

Article

Data Fusion for the Prediction of Elemental Concentrations in Polymetallic Sulphide Ore Using Mid-Wave Infrared and Long-Wave Infrared Reflectance Data

Feven Desta ^{1,*}, Mike Buxton ¹ and Jeroen Jansen ²

¹ Resource Engineering Section, Department of Geoscience and Engineering, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands; m.w.n.buxton@tudelft.nl

² Department of Analytical Chemistry: Chemometrics, Faculty of Science, Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands; jj.jansen@science.ru.nl

* Correspondence: f.s.desta@tudelft.nl; Tel.: +31-15-278-4837

Received: 3 January 2020; Accepted: 3 March 2020; Published: 5 March 2020



Abstract: The increasing availability of complex multivariate data yielded by sensor technologies permits qualitative and quantitative data analysis for material characterization. Multivariate data are hard to understand by visual inspection and intuition. Thus, data-driven models are required to derive study-specific insights from large datasets. In the present study, a partial least squares regression (PLSR) model was used for the prediction of elemental concentrations using the mineralogical techniques mid-wave infrared (MWIR) and long-wave infrared (LWIR) combined with data fusion approaches. In achieving the study objectives, the usability of the individual MWIR and LWIR datasets for the prediction of the concentration of elements in a polymetallic sulphide deposit was assessed, and the results were compared with the outputs of low- and mid-level data fusion methods. Prior to low-level data fusion implementation, data filtering techniques were applied to the MWIR and LWIR datasets. The pre-processed data were concatenated and a PLSR model was developed using the fused data. The mid-level data fusion was implemented by extracting features using principal component analysis (PCA) scores. As the models were applied to the MWIR, LWIR, and fused datasets, an improved prediction was achieved using the low-level data fusion approach. Overall, the acquired results indicate that the MWIR data can be used to reliably predict a combined Pb–Zn concentration, whereas LWIR data has a good correlation with the Fe concentration. The proposed approach could be extended for generating indicative element concentrations in polymetallic sulphide deposits in real-time using infrared reflectance data. Thus, it is beneficial in providing elemental concentration insights in mining operations.

Keywords: MWIR; LWIR; data fusion; chemometrics; sulphide ore; iron; lead–zinc

1. Introduction

The demand for accurate and reliable resource estimation in mining operations is growing rapidly, as it is advantageous to be able to obtain more readily information on the potential economic and environmental operability of commercial mines. This requires the use of different sensor technologies for raw material characterization. Each sensor technology operates over a certain range of the electromagnetic spectrum, associated with a subset of molecules or chemical structures of potential interest for mining extraction. However, a single sensor might not provide a sufficiently comprehensive description of a material's composition. This introduces uncertainty into both resource estimation and defining requirements for mineral processing. Thus, it is necessary to utilize strategic sensor

combinations to improve accuracy, minimize uncertainty, and enhance specific inferences for raw material characterization.

Sensor technologies may be integrated physically or digitally (through data fusion). Physical integration is achieved when two or more sensors are combined on a single platform. For example, Sharma et al. [1] integrated a remote-laser-induced breakdown spectroscopy (LIBS) and Raman system for analysis of minerals (e.g., carbonates, sulphates, iron oxides) with a single 532 nm laser. The integrated system was capable of simultaneously providing elemental and mineralogical information. Fernández et al. [2] physically integrated RGB and multispectral imagery sensors for an improved classification of the Cabernet Sauvignon grapevine elements. Hoehse et al. [3] demonstrated the utility of mapping heterogeneous minerals using a single unified LIBS and Raman setup for a combined analysis of molecular and elemental information that resulted in a comprehensive material characterization.

Data fusion is a process of integrating data blocks from multiple sources or sensors into a single comprehensive model [4]. It can be implemented using probabilistic descriptions of observations and processes. One such method is the Bayes' rule, which is the most commonly employed method in robotics [5] and military applications [6]. However, when probabilistic methods are employed, a large number of probabilities needs to be specified, thus introducing inconsistencies into the specification of a consistent set of beliefs in terms of probability and model precision. Even though these limitations can be partially mitigated using alternative techniques, such as fuzzy logic [7], this approach was not considered in the present study due to its high complexity.

Data fusion can also be implemented using chemometric techniques, which require the collected data to have a relatively high quality. This approach is widely applied in analytical chemistry and metabolomics [8–12]. Data fusion using chemometrics can be realized at three levels, designated as low-level, mid-level, and high-level fusion. Low-level data fusion involves the integration of multiple data sources by concatenating data blocks of different nature [13–15]. In this approach, correlations among all variables collected in the different data blocks are considered. This results in a single matrix, which yields a single classification or prediction model. Mid-level or hierarchical data fusion is a feature-level fusion that involves variable screening [9,15]. Prior to the fusion process, important variables are extracted separately as features from each data source using decomposition (or resolution) or variables selection techniques [16]. These features can subsequently be used in classification and prediction. One of the advantages of mid-level data fusion is that it reduces data dimensionality, thereby allowing for data blocks to be treated individually, avoiding the influence of other datasets. Mid-level data fusion, however, requires an optimal combination of extracted features that describe most of the variation in the data. Whereas both low- and mid-level data fusion methods combine the data sources at the data level, high-level data fusion is a decision-level fusion that combines model outputs to produce a final fused response. For each available block of data, separate models are developed and their individual responses are combined to produce a final fused response [10].

Low-level data fusion has been widely used to authenticate the origin of foods and beverages and assess their quality. Available evidence indicates that low-level data fusion results in better prediction and classification models than individual techniques. For example, low-level data fusion has been shown to improve the quality assessment of wines [17], the identification of the origin of olive oils [18], the predictive model accuracy of the quality parameters of soybean flour [19], and apples characterization [20,21], as well as to enhance the robustness in the discrimination of coffee varieties [22]. Mid-level data fusion approaches have also been proven to be useful in a wide range of applications, such as in the identification of food fraud [23], analysis of geographic traceability of soil samples [11], and discrimination of different types of beer [24,25]. To date, high-level fusion has mainly been applied to classification tasks pertaining to, for example, fruit quality assessment [24,26] and identification of food fraud [23]. Studies indicated that the use of data fusion for mineralogical investigations resulted in enhanced characterization of materials. For example, a low-level fusion of LIBS and visible/near-infrared (VNIR) data resulted in enhanced quantitative identification of

hematite, magnetite, and quartz minerals in tailing slurries of an iron ore [27]. However, at present, the application of data fusion using chemometrics for mineralogical investigations remains very limited.

Sensor technologies provide qualitative, quantitative, and semi-quantitative information on different aspects of material properties. Such information can be broadly categorized into geochemistry, mineralogy, and texture but is not limited to these attributes of materials. For example, technologies such as LIBS, X-ray fluorescence (XRF), and inductively coupled plasma optical emission spectrometry/mass spectrometry (ICP-ES/MS) provide qualitative and quantitative geochemical data, whereas infrared (IR) technologies (e.g., mid-wave infrared (MWIR) and long-wave infrared (LWIR)) measure the signals that result from the molecular bond vibrations of IR active minerals. Each molecule in the IR active minerals follows different vibrational modes such that the IR can be analyzed to acquire information on various minerals. Therefore, IR technologies provide mineralogical information, and from the different IR analytical methodologies, the choice of the IR region depends on the type of mineral under investigation. In this study, the MWIR and LWIR technologies that operate in the wavelength ranges of 2.5–7 μm and 7–15 μm , respectively, were considered. LWIR can provide accurate identification of rock-forming minerals [28,29]. The MWIR region of the electromagnetic spectrum is under-investigated due to limited instrumental development. However, recent studies [30,31] showed that the MWIR region of the electromagnetic spectrum has a good potential for the characterization of materials.

Infrared spectroscopy is a well-known analytical technique that responds to the concentration of the analyte. The measurements include instrumental artefacts, which are generally compensated for by data pre-processing. Pre-processing techniques generally use very simple models, based on first principles knowledge, to digitally remove such artefacts. However, it has proven to be very challenging to find the (most) optimal pre-processing technique for a dataset—i.e. that provides the best chemical predictions [32]. Of specific interest in data fusion is, however, whether the pre-processing that is most effective on the dataset of one sensor is the same as for data collected with another sensor. As predictions based on spectroscopic data are generally performed with multivariate calibration that takes correlations between all variables into account, pre-processing may affect these correlations, both within data from one platform but also in data from different platforms.

Mineralogical techniques (MWIR and LWIR) have not been previously employed for the indication of element concentrations in polymetallic sulphide ore. This gap in the current analytical approaches and the promising findings reported recently [30] have motivated the present study. The objective of this investigation was twofold: to develop prediction models that indicate the elemental concentration in polymetallic sulphide ore using the mineralogical techniques (i.e., MWIR and LWIR reflectance spectral data) separately, and to assess the improvements in the predictive models using low-level and mid-level data fusion approaches.

2. Materials and Data Acquisition

2.1. Materials

2.1.1. Case Study Area

The case study is the Reiche Zeche underground mine located in the Freiberg district, eastern Erzgebirge, Germany. It is one of the oldest mines in Europe, mined for silver, copper, lead, arsenic, zinc, and pyrite [33,34]. It ceased operating in 1969. The “Reiche Zeche” and “Alte Elisabeth” shafts were reconstructed and reopened as a research and teaching mine in 1976. The area is characterized by a polymetallic vein-type hydrothermal deposit formed by two mineralization events of Late-Variscan and Post-Variscan age [35]. The Late-Variscan mineralization is rich in iron, sulphur, zinc, lead, and copper. Typical ore minerals include pyrite, galena, arsenopyrite, chalcopyrite, and sphalerite, along with quartz and minor carbonate gangue. The Post-Variscan mineralization event consists of a fluorite barite/lead ore assemblage, mainly containing sphalerite, pyrite, galena, chalcopyrite, and marcasite, as well as quartz, fluorite, carbonates, and barite as gangue [35,36]. The ore minerals of this younger mineralization event contain smaller amounts of Cu, Zn, and Fe, while it is rich in Ca

and Mg. For the present study, elements mainly sourced by ore minerals (e.g., sphalerite, galena, and chalcopyrite) and waste materials (e.g., carbonates) were considered. These are elements of economic interest (Cu, Pb, Zn, and Fe), penalty elements in mineral processing (As), and elements mainly sourced by waste materials (Ca and Mg).

