Title: **Quantification of synthetic non-metallic inclusion multiphase mixtures from a CaO–Al₂O₃–MgO–CaS system using Raman spectroscopy.**

Francis Gyakwaa^{1)}, Matti Aula¹⁾, Tuomas Alatarvas¹⁾, Qifeng Shu¹⁾, Marko Huttula²⁾ and Timo Fabritius¹⁾*

1) Process Metallurgy Research Unit, University of Oulu, P.O. Box 4300, FI-90014, University of Oulu, Finland.

2) Nano and Molecular Systems Research Unit, University of Oulu, P.O. Box 4300, FI-90014, University of Oulu, Finland.

Corresponding author: Francis Gyakwaa, E-mail: Francis.gyakwaa@oulu.fi, Process

Metallurgy Research Unit, University of Oulu, P.O. Box 4300, FI-90014, University of Oulu, Finland.

Abstract

Raman spectroscopy has features such as its relatively easy sample preparation. Therefore, its application could play a role in non-metallic inclusion studies. This study assessed the use of Raman spectroscopy to quantify multiphase synthetic inclusion mixtures consisting of C12A7, CA, C3A, CaS and MgO·Al₂O₃. Partial least squares (PLS) regression was used to obtain a calibration model, and the enhancement of the model performance was done using standard normal variate (SNV). From the calibration model, the root mean standard error in cross-validation (RMSECV) values ranged between 3.82–6.13 wt%, the root mean square error of prediction (RMSEP) was within 2.0–4.04 wt%, and the R² value was between 0.93–0.99 for estimating the phases based on SNV Raman data. The raw Raman spectra data had RMSECV values between 9.83–16.46 wt%, the RMSEP ranged between 9.41–15.33 wt% and the R² was estimated within 0.71–0.95. The PLS regression had a satisfactory prediction performance with high range error ratio (RER) and ratio of prediction to deviation (RPD) values for SNV Raman data. The study demonstrates the use of Raman spectroscopy and a calibration model to quantify the specific phase in a multiphase synthetic inclusion mixture in CaO–Al₂O₃–MgO–CaS system.

KEYWORDS: Raman spectroscopy; MgO·Al₂O₃ spinel; calcium aluminate; non-metallic inclusion; CaS.

1. Introduction

The need to ensure steel cleanliness continues to gain the attention of steelmakers and steel product end-users. Non-metallic inclusions form an essential parameter for assessing steel cleanliness and for process control during the steelmaking process^[1,2]. For aluminium killed steels, calcium treatment is mostly used for transforming solid Al_2O_3 inclusions into low melting calcium aluminate inclusions of C12A7 or phases close to C12A7 such as CA and C3A, and also to reduce the total oxygen content^[2,3]. However, the presence and high content of elements such as Mg and S in the molten steel, as well as the effectiveness of the deoxidation and Ca treatment processes, could result in the formation of potentially detrimental inclusions such as magnesium aluminate spinel ($\text{MgO} \cdot \text{Al}_2\text{O}_3$ or MA) and solid CaS and could cause process issues such as submerged entry nozzle (SEN) clogging^[1].

Numerous works^[4–12] in the literature had been dedicated to the study of inclusion formation, evolution, and modification for Al-killed calcium treated steels with the aim to ensure steel cleanliness and process control. Studies by Gollapalli et al.^[4] demonstrated that for high-sulphur Al-killed steel samples, ineffective calcium treatment could generate oxy-sulphide non-metallic inclusions containing CaS. Some possible ways that CaS could be distributed or interact with oxides (such as calcium aluminates and MA spinel) include possible collision with the oxides surrounding the oxide phases or uniformly distributed within the oxide inclusion^[5,12]. Zhang et al.^[5] further suggest that CaS collision with oxides is usually generated after the addition of Ca, CaS surrounds or envelops oxides during the evolution process when Al_2O_3 transforms into calcium aluminates, while the uniform distribution of CaS and oxides is a result of the precipitation of CaS dissolved in the calcium aluminate as the casting temperature reduces. Studies^[4, 5, 8, 12] conducted on steel samples have also demonstrated that oxy-sulphide, non-metallic inclusions types could be formed within $\text{CaO-Al}_2\text{O}_3\text{-CaS}$ and $\text{CaO-Al}_2\text{O}_3\text{-MgO-CaS}$

systems. At steelmaking temperatures, a composition near C12A7 is fully liquid, whereas C12A7–C3A and C12A7–CA ranges also contain partially liquid regions.

The use of various analytical characterisation techniques for the identification and classification of inclusions have been outlined by Zhang et al.^[13]. Despite the use of these conventional analytical methods, some of them are time consuming regarding the sample preparation or the analyses, and some are sample destructive, potentially limiting their use. The utilisation of Raman spectroscopy performs better and requires less effort and requires a shorter time for the sample preparation and acquisition and analysis of the measurement results. Additionally, calcium aluminate phases,^[14–16] magnesium aluminate spinel^[17] and some sulphides such as CaS^[18,19] have been demonstrated to be Raman active. Furthermore, Li and Hihara^[20] have used Raman spectroscopy techniques to characterise some calcium aluminate inclusion phases such as CA2 and CA in steel samples.

