



Published in final edited form as:

*Appl Spectrosc.* 2025 March ; 79(3): 364–375. doi:10.1177/00037028241296158.

## Applying Finite Mixture Models to Quantify Respirable Dust Mass in Coal and Metal–Nonmetal Mines Using Fourier Transform Infrared Spectroscopy

Andrew T. Weakley<sup>1</sup>, David A. Parks<sup>2</sup>, Arthur L. Miller<sup>2</sup>

<sup>1</sup>Department of Neurology, University of California Davis Health, Sacramento, California, USA

<sup>2</sup>Office of Mine Safety and Health Research, National Institute for Occupational Safety and Health (NIOSH), Spokane, Washington, USA

### Abstract

Respirable dust mass is a prevalent occupational health hazard to the mining workforce. Mineral matrices observed in the mine environment are complex, time varying, and heterogeneous. This poses a challenge to assessing dust exposure using Fourier transform infrared (FT-IR) spectrometry as calibrations for constituent dust species (e.g., crystalline silica) have historically been trained using homogeneous standards or simple mixtures therein. Investigations have considered direct-on-filter analysis, which collects FT-IR spectra directly from sampling filters for calibration, as an alternative. Direct-on-filter analysis using a partial least squares (PLS) method has gained particular interest recently due to the potential to rapidly quantify multiple species from a single filter at the mine site. By design, heterogeneity, and its presumed impact on method accuracy, cannot be addressed in the laboratory when using a direct-on-filter approach motivating the need for more advanced calibration approaches. When heterogeneity is present, mixture of experts (MoE) finite mixture models offer a promising and novel alternative to PLS direct-on-filter analysis as MoE incorporates cluster discovery, regression, and outlier identification into model fitting. Three MoE models of increasing complexity were tasked with determining respirable dust mass in 243 field samples from thirteen active coal, limestone, sandstone, and silver mines. All MoE models, including those using only “expert” spectroscopic predictors or a combination of expert and categorical “gate” variables (e.g., mine type), significantly outperform PLS in terms of accuracy ( $\alpha = 0.05$ ). Decomposing bias by mine type shows that accuracy generally improves across all types considered when MoE models are not overfitted. The MoE method’s effectiveness was linked to its ability to endogenously classify outliers as well as possibly to the use of an additional cluster model for mass predictions. Overall, MoE methods appear as a capable and novel tool to addressing problems of heterogeneity for direct-on-filter quantitative analysis.

---

**Corresponding Author:** Andrew T. Weakley, UC Davis Alzheimer’s Disease Research Center, 1651 Alhambra Blvd #200a, Sacramento, California, 95816, USA. weak5174@gmail.com.

**Declaration of Conflicting Interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Supplemental Material**

All supplemental material mentioned in the text is available in the online version of the journal.

## Keywords

Mixture of experts; model-based clustering; occupational health; Fourier transform infrared spectrometry; FT-IR

---

## Introduction

Occupational exposure to respirable dust is a health hazard to workers, especially if particulate matter is concentrated with respirable crystalline silica (RCS) as is frequently the case in the mine environment.<sup>1</sup> Respirable dust resulting from mining operations has been regulated in the United States for several decades by the Mine Safety and Health Administration (30 Code of Federal Regulations (CFR)) § 70.100, 70.101, 90.100, 90.101). These regulations specifically address RCS exposure as it is considered more toxic than other dusts, with the permissible exposure limit (PEL) recently set to 50 µg/m<sup>3</sup> over a time-weighted average working shift (89 FR 28218). Acute overexposures and/or prolonged exposures to RCS can lead to adverse health outcomes including silicosis,<sup>2–4</sup> lung cancer,<sup>5,6</sup> and other respiratory and autoimmune diseases.<sup>7,8</sup> Furthermore, the fine fraction of ambient aerosol (a.k.a., particulate matter with less than 2.5 µm diameter) has been regulated under the National Ambient Air Quality Standards for several decades under the purview of the Clean Air Act (50 CFR § 50.7). Given the preponderance of epidemiological evidence concerning the health risks associated with particulate matter in and outside the mine environment,<sup>9,10</sup> it is reasonable to assume that refinements to existing mine safety standards will continue on the basis of particle size, aerosol composition, and other particulate matter sources, e.g., diesel particulate matter (DPM).

Investigations by the National Institute for Occupational Safety and Health (NIOSH) have evaluated the potential benefits of a field-portable, real-time, or end-of-shift infrared measurement to quantify RCS and DPM particle concentrations.<sup>11–15</sup> The hypothetical primacy of a field-portable infrared method over current approaches derives from the former's ability to measure specific dust constituents rapidly, in contrast to laboratory evaluations that lag behind mining operations by days or possibly weeks. The non-infrared personal dust monitor (PDM) has shown that direct measurement of mine dust helps miners modify work activities to minimize exposure.<sup>16</sup> Field-portable infrared monitoring therefore complements existing approaches like the PDM while expanding the range of dust species (e.g., RCS and DPM) quantified in real- or near-real time at the mine site.

Infrared methods measuring RCS and DPM directly on sampling filters have been previously studied.<sup>12,17–19</sup> These studies demonstrate two key aspects of infrared methods: They are sensitive to the many infrared-active species present on a filter and are specific to a target substance following calibration.<sup>20</sup> This turns out to be advantageous from an instrument design standpoint as the high resolution measurements of absorbance over a wide range of wavelengths in an infrared spectrum may be used to develop multivariate calibrations that quantify aggregate species (e.g., total dust mass), indirectly measured proxies such as aluminum mass,<sup>21</sup> infrared-active functional groups,<sup>22</sup> and molecules such as RCS.<sup>23</sup> However, studies sometimes differ in their use of field samples, analytical

standards, or both to develop calibrations. Field samples are advantageous as they provide a snapshot of a real mine environment which forces calibration parameters to adjust predictions for interferences and background. While mine dusts are aerosolized and measured at NIOSH Spokane, Washington, on a regular basis, training a calibration using field samples frees the analyst from having to identify and select the correct mineral standards to use for calibration. Selecting all the appropriate standards for any calibration is not a trivial task as, for example, Walker and colleagues identified 29 mineral phases present in 130 field samples collected from seven mining commodities using X-ray diffraction.<sup>24</sup> Most samples contained at least five mineral phase illustrating the complexity of these sample matrices. On the other hand, a major disadvantage of using field samples stems from possibly introducing uncontrolled (and unknown) factors into the predictive model leading to increased inaccuracy and/or imprecision in some or all sample predictions.

Quantifying total dust mass, or other complex species such as elemental carbon (EC), may pose a particular challenge to infrared methods as any unique aerosol source may disproportionately impact only a subset of samples used to train a calibration. For example, a single infrared calibration was incapable of accurately predicting EC in a subset of “atypical” samples impacted by DPM from mobile sources.<sup>25</sup> A two-calibration partial least squares (PLS) approach was needed to discriminate and then allocate samples to a “high DPM” or “low DPM” calibration. A similar calibration-after-classification scheme was also needed to quantify EC heavily impacted by wildfire smoke.<sup>21</sup> In both cases, infrared spectra did not cluster distinctly on principal component or PLS score plots and required metrics derived from the response variable (e.g., prediction errors) to guide cluster discovery.

These studies illustrate that the response variable and qualitative information is sometimes needed to identify poorly predicted sample clusters. Although these two-step procedures proved successful in their respective applications, there are methodological limitations to such approaches. First, the underlying factor required to discriminate samples must be known a priori (e.g., “wildfire smoke”) or identified post hoc to classify and predict samples using an appropriate model. Methods employed by Debus et al.<sup>21</sup> placed considerable effort on defining and refining sensible criteria to distinguish “wildfire” from “non-wildfire” samples following the failure of the “global” calibration. In other words, clusters were identified only after observing the fit between the response variable and the infrared spectra. Although defensible on scientific grounds, criteria were exogenous to modeling and thus not involved directly in estimating (or “adjusting”) calibration parameters to handle wildfire smoke (or the absence therein). Another limitation relates to both studies using “hard” (either/or) classification boundaries to assign a sample to a calibration. Such an approach may prove suboptimal as aerosol composition exists on a continuum; i.e., hard classification risks partitioning samples (and their spectral variance) such that variation useful to model all samples is restricted to only one calibration.

