature change over the medium is relatively small. It is also assumed that K_x and K_y, the directional permeabilities depend on the pressure and position coordinate only. Finally, the last assumption was that flow would occur along the seam except when a production and injection well term is involved. The applications of this mathematical model were demonstrated successfully. However, further research proved that this model only was able to work reliably for shallow mines (Dixon, 1992).

Sung et al. (1987) presented a two-dimensional finite-difference numerical model to predict the methane emission and concentration rates into active underground coal mine working. The authors implemented Darcy's law to describe the methane gas flow on the coal macropore structure, while Fick's law was used to explain the methane flow in the coal seams' micropore structure. In this investigation, some coal seam properties such as porosity, permeability, thickness, and sorption characteristics were identified and separated to study and understand each parameter's influence on the methane gas emissions rates. This research discovered that

coal seams' permeability is the most leading parameter influencing the mine face's methane gas inflow rate.

Ediz and Edwards (1991) established a numerical method to estimate methane emission from source beds to the underground coal mine workings. This research focused on solving the time-dependent gas flow equation (Equation (4) concerning a medium having variable anisotropic permeability using finite element analysis to give time-dependent gas pressures. The researchers implemented the finite element package PAFEC'75, which is employed to simplify gas pressure distributions, calculate gas flow for a given boundary, and simulate boreholes. This research concluded that the model results are very similar to those achieved from physical considerations. Finally, the authors concluded that coal seams' permeability regulates the methane gas flow through strata.

$$\frac{\partial p}{\partial t} = \frac{1}{2\mu\emptyset}\left[\frac{\partial}{\partial x_1}\left(K_1\frac{\partial p^2}{\partial x_1}\right) + \frac{\partial}{\partial x_2}\left(K_2\frac{\partial p^2}{\partial x_2}\right) + \frac{\partial}{\partial x_3}\left(K_3\frac{\partial p^2}{\partial x_3}\right)\right] \qquad (4)$$

where p = pressure, t= time, μ = gas viscosity, K_i = permeability in the i direction, x = space coordinate, and Ø = porosity of the material.

Tauziede and Pokryszka (1993) proposed a dynamic mathematical model to forecast methane gas emissions in longwall mines on a daily or weekly basis. This model was based on two main parameters. The first parameter was the strata surrounding the mined seam, and the second factor was the methane gas emissions as a function of coal stratigraphy. The authors stated that the total methane emission during a given period on a given day could be expressed using equation (5). Finally, it was concluded that it is challenging to design a reliable methodology for methane gas prediction due to the number of parameters and variables involved in each particular case.

$$G(j) = \frac{D_s}{100}\left[P_{MS}\,A(j) + \sum_{i=1}^{R} P_R(i) \sum_{k=1}^{j} A(k) \int_{j-k}^{j-k+1} f_i^R(t)dt + \sum_{i=1}^{F} P_F(i) \sum_{k=1}^{j} A(k) \int_{j-k}^{j-k+1} f_i^F(t)dt\right] \qquad (5)$$

where D_s = specific emission calculated in advance (m^3/m), P_{MS} = share of specific emission generated by the mined seam (%). P_R (i) and P_F (i) = share of specific emission generated by bed number i at the roof and the floor, respectively. f_i^R and f_i^F = gas release rate for that part of the bed being considered, at the roof and the floor, respectively. R = is the number of beds on the roof, F = number of beds on the floor, A(j) and A(k) = faces advances on days j and k, respectively.

Karacan et al. (2005) explored an advanced numerical model that can simulate the rock mass and the gas flow responses to predict methane gas flow and emissions when the longwall face width

increases. This research was divided into two main phases. The first phase involved using Fast Lagrangian Analysis of Continua 2D (FLAC 2D) to simulate the rock's geomechanical characteristics, such as permeability, fractures, stress, and strain. 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.

Guo et al. (2008) recommended a three-dimensional numerical model to forecast mine gas emission, rock mass deformation, water inflow, and mine stability in underground longwall coal mining. This numerical model used a combined 3D mechanical deformation and dueled porosity multiphase flow finite element code-named COSFLOW, developed at CSIRO. The authors stated that COSFLOW has unique characteristics because it incorporates Cosserat continuum theory, which allows obtaining a compelling description of mechanical deformation in weak layered rock and stress changes.

Luxbacher et al. (2009) proposed a model that simulates the methane gas flow and atmospheric ventilation needs in an underground continuous miner section. This research is focused on the effects of permeability and porosity variations of the coal seams on methane emissions and concentration. Coalbed thickness and pressure, permeability, effective porosity, water saturation, sorption time were some of the coalbed's initial properties selected to run the model. The authors concluded that the highest amount of methane gas emissions occurred when the reservoir's porosity and permeability present a gradual change due to induced stress.

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 (Brillinger, 2001; Shumway and Stoffer, 2006; Brockwell and Davis, 2016). 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 literature review of time series analysis for forecasting and controlling atmospheric monitoring systems, methane concentration, and emissions in underground coal mines are presented below.

Kaffanke (1980, cited in Dixon, 1992) developed a methodology for predicting concentration and emissions of methane gas for a period between one day to one month in length by using discrete multiple linear regression. The most relevant variables included in the model were: daily output (tons/day), accumulated output (tons), previous day methane gas flow (m^3/s), previous Sunday methane flow (m^3/s), number of no working days, and desorbed gas content (m^3/t). Methane emissions and concentration forecast were done by implementing seven equations, one for each day of the week. Figure 68 presents an example of the results obtained in this research. The author made two important conclusions: first that reliable methane emission prediction can be achieved by implementing statistical methods, in this particular case, discrete multiple linear

regression. The author recommends further research in implementing different statistical methods for methane prediction, such as linear filter theory, spectral analysis, and time series analysis.

Tructin and Wasilewski (1987) established an approach for modeling airflow in underground coal mine workings by studying atmospheric data implementing time series analysis and digital filtering techniques. In this particular case, the low-pass first-order recursive filter was implemented. This research's primary objective was to identify and separate signals of different amplitudes and durations that disturb monitoring ventilation systems. 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 high frequency above $2.77x10^{-4}$ Hz, disturbances with a middle frequency between $1.15x10^{-5}$ Hz and $2.77x10^{-4}$, and the last group is disturbances with a low frequency below $1.15x10^{-5}$ Hz.

Dixon (1992) described a model for predicting methane concentration and emissions 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 research implemented the AutoRegressive Integrated Moving Average models, also known as the ARIMA model, for describing stationary and non-stationary time series. This technique's implementation identified the relationship between methane concentration and its explanatory variables such as coal production, barometric pressure, and airflow velocity. The author highlighted that this model could be constructed without any previous knowledge of the series itself. This model is built from a consideration of past values. Finally, the author recommended time series analysis for mining process control and forecasting methane concentration and underground coal mining emissions. Figure 69 illustrates hourly average methane concentration forecasts.

Tauziede and Pokryszka (1993) presented a dynamic statistical methodology for predicting methane gas on a daily or weekly basis. The researchers' study was based on data analysis from 14 longwall working mines in the Lorraine Basin in France. The authors stated that methane emissions depend on two significant factors: the rate advance in the current week face and the rate of advance on previous weeks. This idea was studied and verified by implementing linear regression. It was found that methane emission can be estimated according to the face's advance being considered by applying equation (6). Some of the results obtained implementing this stochastic model are illustrated in Figure 70. It was concluded that the statistical method used showed remarkable results, but the face's starting period was not represented correctly.

$$D_n = D_s[306\,A_n + 150\,A_{n-1} + 75\,A_{n-2} + 5{,}470] \qquad (6)$$

where Dn = expected volume of methane for week n (m^3), Ds = specific emission of methane (m^3) per meters of advance, An = planned advance per week n (m), An-1 and An-2 = real advances for weeks n-1 and n-2 (m).

Dixon and Longson (1993) established a statistical method for short-term prediction of the methane gas concentration in longwall coal mines based on time series analysis developed using

data obtained from atmospheric monitoring systems. The model takes into account the most important sources of methane gas, such as the desorbable gas content of the working coal seam, the stratigraphy above and below the working seam, and the degree of methane emissions from the adjacent seams and strata. Figure 71 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.