The present authors previously conducted studies concerning the use of Raman spectroscopy that were focused on characterising calcium-aluminate (CaO–Al₂O₃) phases around the liquidus region in the CaO–Al₂O₃ binary system^[14], as well as in the magnesium aluminate spinel (MgO·Al₂O₃)–calcium aluminate binary system^[17], and also for duplex oxide-sulphide synthetic inclusions^[19]. It has been suggested^[21, 22] that inclusions could occur as a multi-component system consisting of CaO–Al₂O₃ (such as CA, C12A7 and C3A), CaS and MgAl₂O₄. **Figure 1** connects our most recent studies^[14, 17, 19, 23] conducted using Raman spectroscopy for characterising the synthetic inclusions and this current study, and also shows inclusion classification conducted by Thunman and Sichen^[11]. They suggested that inclusions studied in **Figure 1 a)** are mostly found during ladle treatment (after tapping, through to after degassing). They also point out that the inclusions indicated in Fig1. b) and d) could be formed in the ladle glaze (after tapping and degassing), while the inclusion types in **Figure 1 c)** are

generally noticed at later stages in the ladle treatment, and finally the inclusions in **Figure 1 e)** are formed after degassing and before casting is done.

Furthermore, Guo et al.^[21] have shown that the route of inclusion modification and also the calcium content in the steel melt might affect the final inclusion composition. However, previous studies concerning the application of Raman spectroscopy for synthetic^[14, 17, 19, 23], or steel samples^[20] and non-metallic inclusion studies have been limited to characterising only binary system between (CaO–Al₂O₃) and the (CaS, Al₂O₃ or MgAl₂O₄) or MgO–CaO–Al₂O₃ systems. Consequently, further studies regarding the use of Raman spectroscopy for characterising a mixture of phases found in CaO–Al₂O₃–CaS and CaO–Al₂O₃–MgO–CaS systems will contribute to expanding the potential use of this technique for non-metallic inclusions analysis.

This work examines the application of Raman spectroscopy for the identification and quantification using a multivariate technique for evaluating the phase content of CA, C12A7 and C3A, CaS and MgAl₂O₄ spinel (MA) in the ternary system of C12A7–CaS–CA, MA–C3A–CaS, MA–CA–CaS, C12A7–CaS–C3A and MA–C12A7–CaS. Additionally, the prospective use of this work for inclusions in the steel matrix in future studies was examined.

Figure 1. Schematic chart illustrating how this current study is associated with the authors' previous studies and basic inclusion formation classification.

2.0. Experimental

2.1. Sample preparation

Chemical powders that consist of aluminium oxide (Al_2O_3), calcium sulphide (CaS), calcium oxide (CaO), aluminate spinel, $\text{MgO} \cdot \text{Al}_2\text{O}_3$ (MA) and monocalcium aluminate $\text{CaO} \cdot \text{Al}_2\text{O}_3$ (CA) used for this study were obtained from Alfa Aesar and had a purity that ranged from 99.7% to 99.9%.

2.1.1. Preparation of synthetic multiphase samples

C3A and C12A7 phases were not commercially available; therefore, these were prepared by weighing a required ratio of $\text{CaO}/\text{Al}_2\text{O}_3$ of calcium oxide (CaO) and aluminium oxide (Al_2O_3). The weighed components were pressed into pellets and sintered at a temperature of 1,350 °C in a chamber furnace and then the sintered samples were milled. The process was repeated until a purity of approximately 99% was achieved for C12A7 and C3A phase samples. Other phases [$\text{CaO} \cdot \text{Al}_2\text{O}_3$ (CA), $\text{MgO} \cdot \text{Al}_2\text{O}_3$ (MA) and (CaS)] used for this study were directly from Alfa Aesar with relatively good purity.

The synthetic multi-component samples were prepared by weighing the required proportions of the phases, which were CaS , C12A7, $\text{MgO} \cdot \text{Al}_2\text{O}_3$, CA, and C3A. The prepared multiphase synthetic (binary and ternary) samples were not further sintered at a higher temperature since the synthetic phase of CaS is considered very sensitive to heat as recommended by the supplier (Alfa Aesar) and as suggested by Avril et al. ^[18]. The sample mixtures were milled to approximately the same grain size, with careful and thorough mixing of the samples carried out several times to ensure that homogenous sample mixtures were obtained. The samples were milled to small grain size and mixed to get multiphase inclusions samples with small particles of the phases as possible and also ensures that inside each laser shot there will be multiple phases during the Raman spectroscopy measurements. The homogeneity of the samples was verified using both X-ray fluorescence (XRF), and X-ray diffraction (XRD).

2.1.2. Multiphase sample design

The five sets of multiphase synthetic inclusion samples (C12A7–CaS–CA, MA–C3A–CaS, MA–CA–CaS, C12A7–CaS–C3A and MA–C12A7–CaS) used for this study were prepared based on a full factorial experimental design. The complete design matrix was carried out using k^n compositions, with n denoting the design variables or a number of phases and k shows the number of design levels. The sample matrix was structured such that each set of the multiphase sample had a specific phase content (weight percentage, wt%) ranging between 0 wt% to 90 wt%. Furthermore, the samples matrix was designed to reflect a general oxy-sulphide non-metallic inclusion composition. As demonstrated by researchers^[4, 5, 8, 12], inclusions could be generated to consist of a multiphase (binary or ternary) from a relatively complex system of MgO–CaO–Al₂O₃ and MgO–CaO–Al₂O₃–CaS during inclusion formation and evolution processes. Examples of this include studies conducted on the evolution of inclusions for slab hot rolled plates^[8] and studies of the types of inclusions at different stages of casting^[11, 12]. Therefore, the sample matrix which we designed in this study could contribute to characterising inclusion mixtures formed after degassing during casting, and in cast steel products.