The mixture of experts (MoE) finite mixture model integrates calibration, cluster discovery, and outlier identification into a single cohesive step thereby eliminating the need to identify and then independently model heterogeneous samples following the failure of a calibration.<sup>26,27</sup> Against the limitations of hard classification, finite mixture models also employ “soft” (probabilistic) clustering such that the spectral variance used for mass

predictions may be weighted and shared between cluster models using Bayesian model averaging.<sup>28</sup> This may prove particularly advantageous for multimodel predictions as a sample collected from one mine (e.g., limestone mine) may occasionally contain absorption bands and mineral phases more typical of those found at another mine. In such an instance, model averaging ensures that a higher weight is given to the calibration best fitting the pattern of sample absorption, irrespective of mine source. The MoE method also handles both continuous and categorical variables, allowing qualitative information (e.g., “mine type”) to endogenously inform parameter estimation. The MoE framework also allows outliers to be identified during model fitting. This decouples outlier identification from the need to use post hoc regression diagnostic plots which often assume a homogenous and normally distributed sample.<sup>29</sup>

The current study presents the novel application of MoE cluster models to quantify total respirable mass on polyvinylchloride (PVC) filters using Fourier-transform infrared (FT-IR) spectroscopy. Samples were collected at both coal and metal/nonmetal (MNM) mines, exposing calibrations to geologies encountered while mining four types of commodities including coal, limestone, sandstone, and silver. We first hypothesize that prediction accuracy (or conversely prediction bias) depends on commodity/mine type. A second hypothesis considers whether the type of sampling cassette used to collect dust onto PVC filters affects prediction accuracy, since deposition patterning on PVC filters varies based on the cassette type.<sup>30</sup> Hypotheses are tested in turn by first calibrating FT-IR spectra to gravimetric mass using a single PLS regression. The PLS model is treated as the standard of comparison by which to judge subsequent MoE models. PLS scores are then extracted and used as “expert” variables in MoE models of increasing complexity. First, a simple MoE model is fitted to test whether multiple clusters (including a “noise” cluster) are required to improve mass prediction. Next, we investigate whether including a categorical variable for “mine type” improves predictions. Finally, both “mine” and “cassette type” are used as categorical variables and compared to the PLS, simple MoE model, and MoE model that uses only “mine type” as a categorical variable. As bias minimization is our chief concern, 95% bootstrap confidence intervals then assess whether median bias deviates significantly from zero for each model. A parsimonious model showing unbiased predictions at coal, limestone, sandstone, and silver mines is considered optimal.

## Experimental

### Materials and Methods

**Sampling Locations and Methodology.**—A total of 156 coal mine dust samples were collected from nine active coal mines using standardized personal sampling trains as per CFR 30, Part 70.2. Coal mine dust samplers were worn by NIOSH personnel in each active mine and configured to collect only the respirable fraction of airborne particulate matter onto pre-weighed 37 mm-diameter PVC filters secured in sealed two- or three-piece cassettes (MSA, Inc.). Coarse particulates were removed prior to reaching the filter using a 10 mm Dorr–Oliver cyclone (50% cut point at 3.5  $\mu\text{m}$  aerodynamic diameter).<sup>31</sup> Sampling pumps were operated at a flow rate of 2.0 L·min<sup>−1</sup> for a standard eight-hour shift (Zefon, Inc.).

A total of 87 samples were collected from one limestone, two sandstone, and one silver mine using sampling trains standardized for MNM mines.<sup>32</sup> Samplers were configured identically to coal mines with the exception of a sampling pump flow rate of  $1.7 \text{ L} \cdot \text{min}^{-1}$ . After collection, all cassettes were removed from samplers and sealed in plastic bags for gravimetric analysis.

**Analytical Methods: Mass Measurements and FT-IR**—Gravimetric analysis was conducted according to NIOSH method 600.<sup>33</sup> The difference between filter pre- and post-weight was used to determine respirable dust mass (milligram, mg). Silica masses were also determined using NIOSH method 7500 in the case of the MNM samples and the NIOSH method 7603 in the case of coal dust samples. The distribution of sample mass according to mine type is presented in Table I.

A Bruker Alpha FT-IR spectrometer operating in transmission mode was used to measure mine dust spectra on PVC filters. The filters were mounted in a stainless-steel holder in the instrument sampling compartment such that the infrared beam interrogated the center of the sample at normal incidence.<sup>30</sup> The instrument was configured to operate at  $4 \text{ cm}^{-1}$  resolution and averaging 40 scans per sample. A spectrum was acquired using a room temperature pyroelectric detector in approximately one minute once the scan sequence was initiated.

Interferograms were zero-filled prior to Fourier transformation at a factor of two and apodized using the Blackman–Harris three-term function. Sample transmittance was ratioed against blank filter transmittance to adjust the measurement for filter and chamber background. The absorption spectrum was calculated as the negative decadic logarithm of this ratio. As only a single blank filter was used as background for all samples some artifacts were present in the absorption spectra, conditional upon deviations in filter thickness between the sample spectrum and reference blank. The final spectrum contained 2542 absorption channels over a wavenumber range of  $4000\text{--}400 \text{ cm}^{-1}$ .

Spectra from the thirteen mines were collected into a design matrix for multivariate calibration. This design matrix is organized such that the 243 rows are sample observations ( $N$ ), and the 2542 columns are the input variables/predictors supplied to the calibration models ( $p$ ). This  $N \times p$  design matrix may be presented in matrix and vector notation as:

$$X = (x_1, x_2, \dots, x_{N-1}, x_N)^T \text{ for } i = 1, \dots, N, \quad (1)$$

where  $X$  is the design matrix comprised of a collection of  $N$  vectors of length  $p$ . The  $i$ th sample (row) is regularly indicated as  $x_i$  throughout this work as context dictates. Furthermore, the dependent/response variable ( $y$ ) of gravimetric mass measurements is represented as an  $N$ -length vector with elements ( $y_i$ ) paired to the appropriate rows in the design matrix ( $x_i$ ). Following the partitioning of samples into training and testing sets, the subscript “c” or “t” is used to denote affiliation with the calibration and test set, respectively.

**Model Development and Validation**—Second derivative spectra were generated from absorption spectra using a second-order Savitzky–Golay filter with a 21-point window.<sup>34</sup> Second derivative spectroscopy has proven beneficial in minimizing the impact of background absorption on multivariate calibrations, particularly those that utilize dust-laden PVC filter spectra.<sup>23</sup> The design matrix presented in Eq. 1 was regenerated following second derivative transformation with the number of rows and columns remaining unchanged.

A principal component analysis (PCA) was next performed on mean-centered derivative spectra to ascertain the predominant absorption variations, clusters present among spectra, and to identify potential outliers.<sup>35</sup> The principal components explaining at least 95% of the variance in the spectra were considered for the subsequent exploratory analysis. Exploratory analysis involved plotted eigenvectors against wavenumbers whereby infrared vibrations were assigned to features (minima) known to absorb in mine spectra.<sup>36,37</sup> Interpreting eigenvectors was aided by studying PC-score plots that were pseudo-colored using qualitative information (e.g., mine type, sampling location) about the spectra.<sup>38</sup>

Samples were then grouped by site identifier and sorted by sample mass. Every third sample in each of the 13 groups (9 coal + 4 MNM sites) was placed in the test set yielding 156 and 87 samples for training/calibration and testing, respectively. Partitioning samples this way ensured that the mass distribution of the training and test samples were as similar as possible. This is particularly important as the median mass and range associated with each type of mine varied widely (Table I).

Partial least squares regression was chosen as the basis to compare subsequent models as it is arguably the de facto standard for multivariate spectroscopic calibrations.<sup>39</sup> Moreover, PLS regression has proven accurate and precise in determining aerosol species collected on polymeric filters in mine dusts and atmospheric samples.<sup>19,21,23,38,40–43</sup> The popularity of PLS stems from its “optimal” use of infrared absorption for calibration as well as the ease in visualizing the key relationships within and between the spectra.<sup>44</sup> Solution optimality follows from an orthogonal transformation of the spectroscopic predictors (i.e., columns of  $X$ ), constrained to maximize the covariance between the response variable and spectra. Orthogonal transformation solves the collinearity problem in spectra-based calibrations while covariance maximization ensures that components are information rich in terms of the response.<sup>45</sup>

Five-fold interleaved cross validation was used to select the number of PLS components (i.e., latent variables or LVs). Training samples were organized into groups (by site) and sorted to ensure that the cross-validation training and “test” sets exhibit analogous mass distributions during model selection. The number of PLS components were selected by minimizing the root mean square error of cross-validation.

