Insights on the characteristics and sources of gas from an underground coal mine using compositional data analysis

C. Özgen Karacan[1,], Josep Antoni Martín-Fernández[2], Leslie F. Ruppert[1], Ricardo A. Olea[1]*

[1]United States Geological Survey, Geology, Energy and Minerals Science Center, Reston, VA, USA

[2]Univ. of Girona, Dept. of Computer Science, Applied Mathematics and Statistics, Girona, Spain

Abstract

Coal mine gas originates from the gas emission zone of the mine, as well as the longwall face and pillars. Gas emissions are controlled directly at the sources using horizontal or vertical boreholes drilled from surface or from the entries in advance of mining, or it is captured from the gob using ventholes during mining. The rest of the gas, especially that gas that originates from the longwall face and caved zone, mixes with the ventilation air and travels through bleeder and return entries before being exhausted to atmosphere from air shafts.

Although the gas associated with mining mostly focuses on methane, the gas is not pure methane but it is a mixture (composition), where both the components and their quantities vary depending on the sampling location. Understanding the evolution of the composition of the gas from source to different sampling and evaluation points in mines using proper statistical analysis and interpretation methods can lead to better designed degasification and ventilation systems and the selection of the most adequate utilization method for generating energy from the gas.

In this work, we present the results of compositional data analysis (CoDa) of gases sampled at different locations in a longwall mine operating Pennsylvanian coal- bearing strata in the Northern Appalachian coal basin. Sampling locations were from accessible parts of the bleeder entries, returns, bleeder evaluation points and shafts within the mine, and also from gob gas ventholes and coal degasification boreholes drilled in the panel areas. In addition, desorbed gas samples from seams that are important for the mine were included in the analyses for comparison. The compositional data analysis showed that the gas composition shifts based on the sampling location, putting in-mine and "pure" coalbed gases on the opposite ends. Removal of air from the samples did not change this observation suggesting that oxygen is depleted especially in in-mine samples due to oxidation. Results also suggested that desorbed gas samples may not represent the composition of the coal seam gas.

Keywords: Coal mine methane, mine ventilation, Aitchison distance, simplex, logratio

1. Introduction

Coal was the driving force behind the Industrial Revolution, and it has played a key role in the development of many countries. It continues to be an important source of energy for electric and power generation around the world, especially in developing countries with abundant coal resources. Recently, both small coal mines and the ones that operate under challenging geological conditions at high costs have been closed due to competitive prices of other forms of energy. Also, many coal-fired power plants were converted to gas and they are unlikely to be converted back to coal. Environmental concerns that evolved around coal use and coal mines, such as disruption to natural landscape and methane emissions, have not had a positive impact on the perception towards coal either. The International Panel for Climate Change (IPCC) recognizes methane (CH_4) as the second most important anthropogenic

greenhouse gas following carbon dioxide (CO_2) due to its higher (25-32 times that of CO_2) global warming potential despite lower atmospheric concentration (IPCC, 2013).

Despite recent mine closures, and even more that are yet to be closed and abandoned, there are still productive coal mines in many countries and methane continues to be one of the concerns due to its safety and environmental implications (Karacan et al, 2011; Hummel et al, 2018; Kędzior and Dreger, 2019; Swolkień, 2020). Large active mines continue to generate significant quantities of methane, which was historically been responsible for many deadly mine explosions in the U.S. (e.g. MSHA, 2006) and globally. Mine ventilation is the primary means of controlling methane within the mine to reduce its concentration, and emissions from ventilation shafts make coal mines one of the largest point sources contributing to global methane emissions (Global Methane Initiative, 2020). In cases where ventilation systems alone may not be enough to get excessive methane emissions under control, additional measures including gas drainage from the source coal seams and gas capture from disturbed strata using ventholes are employed (Karacan, 2009). Boreholes produce gas with high methane content, with compositions the same or similar to the specific coal seam being mined, whereas ventholes produce gas that is a mix from different coal seams and mine air extracted through fracture networks in the overlying disturbed strata. It is realistically not possible to capture all of the methane from an area affected by the mine when the mine is active. Any uncaptured methane may accumulate in abandoned and sealed sections and may migrate to the surface through fractures with changes in ground-water level (Sechman et al., 2013), through geological discontinuities (Voltattorni et al. 2015) or from shafts (Sechman et al., 2019) causing health risks in nearby communities and atmospheric emissions. Kholod et al (2020) argue that emissions from abandoned mines will continue through the 21st century, even though coal production will decline: they estimate that abandoned mine methane emissions will increase by a factor of 8 by the year 2100.

If methane emissions from different sources of active mines and the gas accumulation in abandoned mines are captured and utilized, not only can the harmful effects of methane be mitigated but also they become an energy source (Karacan et al, 2011; Global Methane Initiative, 2020). In order to take advantage of these benefits, national greenhouse gas emission inventories should be prepared by using reliable emission factors, potentially at different scales by relying on improved characterization of sources (National Academies of Sciences, Engineering and Medicine, 2018) and taking uncertainty into consideration (Wang et al., 2018; Kholod et al., 2020). Such work often requires monitoring, geological understanding, and identification of the source of methane. One additional, and important, consideration is the fact that although in many ventilation, inventory, gas capture and utilization studies the emphasize is given to methane due to the aforementioned reasons, coal mines and the associated disturbed strata generate and accumulate gas from multiple coal seams and ventilation air. Therefore, the mine gas is a gas mixture from different sources and thus must be thought of as a 'composition' as it carries relative information about the parts of the mixture. This type of data is called "compositional data" (CoDa) (Aitchison, 1986). In working with CoDa, proper metrics and statistics for adequate quantification must be followed since simply applying univariate analysis or classical statistical methods to gas fractions directly is inappropriate, as CoDa caries only relative information, which is contained in the ratios between the parts of the composition (Pawlowsky-Glahn et al., 2015). This is especially important from a data analysis point of view, which is usually one of the critical steps in inventory building or estimating an engineering system's capacity even just by using basic metrics of gas components such as mean, standard deviation and correlation coefficients (Filzmoser et al., 2009;

Edjabou et al, 2017). The same consideration also applies to coal and shale geochemistry, studies of gas and water compositions and their potential relations to geological processes and sources (e.g. Engle et al, 2016). Basically, any type of data that is part of a whole and is represented as percent, ppm, etc can be considered CoDa and requires special analysis techniques.

Treating the gas measurements with adequate data analysis methods and understanding the composition differences between different sampling locations may help identifying the source of the gas and lead to the design of more effective capture and utilization systems, as well as building more accurate emission inventories for forecasting. In this study, centered log ratio (clr) biplots, cluster analysis and analysis of shifts are used in an attempt to investigate the composition differences between sampling locations and to explore the statistical differences between gases sampled at different locations within the mine, capture points and sources. Such analyses have been used before for geochemical analysis of soil samples (Tolosana-Delgado and McKinley, 2016) to explore geo-environmental conditions and associations between geochemical species and to investigate the source of different chemical components observed in samples from monitoring stations along the Llobregat River and its tributaries in Spain (Otero et al., 2005), besides other applications. However, to our knowledge, this is the first application of a CoDa approach for the analysis of mine gases. In that respect, we believe that the paper will contribute to the literature by demonstrating the compositional shift of mine gases sampled at different locations from a CoDa perspective.

2. Study area and samples

The mine where the gas samples were acquired is located in Greene County, southwestern Pennsylvania. This geographic location is in the Northern Appalachian coal basin, which used to be a very active coal mining and coalbed methane producing region. The following sections describe the general geology of the studied mine's location and mine characteristics pertinent to this work and the samples studied in this paper.

2.1. Geology of the area relevant to mine methane generation

The geology of the general area pertinent to this study and the importance of different formations and coal seams from mine gas generation and control points of view are discussed in previous publications (Karacan and Goodman, 2011; Karacan et al., 2012; Karacan and Warwick, 2019). Although the studied mine here is different than the one in those publications, the geology and the general gas control measures due to mining-induced disturbances are very similar. Therefore, they will not be repeated here. However, for completeness, a brief information about the general geology and gas control practices will be given, while the readers are referred to location maps and general stratigraphic sections in the previous publications.

The study mine operates in the Pittsburgh main coal seam at about a depth of 300 m (900 ft). The Pittsburgh main seam is the lowest member of the Monongahela Group and is usually overlain by the Pittsburgh rider seam at 0.3-1 m (1-3 ft). The Redstone coal seam, when present, is 5.1-14.8 m (18-45 ft) above the Pittsburgh seam and is not uniform in thickness and is usually thin (~0.2 m) in the study area. The Sewickley coals, which include the Sewickley coal itself and any riders and splits of the main bench, are the second most important economically important coal beds in the Monongahela Group in the study area due to their thickness, uniformity and proximity to the Pittsburgh coal. The Sewickley coal is laterally persistent and generally 0.3-1.8 m (1-6 ft) thick and occurs approximately 60 m (180 ft) above

the Pittsburgh main coal. Other coals that are within the Monongahela Group are the Uniontown and Waynesburg coals. The Uniontown coal is approximately 89-98 m (270-300 ft) above the Pittsburgh coal and is usually not continuous with a varying thickness around 0.15 m (0.5 ft), when present. The Waynesburg coals, on the other hand, are at the top of the Monongahela Group and consist of the Waynesburg A and B coals, which are laterally persistent and usually multiple-bedded.

As previous publications also demonstrate (Palchik, 2003; Schatzel et al, 2012), longwall mining creates a large volume of disturbed strata due to caving and fracturing. This thick interval of formations, called gob, is also the source of mine gas due to the fractured coal seams and is termed as the gas emission zone (GEZ). The height of different zones and disturbances within the GEZ varies with the general geology and the thickness of the mined coal, as well as the depth and the dimensions of the longwall panels (Schatzel et al, 2012). Gas emission potential from the GEZ is a function of varying distance from the mined coal bed in overlying and underlying formations (Duda and Krzemień, 2018). Using rock displacement measurements with formation testers, Karacan and Goodman (2011) and Watkins et al. (2020) predicted that the height of the gas emission zone of the typical longwall mine panels might extend as high as 88-131 m (350-400 ft) above the Pittsburgh coal seam. Also, these and previous studies established three major zones of deformation within the GEZ: a caved zone, a highly fractured zone, and a bending zone. These zones and their characteristics are important for gas emissions and transport into the mines from the existing sources (mostly coal) within these intervals.

2.2. Source of gas emissions and gas mixtures in different parts of GEZ

In order to estimate the potential sources of gas emissions into the mines, either for safety purposes or to capture the gas for greenhouse gas management, coal seam locations in the GEZ and their relation to the mine need to be analyzed.

In the study location, the Pittsburgh rider coal is usually not extracted during mining and thus it is within the caved volume due to the its proximity to the mine. Therefore, any gas emitted from this coal and the Pittsburgh seam exposed at the longwall face, as well as any gas emissions that enter the caved zone from upper intervals through fractures, are handled by the ventilation air in the bleeder system, which is designed to transport the contaminants from the active areas as well as inaccessible parts of the mine, including the caved zone, to a bleeder fan. The Redstone coal is usually thin and may not be uniformly present in a large area. However, when it is present, it is close to the upper part of caved zone and its gas contribution can be handled by both the bleeder ventilation and gob venthole systems. Coal seam gas that mixes with ventilation air is monitored in the mines at shafts and in return and bleeder evaluation points, mostly for safety reasons, to check methane and oxygen levels as well as any other gases that may indicate oxidation.

