Analyst

A Novel Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) Methodology for Application in Hyperspectral Raman Imaging Analysis

SCHOLARONE™
Manuscripts

A Novel Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) Methodology for Application in Hyperspectral Raman Imaging Analysis

Joseph P. Smith*[1], Erin C. Holahan[2], Frank C. Smith[3], Veronica Marrero[2], and Karl S. Booksh[2]*

[1]*Analytical Research & Development, Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065*

[2]*Department of Chemistry & Biochemistry, University of Delaware, Newark DE 19716*

[3]*Department of Geological Sciences, University of Delaware, Newark DE 19716*

   *Corresponding Authors:

   Joseph P. Smith. Tel: (732) 594-0194. Email: joseph.smith@merck.com

   Karl S. Booksh. Tel: (302) 831-2561. Email: kbooksh@udel.edu

ABSTRACT

   Multivariate curve resolution-alternating least squares (MCR-ALS) applied to hyperspectral Raman imaging is extensively used to spatially and spectrally resolve the individual, pure chemical species within complex, heterogeneous samples. A critical aspect of performing MCR-ALS with hyperspectral Raman imaging is the selection of the number of chemical components within the experimental data. Several methods have previously been proposed to determine the number of chemical components, but it remains a challenging task that if done incorrectly, can lead to the loss of chemical information. In this work, we show that the choice of 'optimal' number of factors in the MCR-ALS model may vary depending on the relative contribution of the targeted species to the overall spectral intensity. In a data set consisting of 27 hyperspectral Raman images of TiO_2 polymorphs, it was observed that the more dominant species were best resolved with a parsimonious model. However, species with intensities near the noise level often needed more factors to be resolved than was predicted by standard methods. Based on the observations in this data set, we propose a new method that employs approximate reference spectra for determining optimal model complexity for identifying minor constituents with MCR-ALS.

KEYWORDS

Chemometric; Multivariate curve resolution; MCR; Raman; Spectroscopy; Chemical imaging; Titanium dioxide; TiO_2-II; Polymorph

1. INTRODUCTION

Hyperspectral Raman imaging and multivariate analysis have been extensively utilized in pharmaceuticals[1-3], food safety and quality[4], biological analysis[5], environmental and national security[6], geological and planetary science[7, 8], and other applications. Hyperspectral Raman imaging is often employed to discern spatially-resolved chemical information about heterogeneous samples. By collecting hundreds to thousands of independent spectra at defined spatial locations (or pixels), hyperspectral images can produce the distribution of the chemical species within a sample. Furthermore, hyperspectral Raman imaging can characterize complex materials, down to diffraction limited spatial resolution, with no sample preparation, high sensitivity, high specificity, extensive sampling versatility, and low sensitivity to water.[9-11] Overall, hyperspectral Raman imaging ultimately can enable the visualization of the chemical, molecular, and structural properties of heterogeneous samples.

The analysis of experimental hyperspectral Raman images remains a significant challenge, especially with respect to the effective extraction of all useful chemical information concerning a sample.[7, 8, 12, 13] Hyperspectral Raman imaging typically generates three-dimensional experimental data containing spatial (x and y) and spectral (λ or v) information.[9] Univariate analysis is the most commonly employed strategy for visualizing this experimental data.[7, 8, 12-14] In univariate analysis, chemical maps from hyperspectral images can be generated based on the integration or intensity of a Raman band or a series of Raman bands characteristic of a known analyte of interest.[7, 8, 12-14] However, univariate analysis may provide a limited representation of the hyperspectral Raman imaging data, especially when interference effects from spectral overlap, background interference, fluorescence, laser power fluctuation, loss of focus, sample roughness, or sample opacity are encountered.[7, 8, 12-14] In these cases, potentially valuable chemical information within the hyperspectral Raman imaging data can go unused or be misrepresented when univariate analysis is utilized.

Multivariate analysis methods, such as multivariate curve resolution-alternating least squares (MCR-ALS), however, can better overcome these interference effects and provide significant advantages relative to univariate analyses. These advantages include increased selectivity, by better parsing chemical information form interference effects, and increased signal-to-noise ratios, which lead to better visualization of chemical distribution.[7, 8, 12-14] MCR-ALS provides a bilinear decomposition of mixed experimental data into estimates of the chemically meaningful profiles of the respective chemical species.[4, 15-17] MCR-ALS applied to hyperspectral Raman imaging data can generate spatially-resolved chemical images and corresponding resolved Raman spectra of the pure, individual chemical species within a complex mixture.

Successful application of MCR-ALS hinges, in part, on the determination of the number of factors to include in the MCR-ALS model.[4, 15-18] Selection of the number of chemical components is a challenging task, especially when a number of the chemical species to be detected are near the detection limit or there is a large, non-bilinear baseline present. Strategies to determine the most appropriate number of factors in the MCR-ALS model frequently require the building of many MCR-ALS models while exploring proper quality-of-fit and interpretability of resolved chemical information.[16] The number of chemical components can be estimated on the basis of prior knowledge of the system or by using a variety of multivariate analysis methods[7, 8, 12, 15, 16, 19, 20], including singular-value decomposition[16], principal component analysis[13], and cluster-aided methods[18]. If the number of components is incorrectly or inconsistently determined, chemical

information can be misrepresented or go unnoticed and unused, due to under- or over-estimations of the resulting MCR-ALS models.[4, 15-18]

Improvement of MCR-ALS models has been demonstrating using spectral information of reference materials. Multiset analysis[5, 15, 21, 22], which employs augmented data matrices that can contain spectra of reference materials, have been utilized with MCR-ALS for the spatial and spectral resolution of complex experimental data sets. Other iterative methods, such as iterative target transform factor analysis (ITTFA)[23-25], have been utilized with reference materials to unmix concentration and spectral profiles under targeted constraints. Using these approaches, however, requires a true reference spectrum for each known analyte of interest. Moreover, this can cause significant challenges as a true reference spectrum should be collected using the same instrumental, experimental parameters, and sample configurations, and typically, multiple reference materials are available for one chemical species of interest. In this work, a novel approach is offered for MCR-ALS by utilizing library-based target Raman spectra of reference materials. This proposed methodology does not require true reference spectra, allows for vast analysis of targeted reference materials that do not have to be known *a priori*, and promotes the use of spectral libraries and databases for evaluation of potential chemical species present in unknown, heterogenous samples and materials.

In this work, we propose a novel methodology for determining the appropriate number of latent factors with MCR-ALS. This methodology involves initially building MCR-ALS models with low numbers of chemical components and systematically increasing the number of components until both the quantitative 'goodness of fit' and the correlation of the MCR-ALS-generated resolved Raman spectra with the potential target estimates of the Raman spectra collected from reference materials is satisfactory. Using this approach, the chemically-relevant components can be determined for MCR-ALS in a facile, convenient process that is driven by the experimental data and is verified by correlation with reference materials. It is significant that these target estimates need not be identical to the 'true' intrinsic Raman profiles in the hyperspectral image. Target spectra libraries collected on different instrumentation or under different experimental conditions can be employed.

With application of this method we are observing two surprising and powerful results. First, we observe that the most 'appropriate' number of factors in a MCR-ALS model depends on the overall signal strength of the target analyte. More predominant species tend to be best resolved with fewer factors; species with overall signal strengths near that of non-bilinear baseline fluctuations tend to require more components in the model to be well resolved. Second, we are requiring more factors than is justified by goodness of fit statistics to be able to identify the spectroscopic signature of species of these minor components present at levels of the non-bilinear baseline fluctuations.

In order to evaluate our methodology, we tested 27 hyperspectral Raman imaging data sets that are directly available in Smith et al.[8]. These data sets were generated by performing hyperspectral Raman imaging on 11 natural rutile (TiO_2) grains (~100 μm in size) that were recovered from four Neoarchean spherule layers deposited between ~2.65 and ~2.54 billion years ago.[7, 8, 26] Previous work[7, 8, 26] documented rutile, ± TiO_2-II, ± anatase (TiO_2), ± quartz (SiO_2), ± substrate-adhesive epoxy in these grains. The presence of TiO_2-II, a high-pressure, α-PbO_2-structured polymorph of titanium dioxide (TiO_2)[27], is geologically important, as it provides physical evidence to further support an impact origin of these four spherule layers.[26] Therefore, these grains provide a textbook example to investigate the spatial and spectral resolution of

complex, heterogeneous mixtures using hyperspectral Raman imaging with MCR-ALS. To our knowledge, this is the *first* report of MCR-ALS with hyperspectral Raman imaging being accomplished using the quantitative correlation of the resolved Raman spectra generated from MCR-ALS with the target Raman spectra of reference materials during the systematic increase in the number of components for MCR-ALS. Using this methodology, MCR-ALS can be more widely utilized by the scientific community for the analysis of hyperspectral Raman imaging data in a data-driven and quantitative fashion.

2. SAMPLES AND METHODS

2.1 SAMPLES AND SAMPLE PREPARATION. For this study, grains recovered from carbonate-rich hand samples that were collected from outcrops of the Carawine spherule layer (CSL), Jeerinah spherule layer (JSL), Monteville spherule layer (MSL), and Bee Gorge spherule layer (BGSL) were used. Detailed geologic descriptions of the spherule layers are given in Glass and Simonson[28]. Hand samples X38-2 (CSL), X21-1 (JSL), V111 (MSL), 96714A (BGSL), and 96714B (BGSL) were used[26], the original weights of which range from ~177 g (96714B) to ~448 g (X21-1). The locations, lithologies, and detailed preparation steps for the hand samples are given in Smith et al.[26]. In brief, the original hand samples were trimmed and sawn into smaller pieces that were rinsed with water and subsequently dried at ~60 °C for 12 hours. For each sample, subsamples, composed of one or more pieces, were weighed prior to acid digestion. The weights of the subsamples used for this study are 144.3 g (X38-2), 34.2206 g (X21-1), 219.9 g (V111), 8.7879 g (96714A), and 14.2225 g (96714B). The subsamples underwent acid digestion at ~60 °C using hydrochloric acid (Fisher Scientific, 12.1 N). The acid-insoluble residues were wet sieved, in conjunction with ultrasonic agitation, typically into five size fractions: <38 μm, 38-63 μm, 63-125 μm, 125-250 μm, and >250 μm. The 63-125 μm size fractions went through heavy liquid (ρ = 2.96 g/cm^3) separation using 1,1,2,2-tetrabromoethane (Fisher Scientific, 99%), and the heavy mineral (ρ > 2.96 g/cm^3) separates were collected on filter paper (Whatman, 11 μm pore size) then rinsed with acetone (Fisher Scientific, 99%). The heavy mineral separates were transferred to glass cavity slides.

The identification of the grains in the separates was done using Raman microspectroscopy.[26] Grain mounts were prepared for selected rutile and TiO$_2$-II-bearing grains, and the details of the grain mount preparation are given in Smith et al.[7]. In brief, grains were fixed to a circular glass substrate (Buehler; 25.4 mm diameter x 1.0 mm thick) using epoxy (Buehler), and the grain mounts were polished using paper strips (600 μm to 3 μm coarseness) with a water/detergent/glycerol solution (~70/15/15 v/v%) to expose cross-sections of the grains. Nine grains are exposed, but grains z1-5 and z4-2 (**Table 1**) are not exposed, in which most portions of these two grains are covered by substrate-adhesive epoxy. A final polish of a grain mount was performed using Al$_2$O$_3$ powder (Excel Metallurgical, 0.05 μm-sized, 99%) wetted with water. Most of the final samples analyzed are therefore cross-sectional representations of grains that are 63-125 μm in size. Grains from samples X38-2 (CSL), X21-1 (JSL), and V111 (MSL) are mounted on grain mounts Z1, Z2, and Z3, respectively, and grains from samples 96714A (BGSL) and 96714B (BGSL) are mounted on grain mount Z4. In grain mount Z4, grain z4-1 is from sample 96714A and grain z4-2 is from sample 96714B.