Regression diagnostic plots aided in assessing heterogeneity in the sample data, the potential impact on calibration, and thus motivated the use of MoE models. First, each sample’s  $Q$  residual and Hotelling  $T^2$  value (i.e., squared Mahalanobis distance) are plotted, comparing the residual/unused spectral variance to the spectral variance used by the PLS model, respectively.<sup>46</sup> Critical values were calculated for both measures to flag extreme samples

( $\alpha = 0.05$ ). Calculating a critical limit is straightforward for  $T^2$  values as they are assumed to follow a chi-squared distribution.<sup>46</sup> The Wilson–Hilferty approximation estimates the critical value for  $Q$  residuals as they follow a noncentral chi-squared distribution.<sup>47,48</sup>

A second diagnostic plot compared Cook’s distance ( $D_{(i)}$ ) to  $Q$  residuals to visualize residual absorption in terms of a sample’s “influence” on the PLS calibration coefficients.<sup>49</sup> As a leave-one-out measure, the  $i$ th sample’s Cook’s distance was calculated by (i) deleting the  $i$ th case, (ii) fitting the PLS model, (iii) calculating the residuals, (iv) estimating the sum of square residuals, and (v) normalizing the result by mean square error of calibration scaled by the regression degrees of freedom ( $df_{PLS}$ ). Degrees of freedom were estimated using the kernel PLS algorithm.<sup>50</sup> Two cutoff values described a sample as either highly influential using an F-distribution ( $D_{(i)} \geq F(0.5, df_{PLS}, N_c - df_{PLS})$ ) or as moderately influential using a common heuristic ( $D_{(i)} > 4/N_c$ ).<sup>51</sup>

The univariate response case of the MoE model was implemented in this study and given as

$$f(y_i | x_i^G, x_i^E) = \sum_{g=1}^G \tau_g(x_i^G) \phi(y_i | \theta_g(\tilde{x}_i^E)) + \frac{\tau_0(x_i^G)}{V} \quad (2)$$

where  $f(y_i | x_i^G, x_i^E)$  describes the conditional probability density of  $y_i$  given the input variables  $x_i$  for the  $i$ th sample,  $\tau_g(x_i^G)$  the mixture component density for the  $g$ th cluster dependent on the “gate” variables  $x_i^G$ , and  $\phi(y_i | \theta_g(\tilde{x}_i^E))$  the  $g$ th Gaussian cluster estimated from the sample data conditioned on parameters  $\theta_g$  and the “expert” variables  $\tilde{x}_i^E$ .<sup>26</sup> The tilde ( $\sim$ ) over  $\tilde{x}_i^E$  indicates the  $i$ th observation is augmented to include a model intercept. A noise component,  $\tau_0(x_i^G)/V$ , is included in each MoE model to identify outliers and uses the gate variables to guide estimation when specified. A gated noise component assumes that the gate variables influence the fitted noise probabilities and leaves Eq. 2 unmodified. The so-called “expert only” models use a non-gated noise term ( $\tau_0(x_i^G) \rightarrow \tau_0$ ) that assumes noise probabilities are constant across samples. For all models tested, the hypervolume ( $V$ ) was set at five times MSHA’s PEL of respirable coal dust exposure per working shift ( $5 \cdot 1.5 \text{ mg/m}^3$ ; 30 CFR, 70.100) as we do not anticipate observing values larger than this in practice ( $\sim 6 \text{ mg/filter}$ ) nor do we expect samples this heavily loaded to obey linear absorption laws.

Gaussian (normal) clusters are parameterized as  $\theta_g(\tilde{x}_i^E) = \{\tilde{x}_i^E \gamma_g, \sigma_g^2\}$  such that  $\gamma_g$  are the linear calibration coefficients used to predict sample mass. This takes the familiar form:

$$\hat{y}_{ig} = \tilde{x}_i^E \hat{\gamma}_g \quad (3)$$

where  $\hat{y}_{ig}$  corresponds to the  $i$ th mass prediction for the  $g$ th cluster model and  $\hat{\gamma}_g$  the estimated calibration coefficients for the  $g$ th cluster model. Gate parameters (not shown) are estimated assuming a multinomial logistic form. Cluster variances are constrained

as equal for all models ( $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_G^2$ ). Cluster variances, mixture components, gate, and expert parameters are estimated by maximizing the log likelihood of  $f(y_i | x_i^G, x_i^E)$  via the expectation maximization (EM) algorithm.<sup>52</sup> The EM algorithm was initialized using hierarchical clustering.

Scores taken from the PLS calibration are used as expert variables in the MoE models ( $\tilde{x}_i^E$ ). In this way, the MoE models reflect increasingly more complicated perturbations of the original PLS calibration dependent on: (i) The inclusion/exclusion of gating variables, (ii) the inclusion/exclusion of a noise component, (iii) the number of clusters selected, and (iv) the variance constraints placed on the clusters ( $\sigma_s^2$ ). In fact, as the PLS scores are used as the expert variables, it is straightforward to show that the original PLS calibration corresponds to a special case of the MoE model, i.e., one that uses a single-cluster ( $G = 1$ ), no gate variables, and fits no noise component to the data.

Three MoE models were fitted and tested to systematically assess the degree that gate variables and the noise component improve method accuracy. First, an “expert only” MoE model was fitted to assess whether multiple clusters (and a noise term) improve the prediction of sample mass. The second MoE model used “mine type” as a categorical gate variable with the four levels following directly from the “Mine Type” column in Table I. The “cassette type” covariate was then included as a second gate variable with the three levels given in the “Cassette Type” column in Table I. For all fitted models, the optimal number of clusters ( $G$ ) was identified as the one minimizing the Bayesian information criterion.<sup>53</sup>

Models are validated and compared on figures of merit calculated from test set data. Non-robust metrics include the coefficient of determination ( $R^2$ ) and root mean square error of prediction (RMSEP).<sup>54</sup> Robust measures include the median absolute error (MAE), median absolute percentage error, median bias, and median percentage bias.<sup>37</sup> The MAEs are calculated as  $k$  times the median absolute difference between the measured and predicted mass or

$$MAE = k \times \text{median}(|y - \hat{y}|) \quad (4)$$

where the scalar  $k = 1.4826$  makes MAE a consistent estimator of average prediction error, i.e., MAE and RMSEP are asymptotically equivalent when errors are unbiased and follow a normal distribution.<sup>55</sup> Median absolute percentage error (MAPE) is calculated by normalizing the term  $y - \hat{y}$  by  $y$  in Eq. 4 with  $k = 100$  to convert to a percentage basis. Median bias and median percentage bias follow analogously from Eq. 4 by dropping the absolute value from the calculation. Again, errors are normalized and  $k$  set equal to 100 for median percentage bias.

Ninety-five percent confidence intervals are calculated for median and median percentage bias using a nonparametric bootstrap percentile approach.<sup>56</sup> Intervals were calculated by resampling training set residuals to generate 5000 bootstrap replicates. Confidence interval endpoint adjustments used bias correction and acceleration (i.e., BCA). Bootstrap intervals

are used to formally test the two-sided null hypothesis that median biases are equal to zero. If confidence intervals around bias estimates do not contain a zero the null hypothesis is rejected at  $\alpha = 0.05$  probability level.

Mass predictions for any MoE model are possible using two techniques when more than a single cluster is used for modelling. The first approach predicts mass in each sample using all available calibrations, providing  $G$  predictions of mass for a sample. Predictions are then multiplied by mixing probabilities ( $\hat{\tau}_g$ ) and summed to generate a final value of sample mass. Mass predictions under this framework are given as

$$\hat{y}_i = \sum_{g=1}^G \hat{\tau}_g \hat{y}_{ig} \quad (5)$$

With  $\hat{y}_i$  indicating the final estimate of sample mass. Alternatively, sample mass may be determined using the cluster model having the largest mixing probability or  $\max(\hat{\tau}_{ig})$ . We adopted the approach in Eq. 5, known as Bayesian model averaging,<sup>28</sup> as it protects against the possibility of false cluster assignment, ambiguous cluster composition, or the belief that a single “correct” cluster exists to determine sample mass under the fitted model. The latter point is particularly important as the levels of the “mine type” and “cassette type” gate variables are not strong proxies for mine geology and sampling deposition biases, respectively. We therefore expect MoE clusters to reflect information from mixed geologic sources and a model that imperfectly adjusts calibration factors for deposition differences (at best). Note that MoE outliers are still quantified for all test samples using Eq. 5; albeit, with probabilities renormalized on the  $G$  clusters (excluding the outlier cluster). Leaving outliers in the test set keeps a comparison between MoE models and the base-case PLS calibration as fair as possible as all models are compared on the same samples.

**Software**—All analyses were performed using R Statistical Software (v4.1.2; R Core Team). The BCA bootstrap percentile method was coded by the authors with computation time reduced using the cluster computing functions available in the parallel package (v4.1.2). The PCA and visualization were performed using the PCA Tools package (v2.5.15). The kernel PLS algorithm was available in the PLS package (v2.8-1). MoE models were fitted using the MoEClust package (v1.5.1). The PLS model was wrapped into the cross-validation search routine using control functions available in the caret package (v6.0-94).