Tominaga and Bandopadhyay (2002) introduced a model for predicting spontaneous combustion in underground coal mining based on measured time series data. Fick's second law of diffusion was implemented to identify the characteristics of time-series data, such as the concentration-time curve. The main objective of this research was to precisely predict the concentration and ubication of potential carbon monoxide sources (CO) in a longwall coal mine located in Hokkaido, Japan.

Zagorecki (2015) developed 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.

Badura et al. (2020) proposed a short-term method to forecast methane concentration and emissions in longwall coal mining employing available atmospheric data based on time series analysis. This research's main objective was to develop a one-day forecast of methane gas concentration at the sensor location up to 10 m in front of the longwall face and at the longwall outlet. The authors concluded that more reliable results were obtained at the airway of the longwall face than at the longwall outlet. Some of the results of this research are illustrated in Figure 72

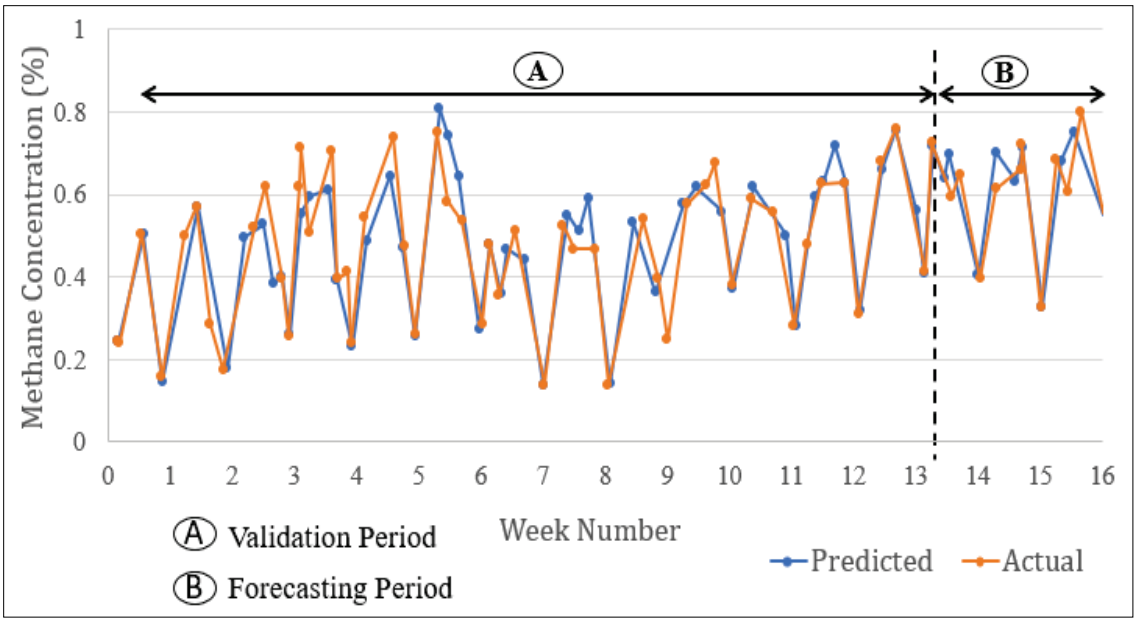


Figure 71: Methane forecasting after Kaffanke (1980, cited in Dixon, 1992)

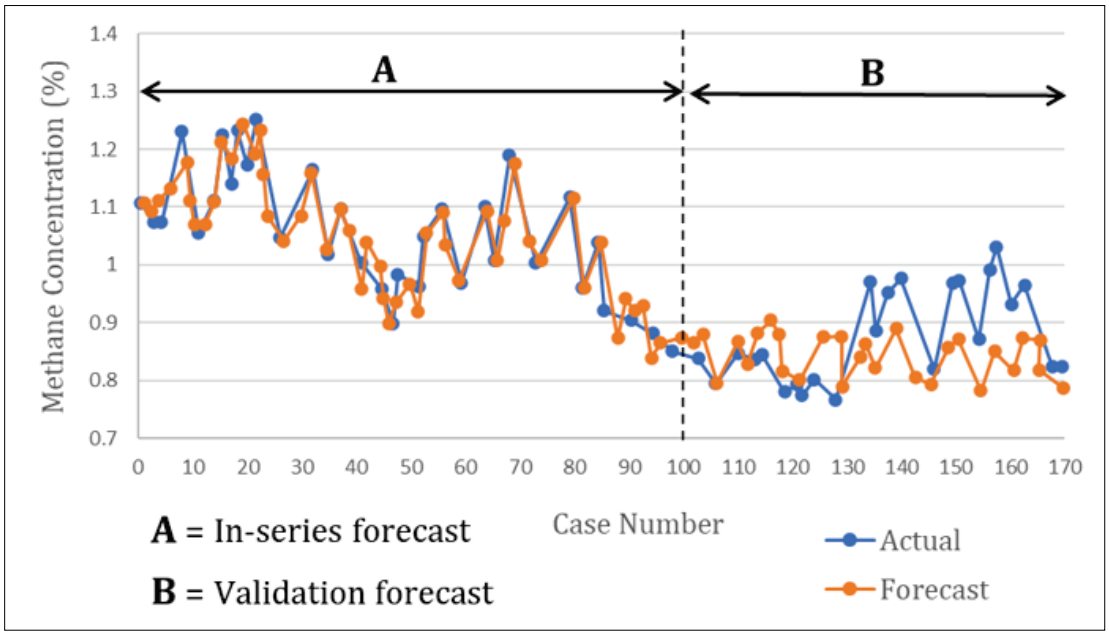


Figure 72: Hourly average methane concentration forecasts After Dixon (1992)

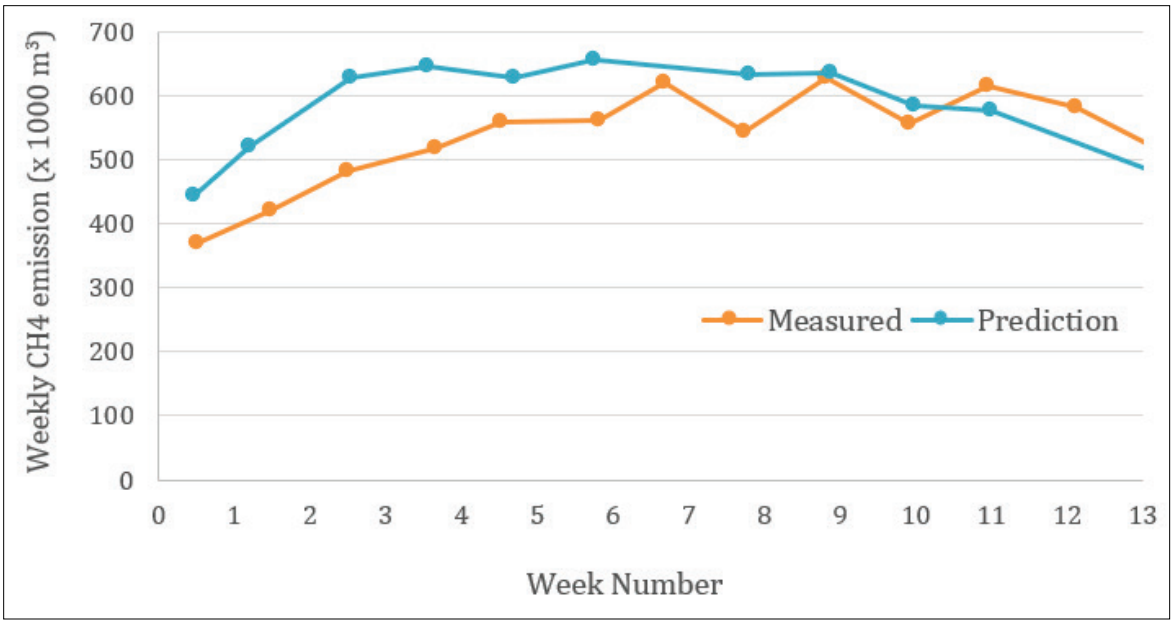


Figure 73: Methane gas forecasting after Tauziede and Pokryszka (1993)