2.2 HYPERSPECTRAL RAMAN IMAGING. All hyperspectral Raman imaging data sets were collected using a Senterra Raman spectrometer (Bruker Optics) coupled to a BX-51

optical microscope (Olympus). Details of the methodology, instrumentation, and procedures for performing hyperspectral Raman imaging on these grains are outlined in Smith et al.[7, 8] In brief, a Nd:YAG laser frequency-doubled to 532 nm was used as the excitation source. A 20× (0.40 numerical aperture, NA), 50× (0.75 NA), or 100× (0.80 NA) objective lens (Olympus) was used to focus the laser onto the sample's surface, yielding a spot size of ~5 μm, ~2 μm, and ~1 μm in diameter, respectively. Laser exposure times of 1 to 15 seconds with 0 to 3 co-additions were utilized. The laser power was held constant and ranged from 2 to 10 mW. Scattered light was collected by the objective lens, filtered by a 50×1000 μm slit aperture, and dispersed by a 1200 grooves/mm grating onto a charge-coupled device (CCD) detector (Bruker Optics) thermoelectrically-cooled to a temperature of −65 ˚C. The spectral resolution was 3 to 5 cm^{-1}. The probed spectral ranges were either 70 to 1550 cm^{-1} or 70 to 2650 cm^{-1} and were covered under one or two grating positions, respectively. Background measurements, Raman shift calibration, and source wavelength calibration were performed prior to each measurement. Spectral grids using dimensions of 10×10 (100 total Raman spectra), 15×15 (225 total Raman spectra), 20×20 (400 total Raman spectra), or 30×30 (900 total Raman spectra) were employed. The spectral grids were rectangular in shape based on the dimensionality of the grain. Raman spectra were collected at defined *x* and *y* locations about the grain at a constant height. The OPUS 7.2 program (Bruker Optics) was used for all hyperspectral Raman imaging measurements.

2.3 MULTIVARIATE CURVE RESOLUTION-ALTERNATING LEAST SQUARES (MCR-ALS). MCR-ALS is a self-modeling mixture analysis method. By applying MCR-ALS to hyperspectral Raman imaging data sets, spatially-resolved chemical images and corresponding resolved Raman spectra of the individual, pure chemical components within complex, heterogeneous samples of unknown composition can be produced. Specifically, MCR-ALS decomposes an experimental data matrix, **D**, as follows:

$$D = CS^T + E \qquad (1)$$

where *C* is the concentration profile matrix, S^T is the resolved spectral matrix, and *E* is the residual error matrix. The experimental data generated from hyperspectral Raman imaging is three-dimensional and includes spatial (*x* and *y*) and spectral (λ or v) information. Prior to MCR-ALS, this three-dimensional experimental data is unfolded into the two-dimensional experimental data matrix, **D**, containing combined spatial (both *x* and *y* together) and spectral (λ or v) information.

A variety of multivariate methods[7, 8, 12, 13, 15, 16, 18-20] have been employed to estimate the appropriate number of latent factors for MCR-ALS and to provide an initial estimation of *C* and S^T. The subsequent optimization of *C* and S^T is performed using ALS until convergence is achieved. Using chemically and physically meaningful constraints, *C* and S^T can be readily guided to the respective solution during ALS. Possible constraints include non-negative concentration and non-negative spectral intensities which are especially valid for hyperspectral Raman imaging because neither the concentration nor the spectral intensity of a chemical component should be negative at any pixel. Correlation of resolved spectral estimates in S^T to target reference spectra are employed to help ascertain which components are related to chemical species and which components model the non-bilinear baseline fluctuations.

In this work, 27 hyperspectral Raman imaging data sets were analyzed using MCR-ALS. Each hyperspectral Raman imaging data set was first unfolded into the two-dimensional experimental data matrix, **D**, to allow for the application of the bilinear MCR-ALS model. In **D**, each row is the intensity at various wavenumbers. The most appropriate number of chemical

components for MCR-ALS was selected by quantitatively comparing and correlating the resolved Raman spectra, in $\boldsymbol{S^T}$, with target Raman spectra of reference materials. An initial estimation of $\boldsymbol{S^T}$ was determined using distance criteria. As such, the first Raman spectrum estimate for initialization was selected to be furthest from the mean, and subsequent Raman spectra were selected to be furthest from the mean and all prior selected Raman spectra. During ALS, constraints of non-negative concentration and non-negative spectral intensity were both employed. A convergence of 0.01% was achieved for all MCR-ALS models. MCR-ALS methods were utilized in Matlab 7.12 (MathWorks) with the PLS Toolbox (Eigenvector Research) to generate the concentration profile matrix, $\boldsymbol{C}$, and the resolved spectral matrix, $\boldsymbol{S^T}$. MCR-ALS was performed using the full spectral range of the given hyperspectral Raman imaging data set. No preprocessing was performed prior to application of MCR-ALS.

2.4 DETERMINATION AND SELECTION OF THE NUMBER OF CHEMICAL COMPONENTS FOR MCR-ALS. A total of 16 target Raman spectra of reference materials were considered. Ten Raman spectra were obtained directly from the RRUFF database[29]—two for rutile, two for anatase, one for quartz, one for aragonite ($CaCO_3$), one for calcite ($CaCO_3$), one for dolomite ($CaMg(CO_3)_2$), one for coesite (SiO_2), and one for stishovite (SiO_2). Six Raman spectra were collected in-house—five for TiO_2-II at different laser powers and one for substrate-adhesive epoxy. Raman spectra of reference materials were collected using the same methodology, instrumentation, and procedures as were used for the generation of hyperspectral Raman imaging data (**Section 2.2**). For the chemical species previously identified[7, 8, 26] in these grains, a total of 11 target Raman spectra of reference materials were used—two for rutile, five for TiO_2-II, two for anatase, one for quartz, and one for substrate-adhesive epoxy. For each of the five chemical species—aragonite, calcite, dolomite, coesite, and stishovite—known not to be present within the grains, a target Raman spectrum of a reference material was used to act as a negative control. For all MCR-ALS models, correlation coefficients were generated from the comparison of the MCR-ALS-based resolved Raman spectra with these 16 total target Raman spectra of reference materials.

For each of the 27 hyperspectral Raman imaging data sets, MCR-ALS models were built using 1–20 latent factors. Therefore, 20 MCR-ALS models were generated for each hyperspectral Raman imaging data set, and in total, 540 MCR-ALS models were built and analyzed. For a given hyperspectral Raman imaging data set, the final selection of the best MCR-ALS model was chosen by first identifying the set of models that resolved the most target chemical species and then selecting, from that set, the single model that is both stable and has high correlations between the resolved Raman spectra generated from MCR-ALS and the target Raman spectra of reference materials for these identified chemical species. Stability of an MCR-ALS model is achieved when the correlation coefficients between the resolved Raman spectra generated from MCR-ALS and the target Raman spectra of reference materials remain statistically constant during the increase in the number of components. If, for example, we compare an MCR-ALS model built using three latent components that resolves only rutile with a correlation coefficient of 0.99 with an MCR-ALS model built using six latent components that resolves rutile and TiO_2-II with correlation coefficients of 0.85 and 0.86, respectively, the latter model is selected as the better model because it has identified all the chemical species present (rutile and TiO_2-II), even though the three component model has a higher correlation coefficient for rutile. However, it is noteworthy that the best correlation between the resolved and reference rutile spectra occurred at 2, not 6, intrinsic factors.

2.5 GENERATION OF CORRELATION COEFFICIENTS FROM THE COMPARISON OF MCR-ALS-BASED RESOLVED RAMAN SPECTRA WITH THE TARGET RAMAN SPECTRA OF REFERENCE MATERIALS. Correlation of the resolved Raman spectra generated from MCR-ALS with the target Raman spectra of reference materials was conducted in RStudio AGPL v3. The correlation results were generated as correlation coefficients (R^2). The maximum correlation coefficient is determined and plotted for each chemical species within the given MCR-ALS model. The spectral range of the target Raman spectra of reference materials, directly obtained from the RRUFF database[29], was adjusted to reflect the 70 to 1550 cm^{-1} spectral range, with increments of 0.5 cm^{-1}, that was utilized for the hyperspectral Raman imaging data sets. Offsets in the RUFF database and collected Raman spectra wavenumbers were rectified by interpolation of a moving 6-point, 2nd-order polynomial fit to the RUFF spectra prior to generating the correlation coefficients.

3. RESULTS AND DISCUSSION

3.1 RAMAN SPECTRA, RAMAN OPTICAL MODES, AND CRYSTALLOGRAPHY OF THE CHEMICAL SPECIES WITHIN THE GRAINS. Previous studies[7, 8, 26] documented the presence of three TiO_2 polymorphs (rutile, anatase, and TiO_2-II), quartz (SiO_2), and substrate-adhesive epoxy within the grains investigated in this work. Rutile, anatase, and brookite are the three main polymorphs of natural TiO_2, with rutile being the most abundant polymorph.[30] Rutile has a tetragonal structure with the *P4$_2$/mnm* space group.[31] The primitive unit cell of rutile has 15 optical modes, four of which are Raman active.[31] The Raman spectral bands (cm^{-1}) and corresponding optical mode assignments of rutile are 143 cm^{-1} (B_{1g}), 237 cm^{-1} (high-order Raman band), 443 cm^{-1} (E_g), 611 cm^{-1} (A_{1g}), and 815 cm^{-1} (B_{2g}).[31, 32] Anatase has a tetragonal structure with the *I4$_1$/amd* space group.[33, 34] The primitive unit cell of anatase has 15 optical modes, six of which are Raman active.[33, 34] The Raman spectral bands and corresponding optical mode assignments of anatase are 145 cm^{-1} (E_g), 198 cm^{-1} (E_g), 396 cm^{-1} (B_{1g}), 513 cm^{-1} (B_{1g}), and 637 cm^{-1} (E_g).[33-35]

High-pressure polymorphs of TiO_2 can be produced under pressures and temperatures of 4-12 GPa and 400-1500 °C.[27, 36, 37] Five high-pressure polymorphs of TiO_2 are presently known, and one of these polymorphs is termed TiO2-II.[38, 39] TiO_2-II is an important high-pressure polymorph due to its stability at ambient conditions.[40] TiO_2-II is isostructural with the orthorhombic α-PbO_2 phase and has space group *Pbcn*.[41] The primitive unit cell of TiO_2-II has 36 optical modes, four of which are Raman active.[41] The transformation of ambient TiO_2 to TiO_2-II is kinetically and thermodynamically limited[40, 42], and evidence of precursory TiO_2 is consistently reported in both natural and synthetic TiO_2-II.[43-47] Therefore, it has been difficult to obtain a Raman spectrum of *pure* TiO_2-II. Smith et al.[7] recently provided an estimated Raman spectrum of *pure* TiO_2-II that is in agreement with the Raman spectrum of synthetic TiO_2-II.[40] The Raman spectral bands and corresponding optical mode assignments of TiO_2-II are 152 cm^{-1} (B_{3g}), 174 cm^{-1} (A_g), 286 cm^{-1} (B_{1g}), 315 cm^{-1} (B_{1g}), 341 cm^{-1} (B_{2g}), 357 cm^{-1} (B_{3g}), 412 cm^{-1} (B_{2g}), 426 cm^{-1} (A_g), 531 cm^{-1} (A_g), and 572 cm^{-1} (B_{1g}).[7, 41]

Crystalline quartz undergoes a reversible solid-state phase transition at approximately 574 °C, in which the low-temperature phase is referred to as α-quartz and the high-temperature phase is referred to as β-quartz.[48] In this work, the Raman active modes and Raman spectra of

ambient quartz, i.e., α-quartz, will be discussed, since this was identified in previous studies[7, 8, 26] of these grains. Quartz is trigonal-structured and has space group $P3_221$.[48] Quartz has 27 optical modes, including $4A_1$ Raman active modes and 8 E_T/E_L Raman active modes.[49] The Raman spectral bands and corresponding optical mode assignments of quartz are 201 cm^{-1} (A_1), 264 cm^{-1} (E_T), 356 cm^{-1} (A_1), 395 cm^{-1} (E_T), 403 cm^{-1} (E_L), and 464 cm^{-1} (A_1).[48] Substrate-adhesive epoxy, the fifth chemical species identified in these grains during previous studies[7, 8], is a commonly used substrate adhesive containing polymeric epoxide and has been extensively studied using Raman spectroscopy.[50, 51] The Raman spectrum of substrate-adhesive epoxy is characterized by five major Raman bands at the following spectral locations: 640 cm^{-1}, 822 cm^{-1}, 1113 cm^{-1}, 1185 cm^{-1}, and 1462 cm^{-1}.[50, 51]

3.2 MULTIVARIATE CURVE RESOLUTION-ALTERNATING LEAST SQUARES WITH HYPERSPECTRAL RAMAN IMAGING. In total, up to five distinct chemical species were resolved within the hyperspectral Raman imaging data sets employing MCR-ALS models with, generally, 8 to 12 factors. These five chemical species include three polymorphs of TiO_2 (rutile, anatase, and TiO_2-II), SiO_2 (quartz), and substrate-adhesive epoxy (**Table 1**). Of the 27 hyperspectral Raman imaging data sets analyzed (**Figures S5-S29**), six data sets were shown to contain these five chemical species. Furthermore, MCR-ALS applied to the hyperspectral Raman imaging data sets revealed that two data sets contain only one chemical species (rutile), one data set contains two chemical species (rutile/substrate-adhesive epoxy), eight data sets contain three chemical species (rutile/TiO_2-II/substrate-adhesive epoxy or rutile/anatase/substrate-adhesive epoxy), and ten data sets contain four chemical species (rutile/anatase/quartz/substrate-adhesive epoxy or rutile/TiO_2-II/anatase/substrate-adhesive epoxy). In this work, we therefore evaluate our MCR-ALS methodology applied to hyperspectral Raman imaging, in which two selected hyperspectral Raman imaging data sets are presented and discussed.