The total number of samples for the multiphase system were 43 for C12A7–CaS–CA, 43 for C12A7–CaS–C3A, 32 for MA–C12A7–CaS and with MA–C3A–CaS and MA–CA–CaS each with 35 samples. Table 1 shows the details concerning the multiphase system and the number of samples split (binary and ternary). In **Table 1** the numbering i–v is further classified into potential systems from which the set of synthetic inclusions samples were prepared. The labels i - ii in **Table 1** are samples prepared from the CaO–Al₂O₃–CaS system and iii–v are sample mixtures obtained from the MgO–CaO–Al₂O₃–CaS system.

Table 1. The sample matrix used for this study

No.	Sample system	Number of samples		
		Binary	Ternary	Total
i	C12A7–CaS–CA	22	23	45
ii	C12A7–CaS–C3A	18	25	43
iii	MA–C12A7–CaS	12	20	32
iv	MA–C3A–CaS	15	20	35
v	MA–CA–CaS	10	25	35

2.3. Characterisation Techniques

XRD and XRF instruments were used to evaluate and verify the phase content and elemental composition in the samples. The XRD was used for estimating the sample weight percentage for each specific phase present in the sample. The XRF was used to assess the elemental composition of the samples prepared. The XRD and XRF instruments' setup parameters can be found in other studies ^[14].

2.3.1. Raman spectroscopy

Raman spectroscopy is a type of vibrational spectroscopic technique that operates using an inelastic scattering of monochromatic light from a laser source, which results in a change in energy which can be used to analyse the vibrational and rotational modes of the excited molecules in a material^[24]. Changes in the energy in a sample are characteristic of the vibrational modes of the molecule in the material^[24]. The Raman spectrum, thus acquired from a sample is considered to be a fingerprint of the individual component present in the sample. The use of Raman spectroscopy for metallurgical slag studies^[25, 26] and others^[27, 28] has been demonstrated to be applicable for the identification and characterisation of materials.

A TimeGated® Raman spectrometer (TG532 M1) supplied by TimeGated instruments Ltd., Oulu, Finland was used in this study. Details concerning the features of Raman spectroscopy can be found in other literature ^[14]. The same analytical parameters associated with the Raman

spectroscopy such as laser power (intensity), spectra acquisition and resolution were used during the sample measurements. The Raman spectroscopy measurements were carried out in ambient conditions. Raman shift region or wavenumbers of 100 -1,100 cm^{-1} were used for this study. A spectra acquisition time of 1-3 minutes, a resolution of 10 cm^{-1} and total spectra with at least four replicates were collected for each sample. A rotating sample holder stage was used to obtain an average Raman spectrum for each batch of the multiphase sample measured. These procedures were used during the Raman spectroscopy measurements to ensure relatively good repeatability and reduce the potential effects of sample inhomogeneity during the multi-component preparation stage.

The Raman spectra acquired from each sample were shown to be relatively stable; therefore, the variability of the Raman signal within each sample measured in the Raman data is considered to be relatively insignificant. The pre-processing software supplied by TimeGated instrument Ltd. was used to treat the signal using a background subtraction. Additionally, the samples used for this study are not known to be affected by the fluorescence effect. However, the presence of any potential fluorescence effect on the acquired spectra could be effectively controlled by the use of time-gated Raman spectroscopy (TimeGated®). This type of Raman spectroscopy was designed for effective fluorescence suppression compared to the use of conventional Raman spectroscopy^[29].

2.4. Multivariate Data analysis

Previous studies have demonstrated the use of linear regression based on the relative intensity of the Raman peak and the phase content in binary samples can be used for estimating a specific phase^[14]. However, for sample mixtures containing more than two components such as a ternary phase system, more elaborated mathematical tools should be used for treating the Raman spectra data for a better quantitative analysis of multiphase samples^[27, 30, 31]. Additionally, our previous studies^[23] also showed that the utilisation of advanced modelling methods was

required for more than two phases in a sample to estimate the specific phase present. For multiphase (more than two phases) Raman spectra analysis, peak overlapping can be a problem due to the increased number of Raman peaks. Additionally, according to Grinberg et al.^[30] and Wold et al.^[31], the use of multiple linear regression (MLR) methods for treating original Raman data could potentially be very unstable or give inaccurate estimates of the individual phases in the samples.

This challenges associated with MLR could be linked to potential high variable collinearity found in some of the Raman spectra data as the number of phases in the samples increases^[30, 31]. These challenges can be mitigated with the use of a robust multivariate technique such as principal component analysis (PCA), with partial least squares (PLS) regression to extract quantitative information from the Raman spectra data obtained from the samples^[28, 32, 33]. The PLS regression technique makes it possible to simultaneously analyse and quantify a specific phase in a sample mixture in a situation where there is a potential peak overlap.

Preliminary, PCA was used to reveal the Raman spectral data structure and for data reduction. The PLS regression method was applied to both the raw Raman spectral data and pre-processed Raman spectra data using an algorithm known as standard normal variate (SNV). SNV was employed to eliminate potential sources of non-linearity or reduced features uncorrelated with the concentrations of the analyses from the measured Raman spectra^[34]. Additionally, SNV can be used to correct potential baseline shifting or background effects^[34]. This study also examines the performance of a PLS regression based on using the raw Raman spectra data and pre-treated Raman spectra data measured from the samples.