The Sewickley, Uniontown, and Waynesburg coals, on the other hand, are within the fractured interval of the GEZ, and they are believed to contribute to the gas in the fractured system. The Sewickley coal, due to its thickness and proximity to the mine is the most important. Recent studies (Watkins et al, 2020; Su et al, 2019) showed that the lateral displacements and the hydraulic conductivity enhancements are highest at the Sewickley coal interval. Therefore, emissions from the Sewickley coal and any potential contributions from Uniontown and Waynesburg coals are captured using gob gas ventholes (GGV), the bottom of which are usually set at around 10-15 m (30-45 ft) to the top of the Pittsburgh coal to withdraw gas from a 60 m (200 ft) fractured interval through slotted casing. However,

since the bottom of the slotted casing is close to the caved zone, it is common that these ventholes withdraw gas mixture from the caved zone too.

Average gas contents of coals, which were determined using the direct method of testing, gas desorption (Diamond and Schatzel, 1998; Diamond et al, 1986), discussed above are given in Table 1. This table shows that there is a general increase in total gas content with depth, and that the Pittsburgh rider and Redstone coals have comparable gas contents. However, since they are much thinner compared to the other coals, their in-place gas capacity and, consequently, their total emissions to caved zone may not be extensive. Other non-coal formations of the Monongahela Group, such as limestones, mostly affect caving and fracturing characteristics of the gas emission zone, while sandstones and thin shales interbedded with coals may contain uneconomic quantities of gas that may release those upon fracturing. More detailed information about the geology of Northern Appalachian coal basin and the Monongahela group in relation to coalbed methane resources can be found in Bruner et al (1998).

Table 1. Average gas content (m^3/t) of Monongahela Group coals – in stratigraphic order – in Green County, Pa, USA.

Coal	Desorbed gas	Lost+desorbed gas	Residual gas	Total gas
Waynesburg	1.54	1.61	1.41	3.02
Uniontown	1.34	1.41	1.70	3.11
Sewickley	1.86	1.98	2.11	4.10
Redstone	3.30	3.37	0.25	4.00
Pittsburgh rider	3.55	3.75	0.29	4.57
Pittsburgh	3.22	3.37	2.64	6.00

2.3. Mine characteristics and general methane control measures

The coal mine, where the study was conducted had overburden depths ranging from 240 to 300 m (800 to 1,000 ft). The study was conducted on two adjacent longwall districts of the mine (Figure 1), where the panel widths were 410 m (1,350 ft) with varying lengths from 2,100 to 2,650 m (6,900 to 8,700 ft). Upon completion of the western district, it has been isolated from the eastern district by mine seals.

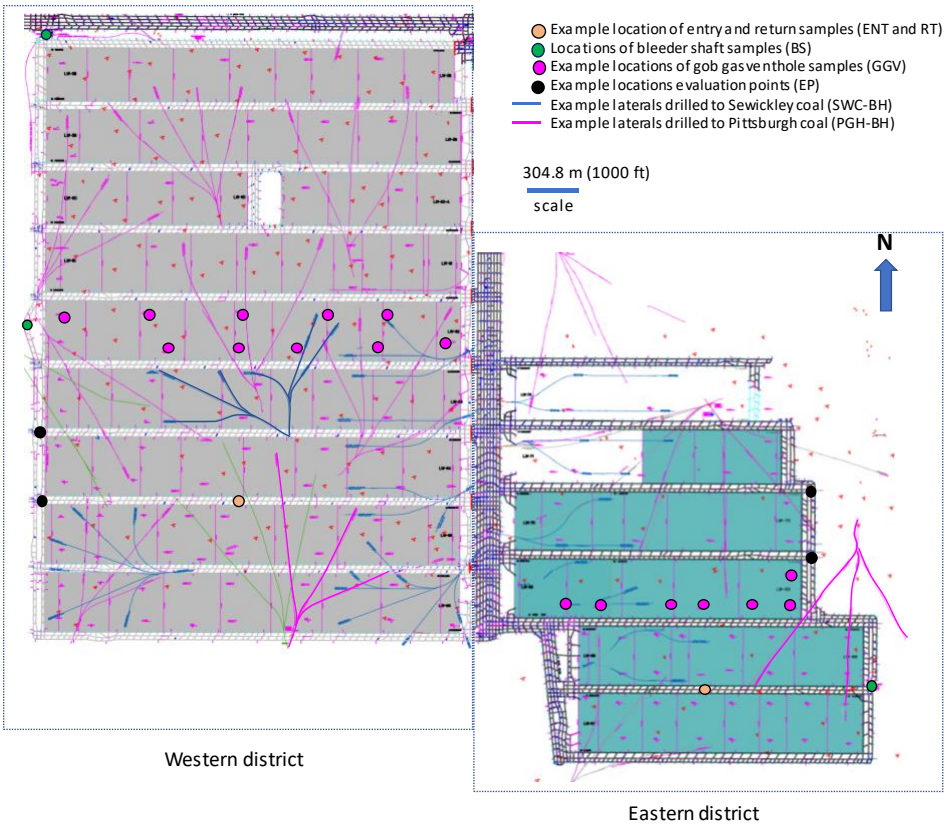

Figure 1. The mine area where the study was conducted. Samples were collected from various locations the flow and concentrations are checked on regular basis. Example locations are shown in the map.

The mine used a three-entry gateroad system in headgate and tailgate entries to ventilate the panel to prevent any hazardous gas accumulations. All gateroad entries surrounding the panels were developed with 4.9 m x 2.6 m (16 ft x 8.5 ft) dimensions. However, gateroad dimensions may change through time due to various factors, such as geomechanical properties of rock strata, depth of cover and prevailing rock stresses. The effective size of the openings over the life of the gateroad dramatically affects the amount of airflow that can be provided towards the back of the panel and to the bleeder shaft. In this mine, a bleeder design with parallel entries serving as separate airflow paths at the back of the panels was used. A negative pressure differential was created from the longwall face to the back of the bleeder system using ventilation controls to allow airflow through the caved material and the gateroad entries towards the bleeder shafts. Air quantity that reached the bleeder shafts (marked in Figure 1) from different legs were between 55.1 m^3/sec (118,000 cfm) and 102.5 m^3/sec (217,000 cfm). Since airflow occurs from high-pressure locations to lower-pressure ones, any disruption in gateroads results in changes in pressures and airflow quantity, which can be the limiting factor to the bleeder system's ability to remove methane from the longwall face and to eliminate any hazardous methane accumulations in the mine, including in the caved zone (Krog, 2016). Therefore, in this mine, the entries were supported using pumpable cribs and standing supports, as needed.

In order to support ventilation and the performance of bleeder system to maintain safety, the mine removes gas from the fractured zone of the panels using GGVs drilled from the surface (Krog, 2016; Karacan and Warwick, 2019). The number of GGVs in each panel was 5-7 drilled generally on the

tailgate margin, except at the panel start, where a headgate GGV was also drilled to better control high amount of emissions that may occur due to initial caving when the panels started to be mined (Figure 1).

The mine also drained gas from coal seams (Pittsburgh and Sewickley) from a few months to a couple of years in advance of mining using lateral boreholes. Lateral boreholes were either drilled from the surface or from development entries of the mine and produced coal seam gas that is not contaminated by mine air. These drained gases can be considered as the purest form of the gas that can be extracted from the seams. Pittsburgh and Sewickley coals were the main targets to drain gas using laterals to decrease their emission load in the ventilation system. These laterals are terminated and plugged before the mining face intersects them. Example gob gas venthole locations and the laterals drilled into different seams at this mine location are shown in Figure 1.

3. Gas samples and their classification for analysis

This study employed gas samples acquired by the mine personnel at different locations in the mine as part of regular ventilation surveys and from gob gas ventholes and lateral boreholes for production monitoring. In addition, the atmosphere in critical accessible in-mine locations and GGVs are monitored for system performance and also to check any indication of spontaneous combustion. These in-mine locations include walkable entries (ENT), bleeder evaluation points (EP) and shafts (BS) as well as returns (RT). In order not to disclose the identity of the mine for confidentiality reasons, the exact names of the locations are not given in the figure, but some are shown in Figure 1 as examples and to help with understanding the sample classification described later in this section.

Gas samples were collected with sampling bags and were analyzed by a commercial laboratory using gas chromatography to determine concentrations (as volume %) of N_2, O_2, He, H_2, Ar, CO_2, CH_4, C_2H_6, C_2H_4, CO, C_{2+} hydrocarbons. In addition, compositions of desorbed gas from coal cores retrieved from exploration boreholes around the mine area were included in the dataset. These gas samples were obtained and analyzed during the direct method of testing by NIOSH (National Institute for Occupational Safety and Health), except for the Waynesburg coal gas composition, which was obtained from Markowski (2001). It has been suggested (Jin et al, 2010), however, that any air contamination in desorption canisters results in oxidation and thus consumption of oxygen as the tests prolong, changing the O_2 and N_2 balance due to air and resulting an overestimation of nitrogen.

The entire dataset acquired for this work consisted of 325 gas samples from different locations and the suit of compositions listed above. However, not all the components were present or detected in all samples. For instance, compositions of desorbed gas samples were already corrected for air and the compositions were closed to 100%. Similarly, borehole and in-mine samples were analyzed for the entire suit of components, but especially borehole samples were not analyzed for CO, C_2H_4 and H_2, and in-mine samples were not analyzed for hydrocarbons heavier than CH_4. Therefore, to unify the samples for analyses, main components that were detected with varying quantities in all samples were reduced to N_2, O_2, CO_2, CH_4, C_2H_6 and “other gases”, as the balance of closure to 100%. Any sample that reported zero or not-detected quantities for any of the main components was eliminated from the database, as a requirement of the log ratio methods in short of applying zero replacement. The elimination was performed after a statistical inspection in the pre-processing step that showed no relevant differences between removed samples and full observed samples as regards the common recorded parts to avoid bias.

Overall, two sets of data were compiled for analyses and comparison.

- The first set included 238 samples with a 6-part composition (N_2, O_2, CO_2, CH_4, C_2H_6 and other gases) from different sources (except desorbed gas samples from coals).
- The second set included 256 samples with 5-part compositions after an air correction procedure based on air O_2 and N_2 ratios to evaluate the effect of air on compositional differences. This set also included desorbed gas samples from coals, since they were already corrected for air, to compare the rest of the samples to coals.