(Insert Table 1 here)

3.3 MCR-ALS APPLIED TO HYPERSPECTRAL RAMAN IMAGING DATA SET RA15: CORRELATION OF MCR-ALS RESOLVED RAMAN SPECTRA WITH TARGET RAMAN SPECTRA OF REFERENCE MATERIALS FOR MODEL SELECTION AND EVALUATION. Analyses of the hyperspectral Raman image collected on Monteville spherule layer grain z3-13 highlight the necessity to better extract trace chemical signatures from highly variable, noisy baselines. Preliminary analysis by principal component analysis (PCA) predicts six components, at most, are needed to model the non-systematic variance of this image. A four principal component (PC) model captures greater than 99% of the cumulative variance and a six PC model captures greater than 99.9% of the variance. Plots of the eigenvalues vs. PC number and root mean squared error of cross validation (RMSECV) vs. PC number similarly pick between four and six components, with more than six components not presenting a statistical improvement in model fit. Were the MCR-ALS analyses limited to six components, the spectra of substrate-adhesive epoxy and quartz would not be extracted and the best estimates of the TiO_2 polymorph spectra would not be resolved.

(Insert Figure 1 here)

Correlation of the resolved Raman spectra to standard spectra from the RRUFF database[29] shows that each of the five chemical components identified in the image is best resolved with a different complexity of MCR-ALS model (**Figure 1**). The chemical species resolved from MCR-

ALS are tracked by correlation (**Figure 1**), not by absolute component number within the MCR-ALS model. Additionally, all models that show highest correlation with the standard spectra have more than the four to six components predicted by PCA analytics. Four of the five components correlate to standard spectra with coefficients, R^2, greater than 0.80. Only the residual substrate-adhesive epoxy spectrum has a poor fit to the standard spectrum employed with a 0.45 correlation coefficient. By way of comparison, no resolved Raman spectrum from any of the 20 different models correlated well to the RRUFF standard Raman spectra of five other target minerals – aragonite, calcite, coesite, dolomite, and stishovite (**Figure 2**). Of the 199 factors produced by the 20 different models, the highest correlation coefficient observed was 0.19 for one factor in the 20 component model with coesite. Consequently, it was determined that these five minerals are not present in the analyzed grain.

(Insert Figure 2 here)

For the five identified chemical components, TiO_2-II is first resolved first with a seven component model and its correlation coefficients are stable with models having between seven and 15 components. Based on correlation alone, it is impossible to state which of the five TiO_2-II standard spectra are most appropriate to characterize the TiO_2-II in the sample. TiO_2-II is known to have a temperature dependent Raman spectrum that broadens with increasing laser power during interrogation.[7] The resolved TiO_2-II Raman spectrum has a slightly higher correlation coefficient with the standard spectra collected at 5 mW and at 10 mW than with the standard spectra collected with the highest (20 mW) or lowest (0.2 mW) laser power. The resolved anatase Raman spectrum is recognizable with a seven factor model (0.78 R^2) but is best resolved with a nine to 15 component model. Quartz is not resolved until 10 or more factors are employed in the MCR-ALS model. Rutile is identifiable in the resolved Raman spectra starting with a six component model (0.8 R^2) and its resolved spectrum rises to a 0.9 correlation coefficient with the chosen standard when an 11 factor model is employed. Analysis of **Figure 1** shows that a 10 component model is the most parsimonious model that resolves all five identified components. The evolution of resolved spectra of the identified components shows that including more factors tends to better remove a spectral baseline from each component (**Figure 3**). This trend is most evident in the removal of the baseline of the rutile spectrum by changing from a four to a 10 factor model (**Figure 3**). The slight baseline in the eight factor anatase spectrum, relative to the 10 factor anatase spectrum, accounts for the slight difference in the respective correlation coefficients.

(Insert Figure 3 here)

Full results for the 10 component MCR-ALS model applied to hyperspectral Raman imaging data set Ra15 are shown in **Figure S3**. The spatial distribution of the resolved chemical species (**Figure 4**) was generated from the application of a 10 component MCR-ALS model to hyperspectral Raman imaging data set Ra15. The optical and backscattered electron (BSE) images of this grain (**Figure S1**) highlight the heterogeneity of this grain. The BSE image (**Figure S1B**) displays polydispersed-sized particles of varying shapes within grain z3-13. Moreover, the optical and BSE images of grain z3-13 (**Figure S1**) can be used to directly compare the physical properties of the grain to the spatially-resolved chemical images generated from MCR-ALS. The spatial distribution of rutile (**Figure 4A**) demonstrates that rutile is a major chemical species within this grain and is primarily located towards the top-middle and bottom-middle portions of the grain. The spatial distribution of TiO_2-II (**Figure 4B**) demonstrates that TiO_2-II is also a predominant chemical species and is primarily located in the top-middle and bottom-middle portions of the grain in a complementary fashion to that of rutile. The spatial distribution of anatase (**Figure 4C**)

demonstrates that anatase is highly localized and is predominantly located towards the top-middle and bottom-right portions of the grain. The spatial distribution of quartz (**Figure 4D**) demonstrates that quartz is located primarily towards the bottom-left portion of the grain in a highly localized manner. The spatial distribution of substrate-adhesive epoxy (**Figure 4E**) demonstrates that substrate-adhesive epoxy is located primarily exterior to the grain.

(Insert Figure 4 here)

The resolved Raman spectra generated from a 10 component MCR-ALS model applied to hyperspectral imaging data set Ra15 (**Figure 4**) are in agreement with the target Raman spectra of reference materials. Specifically, the resolved Raman spectrum of rutile (**Figure 4A**) demonstrates a 0.87 correlation coefficient with that of the RRUFF database reference material, in which the Raman bands characteristic of rutile at 446 cm^{-1} and 610 cm^{-1} are observed. The resolved Raman spectrum of TiO_2-II (**Figure 4B**) demonstrates a 0.96 (5 mW) correlation coefficient with that of the reference material, in which the Raman bands characteristic of TiO_2-II at 151 cm^{-1}, 175 cm^{-1}, 287 cm^{-1}, 315 cm^{-1}, 339 cm^{-1}, 357 cm^{-1}, 428 cm^{-1}, 533 cm^{-1}, and 574 cm^{-1} are observed. The resolved Raman spectrum of anatase (**Figure 4C**) demonstrates a 0.93 correlation coefficient with that of the reference material from the RRUFF database, in which the Raman bands characteristic of anatase at 143 cm^{-1}, 397 cm^{-1}, 514 cm^{-1}, and 637 cm^{-1} are observed. The resolved Raman spectrum of quartz (**Figure 4D**) demonstrates a 0.83 correlation coefficient with that of the reference material from the RRUFF database, in which the Raman bands characteristic of quartz at 127 cm^{-1}, 206 cm^{-1}, and 464 cm^{-1} are observed. The resolved Raman spectrum of substrate-adhesive epoxy (**Figure 4E**) demonstrates a 0.38 correlation coefficient with that of the reference material, in which the Raman bands characteristic of substrate-adhesive epoxy at 640 cm^{-1}, 822 cm^{-1}, 1113 cm^{-1}, 1187 cm^{-1}, and 1462 cm^{-1} are observed. Thus, for the identified chemical species, the resolved Raman spectra generated from MCR-ALS are both in agreement and have high correlations with the target Raman spectra of the respective reference materials.

3.4 MCR-ALS APPLIED TO HYPERSPECTRAL RAMAN IMAGING DATA SET RA23: CORRELATION OF MCR-ALS RESOLVED RAMAN SPECTRA WITH TARGET RAMAN SPECTRA OF REFERENCE MATERIALS FOR MODEL SELECTION AND EVALUATION. MCR-ALS was applied to the hyperspectral Raman imaging data set Ra23, collected on grain z4-1, that was recovered from the Bee Gorge spherule layer[26]. Previous MCR-ALS analyses discerned four identifiable Raman spectra on this grain (**Table 1**)—rutile, TiO_2-II, anatase, and substrate-adhesive epoxy.[7, 8, 26] Preliminary analyses by PCA predicts a three to five component model is appropriate for describing the non-systematic spectral variance in the hyperspectral image. A three PC model describes 99.5% of the variance, while a four PC model describes 99.9% of the spectral variance. Eigenvalue, root mean squared error of calibration (RMSEC), and RMSECV plots estimate either four or five PCs as appropriate.

(Insert Figure 5 here)

However, if a four or five factor MCR-ALS model were employed, then anatase, a minor chemical species in this grain, would not be detected. Quantitative comparison between the resolved Raman spectra and Raman spectra of target standards show that three of the chemical components are well resolved with a four factor MCR-ALS model (**Figure 5**), which exhibits correlation coefficients of 0.70, 0.96 (5 mW), and 0.89 for rutile, TiO_2-II, and substrate-adhesive epoxy, respectively. The substrate-adhesive epoxy is not resolved with a three factor MCR-ALS model. The resolved rutile Raman spectra are not as well correlated to those of the RRUFF rutile

target standards compared to the correlations of the other resolved species. As discussed below, this is likely to be a consequence of the sub-micrometer-scale crystal size of minerals within the grain and the orientation dependence of the Raman spectrum for rutile. The rutile Raman spectrum depends on both the crystal face presented to the laser and the rotation of the crystal relative to the laser polarization. Consequently, there is a bounded continuum of rutile Raman spectra from the grains analyzed in this work that prohibits comparison to a true standard. The seven target rutile standards employed represent seven possibilities (and ideally end members) within this continuum. The fourth identified chemical species, anatase, is first observed using an 11 factor MCR-ALS model. Here the resolved Raman spectrum of anatase has a 0.91 correlation coefficient with that of the RRUFF anatase standards.

(Insert Figure 6 here)

The resolved Raman spectra generated from MCR-ALS applied to hyperspectral Raman imaging data set Ra23 are in good agreement with the target Raman spectra of reference materials expected to be found in this sample (**Figure 5**). Furthermore, these resolved Raman spectra exhibit poor correlation with target Raman spectra of reference materials from the RRUFF database that are not observed in this sample (**Figure 6**). Five minerals (aragonite, calcite, coesite, dolomite, and stishovite) not observed in the grain all exhibit correlation coefficients less than 0.2 for all the MCR-ALS models having one to 20 factors. For quartz, however, the correlation coefficient was less than 0.15 for MCR-ALS models having one through 15 factors, but jumped to approximately 0.25 for 16 through MCR-ALS models having 20 factors. Given that quartz was observed in other samples, these results suggest a trace quartz contribution within the sample. If the presence of quartz was particularly interesting or worthy of study, these results would be an indication to collect spectra at greater signal to noise were the putative quartz contribution dispersed and/or at greater pixel density in a small area were the putative quartz contribution highly localized.