The original Raman spectral data were randomly split into a calibration set for creating the quantification model, and the prediction set used to test the model. The calibration set or training set was made up of 75% of the samples and 25% of the remaining samples as the prediction or testing (external validation) set. In carrying out the PLS analysis, the covariance

between the Raman spectra data (X variables - spectra intensities) and concentration of the different phases (as Y variables) was calculated. In the calibration procedure, the Raman spectra acquired from the samples had X variables marked as an independent or instrumental response where n represents the number of samples, and w represents the variable number. Additionally, the predicted variables Y, (corresponding to the phase of interest), n the number of the samples for the and z is the number of constituents. Equation 1 shows the matrix used for the X variables (independent) and the dependent variables Y are represented by equation 1 and 2 to explain the PLS regression^[35]. Therefore, briefly PLS the X and Y variables are expressed as^[35]:

$$X = TP^T + E_X, \quad 1$$

$$Y = UQ' + E_Y, \quad 2$$

where T is the score matrix for X variables, E_X is the residue for X, P^T is the loading for the X variables. Additionally, U is the score matrix for the Y variables, E_Y is the residue for Y, Q' is the loading for the Y variables.

The cross-validation process was carried out by using the leave-one-out cross-validation technique to assess the accuracy of the model. Additionally, the leave-one-out cross-validation technique was used to evaluate the optimal latent variables or PLS factors. The Raman spectral range of 100–1000 cm^{-1} was found to be suitable for acquiring a Raman spectrum that contained the main Raman peaks (bands) for the prepared samples and was also used for the PLS model for this study.

The performance for the calibration model was assessed by using the coefficient of determination (R^2), root mean standard error in calibration (RMSEC), and the root mean standard error in cross-validation (RMSECV) for the calibration set^[27, 36]. The RMSECV can be explained as a relation between the spectral data and reference values in a cross-validation

sample set based on the standard deviation of differences. The root mean standard error (RMSE) is expressed as^[37]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad 3$$

Where y_i represents the reference value, and the number of samples marked as n and $\hat{y}$ denote the calculated value.

For the test set (external validation) or prediction set, the root mean square error of prediction (RMSEP), determination coefficient, the ratio of prediction to deviation (RPD) and the range error ratio (RER)^[27, 37–39] were used to assess the external validation (prediction) performance.

Table 2 summarises the model performance assessment that shows the suitability of the calibrations, as suggested by other studies^[36]. According to some research studies^[36–38], calibration performance ranking could be done with RPD value where < 1.5 is unsatisfactory, 3–4.9 is generally satisfactory and > 5 considered as excellent. Similarly, these researchers^[36–38] demonstrated that an RER value < 4 is unsatisfactory, ≥ 10 is satisfactory, and ≥ 15 is excellent for calibration performance ranking. The range error ratio (RER) and the ratio of prediction to deviation (RPD) according to Teixeira dos Santos *et al.*,^[36] can be expressed as:

$$RER = \frac{(Max - Min)_{refv}}{RMSEP}, \quad 4$$

Where Max is maximum, Min is minimum, and $refv$ denotes reference value.

$$RPD_{pred} = \frac{SD_{PRD}}{RMSEP}, \quad 5$$

Where SD_{PRD} is the standard deviation for the prediction set.

The Raman spectra pre-processing and multivariate analysis (PCA and PLS regression) were performed by using UNSCRAMBLER (version 11.0), which is a statistical software package

made by Camo Software, Oslo, Norway. This statistical software has been demonstrated to be a highly promising spectroscopic multivariate analysis by other researchers^[33, 39].

3.0. Results and Discussion

3.1. Raman spectra analysis based on phase content

The Raman shift (bands) assigned to each phase used for preparing the multiphase samples and the referenced literature studies are shown in **Table 2**. Details concerning the bonding structure for the Raman band (peak) characteristic and identified for each phase's component are presented in previous work^[23]. The selection of Raman peaks candidates for distinguishing the phases present in a sample mixture was carried out such that to reduce the effect of peak overlapping by utilising peaks in a different region. For example, within 500 – 600 cm⁻¹ both CA and C12A7 have some Raman peaks associated these phases within this region, however previous work^[14] had shown that it is possible to use different Raman peaks such as at 312 – 333 cm⁻¹, 772 – 781 cm⁻¹ (assigned to C12A7) for distinguishing these phases.

Table 2. Measurement Raman bands for the synthetic phase used for this study.

Phase	Measured peaks [Raman shift, cm ⁻¹]	Reference [Raman shift, cm ⁻¹]	Reference
CaS	160 <i>s</i> , 200 <i>m</i> , 478 <i>w</i>	160 <i>s</i> , 185 <i>m</i> , 215 (±15), 485 (±10) <i>w</i>	[15]
CA	525 <i>s</i> , 549 <i>m</i> , 790 <i>w</i>	520–521 <i>s</i> , 545–547 <i>m</i> , 790–793 <i>w</i>	[12], [13]
C12A7	522 <i>s</i> , 781 <i>m</i> , 314 <i>m</i>	312–333 <i>m</i> , 516–517 <i>m</i> , 772 <i>m</i> , 779 <i>m</i>	[12], [13]
C3A	756 <i>s</i> , 508 <i>m</i> ,	756–757 <i>s</i> , 140–150 <i>w</i> , 506–508 <i>m</i>	[12], [13]
MgO.Al ₂ O ₃	416 <i>s</i> , 674 <i>m</i> , 773 <i>m</i>	409–412 <i>s</i> , 767–772 <i>m</i> , 666–674 <i>m</i> , 312–313 <i>m</i>	[40]

Note: *s* denotes a strong Raman band, *m* is medium, and *w* shows a weak Raman band.

