

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

STACEY, Peter <<http://orcid.org/0000-0002-2689-2890>>, CLEGG, Francis <<http://orcid.org/0000-0002-9566-5739>> and SAMMON, Chris <<http://orcid.org/0000-0003-1714-1726>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/30152/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

STACEY, Peter, CLEGG, Francis and SAMMON, Chris (2021). 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. *Annals of Work Exposures and Health*.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Original Article

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

Peter Stacey^{1,2,*}, Francis Clegg² and Christopher Sammon²

¹Health and Safety Executive, Harpur Hill, Buxton, Derbyshire, SK17 9JN, United Kingdom; ²Sheffield Hallam University, Materials and Engineering Research Institute, Sheffield, S1 1WB, United Kingdom.

*Author to whom correspondence should be addressed. E-mail: peter.stacey@hse.gov.uk

Submitted 12 March 2021; revised 4 September 2021; editorial decision 5 September 2021; revised version accepted 10 September 2021.

Abstract

Exposure to respirable crystalline silica (RCS) is potentially hazardous to the health of thousands of workers in Great Britain. Both X-ray diffraction (XRD) and Fourier transform infrared (FTIR) spectroscopy can be used to measure RCS to assess exposures. The current method outlined in the Health and Safety Executive's (HSE) Methods for the Determination of Hazardous Substances (MDHS) guidance series is 'MDHS 101 Crystalline silica in respirable airborne dust - Direct-on-filter analyses by infrared spectroscopy or x-ray'. This describes a procedure for the determination of time-weighted average concentrations of RCS either as quartz or cristobalite in airborne dust. FTIR is more commonly employed because it is less expensive, potentially portable and relatively easy to use. However, the FTIR analysis of RCS is affected by spectral interference from silicates. Chemometric techniques, known as Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR), are two computational processes that have the capability to remove spectral interference from FTIR spectra and correlate spectral features with constituent concentrations. These two common chemometric processes were tested on artificial mixtures of quartz and kaolinite in coal dust using the same commercially available software package. Calibration, validation and prediction samples were prepared by collecting aerosols of these dusts onto polyvinylchloride (PVC) filters using a Safety in Mines Personal Dust Sampler (SIMPEDS) respirable cyclone. PCR and PLSR analyses were compared when processing the same spectra. Good correlations between the target values, measured using XRD, were obtained for both the PCR and PLSR models e.g. 0.98–0.99 (quartz), 0.98–0.98 (kaolinite) and 0.96–0.97 (coal). The level of agreement between PCR and PLSR was within the 95% confidence value for each analyte. Slight differences observed between predicted PCR and PLSR values were due to

What's Important About This Paper?

Incidence of worker ill health caused by exposure to aerosols containing hazardous particulate, like respirable crystalline silica (RCS), remains an issue of international concern. Fourier Transform Infrared (FTIR) instruments are relatively inexpensive, portable and potentially measure many agents simultaneously, but interference can limit accuracy. This work explores the use of two chemometric data processing methods—principal component and partial least squares regression—that use commercially available software, to strip out interference from FTIR spectra. Both methods improved the accuracy of RCS measurement and provided a more complete characterization of a worker's exposure to a number of substances with a single measurement on one sample.

the number of optimal principal components applied to each chemometric process. The presence of kaolinite in these samples caused an 18% overestimation of quartz, for the FTIR, when following MDHS 101 without a chemometric method. Chemometric methods are a useful approach to obtain interference-free results for the measurement of RCS from some workplace environments and to provide a multicomponent analysis to better characterise exposures of workers.

Keywords: quartz, respirable crystalline silica, coal dust, kaolinite, principal component regression, partial least squares regression

Introduction

Crystalline silica is a generic name for a group of minerals that have the same chemical composition of silicon (Si) and oxygen (O) but with different crystalline structures known as polymorphs. The most common polymorphs of crystalline silica is quartz. Respirable dust is a health-related term for aerodynamic sizes of particles that enter the alveoli of the lung. The respirable definition includes particles with aerodynamic diameters of less than 16 μm and a median of 4 μm (CEN, 1993). Inhalation of respirable crystalline silica (RCS) is a potential hazard to the health of 560,000 workers in Great Britain (Brown *et al.*, 2012) because exposure to aerosols containing RCS can cause lung diseases like silicosis (HSE, 2003), lung cancer (Brown *et al.*, 2012) and chronic obstructive pulmonary disease (COPD) (Bergdahl *et al.*, 2004). One established approach to assess exposures is to measure a portion of the aerosol entering the breathing zones of workers. The workplace exposure limit for an 8-h time-weighted average exposure to RCS during a working shift in Great Britain is currently 100 $\mu\text{g m}^{-3}$ (HSE, 2005). Samples of aerosol are collected onto a filter using a respirable sampler. The most common respirable sampler used by the Health and Safety Executive is the Safety in Mines Personal Dust Sampler (SIMPEDS) which uses a cyclone action at a flow rate of 2.2 L min^{-1} . The two most common analytical techniques universally employed for RCS measurement of aerosol particulate on filters are x-ray diffraction (XRD) and Fourier transform infrared spectroscopy (FTIR). FTIR instruments measure

the absorption of infrared energy from Si–O vibrations, whereas XRD measures the reflections from the crystalline structure of the silica. Infrared measurements of RCS are more affected by potential mineral interference from other silicates present in the sample than XRD, because silicate mineral structures contain similar tetrahedral arrangements of Si and O atoms and so have similar Si–O vibrations. In addition, crystalline silica polymorphs and amorphous forms of silica also have similar convergent FTIR spectra patterns (Bye, 1994) affecting accurate quantification. FTIR instruments are more widely used to measure RCS in aerosol dust collected onto filters because they are currently much cheaper than using XRD and some models are more portable.