The dataset used for this work included, as mentioned before, samples from different locations. Especially the in-mine samples were very diverse in terms of exact locations and sampling frequency. In order to make the analysis possible and compare different types of samples as groups of similar origin, they were classified based on the general sampling locations rather than the individual points. For example, different legs of different bleeder shafts were sampled multiple times at different dates, as were the returns and the samples from different entries of the bleeder and gateroad system. Therefore, the following classification and naming convention was used (please refer to Figure 1 for locations shown as examples):

- The bleeder shaft samples, regardless of which leg of the shaft it belonged to were classified as “BS”. Similarly, return samples were called as “RT”, samples from bleeder evaluation points were named as “EP”, and the samples that belonged to gateroads and walkable entries of the bleeder system were classified as “ENT”.
- Gob gas venthole samples, regardless of which venthole of the panel, were called “GGV”.
- Lateral borehole samples were named with the coal seam they produced gas from: SWC-BH and PGH-BH represented the laterals that produced from Sewickley and Pittsburgh coal seams, respectively.
- Gas samples from desorption experiments were named RDR, RDS, SWC, PGH and WBG to refer to Pittsburgh rider, Redstone, Sewickley, Pittsburgh and Waynesburg coals, respectively. Unfortunately, gas composition for the Uniontown coal was not available to include in the data.

Table 2 gives the median (or second quartile Q_2) and interquartile range (IQR= Q_3-Q_1) values of components and samples in general categories. The median of component values in each category of samples is an estimator of the central location of the random samples. The IQR is a measure of dispersion around the median, similar to standard deviation, but it is more robust than standard deviation, especially in the presence of potential outliers. Gas samples before correcting for air (Table 2) shows that borehole samples (PGB-BH+SWC-BH) that are merged into a single borehole class have a high (1.85) IQR for CO_2, potentially due to PGH-BH samples generally having CO_2 values around 10%, while SWC-BH samples have 3-4%. Similarly, O_2, N_2 and CH_4 from GGVs have high IQRs (12.10, 27.87 and 36.98 respectively), possibly due to variability of compositions in terms of air and methane produced from GGVs drilled at different locations over the panels and at different panels in the district. After O_2 is removed from the data set (Table 2), “other gases” shows more dispersion in in-mine samples (BS, RT, ENT and EP), and the dispersion of especially N_2 and CH_4 increase in almost all sampling locations.

Table 2. Median and interquartile range (median, IQR) of all gas samples in both datasets. The unit is %.

Gas samples before correcting for O_2 (n=238)							
Location	Sample quantity	CO_2	O_2	N_2	C_2H_6	CH_4	Other gases
BS	39	0.35, 0.14	19.72, 0.56	78.14, 0.28	0.02, 0.02	0.72, 0.71	0.88, 0.02
ENT	122	0.39, 0.26	19.74, 0.93	77.95, 0.54	0.02, 0.04	0.86, 1.36	0.89, 0.04
EP	18	0.48, 0.51	19.22, 1.84	78.03, 0.92	0.03, 0.03	1.36, 1.34	0.85, 0.08
GGV	30	0.53, 0.77	7.45, 12.10	39.96, 27.87	0.44, 0.38	52.49, 36.98	0.46, 0.40
PGB-BH+SWC-BH	23	8.27, 1.85	0.07, 1.52	0.58, 5.66	0.21, 0.29	88.49, 7.87	0.01, 0.06
RT	6	0.19, 0.42	19.18, 1.51	78.23, 0.11	0.02, 0.03	1.47, 1.14	0.85, 0.07
Gas samples after correcting for O_2 (n=256)							
BS	39	5.45, 1.75		65.28, 8.43	0.26, 0.21	11.09, 6.48	12.94, 7.91
ENT	122	5.79, 1.59		64.37, 18.47	0.23, 0.29	11.69, 8.85	12.67, 10.54
EP	18	5.47, 0.88		72.13, 11.14	0.35, 0.15	15.77, 8.59	9.43, 7.46
GGV	30	1.12, 0.94		15.33, 17.36	0.78, 0.10	79.92, 19.99	0.62, 1.79
PGB-BH+SWC-BH	23	8.27, 1.71		0.15, 0.18	0.22, 0.31	89.99, 1.30	0.01, 0.06
RT	6	3.45, 2.18		69.41, 13.06	0.16, 0.31	11.44, 9.31	9.44, 22.55
Desorbed gas	18	7.61, 2.99		3.42, 4.95	0.11, 0.10	88.69, 7.43	0.17, 0.21

4. Methodology of data analysis

The type of multivariate data used in this work is CoDa, a set of real vectors $\mathbf{x} \in R^D$ with positive components. The data are usually represented with a constant sum (closure constant), i.e. with sample space a *D-part* simplex (S^D), and carry information about relative abundances of each part in the total, rather than absolute magnitudes (Aitchison, 1986). The relative nature of the CoDa gives it a multivariate character with D-1 degrees of freedom, as the parts should be compared at least by pairs. The relativeness of CoDa also makes it possible to take a sub composition of any two or more parts and close it to the desired constant. From a practical point of view, the relative information contained in the CoDa makes classical statistical methods, particularly those based on the covariance structure, not appropriate if compositions are treated as absolute values and the distances to explore their differences are handled with Euclidean geometry (Otero et al, 2010). The issues related to compositions and the use of classical statistics was addressed by Aitchison (1982), who introduced the log-ratio scores for representing a composition. In this paper, we used centered log ratio (clr) scores (Aitchison, 1986)

$$clr_i(\mathbf{x}) = ln\frac{x_i}{g(\mathbf{x})} = \frac{1}{D}\sum_{\substack{k=1;\\ k\neq i}}^{D} ln\frac{x_i}{x_k} \qquad i = 1, 2 \dots, D, \tag{1}$$

where $g(\mathbf{x}) = \sqrt[D]{x_1 x_2 \dots x_D}$ is the geometric mean of a composition $\mathbf{x}$. The clr-score, $clr_i(\mathbf{x})$, informs the relative importance, in logarithmic scale, of the chemical element x_i with respect to all other elements. Clr-scores are used for defining a distance concept suitable to compute differences between compositions, the Aitchison distance (d_A):

$$d_A(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{D}(clr_i(\mathbf{x}) - clr_i(\mathbf{y}))^2} \tag{2}$$

which is the Euclidean distance between the clr-vectors *clr*(**x**) and *clr*(**y**). Although, depending on the nature of the data, the choice of the appropriate distance and the methods to analyze them lies with the analyst, one should recognize that the amplitude of the differences between CoDa should not be treated as absolute, as this decision may affect statistical results due to the inappropriate choice of the metric for the sample space (Palarea-Albaladejo and Martín-Fernández, 2012). Therefore, after closing the data, an appropriate distance should be chosen followed by an isometric transformation to bring the data into real vector space to apply classical statistical methods (Pawlowsky-Glahn et al., 2015).

A *form clr-biplot* (biplot of clr-vectors) (Pawlowsky-Glahn et al., 2015) is similar to biplot of PCA (Principal Component Analysis) in classical multivariate statistics with adaptations to the properties of CoDa, and collectively represent clr-variables as rays and the samples as dots. Different from classical biplots, in a clr-biplot the vertices of rays and the links between those rays are important to interpret because the cosine of the angle between two links approximates the correlation coefficient between the two pairwise log-ratios involved. The length of a link between two clr-transformed variables approximates –up to the projection- the standard deviation of the logarithm of the ratio of the variables. Therefore, if two vertices are close to each other, their quotient might be constant, and thus they might be proportional. The orthogonal links between four different clr-variables may indicate the independence of pairwise log-quotients of the parts in the corresponding numerators. And, if three or more vertexes can be connected by the same link, the corresponding sub-composition might have one degree of freedom. Both classical and clr-biplot share that the axes origin represents the center of the data set. However, in a classical biplot the axes origin represents the arithmetic mean of the variables, whereas it represents the geometric mean of parts in a clr-biplot.

In this work, we have used exploratory CoDa analysis, clr-biplots and cluster analysis performed with Aitchison distance using both data sets to statistically investigate the parts and the gas samples for their associations and similarities. Sample classifications and naming convention used throughout the analysis is as described in section 3.

5. Results and discussion

5.1. The dataset without correction for air (6-part composition)

Variation array reports pairwise log-ratio of variances (variance $\ln(x_i/x_j)$) and means (mean $\ln(x_j/x_i)$) in the upper and lower triangles, respectively. In the variation array, (x_i) is the part in the columns, whereas the (x_j) is the part in the row during pairwise comparisons. The (+/-) signs of the log-ratio mean values indicates the direction of the ratio between the corresponding components. In the array, small variance values suggest a possibly constant ratio between parts, whereas high values indicate no relationship between them. Therefore, variance $\ln(x_i/x_j)$ is a measure of association between two parts. However, the lack of a scale in the variation array may make a more in-depth interpretation difficult, especially if there are very high and very small (close to zero) values. To ease this potential issue, entries of the variation matrix can be normalized by introducing a reference composition (Egozcue et al, 2018). In this work, normalization was not performed since observed variances could be interpreted without the need for scaling.

Table 3 gives geometric mean and the variation array of closed parts constructed for the dataset with O_2 (n=238) included. The geometric mean data shows that N_2 has the largest relative abundance of all parts, followed by O_2 due to in-mine samples. Variation array shows that the possibly highly associated (proportional) parts in these samples are O_2 and N_2, and C_2H_6 and CH_4 as indicated by the low values of 0.11 and 0.94, respectively. The expected value of the logratio $ln(N_2/O_2)$ is 1.49 indicating the expected proportionality is approximately $N_2 \approx 4.43 \cdot O_2$. Given the fact that O_2 is not a component of pure coalbed gas, this indicates its association to air and the N_2 in samples. However, it should be noted that the ratio of N_2 to O_2 in air is ~3.7. Therefore, a ratio of 4.48 indicates an overall O_2 depletion in the sample set. C_2H_6 and CH_4 are components of the coal gas, which explains their proportionality. Additionally, “other gases” has relatively low log-ratio variances with O_2 and N_2 indicating that this part

is more proportional to O_2 and N_2, and thus to air and other potential sources such as coal oxidation products, rather than the two significant parts of coal gas (C_2H_6 and CH_4). This association can be explained with the composition of "other gases" that includes Ar, CO, H_2, C_2H_4. As for CO_2, it might be weakly proportional to CH_4 and C_2H_6 in these gas samples although it is also a constituent of coal gas, which may indicate other sources for CO_2.

Table 3. Geometric mean and variation array of pairwise logratios, except desorbed gas, samples (n=238) before O_2 correction.

Geometric mean

CO_2	O_2	N_2	C_2H_6	CH_4	Other gases
0.8332	17.4714	77.3338	0.0476	3.4827	0.8311

Variation array (Lower diagonal: Expected Value, Upper diagonal: Variance)

	CO_2	O_2	N_2	C_2H_6	CH_4	Other gases
CO_2	0.00	6.87	5.97	1.57	1.77	6.96
O_2	3.04	0.00	0.11	8.40	11.32	0.27
N_2	4.53	1.49	0.00	7.28	10.01	0.24
C_2H_6	-2.86	-5.90	-7.39	0.00	0.94	8.07
CH_4	1.43	-1.61	-3.10	4.29	0.00	11.10
Other gases	0.00	-3.05	-4.53	2.86	-1.43	0.00

It should be kept in mind that the variation array for this dataset was constructed using all samples, except desorbed gas, without a distinction of sampling locations and their individual compositions. Therefore, it has an exploratory value to understand the general relative abundance of the parts and their association considering all samples. In order to explore the classification of parts and gases at different locations, and their associations with each other, clr-biplot was employed (Figure 2).