3.2. Observations based on variation in the phase content from the Raman measurements

The Raman spectra obtained from the samples were observed to have a characteristic Raman band with relatively good Raman peak intensity. Consequently, they could be used to estimate constituent (phase) variation in the samples. **Figure 2 (i) and (ii)** illustrates examples of the Raman spectra measured from the ternary samples, where the effect of Raman peak intensities ratios could be related to the change in phase content (wt%) in the samples. A variation in the phase content (wt%) in the samples measured using Raman spectroscopy had a corresponding change in the relative intensity of the peaks which were characteristic of a specified phase that made up the sample, thus, offering the opportunity to estimate sample phases qualitatively.

The C12A7–CaS–C3A samples matrix values the composition for the Raman spectra in **Figure 2(i)** (in wt%), are 70C12A7–10CaS–20C3A for a), 10C12A7–70CaS–20C3A for b), 20C12A7–10CaS–70C3A for c). Similarly, **Figure 2 (ii)** shows the Raman spectra for MA–C12A7–CaS samples with the phase content (wt%) as 70MA–20C12A7–10CaS for d), 35MA–45C12A7–20CaS for e, and 45MA–10C12A7–45CaS for f.

For the C12A7 phase content in the ternary sample presented in **Figure 2 (i)**, it could be observed from spectra label **a** (70 wt% C12A7) that when a relative Raman band intensity in the region of 517–522 cm^{-1} decreased toward **c** (20 wt% C12A7), there was a corresponding effect when the phase for C12A7 in the samples was reduced. This is because the most intense Raman peak which is also a characteristic feature of the C12A7 phase is located at 517 cm^{-1} . A similar trend can be seen between the relative Raman band intensity and phase content for C12A7 in **Figure 2 (ii)** in the Raman spectra **e** and **f**.

Additionally, Raman spectra **a** and **b** in **Figure 2 (i)** illustrate how a change in the CaS phase content in the C12A7–CaS–C3A samples affected the relative Raman intensity. Furthermore, **Figure 2 (i) a and b** show that an increment in the CaS phase content in the C12A7–CaS–C3A samples demonstrated a corresponding rise in the relative Raman band intensity within a region

of $157\text{--}162\text{ cm}^{-1}$. Thus, it could be attributable to CaS since its most intense Raman peak was approximately 157 cm^{-1} .

Figure 2. Raman spectra for **i)** C12A7-CaS-C3A and **ii)** MA-C12A7-CaS ternary samples. The phase composition for C12A7-CaS-C3A samples are 70C12A7-10CaS-20C3A for a), 10C12A7-70CaS-20C3A for b), 20C12A7-10CaS-70C3A for c), and for MA-C12A7-CaS samples; 70MA-20C12A7-10CaS for d), 35MA-45C12A7-20CaS for e), and 45MA-10C12A7-45CaS for f) phase composition.

Another illustrative example for qualitative analysis relates MgO·Al₂O₃ spinel (MA) phase contents to the relative Raman intensity to estimate the MA amount in MA-C12A7-CaS samples. This phenomenon could be noticed in the Raman spectra labelled **d** and **e**, as shown in **Figure 2 (ii)**. MA has its most intense Raman shift (band) in the region of $410\text{--}420\text{ cm}^{-1}$, hence, in **Figure 2 (ii)** the Raman peak intensity increases relatively higher for spectra **d** (with 75 wt% of MA) compared to spectra **e** (with 30 wt% of MA). Consequently, these observations were made for the other ternary samples of C12A7-CaS-CA, MA-C3A-CaS and MA-CA-CaS in this study, and these all followed similar phase content and relative Raman band intensity. Therefore, based on the examples presented in **Figure 2 (i and ii)**, a qualitative observation could be made based on the change in the phase content in the context of the ternary samples with the corresponding variation with the relative Raman shift (band) intensity.

3.3. Sample phase content estimation using a calibration model

Evaluation of the phases contained in the multiphase samples was carried out using a PLS regression model to predict the phase content (wt%) based on their Raman spectra data. The PLS regression model was used to establish the relationship between the Raman signals and the phase concentration (wt%) in the samples prepared. A detailed description of the multivariate technique is found in **section 2.4**. The (PLS) regression model was performed on both the raw Raman spectral data and pre-processed Raman spectra for quantification of the phases present in the samples.

Tables 3 and 4 show the estimated values for the coefficient of determination R^2 (for calibration, validation, and prediction), the root mean square of calibration (RMSEC), the root mean square of cross-validation (RMSECV), and the root mean square error of prediction (RMSEP). These are model performance or statistical parameters for the raw Raman spectra data and Raman spectra treated with SNV. Additionally, **Figures 3 and 4** are the calibration graphs showing the relation between the weight percentage, wt% (concentration) and weight percentage predicted in the multiphase system, consisting of two or three phases of C12A7, CA, CaS, C3A and MA.

3.3.1. *CaO–Al₂O₃–CaS system*

Ternary samples presented in this section consist of C12A7, CaS, CA and C3A phases, and **Figure 3 (i - iv)** illustrates the model performance values based on a PLS regression for C12A7–CaS–C3A and C12A7–CaS–CA analysis. The estimated phase content for C12A7 in C12A7–CaS–C3A samples is shown in **Figure 3 (i and ii)**, where *i* is calibration graph for the raw Raman spectral data and *ii* denotes the pre-processed Raman spectral data using SNV. Similarly, **Figure 3 (iii)** is a calibration graph for the raw Raman, and *iv*) represents that of SNV treated Raman data for C12A7–CaS–CA samples. The estimated coefficient of determination for the calibration (R^2_{CAL}) and for the cross-validation (R^2_{CV}) values ranged from as low as 0.33

and less than 0.95 for the raw Raman spectra data. For the Raman data pre-processed with SNV method, the evaluated values for R^2_{CAL} and R^2_{CV} ranged between 0.95–0.99.