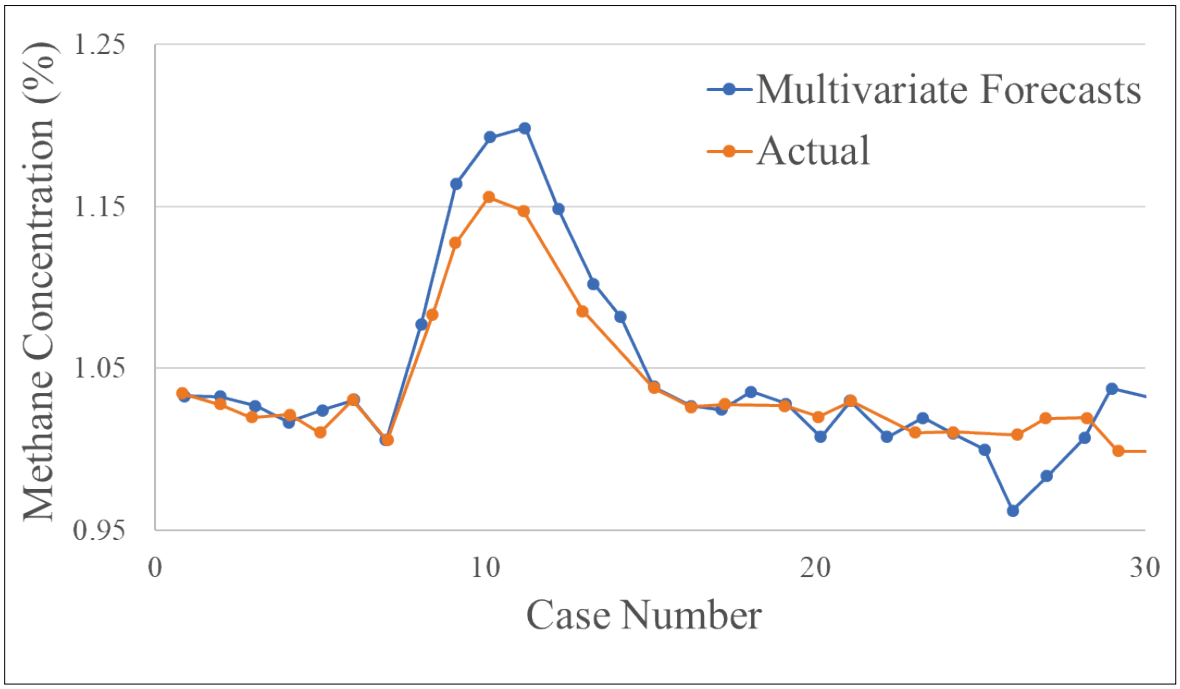


Figure 74: Methane multivariate forecasting after Dixon and Longson (1993)

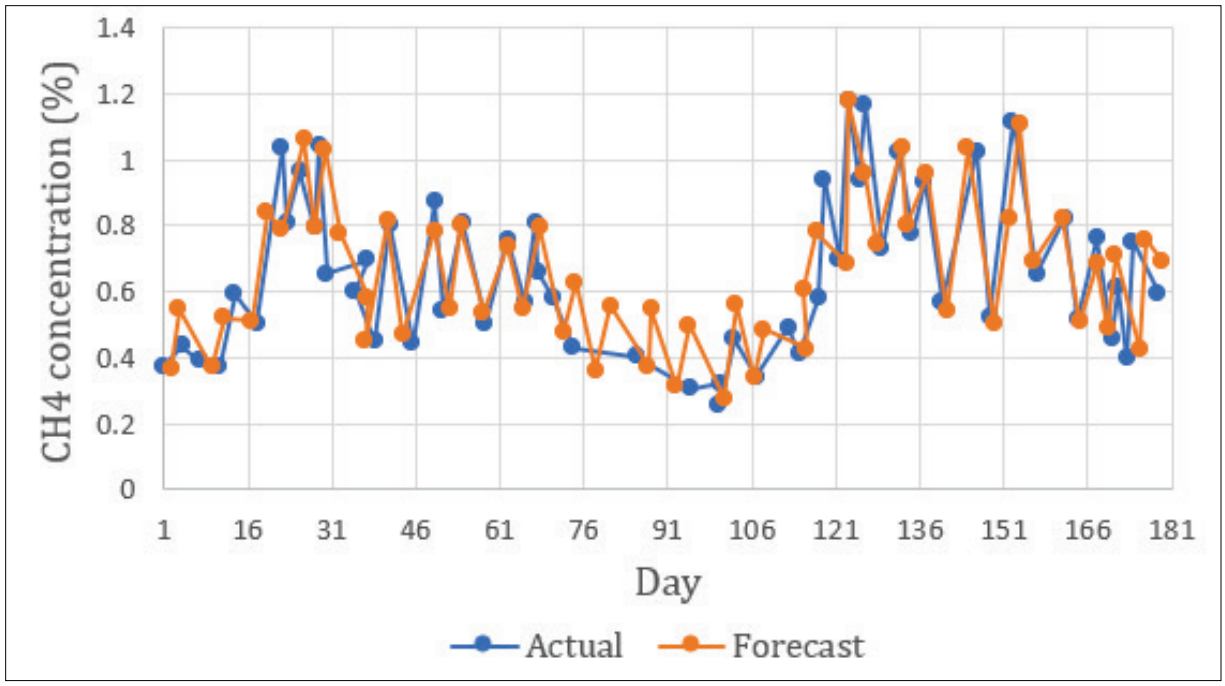


Figure 75: Methane prediction at the airway up to 10 m in front of the longwall after Badura et al. (2020)

Appendix 4: The AMANDA Software Package

The following figures highlight some of the capabilities of the AMANDA system. Although the initial version of AMANDA was developed several years ago, it has been updated to be able to handle the data for this project. During this project, the AMANDA program has been updated multiple times.