## Results and Discussion

### Exploratory Analysis

Exploratory PCA indicates a clear contrast between spectra collected at the coal, silver, and sandstone mines ( $\approx$ PC2) to those collected at the limestone mine ( $\approx$ PC1; Figure 1a). The mass of limestone mine samples shows a very strong correlation to PC1 ( $\hat{r} = 0.933$ ) which may be visualized from the distribution of limestone mine spectra in Figure 1b. The eigenvector associated with PC1 shows that carbonate ( $\text{CO}_3$ ) absorption bands are primarily

responsible for the distribution of spectra on the first PC (see Section S1, Supplemental Material, for PCA eigenvectors). As limestone is rich in calcium carbonates, the mass range of limestone mine samples is highest in these samples (Table I), and carbonate absorption is very strong in the mid-infrared it is not surprising that these variations dominate the PCA. As samples from the other mines tend to show little to no loading onto the first PC, we may infer that this component is a chemical factor associated with the carbonate-rich absorption prominent in limestone mine samples.<sup>57</sup>

Aside from the strong correlation between PC2 and non-limestone mass ( $\hat{r} > |0.75|$ ), the lesser principal components (i.e., PC3, PC4,...) are not easily interpreted and track absorption associated with silica, clay, and organic matter. Given their ambiguity, their presentation and discussion are available in supporting information. Ultimately, PCA frames our understanding of the information accessible to multivariate calibration in general and cluster modeling in particular: we expect the PLS calibration to require few components given the strong association between sample mass from all mine sites and the variations in principal infrared absorption. We also expect that valid MoE models will likely use few clusters for modeling.

### Multivariate Calibration

Cross-validation identified a two-component PLS calibration as optimal to determine sample mass. Consistent with Figure 1a, the PLS scores differentiate limestone mine samples from the other samples; albeit, with the PLS scores rotated away from the ordinate and abscissa (see Figure S6, Supplemental Material). In terms of percentage of variance explained, the PLS and PCA components are nearly identical reflecting that PLS affects a straightforward rotation of the subspace spanned by the first two PCA components.

Figure 2a presents  $Q$  residuals plotted against  $T^2$  values. Heterogeneity is immediately apparent as spectra cluster by mine type (shapes/symbols). Although violations to normality are present, we see that all samples with total mass above the 95th percentile fall beyond the 95%  $T^2$  critical value illustrating a clear relationship between the modeled infrared absorption and total mass. Notably, this relationship is imperfect as a subset of coal mine samples with mass between the 75th and 95th percentile fall above the  $T^2$  critical limit, again highlighting sample heterogeneity. In contrast,  $Q$  residuals do not show a straightforward connection to total mass as high  $Q$  residual spectra span a large mass range and fall above the 75th percentile. Notably, all limestone mine samples (filled circles) fall above the  $Q$  residual critical value suggesting that a relatively significant part of the infrared absorption in limestone mine samples was not modeled.

Cook's distance plotted against  $Q$  residuals provides a more refined understanding of the connection between a sample's residual absorption and influence on the PLS parameter estimates (Figure 2b). Samples considered moderately influential always had significant  $Q$  residuals indicating that large  $Q$  residuals were a necessary but not sufficient indicator a sample's importance to calibration. This point is particularly salient with respect to limestone mine samples as many of the largest  $Q$  residuals are derived from limestone samples but only two are flagged as influential. In fact, moderately influential samples

were collected from sandstone, limestone, and coal mines suggesting that Cook's distance provides a more granular understanding of a sample's importance, especially when  $Q$  residuals alone prove difficult to interpret.

Table II presents figures of merit for the PLS and MoE cluster models. First, we see a strong linear association between the spectra and response variable ( $R^2$ ) indicating that the mid-infrared spectra contains the necessary information to predict sample mass using all methods tested. Model sufficiency is best determined and differentiated using robust measures of method performance, particularly the median absolute percentage error (MAPE) and bias metrics. Notably, the PLS method exhibits the largest percentage error relative to the cluster models as well as a statistically significant bias according to confidence intervals. The underperformance of PLS provides subsequent justification to employ the more complicated cluster modeling technique.

A caveat is worth mentioning before judging the relative size and importance of bias statistics. The figures of merit calculated in this study implicitly assume that the gravimetric mass measurements are treated as ground truth. Therefore, method bias (and conversely discussions of accuracy) as presented here does not reflect measurement bias in the strictest sense, i.e., a deviation in predicted mass from the true concentration of respirable mass present in the mine environment over a working shift. Rather, method bias reflects only a deviation between a calibration model's prediction and the gravimetric mass measurement, within which any (true) measurement or sampling biases may be embedded. True measurement biases, such as those associated with aerosol sampling, for example,<sup>58</sup> cannot be remedied using these calibration methods and may only be addressed by improving field and laboratory methodologies.

Table II shows that all MoE cluster models outperform the PLS method in terms of bias. Method bias improves immediately using the "expert only" MoE model (row 2), which only differs from the PLS solution by including a noise component to identify outliers. Table III illustrates that this model's ability to identify and ignore five training samples as "noise" (outliers in MoE nomenclature) during model fitting had a small impact on the size of the model coefficients but a consequently large impact on method accuracy. Four of the five outliers can be seen in Figure 2b as falling above the  $Q$  residual critical value and within the moderately influential range on Cook's distance. The fifth sandstone sample falls just outside and to the upper left of the mass of points in Figure 2b and near the centroid of Figure 2a. This illustrates that MoE models identify so-called inliers in addition to the more obvious outliers observed on regression diagnostic plots.<sup>59</sup> Moreover, comparing the width of confidence intervals indicates that the bias estimates associated with MoE models are more certain (narrower) than the PLS method which likely follows primarily from outlier/inlier identification.

Adding the "mine type" gate variable into the MoE model results in a two-cluster solution with one cluster capturing limestone mine samples and the other handling most of the remaining samples. All metrics again improve relative to the PLS calibration. This model shows mixed improvement in performance relative to the "expert only" MoE model; namely,

$R^2$  and RMSEP are slightly lowered and elevated, respectively, while robust measures such as MAE and MAPE improve (Table II). Bias metrics are negligibly different according to confidence intervals. This may suggest that outlier identification is the crucial aspect of the MoE method applied to these data. However, the 3% drop in MAPE is noteworthy and perhaps indicate that two clusters are indeed optimal for mass prediction. Adding the “cassette type” gate variable to the problem does not confer any clear gains in overall method performance with the exception of a small improvement in MAE. This possibly indicates an overfitted model. To test this, Cramér’s  $V$ , a measure of association between non-ordinal categorical variables, was calculated between the “mine type” and “cassette type” gate variables. A strong association was indicated ( $\phi_c = 0.901$ ) suggesting that “cassette type” is likely redundant; namely, the categories of the “cassette type” variable act as very good substitutes for the categories of the “mine type” variable in this dataset (see Section S2, Supplemental Material, for a detailed elaboration of this topic).

Figure 3 presents the regression analysis cross-plot comparing the PLS model to the “mine type” gated MoE model. Only an improvement in limestone mine predictions with large mass is noticeable at this scale. Four of the seven samples described as noise by the MoE model are obvious in Figure 3b as three are inliers with masses below 2 mg. Therefore, to appreciate the effect of this two-cluster MoE solution on the accuracy of test sample predictions we must turn to the decomposition and assessment of bias statistics by mine type.

### Method Accuracy and Mine Type

Table II provides prima facie evidence that method bias improves following cluster modeling. However, these metrics fail to assess, in fine detail, the impact of source heterogeneity on sample predictions since metrics provide only an overall/average measure of bias. Table IV presents median bias decomposed in terms of the “mine type” variable. Aside from the most complex MoE model, coal mine predictions are statistically unbiased. This important result indicates that coal mine samples are not the main drivers of average PLS bias presented in Table II. Limestone and sandstone mine biases are large and significant for the PLS calibration implicating these mine types as responsible for elevated bias, on average. Table IV also shows that the most complex MoE model is indeed likely overfitted as coal samples actually become biased following the inclusion of the “cassette type” gate variable into the fitting procedure.

All MoE models produce unbiased predictions for limestone mine samples indicating that carbonate and noncarbonate-rich absorption are adequately modeled after flagging and removing crucial training set outliers. It is important to remember that all test samples are predicted and used to calculate bias, regardless of their outlier status, in order to fairly assess method performance. Table S7, (Supplemental material) presents the training set outliers identified by the MoE models—many of which are inliers when judged by the cutoff values in Figure 2 (see Section S5, Supplemental Material). Outliers identified in all models (“All”) contained very high sample and RCS mass with all but one falling beyond the 90th and 86th percentile, respectively. Identifying and setting aside these samples during model training appears to be the crucial factor in eliminating bias in limestone predictions as well

as reducing sandstone mine bias. Two additional outliers common to both gated models show very high RCS content but moderate sample mass (Gate only). Again, we might infer that removing these two sandstone mine samples from training may have resulted in more accurate predictions; although, one cannot rule out the effect of including an additional cluster to improve performance. The most complex MoE model flagged six additional coal samples with very low RCS-to-mass ratio samples (Mine and Cassette Only). The removal of these pertinent low mass coal samples from the training set explains the significant shift in coal mine sample bias, again providing evidence of an overfitted solution.