(Insert Figure 7 here)

The optimal MCR-ALS model for hyperspectral Raman imaging data set Ra23 was determined to have 11 factors. This conclusion is based on the high correlation of the resolved Raman spectra generated from MCR-ALS with the target Raman spectra of reference materials (**Figures 5 and 7**) and the stability of the correlation coefficients from factor to factor for this model (**Figure 5**). Full results for the 11 component MCR-ALS model applied to hyperspectral Raman imaging data set Ra23 are shown in **Figure S4**. Chemical images constructed from the 11 component MCR-ALS model (**Figure 7**) show the spatial distribution of the four resolved chemical components. The optical and scanning electron microscopy images (**Figure S2**) of this grain (z4-1; **Table 1**) indicate that it is heterogeneous and comprised of polydispersed particles of varying shapes and orientations. The imaged section of the grain is dominated by rutile and TiO_2-II. Rutile is concentrated in the upper-left portion of the imaged area and TiO_2-II is the primary constituent in the rest of the imaged grain. Anatase is localized in a few areas along the grain perimeter, with the most intense Raman spectra extracted from the lower-left edge. As expected, the substrate-adhesive epoxy is found exterior to, and not within the polished grain, and mineral Raman spectra were not extracted from the region of the substrate-adhesive epoxy.

Selected Raman spectra within hyperspectral Raman imaging data set Ra23 are shown in **Figure S30**. These Raman spectra demonstrate the varying baselines observed throughout Ra23. Moreover, non-bilinear spectral baselines that are much larger in their variance than traditional white noise baselines are observed. To resolve these non-bilinear baselines, additional components

in the MCR-ALS model may be needed, especially when investigating trace or minor chemical species. Performing minimal baseline corrections to ensure no chemical information concerning the minor or trace species is removed can further this effect, in which the model can now account for the nonrandom variance in the baseline by using additional components for MCR-ALS.

3.5 INVESTIGATION OF THE TARGET RAMAN SPECTRA OF REFERENCE MATERIALS FOR MCR-ALS MODEL DETERMINATION. The method described here is applicable when there is no exact reference target in the library, as is the case for TiO_2-II. As discussed above, the TiO_2-II Raman spectrum broadens with increasing laser power. Similarly, the RRUFF database offers four distinct anatase Raman spectra (**Figure 8A**) and seven distinct rutile spectra (**Figure 9A**) collected with a 532nm depolarized laser.[29] Both anatase and rutile occur in the ditetragonal-dipyramidal (4/*mmm*) crystal class. However, anatase belongs to the $I4_1/amd$ space group and rutile belongs to the $P4_2/mnm$ space group. Consequently, these polymorphs exhibit different Raman spectra. Anatase presents an intense peak at 143 cm[-1] and other peaks at 397 cm[-1], 514 cm[-1], and 637 cm[-1].[29] Of the four RRUFF anatase spectra, three have the 143 cm[-1] peak truncated. Rutile has three main Raman bands—235 cm[-1], 446 cm[-1], and 610 cm[-1]. The location of these bands does not shift among the RRUFF Raman spectra; however, their relative intensities vary greatly with the orientation of the crystal relative to the polarization of the laser. The effect of laser scattering relative to crystal orientation is evident in the Raman spectra of the RRUFF rutile standards. With the laser parallel to the (1 0 0) face of rutile, all three bands are visible with the 444 cm[-1] and 608 cm[-1] peaks having intense, comparable heights. When Raman spectra are collected with the excitation laser parallel to the (0 0 1) face, the 608 cm[-1] peak is more intense while the 444 cm[-1] peak is decreased. Additional effects are seen in the polarization of the laser relative to the rutile crystallographic axes. In this work, the depolarized and unoriented RRUFF standards were employed for both anatase and rutile comparisons. However, the set of seven depolarized, unoriented rutile Raman spectra show characteristics of being variously aligned with the (1 0 0) or (0 0 1) crystal faces.

(Insert Figure 8 here)

While the most appropriate anatase standard to employ for spectral validation is not known *a priori*, plots of correlation coefficients vs. number of factors clearly indicate that the R070582 RRUFF standard Raman spectrum has the best correlation to the resolved anatase spectra in both hyperspectral Raman imaging data sets (**Figure 8 B/C**). Both the intense 143 cm[-1] band and the less intense 637 cm[-1] band match for the standard and resolved Raman spectra (**Figures 4C and 7C**). The other three anatase standards have significantly lower correlations due to differences in the relative intensities of the four bands between the RRUFF standards and the resolved anatase Raman spectra. Large correlation coefficients, around 0.9, are realized despite the fact that anatase is a minor component in these grains. Anatase is stable at ambient temperature, but is not the equilibrium polymorph of TiO_2. Between 500 °C and 1000 °C anatase converts to rutile. Given that TiO_2-II forms around 5-12 GPa and 500 °C-1200 °C, the presence of TiO_2-II indicates a formational environment in which anatase would convert to rutile.

(Insert Figure 9 here)

The seven unoriented rutile Raman spectra from the RRUFF database fall into three classes (**Figure 9A**). Four of the spectra (R040049, R050031, R050417, and R060493) exhibit strong spectral features at both 444 cm[-1] and 608 cm[-1], consistent with spectra collected with the laser parallel to the (1 0 0) face. Two RRUFF standards (R060745 and R110109) have a very strong

feature at 608 cm^{-1} and lesser intensity at 444 cm^{-1}, consistent with Raman spectra collected with the laser parallel to the (0 0 1) face. The seventh spectra (R120008) is difficult to classify.

Plots of correlation coefficients for the Raman spectra of seven rutile standards with the resolved Raman spectra versus the number of factors in the MCR-ALS model show the best match with the RRUFF spectra having intense features at both 444 cm^{-1} and 608 cm^{-1} for models generated with less than 15 factors (**Figure 9**). RRUFF standards R040049, R050031, and R060493 fit equivalently to the resolved spectra, while standard R050417 matches slightly better for data set Ra23 and slightly worse for data set Ra15. For data set Ra15, the RRUFF Raman spectra (R060745 and R110109) having a single intense feature at 608 cm^{-1} do not correlate well with the resolved Raman spectra until a 17 factor MCR-ALS model is employed. For data set Ra23, there is a steady increase in the correlation for RRUFF standards R060745 and R110109, with these 'one Raman band' standards having a greater correlation with resolved components than the 'two Raman band' standards when the models have 15 or more factors. One possible explanation for these results is the forced splitting into degenerate factors when many factors are employed. However, were this the case, the resolved spectral profiles would be unlikely to have a high correlation with the rutile standards, a situation in disagreement with the result that the two standards have consistently high correlations for models employing 16 through 20 factors.

An alternative explanation is that MCR-ALS is able to resolve spectral profiles of two different rutile orientations – one with the (1 0 0) plane parallel to the laser and the other with the (0 0 1) plane parallel to the laser. Scanning electron microscope (SEM) analyses show that the TiO$_2$ particles and crystals are very small relative to the 1-5 µm-diameter probing area of the laser employed for acquisition of the hyperspectral images. Consequently, a random sampling of rutile orientations are presented at each pixel in the hyperspectral image. The slight variation in (1 0 0) to (0 0 1) ratio necessitated MCR-ALS models with 15 or more factors to resolve this difference in the presence of other baseline variations. Converting the relative resolved intensities of these two spectral profiles to a chemical image shows that the (1 0 0) face of rutile is predominantly parallel to the laser in the upper right of the chemical image for data set Ra15 (**Figure 10A**), while the contribution by rutile oriented with the (0 0 1) face parallel to the laser is consistent across the chemical image (**Figure 10B**). Similarly, for data set Ra23, rutile oriented with the (1 0 0) face parallel to the laser is slightly more prevalent in the lower half of the grain (**Figure 10C**), while rutile oriented with the (0 0 1) face parallel to the laser is most predominant in the upper third of the grain (**Figure 10D**). This explanation is also supported by multiple hyperspectral Raman imaging data sets producing MCR-ALS models with resolved Raman spectra that correlate well with both RRUFF rutile standard spectra.

4. CONCLUSIONS

Demonstrated here are recent advances and observations on the extraction and identification of chemically significant spectral profiles by MCR-ALS. Correlations to library spectra are employed to identify the most appropriate models for feature identification. We have shown that this method works even when there is no 'true' spectral profile in the library. It is observed that MCR-ALS models which are more complex than is optimally predicted by traditional PCA screening methods may contain chemically descriptive factors that are not well resolved in more parsimonious models. This is especially true when samples have significant baseline variance that is not well modeled by MCR-ALS. It is additionally observed that there is

no single 'best' model for the extraction of particular chemical profiles within a data set among the given components of interest; more prominent chemical profiles tend to be more accurately resolved with fewer components in the model, while more minor components need more complex models to realize the most accurate spectral estimation. Future work includes testing to see if these observations are generalizable to other linear ALS modelling methods such, as PARAFAC or Tucker models.

ACKNOWLEDGMENTS

REFERENCES

1.　Kandpal, L. M.; Tewari, J.; Gopinathan, N.; Boulas, P.; Cho, B. K., In-Process Control Assay of Pharmaceutical Microtablets Using Hyperspectral Imaging Coupled with Multivariate Analysis. *Analytical Chemistry* **2016,** *88* (22), 11055-11061.

2.　Piqueras, S.; Duponchel, L.; Tauler, R.; de Juana, A., Monitoring Polymorphic Transformations by using in situ Raman Hyperspectral Imaging and Image Multiset Analysis. *Analytica Chimica Acta* **2014,** *819*, 15-25.

3.　Boiret, M.; de Juan, A.; Gorretta, N.; Ginot, Y. M.; Roger, J. M., Distribution of a low dose compound within pharmaceutical tablet by using multivariate curve resolution on Raman hyperspectral images. *Journal of Pharmaceutical and Biomedical Analysis* **2015,** *103*, 35-43.

4.　Zhang, X.; de Juan, A.; Tauler, R., Multivariate Curve Resolution Applied to Hyperspectral Imaging Analysis of Chocolate Samples. *Applied Spectroscopy* **2015,** *69* (8), 993-1003.

5.　Piqueras, S.; Krafft, C.; Beleites, C.; Egodage, K.; von Eggeling, F.; Guntinas-Lichius, O.; Popp, J.; Tauler, R.; de Juan, A., Combining multiset resolution and segmentation for hyperspectral image analysis of biological tissues. *Analytica Chimica Acta* **2015,** *881*, 24-36.

6.　Almeida, M. R.; Correa, D. N.; Zacca, J. J.; Logrado, L. P. L.; Poppi, R. J., Detection of explosives on the surface of banknotes by Raman hyperspectral imaging and independent component analysis. *Analytica Chimica Acta* **2015,** *860*, 15-22.

7.　Smith, J. P.; Smith, F. C.; Ottaway, J.; Krull-Davatzes, A. E.; Simonson, B. M.; Glass, B. P.; Booksh, K. S., Raman Microspectroscopic Mapping with Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) Applied to the High-Pressure Polymorph of Titanium Dioxide, TiO_2-II. *Applied Spectroscopy* **2017**.

8.　Smith, J. P.; Smith, F. C.; Krull-Davatzes, A. E.; Simonson, B. M.; Glass, B. P.; Booksh, K. S., Raman Microspectroscopic Mapping with Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) of the High-Pressure, α-PbO_2-Structured Polymorph of Titanium Dioxide, TiO_2-II. *Chemical Data Collections* **2017,** *9-10*, 35-43.

9.　Stewart, S.; Priore, R. J.; Nelson, M. P.; Treado, P. J., Raman Imaging. In *Annual Review of Analytical Chemistry, Vol 5*, Cooks, R. G.; Yeung, E. S., Eds. Annual Reviews: Palo Alto, 2012; Vol. 5, pp 337-360.