The figure of merit (coefficient of determination, R^2) presented in **Tables 3 (i and ii)** corresponds to the quality of the calibration graphs for the raw Raman spectra data and the pre-treated (SNV) Raman data. The model performance assessments for estimating the phase content in C12A7–CaS–CA samples are given in **Table 3 i)** and for C12A7–CaS–C3A samples they are presented in **Table 3 ii)**. **Table 3** provides the evaluated RMSEC values ranging between 6.14–14.94 (wt%) and RMSECV values between 9.09–19.50 (wt%) for the raw Raman spectral data. The PLS regression model analysis carried out based on the SNV Raman data had narrower values that ranged between 1.19–2.87 (wt%) for the RMSEC values and varied between 2.50–5.16 (wt%) for the RMSECV values, which demonstrates that the deviation in the model accuracy was relatively sound compared to the raw Raman data analysed.

Furthermore, the external validation carried out for the calibration model, demonstrated in **Table 3 (i and ii)** that the values for the coefficient determination of prediction (R^2_{pred}) and RMSEP associated with SNV performed much better than Raw Raman spectral data for analysing the individual phase content in C12A7–CaS–CA and C12A7–CaS–C3A samples. Therefore, the overall model performance based on these parameters for the phase quantification using the SNV method was better than the raw Raman spectral data using the PLS regression model.

Table 3. Parameters and figures of merit for calibration and validation based on the PLS models for i) C12A7–CaS–CA and ii) C12A7–CaS–C3A

	Phase	Method	PLS factor	R^2_{CAL}	R^2_{CV}	$R^2_{\text{PRED.}}$	RMSEC, (wt%)	RMSECV, (wt%)	RMSEP, (wt%)
i) C12A7–CaS–CA	C12A7	Raw Data	7	0.72	0.49	0.55	13.13	18.25	16.85
		SNV	5	0.99	0.97	0.99	2.49	3.75	2.48
	CA	Raw Data	7	0.58	0.33	0.49	14.94	19.50	16.75
		SNV	5	0.98	0.95	0.97	2.87	5.16	3.40
	CaS	Raw Data	7	0.72	0.59	0.64	13.88	17.03	15.78
		SNV	5	0.99	0.96	0.99	2.64	4.62	2.63

ii) C12A7–CaS–C3A	C12A7	Raw Data	7	0.94	0.87	0.87	6.14	9.09	8.75
		SNV	5	0.99	0.98	0.99	1.47	2.90	2.37
	C3A	Raw Data	7	0.93	0.83	0.71	6.18	11.46	13.17
		SNV	5	0.99	0.98	0.99	1.19	2.50	2.91
	CaS	Raw Data	7	0.84	0.78	0.78	9.06	12.13	10.69
		SNV	7	0.99	0.98	0.99	1.45	2.91	1.98

Figure 3. Comparison between weighted and estimated phase contents of ternary mixtures using PLS regression, where *i*) is the C12A7 content in C12A7–C3A–CaS and *iii*) is the C3A content in C12A7–C3A–CaS analysed using the raw Raman data, *ii*) is the C12A7 content in C12A7–C3A–CaS and *iv*) is the C3A content in C12A7–C3A–CaS using SNV Raman spectra data.

3.3.2. (CaO–Al₂O₃)–MgO–Al₂O₃–CaS system

Table 4 (a-c) show the figures of merit for the model assessment. **Table 4 a)** shows the figures for MA–C12A7–CaS, **Table 4 b)** shows the figures for MA–C3A–CaS and **Table 4 c)** shows

the figures for MA–CA–CaS. It can be observed from the estimated values in **Table 4 (a-c)** that the (R^2_{CAL} and R^2_{CV}), RMSEC and RMSECV values for the SNV were better than the raw Raman data in terms of evaluating the individual phase content in the MA–C12A7–CaS, MA–C3A–CaS and MA–CA–CaS ternary samples.

For example, the MA phase content based on **Table 4 (a-c)** could be evaluated in all three sets of samples prepared from a (CaO–Al₂O₃)–MgO.Al₂O₃–CaS system. All the R^2 values for the raw Raman spectral data ranged between 0.70–0.97 and the Raman data pre-processed with SNV had R^2 values between 0.93–0.99 when quantifying the MA content from the samples prepared of the (CaO–Al₂O₃)–MgO.Al₂O₃–CaS system. Furthermore, based on the RMSEC (1.23–3.57), RMSECV (3.9–5.96) and RMSEP (2.75–4.66) values, the SNV method performs better compared to the RMSEC (3.10–9.12), RMSECV (6.48–15.69) and RMSEP (7.62–15.13) values obtained from raw Raman spectral analysis for evaluating the MA phase content (wt%).

The analysis results for specific calcium aluminate phases of C12A7, C3A and CA contained in the MA–C12A7–CaS, MA–C3A–CaS and MA–CA–CaS samples generally had a very impressive figure of merit as shown in **Table 4**. All the R^2 (R^2_{CAL} , R^2_{CV} and R^2_{PRED}) values for quantifying the individual phases of C12A7, C3A and CA were estimated to range between 0.78–0.98 for the raw Raman data and 0.87–0.99 for the SNV treated Raman data used for the PLS regression model. The root mean standard error (RMSE) for the calibration, internal and external validation also varied between 1.57–13.90 for the raw Raman data and for the SNV Raman data between 1.14–7.09. The raw Raman data had a broader range of values for the root mean standard error which was a less accurate model analysis compared to the SNV Raman data. In evaluating the performance for individual calcium aluminates phases in the (CaO–Al₂O₃)–MgO.Al₂O₃–CaS system, the C12A7 and C3A components had a better figure of merit which is shown in **Table 4** compared to CA, particularly for the raw Raman data. This phenomenon associated with the CA values could be related to the potentially high variable

collinearity of the Raman spectra data ^[31]. However, this effect was mitigated when the raw data was pre-processed using the SNV method.