## Conclusion

Method of experts finite mixture models were applied to field samples collected from U.S.-based coal, limestone, sandstone, and silver mines to determine their efficacy in predicting sample mass using FT-IR spectra. Determining sample mass using an FT-IR approach may prove valuable for field applications at U.S. mines given the low cost and high throughput associated with measuring an infrared spectrum. Furthermore, a principal aim of this work was to determine whether accurate mass predictions, as evaluated formally using 95% nonparametric bootstrapped confidence intervals, were possible for these samples given both their mixed origin (coal versus MNM mines) and the difference between aerosol sampling in coal and noncoal mines.

Overall, it was possible to demonstrate systematically that MoE cluster models provide an improvement over a PLS calibration, trained and tested on the same samples, in terms of the MAE and a statistically significant improvement in model accuracy (prediction bias). An evaluation of the MoE parameters highlighted that the outlier identification capabilities of the MoE method, used to flag and isolate samples during model fitting, likely had the greatest effect on method performance. This result was particularly important as the only difference between the PLS calibration and the simplest MoE model is the inclusion of a noise component suggesting that, at minimum, one-cluster MoE models with appropriate noise constraints provide a benefit to the now routine application of PLS calibrations to spectroscopic data.

A more complex, two-cluster MoE model was required to accurately predict sandstone mine sample mass, with this solution guided by a qualitative gate variable related to the type of mine the sample was collected from ("mine type"). Prediction accuracy declined after including another gate variable for the type of cassette used during aerosol sampling ("cassette type"), likely due to its redundancy to the model as well as the consequentially damaging impact on outlier identification. Generally speaking, this study demonstrated the value of the MoE method's ability to incorporate categorical variables into multivariate calibrations. The implications of this finding should not be understated: the MoE framework provides a simple approach to introduce (often abundant) laboratory and field-sample metadata into a calibration to inform predictive modeling. Ultimately, we recommend that future work explore whether cluster models provide any benefit to improving method accuracy in calibrations impacted by mineral absorption/composition heterogeneity.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

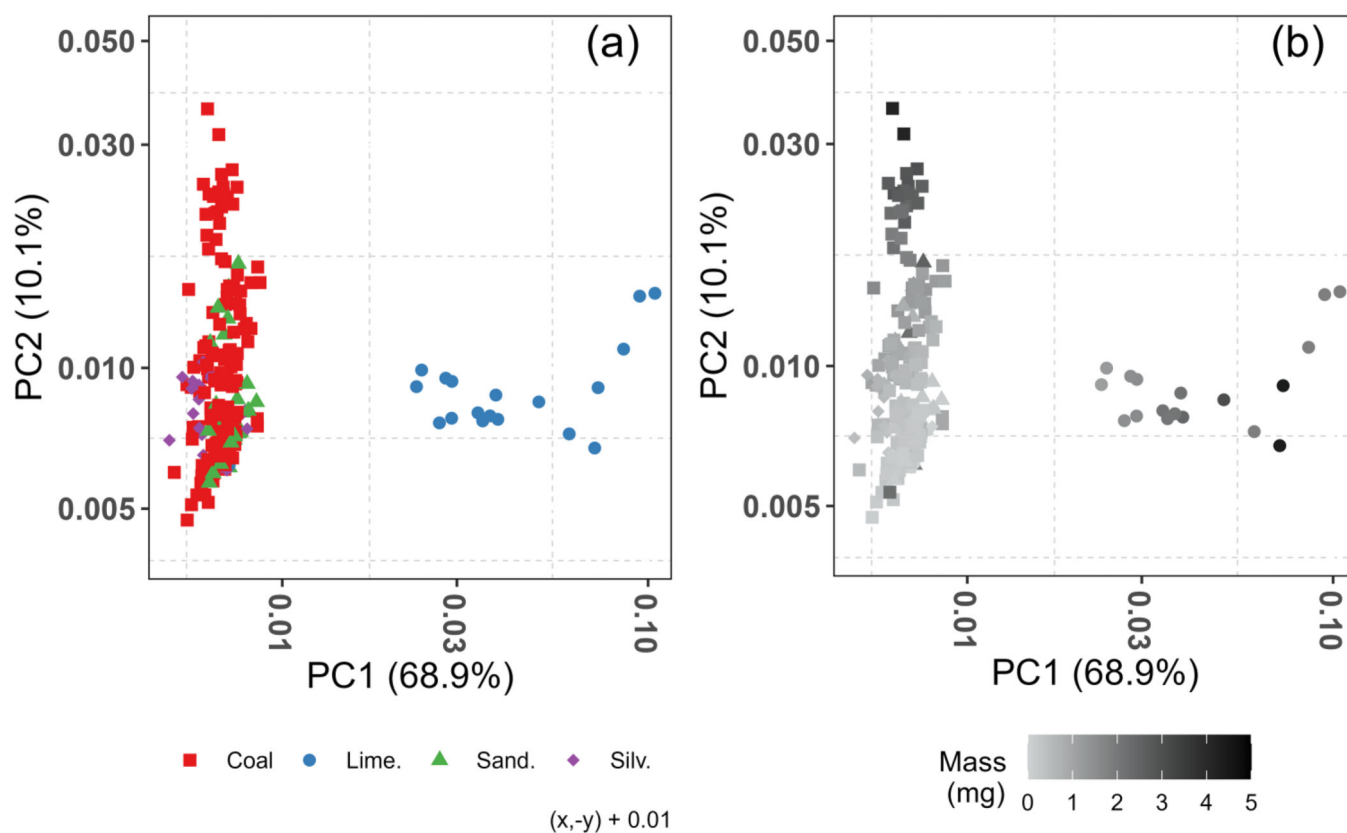
## References

1. National Academies of Sciences, Engineering, and Medicine. Monitoring and Sampling Approaches to Assess Underground Coal Mine Dust Exposures. Washington, DC: The National Academies Press, 2018.
2. Fazio JC, Gandhi SA, Flattery J, Heinzerling A, et al. "Silicosis Among Immigrant Engineered Stone (Quartz) Countertop Fabrication Workers in California". *JAMA Intern. Med* 2023. 183(9): 991–998. doi:10.1001/jamainternmed.2023.3295. [PubMed: 37486642]
3. Antao VC, Petsonk EL, Sokolow LZ, Wolfe AL, et al. "Rapidly Progressive Coal Workers' Pneumoconiosis in the United States: Geographic Clustering and Other Factors". *Occup. Environ. Med* 2005. 62(10): 670–674. doi:10.1136/oem.2004.019679. [PubMed: 16169911]
4. Leung CC, Yu ITS, Chen W. "Silicosis". *Lancet*. 2012. 379 (9830): 2008–2018. doi:10.1016/S0140-6736(12)60235-9. [PubMed: 22534002]
5. Sato T, Shimosato T, Klinman DM. "Silicosis and Lung Cancer: Current Perspectives". *Lung Cancer (Auckl)*. 2018. 9: 91–101. doi:10.2147/LCTT.S156376. [PubMed: 30498384]
6. Straif K, Benbrahim-Tallaa L, Baan R, Grosse Y, et al. "A Review of Human Carcinogens—Part C: Metals, Arsenic, Dusts, and Fibres". *Lancet Oncol*. 2009. 10(5): 453–454. doi:10.1016/S1470-2045(09)70134-2. [PubMed: 19418618]
7. Ehrlich R, Akugizibwe P, Siegfried N, Rees D. "The Association Between Silica Exposure, Silicosis and Tuberculosis: A Systematic Review and Meta-Analysis". *BMC Public Health*. 2021. 21(1): 953. doi:10.1186/s12889-021-10711-1. [PubMed: 34016067]
8. Boudigaard SH, Schlünssen V, Vestergaard JM, Søndergaard K, et al. "Occupational Exposure to Respirable Crystalline Silica and Risk of Autoimmune Rheumatic Diseases: A Nationwide Cohort Study". *Int. J. Epidemiol* 2021. 50(4): 1213–1226. doi:10.1093/ije/dyaa287. [PubMed: 33462590]
9. Feng S, Gao D, Liao F, Zhou F, Wang X. "The Health Effects of Ambient PM<sub>2.5</sub> and Potential Mechanisms". *Ecotoxicol. Environ. Saf* 2016. 128: 67–74. doi:10.1016/j.ecoenv.2016.01.030. [PubMed: 26896893]
10. Atkinson RW, Kang S, Anderson HR, Mills IC, Walton HA. "Epidemiological Time Series Studies of PM<sub>2.5</sub> and Daily Mortality and Hospital Admissions: A Systematic Review and Meta-Analysis". *Thorax*. 2014. 69(7): 660–665. doi:10.1136/thoraxjnl-2013-204492. [PubMed: 24706041]
11. Noll J, Janisko S, Mischler SE. "Real-Time Diesel Particulate Monitor for Underground Mines". *Anal. Methods* 2013. 5 (12): 2954–2963. doi:10.1039/C3AY40083B.
12. Miller AL, Drake PL, Murphy NC, Noll JD, Volkwein JC. "Evaluating Portable Infrared Spectrometers for Measuring the Silica Content of Coal Dust". *J. Environ. Monit* 2012. 14(1): 48–55. doi:10.1039/c1em10678c. [PubMed: 22130611]
13. Cauda E, Miller A, Drake P. "Promoting Early Exposure Monitoring for Respirable Crystalline Silica: Taking the Laboratory to the Mine Site". *J. Occup. Environ. Hyg* 2016. 13(3): D39–D45. doi:10.1080/15459624.2015.1116691. [PubMed: 26558490]
14. Cauda E, Dolan E, Cecala A, Louk K, Yekich M, et al. "Benefits and Limitations of Field-Based Monitoring Approaches for Respirable Dust and Crystalline Silica Applied in a Sandstone Quarry". *J. Occup. Environ. Hyg* 2022. 19(12): 730–741. doi:10.1080/15459624.2022.2132257. [PubMed: 36219680]
15. Pampena JD, Cauda EG, Chubb LG, Meadows JJ. "Use of the Field-Based Silica Monitoring Technique in a Coal Mine: A Case Study". *Min. Metall. Explor* 2020. 37(2): 717–726. doi:10.1007/s42461-019-00161-0. [PubMed: 35836821]