2.1.2. Samples

One hundred and seventeen representative samples were acquired from the test case area. These samples included 62 channel samples acquired from the mine face, 45 samples that were systematically collected from eight drill cores each having a length of approximately 2 to 3 m, as well as 10 muck pile samples collected after blasting part of the mine face (Figure 1). Channel sampling is advantageous for capturing different lithotypes and variations in their abundance and distribution. Drill core samples provide continuous information across the entire drill profile depth, while the blasted rock pile (muck pile) samples are disturbed samples since the rock fragments are moved from their original location. The collected samples were powdered and the measurements were performed using the powdered samples.

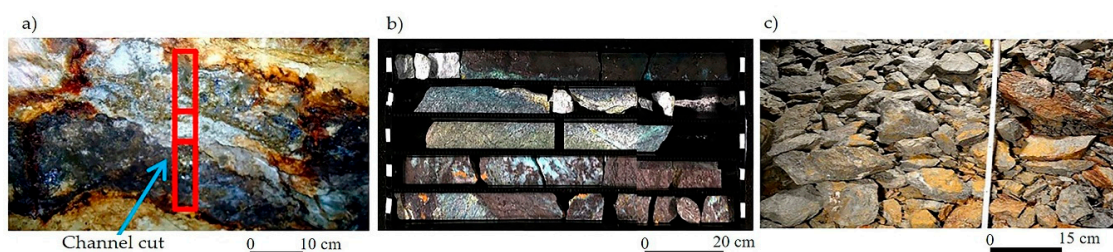


Figure 1. Images of (a) the mine face showing a channel cut with different intervals (the red boxes); (b) drill core samples; and (c) muck pile samples.

2.2. Instrumentation and Datasets

2.2.1. MWIR and LWIR

The Agilent 4300 Fourier-transform infrared spectroscopy (FTIR) analyzer (Edinburgh, UK) was used to acquire reflectance spectral data using the collected 117 samples. The FTIR analyzer obtains spectra over the ~ 1.9 to ~ 15.0 μm wavelength range. The instrument is a portable handheld device that provides point data at high data acquisition speed (less than 30 s). Its relatively lightweight (under 2.2 kg) and compact dimensions ensure its effective use in a wide range of in-situ applications. The spectral profile of each sample was acquired as the mean of 64 sample scans at a resolution of 4 cm^{-1} using the diffuse reflectance-sampling interface. The instrument was calibrated every 15 min with 126 background scans. For the test case materials, the spectral range from 1.9 to 2.5 μm yielded noisy results and was excluded from further analyses. Three sub-datasets were prepared prior to modeling: the full FTIR data (excluding the range from 1.9 to 2.5 μm), the MWIR (2.5 to 7 μm) data, and the LWIR (7 to 15 μm) data. To accommodate the samples' heterogeneity, multiple spectra (7 to 10 measurements) were collected from each sample and the averages were computed.

2.2.2. Chemical Analyses

Conventional data acquisition techniques—X-ray fluorescence (XRF) and inductively coupled plasma mass spectrometry (ICP-MS)—were used to obtain the data that were used as response variables and in the validation of the element concentrations prediction results. The XRF and the ICP-MS analysis were performed at Acme analytical laboratory, Ankara, Turkey. The two techniques provide quantitative elemental information. The datasets of Pb, Zn, and Cu comprised XRF and ICP-MS measurements performed using 117 samples. The analysis of As, Fe, Mg, and Ca was performed using the ICP-MS technique and 89 samples. In this study, the number of samples analyzed using the

FTIR analyzer was 117. All 117 samples were analyzed for Cu, Pb, and Zn. However, 89 out of the 117 samples were analyzed for Fe, As, Ca, and Mg. Therefore, the same samples were used; however, the number of samples used to analyze the elements differed. Therefore, the Cu, Pb, and Zn dataset consisted of 117 sample measurements whereas the Fe, As, Ca, and Mg dataset consisted of 89 sample measurements. Due to their co-occurrence, Pb and Zn are mainly mined simultaneously. Thus, Pb and Zn were combined (the concentrations of the two elements were summed) in this study. The analysis results pertaining to 117 powder samples of Cu and the combined Pb–Zn and 89 powder samples of Fe, As, Ca, and Mg are discussed in the sections that follow.

3. Methodology

Prediction of elemental concentration in polymetallic sulphide ore using MWIR and LWIR reflectance spectra requires an integrated methodology. Therefore, the methodological approach developed as a part of the present study comprises a multi-step process that commences with data exploration, followed by data pre-processing and data fusion, and culminates with data modeling and model validation (Figure 2). The unscrambler software was employed for the analysis. The data exploration task includes outlier detection and data splitting (i.e., into calibration and validation datasets). In the present study, data pre-processing was performed using different data filtering techniques. The pre-processed data were subsequently used to develop a series of calibration and prediction models for MWIR and LWIR datasets separately. The low-level (the red box in Figure 2) and mid-level (the green box in Figure 2) data fusion approaches were implemented using the concatenated pre-processed MWIR and LWIR data blocks, and the features extracted from the two data blocks, respectively. All prediction models were validated using independent datasets. The use of the individual datasets and the fused dataset for the prediction of element concentrations in polymetallic sulphide ore was evaluated and compared. Each of these steps is described in detail below.

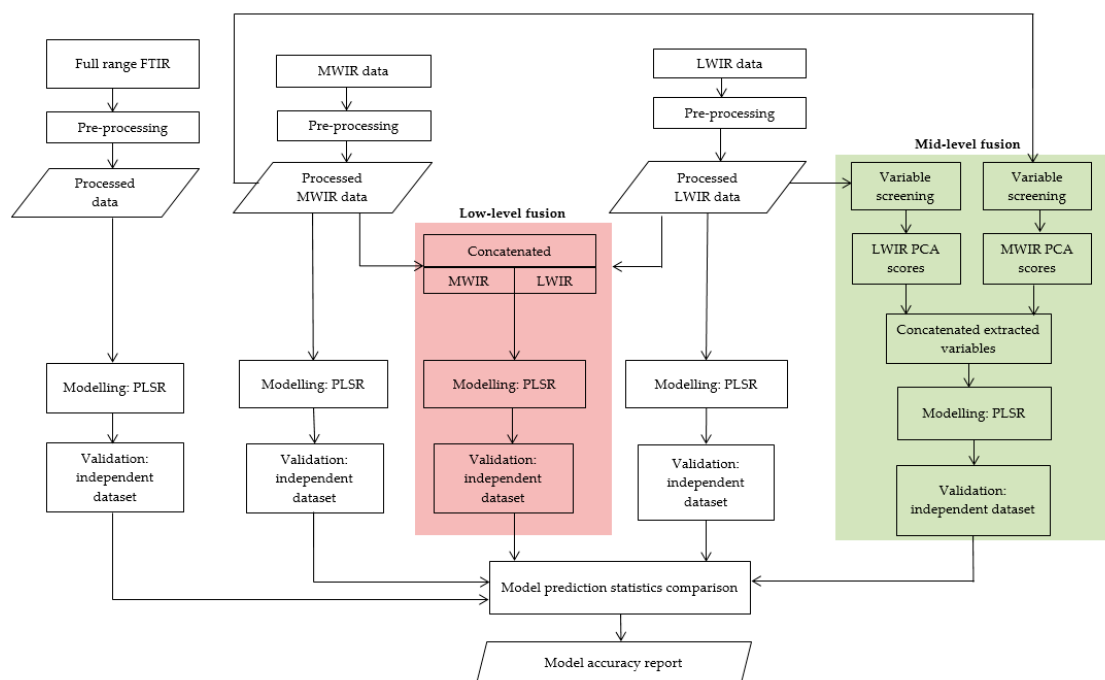


Figure 2. A workflow diagram depicting the steps of the research approach.

3.1. Data Exploration

3.1.1. Descriptive Model—Principal Component Analysis (PCA)

PCA is one of the most powerful explanatory data analysis techniques that can be applied to identify patterns in samples, potential subtle outliers, and informative variables (i.e., those responsible for the observed patterns). In this study, PCA models were developed using the full-range FTIR data. Potential outliers were identified using Hotelling's T^2 of the PCA model with the critical limit (p -value) of 5%. Therefore, the 95% confidence ellipse was included in the score plots of the PCA models to reveal potential outliers (i.e., data points located outside the ellipse contour). Potential outliers were further investigated using the influence plot. This plot is useful for detecting influential samples and dangerous outliers. For example, samples with high leverage and high residual variance are deemed to contain the most dangerous outliers. The pattern of the spectral data with respect to the elemental concentration was assessed using PCA.

3.1.2. Covariance between Elements and IR Spectra

The proportion of the variance in the y -variable (the quantitative elemental data) that is predictable from the x -variables (full-range FTIR reflectance spectra) was computed using partial least squares regression (PLSR) and each target element, separately. Elements that showed a weak linear relationship with the spectral data (a coefficient of determination less than 0.4) were excluded from further analysis.