For estimating the CaS phase content in the sample mixtures prepared from the (CaO–Al₂O₃)–MgO–Al₂O₃–CaS system, the analysed data shows that all three sets of samples (MA–C12A7–CaS, MA–C3A–CaS and MA–CA–CaS) provided similar figures of merit for both the raw and SNV Raman data. For instance, MA–C12A7–CaS (**Table 4 a**) had the values $R^2_{\text{CAL}} = 0.98$ (for raw Raman data), $R^2_{\text{CAL}} = 0.99$ (for SNV Raman data) and $R^2_{\text{CV}} = 0.95$ (for raw Raman data) and $R^2_{\text{CV}} = 0.95$ (for SNV Raman data). Additionally, the PLS regression model performed on the raw, and SNV Raman data evaluated respective values such as the R^2 , and root mean standard error features (RMSEC, RMSECV and RMSEP) very close to each other. For example, sample mixtures of MA–C12A7–CaS (**Table 4a**) had values such as $R^2_{\text{CAL}} = 0.98$ and MA–C3A–CaS (**Table 4b**) had $R^2_{\text{CAL}} = 0.97$ for the raw Raman data, and for SNV Raman data the values were $R^2_{\text{CAL}} = 0.99$ for MA–C12A7–CaS and $R^2_{\text{CAL}} = 0.98$ for MA–C3A–CaS.

Table 4. Parameters and figures of merit for calibration and validation based on the PLS models for a) MA–C12A7–CaS, b) MA–C3A–CaS and c) MA–CA–CaS.

	Phase	Method	PLS factor	R^2_{CAL}	R^2_{CV}	R^2_{PRED}	RMSEC, (wt%)	RMSECV, (wt%)	RMSEP, (wt%)
a) MA–C12A7–CaS	MA	Raw Data	7	0.97	0.90	0.84	3.10	6.93	8.11
		SNV	6	0.98	0.96	0.98	1.23	3.9	2.75
	C12A7	Raw Data	7	0.98	0.83	0.90	1.57	6.01	4.35
		SNV	6	0.99	0.89	0.99	1.14	4.80	1.14
	CaS	Raw Data	7	0.98	0.95	0.88	2.58	5.70	8.14
		SNV	6	0.99	0.95	0.98	2.25	5.71	3.10
b) MA–C3A–CaS	MA	Raw Data	7	0.97	0.92	0.88	3.34	6.48	7.62
		SNV	6	0.99	0.93	0.95	2.25	5.96	4.66
	CaS	Raw Data	7	0.97	0.83	0.90	2.90	9.01	5.65
		SNV	6	0.99	0.94	0.98	2.27	6.43	2.93
	C3A	Raw Data	7	0.97	0.83	0.91	2.90	8.20	5.64
		SNV	6	0.98	0.87	0.95	2.38	7.57	4.04
c) MA–CA–CaS	MA	Raw Data	7	0.89	0.70	0.71	9.12	15.69	15.13
		SNV	6	0.98	0.97	0.97	3.57	5.09	4.58
	CA	Raw Data	7	0.83	0.78	0.83	11.58	13.41	13.90
		SNV	6	0.99	0.94	0.99	3.24	7.09	4.64

CaS	Raw Data	7	0.81	0.58	0.67	11.68	17.87	15.70
	SNV	6	0.98	0.93	0.96	3.24	6.87	4.89

Figure 4 shows the plot of reference against the predicted values of calibration and cross-validation for specific component estimation in the ternary system. In **Figure 4**, label **i** is for the raw Raman data and **ii** is for the SNV Raman data, and the calibration curve estimates the MA content in the MA–C12A7–CaS. Additionally, in **Figure 4**, the CaS content is evaluated in MA–CA–CaS based on the calibration curve. **Figure 4 iii)** shows the Raw Raman data and **iv** shows the SNV Raman data.

Figure 4. Comparison between weighted and estimated phase content of ternary mixtures using PLS regression, where (i) is for analysis based the raw Raman data and (ii) is for analysis using SNV Raman spectra data for MA content in MA–CA–CaS; (iii) is for analysis using the raw Raman data and (iv) is for analysis using SNV Raman spectra data for CaS content in MA–CA–CaS samples.

Figures 3 and 4 show how the scattering effect in the calibration model between raw and pre-processed Raman data can be observed. Consequently, based on the figures of merit in **Tables 3–4** and **Figures 3 and 4**, the use of a pre-processing method such as SNV gives enhanced results which can be used to estimate constituents found in the ternary samples of C12A7–CaS–C3A, C12A7–CaS–CA, MA–C12A7–CaS, MA–C3A–CaS and MA–CA–CaS.