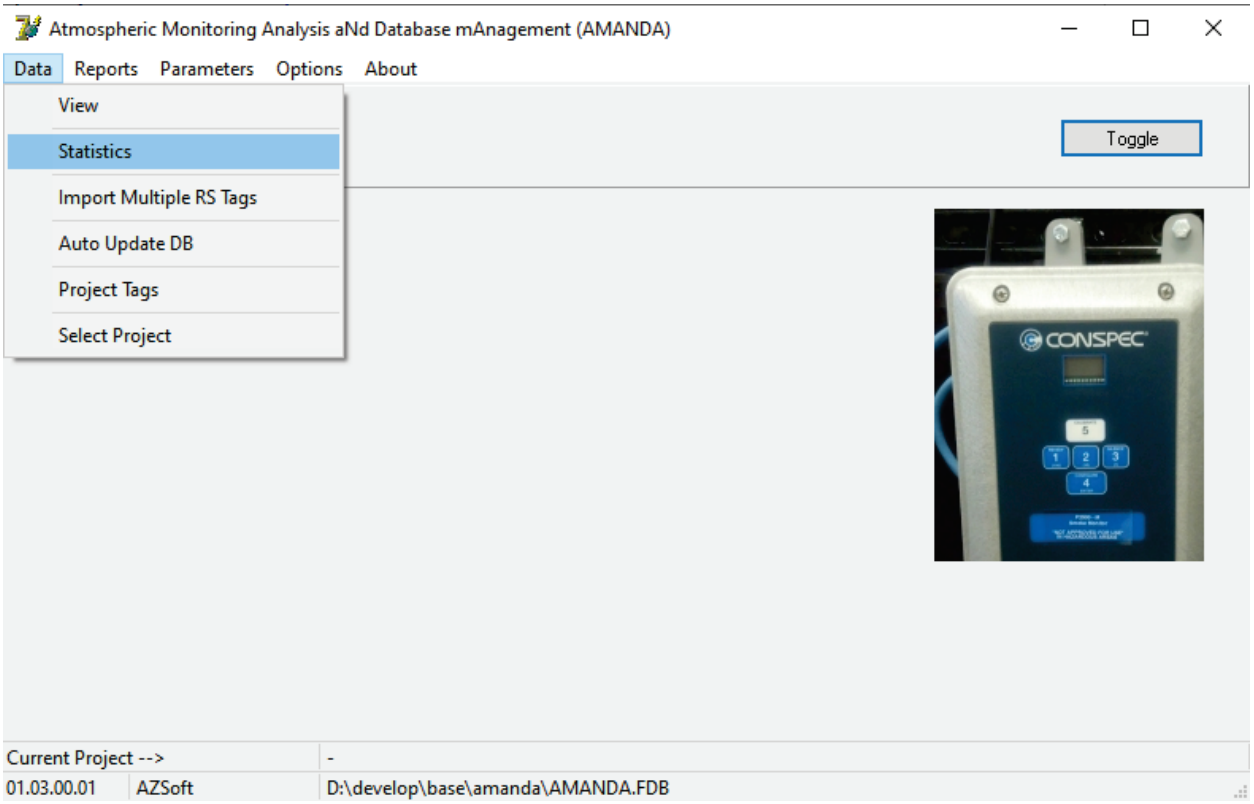


Figure 76: AMANDA main menu

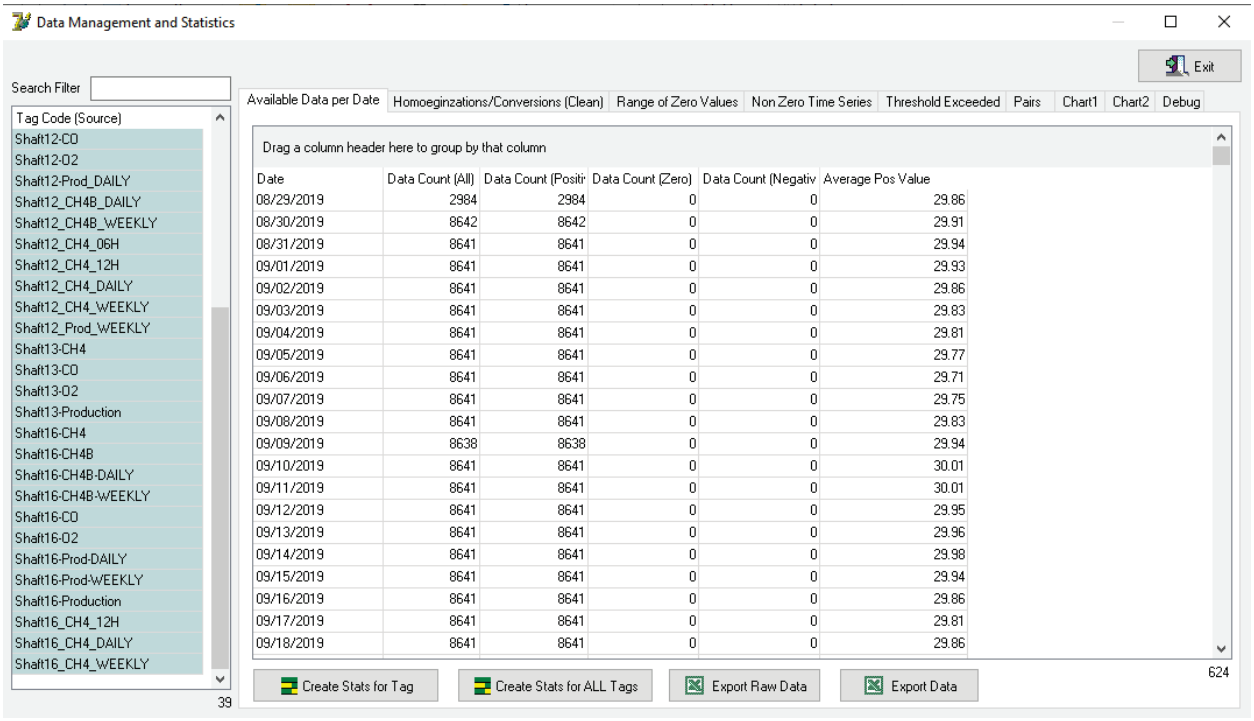


Figure 77: Data statistics

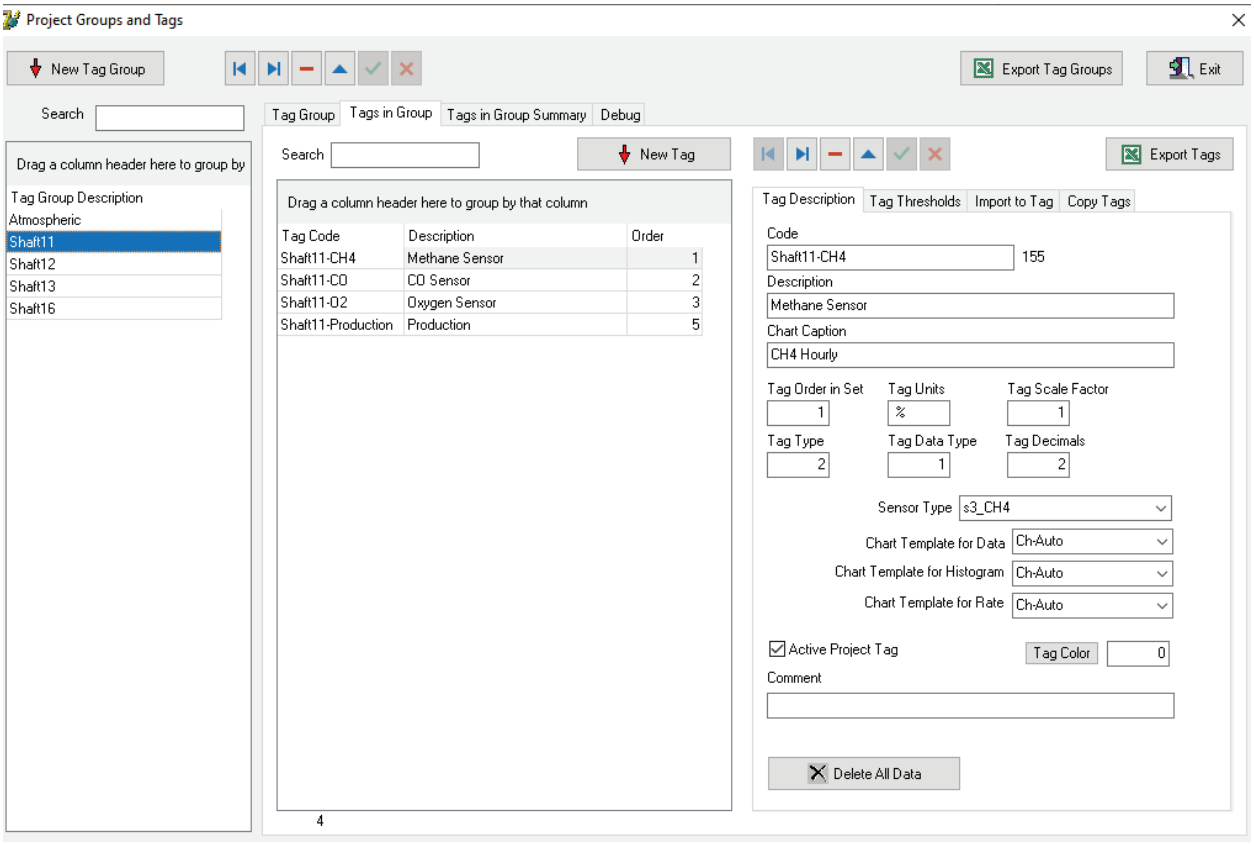


Figure 78: Tag definition per location as well as parameters per tag (tag grouping for Shaft 11)