The development of portable (often battery powered) and low-cost infrared instruments enables the potential for the measurement of end-of-shift samples at the worksite (Cauda *et al.*, 2016). Portable instruments have an advantage since the hygienist can potentially measure and report exposure results soon after a work shift or work task is completed. The time taken to improve emission controls and lower airborne RCS concentrations emitted from the process can be significantly reduced, since it is not necessary to wait for a result from an external laboratory, which can take several weeks. On-site instruments rely on the application of a direct on-aerosol filter measurement method, which is the standard RCS analysis approach used in the series of methods developed by the Health and Safety Executive (HSE) in Great Britain (Foster and Walker, 1984, HSE, 2014) known as Methods

for the determination of hazardous substances (MDHS 101: Respirable crystalline silica by infrared and x-ray). Hence, the MDHS 101 measurement approach for RCS involves the analysis of the dust deposited on the aerosol sampling filter without any attempt to treat the dust to remove interference. Early work (Pickard *et al.*, 1985) demonstrated no significant difference in direct on-aerosol filter FTIR results compared with a direct on-aerosol-filter XRD method when measuring the same samples from foundries, brickworks and potteries. However, other researchers (Lee *et al.*, 2013) measuring quartz in coal dust samples, found significant interference from kaolinite when using indirect methods (0.2–23%); where the dust is recovered from the aerosol filter and concentrated onto a smaller filter for analysis (MSHA, 1994). In Great Britain, analysis of samples of coal for quartz followed an indirect infrared potassium bromide (KBr) disc analysis approach involving a furnace to heat the dust and decompose the kaolinite (Dodgson and Whitaker, 1973). However, this process involves the potential for sample loss during sample handling and processing. Currently, skilled judgement from a trained analyst is needed to interpret FTIR spectra with spectral interference. The subjective nature of this may add to the variability of results. The increased availability of low-cost FTIR portable equipment for on-site measurement may diversify the range of people involved in the interpretation of results, including those in roles who are not previously trained for this task. The potential for chemometric methods to accurately measure without the need for further spectral interpretation may lead to a ‘black-box’ approach facilitating an instant and less expensive analysis of a worker’s exposure. Currently, the chemometrics model may need a check to verify that it works in a particular workplace because it is not yet known if they universally applicable. However, the addition of workplace-specific samples, where target values are derived from a reference laboratory, could resolve any site-specific issues.

The purpose of this article is to compare the performance of both partial least squares regression (PLSR) and principal component regression (PCR) for the measurement of quartz collected on aerosol sampling filters and to compare the XRD method with the FTIR method in MDHS 101. Coal dust itself is also a potential cause of lung disease (Ross and Murray, 2004) so we also assessed if the parameters selected by the chemometric models can also quantify other major and potentially hazardous components of these samples.

Chemometric methods

Current regulatory analytical methods for the measurement of RCS using infrared instruments (NIOSH, 2004,

HSE, 2014) measure the peak height, from the base line to the maxima, of the characteristic absorbance bands for quartz at 800 cm^{-1} and 780 cm^{-1} . These methods use the Beer–Lambert Law assumption which states the attenuation of absorbance is directly proportional to the concentration of the substance and a constant known as the molar extinction coefficient. The model for the calibration is based on a simple regression between two sets of coordinates, where x is the mass of the analyte and y is the measured response. This approach is applied because most calibrations are conducted on test samples containing pure RCS (quartz or cristobalite). Correction due to the presence of other components is addressed at a later stage in the analytical method.

Chemometrics is a mathematical process that can examine the whole FTIR spectrum and access the influence of different infrared active components with overlapping absorbance from interferences (Breen *et al.*, 2008, Weakley *et al.*, 2014). The application of chemometric models to resolve the influence of amorphous silica on the quantification of crystalline silica in an FTIR spectrum was first applied in the 1990’s by Bye (1994), who employed PLSR to resolve artificial mixtures of crystalline silica (quartz and cristobalite) and amorphous silica. More recently, (Miller *et al.*, 2016) used PLSR to resolve the influence of kaolinite from the measurement of quartz on coal dust samples when using portable infrared instruments.

Chemometric techniques, like PLSR and PCR, are based on a technique known as Principal Component Analysis (PCA). The PCA is a way of reducing large amounts of data to a limited number of characteristic vector quantities (Jolliffe and Cadima, 2016). Spectral data, consisting of a series of paired points representing an absorbance at specific wavenumbers, is an ideal format for PCA. This technique is usually applied when there are multiple variables in both x and y . In this case discussed herein, y represents the multiple combinations of different constituents (masses of coal dust, kaolinite and quartz) and x are multiple versions of absorbencies from the resulting scans obtained. The PCA process introduces new coordinates by maximising the co-variance of these data and reducing the multiple points of x or y to vectors known as principal components (PC). The technique essentially provides vector coordinates for the bands that have a magnitude in the assembled spectra and are most important for each component. The PCA can provide a vector value (score) for each matrix component by fitting a line to the variables and determining the distance of a point to an origin. In PCR, the first stage of the process is to determine the direction of each absorbance from the mean value. This

PCR process then establishes a relationship for each substance in the matrix between the calibration spectra (associated with mass of coal dust, kaolinite, and quartz) with the most correlated vector values. The difference between PCR and PLSR is that PCR interrogates only the spectral information i.e. the absorbencies, whereas PLSR examines both spectral and concentration data (Hemmateenejad *et al.*, 2007). Both processes analyse multiple regressions to test the resulting models, however, PLSR contains separate calibrations for each component. PLSR has been widely used with infrared instruments to measure complex mixtures of components including recent studies investigating the measurement of quartz in Pittsburgh coal dust (Miller *et al.*, 2016) and a PLSR using a modified Monte Carlo process to eliminate unimportant spectral variables (Weakley *et al.*, 2014). The performance of PCR and PLSR has been studied and the differences between results are not generally considered significant although PLSR is generally preferred for FTIR analysis. The main difference is that the PLSR calibration often requires fewer latent variables for its prediction (Wentzell and Vega Montoto, 2003) which may increase the PLSR models efficiency when measuring mass values close to the background, since this reduces the likelihood of over modelling the background noise.

Method

Overview of the analytical process

A schematic representation of the analytical measurement process used in this study is shown in Figure 1.

The PVC filter is itself a spectral component that affects absorbance, so current direct on-aerosol filter analysis methods for quartz (e.g. MDHS 101) involve the subtraction of a spectrum of an unused PVC filter (HSE, 2014; Hart *et al.*, 2018). Before sample collection, unused PVC filter blanks were analysed using FTIR to determine their spectral profile. The unused filters were then weighed. Aerosols were generated and the filters were loaded with test mixtures following the description outlined below. Gravimetric analysis was performed to assess the total gravimetric mass of the loaded material and to later calculate the mass of coal dust. XRD was used to determine the mass of quartz and kaolinite in the test mixtures and the samples were later scanned using FTIR. The XRD measured values for the quartz and kaolinite were used as the target values assigned to each sample for use in the chemometric modelling and for comparison with FTIR values determined using MDHS 101.