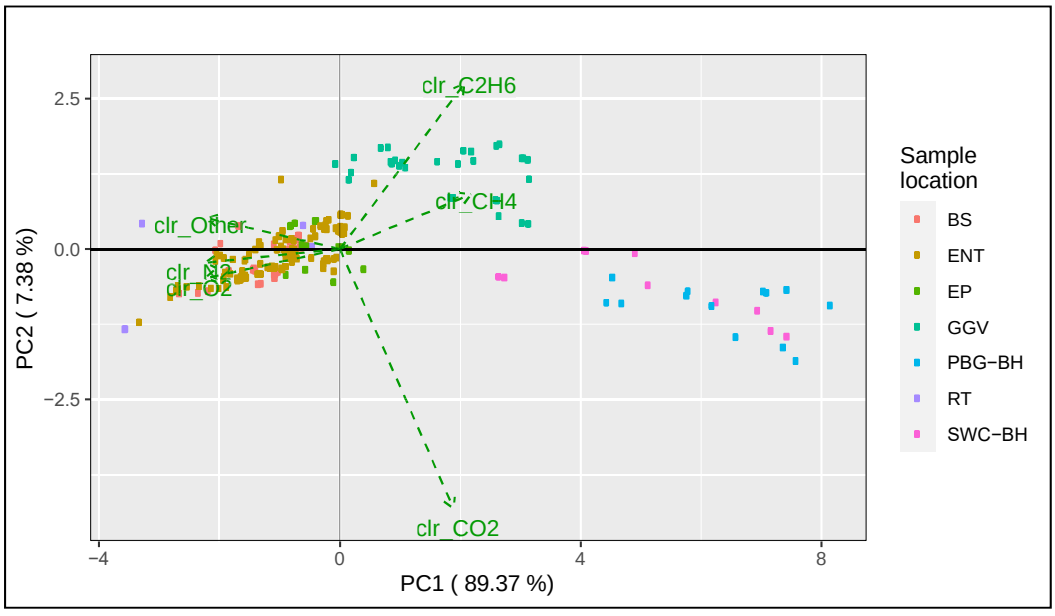

Figure 2. Form clr-biplot of gas samples (n=238) with the parts and sampling locations.

This plot retains approximately 97% of the variance, distributed between the first principal component (PC1=89.37%) and the second (PC2=7.38%). Also, given that the first axis (PC1) can describe most of the variability, it can also be used to discuss the distribution of the samples between negative and positive values. Negative values that are associated mostly with relatively high amounts of N_2, O_2 and "other gases" (clr-Other on Fig. 2) in samples are populated with samples collected from entries (ENT), bleeder shafts (BS) and evaluation points (EP). This grouping is meaningful because these are the in-mine samples, which have significant amounts of ventilation air in their composition. Also, this plot shows that there is not any discernible compositional difference between these samples to distinguish them from each other by using the 6-part composition studied here. Also, the gas samples from the lateral boreholes of Pittsburgh and Sewickley coal seams, PGH-BH and SWC-BH respectively, position themselves on the positive values of the same axis, and they are relatively rich in CH_4, C_2H_6 and CO_2 relative to the other three parts in the composition considered. These samples can potentially be characterized as closest to the composition of the gases from the respective coal seams too. Another noteworthy observation is, from a compositional point of view, SWC-BH and PGH-BH samples do not show differences. Therefore, the two opposite sides of PC1 represent samples with different compositional character and sources: the negative direction is for samples contaminated with ventilation air, and also possibly with coal oxidation gases that are lumped in "other gases", and the positive direction is for "pure" coalbed gases represented by borehole samples that are relatively richer in CH_4, C_2H_6 and CO_2. In between, GGV samples are located. From a gas source point of view, this makes sense because GGV gases are a mixture of coal seam gas, due to their completion at SWC and RDS coal intervals, and mine gas, due to the proximity of the well bottoms to the top of the caved zone, from which they likely withdraw in-mine gases similar to BS or ENT. Figure 3 also shows that the vertices of the rays of O_2, N_2 and "other gases", and also those of CH_4 and C_2H_6, can be connected by relatively short links. This indicates that these parts are related and can be treated as sub-compositions, in accordance with the previous interpretations, with single degrees of freedom.

In order to explore the compositional similarities of samples and their groupings based on source, or origin, we have used Ward's hierarchical cluster analysis with Aitchison distance (Eqn. 2). The Ward's method minimizes the sum of squares of distance between any two clusters that can be formed at each step based on the given distance criteria. Figure 3 presents the resulting dendrogram for samples with 6-part composition and preliminary suggests 3 or 4 cluster structure within the data.

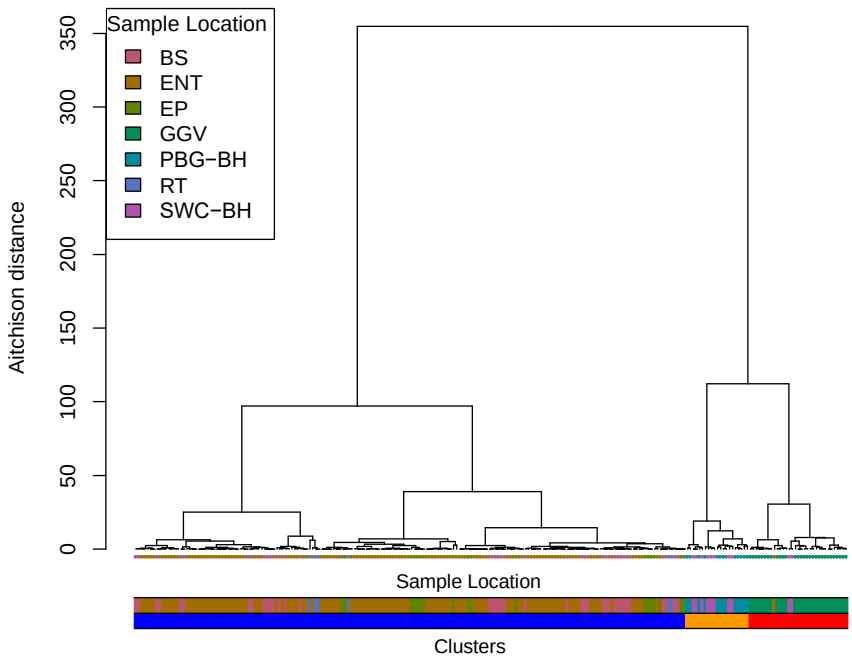

Figure 3. Dendrogram of Ward's hierarchical cluster analysis on 6-part composition samples.

For the selection of meaningful number of clusters, a cluster validation study needs to be performed. Cluster validation, with the support of expert opinion about the data, evaluates the goodness of clustering results by using either external or internal validation. The main difference is whether external information is provided, or only the hidden information within the data set is used for clustering. In this work, we used an internal validation approach by utilizing three different indices, i.e. Calinski-Harabasz index (Calinski and Harabasz, 1974), Dunn index (Dunn, 1974) and Silhouette index (Rousseeuw, 1987), to choose the optimal number of clusters without any additional information.

The Calinski-Harabasz (CH) index, also known as the variance ratio criterion, measures the intra-clusters and inter-cluster dispersions between points and cluster centers for all clusters created at any step, and evaluates cluster validity in terms of an index of cluster separation. A higher score reflects the optimal number of clusters. Dunn index evaluates the variance between members of a cluster as a measure of compactness and the distance between different clusters in comparison to inter-cluster cluster variance as separation. The number of clusters that maximizes the Dunn index is considered as the optimal number. The third index that was used in this work was the Silhouette index, which evaluates the silhouette width of each data, an average width for each cluster and an overall width for the entire data set according to a dissimilarity formula, whose result varies between -1 and 1. An index value close to 1 refers to a point being in the correct cluster, whereas a value close to zero means that the data is equidistant and can be in either of the closest clusters. A value close to –1 means that data is misclassified. In essence, all indices aim to identify sets of clusters that are compact and well separated, but since they may result in different cluster formations, their results can be interpreted together with the aid of what is known about the data to decide the number of clusters. Therefore, in this work, all thee indices were utilized to make a decision on the number of clusters.

Figure 4 presents the results of three indices used for identifying optimum number of clusters for the 6-part compositional data of gases discussed in this section. The results show that the CH index

suggests 3, 4, or 8 clusters, Dunn index indicates 2 or 3 and Silhouette index suggests 2 or 3 clusters for this data set. Since 2 clusters are too distant (Figures 2 and 4) and 7 or 8 are too many to explain the data reasonably well, a 3-cluster structure was chosen as the optimum (Figure 3). These obvious clusters belong to in-mine samples, GGV samples and lateral coal gas production boreholes (SWC-BH and PBG-BH) due to distinct differences in compositions of these groups.

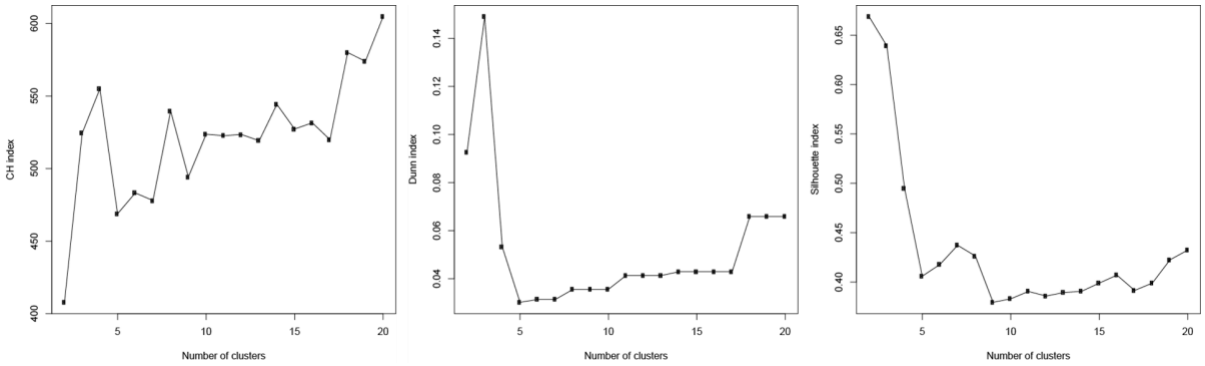

Figure 4. CH, Dunn and Silhouette index plots as functions of number of generated clusters for the gas samples with 6-part compositions (data without correction for O_2).

Individual Silhouette widths of each sample within clusters were further analyzed for their distribution and to explore samples' belongings to clusters. Figure 5 shows mapping of individual Silhouette widths of each sample within 3 clusters. These clusters have 184, 33 and 21 samples corresponding to in-mine, GGV samples and lateral boreholes, respectively. Individual widths vary between a minimum value of -0.3923 and a maximum of 0.7793, with a mean of 0.6391 and median of 0.7029 for the entire dataset. The mean values for each of the clusters were 0.68 (cluster 1), 0.47 (cluster 2) and 0.52 (cluster 3). Figure 6 shows that only one sample of cluster 2, which is one of the in-mine samples collected from entries (ENT), had a negative width value, corresponding also to the minimum value of the entire distribution, which indicates misclassification of this sample.