10.   Opilik, L.; Schmid, T.; Zenobi, R., Modern Raman Imaging: Vibrational Spectroscopy on the Micrometer and Nanometer Scales. In *Annual Review of Analytical Chemistry, Vol 6*, Cooks, R. G.; Pemberton, J. E., Eds. Annual Reviews: Palo Alto, 2013; Vol. 6, pp 379-398.

11.   Gendrin, C.; Roggo, Y.; Collet, C., Pharmaceutical Applications of Vibrational Chemical Imaging and Chemometrics: A Review. *Journal of Pharmaceutical and Biomedical Analysis* **2008,** *48* (3), 533-553.

12.   Smith, J. P.; Smith, F. C.; Booksh, K. S., Spatial and Spectral Resolution of Carbonaceous Material from Hematite (α-Fe_2O_3) using Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) with Raman Microspectroscopic Mapping: Implications for the Search for Life on Mars. *Analyst* **2017**.

13.   Smith, J. P.; Smith, F. C.; Booksh, K. S., Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) with Raman Imaging Applied to Lunar Meteorites. *Applied Spectroscopy* **2017**.

14.   Stellman, C. M.; Booksh, K. S.; Muroski, A. R.; Nelson, M. P.; Myrick, M. L., Principal Component Mapping Applied to Raman Microspectroscopy of Fiber-reinforced Polymer Composites. *Science and Engineering of Composite Materials* **1998,** *7* (1-2), 51-80.

15.   de Juan, A.; Jaumot, J.; Tauler, R. A., Multivariate Curve Resolution (MCR). Solving the Mixture Analysis Problem. *Analytical Methods* **2014,** *6* (14), 4964-4976.

16.   Felten, J.; Hall, H.; Jaumot, J.; Tauler, R.; de Juan, A.; Gorzsas, A., Vibrational Spectroscopic Image Analysis of Biological Material using Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS). *Nature Protocols* **2015,** *10* (2), 217-240.

17.   Zhang, X.; Tauler, R., Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to Remote Sensing Hyperspectral Imaging. *Analytica Chimica Acta* **2013,** *762*, 25-38.

18.   Motegi, H.; Tsuboi, Y.; Saga, A.; Kagami, T.; Inoue, M.; Toki, H.; Minowa, O.; Noda, T.; Kikuchi, J., Identification of Reliable Components in Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS): A Data-Driven Approach across Metabolic Processes. *Scientific Reports* **2015,** *5*, 12.

19.   Vajna, B.; Patyi, G.; Nagy, Z.; Bodis, A.; Farkas, A.; Marosi, G., Comparison of Chemometric Methods in the Analysis of Pharmaceuticals with Hyperspectral Raman Imaging. *Journal of Raman Spectroscopy* **2011,** *42* (11), 1977-1986.

20.   Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R., A Graphical User-Friendly Interface for MCR-ALS: A New Tool for Multivariate Curve Resolution in Matlab. *Chemometrics and Intelligent Laboratory Systems* **2005,** *76* (1), 101-110.

21.   De Luca, M.; Ragno, G.; Ioele, G.; Tauler, R., Multivariate curve resolution of incomplete fused multiset data from chromatographic and spectrophotometric analyses for drug photostability studies. *Analytica Chimica Acta* **2014,** *837*, 31-37.

22.   Tauler, R.; Smilde, A.; Kowalski, B., Selectivity, Local Rank, 3-Way Data-Analysis and Ambiguity in Multivariate Curve Resolution. *Journal of Chemometrics* **1995,** *9* (1), 31-58.

23.   Vandeginste, B. G. M.; Derks, W.; Kateman, G., Multicomponent Self-Modeling Curve Resolution in High-Performance Liquid-Chromatography by Iterative Target Transformation Analysis. *Analytica Chimica Acta* **1985,** *173* (JUL), 253-264.

24.   Liang, X. H.; Andrews, J. E.; deHaseth, J. A., Resolution of mixture components by target transformation factor analysis and determinant analysis for the selection of targets. *Analytical Chemistry* **1996,** *68* (2), 378-385.

25.   Zhu, Z. L.; Cheng, W. Z.; Zhao, Y., Iterative target transformation factor analysis for the resolution of kinetic-spectral data with an unknown kinetic model. *Chemometrics and Intelligent Laboratory Systems* **2002,** *64* (2), 157-167.

26.   Smith, F. C.; Glass, B. P.; Simonson, B. M.; Smith, J. P.; Krull-Davatzes, A. E.; Booksh, K. S., Shock-metamorphosed Rutile Grains containing the High-pressure Polymorph TiO2-II in Four Neoarchean Spherule Layers. *Geology* **2016,** *44* (9), 775-778.

27.   Withers, A. C.; Essene, E. J.; Zhang, Y. X., Rutile/TiO2II Phase Equilibria. *Contributions to Mineralogy and Petrology* **2003,** *145* (2), 199-204.

28.   Glass, B. P.; Simonson, B. M., *Distal Impact Ejecta Layers: A Record of Large Impacts in Sedimentary Deposits*. Springer: Berlin, Germany, 2013.

29.   Lafuente, B.; Downs, R. T.; Yang, H.; Stone, N., The Power of Databases: The RRUFF Project. In *Highlights in Mineralogical Crystallography*, Armbruster, T.; Danisi, R. M., Eds. De Gruyter: Berlin, Germany, 2015; pp 1-30.

30.   Meinhold, G., Rutile and its Applications in Earth Sciences. *Earth-Science Reviews* **2010,** *102* (1-2), 1-28.

31.   Lukacevic, I.; Gupta, S. K.; Jha, P. K.; Kirin, D., Lattice Dynamics and Raman Spectrum of Rutile TiO2: The Role of Soft Phonon Modes in Pressure Induced Phase Transition. *Materials Chemistry and Physics* **2012,** *137* (1), 282-289.

32.   Lan, T.; Li, C. W.; Fultz, B., Phonon anharmonicity of rutile SnO2 studied by Raman spectrometry and first principles calculations of the kinematics of phonon-phonon interactions. *Physical Review B* **2012,** *86* (13), 7.

33.   Ohsaka, T.; Izumi, F.; Fujiki, Y., Raman-Spectrum of Anatase, TiO2. *Journal of Raman Spectroscopy* **1978,** *7* (6), 321-324.

34.   Giarola, M.; Sanson, A.; Monti, F.; Mariotto, G.; Bettinelli, M.; Speghini, A.; Salviulo, G., Vibrational Dynamics of Anatase TiO2: Polarized Raman Spectroscopy and Ab Initio Calculations. *Physical Review B* **2010,** *81* (17), 7.

35.   Tian, F.; Zhang, Y. P.; Zhang, J.; Pan, C. X., Raman Spectroscopy: A New Approach to Measure the Percentage of Anatase TiO2 Exposed (001) Facets. *Journal of Physical Chemistry C* **2012,** *116* (13), 7515-7519.

36.   Yong, W. J.; Dachs, E.; Benisek, A.; Secco, R. A., Heat Capacity and Entropy of Rutile and TiO2II: Thermodynamic Calculation of Rutile-TiO2II Transition Boundary. *Physics of the Earth and Planetary Interiors* **2014,** *226*, 39-47.

37.   Olsen, J. S.; Gerward, L.; Jiang, J. Z., On the Rutile/alpha-PbO2-type Phase Boundary of TiO2. *J. Phys. Chem. Solids* **1999,** *60* (2), 229-233.

38.   Zhu, T.; Gao, S. P., The Stability, Electronic Structure, and Optical Property of TiO2 Polymorphs. *Journal of Physical Chemistry C* **2014,** *118* (21), 11385-11396.

39.   Lan, Y. C.; Lu, Y. L.; Ren, Z. F., Mini Review on Photocatalysis of Titanium Dioxide Nanoparticles and their Solar Applications. *Nano Energy* **2013,** *2* (5), 1031-1045.

40.   Spektor, K.; Tran, D. T.; Leinenweber, K.; Haussermann, U., Transformation of Rutile to TiO2-II in a High Pressure Hydrothermal Environment. *Journal of Solid State Chemistry* **2013,** *206*, 209-216.

41.   Cai, Y. Q.; Zhang, C.; Feng, Y. P., Dielectric Properties and Lattice Dynamics of alpha-PbO2-type TiO2: The Role of Soft Phonon Modes in Pressure-induced Phase Transition to baddeleyite-type TiO2. *Physical Review B* **2011,** *84* (9), 8.

42.   Al-Khatatbeh, Y.; Lee, K. K. M.; Kiefer, B., High-pressure Behavior of TiO2 as Determined by Experiment and Theory. *Physical Review B* **2009,** *79* (13), 9.

43.　　Kusaba, K.; Kikuchi, M.; Fukuoka, K.; Syono, Y., Anisotropic Phase-transition of Rutile under Shock Compression. *Phys. Chem. Miner.* **1988,** *15* (3), 238-245.

44.　　Chen, M.; Gu, X. P.; Xie, X. D.; Yin, F., High-pressure Polymorph of TiO2-II from the Xiuyan Crater of China. *Chin. Sci. Bull.* **2013,** *58* (36), 4655-4662.

45.　　El Goresy, A.; Chen, M.; Gillet, P.; Dubrovinsky, L.; Graup, G.; Ahuja, R., A Natural Shock-induced Dense Polymorph of Rutile with alpha-PbO2 Structure in the Suevite from the Ries Crater in Germany. *Earth. Planet. Sci. Lett.* **2001,** *192* (4), 485-495.

46.　　Mammone, J. F.; Sharma, S. K.; Nicol, M., Raman-study of Rutile (TiO2) at High-pressures. *Solid State Commun.* **1980,** *34* (10), 799-802.

47.　　Mammone, J. F.; Nicol, M.; Sharma, S. K., Raman-spectra of TiO2-II, TiO2-III, SnO2, and GeO2 at High-pressure. *J. Phys. Chem. Solids* **1981,** *42* (5), 379-384.

48.　　Umari, P.; Pasquarello, A.; Dal Corso, A., Raman Scattering Intensities in alpha-quartz: A First-principles Investigation. *Physical Review B* **2001,** *63* (9), 9.

49.　　Dean, K. J.; Sherman, W. F.; Wilkinson, G. R., Temperature and Pressure Dependence of the Raman Active Modes of Vibration of α-Quartz. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **1982,** *38* (10), 1105-1108.

50.　　Chike, K. E.; Myrick, M. L.; Lyon, R. E.; Angel, S. M., Raman and Near-Infrared Studies of an Epoxy-Resin. *Applied Spectroscopy* **1993,** *47* (10), 1631-1635.

51.　　Lyon, R. E.; Chike, K. E.; Angel, S. M., In Situ Cure Monitoring of Epoxy Resins using Fiber-Optic Raman Spectroscopy. *Journal of Applied Polymer Science* **1994,** *53* (13), 1805-1812.

TABLES AND FIGURES

Table 1: Results of MCR-ALS applied to 27 hyperspectral Raman imaging data sets. These results were determined by comparison of the resolved Raman spectra generated from MCR-ALS with the target Raman spectra of reference materials. This comparison includes Raman band positions, intensities, and Raman spectral shape. The hyperspectral Raman imaging data sets are directly available in Smith et al.[8], in which the "Data Set ID" is identical to that in Smith et al.[8].