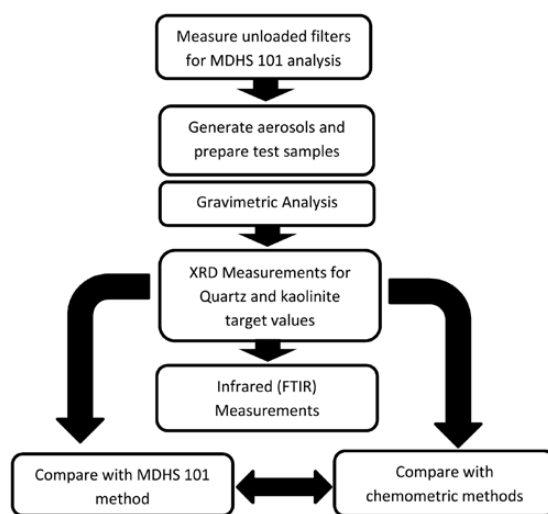


Figure 1. The analytical process.

Sample preparation

Several approaches for the preparation of calibration standards are possible; however, to be predictive of samples from different workplaces, the calibration and validation mixtures must contain representative proportions of relevant components i.e. the constituent minerals.

In this study, artificial mixtures were generated so that each component of interest could be evaluated and quantifiable. The proportions of each significant component were altered to decrease any collinearity within the model and increase the variability of component concentrations. Test filters containing mixtures of respirable dust were prepared following the method described in MDHS 101 (HSE, 2014) and Stacey (2019). Aerosols of a single constituent or mixtures of constituent minerals were generated in a glass dust cloud generator of the type shown in MDHS 101 (HSE, 2014). The test dust was placed in the bowl at the bottom of the generator and aerosolised using a short burst of pressurised air. The apparatus was earthed (grounded) to reduce the agglomeration of particles due to charge. Test samples of these aerosols were collected onto 25 mm diameter 5 µm pore size PVC filters (SKC Ltd, Blandford Forum, Dorset, UK) using a SIMPEDS (Casella UK Ltd, Bedford, United Kingdom) at a flow rate of 2.2 L min⁻¹. The flow rate through the respirable sampler was calibrated to its nominal value using a TSI 4100 flow calibrator (TSI Inc, Shoreview, USA). The bulk powders used for aerosol generation and their origin are listed in Table 1. Mineral samples can potentially contain other trace minerals as contaminants, which may affect the reliability of their quantification. Each powder was scanned using XRD

Table 1. Respirable powders included in the study and their origin.

Powder	Origin
Respirable Quartz	HSE reference power A9950, Buxton, United Kingdom (UK)
Kaolinite	Georgia, United States of America
Coal Dust (Coal 1)	Durrans Graded Coal Dust 30, James Durrans & Sons Ltd, Sheffield, UK
Coal Dust (Coal 2)	Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa

to assess the presence of other potential mineral components. XRD scans for coal dust from a variety of sources are shown in [Figure S1 \(Supplementary Information\)](#). Coal dusts identified as Polish, South African, Pittsburgh (United States of America), Coal Rank 502 (a coke coal) and Durrans were measured. The South African coal dust differs from the others because it was collected from ventilation filters in a mine, whereas the other samples were collected from pulverised coal. The mineral composition was similar for each coal dust. Quartz (SiO_2), kaolinite ($\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$), a mica type named muscovite 2M1 or 3T ($\text{KAl}_2(\text{AlSi}_3\text{O}_{10})(\text{OH})_2$), pyrite (Fe_2S) and some trace reflections that could be associated with calcite (CaCO_3), and a microcline (KAlSi_3O_8) were identified in the majority of samples. The most significant reflections were for quartz and kaolinite. The coal identified as Polish was relatively quartz-free although it had noticeable kaolinite reflections. Durrans coal was selected for further work because it had lower quartz and kaolinite reflections. FTIR scans for Durrans coal deposited onto a PVC filter are shown in [Figure 2](#) and compared with an unused filter.

Determination of the sample 'assigned' independent target value

XRD following the MDHS 101 method ([HSE, 2014](#)) was used as the reference method for the assigned values since XRD is less influenced by the presence of kaolinite as interference. The quantity of quartz and kaolinite in all samples was determined using an X-pert Pro MPD x-ray diffractometer (Malvern PANalytical Ltd, Cambridge, UK) with focusing Bragg-Brentano geometry. The XRD measurement conditions are described in the [supplementary information](#). The calibrations for quartz and kaolinite are shown in [Figures S2 and S3 \(Supplementary information\)](#).

The mass of coal was determined by subtracting the mass values for quartz and kaolinite from the gravimetric mass of the dust loaded onto each PVC filter. Filter weighing was conducted using a XP2U Mettler microbalance with a readability of $1 \mu\text{g}$ (Mettler-Toledo Ltd, Leicester, United Kingdom) in a balance room with controlled temperature and humidity maintained at 25°C and 50% relative humidity. Static charge was eliminated from the PVC filters before weighing using an Haug deioniser (Mettler-Toledo Ltd, Leicester, United Kingdom).

FTIR analysis

A Perkin Elmer Frontier FTIR instrument was used with Quant software (Perkin Elmer UK Ltd, Beaconsfield, UK). The standard conditions in MDHS 101 were used for FTIR analysis ([HSE, 2014](#)). Thirty-two scans were collected from 560 cm^{-1} to 1000 cm^{-1} with a resolution of 4 cm^{-1} . A previous study ([Miller et al., 2016](#)) used a resolution of 2 cm^{-1} but this takes longer and was found to make little or no difference to sensitivity ([Figure S4, Supplementary Information](#)). With a scan speed of 2 cm s^{-1} the total sample measurement time was less than 4 min. Spectral effects due to external factors were compensated for by running a background scan prior to analysis of each batch of filters.

Background correction

The first-order derivative background treatment was compared with the spectral subtraction approach of MDHS 101 using PVC filters containing pure quartz standard A9950; to assess if the first-order derivative approach could replace the current procedure. The same standards were used for calibration and prediction and both approaches used the Beer-lambert approach where the absorbance is proportional to the concentration. The only difference between the two approaches was the processing of the spectra before analysis. The absolute and relative differences were compared.