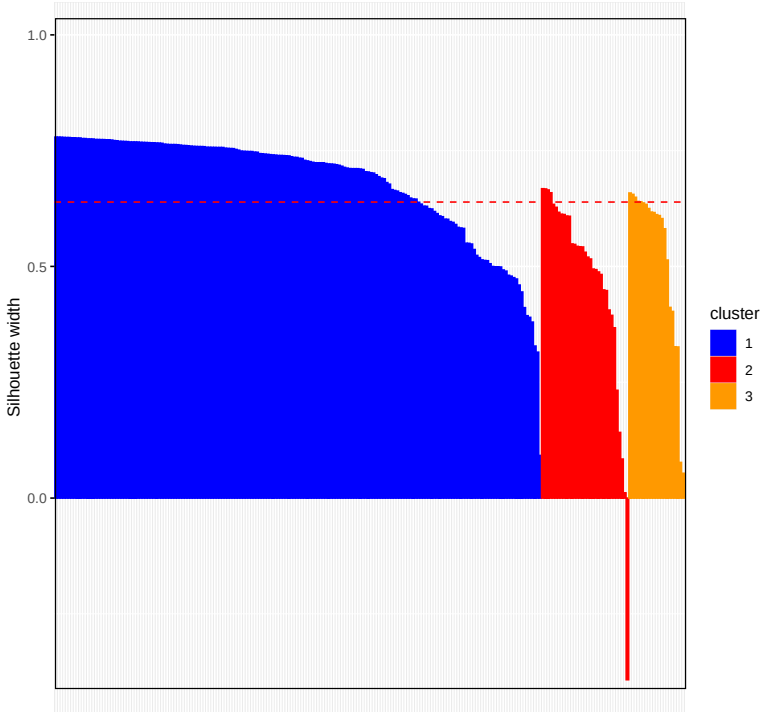

Figure 5. Silhouette width distributions of samples within each of the clusters. Red dashed line is the mean width (0.6391) of all samples.

The three identified clusters were plotted in a clr-biplot, with the addition of the 95% level predictive region (Figure 6), as a measure of uncertainty, within which one would expect 95% of samples of each of the clusters would be included. The assumption here is that the samples in each cluster are log-ratio normal distributed. Figure 6 shows that the clusters are well separated, mostly influenced by the values of N_2, CH_4 and $C2H_6$, and CO_2, indicating the significant differences in composition of the samples from different sources – or sampling locations. Predictive regions in this plot also mean that any future sample collected from these locations will probably fall into the corresponding cluster and will be within the predictive region, and thus can be interpreted as before from a compositional point of view.

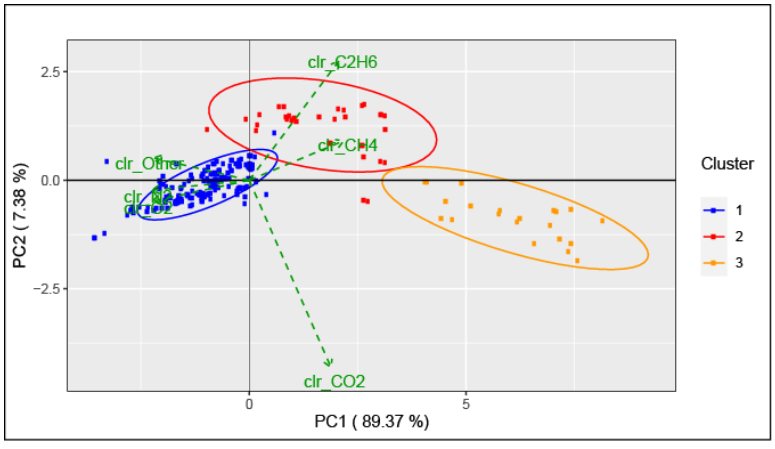

Figure 6. Form clr-biplot of samples and clusters with 95% predictive region (6-part composition).

Another point of interest to understand the compositional changes in mine gas data from a CoDa point of view is to highlight the differences between clusters and the overall dataset, as well as the parts that are responsible for the differences between the clusters, that the clr-biplot is suggesting. Therefore, we looked at the logratio between the geometric mean of individual clusters and the overall geometric mean of the whole data, i.e. the perturbation difference between the center of each cluster from the center of the whole data set, and represented the results in terms of a geometric mean barplot in log-scale. A geometric mean barplot is a representation of perturbation differences ($\ln(g_{k1}/g_1)$, $\ln(g_{k2}/g_2)$,..., $\ln(g_{kD}/g_D)$) where $\mathbf{g}_k=(g_{k1}, g_{k2}, \ldots, g_{kD})$ is the geometric center of the corresponding cluster k, and $\mathbf{g}=(g_1, g_2, \ldots, g_D)$ is the geometric center of the entire data set (Martin-Fernandez et al, 2015). In this plot, when the center of a cluster is equal to the center of the whole data, the ratio of each part is one and the corresponding logarithm is zero. If one part of the center of the cluster is greater or smaller than the corresponding part of the whole center, then the logratio is positive or negative, respectively. Larger bars (positive or negative) will indicate large differences between the geometric means.

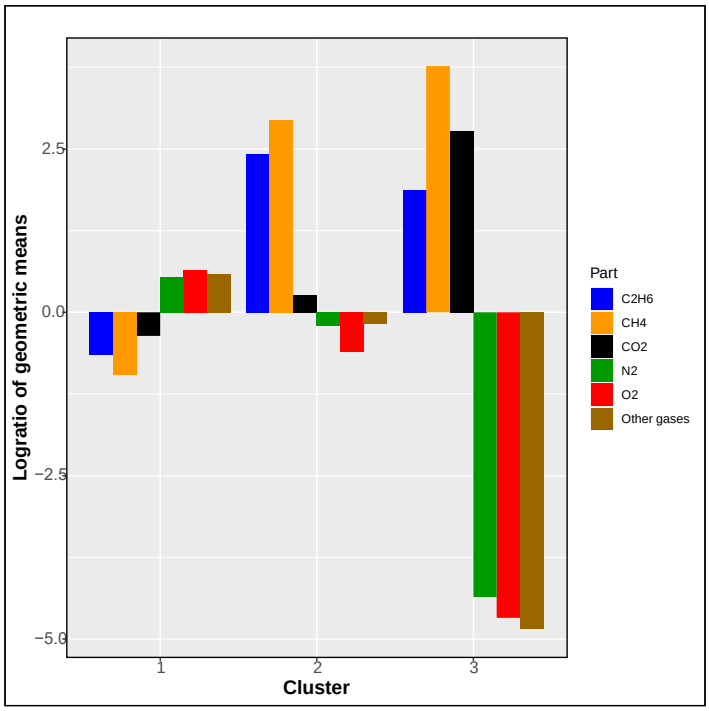

Figure 7. Geometric mean barplot for the gas samples (n=238) with 6-part compositions (data without correction for O_2)

The geometric mean barplot for the 6-part composition of gases is shown in Figure 7 to reflect the compositional differences between clusters in terms of the values of the parts. This figure indicates that the geometric mean of cluster 1 (in-mine gases) is most similar to the overall geometric mean for all parts. This may be due to the large number of samples (183) in cluster 1, and therefore the geometric means of parts are close to the overall geometric mean. In a way, this cluster can be the reference to which others can be compared to for the differences. Consistent with the clr-biplot presented in Figure 6, samples in cluster 1 have, on average, relatively large values in parts O_2, N_2 and other gases (positive bars). For cluster 2 (GGV samples), geometric means of CH_4 and C_2H_6 are larger than overall geometric means whereas the means of other parts are similar to the mean of whole data, or cluster 1. Samples of

cluster 2 are close to the rays associated to these two parts. This is due to increased CH_4 and C_2H_6 from coal and shale beds in GGV gas. In cluster 3, on the other hand, where the PGH-BH and SWC-BH samples are grouped, the geometric means of CH_4, C_2H_6 and CO_2 are larger than the whole data, indicating the enrichment of these samples with coal gas and depletion with components of air and oxidation, as expected. As suggested by the clr-biplot (Figure 6) the samples in cluster 3 have the largest positive values in the first PC-axis, and the positive side is associated to the rays of CH_4, C_2H_6 and CO_2. One note of interest is that C_2H_6's geometric mean is less than the whole data in cluster 3 compared to cluster 2, indicating that this part is reduced in borehole samples, which is probably due to sampling only from coals rather than possible shale beds interlayered with coals. So, this may invoke the question how geometric means of parts may be "shifting" between different clusters. Therefore, we computed the shifts as perturbation differences between the centers of the clusters and we represent them in a geometric mean barplot (Figure 8).

Figure 8 shows the perturbation differences between centers of clusters in log-scale to make the relative shifts clearer. This figure indicates that the composition of borehole samples (SWC-BH and PBG-BH) in cluster 3 shows a general relative increase, especially in CO_2 compared to GGV samples (cluster 2), and a general relative decrease in air and in other gases. The ratio of the centers of borehole samples (cluster 3) to in-mine samples (cluster 1) shows a relative increase in CH_4, C_2H_6 and CO_2 and a decrease of other parts. A similar comparison between GGV samples (cluster 2) and in-mine samples (cluster 1) indicates a relative increase in especially CH_4 and C_2H_6.

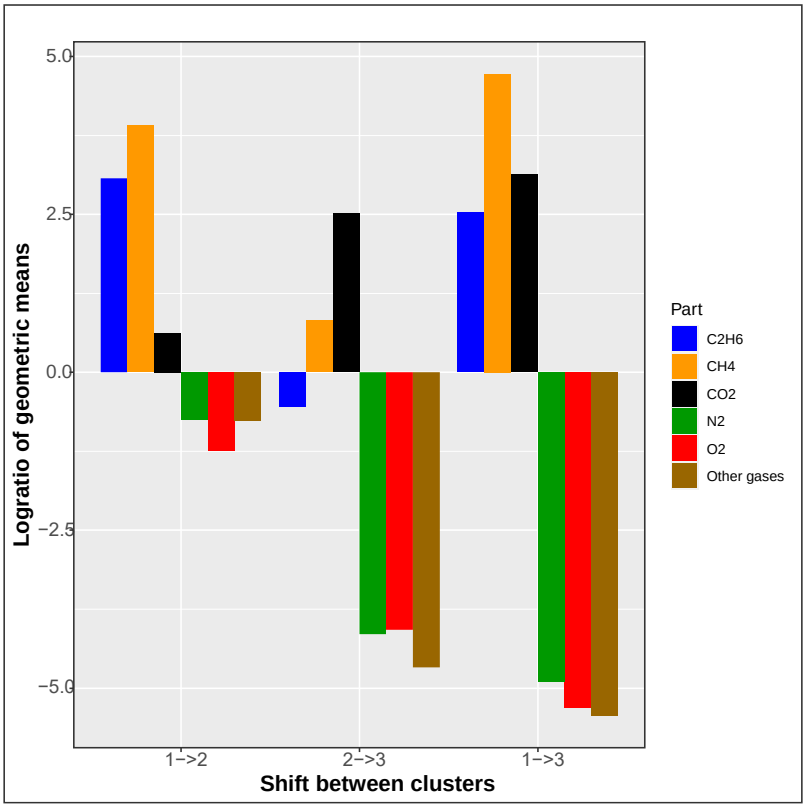

Figure 8. Geometric mean barplot of pairwise difference between centers of clusters for the gas samples with 6-part compositions (data without correction for O_2)

5.2. The dataset after correction for air (5-part composition)

As explained in section 3 and also mentioned in the earlier discussion, in-mine samples had a significant contribution of mine air due to their nature and as evidenced by the presence of high amounts of N_2 and O_2, which is normally not part of the coal gas (Jin et al, 2010). From the statistical point of view, high amounts of air, as O_2 and N_2, in the samples can still be considered as another component that dilutes the other parts to close the composition to 100%. However, from a physical point of view, it may be better to remove it to amplify the contributions of the other parts in the samples in order to be able to analyze the relations better. Also, from the desorbed gas samples of different coals pertinent to this mine, air had already been removed based on N_2 and O_2 ratios in air. Therefore, in order to remove the dilution factor from samples and to be able to include desorbed gas samples in the dataset for analysis, the ratio-based air removal procedure was applied to all samples. The resulting dataset after including desorbed gases had 256 gas samples with 5-part composition.