3.1.3. Data Splitting

For the combined Pb–Zn dataset, the full-range FTIR, MWIR, and LWIR datasets, each containing 113 samples, were split into calibration (72 measurements) and validation (35 measurements) subsets using a random sample selection algorithm. To encompass the entire variability domain and allow for direct model comparisons, the same split was maintained in the calibration and validation datasets pertaining to the three data blocks. The same procedure was followed for the Fe dataset to obtain calibration (64 measurements) and validation (23 measurements) datasets.

3.2. Pre-Processing

Data from different sensors should be treated based on the specific characteristics of the spectral data and the ultimate goal of data analysis in order to remove undesired variations (e.g., instrumental artefacts). Accordingly, prior to modeling, the data matrices were treated with normalization, baseline correction, mean centering (MC), standard normal variate (SNV), and smoothing (Gaussian filter smoothing). Normalization removes undesired intensity variation caused by multiplicative effects [37]. Baseline correction is a signal correction method that subtracts the unwanted spectral background from the main signal information. MC represents variation around a mean by subtracting the sample mean from each data value [37,38]. MC was performed in combination with each pre-processing technique. SNV normalizes by deducting the spectrum mean value from each variable in the spectrum and dividing each resulting value by the spectrum standard deviation [37,39]. It minimizes the light scattering effect and particle size effects from spectral data, and is commonly used with diffuse reflectance data. Smoothing is based on averaging the neighborhood points in order to minimize random noise [37].

3.3. Prediction Models

PLSR is a data prediction technique commonly applied in multivariate data analysis. It maximizes the covariance between the predictor (X) and the response (Y) blocks. It also allows for the identification of new predictor variables, called factors, and the optimal factors for the proposed linear relationship [40]. In this study, PLSR models were developed using the IR spectral data as explanatory variables (X) and the concentration of elements as response variables (Y). Models were developed for the prediction of

each element concentration, separately. Thus, the usability of the IR spectra for the prediction of the individual element under investigation was assessed. A series of PLSR models were developed using the pre-processed data. The models were developed using calibration datasets and the performance of the models was validated using independent datasets (validation datasets). The predictive performance of the models was evaluated in terms of statistical error terms (root mean square error of cross validation (RMSECV) and root mean square error of prediction (RMSEP)) and the coefficient of determination (R^2) of the prediction values. RMSECV is calculated on the test split using a 10-fold cross validation. The optimum number of PLS factors was determined using the RMSECV values. RMSEP represents the prediction error based on a comparison of real cases with reference values, in this case, an independent dataset. In sum, RMSEP indicates how well the model built using calibration data performs when applied to unknown cases. R^2 describes the percentage of variance in a dependent variable that is explained by the independent variable(s) in the model. The optimal prediction model is the one with the minimum RMSECV and RMSEP and a higher R^2 value.

3.3.1. Model Optimization and Validation

Model optimization was performed taking into account the different data filtering techniques adopted in this work. The optimal pre-processing technique for an improved prediction of elemental concentrations was selected based on the performances of the calibration models and the predictive ability of the models after each pre-processing technique applied to the datasets. Therefore, the RMSECV, RMSEP, and R^2 values were used to evaluate the performances of each model. Model validation is important to ensure that models are not prone to overfitting. Thus, validation using independent datasets was performed for the five datasets, namely MWIR, LWIR, full-range FTIR, the concatenated pre-processed MWIR and LWIR data, and the concatenated extracted new scores of variables from the two data blocks. The methodological approaches for the prediction of elemental concentrations using the datasets mentioned above are described in the following sections.

3.3.2. MWIR and LWIR Data Models

PLSR models were developed using the MWIR and LWIR reflectance spectral data separately. A series of models was developed using the pre-processed MWIR data after applying the aforementioned data-filtering techniques. Similarly, a series of prediction models was developed using the pre-processed LWIR data. The prediction performance of each model was evaluated using independent datasets. The predictive performances of MWIR and LWIR data models for the prediction of elemental concentrations were compared.

3.3.3. Low-level Data Fusion

Potential correlations between wavelengths in different IR ranges (e.g., MWIR and LWIR) require data fusion to capture latent information that cannot be extracted by analysis of individual data blocks separately. In low-level data fusion, the way in which the data matrices from different sources are concatenated, normalized, and scaled, and the variation in variable magnitude, potentially have a profound effect on model performance in terms of both prediction and classification. For the choice of the optimal model, a “systematic approach” is recommended to understand the effect of the pre-processing techniques on model accuracy [32]. The pre-processed data from the MWIR and LWIR data blocks were concatenated to form a series of new data blocks. The unified (fused) data were mean-centered and a single model that provides the final elemental concentration prediction was developed (indicated by the red box in Figure 2). The PLSR models were developed using calibration datasets and were validated using the independent datasets. The pre-processing techniques applied to the datasets prior to modeling resulted in different model performances. Thus, the optimal pre-processing method was selected by comparing the performance of the prediction models using RMSECV, RMSEP, and R^2 values.

For data blocks based on a different measurement scale, low-level fusion may require additional pre-processing (e.g., block scaling with log scaling, root square scaling) of data blocks to compensate for the scale differences. However, as the MWIR and LWIR data blocks were obtained by adopting the same measuring scale (reflectance value in the 0–100 range), the chance of one technique being dominant was very low. The MWIR and LWIR data considered in the present study were acquired using a single instrument. Therefore, to fully elucidate the importance of the low-level fusion approach, the performance of the predictive models based on the full-range FTIR data and low-level fusion results was compared.

3.3.4. Mid-level Data Fusion

Mid-level data fusion consists of feature extraction from the signals pertaining to each data block. Prior to the fusion process, important variables are extracted separately as features from each data source using decomposition (or resolution) or variables selection techniques [16]. These extracted features are fused to develop a final model. Variable screening requires the most efficient data reduction method to retain the important variables in the model. In this study, a PCA model was used as a data reduction method (Figure 2). The most informative variables that explain most of the variations in the spectral data were extracted from the results yielded by the PCA models developed using MWIR and LWIR data. Prior to conducting PCA, different data pre-processing techniques (discussed in Section 2.2) were applied.

The number of PCs that explain most of the total variance in the data was chosen based on the minimum RMSECV value. The scores of the extracted PCs were concatenated to build a final fused data block of the calibration datasets. Similarly, the same procedures were followed to build the validation datasets separately. The fused data block was mean-centered and a prediction model was developed using PLSR (the green box in Figure 2). The new variable blocks formed matrices of the elemental concentration and PC scores that were used in the models as response and explanatory variables, respectively. Finally, the performance of the models developed using low-level and mid-level data fusion approaches was compared.

4. Results and Discussion

This section consists of seven parts. In the first two, the exploratory data analysis results and the results of the MWIR and LWIR prediction models are presented. In the subsequent two parts, the results of the low-level and mid-level data fusion for the prediction of element concentrations are described in detail. In the remaining three parts, the comparisons of the adopted approaches, assessment of IR spectra usability for prediction of elemental concentrations, and benefits and limitations of the current approach for in-situ applications are discussed, respectively.

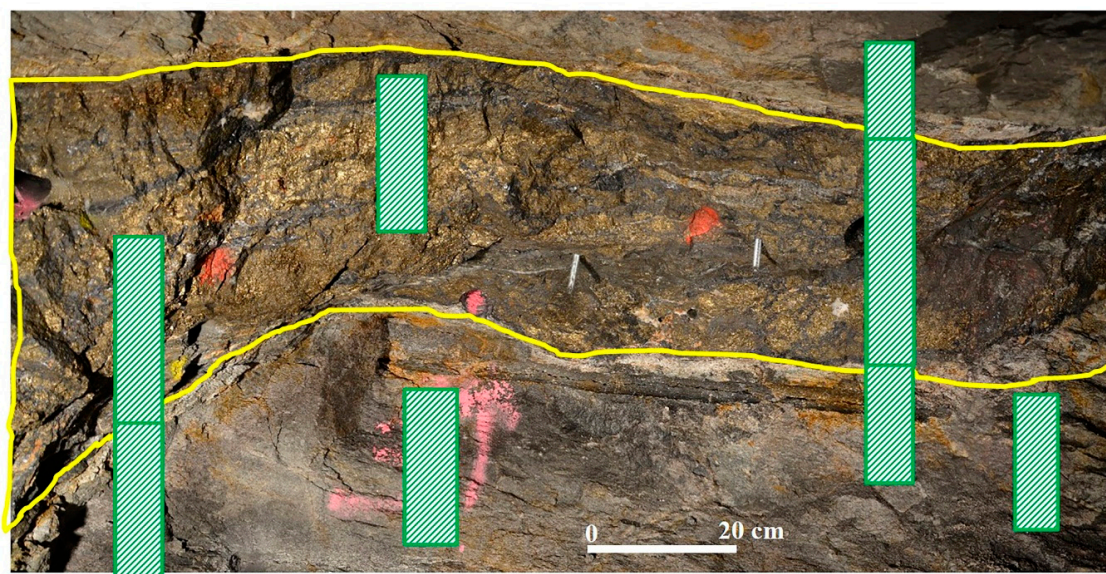
4.1. Exploratory Data Analysis and Pre-processing

Descriptive statistics were used in the present study to describe and summarize the basic features of the elemental concentrations in the analyzed samples (Table 1). The results reported in Table 1 indicate that the samples exhibit high variability in elemental concentrations. This likely stems from the sample collection strategy, whereby the ore zone region and waste (host rock and weathered) material were obtained to ensure material representativity (Figure 3). However, variability was also observed within the ore material.

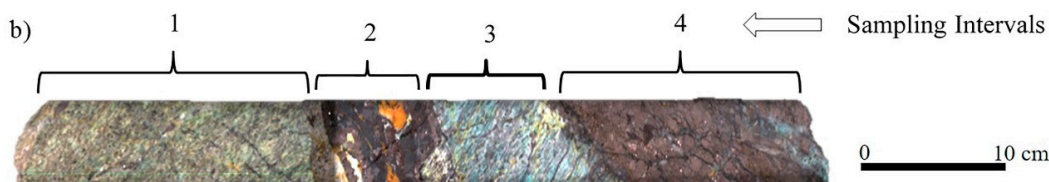
Table 1. Summary statistics of elemental concentrations (in wt.%) in the analyzed samples.