The chemometric process

The chemometric process involved the collection and measurement of three independent sets of samples; one for calibration, one for validation of the model and another to serve as a prediction set to assess how well the models compare. The calibration test samples included samples of pure analyte and combinations of mixtures of the constituent components (quartz, kaolinite and Durrans coal dust). The validation and prediction samples generally included all three constituents, although a small number of samples were included with, no kaolinite, just kaolinite and some samples containing only quartz obtained from the AIR

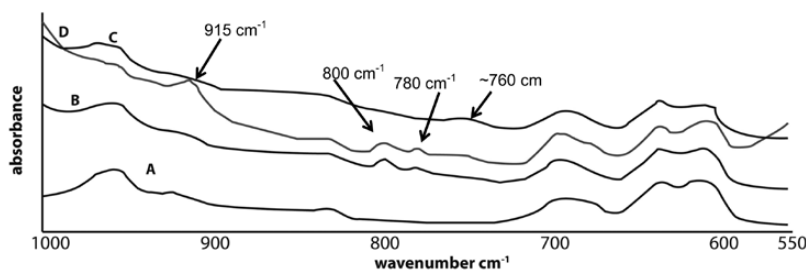


Figure 2. FTIR scans between 560 cm^{-1} and 1000 cm^{-1} wavenumbers showing a unloaded blank PVC filter (A), a PVC filter loaded with about 1.3 mg of Durran's coal dust and a mass of quartz (B), a PVC filter loaded with 1 mg of Durran's coal dust (C), and a sample of about 1.3 mg of coal dust from South Africa (D).

proficiency testing (PT) programme (LGC Ltd, Teddington, UK). AIR PT is a quality assurance programme that distributes samples to assess the proficiency of laboratories measuring contaminants in workplace atmospheres. Durran's coal dust (Coal 1) was used for the preparation of calibration and validation samples. Prediction samples to test the models also included coal dust from a South African mine (Coal 2). A list of the samples, loaded masses and their constituent proportions are presented in [Tables S2 to S4 \(Supplementary Information\)](#). The same samples were used for the PCR and the PLSR model.

Processing parameters

The spectral range was constrained in favour of obtaining a scan of the area of interest for the measurement of quartz. [Miller et al. \(2016\)](#) found that wavenumbers between 560 cm^{-1} and 1000 cm^{-1} contained the wavenumbers of most significance for the measurement of quartz in coal dust. There were few spectral wavenumber ranges that were not affected by the mineral constituents. The measured spectrum in the analysis was not subdivided further (e.g. to specifically reduce the proportion of redundant spectra to improve kaolinite or coal dust measurement). Unlike other work ([Weakley et al., 2014](#), [Miller et al., 2016](#)), the spectrum of the loaded test sample was not corrected by subtraction of a spectrum of an unused PVC filter. A first-order derivative was applied to each scan to normalise the influence of the spectral bands from the filter material. A beneficial effect of the application of the first-order derivative is to suppress off-sets (i.e. variation about the expected vibrational absorbencies). The most common reason for this to occur in this work was the difference in absorbance due to the depth of the PVC filter and the sample loading i.e. the smaller the sample loading the more spectral interference from the filter. A second-order derivative was attempted because this process is potentially more sensitive to low analyte mass values; however, it tended to underestimate the higher loadings and was not used because it was not applicable

to the wider analytical range. No further weighting or normalisation was applied for processing the spectrum before applying the PLSR or PCR model.

All data were centred around the mean for the principal component analysis. The PCR model was centred around the mean for y (absorbance) and the PLSR model was centred around the mean for both x (constituents and masses) and y (constituents and absorbance). The maximum possible number of components was not restricted. The commercial software uses a validation protocol during the calibration stage, as well as an independent validation stage after calibration, to assess its predictive ability. Each calibration spectrum is sequentially removed and predicted from the recalculated model. The prediction error is a measure of all the differences obtained by the model during the calibration validation process. The validation set included both samples prepared with Durran's and South African coal dust and the results were corrected for a slight bias for kaolinite which the software applied to all analytes.

The influence of the number of principal components on the values reported for quartz in the prediction set was investigated using the PCR model. The similarity between values predicted by PCR with different numbers of PCs and the XRD assigned values were assessed from the 95% level of confidence intervals for the slope and intercept of their regressions obtained from Excel (Microsoft Corp, United States). Ideally, the slope value should not be significantly different from 1.00 and the intercept should not be significantly different from zero. The similarity between the PCR and the PLSR models was also determined in the same way, once the models were established with numbers of PCs that were a close match to the XRD assigned values.

Results and discussion

Calibration results

There were 62 and 19 samples in the calibration and validation sets, respectively. Statistical estimates generated

by the software for the quantification of each component using all 62 samples are shown in [Table 2](#).

The calculated values from PCR and PLSR calibration and validation sets compared with the assigned value of quartz determined using XRD are shown in [Figure 3](#).

A set of thirty samples used in the prediction set containing quartz in artificial mixtures with added kaolinite and South African coal dust (Coal 2) were prepared and evaluated against their measured quartz values ([Table S3 supplementary information](#)).

Evaluation of the number of principal components (PC)

The software package computationally processed the PCR model for quartz with six PCs, however, the software options for PCR allowed the user to specify the number of principle components for each analyte. Therefore different numbers of principle components were assessed to check if six was the best for prediction as well as to assess trends in the prediction values. Prediction values, obtained by processing PCR calibration with different numbers of PCs (3 to 7), were plotted against the assigned target values ([Figure S5, Supplementary Information](#)). The regression and average differences are provided in [Table S1 \(Supplementary Information\)](#). Changing the number of PCs alters the slope and intercepts coefficients. The slope of the regression line between the predicted model value and the assigned XRD value for the trend line with five PCs was one of two PCRs that were not significantly different from the ideal 1:1 trend line slope from the measured XRD values indicating a close statistical relationship (95% confidence interval of the slope was from 0.99 to 1.11). The average difference was $-6.5 \mu\text{g}$ with values for the slope and intercept of 1.05 and $-10.7 \mu\text{g}$, respectively. Predicted values from the PCR model with three PCs were also not significantly different from the XRD assigned values. The model with 3 PCs had the least

average difference with the XRD values ($0.86 \mu\text{g}$) and a better fitted linear relationship with the target values (Slope = 0.996 and Intercept = $1.1 \mu\text{g}$). The PCR model with five PCs was selected for further work to compare with the equal number of PCs selected by the PLSR method.