Table 4. Geometric mean and variation array of pairwise logratios of all samples, including desorbed gas samples, (n=256) after O_2 correction.

Geometric mean

CO_2	N_2	C_2H_6	CH_4	Other gases
8.7563	44.3392	0.4724	39.6529	6.7791

Variation array (Lower diagonal: Expected Value, Upper diagonal: Variance)

	CO_2	N_2	C_2H_6	CH_4	Other gases
CO_2	0.00	4.81	1.70	1.76	7.28
N_2	1.62	0.00	5.02	7.81	1.63
C_2H_6	-2.92	-4.54	0.00	1.29	8.11
CH_4	1.51	-0.11	4.43	0.00	11.75
Other gases	-0.26	-1.88	2.66	-1.77	0.00

Removal of air with the ratio method resulted in the loss of O_2 in the samples with the corresponding reduction of N_2, as seen in its relative abundance in Table 4, which decreased from 77.3% to 44.3%. While coal seam gas usually includes N_2 in its composition and it would still be expected to contain N_2, its relative abundance observed after air removal is high. This can be attributed to overestimation of N_2 due to consumption of O_2 in the samples, as argued by Jin et al (2010), possibly due to oxidation. In addition, in-mine gases interact with expansive coal surfaces in accessible and inaccessible parts of the ventilation system before they are sampled (i.e. at BS, ENT, EP). Generally, the presence of C_2H_4, H_2, CO_2 and CO in mine samples, which are lumped in “other gases” here, are good indicators of oxidation and self-heating, although their concentrations usually remain low unless there is a thermal or spontaneous combustion event (Xie et al, 2011). The presence of these components supports the argument for the consumption of oxygen. Even the GGV samples can show higher than expected N_2 due to loss of oxygen and mixing with in-mine gases before sampling. Therefore, it can be argued that the excess N_2 calculated in these samples reflects both the contribution from coal gas and the balance of N_2 from air due to loss of oxygen. A similar argument can also be put forward for CO_2,

which is a component of coal gas but there may be a contribution from coal oxidation as well. The rest of the components are represented in “other gases”, as explained in section 3. Nevertheless, after air correction and inclusion of desorbed gases, the relative abundance of CO_2, CH_4 and “other gases” increased in the dataset.

Variation array (Table 4) shows that the lowest log-ratio variances, and thus highest associations, are between CH_4 and C_2H_6 (1.29), and “other gases” and N_2. The association between N_2 and “other gases” can be due to the presence of Ar (which was not analyzed for in these samples) and the oxidation components that represent the consumed O_2 and thus excess N_2.

The clr-biplot (Figure 9) shows the associations of different samples with different parts better than the clr-biplot of the 6-part samples. This plot represents 90.6% of the total variability of the data in compositional space. After correction for air and inclusion of desorbed gas samples from different coal seams in the analysis, PC1 represents 75.1% of the variability and PC2 represents 15.5%. Similar to the previous dataset discussed, all in-mine samples are associated with N_2 and “other gases” and are located along the negative values of PC1, and air removal did not change their association with parts in the compositional space. The reason for this association is explained in the previous paragraph. All desorbed coal gas samples and the ones from lateral boreholes (SWC-BH and PGH-BH) are associated more with CH_4 (borehole samples) and CO_2 (desorbed gases) and are aligned on the positive side of the PC1, indicating similar origins of these gases, i.e. coal. Therefore, positive values of PC1 might indicate coal seam gases. However, despite the compositional similarity in terms of source, it is still evident that desorbed gas samples from different coals and their borehole samples tend to group separately in the plot suggesting compositional differences due to source and potential effect of oxidation in desorption canisters. Desorbed gas samples are more associated with CO_2 than CH_4 and the borehole samples are just the opposite. GGV samples, on the other hand, are placed between coal gases and in-mine gases, potentially indicating a compositional mix between coal samples and in-mine samples. Also, they lay along the C_2H_6 ray, which may indicate potential contributions from other sources richer in C_2H_6 to GGV gas, such as shale formations interbedded with coal seams, which have typically higher ethane concentrations than coal gas (Bullin and Krouskop, 2008). Although those formations are not thick and mature enough for economic gas production in the study area at depths relevant to this work, their potential contribution to GEZ and thus to GGV gas may be detected in composition analysis. Unfortunately, there was no gas sample from those interbedded shales available to include in this work to support this. However, considering the size of the GEZ, this is not an unlikely possibility to consider.

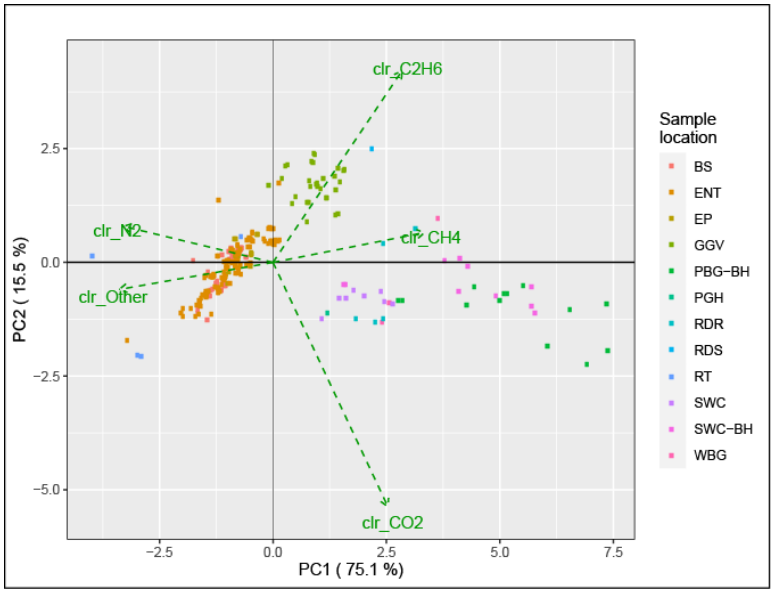

Figure 9. Form clr-biplot of gas samples (n=256) with the parts after O_2 correction and sampling locations.

In order to investigate associations of samples after O_2 removal and inclusion of desorbed gas samples in the dataset, a similar clustering analysis applied to the 6-part data was applied to this dataset too. Results of Ward hierarchical clustering and indices of cluster validation tests are presented in Figures 10 and 11, respectively. Different indices presented in Figure 11 suggest 2, 3, 4 or 5 (CH), 2 or 3 (Dunn) and 2, 3 or 5 (Silhouette) potential clusters in the data. Following the previous approach of looking for a commonly maximized value for all indices and leveraging from what we know about the source of gases, 3 clusters were decided as the optimum cluster number.

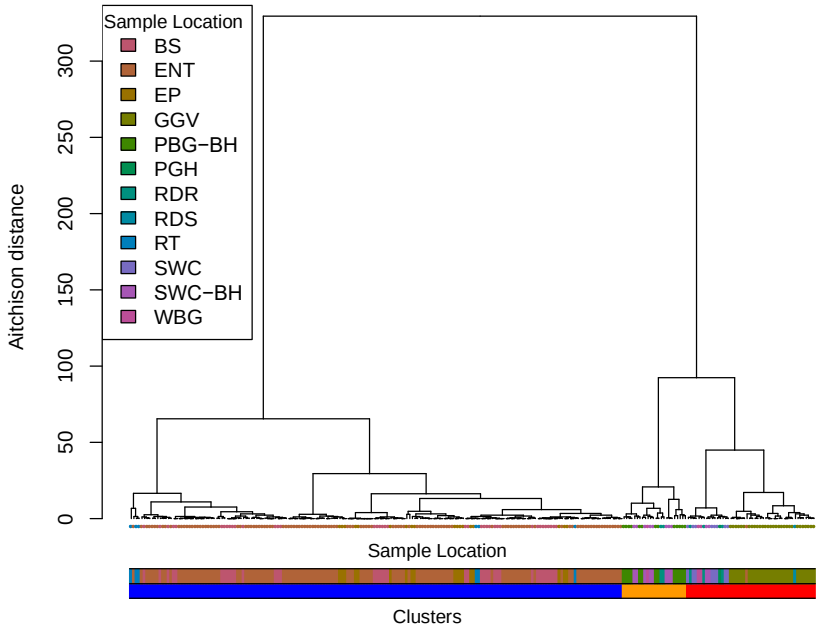

Figure 10. Dendrogram of Ward's hierarchical cluster analysis on 5-part composition generated after correcting for air and inclusion of desorbed gas samples.

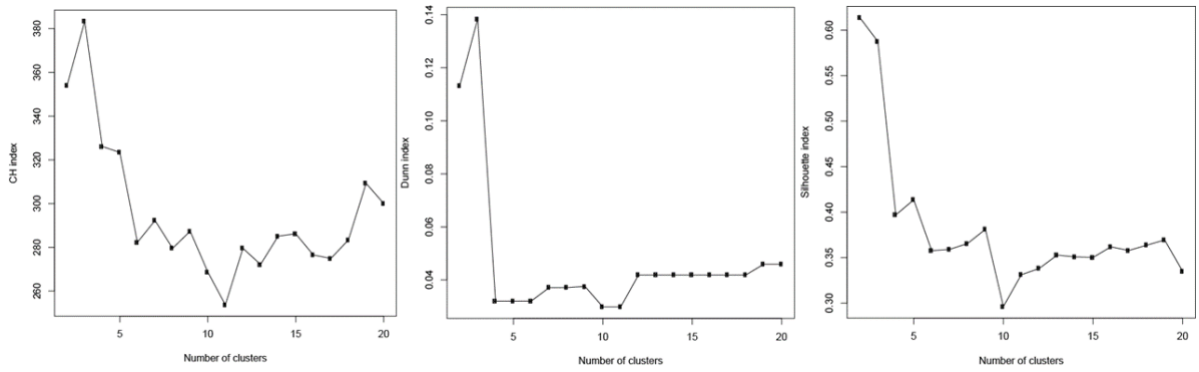

Figure 11. CH, Dunn and Silhouette index plots as functions of number of generated clusters for the gas samples with 5-part compositions after correction for O_2, and inclusion of desorbed gas samples in the dataset.