Statistics	Fe	Cu	As	Mg	Ca	Pb–Zn
Mean	13.72	0.1	0.27	0.25	0.69	6.43
Std Deviation	10.45	0.12	0.46	0.21	0.56	5.87
Min	2.30	0.002	0.006	0.03	0.08	0.04
Median	10.22	0.08	0.13	0.22	0.52	5.08
Max	40.00	0.9	2.44	1.4	3.33	30.32

a)



b)

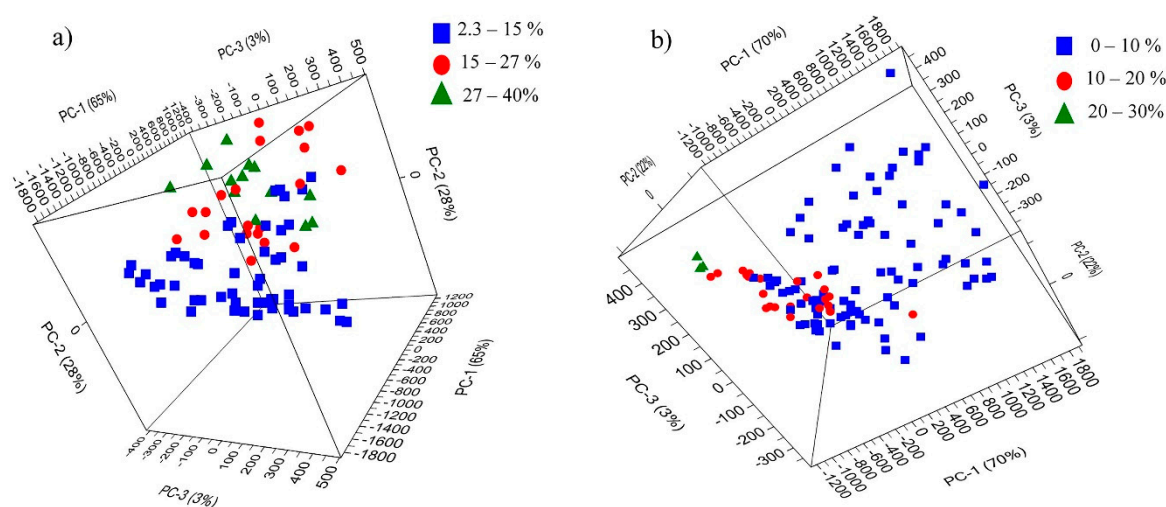
**Figure 3.** Photographs showing some of the sampling locations: (a) channel cuts at the mine face (the green boxes). The yellow line shows the ore zone boundary; and (b) a drill core showing the different lithotypes that were sampled separately.

As shown in Table 2, the use of the IR spectra for the prediction of the As, Cu, Ca, and Mg concentrations resulted in lower R^2 values (<0.4). Therefore, these elements were excluded from further analysis. Conversely, Fe and the combined Pb–Zn concentration showed a strong association (linear relationship) with the IR reflectance data. Therefore, the analysis results presented in this and the subsequent sections focus on Fe and the combined Pb–Zn only. The lower relationships of As, Cu, Ca, and Mg with the IR reflectance spectra are likely attributed to the lower concentration of the sourcing minerals in the samples that resulted in a lower spectral signal. It is also possible that the concentration of the sourcing minerals was below the detection limit of the instrument, or that the spectral signals of the sourcing minerals in the IR reflectance spectra were very weak.

Table 2. Coefficient of determination values for the prediction of each target element using the full-range FTIR reflectance spectra and partial least squares regression (PLSR).

Element	As	Ca	Mg	Cu	Fe	Pb–Zn
R ²	0.07	0.2	0.19	0.35	0.80	0.81

Figure 4 shows the PCA model score plots of the full-range FTIR data for Fe and the combined Pb–Zn datasets. The plots provide information on the potential patterns that are related to elemental concentration. As shown in Figure 4, elemental concentrations (Fe and the combined Pb–Zn) were categorized into three ranges. The plots further reveal the presence of patterns that can be related to the elemental concentration. Potential outliers were identified. As a result, two outliers were excluded from the Fe dataset, whereas four outliers were excluded from the combined Pb–Zn dataset.

**Figure 4.** Principal Component Analysis (PCA) model score plots of (a) the Fe concentration categorized into three ranges; and (b) the combined Pb–Zn concentration categorized into three ranges (the concentrations are expressed in wt.%).

4.2. MWIR and LWIR Datasets

As shown in Table 3, the model based on the MWIR dataset yields a better Fe prediction, with the RMSEP of 4.80 and the R² value of 0.69, after data normalization. Compared with the model performance based on raw MWIR data (the RMSEP of 5.63 and the R² of 0.58), data filtering using normalization improved the results. Likewise, the model developed using the normalized LWIR data resulted in RMSEP = 3.38 and R² = 0.85, which is a remarkable improvement relative to the RMSEP = 4.59 and the R² = 0.72 obtained using the raw LWIR data. Therefore, for Fe prediction, data normalization resulted in a better prediction performance for both MWIR and LWIR models. This shows that the removal of the undesired intensity variation in the spectra caused by multiplicative effects enhances the signals from the Fe-bearing minerals. Comparing the two datasets in terms of Fe prediction, the LWIR data model is superior to the MWIR model. The Fe-bearing minerals include oxides (e.g., hematite, goethite), carbonates (e.g., siderite), sulphides (e.g., pyrite, chalcopyrite), and silicates (pyroxene, olivine). LWIR is suited for rock-forming mineral and carbonate identification [28,29,41]. This is likely one of the reasons for the better prediction of Fe concentration based on the LWIR relative to MWIR data. The other possible reason is the presence of certain minerals that have a good correlation with both IR spectra and Fe concentration but do not contain Fe. For example, minerals formed as a result of retrogressive mineralization during hydrothermal events have a high chance of correlation.

Table 3. Summary of the predictive performances of mid-wave infrared (MWIR), long-wave infrared (LWIR), full-range FTIR, and low-level data fusion models for Fe concentration prediction.

Filtering Techniques	Statistics	MWIR	LWIR	Full Range FTIR	Low-Level Fusion
Raw data	RMSECV	3.76	4.54	3.75	
	RMSEP	5.63	4.59	3.87	
	R ²	0.58	0.72	0.80	
SNV	RMSECV	3.57	3.96	3.89	3.81
	RMSEP	5.49	4.13	4.43	4.12
	R ²	0.60	0.77	0.74	0.77
Normalize	RMSECV	3.46	3.51	2.88	2.71
	RMSEP	4.80	3.38	3.18	2.71
	R ²	0.69	0.85	0.87	0.90
Baseline	RMSECV	4.27	4.87	3.80	3.85
	RMSEP	5.73	4.20	4.50	4.11
	R ²	0.56	0.77	0.73	0.77
Gaussian	RMSECV	3.93	4.35	3.09	2.97
	RMSEP	5.47	4.44	3.40	3.43
	R ²	0.60	0.74	0.84	0.84

Model performance in terms of predicting the combined Pb–Zn concentration using MWIR and LWIR data is summarized in Table 4. The model based on the MWIR data yielded a better prediction after the data were treated using baseline correction (RMSEP = 2.23 and R² = 0.83). The model prediction accuracy obtained using raw MWIR data was lower than that based on the pre-processed data (RMSEP = 2.46 and R² = 0.80). The prediction accuracy of the model based on the LWIR data treated using baseline correction is reflected in RMSEP = 3.25 and R² = 0.65. As shown in Table 4, a better prediction was attained using the LWIR data subjected to baseline correction. A combination of the pre-processing techniques was analyzed for the prediction of both Fe and the combined Pb–Zn concentrations. However, the prediction performances of the models were not improved; thus, the results were not included in this paper.

Table 4. Summary statistics of the predictive performances of MWIR, LWIR, full-range FTIR, and low-level data fusion models for the combined Pb–Zn concentration prediction.

Filtering Techniques	Statistics	MWIR	LWIR	Full Range FTIR	Low-Level Fusion
Raw data	RMSECV	1.97	2.94	2.04	
	RMSEP	2.46	3.48	2.37	
	R ²	0.80	0.60	0.81	
SNV	RMSECV	2.37	3.65	2.41	1.96
	RMSEP	3.11	3.39	2.79	2.86
	R ²	0.68	0.62	0.74	0.73
Normalize	RMSECV	1.87	3.32	1.88	1.90
	RMSEP	2.33	3.28	2.19	2.24
	R ²	0.82	0.64	0.84	0.83
Baseline	RMSECV	2.09	3.04	2.07	1.89
	RMSEP	2.23	3.25	2.32	2.10
	R ²	0.83	0.65	0.82	0.85
Gaussian	RMSECV	2.04	3.15	2.09	1.98
	RMSEP	2.34	3.34	2.28	2.27
	R ²	0.82	0.63	0.83	0.83

Comparing the model performance in terms of the combined Pb–Zn concentration prediction, the model based on the MWIR data is superior to that using the LWIR data. The primary mineral sources of Pb and Zn in the test case deposit are galena and sphalerite, respectively. This is one of the likely reasons for a better prediction of the combined Pb–Zn concentration using the MWIR data than the LWIR data. For example, sphalerite has a lower reflection pattern in the spectral data in the 2.33–3.78 μm range and at 14.9 μm [42]. Similarly, for galena, higher reflection points are found in the MWIR region (Figure 5). Therefore, the amount of spectral information (signals) in the MWIR dataset is greater (more pronounced) than in the LWIR dataset, resulting in a superior combined Pb–Zn concentration prediction yielded by the model based on MWIR data.