Prediction close to the XRD limit of detection (LOD)

Six prediction samples of South African coal (shown in [Figure 4](#)) dust contained very low masses of quartz ($7.7\text{--}9.5 \mu\text{g}$). These samples provided an opportunity to test whether the reduction in the number of PCs improved the LOD for quartz. Large negative values for these six samples (about $-20 \mu\text{g}$) occurred when the number of PCs was high ($n = 7$), which is associated with the larger intercept of this model ([Table S1, Supplementary Information](#)). However, poor predictions can also occur if the number of PCs is too low since the variances in data might not be represented ([Breen et al., 2008](#)). For the 5 PC model, the predicted quartz values for the six samples were close to zero and ranged from $-6.1 \mu\text{g}$ to $1.1 \mu\text{g}$ (average was $-1.5 \mu\text{g}$), suggesting that the method is unable to reliably predict very low masses of quartz close to the XRD limit of detection of about $8 \mu\text{g}$. When using 3 PCs, the predicted were closer to the XRD target values (average absolute difference was $2 \mu\text{g}$ compared with $8 \mu\text{g}$), indicating 3 PCs were better match with the XRD target values.

The need for blank matching

An added advantage of the chemometric software package was the opportunity to apply derivatives to normalise the spectra before processing, i.e. tangent lines to the graph of the function at that point. The comparability of the predicted values obtained from the models with the linear 1:1 relationship demonstrates that the derivatisation did not have a negative impact on the results. A separate experiment compared eight results

Table 2. Calibration data for quartz, kaolinite and coal dust.

Model	Analyte	Number of principal components	Percentage of explained variance	Estimated prediction error (μg)
Principal component regression	Quartz	5	99.5%	12.6
	Kaolinite	5	98.4%	17.6
	Coal dust	10	97.1%	36.5
Partial least squares regression	Quartz	5	99.6%	11.6
	Kaolinite	1	96.8%	24.4
	Coal Dust	5	94.8%	50.5

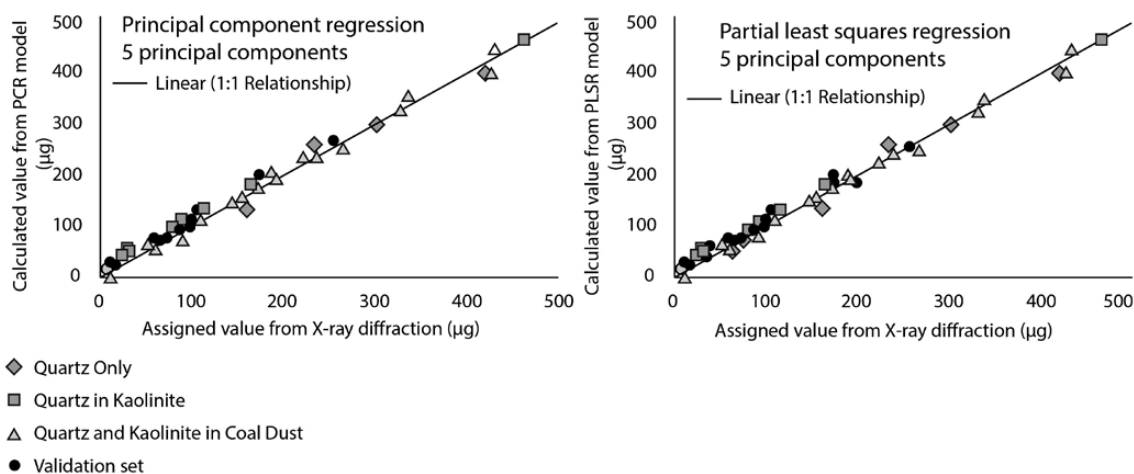


Figure 3. Calibration and validation calculated values for quartz.

from two Beers law models; one following MDHS 101 with blank filter subtraction and the other using a first-order derivative background treatment (Table S4, Supplementary Information). Most differences when comparing these background treatment approaches were relatively small (-3% to -7%). Other studies (Weakley *et al.*, 2014, Miller *et al.*, 2016) removed the spectral bands from the filter by subtracting the spectrum recorded from a clean unused filter. However, such blank subtraction proved unnecessary and risks the occurrence of spectral artefacts remaining; which may hinder the determination of trace levels of other components. The derivatisation removes the need for loaded samples to be matched with blank filters before analysis simplifying the analytical process and may also further reduce the cost of analysis. Blank matching can be the most laborious part of the FTIR analytical process specified in MDHS 101.

Comparison with the MDHS 101 method

A comparison of the predicted quartz concentrations obtained from the samples in the PCR calibration set containing coal dust and kaolinite was made with those values obtained following MDHS 101 method (Figure 5). MDHS 101 uses a calibration based on the Beer-Lambert law i.e. the absorbance response is directly proportional to the concentrations of quartz present in the calibration standards of pure quartz collected onto filters.

To calculate the corrected peak height absorbance, a baseline correction under the quartz absorbance at 780 cm^{-1} and 800 cm^{-1} was applied. This followed the current guidance in the MDHS to 'draw baselines on the blank corrected spectra tangential to the absorbance

minima at about 730 cm^{-1} and 830 cm^{-1} ' and did not have any special regard to the presence of kaolinite. The trend line intercept for the comparison of XRD measured quartz values with the infrared values, determined following MDHS 101 method, was not significantly different from zero (95% level of confidence was from -10 to $+33$). The trend with results from the Beer-Lambert law calibration was linear with a significant positive bias (about $+18\%$ at the current workplace exposure limit in Great Britain) and with a slightly poorer coefficient of determination (0.97). The range of bias was from 0% to 23% and exactly matches that found by (Lee *et al.*, 2013) when using peak height measurement with the National Institute for Occupational Safety and Health (NIOSH) infrared method 7603 (NIOSH, 2004), which is similar to the regulatory method P7 for the Mines Safety and Health Administration (MSHA) (MSHA, 1994). An RCS concentration value that is 18% higher than the 'true' exposure concentration might result in unnecessary further corrective action to control exposures. A plot of the percentage difference between the measured XRD values and the results obtained using FTIR and following MDHS 101 with the varying proportions of quartz and kaolinite showed a negatively sloped relationship (Figure S6, Supplementary Information). The larger proportions of kaolinite (with ratios close to 1:1 with quartz) were responsible for the largest relative errors. The trend line relationship between the percentage differences from the XRD assigned values and the ratio of quartz and kaolinite was $-11.3 \times (\text{ratio quartz/kaolinite}) + 47.8$ with a coefficient of determination of 0.41. The slope of the trend line was significantly different from zero indicating that the relationship is real (the 95% level of confidence of the slope was from -4.1 to -18). No significant relationship