As with the previous dataset, individual Silhouette widths of each sample were further analyzed for three clusters (Figure 12). These clusters had 184 (in-mine), 48 (GGV and desorbed gas) and 24 (lateral borehole) gas samples. Individual widths varied between a minimum value of -0.3484 and a maximum of 0.7686, with a mean of 0.5874 and median of 0.6667 for the entire dataset. The mean values for each of the clusters were 0.67 (cluster 1), 0.38 (cluster 2) and 0.34 (cluster 3). Figure 12 shows that three samples from PGH-BH, RDR and ENT locations had negative index values, which indicates that the classifications are incorrect.

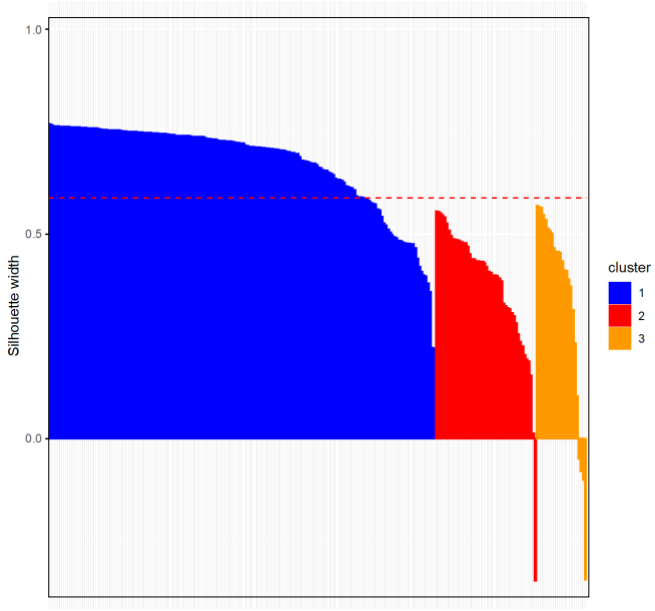

Figure 12. Silhouette width distributions of samples within each of the clusters. Red dashed line is the mean width of all samples (0.5874). The distribution of widths indicates that 3 samples were classified incorrectly (have negative values).

Identified clusters in a clr-biplot, with the addition 95% level predictive region is shown in Figure 13. While the interpretation of the similar plot for 6-part composition (Figure 6) is also valid here in terms of cluster separation and the influence of parts, the clustering results for 5-part composition interestingly shows that GGV samples and desorbed gas samples are located in the same cluster (cluster 2). This is interesting because one would expect the desorbed gas samples to be in the same cluster with the borehole samples rather than with the GGV samples. This may indicate that desorbed gas samples are compositionally more similar to GGV gases in terms of their differences (shifts) from either of the other two clusters. One of the implications of this result is that the composition of desorbed gas samples may not truly represent the composition of the coal gas for resource and inventory calculations.

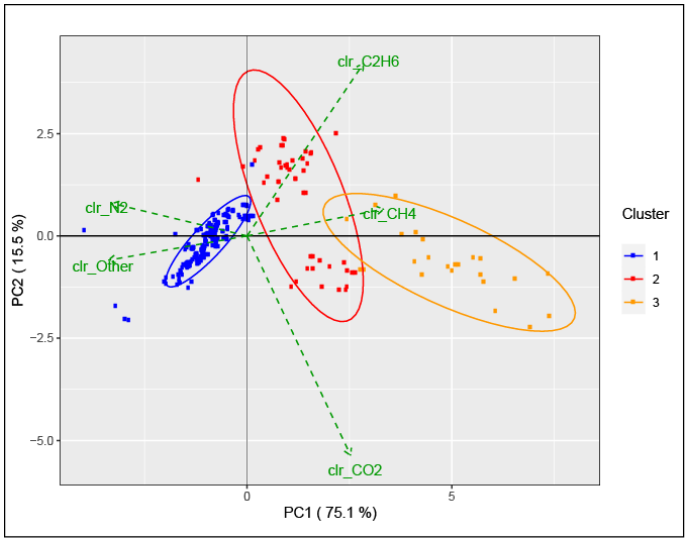

Figure 13. Form clr-biplot of samples and clusters with 95% predictive region (5-part composition).

In-mine samples and borehole samples represent two opposite members in compositional space representing the pure gases and the ones that are heavily contaminated by air and oxidized while circulating within the mine. GGV samples is a mix of these two with potential contribution from gases in other strata. Desorbed gases, on the other hand, do not include in-mine gases but the potential oxidation processes within the canisters separate them from the borehole gases. In this geologic environment, the position of the GGV samples between two main clusters is through the C_2H_6 path. The similarity of compositional distances of GGV and desorbed gas samples puts them apart from the other two groups and joins them in the same cluster (cluster 2).

A similar analysis was performed for the 5-part composition after correcting for air, which gives consistent interpretation with the clr-biplot (Figure 13). The geometric mean barplot for the 5-part composition of gases is presented in Figure 14, which indicates that the geometric mean of cluster 1 (in-mine gases) is similar to the overall geometric mean for all parts, although N_2 and other gases are magnified after removal of O_2. This may indicate consumption of oxygen and a relative increase in oxidation products in these samples, as suggested before. For cluster 2 (GGV samples and the desorbed gas samples), geometric means of CH_4 and C_2H_6 are relatively larger whereas the geometric means of other parts take smaller values than the center of data set. In cluster 3, where the PGH-BH and SWC-BH samples are grouped, the geometric means of CH_4 and CO_2 are larger than the whole data, indicating the relative enrichment of these samples with coal gas and significant depletion of components of air and

oxidation. The perturbation differences between centers of clusters in log-scale were also computed and presented in Figure 15 to show the responsible parts of the shifts between clusters. Figure 15 indicates that the differences between centers of cluster 1 and 2 (in-mine samples and GGV+desorbed gas samples) are mainly created by relatively increases in CH_4 and C_2H_6 and decreases in CO_2, N_2 and other gases, whereas the difference between in-mine samples and borehole samples (clusters 1 and 3) are due to increase in CH_4 and large decreases in N_2 and other gases.

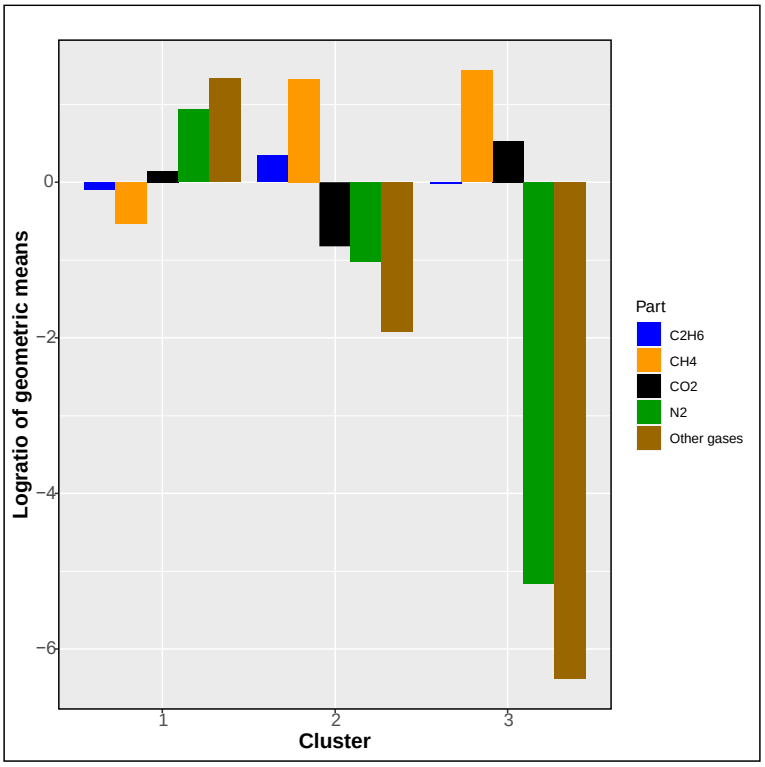

Figure 14. Geometric mean barplot for the gas samples (n=256) with 5-part compositions (data with correction for O_2)

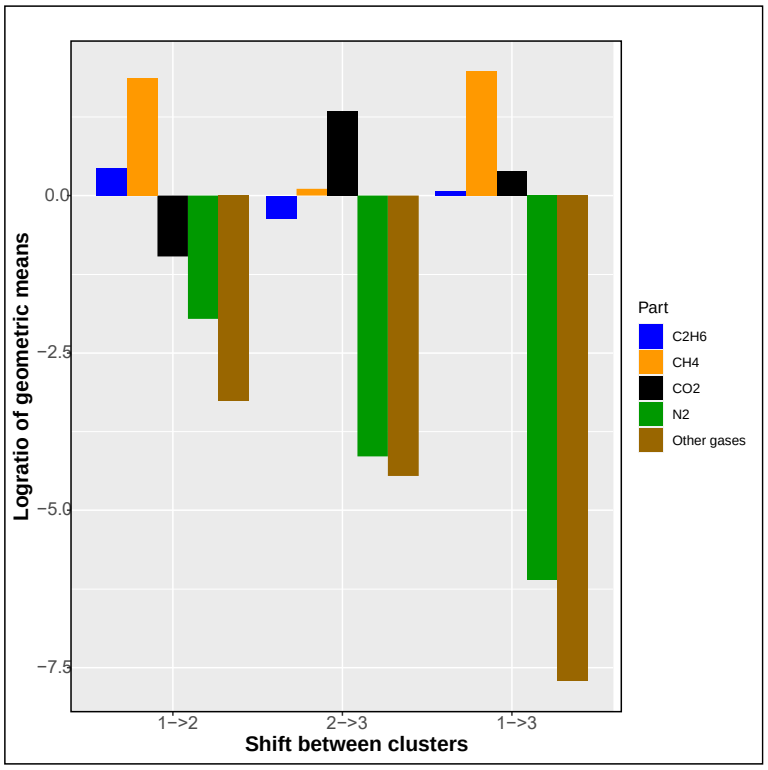

Figure 15. Geometric mean barplot of pairwise difference between centers of clusters for the gas samples with 5-part compositions (data with correction for O_2)

6. Conclusions

Methane in active and abandoned coal mines is a major safety and environmental concern. It originates from the coal seams and other potentially gassy strata within the gas emission zone of mines. Although most monitoring locations and gas control systems are concerned about methane, it is only one of the components in the mine gas mixture that are detectable using analytical methods. Besides methane, mine gas may contain excessive proportional amount of air, other gas hydrocarbons, non-hydrocarbon gases and also oxidation products depending on the sampling location. Therefore, it is important to be able to recognize general characteristics of the gas mixture to identify sources for effective gas control, detect any oxidation events and to design the most effective capture and utilization systems to mitigate methane emissions. CoDa methods may help interpreting the gas compositions to achieve these goals by using adequate statistics that comply with the nature of the data. In this paper, we analyzed mine gases of a longwall coal mine operating in the Northern Appalachian coal basin to investigate the associations of different gas components with sampling locations with and without correction for air.