16. Peters RH, Vaught C, Hall EE, Volkwein JC. "Miners' Views About Personal Dust Monitors". J. Int. Soc. Respir. Prot 2007. 24(III–IV): 74–92.
17. Ashley EL, Cauda E, Chubb LG, Tuchman DP, Rubinstein EN. "Performance Comparison of Four Portable FTIR Instruments for Direct-on-Filter Measurement of Respirable Crystalline Silica". Ann. Work Exposures Health. 2020. 64(5): 536–546. doi:10.1093/annweh/wxaa031.
18. Hart JF, Autenrieth DA, Cauda E, Chubb L, et al. "A Comparison of Respirable Crystalline Silica Concentration Measurements Using a Direct-on-Filter Fourier Transform Infrared (FTIR) Transmission Method vs. a Traditional Laboratory X-ray Diffraction Method". J. Occup. Environ. Hyg 2018. 15(10): 743–754. doi:10.1080/15459624.2018.1495334. [PubMed: 29985762]
19. Parks DA, Griffiths PR, Weakley AT, Miller AL. "Quantifying Elemental and Organic Carbon in Diesel Particulate Matter by Mid-Infrared Spectrometry". Aerosol Sci. Technol 2021. 55 (9): 1014–1027. doi:10.1080/02786826.2021.1917764.
20. Saah AJ, Hoover DR. "'Sensitivity' and 'Specificity' Reconsidered: The Meaning of These Terms in Analytical and Diagnostic Settings". Ann. Intern. Med 1997. 126(1): 91–94. doi:10.7326/0003-4819-126-1-199701010-00026. [PubMed: 8992938]
21. Debus B, Weakley AT, Takahama S, George KM, et al. "Quantification of Major Particulate Matter Species from a Single Filter Type Using Infrared Spectroscopy –Application to a Large-Scale Monitoring Network". Atmos. Meas. Tech 2022. 15(9): 2685–2702. doi:10.5194/amt-15-2685-2022.
22. Boris AJ, Takahama S, Weakley AT, Debus BM, et al. "Quantifying Organic Matter and Functional Groups in Particulate Matter Filter Samples from the Southeastern United States: Part 1: Methods". Atmos. Meas. Tech 2019. 12 (10): 5391–5415. doi:10.5194/amt-12-5391-2019.
23. Weakley AT, Miller AL, Griffiths PR, Bayman SJ. "Quantifying Silica in Filter-Deposited Mine Dusts Using Infrared Spectra and Partial Least Squares Regression". Anal. Bioanal. Chem 2014. 406(19): 4715–4724. doi:10.1007/s00216-014-7856-y. [PubMed: 24830397]
24. Walker RLT, Cauda E, Chubb L, Krebs P, et al. "Complexity of Respirable Dust Found in Mining Operations as Characterized by X-ray Diffraction and FTIR Analysis". Minerals. 2021. 11(4): 383. doi:10.3390/min11040383.
25. Weakley AT, Takahama S, Wexler AS, Dillner AM. "Ambient Aerosol Composition by Infrared Spectroscopy and Partial Least Squares in the Chemical Speciation Network: Multilevel Modeling for Elemental Carbon". Aerosol Sci. Technol 2018. 52(6): 642–654. doi:10.1080/02786826.2018.1439571.
26. Murphy K, Murphy TB. "Gaussian Parsimonious Clustering Models with Covariates and a Noise Component". Adv. Data Anal. Classif 2020. 14(2): 293–325.
27. Lamont AE, Vermunt JK, Van Horn ML. "Regression Mixture Models: Does Modeling the Covariance Between Independent Variables and Latent Classes Improve the Results?". Multivar. Behav. Res 2016. 51(1): 35–52.
28. Hinne M, Gronau QF, Van Den Bergh D, Wagenmakers E-J. "A Conceptual Introduction to Bayesian Model Averaging". Adv. Methods Pract. Psychol. Sci 2020. 3(2): 200–215.
29. Jackson JE, Mudholkar GS. "Control Procedures for Residuals Associated with Principal Component Analysis". Technometrics. 1979. 21(3): 341–349.
30. L Miller A, Drake PL, Murphy NC, Cauda EG, et al. "Deposition Uniformity of Coal Dust on Filters and its Effect on the Accuracy of FTIR Analyses for Silica". Aerosol. Sci. Technol 2013. 47(7): 724–733. [PubMed: 26719603]
31. Colinet JF, Halldin CN, Schall J. Best Practices for Dust Control in Coal Mining. Pittsburgh: U.S. Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Office of Mine Safety and Health Research, 2021. doi:10.26616/NIOSH-PUB2021119.
32. U.S. Department of Labor. "Particulate Sampling Methods". In: Health Inspection Procedures Handbook. Washington, D.C.: Mine Safety and Health Administration, 2020. Chap. 2, Pp. 2–1–2–13. <https://arlweb.msha.gov/READROOM/HANDBOOK/PH20-V-4.pdf> [accessed Oct 15 2024].
33. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. "Particulates Not Otherwise Regulated, Respirable: Method 0600". In: NIOSH Manual of Analytical Methods. Cincinnati, Ohio: National Institute for Occupational Safety and Health, 1998. Pp. 1–6. <https://www.cdc.gov/niosh/docs/2003-154/pdfs/0600.pdf> [accessed Oct 15 2024].