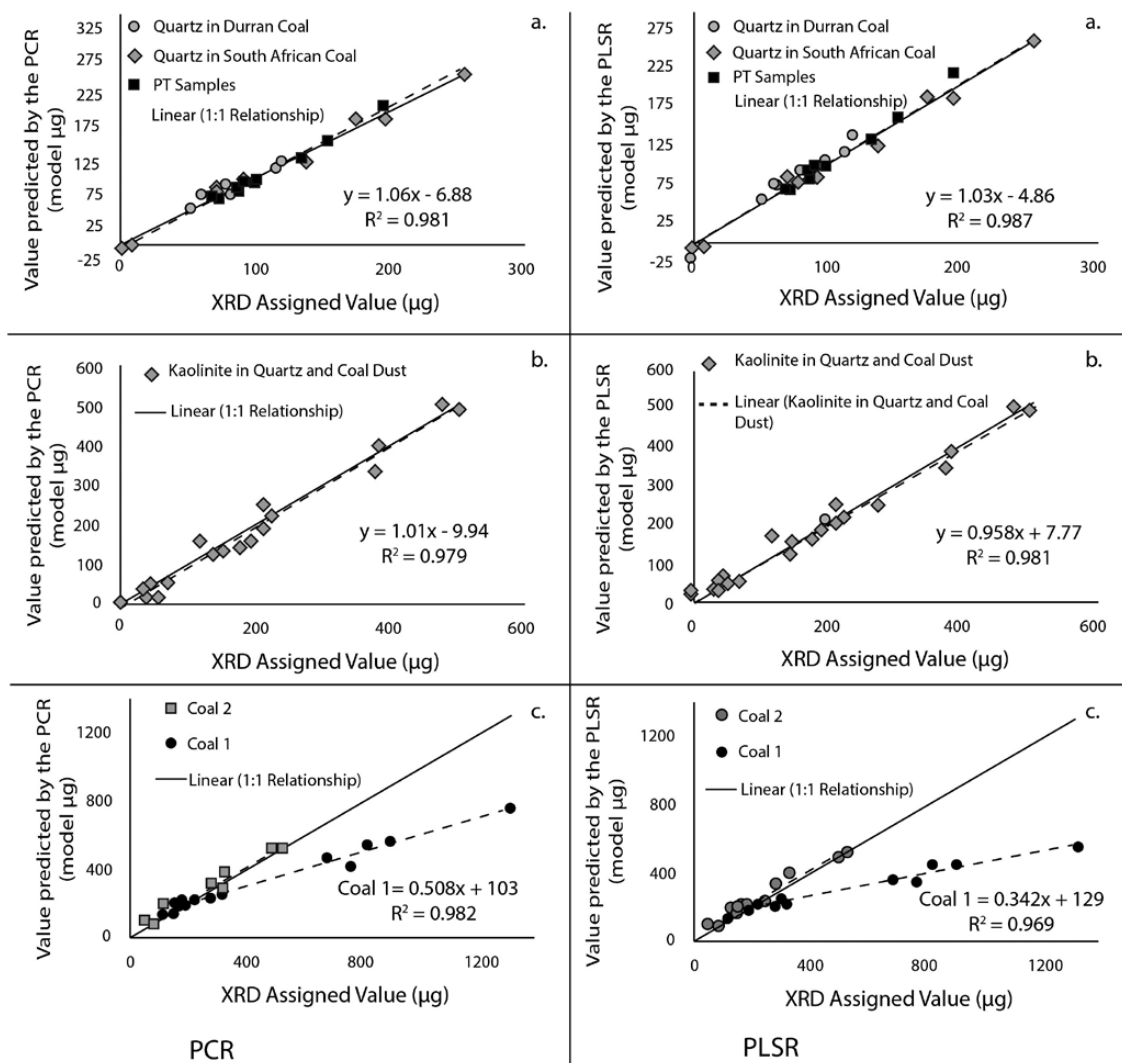


Figure 4. Predicted FTIR values for quartz (a), kaolinite (b) and coal dust (c) from the PCR (left hand side charts) and PLSR (right hand side charts) analysis chemometric models compared with assigned values measured by XRD.

between the percentage differences was found for varying proportions of coal dust in the calibration samples as the confidence interval for the slope included zero (the 95% level of confidence was from -28 to $+59$).

The comparison between PCR and PLSR

The two chemometric methods (PCR and PLSR) produced similar predicted values when compared with the XRD values. Figure 4 compares the predicted values of twenty-eight samples from the PCR and PLSR with the XRD assigned values. All results show good correlations with the assigned values except for coal dust. The coefficients of determination for quartz, kaolinite, South

African coal dust (Coal 2) and Durran's coal dust (Coal 1) were 0.99, 0.98, 0.96 and 0.95 for the PCR method and 0.98, 0.98, 0.96 and 0.97 for the PLSR method, respectively, demonstrating that they are statistically related with mass of analyte. In this work, the two coal dust types appeared to have a different relationship to each other (Figure 4) when the loading is greater than $380 \mu\text{g}$ with high coefficients of determination, i.e. the samples from the different types (ranks) of coal dust are potentially predictable and a correction factor could be applied. For example, the average PCR difference from the target value for coal dust 1 was $169 \mu\text{g}$ whereas it was $26 \mu\text{g}$ for coal dust 2. Different types (ranks) of

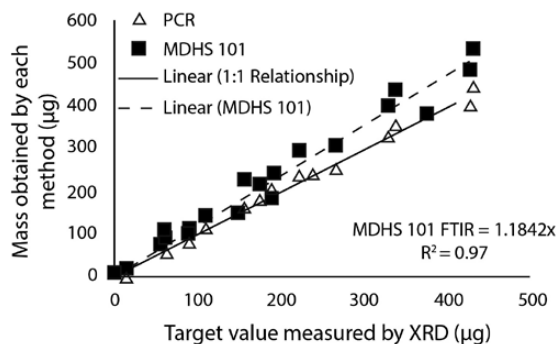


Figure 5. Comparison between predicted PCR results for quartz and those obtained using FTIR following MDHS 101 (Beer–Lambert relationship) and the background subtraction correction for the absorption from the PVC filter.