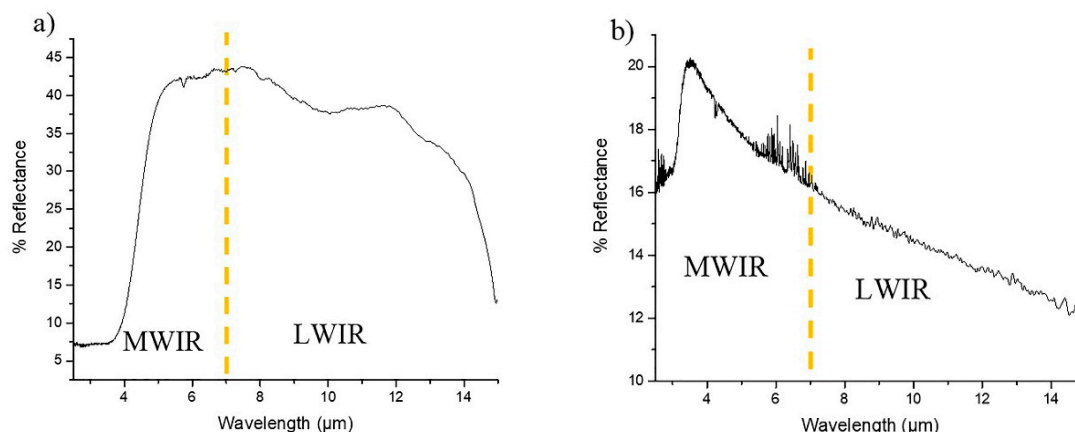


Figure 5. Reflectance spectra of (a) sphalerite and (b) galena (source: NASA spectra database of minerals [42]).

The elemental concentration prediction results yielded by the MWIR and LWIR models show that the IR technique is capable of indicating the concentration of elements in polymetallic sulphide ore semi-quantitatively. The IR technique is a mineralogical technique that provides information on the functional groups of minerals. Thus, elemental information cannot directly be derived from the IR spectra. However, the acquired results show that an indicative elemental concentration can be estimated using multivariate techniques without directly examining the fingerprint of individual elements in the IR spectral signals. The results of the present study indicate that MWIR data are a promising candidate for prediction of combined Pb–Zn concentrations. Thus, MWIR has great potential for use in material characterization.

The models developed using MWIR and LWIR data explained the variations in the Y-variables (concentration of elements) from the variations in the X-variables (reflectance spectral data), and resulted in acceptable prediction model accuracies. The model based on the MWIR data yielded a better prediction of the combined Pb–Zn concentration than that based on the LWIR data, whereas the Fe concentration was more accurately predicted using LWIR rather than MWIR data. Based on the available spectral libraries (e.g., the NASA mineral library), the sulphide minerals showed weak spectral features in the MWIR and LWIR spectral data. Such weak spectral information (signal) is insufficient for a direct interpretation or mineral fingerprinting, especially when target minerals are present in a mixture (co-occurring with other minerals). Moreover, the IR sensors provide mineralogical information but cannot directly identify the elements in the spectral signal. However, such information can be extracted using multivariate data analysis techniques.

4.3. Low-level Data Fusion

Table 3 shows a summary of the Fe prediction model performance when applied to the individual data blocks, full-range FTIR data, and low-level data fusion results. As shown in Table 3 and Figure 6,

better model performance was achieved using low-level data fusion (RMSEP = 2.71 and $R^2 = 0.90$) after the normalized data blocks based on the MWIR and LWIR datasets were fused.

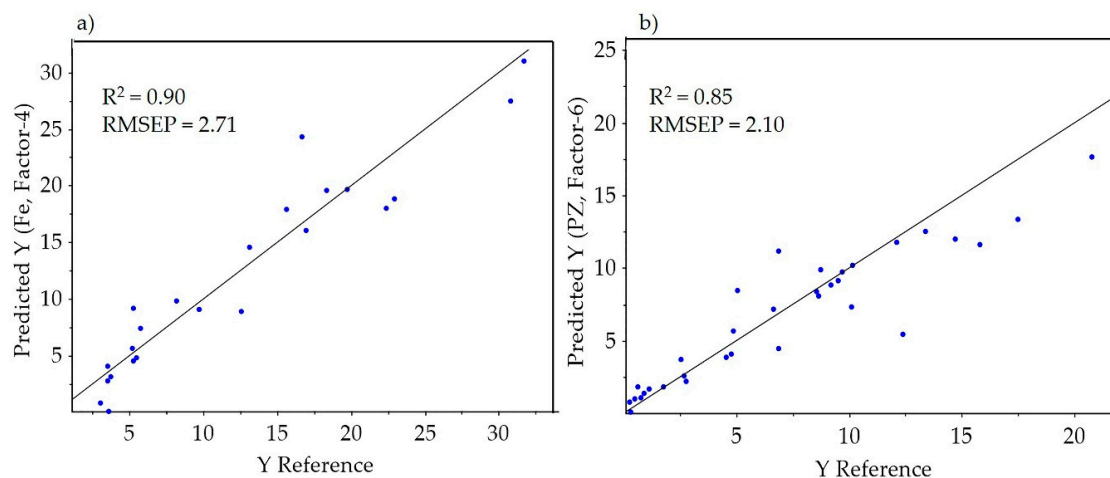


Figure 6. Predicted versus actual (a) Fe concentration values (wt.%); (b) the combined Pb–Zn concentration values (wt.%) computed using a low-level data fusion approach.

As described in Section 2.1.2, the MWIR and LWIR datasets were acquired using a single-sensor FTIR spectrometer, which allowed for the performance of models based on the full-range FTIR data (which include both MWIR and LWIR datasets) to be assessed and compared to the low-level fusion results. This was done to test whether the low-level data fusion approach yielded superior results to those based on the full-range data. Though the performance of the prediction models based on the full-range FTIR data was better than that obtained using the individual datasets (Tables 3 and 4), the prediction performance acquired from the low-level data fusion (after data block normalization) was superior to that acquired from the full-range FTIR data model. Similarly, the combined Pb–Zn prediction using a low-level data fusion was better than the results obtained using the individual and the full-range FTIR data (Table 4).

The portable full-range FTIR sensor is an example of a physically integrated MWIR and LWIR sensor. This is an interesting achievement of the current advancement of IR technologies. The performance of the prediction models based on the full-range FTIR data was better than that obtained using the individual datasets. However, the results of this study suggest data fusion to be a better and comparative option for a combination of multiple sensors. Moreover, the physical integration of multiple sensor technologies into a single platform is challenging and, in terms of practical implementation, it is expensive. Therefore, for the integration of multiple data sources, such as short-wave infrared (SWIR), Raman, and LWIR data (provided each sensor technology has its own advantages in providing useful information), data fusion can be considered to be an economic and practical alternative option.

The Fe and the combined Pb–Zn concentrations yielded by the low-level data fusion approach were highly affected by the choice of pre-processing technique. This is in line with the recommended “systematic approach” [32]. The normalization of the two data blocks prior to data fusion resulted in a better prediction of Fe concentration, whereas baseline correction resulted in a better model prediction for the combined Pb–Zn concentration. The MWIR and LWIR data considered in this study were acquired using a single instrument, thus ensuring that the data scale difference is not an issue. However, treating the noise in the two data blocks separately and fusing the pre-processed data blocks enhanced the prediction performance of the PLSR models.

In low-level data fusion, as the name suggests, data integration occurs at the bottom of the analytical data flow. Indeed, this strategy implies that the matrices describing the individual blocks, after proper pre-processing, are concatenated to build a fused data block, which is then processed by the desired chemometrics technique. Suppose X denotes an $m \times n$ matrix of independent variables,

where m is the number of samples and n is the number of MWIR and LWIR wavelengths together. Y ($m \times p$) is the matrix of dependent variables, where p denotes the elements to be predicted. The purpose is to train a model estimating the concentration of the elements in each polymetallic sulphide ore sample through the matrix of the measured spectra. The problem is formulated by constructing input (X) and output (Y) matrices. For example, the combined MWIR and LWIR reflectance for 89 samples is formed as X (89×7190) and the matrix of outputs for Fe will be Y (89×1). As the fusion occurs at the level of the original data matrices, the resulting data block will typically contain a very high number of variables. Consequently, the main drawback of this strategy is that the increase in information obtained by adding one or more blocks of data to describe the sample may not compensate for irrelevant variability introduced by the addition of the same blocks.

The PLSR model developed using the fused data blocks simultaneously considers the MWIR and LWIR measurements, due to which it takes advantage of having a greater number of informative explanatory variables to perform the elemental concentration prediction with better accuracy than that achieved by the models based on individual data blocks. Low-level fusion is conceptually simple to implement and the acquired results are superior to those yielded when using individual techniques. However, the prediction performance of the individual MWIR and LWIR data models also resulted in a good prediction potential for Fe and the combined Pb–Zn concentration. It is possible that the same minerals were measured in both datasets or the minerals were highly correlated to each other due to their co-occurrence in the deposit. However, the improvement of the predictive model after low-level data fusion implies that, as there is unique spectral information in both datasets, combining the two data blocks results in better prediction performance.

4.4. Mid-level Data Fusion

Feature selection is an important step in mid-level data fusion implementation. Hence, an attempt was made to ensure proper coverage of the existing variation in the reflectance spectra data. For example, for the prediction of the combined Pb–Zn concentration using the MWIR data treated with baseline correction, four PCs that explain 98% of the variance in the data were chosen. Similarly, five PCs that explain 97% of the variation in the data were chosen for the LWIR data treated with baseline correction. These PCs (four from the MWIR data block and five from the LWIR data block) were concatenated to generate the fused data matrix. Using the newly generated variable data matrices, final elemental concentration prediction models were developed by adopting PLSR.