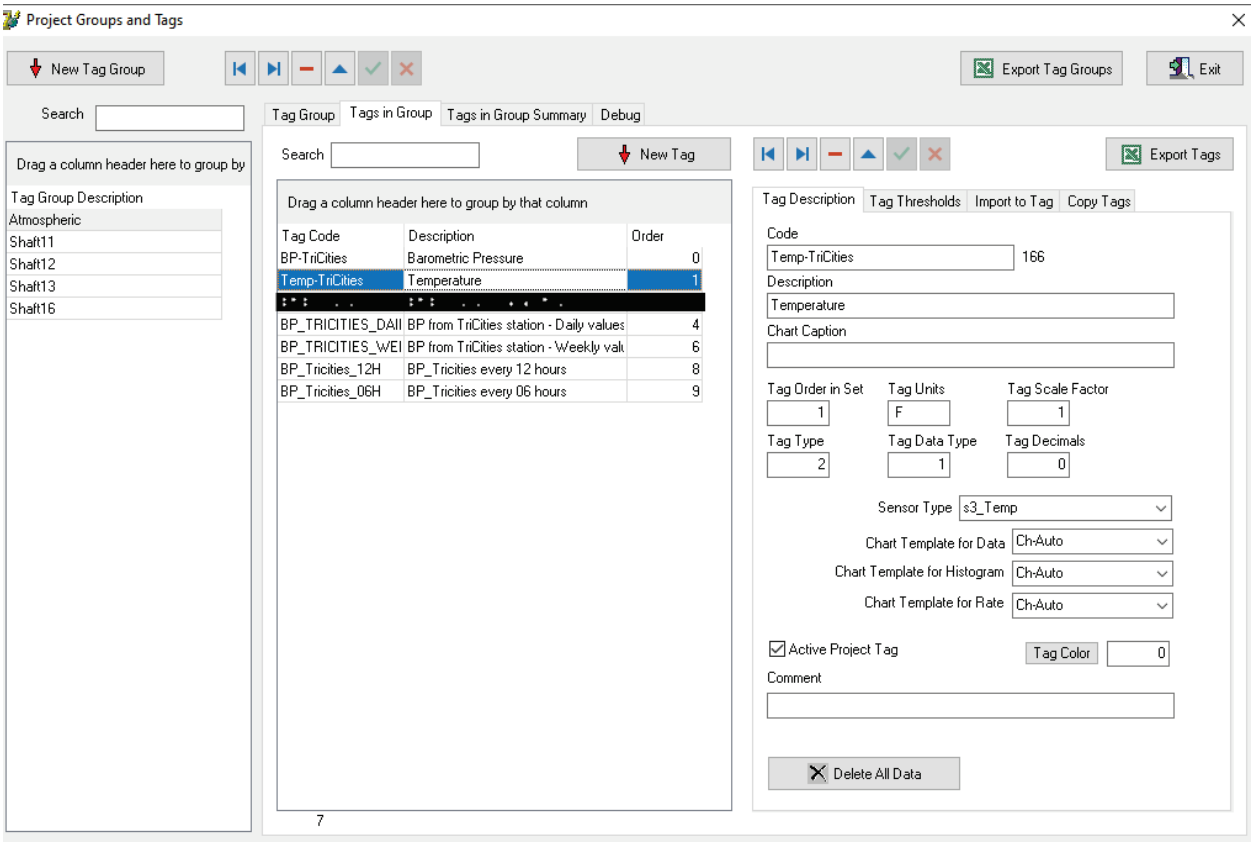


Figure 79: Tag definitions for atmospheric data (black line corresponds to a redacted entry)

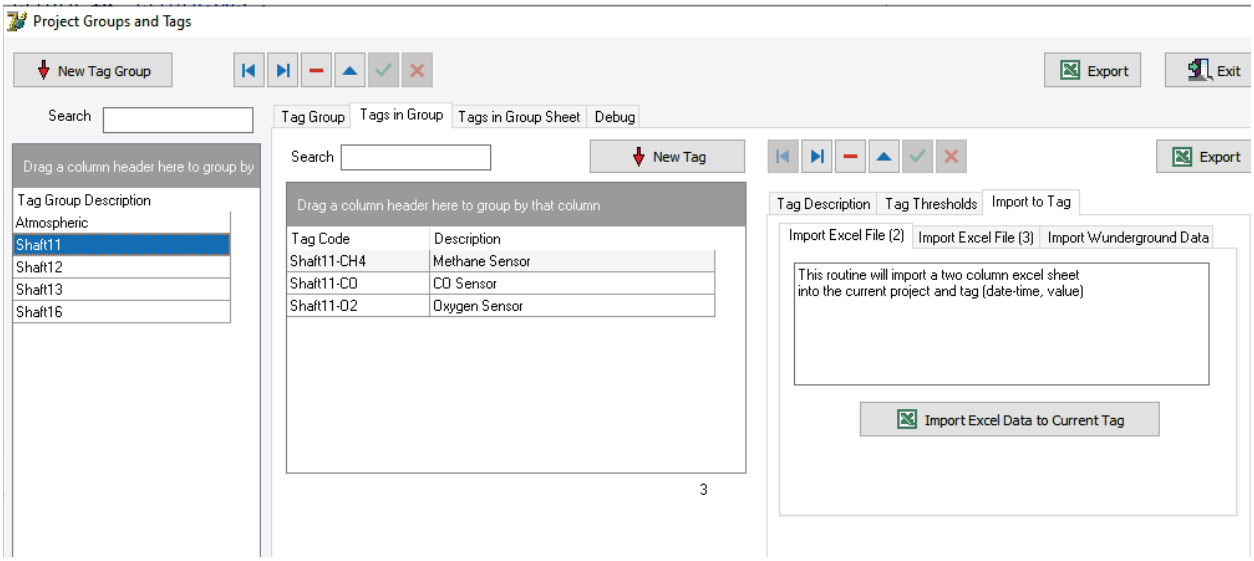


Figure 80: Interface to import a two-column excel sheet (date-time, value)

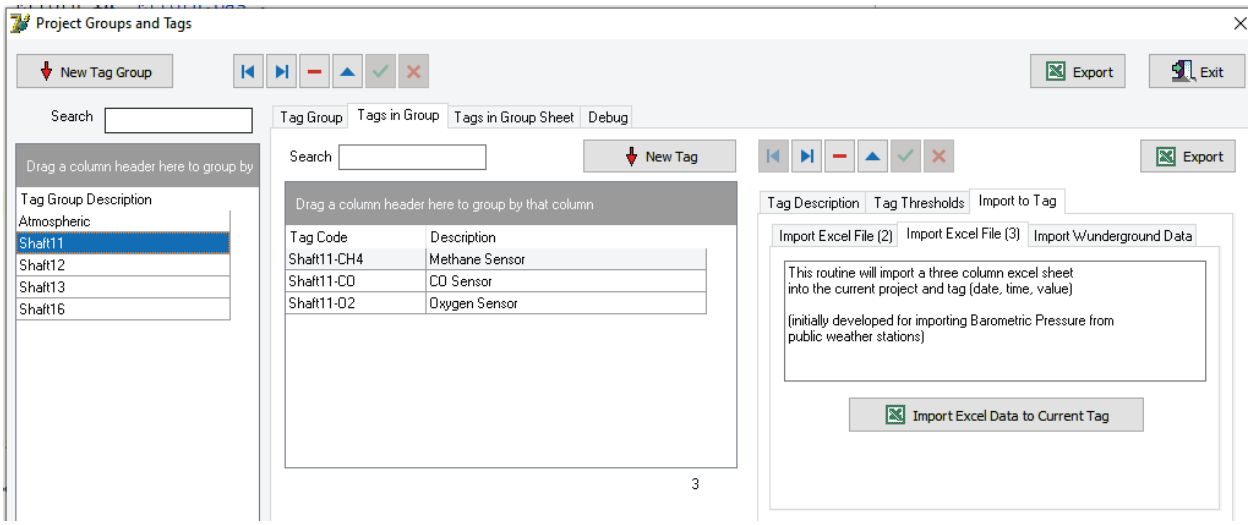


Figure 81: Interface to import a 3-column excel sheet (date, time, value)