Results indicated that gas mixtures monitored at different locations had drastic differences. Clr-biplot analyses showed that in-mine gases grouped together regardless of the sampling location. They associated with gas components of air and other gases which contained heavier hydrocarbons as well as oxidation products, rather than any source of methane. Lateral coalbed methane borehole samples, which may indicate the characteristics of pure coal gas, grouped at the opposite of in-mine gases showing more association with CH_4 and C_2H_6. The gases from different coal seams did not show,

however, distinguishable compositional differences possibly due to their genetic similarities in this geology. Gob gas venthole samples, on the other hand, positioned between in-mine samples and the borehole samples, indicating the mixture of the gob gas from in-mine and coal gas.

Analytical removal of air through the use of oxygen and nitrogen ratios confirmed that in-mine samples, in particular, might have been oxidized due to excessive contact with coal which resulted in depletion of oxygen in relation to nitrogen leading to its over estimation. Oxidation was also concluded to decrease oxygen in desorbed gas samples as these samples were grouped separate from the borehole samples.

Cluster analysis, with the use of clustering index, showed all gas samples optimally formed three groups, supporting clr-biplot analyses. Interestingly though, GGV samples and desorbed gas samples positioned in the same cluster, something that might indicate that these samples have similar compositional differences from the two main clusters.

This study is the first, based on our knowledge, demonstrating the application of CoDa analysis to mine gases for interpretation of their compositions. Further applications of CoDa techniques for coalbed, coal mine and shale gas applications can be identification of gas mixtures, their geological associations and sources, especially in cases when there are enough genetic differences (organic material type, depositional environment etc) between sources, especially when isotope analysis is not available. CoDa techniques may further find application in distinguishing the type of the gas (coal vs. shale) in gas storage sites in the event of a leakage.

Acknowledgements

The studied mine and its personnel are gratefully acknowledged for providing the gas samples and for their help. Prof. Vera Pawlowsky-Glahn is also acknowledged for reviewing an earlier version of this paper as part of the USGS review process, which helped improving the content.

Disclaimer

This paper has been peer reviewed and approved for publication consistent with U.S. Geological Survey Fundamental Science Practices (http://pubs.usgs.gov/circ/1367/). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Aitchison, J., 1982. The statistical analysis of compositional data. Journal of the Royal Statistical Society, Series B (Statistical Methodology) 44 (2), 139–177.

Aitchison J. 1986. The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). Reprinted (2003) with additional material by The Blackburn Press, Caldwell, NJ.

Bullin K, Krouskop P. 2008. Composition variety complicates processing plans for US shale gas. Houston: Gas Processors Association.

Bruner KR, Oldham AV, Repine TE, Markowski AK, Harper JA. 1998. Geological Aspects of Coalbed Methane in the Northern Appalachian Coal Basin, Southwestern Pennsylvania and North-Central West Virginia. Open-File Report 98-13. Pennsylvania Geological Survey, Harrisburg PA. 72 pp.

Calinski T and Harabasz J. 1974. A dendrite method for cluster analysis. Communications in Statistics 3(1), 1-27.

Diamond WP, LaScola JC, Hyman DM. 1986. Results of Direct-Method Determination of the Gas Content of U.S. Coalbeds. Information Circular 9067, U.S. Department of the Interior, Bureau of Mines, Pittsburgh PA, 95 pp.

Diamond WP, Schatzel SJ. 1998. Measuring the gas content of coal: A review. International J. of Coal Geology 35 (1–4), 311-331.

Duda A, Krzemień A. 2018. Forecast of methane emission from closed underground coal mines exploited by longwall mining – A case study of Anna coal mine. Journal of Sustainable Mining 17, 184-194.

Dunn J. 1974. Well separated clusters and optimal fuzzy partitions. J. Cybern. 4 (1), 95– 104.

Edjabou ME, Martín-Fernández JA, Scheutz C, Astrup TF. 2017. Statistical analysis of solid waste composition data: Arithmetic mean, standard deviation and correlation coefficients. Waste Management 69, 13-23.

Egozcue JJ, Pawlowsky-Glahn V, Gloor GB. 2018). Linear Association in Compositional Data Analysis. Austrian Journal of Statistics 47, 3-31.

Engle MA, Reyes FR, Varonka MS, Orem WH, Ma L, Ianno AJ, Schell TM, Xu P, Carroll KC. 2016. Geochemistry of formation waters from the Wolfcamp and "Cline" shales: Insights into brine origin, reservoir connectivity, and fluid flow in the Permian Basin, USA. Chemical Geology 425, 76-92.

Filzmoser P, Hron, K, Reimann, C. 2009. Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. Sci. Total Environ. 407, 6100–6108. http://dx.doi.org/10.1016/j.scitotenv.2009.08.008.

Global Methane Initiative. 2020. Global Methane Emissions and Mitigation Opportunities. https://www.globalmethane.org/documents/gmi-mitigation-factsheet.pdf.

Hummel JA, Ruiz FA, Kelafant JR. 2018. Quantifying the benefits of coal mine methane recovery and use projects: Case study on the application of in-mine horizontal pre-drainage boreholes at gassy coal mines in India and the optimization of drainage system design using reservoir simulation. Environmental Technology & Innovation 10, 223-234.

IPCC, 2013. Stocker T.F., D. Qin, G.-K. Plattner, et al. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1535 pp.

Jin H, Schimmelmann A, Mastalerz M, Pope J, Moore TA. 2010. Coalbed gas desorption in canisters: Consumption of trapped atmospheric oxygen and implications for measured gas quality. Int. J. Coal Geology 81, 64–72.

Karacan CÖ, Warwick PD. 2019. Assessment of coal mine methane (CMM) and abandoned mine methane (AMM) resource potential of longwall mine panels: example from Northern Appalachian Basin, USA. Int. J. of Coal Geology 208, 37-53.

Karacan CÖ, Olea, RA., Goodman GVR. 2012. Geostatistical modeling of gas emission zone and its in-place gas content for Pittsburgh seam mines using sequential Gaussian simulations. International Journal of Coal Geology 90-91, 50-71.

Karacan CÖ, Ruiz FA, Cotè M, Phipps S. 2011. Coal mine methane: a review of capture and utilization practices with benefits to mining safety and to greenhouse gas reduction. Int. J. Coal Geology, 86, 121-156.

Karacan CÖ, Goodman GVR. 2011. Probabilistic modeling using bivariate normal distributions for identification of flow and displacement intervals in longwall overburden, International Journal of Rock Mechanics and Mining Sciences, 48, 27-41.

Karacan CÖ. 2009. Degasification system selection for U.S. longwall mines using an expert classification system. Computers and Geosciences, 35, 515-526.

Kholod N, Evans M, Pilcher RC, Roshchanka V, Ruiz F, Cote M, Collings R. 2020. Global methane emissions from coal mining to continue growing even with declining coal production. Journal of Cleaner Production 256, 120489.

Kędzior S, Dreger M. 2019. Methane occurrence, emissions and hazards in the Upper Silesian Coal Basin, Poland. Int. J. Coal Geology 211, 103226.

Krog RB. 2016. Critical Analysis of Longwall Ventilation Systems and Removal of Methane. PhD Dissertation. West Virginia University, Morgantown, WV. 184 pp.

Markowski AK. 2001. Reconnaissance of the Coalbed Methane Resources in Pennsylvania. Mineral Resource Report 95. Pennsylvania Geological Survey, Harrisburg PA. 134 pp.

Martín-Fernández JA, Daunis-i-Estadella J, Mateu-Figueras G. 2015. On the interpretation of differences between groups for compositional data. Statistics and Operations Research Transactions 39(2), 231-252.

MSHA, 2006. Report of investigation, Fatal Underground Coal Mine Explosion January 2, 2006. Sago Mine, Wolf Run Mining Company http://www.msha.gov/Fatals/2006/Sago/ftl06C1-12.pdf.

National Academies of Sciences, Engineering, and Medicine. 2018. Improving Characterization of Anthropogenic Methane Emissions in the United States. Washington, DC: The National Academies Press. https://doi.org/10.17226/24987.

Otero N, Tolosana-Delgado R, Solera A, Pawlowsky-Glahn V, Canals A. 2005. Relative vs. absolute statistical analysis of compositions:A comparative study of surface waters of a Mediterranean river. Water Research 39, 1404–1414.

Palarea-Albaladejo J, Martín-Fernández JA, and Soto JA. 2012. Dealing with Distances and Transformations for Fuzzy C-Means Clustering of Compositional Data. Journal of Classification 29(2), 144–169.

Palchik V. 2003. Formation of fractured zones in overburden due to longwall mining. Environmental Geology 44(1), 28-38.

Pawlowsky-Glahn V, Egozcue J, Tolosana-Delgado R. 2015. Modeling and Analysis of Compositional Data. John Wiley & Sons.

Rousseeuw P J. 1987. Silhouettes : a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53-65.

Sechman H, Kotarba MJ, Fiszer J, Dzieniewicz M. 2013. Distribution of methane and carbon dioxide concentrations in the near-surface zone and their genetic characterization at the abandoned “Nowa Ruda” coal mine (lower Silesian Coal Basin, SW Poland). Int. J. Coal Geol. 116-117, 1–16.

Sechman H, Kotarba MJ, Kędzior S, Dzieniewicz M, Romanowski T, Twaróg. 2019. Distribution of methane and carbon dioxide concentrations in the near surface zone, genetic implications, and evaluation of gas flux around abandoned shafts in the Jastrzębie-Pszczyna area (southern part of the Upper Silesian Coal Basin), Poland. Int. J. Coal Geol. 204, 51–69.

Schatzel SJ, Karacan CÖ, Dougherty HN, Goodman GVR. 2012. An analysis of reservoir conditions and responses in longwall panel overburden during mining and its effect on gob gas well performance. Engineering Geology 127, 65-74.

Swolkień J. 2020. Polish underground coal mines as point sources of methane emission to the atmosphere. International Journal of Greenhouse Gas Control 94, 102921.

Su DWH, Zhang P, Schatzel SJ, Gangrade V, Watkins E. (2019). Longwall-Induced Subsurface Deformations and Permeability Changes Shale Gas Well Casing Integrity Implication. In: Proceedings of the 38th International Conference on Ground Control in Mining. Morgantown, WV: West Virginia University, pp 49-59.

Tolosana-Delgado R, McKinley J. 2016. Exploring the joint compositional variability of major components and trace elements in the Tellus soil geochemistry survey (Northern Ireland). Applied Geochemistry 75, 273-276.

Voltattorni N, Lombardi S, Beaubien SE. 2015. Gas migration from two mine districts: The Tolfa (Lazio, Central Italy) and the Neves-Corvo (Baixo Alentejo, Portugal) case studies. Journal of Geochemical Exploration 152, 37-53.

Wang K, Zhang J, Cai B, Yu S. 2019. Emission factors of fugitive methane from underground coal mines in China: Estimation and uncertainty. Applied Energy 250, 272-282.

Watkins E, Karacan CÖ, Schatzel SJ, Gangrade V. 2020. Field determination of strata permeability responses to longwall-induced deformations at the location of a hypothetical shale gas well drilled through a longwall abutment pillar. Natural Resources Research (in review).

Xie J, Xue S, Cheng W, Wang G. 2011. Early detection of spontaneous combustion of coal in underground coal mines with development of an ethylene enriching system. Int. J. Coal Geology 85, 123-127.