As shown in Table 5, for the prediction of the combined Pb–Zn concentration, better model performance was achieved ($RMSEP = 3.05$ and $R^2 = 0.69$) after baseline correction. For the Fe concentration prediction, a better model was achieved after the normalized data block features were fused ($RMSEP = 5.09$ and $R^2 = 0.65$). The acquired prediction performance of the models for estimation of the combined Pb–Zn concentration was good. However, model performance in terms of predicting the Fe concentration was lower ($RMSEP = 5.09$). Overall, the performance of the prediction models after mid-level data fusion was lower than that of the models employing individual techniques. Nonetheless, the acquired results are promising and can likely be improved by adding more data to the calibration dataset (thus to fully capture the high variability between the samples) or by considering other feature selection techniques.

Table 5. Summary of the low-level and mid-level data fusion model performance in terms of prediction accuracy of Fe and the combined Pb–Zn concentration. The total number of selected principal components (PCs) shows the number of selected PCs from the principal component analysis (PCA) models of the MWIR and LWIR data.

Element	Data Pre-Processing Technique	Statistics	Low-Level Fusion	Total No. of Selected PCs	Mid-Level Fusion	Factors for the Model after Mid-Level Fusion
Pb–Zn	Normalize	RMSECV	1.90	7	2.37	2
		RMSEP	2.24		3.12	
		R ²	0.83		0.68	
	Baseline	RMSECV	1.89	9	2.09	2
		RMSEP	2.10		3.05	
		R ²	0.85		0.69	
Fe	Normalize	RMSECV	2.71	7	2.87	2
		RMSEP	2.71		5.09	
		R ²	0.90		0.65	
	Baseline	RMSECV	3.67	7	5.08	4
		RMSEP	4.58		5.67	
		R ²	0.72		0.57	

In the mid-level data fusion approach, data reduction using variable screening is a pre-requisite. The data volume of the combined MWIR and LWIR datasets and the new data matrices generated from the extracted features have a significant difference. The combined MWIR and LWIR datasets contain 7190 variables, whereas the new data blocks generated from the extracted features included only seven to nine variables, depending on the amount of the explained variance from the score results of the PCA models. With feature extraction methods, the important information (informative variables in the spectral data) is retained using fewer variables than in the original data blocks (MWIR and LWIR datasets). This is an interesting point, since a large data volume (megavariate or multivariate data) may lead to several computational challenges and, due to the development of high-throughput instrumentation, complex datasets are increasingly becoming available.

The reflectance IR data likely contains a certain amount of noise (e.g., caused by instrument imprecision and/or measurement errors) and, in some instances, may include irrelevant information in the spectra (data that are unrelated to elemental concentrations). Therefore, treating the datasets with the data filtering techniques mentioned above improved the prediction accuracy of the models developed as a part of this study. Irrelevant information can also hinder the predictive ability of the available information; in this case, extracting the important variables can result in better prediction.

4.5. Data Fusion versus Individual Sensors

The prediction model performance based on the individual MWIR and LWIR datasets is favorable in terms of the elemental concentration (Fe and the combined Pb–Zn concentration) prediction accuracy. Owing to the fact that these mineralogical techniques are mainly used for qualitative analysis of mineralogical information, the results obtained in this study are encouraging, indicating that the potential of these techniques, especially the MWIR data, should be explored further, since it is the least-utilized region of the IR spectrum. The approach presented in this study is a data-driven approach as a part of which the hidden information in the spectral signal is transformed into a quantitative indication of elemental concentration.

Comparing the performance of the prediction models based on the individual techniques with the data fusion results, an improved prediction accuracy was achieved using a low-level data fusion approach. Low-level data fusion enhanced the reliability of the prediction models by increasing the prediction performance and minimizing the uncertainty implicit in each individual technique.

Data fusion is effective in identifying the correlations and the similarities/differences (common and distinctive information) among different variables. However, if applied to correlated or noisy data blocks, data fusion might not necessarily improve model performance [14]. In some cases, unnecessary information can also hinder model performance, due to which a better prediction can be achieved from individual data blocks than from the fused dataset.

The prediction performance of the models after low-level data fusion was superior to that yielded by the mid-level data fusion models. This is likely due to the fact that the original information from both data blocks (MWIR and LWIR) is maintained in low-level fusion, making it potentially more accurate. In mid-level data fusion, PCA potentially failed to describe all important variations in the data. However, this issue can be overcome by considering other data decomposition methods, such as multivariate curve resolution/alternating least squares (MCR-ALS) or independent component analysis (ICA). Review studies [9] indicate that there is no preferred level of data fusion, as the choice is always application-dependent.

4.6. Use of Infrared Reflectance Spectra for Elemental Concentration Indication

Understanding the spatial variability in elemental concentration is crucial for strategic extraction of minable products in mining operations. The data-driven approach proposed in this study provides an insight into the quantifiable elemental concentrations in polymetallic sulphide ore using IR reflectance data. PLSR was used to relate the concentration of elements of interest with complex high-dimensional IR data to reveal hidden information in the spectra. Low-level data fusion resulted in better predictive performance of models than could be attained using models based on the individual techniques. This is likely due to the fact that the two data blocks (MWIR and LWIR) contain different chemical information that can be linked to the description of the sourcing minerals containing elements of interest. In addition, the comparable predictive performance of the MWIR and LWIR data models likely indicates the occurrence of similar chemical information as well. Therefore, the two techniques are complementary to each other.

The performance of the prediction models utilizing the MWIR, LWIR, and fused datasets indicates that prediction improves with higher Fe concentration in the samples. This is likely due to the fact that a higher Fe concentration in the samples results in a better spectral signal of the Fe sourcing minerals in the datasets. Thus, a better linear relationship can be established with the IR spectra. Conversely, for the combined Pb–Zn concentration prediction, a lower concentration of these elements was better described by the models. However, the higher the Pb–Zn concentration in the samples, the lower its description by the models, likely due to higher reflectance variation, as more samples are needed in the calibration data to obtain a better description by the models.

In mineralogical or elemental investigations, identification of as many minerals or elements as possible in the material is advantageous to finding indirect proxies of target minerals and understanding the requirements for mineral processing. However, material characterization is a multi-step process that involves addressing different issues at different levels. The entire process, which is comprised of data generation, analysis, information extraction, and the depth of information (or scope of material characterization), is illustrated in Figure 7. Nowadays, most sensors produce high-throughput multivariate or megavariable data. The raw data is complex, making it challenging to interpret and handle the megavariable data obtained from different sources. Therefore, there is a need for machine learning methods, such as chemometrics, for data-driven understanding of the complex datasets. Using multivariate techniques, the hidden information in spectral data can be transformed into useful information, such as key geological parameters (elemental, mineralogical, and textural data) in mining operations. However, based on the information level that can be acquired from the sensor signals, the depth of information or the amount of information that can be generated differs. The depth of information or scope of material characterization from the bottom up can be organized into (i) detection (determines the presence of materials in the entity under investigation), (ii) discrimination (separates

materials into similar groups based on the signals in the spectral data), (iii) identification (determines the fingerprints of individual entities), and (iv) quantification (determines the quantity of materials).

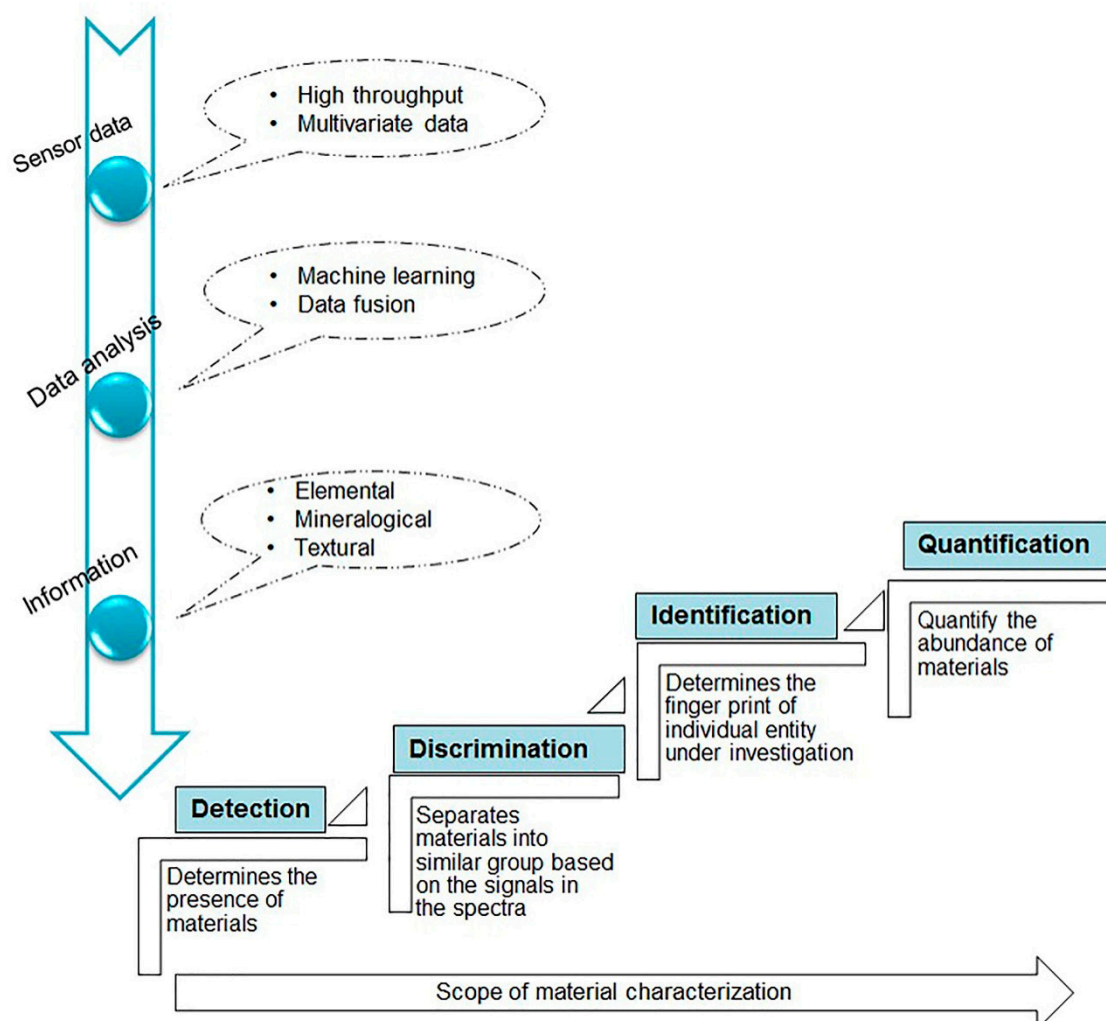


Figure 7. A schematic diagram illustrating the process of knowledge generation from sensor data and the depth of information.