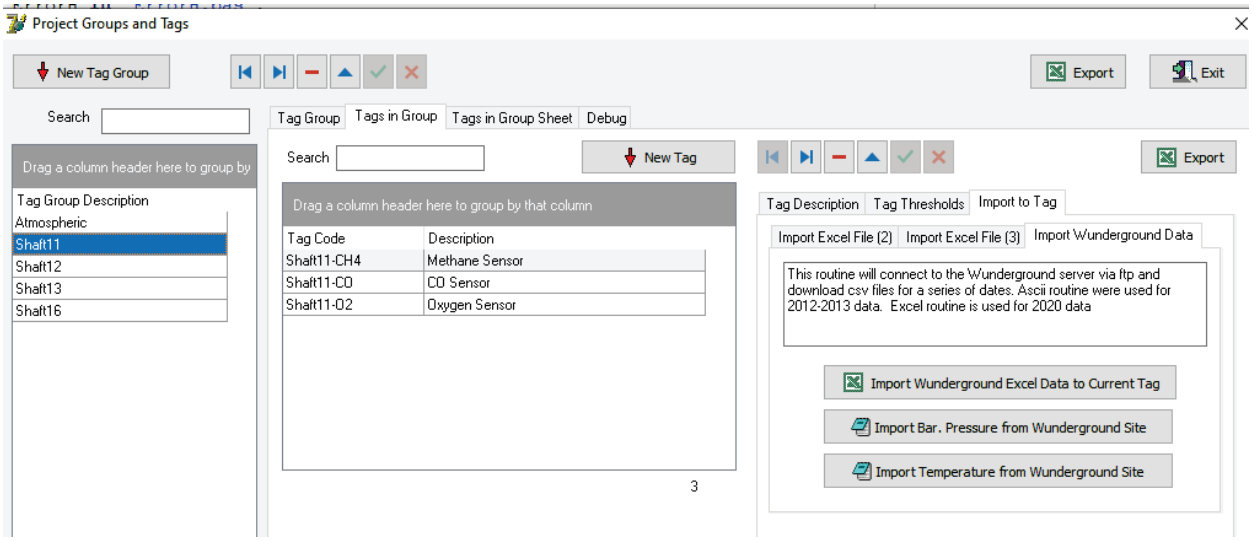


Figure 82: Interface to import Weather underground data to the database

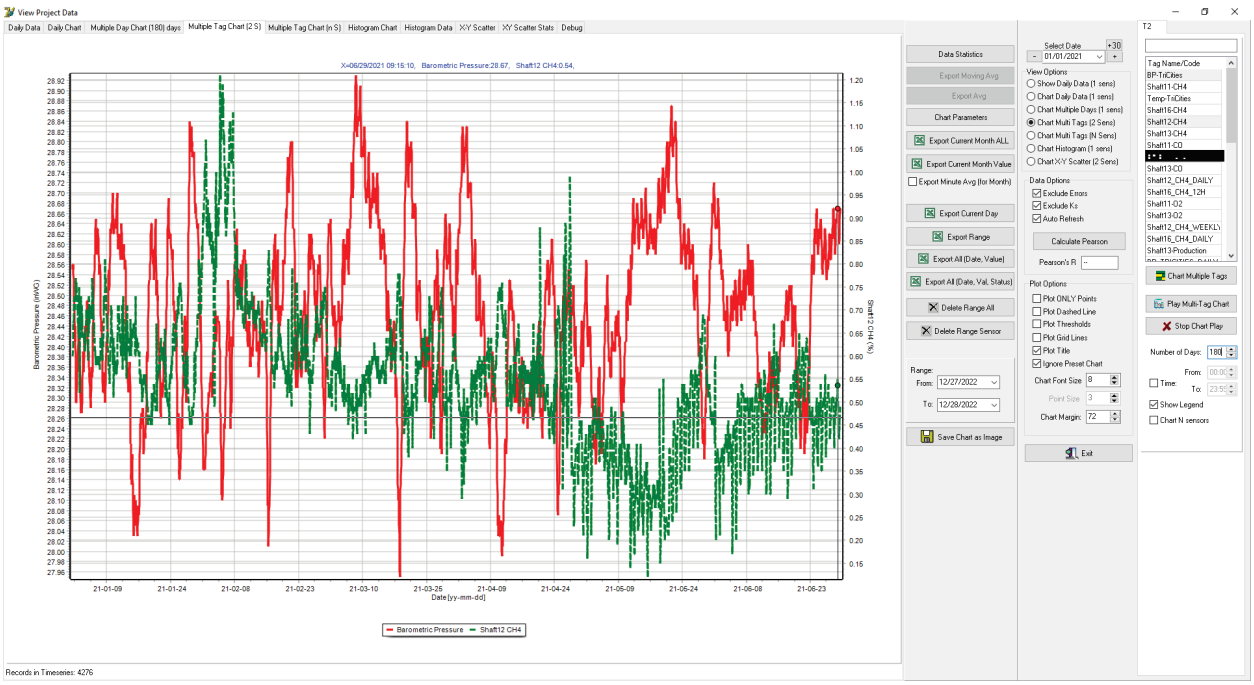


Figure 83: CH_4 vs Barometric pressure for 5 days for Shaft 13 (Mine A)

Appendix 5: ARIMA one-step-ahead model: Preliminary Results

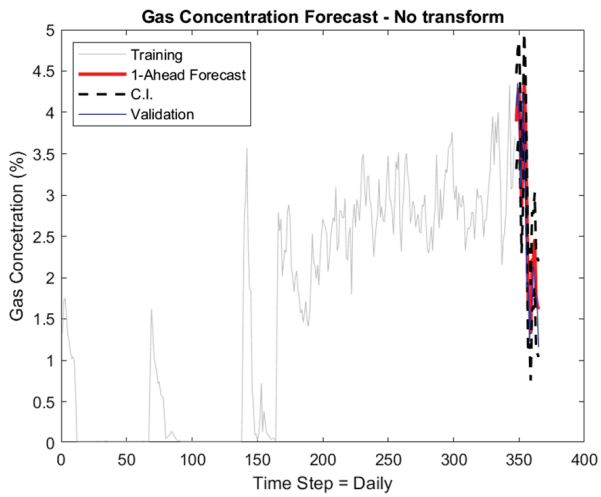
This appendix presents the results obtained from ARIMA one-step-ahead model that can be used to forecast methane gas concentrations in underground coal mines. The following figures correspond to the Mine A and pertain to several data segments with different lengths (e.g., one year and six years) and different time steps (e.g., 12-hours and daily). The graphs were obtained using custom code developed in the MATLAB environment.

Each figure below includes three areas (a to c). Graphs (a) are the forecasts obtained using the ARIMA (p,d,q) one-step-ahead model, plots (b) are magnified versions of the forecasts shown in graphs (a). Area (c) includes a table that shows the results of the validation measurements (Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Correlation (Cor) and Nash-Sutcliffe Efficiency coefficient (NS)).

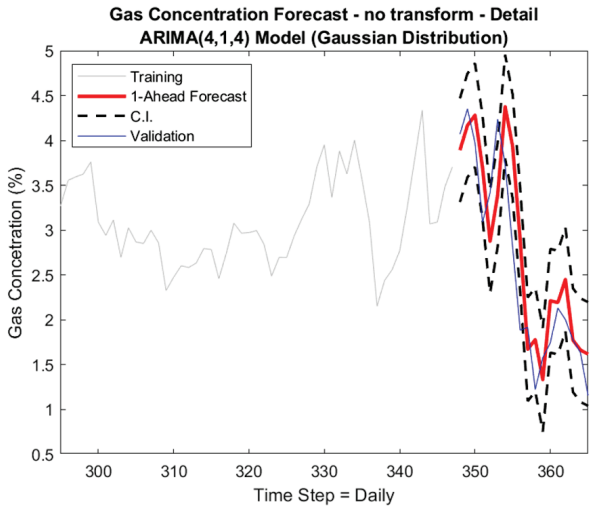
In the graphs, the gray line represents the training data (methane gas time series), the upper and lower black dashed lines signify the boundaries of the 95% Confidence Interval (C.I.), the blue line indicates the time series in the validation set, while the red line represents the forecast. The validation and forecast periods contain five percent (5%) of points used in the training time series.

Figure 81 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 Akaike Information Criterion 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 81b 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, as shown in Figure 81c, implying a strong correlation between the validation data and the forecasts. Therefore, the ARIMA (4,1,4) model provides a reliable forecast. Moreover, the observed values lie within the 95% prediction interval.