coal do have different organic contents (He *et al.*, 2017) which may explain the differences observed by the chemometric modelling at higher mass loadings. He *et al.* (2017) characterised thirteen types of coal from lignite to the more crystalline anthracite and found that they have very similar absorption bands that vary in intensity. The main IR active region for coal, in the spectral range used in this study, is between 700 cm^{-1} and 900 cm^{-1} . These IR active bands for coal are associated with aromatic C–H (several bands within 670 cm^{-1} to 900 cm^{-1}) out of plane bending (Coates, 2006, He *et al.*, 2017) and coincide with those for quartz (800 cm^{-1} and 780 cm^{-1}) and the hydroxyl (Al–OH) stretching vibrations of kaolinite at 915 cm^{-1} (Madejová and Komadel, 2001, Lee *et al.*, 2013). However, the only perceptible spectral differences in the FTIR scan of a clean PVC filter and one loaded with Durrans coal (Coal 1) is a broad band at about 760 cm^{-1} and an increase in absorbance from 990 cm^{-1} to 1000 cm^{-1} , which is probably associated with hydroxide group bending from the additional kaolinite found in the South African coal (Coal 2). These differences indicate that components other than kaolinite and quartz within the coal are potentially distinctive to the type of coal. The hydrocarbon content might also have an influence on the observed spectrum and a chemometric FTIR method may provide an opportunity to more fully characterise a worker’s personal exposure. To achieve this, the scan range may need an extension above 1000 cm^{-1} (He *et al.*, 2017). It is possible that the spectra of different kaolinites might also vary, however, Lee *et al.* (2013) demonstrated that most kaolinite will have a similar IR response for the 915 cm^{-1} band unless they are chemically treated (acid washed) or a slight variant of kaolinite (halloysite, which contains more hydroxyl groups). Therefore, the coal dust models described by this study could be corrected for

different coal dusts because the two other quantifiable components have a relatively consistent absorbance per unit mass.

The masses of quartz and kaolinite are predicted by the models with reasonable levels of precision. The following standard deviations include the difference from the assigned target value. For the PCR model, the standard deviation of percentage differences between the PCR predicted value and the assigned value by XRD was 7.9%, 14.1% and 38.5% for quartz, kaolinite and coal dust, respectively. Eight samples from two rounds of the AIR proficiency testing (PT) programme were added for prediction as well as two samples containing pure quartz prepared using a SIMPEDS sampler. The pure quartz samples, including the two sets of four from the AIR PT, obtained a percentage standard deviation of 4.4% from the target values (range of results was 0.7–11%). The standard deviations of percentage differences were 19.6% for coal dust from South Africa and 36.8% for the Durrans coal dust, i.e. there was more variability for the FTIR results when using PCR about the expected target values, for coal dust and Durrans coal when compared to quartz and kaolinite.

For the PLSR model, the standard deviations of percentage differences were, 8.0%, 13.8% and 35.4% for quartz, kaolinite and coal dust, respectively. The South African coal dust was 23% and the Durrans coal dust was 27%. The standard deviation of percent differences for the same AIR PT samples and quartz standards was 7.0% (range of percent differences was from 0.6% to 10.5%).

The agreement between quartz standards prepared using the SIMPEDS and the AIR proficiency testing (PT) samples is good. The AIR PT samples are prepared using the British Cast Iron Research Association (BCIRA) respirable sampler, which is different from the SIMPEDS sampler used in this work and has a slightly different sampling bias. The deposit of dust on the aerosol filter can alter the sensitivity of the measurement, however, the close agreement was observed for results from these samplers in an international comparison (Stacey *et al.* 2014).

A direct comparison of regression coefficients for the PCR and PLSR models, where x are the values predicted by PLSR and y are the predicted values from PCR when processing the same thirty samples, is shown in Table 3 and plots are provided in Figure S7 (supplementary information).

Table 3, Figure 4 and Figure S7 show that there is generally good relationship with a linear model between results generated by the PCR and PLSR models, except for coal dust. There are some slight differences in the predictions of the two methods. Firstly, the ideal value of 1.00 demonstrating perfect correlation is significantly different for kaolinite and coal dust at the 95% level of confidence.

Table 3. Regression coefficients obtained from a comparison of PCR and PLSR results.

Analyte	Regression coefficient (r^2)	Slope	95% level of confidence for the slope coefficient	Intercept (μg)	95% level of confidence for the intercept coefficient
Quartz	0.997	0.98	0.94–1.01	1.39	–1.5 to + 5.5
Kaolinite	0.997	1.04	1.01–1.07	–15.6	–9.6 to –22
Coal dust	0.969	1.23	1.11–1.37	–52.9	–97 to –17

Coal dust has a bias of about +23% indicating that the PCR is reporting higher values than the target values (see Figure 4). The difference for the intercept and slope coefficients is probably due to the number of PCs applied by each model (e.g. a low number (1 principle component) was selected by the software for the PLSR and a higher number (5) was selected for the PCR model used for kaolinite). However, for each analyte, 92.5% of differences between individual values were within the 95% level of agreement (Bland and Altman, 2003) indicating that there was no practical difference in results reported between models.

A potential flaw, when generating samples onto a filter from an aerosol, is that it is difficult to control the collection of the exact proportions of the minerals of interest (quartz and kaolinite). The proportions for quartz in the calibration and validation sets (generally more than 18%) tended to match the top end of the range (2–30%) found by Miller *et al.* (2016) when measuring samples from six coal mines in the United States; which may compromise the model's predictions for samples with low proportions of quartz.

Conclusions

The results reported herein demonstrate the need for matrix-specific calibration standards when measuring quartz from workplace atmospheres with FTIR where other components are present that will influence the spectra close to the absorbance for quartz. Both PLSR and PCR models were found to provide values comparable with XRD when following the XRD part of the MDHS 101 for the measurement of constituents in coal dust with matrix matched standards, which was more accurate than the current FTIR MDHS 101 approach.

The advantage of these chemometric models is that they offer the potential for multicomponent analysis to provide more comprehensive knowledge of a workers' exposure to the workplace atmosphere and not only a single individual component in the matrix. They also have the potential to provide a black box capability that requires less knowledge or experience to interpret spectra. The next stage of this work will be to examine the universal applicability of these models for predicting quartz in a range of workplace environments.

The use of the first derivative to normalise spectra before processing removes the need for the blank filter subtractions for each spectrum. This will significantly reduce the complexity of a direct on-filter analysis approach.

There are only slight differences between results obtained by the PCR and PLSR methods. Selection of the number of PCs has an important effect on the slope of the relationship between the 'true' value and the value predicted by the chemometric models applied to the FTIR data. It was useful to have the capability to further optimise the number of PCs for the PCR approach and to check its impact on the predicted values by testing the significance of coefficients and differences obtained from a regression with the assigned target XRD values. The software selected 6 PCs, for the PCR; however, 5 or 3 PC provided predicted values that were more comparable with the target values. This capability might be a useful consideration when developing models based on PCR using commercially available software.

Supplementary Data

Supplementary data are available at *Annals of Work Exposures and Health* online.