Thus, quantification of the composition of a material under investigation is at a higher level of the material characterization or information extraction process. Quantification can be achieved after identification of a unique identity of a material in the acquired spectra, or it can be achieved indirectly by developing prediction models based on the correlation between the spectral signal and the entity under investigation (e.g., elemental concentration). In the present study, the use of MWIR and LWIR data for semi-quantification of elemental concentration was analysed. The prediction accuracies acquired from the fused data or individual datasets were very good. The achieved results should be termed ‘semi-quantification’, since the acquired RMSEP was not as low as would be expected for exact prediction. However, the achieved RMSEP gives a very good indication of the elemental concentration in the samples and can be used to indicate high-, medium-, and low-mineralization zones in mining operations. In addition, a lower RMSEP can be achieved using more data in the calibration datasets. One of the advantages of the use of IR data for the indication of elemental concentration stems from the fact that the infrared data can be further analyzed to acquire mineralogical information simultaneously. Thus, IR can be considered to be a complementary technique to technologies that provide elemental information (e.g., XRF or LIBS).

Overall, the prediction accuracies acquired in this study show promise in terms of semi-quantification of elemental concentrations in polymetallic sulphide ore samples using data sources that have historically been used for mineral identification only. Therefore, this study can serve as a baseline to show a framework for the implementation of data fusion using multiple data sources and various classification or regression algorithms.

4.7. Opportunities and Limitations for In-Situ Application

Owing to current technological advancement, rapid and non-destructive sensor technologies are feasible for mineral and mining applications. These technologies produce high-throughput megavariable datasets. As a result, understanding and interpreting the sensor data directly might not be possible. This leads to computational challenges in analyzing the megavariable data and generating knowledge based on fast fingerprinting approaches in order to automate a material characterization process. Therefore, as illustrated in Figure 7, knowledge generation from sensor-derived data requires advances in analytics using machine-learning techniques. The approach presented in this study enables semi-quantification of elemental concentrations using IR spectra and data fusion. The reflectance IR spectra were linearly correlated to the elemental concentrations in polymetallic sulphide ore. Data fusion resulted in improved model prediction accuracies and lower detection errors (uncertainty) than the individual datasets. This is likely due to jointly analyzing data blocks from different sensors, which allowed us to capture the latent information that would not be extracted by analyzing each data block individually.

Data fusion is mainly a data-driven approach in which the preferred level of data fusion depends on the nature of the data. Low-level fusion is conceptually simple, uses a single model, and can capture correlations between variables contained in different blocks. The disadvantages of the low-level data fusion are the high data volume (high variable to sample ratio) and the possible predominance of one data source over the others. The concerns stemming from a high data volume can be resolved using mid-level data fusion. For example, data reduction via feature selection (the scores in this case) in mid-level data fusion reduced the data dimensionality tremendously in the present study. Comparing the number of variables used in the two levels of data fusion, the mid-level data fusion used ~0.14% of the data volume required for low-level fusion. Thus, feature extraction in mid-level fusion reduces data dimensionality (in this case by ~99.86%), which significantly minimizes the time required for computational analysis. The disadvantage of mid-level data fusion is that it requires the development of many models prior to fusion (each dataset must be reduced) and there is a need for an ultimate variable screening method. Therefore, the choice of the preferred level of data fusion is dependent on the performance of prediction or classification models and their applications, and should thus be made using a “systematic” approach.

Polymetallic sulphide ore characterization using IR spectra considers weak IR features, complex matrices due to the presence of multiple sulphide minerals and other associated minerals, peak overlaps, and additional variability that results from physical and chemical weathering. Therefore, the need to reveal the hidden information in IR spectra related to the material of interest mandates the use of multivariate analysis (e.g., integration of data blocks from multiple sensors). Elemental concentration indication into low-, medium-, and high-mineralization zones is advantageous for elemental spatial variability mapping in ore grade control applications. The proposed framework can be extended to simultaneous fingerprinting of the mineralogical information and elemental concentration indication (measurements of concentrations of valuable elements) in order to achieve comprehensive quantitative analysis.

The in-situ application of elemental concentration indication using an IR spectra and data fusion approach is highly advantageous. However, to achieve this aim in practice, a robust system design and an integrated principled tool that integrates efficient data collection, processing, and knowledge generation is necessary. Going forward, automated material characterization is possible with a robust system design (exemplified by a portable and ruggedized system) and efficient software

(test-case-specific mineral libraries) that can be developed using a combined sensor signal. As currently available sensor technologies measure different aspects of material properties, information acquired from each technique likely adds to a holistic description of materials. The application presented in this study was based on a point technique (FTIR spectrometer data), which yields information at a specific point. However, a correlation model can be developed using image data yielded by other sensor technologies, such as hyperspectral imagery or RGB imaging, to predict the elemental concentration in unmeasured areas (to extend the coverage).

5. Conclusions

This study developed prediction models for the indication of elemental concentration in polymetallic sulphide ore using infrared reflectance spectra. Different factors that possibly influence the performance of the prediction models were used to formulate testing scenarios: (1) the usability of the individual datasets (MWIR and LWIR) for the prediction of the elemental concentrations; (2) the effect of the different data pre-processing techniques on the prediction performance; and (3) the potential for improvement in prediction accuracy by applying low-level data fusion and mid-level data fusion.

The results reported in the preceding sections show that both MWIR and LWIR datasets include relevant information that can be related to the concentrations of the elements. Comparing the best-performing models, the prediction performance of the MWIR model after baseline correction (RMSEP = 2.23 and $R^2 = 0.83$) for the combined Pb–Zn elemental concentration prediction was superior to that obtained using LWIR data (RMSEP = 3.25 and $R^2 = 0.65$). On the other hand, the LWIR-based model after data normalization predicted the Fe concentration (RMSEP = 3.38 and $R^2 = 0.85$) better than the MWIR-based model (RMSEP = 4.80 and $R^2 = 0.69$). The MWIR is the least-explored region of the electromagnetic spectrum for mineralogical applications. However, the results yielded by this study show the potential of this technique for material characterization. The use of a low-level data fusion strategy improved the model predictive ability relative to the results yielded by using individual techniques. After the low-level fusion, the best-achieved model performance for the prediction of the combined Pb–Zn concentration was RMSEP = 2.10 and $R^2 = 0.85$ and for the prediction of the Fe concentration was RMSEP = 2.71 and $R^2 = 0.90$. The performance of the prediction models after the mid-level fusion was lower than that of the models employing individual techniques. However, the acquired results are promising and can likely be improved by adding more data to the calibration dataset (thus to fully capture the high variability between the samples) or by considering other feature extraction techniques.

Overall, the low- and mid-level data fusion approaches proposed in this study showed promising results. The outcomes suggest that the approach has a great potential to be extended for the integration of more complex data from a range of sensor technologies, due to the fact that each sensor technology has its own advantages in providing useful information. Prediction or indication of elemental concentration using mineralogical techniques (IR data) is useful to understand compositional properties, both in terms of mineralogy and geochemistry. The individual techniques (MWIR and LWIR) exhibit a great potential for use in elemental concentration characterization in sulphide ore, and the model performance can be improved using data fusion. This study is a baseline for future research that involves multiple sensor integration aimed at developing predictive models that can estimate elemental concentrations in different types of deposits. We recommend enhancing the current work using data from other sensors.

Author Contributions: Conceptualization, F.D. and M.B.; methodology, F.D., M.B., and J.J.; F.D. performed the experimental analysis; F.D. wrote the paper; and M.B. and J.J. reviewed it before submission. All authors have read and agreed to the published version of the manuscript.

Funding: This research (Real-Time Mining Project) has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 641989.