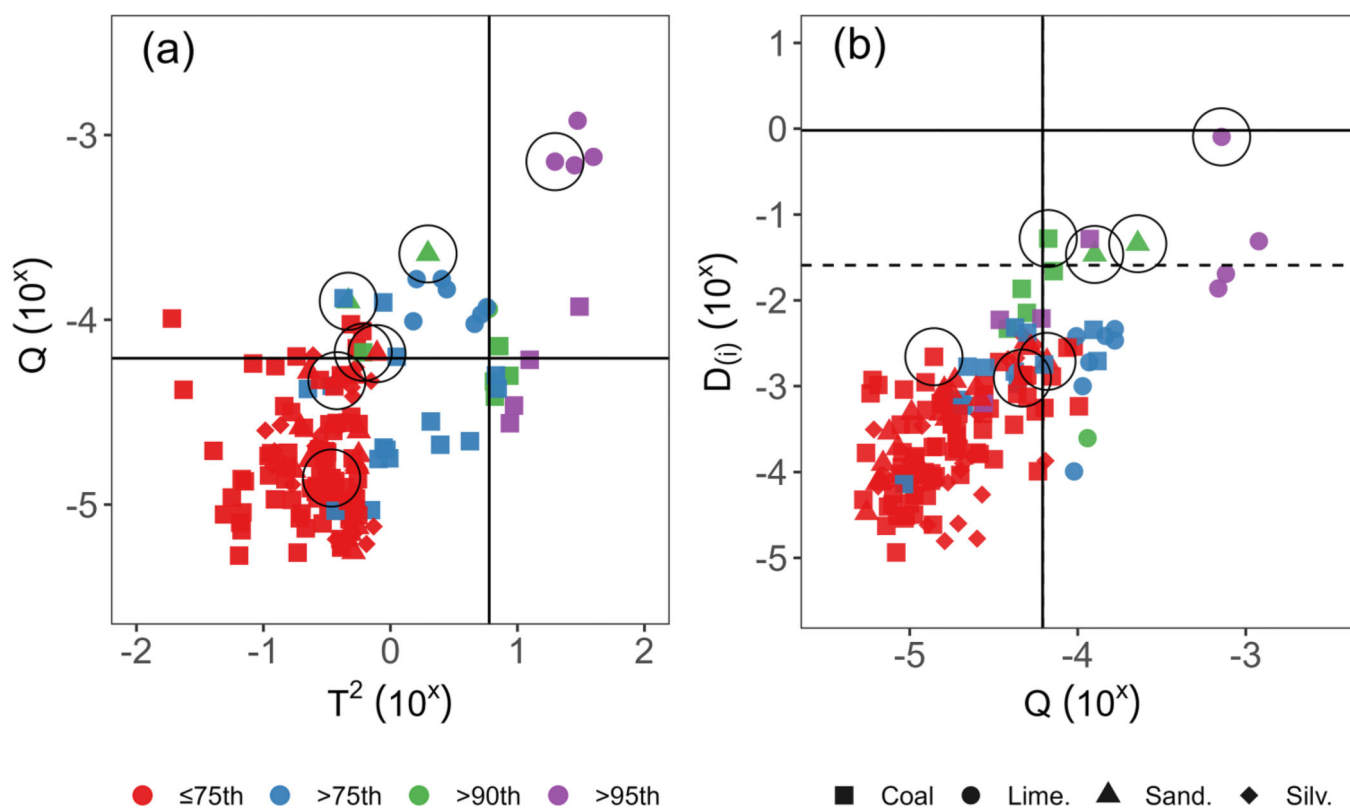
34. Savitzky A, Golay MJE. "Smoothing and Differentiation of Data by Simplified Least Squares Procedures". *Anal. Chem* 1964. 36(8): 1627–1639.
35. Abdi H, Williams LJ. "Principal Component Analysis". *WIREs Comput. Stat* 2010. 2(4): 433–459.
36. Gemperline PJ "Principal Component Analysis". In: Gemperline PJ, editor. *Practical Guide to Chemometrics*. Boca Raton, Florida: CRC/Taylor and Francis, 2006. Chap. 4, Pp. 96–103.
37. Weakley AT, Takahama S, Dillner AM. "Ambient Aerosol Composition by Infrared Spectroscopy and Partial Least-Squares in the Chemical Speciation Network: Organic Carbon with Functional Group Identification". *Aerosol Sci. and Technol* 2016. 50(10): 1096–1114.
38. A.L., Weakley AT, Griffiths PR, Cauda, Bayman. "Direct-on-Filter  $\alpha$ -Quartz Estimation in Respirable Coal Mine Dust Using Transmission Fourier Transform Infrared Spectrometry and Partial Least Squares Regression". *Appl. Spectrosc* 2017. 71(5): 1014–1024. [PubMed: 27645724]
39. Wold S, Sjöström M, Eriksson L. "PLS-Regression: A Basic Tool of Chemometrics". *Chemom. Intell. Lab. Syst* 2001. 58(2): 109–130.
40. Takahama S, Dillner AM, Weakley AT, Reggente M, et al. "Atmospheric Particulate Matter Characterization by Fourier Transform Infrared Spectroscopy: A Review of Statistical Calibration Strategies for Carbonaceous Aerosol Quantification in US Measurement Networks". *Atmos. Meas. Tech* 2019. 12(1): 525–567.
41. Wolfe C, Chubb L, Walker R, Yekich M, Cauda E. "Monitoring Worker Exposure to Respirable Crystalline Silica: Application for Data-Driven Predictive Modeling for End-of-Shift Exposure Assessment". *Ann. Work Exposures Health* 2022. 66(8): 1010–1021.
42. Stacey P, Clegg F, Sammon C. "Multicomponent Measurement of Respirable Quartz, Kaolinite and Coal Dust Using Fourier Transform Infrared Spectroscopy (FTIR): A Comparison Between Partial Least Squares and Principal Component Regressions". *Ann. Work Exposures Health* 2022. 66(5): 644–655.
43. Stach R, Barone T, Cauda E, Mizaikoff B. "A Novel Calibration Method for the Quantification of Respirable Particles in Mining Scenarios Using Fourier Transform Infrared Spectroscopy". *Appl. Spectrosc* 2021. 75(3): 307–316. [PubMed: 33031006]
44. Nadler B, Coifman RR. "Partial Least Squares, Beer's Law and the Net Analyte Signal: Statistical Modeling and Analysis". *J. Chemom* 2005. 19(1): 45–54.
45. Næs T, Mevik B-H. "Understanding the Collinearity Problem in Regression and Discriminant Analysis". *J. Chemom* 2001. 15(4): 413–426.
46. Jackson JE, Morris RH. "An Application of Multivariate Quality Control to Photographic Processing". *J. Am. Stat. Assoc* 1957. 52(278): 186–199.
47. Severo NC, Zelen M. "Normal Approximation to the Chi-Square and Non-Central F Probability Functions". *Biometrika*. 1960. 47(3–4): 411–416.
48. Nomikos P, MacGregor JF. "Multivariate SPC Charts for Monitoring Batch Processes". *Technometrics*. 1995. 37(1): 41–59.
49. Cook RD, Weisberg S. "Assessment of Influence". In: *Residuals and Influence in Regression*. New York: Chapman and Hall, 1982. Chap. 3, Pp. 116.
50. Krämer N, Sugiyama M. "The Degrees of Freedom of Partial Least Squares Regression". *J. Am. Stat. Assoc* 2011. 106(494): 697–705. <https://www.jstor.org/stable/41416403>
51. Altman N, Krzywinski M. "Analyzing Outliers: Influential or Nuisance?". *Nat. Methods* 2016. 13(4): 281–282. [PubMed: 27482566]
52. Dempster AP, Laird NM, Rubin DB. "Maximum Likelihood from Incomplete Data via the EM Algorithm". *J. R. Stat. Soc. Ser. B Methodol* 1977. 39(1): 1–38.
53. Burnham KP, Anderson DR. "Multimodel Inference: Understanding AIC and BIC in Model Selection". *Sociol. Methods Res* 2004. 33(2): 261–304.
54. Næs T, Isaksson T, Fearn T, Davies T. "Validation". In: *A User-Friendly Guide to Multivariate Calibration and Classification*. Chichester, UK: NIR Publications, 2002. Chap. 13, Pp. 155–175.
55. Rousseeuw PJ, Croux C. "Alternatives to the Median Absolute Deviation". *J. Am. Stat. Assoc* 1993. 88(424): 1273–1283.

56. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Boca Raton, Florida: Taylor and Francis, 1994.
57. Malinowski ER. Factor Analysis in Chemistry. New York: Wiley, 2002.
58. Bartley DL, Breuer GM. "Analysis and Optimization of the Performance of the 10 mm Cyclone". Am. Ind. Hyg. Assoc. J 1982. 43(7): 520–528.
59. Workman J. "Process Chemometrics/Spectroscopy Terminology". NIR News. 1996. 7(4): 15–16.