(a)



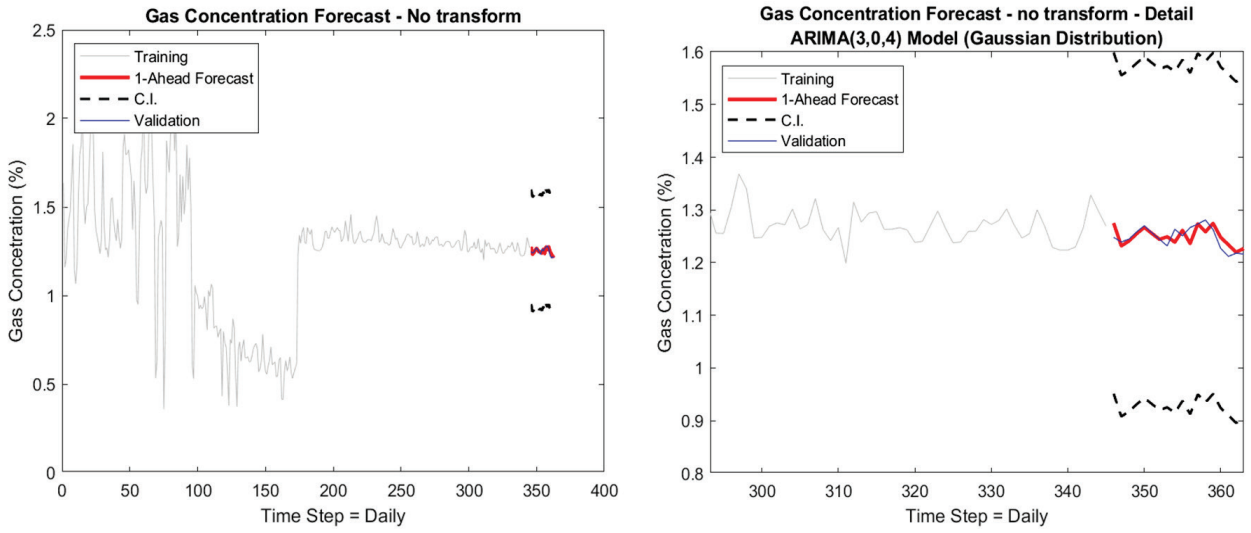
(b)

ME	MAE	RMSE	cor	NS
-0.19261	0.44163	0.54174	0.88586	0.74754

(c)

Figure 84. ARIMA one-step-ahead CH4 concentration forecasts using a daily average time step;(a) Forecasting of segment 1,(b) Magnified view of the forecast in (a), and (c) validation measures

Figure 82a shows the forecast obtained for a time series of methane gas concentrations spanning 360 days using a daily average time step. The ARIMA (3,0,4) model was selected based on the lowest Akaike Information Criterion value. As a result, the autoregressive order of the AR(p) term is four (p=3), the order of the differencing (d) is one (d=0), and the order of the MA(q) term is four (q=4). Figure 82b 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=0.65, as shown in Figure 82c, implying a strong correlation between the validation data and the forecasts. Therefore, the ARIMA (3,0,4) model provides a reliable forecast. Moreover, the observed values lie within the 95% prediction interval.

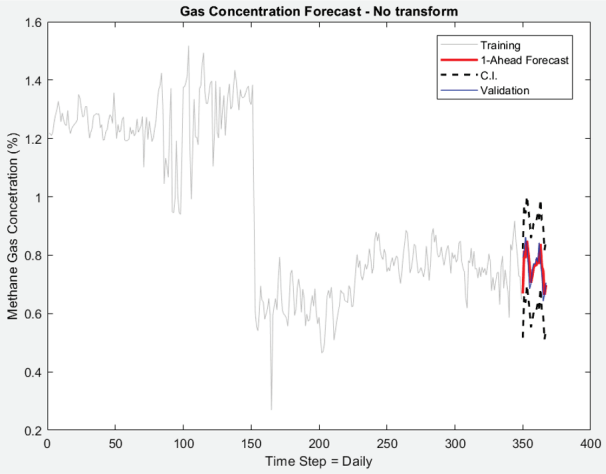


ME	MAE	RMSE	cor	NS
-0.0014491	0.012181	0.015759	0.64561	0.38313

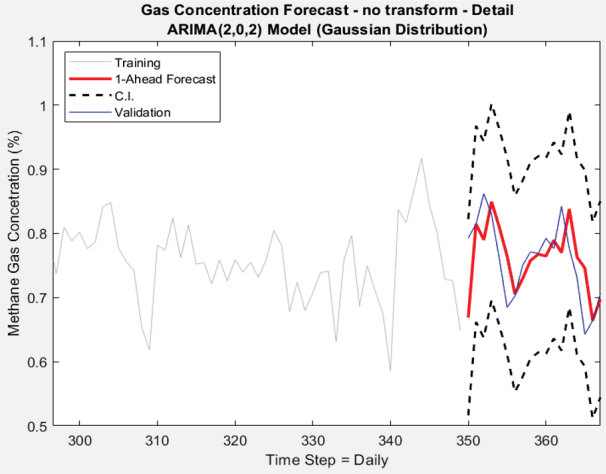
(c)

Figure 85. ARIMA one-step-ahead CH4 concentration forecasts using a daily average time step;(a) Forecasting of segment 2,(b) Magnified view of the forecast in (a), and (c) validation measures

Figure 83a shows the forecast obtained for a time series of methane gas concentrations spanning 360 days using a daily average time step. The ARIMA (2,0,2) model was selected based on the lowest AIC value. As a result, the autoregressive order of the AR(p) term is four (p=2), the order of the differencing (d) is one (d=0), and the order of the MA(q) term is four (q=2). Figure 83b 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=0.54, as shown in Figure 83c, implying a strong correlation between the validation data and the forecasts. Therefore, the ARIMA (2,0,2) model provides a reliable forecast. Moreover, the observed values lie within the 95% prediction interval.



(a)



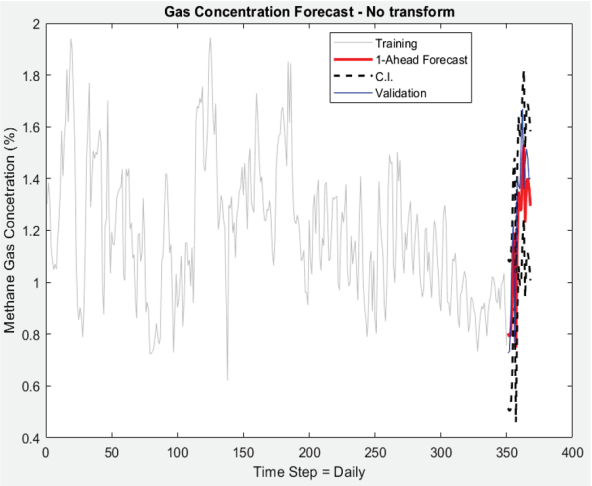
(b)

ME	MAE	RMSE	cor	NS
-0.0012426	0.039049	0.053815	0.53541	0.18209

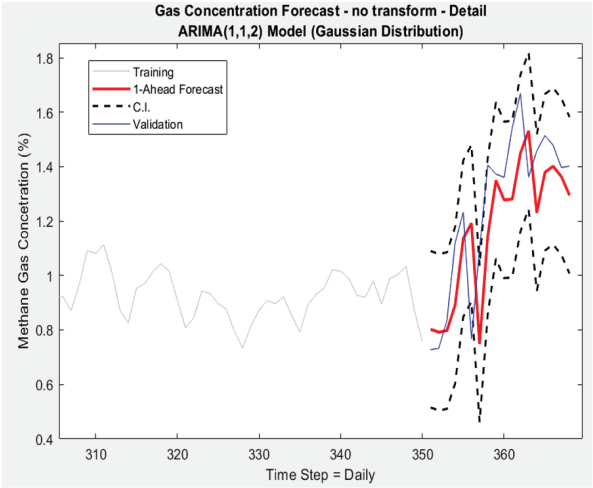
(c)

Figure 86. ARIMA one-step-ahead CH4 concentration forecasts using a daily average time step;(a) Forecasting of segment 3,(b) Magnified view of the forecast in (a), and (c) validation measures

Figure 84a shows the forecast obtained for a time series of methane gas concentrations spanning 360 days using a daily average time step. The ARIMA (1,1,2) model was selected based on the lowest Akaike Information Criterion value. As a result, the autoregressive order of the AR(p) term is four (p=1), the order of the differencing (d) is one (d=1), and the order of the MA(q) term is four (q=2). Figure 84b 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=0.79, as shown in Figure 84c, implying a strong correlation between the validation data and the forecasts. Therefore, the ARIMA (1,1,2) model provides a reliable forecast. Moreover, the observed values lie within the 95% prediction interval.