Data Set ID*	Grain ID*	Total Number of Chemical Species	Chemical Species Resolved
Ra1	z1-2	1	rutile
Ra2	z1-5	5	rutile, TiO_2-II, anatase, quartz, and epoxy
Ra3	z1-19	4	rutile, anatase, quartz, and epoxy
Ra4	z2-4	3	rutile, TiO_2-II, epoxy
Ra5	z2-4	3	rutile, TiO_2-II, epoxy
Ra6	z2-4	3	rutile, TiO_2-II, epoxy
Ra7	z3-2	2	rutile, epoxy
Ra8	z3-2	1	rutile
Ra9	z3-3	3	rutile, anatase, epoxy
Ra10	z3-4	4	rutile, anatase, quartz, and epoxy
Ra11	z3-13	5	rutile, TiO_2-II, anatase, quartz, and epoxy
Ra12	z3-13	5	rutile, TiO_2-II, anatase, quartz, and epoxy
Ra13	z3-13	5	rutile, TiO_2-II, anatase, quartz, and epoxy
Ra14	z3-13	5	rutile, TiO_2-II, anatase, quartz, and epoxy
Ra15	z3-13	5	rutile, TiO_2-II, anatase, quartz, and epoxy
Ra16	z3-14	4	rutile, TiO_2-II, anatase, epoxy
Ra17	z3-14	3	rutile, TiO_2-II, epoxy
Ra18	z3-14	3	rutile, TiO_2-II, epoxy
Ra19	z3-14	3	rutile, TiO_2-II, epoxy
Ra20	z4-1	4	rutile, TiO_2-II, anatase, epoxy
Ra21	z4-1	4	rutile, TiO_2-II, anatase, epoxy
Ra22	z4-1	3	rutile, TiO_2-II, epoxy
Ra23	z4-1	4	rutile, TiO_2-II, anatase, epoxy
Ra24	z4-1	4	rutile, TiO_2-II, anatase, epoxy
Ra25	z4-1	4	rutile, TiO_2-II, anatase, epoxy
Ra26	z4-1	4	rutile, TiO_2-II, anatase, epoxy
Ra27	z4-2	4	rutile, TiO_2-II, anatase, epoxy

*Note: ID = Identification

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
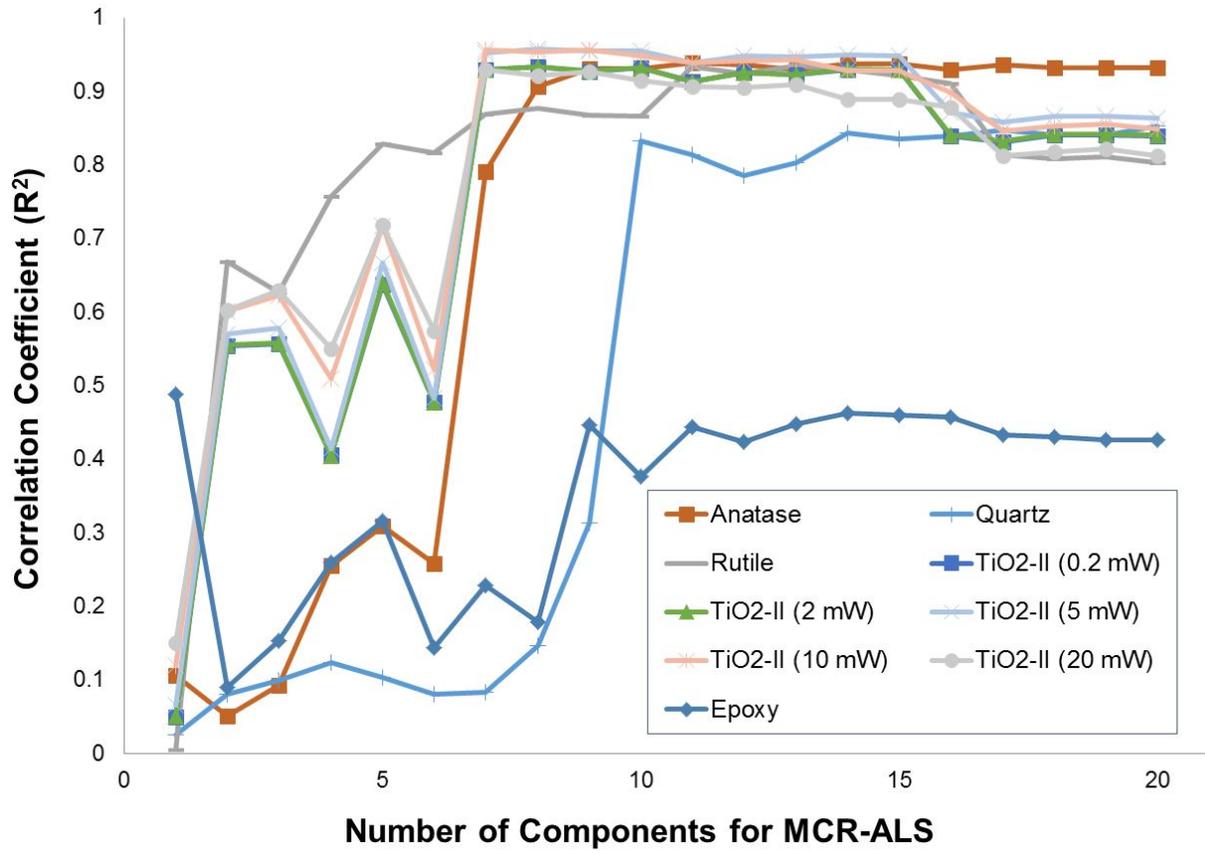
Figure 1: Correlation results for MCR-ALS models applied to hyperspectral Raman imaging data set Ra15. The resolved Raman spectra generated from MCR-ALS were quantitatively compared to target Raman spectra of reference materials, in which correlation coefficients were generated for each individual comparison. The maximum correlation coefficient is plotted for each chemical species within the given MCR-ALS model. The number of chemical components within each MCR-ALS model was varied from one to twenty.