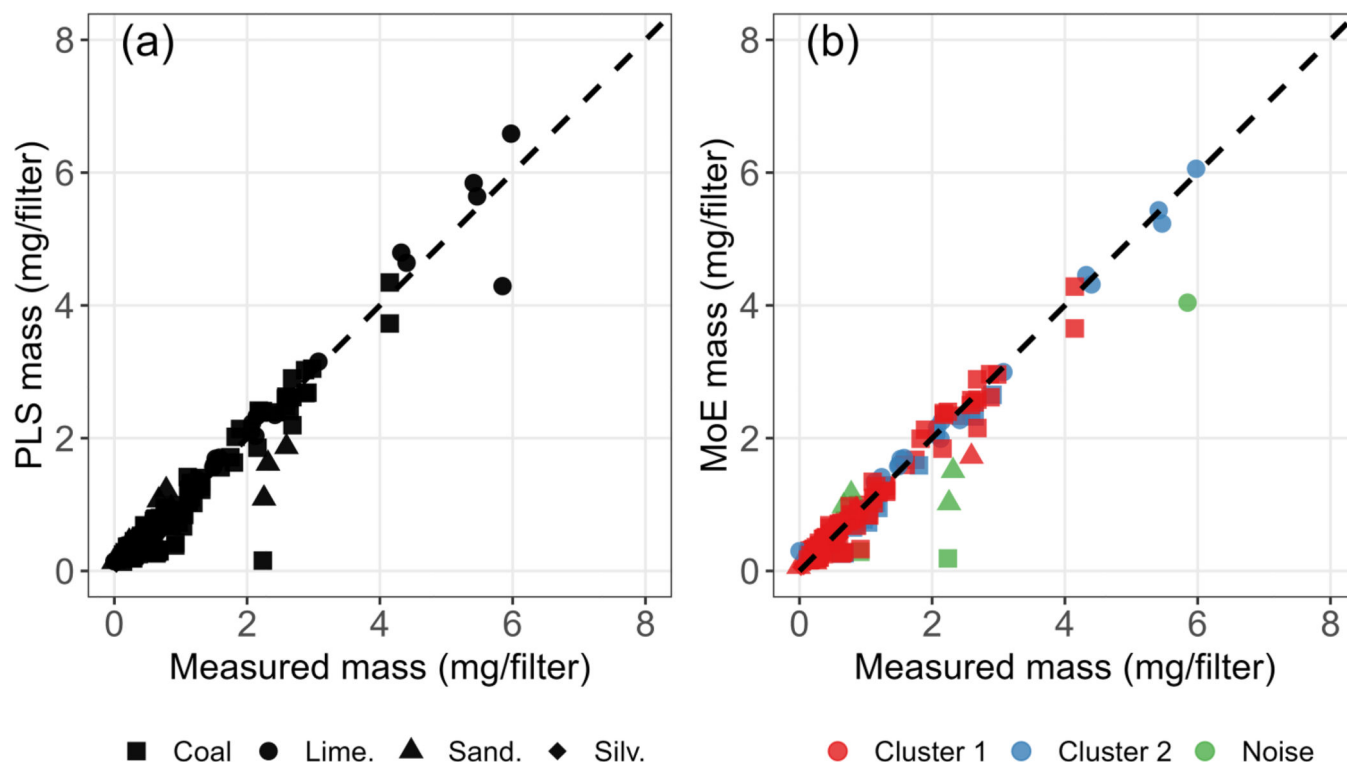
**Figure 1.**

Variation in limestone mine spectra dominated the PCA solution with the first component plainly mapping an increase in limestone absorption/mass (+PC1 direction) while explaining 68.9% of the spectral variance (a). Notably, samples from the other three mine sites loaded together onto PC2 with sample mass generally increasing in the + PC2 direction (after negation). Scores were plotted on a log scale after adding a small constant (0.01) to each score.



**Figure 2.**

(a) Diagnostic plots showing that unmodeled part of the FT-IR spectra (i.e.,  $Q$  residual;  $Q$ ) tends to increase log-linearly with the magnitude of Hotelling  $T^2$  values denoted by  $T^2$ . Lines denote critical values ( $\alpha = 0.05$ ). Spectra are colored according to mass percentiles and shapes denote mine affiliation. Encircled samples were those selected by the "mine type" gated MoE model. (b) Cook's distance and  $Q$  residuals aid in identifying samples with spectra that are poorly modeled on PLS components and influential to estimating calibration coefficients. Dashed and solid horizontal lines bound a region of moderate sample influence.



**Figure 3.**

(a) Regression analysis cross-plots comparing the PLS predicted mass to reference mass in test samples. Here we see positive bias in limestone mine samples for high mass (>4 mg). Several poor mass predictions from coal, sandstone, and limestone sources are apparent as outliers. (b) The “mine type” gated MoE model improves predictions for limestone mine samples while correctly assigning many of the obvious outliers to the “noise” cluster. The four outliers shown most clearly on this figure are also identified by the “expert only” MoE model.

Descriptive statistics calculated for samples collected from each mine type for 13 different mine sites. Coal samples from two sites were collected on three-piece cassettes totaling 16 samples. These are distinguished from MSA cassettes in the Samples column accordingly.

**Table I.**

| Mine type | Cassette type | Samples        | Sites | Mean (mg) | Median (mg) | Min (mg) | Max (mg) | Standard deviation (mg) |
|-----------|---------------|----------------|-------|-----------|-------------|----------|----------|-------------------------|
| Coal      | MSA + 3-Piece | 156 (140 + 16) | 9     | 0.9       | 0.6         | 0.1      | 4.2      | 0.83                    |
| Limestone | 2-Piece Clear | 19             | 1     | 2.8       | 2.1         | 0        | 6        | 1.82                    |
| Sandstone | 2-Piece Clear | 33             | 2     | 0.4       | 0.2         | 0        | 2.6      | 0.66                    |
| Silver    | 3-Piece       | 35             | 1     | 0.3       | 0.3         | 0        | 0.7      | 0.21                    |

Table II.

Test set figures of merit for each model considered. The number of components (latent variables, LV) and input variable number ( $p$ ) are indicated for the PLS model. The number of clusters (G) used by each model are indicated in parentheses (number) and gate variable type in parentheses (e.g., mine).

| Model                    | $R^2$  | RMSEP (mg) | MAE (mg) | MAPE (%) | Bias (mg)          | Bias (%)         |
|--------------------------|--------|------------|----------|----------|--------------------|------------------|
| PLS (LV = 2, $p$ = 2542) | 0.9719 | 0.190      | 0.147    | 16.6     | 0.04* (0.01,0.10)  | 5* (0.7, 13.2)   |
| MoE (1)                  | 0.9746 | 0.181      | 0.128    | 16.5     | 0.01 (−0.04, 0.01) | 0.7 (−5.1, 1.7)  |
| MoE (2) (mine)           | 0.9733 | 0.185      | 0.113    | 13.9     | 0.00 (0.00,0.05)   | −0.4 (−2.5, 6.5) |
| MoE (2) (mine, cassette) | 0.9717 | 0.191      | 0.103    | 14.7     | −0.01 (−0.02,0.02) | −2.1 (−9.6, 0.8) |

\* Significant bias statistics.

Calibration coefficients for PLS and MoE models. Coefficients are distinguished by cluster affiliation and PLS components denoted by the LV number. The number of training samples flagged as noise are indicated ( $N_{noise}$ ), when applicable. AU corresponds to the arbitrary units used to report transformed absorption in derivative spectra.

Table III.

| Model                     | Cluster 1      |             |             | Cluster 2      |             |             | $N_{noise}$ |
|---------------------------|----------------|-------------|-------------|----------------|-------------|-------------|-------------|
|                           | Intercept (mg) | PC1 (mg/AU) | PC2 (mg/AU) | Intercept (mg) | PC1 (mg/AU) | PC2 (mg/AU) |             |
| PLS (LV = 2, $p = 2542$ ) | 0.85           |             | 115.2       | –              | –           | –           | –           |
| MoE (1)                   | 0.81           | 65.7        | 115.7       | –              | –           | –           | 5           |
| MoE (2) (Mine)            | 0.99           | 62.0        | 117.7       | 0.58           | –4.5        | 135.1       | 7           |
| MoE (2) (Mine, cassette)  | 0.85           | 65.7        | 125.0       | 0.51           | –24.1       | 138.9       | 13          |

Method bias decomposed in terms of mine source and method type. For MoE models, the number of clusters are indicated as a number in parentheses and gate variable type in parentheses (e.g., Mine). Median percent bias statistics are presented in Section S3, Supplemental Material.

Table IV.

| Mine      | N  | PLS                            | MOE(1)                         | MOE (2) (Mine)      | MOE (2) (Mine, cassette)         |
|-----------|----|--------------------------------|--------------------------------|---------------------|----------------------------------|
| Coal      | 56 | 0.04 (−0.03, 0.06)             | 0.00 (−0.05, 0.01)             | −0.01 (−0.09, 0.02) | −0.01 <sup>*</sup> (−0.11, 0.00) |
| Limestone | 7  | 0.13 <sup>*</sup> (0.05, 0.24) | 0.02 (−0.09, 0.07)             | 0.08 (−0.35, 0.12)  | 0.03 (−0.39, 0.05)               |
| Sandstone | 12 | 0.11 <sup>*</sup> (0.09, 0.19) | 0.08 <sup>*</sup> (0.07, 0.14) | 0.01 (−0.01, 0.12)  | 0.00 (−0.04, 0.09)               |
| Silver    | 12 | −0.01 (−0.02, 0.08)            | −0.04 (−0.04, 0.02)            | −0.03 (−0.12, 0.07) | −0.04 (−0.13, 0.04)              |

<sup>\*</sup> Statistically significant values.