Acknowledgements

Thanks to Ian Pengelly, Laurie Davies, Margaret Wade, Clare McNicholas and Susan Hambling at HSE who conducted the internal HSE technical and editorial reviews and Dr Chisholm (NIOSH) for his assistance when examining the effect of different FTIR instrument resolutions on the signal to noise ratio (Figure S4).

Disclaimer

The work presented in this article was funded by the Health and Safety Executive (HSE) in the United Kingdom (Project PH00760). The opinions expressed in this article are those of the authors and do not necessarily reflect HSE policy.

Data availability

The data underlying this article are available in the article and in its online supplementary material. The raw data will be shared on reasonable request to the corresponding author.

References

- Bergdahl IA, Torén K, Eriksson K, *et al.* (2004) Increased mortality in COPD among construction workers exposed to inorganic dust. *Eur Res J*; **23**: 402.
- Bland JM, Altman DG. (2003) Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol*; **22**: 85–93.
- Breen C, Clegg F, Herron MM, *et al.* (2008) Bulk mineralogical characterisation of oilfield reservoir rocks and sandstones using diffuse reflectance Infrared Fourier Transform Spectroscopy and partial least squares analysis. *J Petroleum Sci Eng*; **60**: 1–17.
- Brown T, Cherrie JW, Van Tongeren M, *et al.* (2012) The burden of occupational cancer in Great Britain Lung cancer. HSE Research Report RR858 Bootle, Liverpool: Health and Safety Executive.
- Bye E. (1994) Chemometrics in Aerosol Analysis: Quantitative Analysis of silica dust mixtures by multivariate calibration applied to Infrared Spectroscopy. *Ann Occupat Hygiene*; **38**: 519–25.
- Cauda E, Miller A, Drake P. (2016) Promoting early exposure monitoring for respirable crystalline silica: Taking the laboratory to the mine site. *J Occupat Environ Hygiene*; **13**: D39–D45.
- CEN. (1993) *EN 481:1993 Workplace atmospheres — Size fraction definitions for measurement of airborne particles*. Brussels: European Committee for Standardisation.
- Coates J. (2006) Interpretation of infrared spectra, a practical approach. In Mckelvy, R. A. M. A. M. L. editor. *Encyclopedia of Analytical Chemistry*. Online Library: John Wiley & Sons.
- Dodgson J, Whitaker W. (1973) The determination of quartz in respirable dust samples by infrared Spectrophotometry—I: The potassium bromide disc method. *Ann Occupat Hygiene*; **16**: 373–87.
- Foster RD, Walker RF. (1984) Quantitative determination of crystalline silica in respirable-size dust samples by infrared spectrophotometry. *Analyst*; **109**: 1117–27.
- Hart JF, Autenrieth DA, Cauda E, *et al.* (2018) A comparison of respirable crystalline silica concentration measurements using a direct-on-filter Fourier transform infrared (FT-IR) transmission method vs. a traditional laboratory X-ray diffraction method. *J Occupat Environ Hygiene*; **15**: 743–54.
- He X, Liu X, Nie B, Song D. (2017) FTIR and Raman spectroscopy characterization of functional groups in various rank coals. *Fuel*; **206**: 555–63.
- Hemmateenejad B, Akhond M, Samari F. (2007) A comparative study between PCR and PLS in simultaneous spectrophotometric determination of diphenylamine, aniline, and phenol: Effect of wavelength selection. *Spectrochim Acta Part A: Mol Biomol Spectros*; **67**: 958–65.
- HSE. (2003) EH75 Respirable Crystalline Silica, variability in fibrogenic potency and exposure-response relationships for silicosis. Health Safety Execut.
- HSE. (2005) *EH40/2005 Workplace exposure limits, containing the list of workplace exposure limits for use with the control of substances hazardous to health regulations (as amended)*. Norwich, United Kingdom, Health and Safety Executive, Crown.
- HSE. (2014) *Methods for the determination of hazardous substances, MDHS 101/2 crystalline silica in respirable airborne dust, direct on-filter analyses by infrared spectroscopy and X-Ray diffraction*. Norwich, United Kingdom, Crown: Health and Safety Executive (HSE).
- Jolliffe IT, Cadima J. (2016) Principal component analysis: a review and recent developments. *Philos Trans R Soc A: Math Phys Eng Sci*; **374**: 20150202.
- Lee T, Chisholm WP, Kashon M, *et al.* (2013) Consideration of Kaolinite interference correction for quartz measurements in Coal Mine Dust. *J Occupat Environ Hygiene*; **10**: 425–34.
- Madejová J, Komadel P. (2001) Baseline studies of the clay minerals society source clays: Infrared methods. *Clays Clay Miner*; **49**: 410–32.
- Miller AL, Weakley T, Griffiths P, *et al.* (2016) Direct-on-Filter α -Quartz Estimation in respirable Coal Mine dust using transmission Fourier transform infrared spectrometry and partial least squares regression. *Appl Spectrosc*; **71**: 1014–24.
- MSHA. (1994) *Infrared determination of quartz in respirable coal mine dust: Method P-7*. Pittsburgh, PA: MSHA: Mine Safety and Health Administration (MSHA).
- NIOSH. (2004) *Quartz in coal mine dust by Infrared (redeposition) Method 7603*. Cincinnati: Department of Health and Human Services, National Institute for Occupational Safety and Health (NIOSH).
- Pickard KJ, Walker RF, West NG. (1985) A comparison of X-Ray diffraction and infra-red Spectrophotometric methods for the analysis of α -Quartz in Airborne Dusts*. *Ann Occupat Hygiene*; **29**: 149–67.
- Ross MH, Murray J. (2004) Occupational respiratory disease in mining. *Occupat Med*; **54**: 304–10.
- Stacey P. (2019) A study to assess the performance of an “X-ray powder diffraction with Rietveld” approach for measuring the crystalline and amorphous components of inhalable dust collected on aerosol sampling filters. *Powder Diffract*; **34**: 251–9.
- Stacey P, Mecchia M, Verpaele S, *et al.* (2014) Differences between samplers for respirable dust and the analysis of quartz—an international study. In Harper M, Lee T, editors. *Silica and associated respirable mineral particles*. West Conshohocken, PA: ASTM International. pp. 73–102.
- Weakley AT, Miller AL, Griffiths PR, *et al.* (2014) Quantifying silica in filter-deposited mine dusts using infrared spectra and partial least squares regression. *Anal Bioanal Chem*; **406**: 4715–24.
- Wentzell PD, Vega Montoto L. (2003) Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemomet Intell Laboratory Syst*; **65**: 257–79.