31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
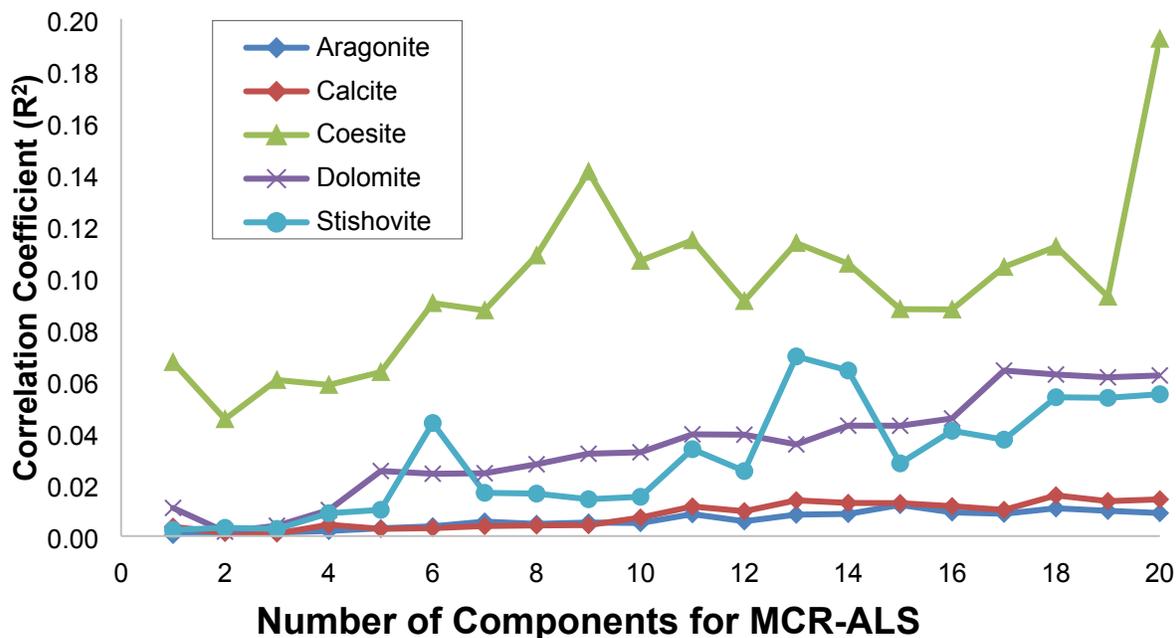
Figure 2: Correlation results for MCR-ALS models applied to hyperspectral Raman imaging data set Ra15. The resolved Raman spectra generated from MCR-ALS were quantitatively compared to target Raman spectra of reference materials not present within data set Ra15, in which correlation coefficients were generated for each individual comparison The maximum correlation coefficient is plotted for each chemical species within the given MCR-ALS model. The number of chemical components within each MCR-ALS model was varied from one to twenty.
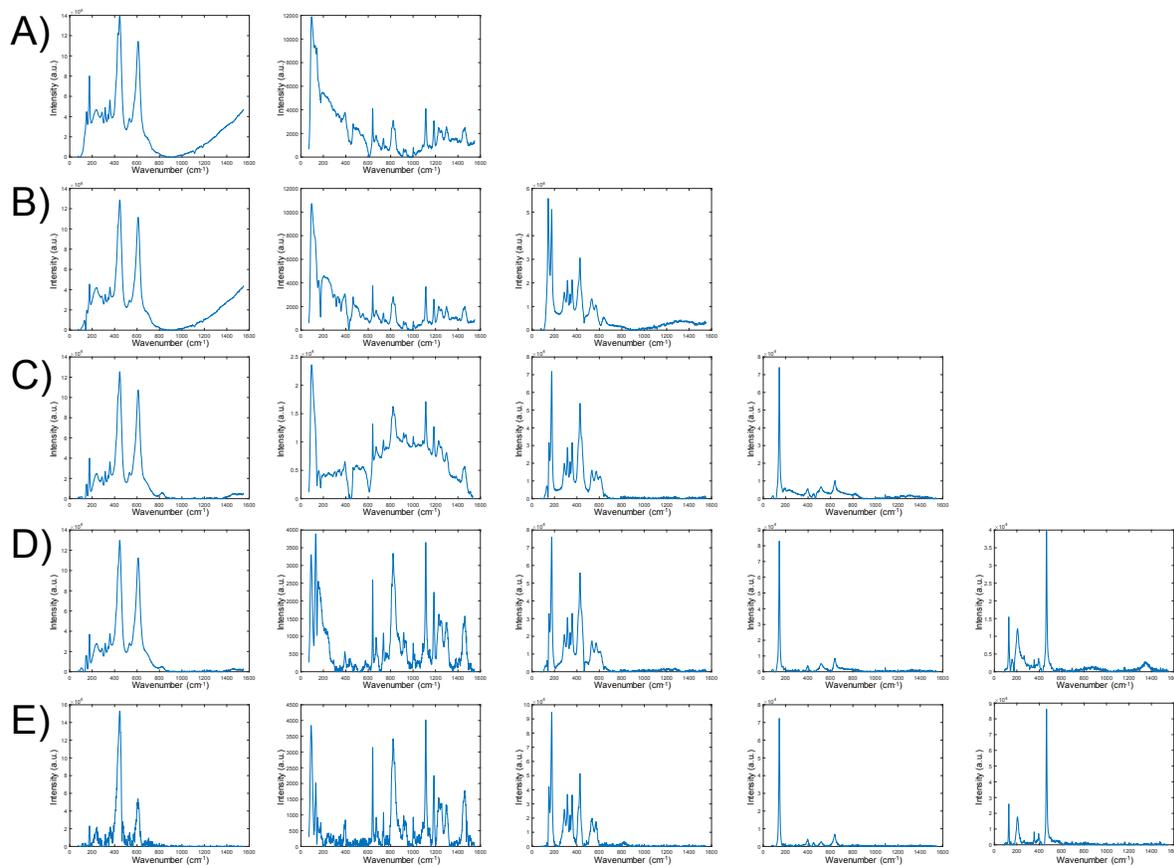
Figure 3: Resolved Raman spectra generated from MCR-ALS applied to hyperspectral Raman imaging data set Ra15. The number of components for MCR-ALS was systematically increased, in which MCR-ALS models with (A) four components, (B) five components, (C) eight components, (D) 10 components, and (E) 20 components were generated. The final MCR-ALS model selected (D) was able to resolve all five chemical species—rutile, substrate-adhesive epoxy, TiO$_2$-II, anatase, and quartz—and was statistically stable with high correlations.
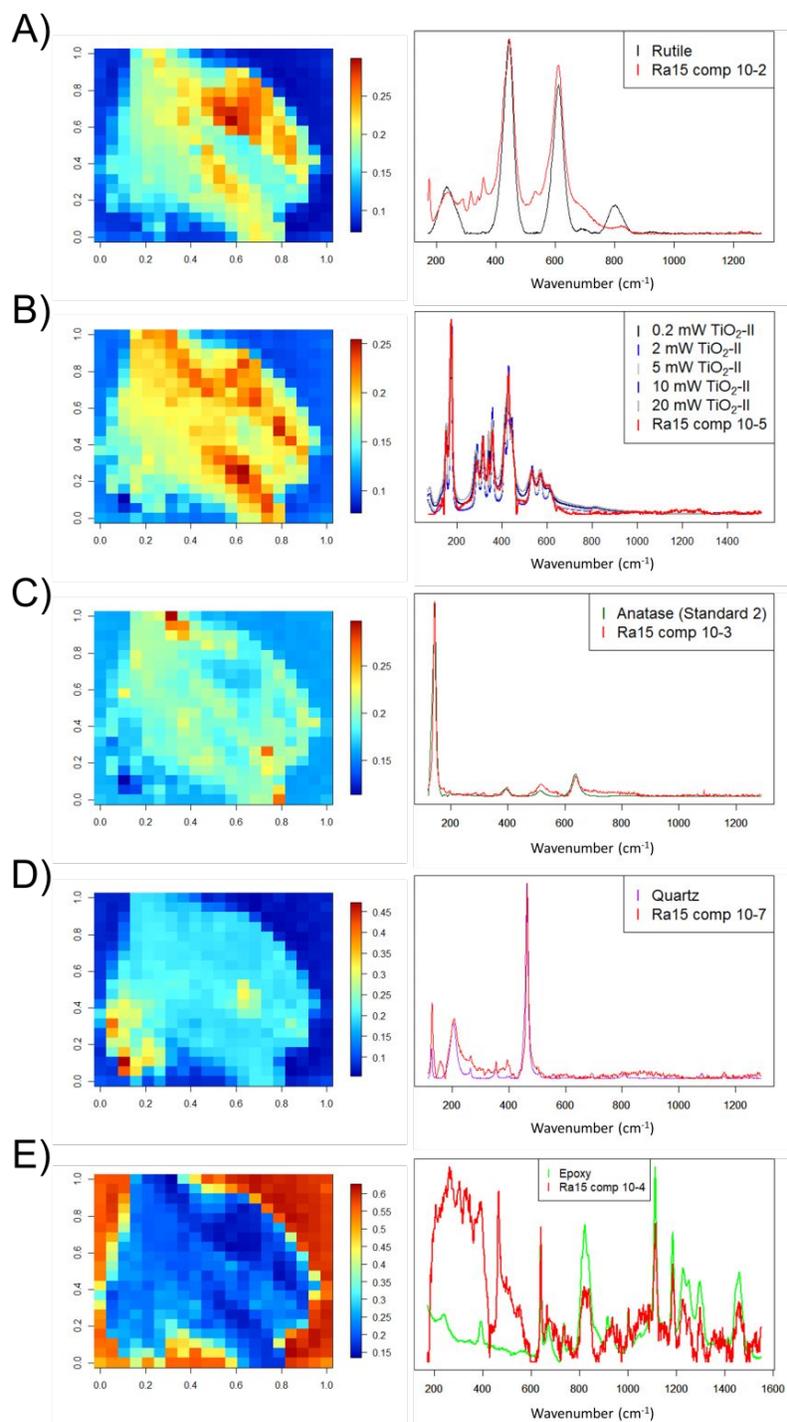
Figure 4: Spatially-resolved chemical images and corresponding resolved Raman spectra generated from a 10 component MCR-ALS model applied to hyperspectral Raman imaging data set Ra15. Chemical images and corresponding resolved Raman spectra are shown for rutile (A), TiO$_2$-II (B), anatase (C), quartz (D), and substrate-adhesive epoxy (E), the five chemical species that were resolved by this 10 component MCR-ALS model. Target Raman spectra of reference materials are displayed in addition to the resolved Raman spectra generated from MCR-ALS.
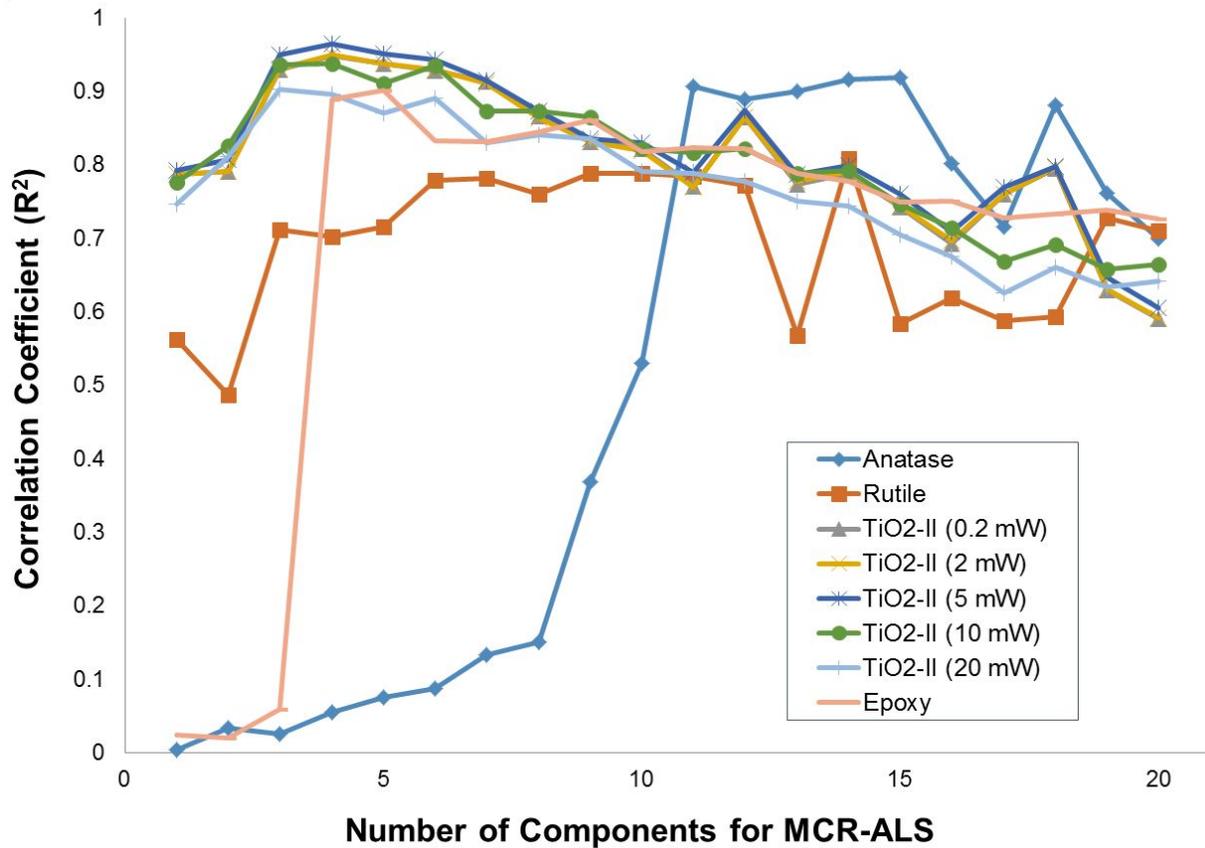
Figure 5: Correlation results for MCR-ALS models applied to hyperspectral Raman imaging data set Ra23. The resolved Raman spectra generated from MCR-ALS were quantitatively compared to target Raman spectra of reference materials, in which correlation coefficients were generated for each individual comparison. The maximum correlation coefficient is plotted for each chemical species within the given MCR-ALS model. The number of chemical components within each MCR-ALS model was varied from one to twenty.
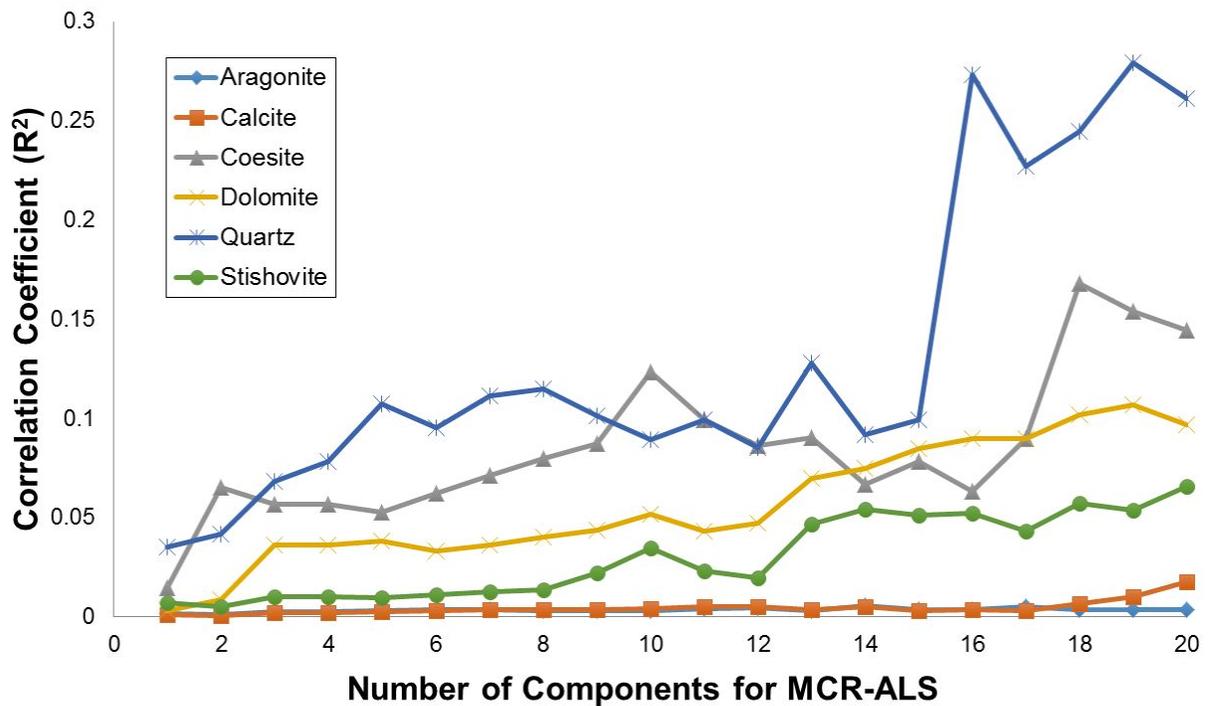
Figure 6: Correlation results for MCR-ALS models applied to hyperspectral Raman imaging data set Ra23. The resolved Raman spectra generated from MCR-ALS were quantitatively compared to target Raman spectra of reference materials not present within data set Ra23, in which correlation coefficients were generated for each individual comparison. The maximum correlation coefficient is plotted for each chemical species within the given MCR-ALS model. The number of chemical components within each MCR-ALS model was varied from one to twenty.