Acknowledgments: We would like to thank our colleagues at TU Bergakademie Freiberg for their support during sample collection field campaigns.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Sharma, S.K.; Misra, A.K.; Lucey, P.G.; Lentz, R.C.F. A combined remote Raman and LIBS instrument for characterizing minerals with 532 nm laser excitation. *Spectrochim. Acta A* **2009**, *73*, 468–476. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fernández, R.; Montes, H.; Salinas, C.; Sarria, J.; Armada, M. Combination of RGB and multispectral imagery for discrimination of cabernet sauvignon grapevine elements. *Sensors* **2013**, *13*, 7838–7859. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hoehse, M.; Mory, D.; Florek, S.; Weritz, F.; Gornushkin, I.; Panne, U. A combined laser-induced breakdown and Raman spectroscopy Echelle system for elemental and molecular microanalysis. *Spectrochim. Acta* **2009**, *64*, 1219–1227. [\[CrossRef\]](#)
- Hall, D.L. *Mathematical Techniques in Multisensor Data Fusion*, 2nd ed.; Artech House: Bristol, UK, 1992; pp. 1–32.
- Thrun, S.; Burgard, W.; Fox, D. *Probabilistic Robotics*; MIT Press: Cambridge, UK, 2005.
- Hall, D.L.; Llinas, J. An introduction to multisensor data fusion. *Proc. IEEE* **1997**, *85*, 6–23. [\[CrossRef\]](#)
- Siciliano, B.; Khatib, O. *Springer Handbook of Robotics*, 2nd ed.; Springer: Berlin, German, 2016.
- Biancolillo, A.; Bucci, R.; Magrì, A.L.; Magrì, A.D.; Marini, F. Data-fusion for multiplatform characterization of an italian craft beer aimed at its authentication. *Anal. Chim. Acta* **2014**, *820*, 23–31. [\[CrossRef\]](#) [\[PubMed\]](#)
- Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto, O. Data fusion methodologies for food and beverage authentication and quality assessment—A review. *Anal. Chim. Acta* **2015**, *891*, 1–14. [\[CrossRef\]](#) [\[PubMed\]](#)
- Doeswijk, T.G.; Smilde, A.K.; Hageman, J.A.; Westerhuis, J.A.; van Eeuwijk, F.A. On the increase of predictive performance with high-level data fusion. *Anal. Chim. Acta* **2011**, *705*, 41–47. [\[CrossRef\]](#)
- Silvestri, M.; Bertacchini, L.; Durante, C.; Marchetti, A.; Salvatore, E.; Cocchi, M. Application of data fusion techniques to direct geographical traceability indicators. *Anal. Chim. Acta* **2013**, *769*, 1–9. [\[CrossRef\]](#)
- Smilde, A.K.; van der Werf, M.J.; Bijlsma, S.; van der Werff-van der Vat, B.J.; Jellema, R.H. Fusion of mass spectrometry-based metabolomics data. *Anal. Chem.* **2005**, *77*, 6729–6736. [\[CrossRef\]](#)
- Castanedo, F. A Review of Data Fusion Techniques. *Sci. World J.* **2013**, 1–19. [\[CrossRef\]](#)
- Forshed, J.; Idborg, H.; Jacobsson, S.P. Evaluation of different techniques for data fusion of LC/MS and ¹H-NMR. *Chemom. Intell. Lab. Syst.* **2007**, *85*, 102–109. [\[CrossRef\]](#)
- Silvestri, M.; Elia, A.; Bertelli, D.; Salvatore, E.; Durante, C.; Li, V.M.; Cocchi, M. A mid-level data fusion strategy for the Varietal Classification of Lambrusco PDO wines. *Chemom. Intell. Lab. Syst.* **2014**, *137*, 181–189. [\[CrossRef\]](#)
- Cocchi, M. Chapter 1 Introduction: Ways and Means to Deal with Data from Multiple Sources. In *Data Handling in Science and Technology*; Cocchi, M., Ed.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 1–26.
- Gil-Sánchez, L.; Soto, J.; Martínez-Máñez, R.; García-Breijo, E.; Ibáñez, J.; Llobet, E. A novel humid electronic nose combined with an electronic tongue for assessing deterioration of wine. *Sens. Actuators A* **2011**, *171*, 152–158. [\[CrossRef\]](#)
- Haddi, Z.; Alami, H.; El Bari, N.; Tounsi, M.; Barhoumi, H.; Maaref, A.; Jaffrezic-Renault, N.; Bouchikhi, B. Electronic nose and tongue combination for improved classification of Moroccan virgin olive oil profiles. *Food Res. Int.* **2013**, *54*, 1488–1498. [\[CrossRef\]](#)
- Brás, L.P.; Bernardino, S.A.; Lopes, J.A.; Menezes, J.C. Multiblock PLS as an approach to compare and combine NIR and MIR spectra in calibrations of soybean flour. *Chemom. Intell. Lab. Syst.* **2005**, *75*, 91–99. [\[CrossRef\]](#)
- Steinmetz, V.; Roger, J.M.; Molto, E.; Blasco, J. On-line fusion of colour camera and spectrophotometer for sugar content prediction of apples. *J. Agric. Eng. Res.* **1997**, *73*, 207–216. [\[CrossRef\]](#)
- Zude, M.; Herold, B.; Roger, J.M.; Bellon-Maurel, V.; Landahl, S. Non-destructive tests on the prediction of apple fruit flesh firmness and soluble solids content on tree and in shelf life. *J. Food Eng.* **2006**, *77*, 254–260. [\[CrossRef\]](#)
- Downey, G.; Briand, R.; Wilson, R.H.; Kemsley, E.K. Near- and mid-infrared spectroscopies in food authentication: Coffee varietal identification. *J. Agric. Food Chem.* **1997**, *45*, 4357–4361. [\[CrossRef\]](#)
- Márquez, C.; López, M.I.; Ruisánchez, I.; Callao, M.P. FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud. *Talanta* **2016**, *161*, 80–86. [\[CrossRef\]](#)

24. Aranda-Sanchez, J.I.; Baltazar, A.; González-Aguilar, G. Implementation of a Bayesian classifier using repeated measurements for discrimination of tomato fruit ripening stages. *Biosyst. Eng.* **2009**, *102*, 274–284. [CrossRef]
25. Vera, L.; Aceña, L.; Guasch, J.; Boqué, R.; Mestres, M.; Busto, O. Discrimination and sensory description of beers through data fusion. *Talanta* **2011**, *87*, 136–142. [CrossRef] [PubMed]
26. Li, C.; Heinemann, P.; Sherry, R. Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection. *Sens. Actuators B* **2007**, *125*, 301–310. [CrossRef]
27. Khajehzadeh, N.; Haavisto, O.; Koresaar, L. On-stream mineral identification of tailing slurries of an iron ore concentrator using data fusion of LIBS, reflectance spectroscopy and XRF measurement techniques. *Miner. Eng.* **2017**, *113*, 83–94. [CrossRef]
28. Clark, R.N. Spectroscopy of rocks and minerals and principles of spectroscopy. In *Manual of Remote Sensing Remote Sensing for the Earth Sciences*; Rencz, A.N., Ed.; Wiley: New York, NY, USA, 1999; Volume 3, pp. 3–58.
29. Hecker, C.; Dilles, J.H.; van der Meijde, M.; van der Meer, F.D. Thermal infrared spectroscopy and partial least squares regression to determine mineral modes of granitoid rocks. *Geochem. Geophys. Geosyst.* **2012**, *13*, 3021. [CrossRef]
30. Desta, F.S.; Buxton, M.W.N. Chemometric Analysis of Mid-Wave Infrared Spectral Reflectance Data for Sulphide Ore Discrimination. *Math. Geosci.* **2018**, *51*, 877–903. [CrossRef]
31. Guatame-Garcia, A.; Buxton, M. The Use of Infrared Spectroscopy to Determine the Quality of Carbonate-Rich Diatomite Ores. *Minerals* **2018**, *8*, 120. [CrossRef]
32. Engel, J.; Gerretzen, J.; Szymańska, E.; Jansen, J.J.; Downey, G.; Blanchet, L.; Buydens, L.M.C. Breaking with trends in pre-processing? TrAC. *Trends Anal. Chem.* **2013**, *50*, 96–106. [CrossRef]
33. Scheinert, M.; Kupsch, H.; Bletz, B. Geochemical investigations of slags from the historical smelting in Freiberg, Erzgebirge (Germany). *Geochemistry* **2009**, *69*, 81–90. [CrossRef]
34. Seifert, T.; Sandmann, D. Mineralogy and geochemistry of indium-bearing polymetallic vein-type deposits: Implications for host minerals from the Freiberg district, Eastern Erzgebirge, Germany. *Ore Geol. Rev.* **2006**, *28*, 1–31. [CrossRef]
35. Seifert, T. *Post-Collisional Magmatism and Its Relationship to Late-Variscan Ore Forming Processes in the Erzgebirge (Bohemian Massif) Germany*; IOS Press BV: Amsterdam, The Netherlands, 2008; pp. 1–128.
36. Benkert, T.; Dietze, A.; Gabriel, P.; Gietzel, J.; Gorz, I.; Grund, K.; Lehmann, H.; Lowe, G.; Mischo, H.; Schaeben, H.; et al. First step towards a virtual mine—Generation of a 3D model of Reiche Zeche in Freiberg. In Proceedings of the 17th Annual Conference of the International Association for Mathematical Geosciences (IAMG), Freiberg, Germany, 5–13 September 2015; pp. 1350–1356.
37. Roussel, S.; Preys, S.; Chauchard, F.; Lallemand, J. Multivariate data analysis (Chemometrics). In *Process Analytical Technology for the Food Industry*; Springer: New York, NY, USA, 2014; pp. 7–59.
38. Bro, R.; Smilde, A.K. Centering and scaling in component analysis. *J. Chemom.* **2003**, *17*, 16–33. [CrossRef]
39. Fearn, T.; Riccioli, C.; Garrido-Varo, A.; Guerrero-Ginel, J.E. On the geometry of SNV and MSC. *Chemom. Intell. Lab. Syst.* **2009**, *96*, 22–26. [CrossRef]
40. Latan, H.; Noonan, R. *Partial Least Squares Path Modelling: Basic Concepts, Methodological Issues and Applications*; Springer: Cham, Switzerland, 2017.
41. Terracore. The Geospectral Image. Available online: www.terracoregeo.com (accessed on 4 December 2018).
42. NASA. ECOSTRESS Spectral Library. Available online: <https://speclib.jpl.nasa.gov> (accessed on 1 September 2017).