(a)



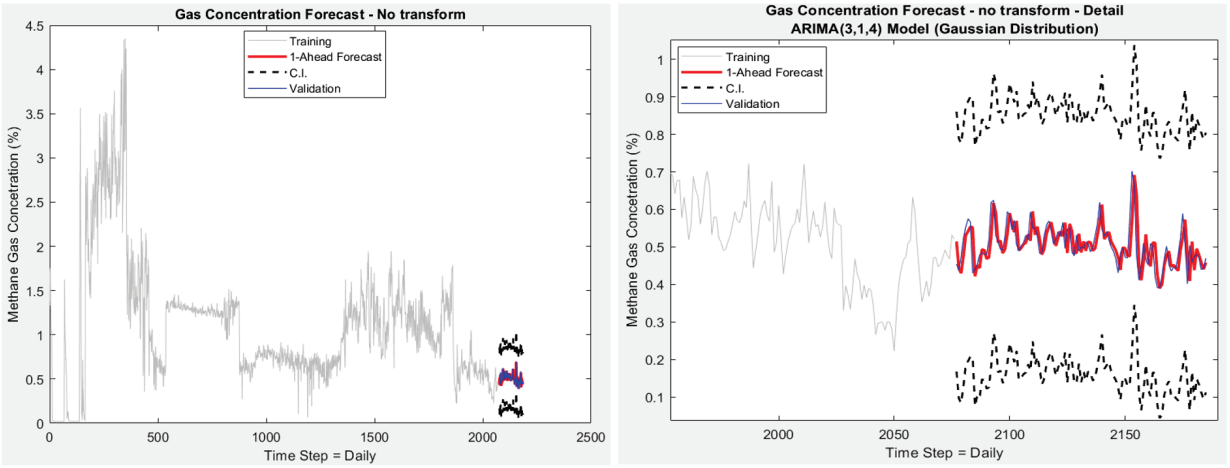
(b)

ME	MAE	RMSE	cor	NS
0.078533	0.15954	0.19529	0.79108	0.54983

(c)

Figure 87. ARIMA one-step-ahead CH4 concentration forecasts using a daily average time step;(a) Forecasting of segment 4,(b) Magnified view of the forecast in (a), and (c) validation measures

Figure 85a shows the forecast obtained for a time series of methane gas concentrations spanning more than 2100 days using a daily average time step. The ARIMA (3,1,4) model was selected based on the lowest AIC value. As a result, the autoregressive order of the AR(p) term is four (p=3), the order of the differencing (d) is one (d=1), and the order of the MA(q) term is four (q=4). Figure 85b 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=0.65, as shown in Figure 85c, implying a strong correlation between the validation data and the forecasts. Therefore, the ARIMA (3,1,4) model provides a reliable forecast. Moreover, the observed values lie within the 95% prediction interval.



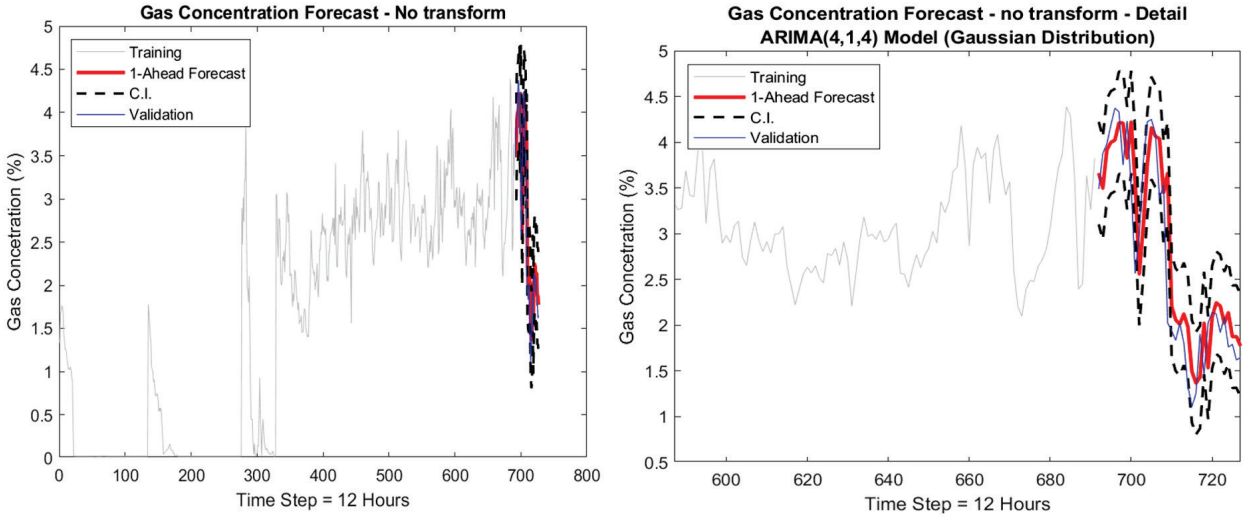
(a) (b)

ME	MAE	RMSE	cor	NS
0.0006874	0.032962	0.044127	0.65283	0.34352

(c)

Figure 88. ARIMA one-step-ahead CH4 concentration forecasts using a daily average time step;(a) Forecasting of segment 5,(b) Magnified view of the forecast in (a), and (c) validation measures

Figure 86a shows the one-step-ahead forecast for the gas concentration time series for the same period (360 days) presented in Figure 81, but using a twelve-hour average time step. The ARIMA (4,1,4) is again the best model. Figure 86b 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 (see Figure 81). Furthermore, the ARIMA (4,1,4) model was selected based on the lowest Akaike Information Criterion 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). Moreover, the observed values lie within the 95% prediction interval.

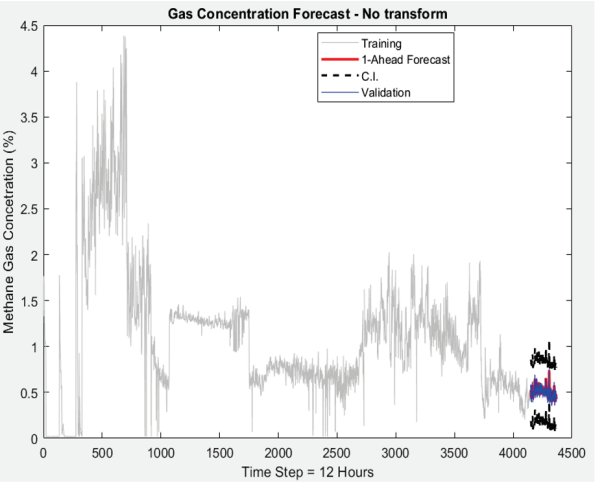


ME	MAE	RMSE	cor	NS
-0.11144	0.35458	0.47483	0.90477	0.80783

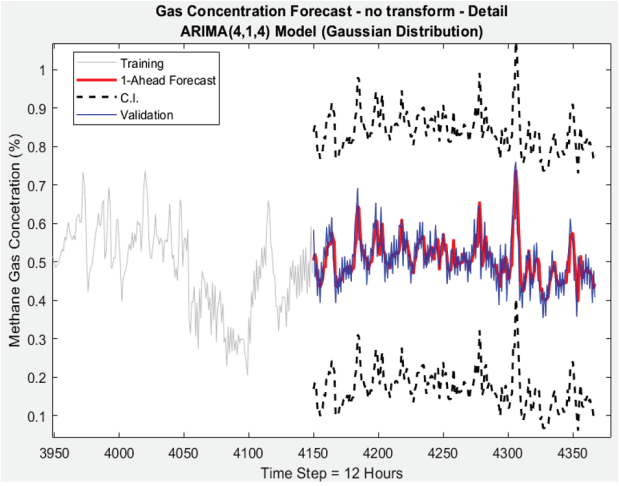
(c)

Figure 89. ARIMA one-step-ahead CH4 concentration forecasts using a 12-Hours average time step;(a) Forecasting of segment 1,(b) Magnified view of the forecast in (a), and (c) validation measures

Figure 87a presents the one-step-ahead forecast for the gas concentration time series for the same period (2100 days) presented in Figure 86 but using a 12-hour average time step, which corresponds to more than 4300 data points. Again, the ARIMA (4,1,4) remains the optimal model. Visual inspection of Figure 87b indicates that the forecast (red line) and the validation data (blue line) are notably similar. In addition, the correlation coefficient was calculated at R=0.71, significantly higher than the respective R for the time series shown in Figure 85c, which means that this forecast, which uses a 12-hour time step, provides a higher approximation to the actual values.



(a)



(b)

ME	MAE	RMSE	cor	NS
0.00051425	0.039326	0.0476	0.70596	0.49187

(c)

Figure 90. ARIMA one-step-ahead CH4 concentration forecasts using a 12-Hours average time step;(a) Forecasting of segment 2,(b) Magnified view of the forecast in (a), and (c) validation measures