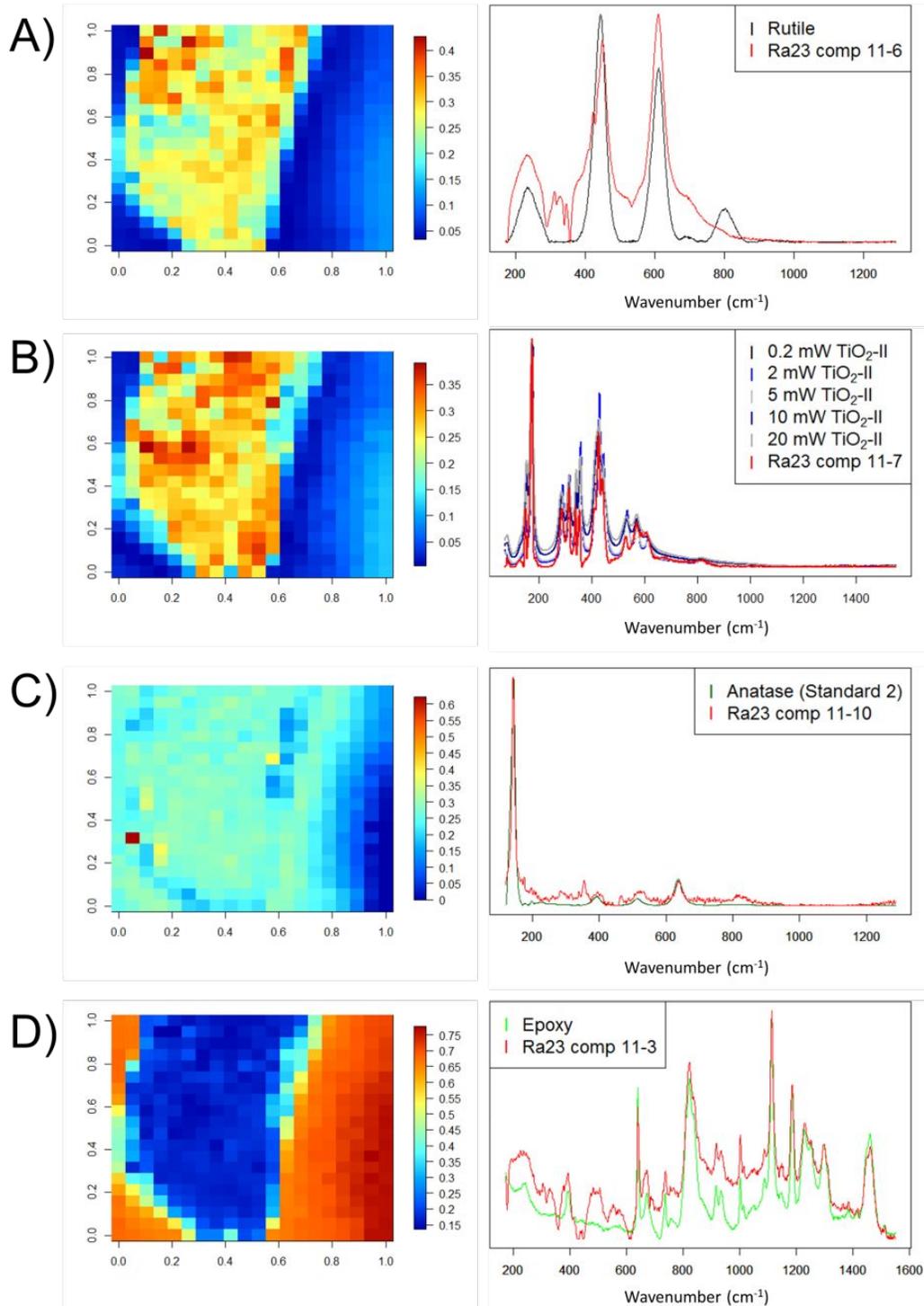
Figure 7: Spatially-resolved chemical images and corresponding resolved Raman spectra generated from an 11 component MCR-ALS model applied to hyperspectral Raman imaging data set Ra23. Chemical images and corresponding resolved Raman spectra are shown for rutile (A), TiO₂-II (B), anatase (C), and substrate-adhesive epoxy (D), the four chemical species resolved by this 11 component MCR-ALS model. Target Raman spectra of reference materials are displayed in addition to the resolved Raman spectra generated from MCR-ALS.
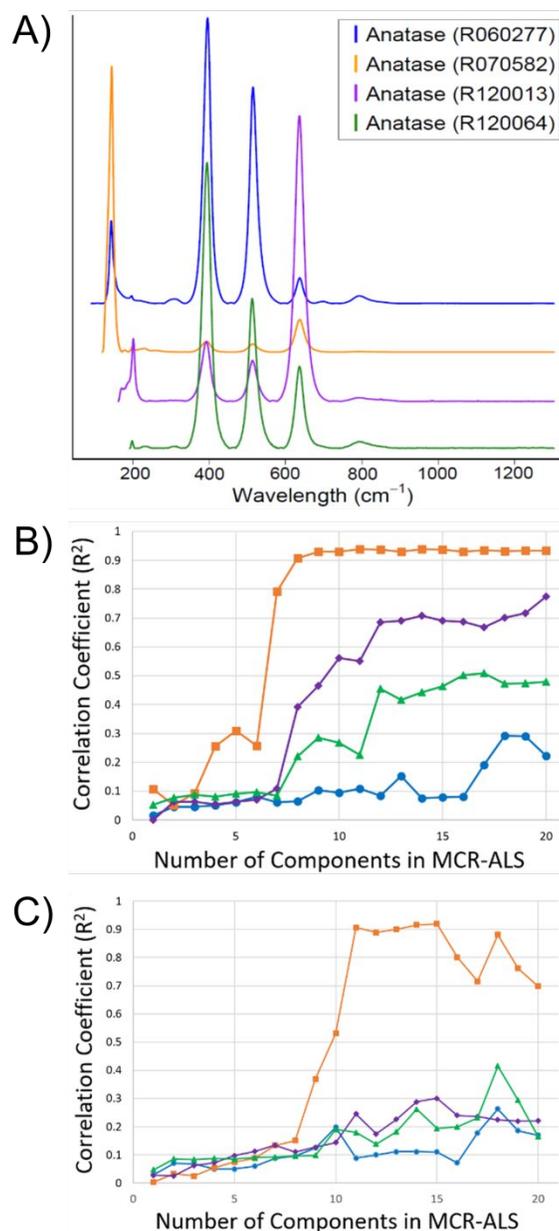
Figure 8: The effects of using different anatase standards from the RRUFF database as reference materials on the MCR-ALS-based results for data sets Ra15 and Ra23. Four Raman spectra for anatase standards from the RUFF database are shown (A). Correlation results for these four reference Raman spectra with the resolved Raman spectra from data sets Ra15 (B) and Ra23 (C) are shown. In (A), (B), and (C), the blue, orange, purple, and green colors correspond to anatase standards R060277, R070582, R120013, and R120064, respectively.
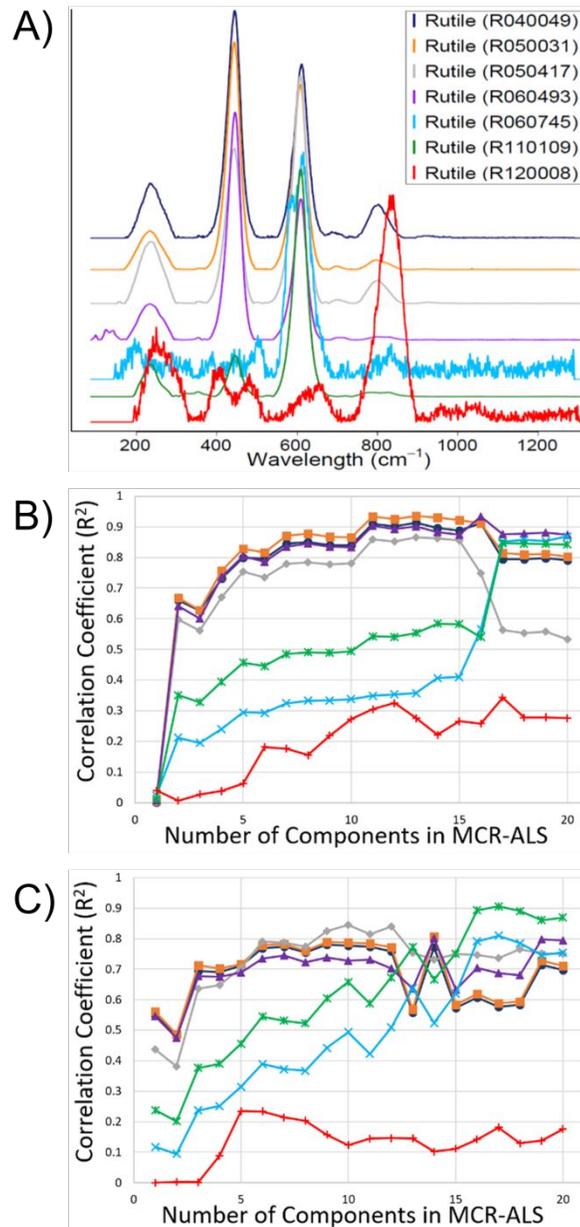
Figure 9: The effects of using different rutile standards from the RUFF database as reference materials on the MCR-ALS-based results for data sets Ra15 and Ra23. Seven Raman spectra for rutile standards from the RRUFF database are shown (A). Correlation results for these seven reference Raman with the resolved Raman spectra from data sets Ra15 (B) and Ra23 (C) are shown. In (A), (B), and (C), the dark blue, orange, gray, purple, light blue, green, and red colors correspond to rutile standards R040049, R050031, R050417, R060493, R060745, R110109, and R120008, respectively.
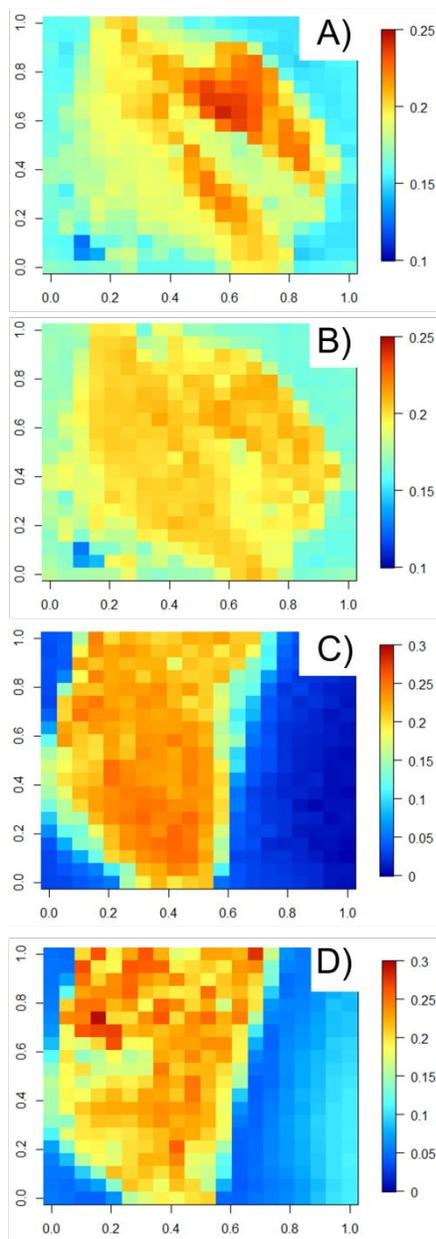
Figure 10: Spatially-resolved chemical images for different rutile reference Raman spectra for data sets Ra15 (A and B) and Ra23 (C and D) are shown. The rutile reference Raman spectra, obtained from the RRUFF database, had either one or two major Raman bands. For those Raman spectra that had two major Raman bands, RRUFF rutile standard R050031 had the highest correlation with data sets Ra15 and Ra23 (A and C, respectively). For those Raman spectra that had one major Raman band, RRUFF rutile standard R110109 had the highest correlation with data sets Ra15 and Ra23 (B and D, respectively).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
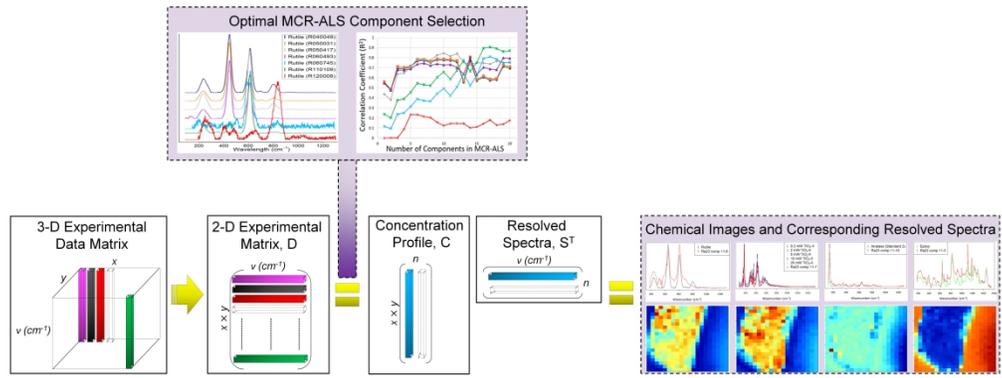
Graphical abstract

2372x880mm (96 x 96 DPI)

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60