The prediction performance for the calibration model was further assessed using statistical parameters such as the ratio of prediction to deviation (RPD) and range error ratio (RER). **Table 5** shows the constituents or the individual phase performances for the ternary samples based on the RPD and RER. **Figure 5 (i – x)** is a graphical representation of the RPD and RER values for five sets of ternary samples demonstrating the performance of estimating each phase present in the sample. **Figure 5 (i and ii)** denotes C12A7–CaS–CA, **Figure 5 (iii – iv)** for MA–CA–CaS and **Figure 5 (v and vi)** for C12A7–CaS–C3A sample set. It can be observed that the RPD and RER values are below the satisfactory mark for the raw Raman data. **Figure 5 (vii and viii)** associated with MA–C12A7–CaS and **Figure 5 (ix and x)** for MA–C3A–CaS generally had a satisfactory score for the raw Raman data. However, for the SNV Raman data, predicting the individual constituent that makes up the samples gave had good RPD and RER values, for all the five sample sets as shown in **Figure 5** and, this corresponds to the merits of the pre-processing spectra data before analysis model using a PLS regression. This could be attributable to the pre-processing technique (SNV) used and for an improved Raman spectra data analysis, as well as correcting the potential baseline shift and background effect^[31, 34].

Table 5. Figures of merit estimated from the calibration model using RDP and RER.

Ternary samples	Phase	RDP		RER	
		Raw	SNV	Raw	SNV
C12A7–CaS–CA	C12A7	1.50	10.16	4.75	32.25
	CaS	1.70	10.17	5.07	30.42
	CA	1.40	6.89	4.18	20.59
MA–C12A7–CaS	MA	2.64	7.79	9.86	29.09
	C12A7	3.29	7.79	18.39	44.44
	CaS	3.06	8.03	9.83	25.80
C12A7–CaS–C3A	C12A7	2.91	10.73	9.14	33.76
	CaS	2.19	11.85	7.48	40.40
	C3A	1.88	8.54	6.07	27.49
MA–C3A–CaS	MA	4.95	4.93	14.16	15.02
	C3A	3.38	4.72	12.41	17.32
	CaS	4.95	9.55	14.16	27.30
MA–CA–CaS	MA	1.88	6.20	5.95	19.65
	CA	2.07	6.16	6.48	19.40
	CaS	1.78	5.66	5.7	18.41

Figure 5. Graphical illustration for the performance of raw Raman data (black columns) and SNV Raman data (grey columns) using RDP and RER; where i–ii denotes C12A7–CaS–CA, iii–iv for MA–CA–CaS, v–vi are for C12A7–CaS–C3A, vii–viii for MA–C12A7–CaS and ix–x for MA–C3A–CaS samples.

As explained in **section 2.4.**, if the RPD value evaluated exceeds three (3) or the estimated RER exceeds ten, then the calibration model can be considered successful in predicting the constituent content used for preparing the sample. Consequently, it can be concluded that based on the calibration model and the performance parameters, the individual phases present in the ternary synthetic inclusion system can be estimated using Raman spectroscopy and give very good results when the Raman spectra data is pre-processed. However, It is important to note the accuracy for estimating the phase content of C12A7, C3A, CA, CaS and $\text{MgO} \cdot \text{Al}_2\text{O}_3$ had weight percent deviations that ranged between 1.5 – 5 wt% for Raman data pre-processed using SNV. For the raw Raman data, the individual phase content could be estimated from the sample mixture with weight percent deviations between 3 – 10 wt%.

3.4. Potential Limitations of the Study

Potential limitations of this work associated with the use of Raman spectroscopy for characterising the samples could cause some slight instrumental variation such as incident laser power and frequency of scattered radiation could potentially have an effect on the intensity of the Raman band. These demerits associated with Raman spectroscopy could be attributed relatively not very satisfactory results obtained from the calibration model in estimating the phase content in the sample mixtures using the raw Raman data and will require pre-processed for enhanced results. It is also important to note that Raman spectroscopic characterisation is generally surface analysis technique, consequently provides an estimation of areal distribution of phases and could be limited in extrapolating analysis results to evaluate the bulk composition or volumetric distribution of phases present in the inclusion based on the sample depth.

Additionally, inclusions found in steel matrix could potentially heterogenous compared synthetic inclusion sample mixtures that are well mixed and homogenous used for this study, therefore could affect the result obtained from the Raman measurement. However, the issue of sample heterogeneity could be mitigated with the used of relatively a wide laser spot size and rotating sample holder stage for a wider surface area for Raman spectroscopy measurements.

3.5. Prospect for practical application

The routes concerning the formation and evolution of inclusions such as $\text{Al}_2\text{O}_3 \rightarrow \text{MgO} \cdot \text{Al}_2\text{O}_3$, $(\text{MA}) \rightarrow \text{MA} - x\text{CaO} \cdot y\text{Al}_2\text{O}_3 \rightarrow \text{MA} - \text{CaS} - \text{CaO} \cdot \text{Al}_2\text{O}_3$ had been demonstrated by various researchers^[4, 5, 8, 12, 21,] as a potential pattern associated with Al-killed calcium treated steels products. This study shows the application of Raman spectroscopy to identify and estimate the

specific phases present in $\text{CaO-Al}_2\text{O}_3\text{-CaS}$ and $\text{CaO-Al}_2\text{O}_3\text{-MgO-CaS}$ systems using a multivariate technique.

The predictability of the individual phase contents (such as C12A7, CA, CaS, MA or C3A) in ternary samples was satisfactory for raw Raman spectral data and the ratio of prediction to deviation (RPD) was on average estimated to be close to 3, while the range error ratio (RER) varied between 5–18. It was also noted that the use of a pre-processing method, such as the standard normal variate (SNV) method performed better. Furthermore, the method used in this study has shown to be very promising for quantifying synthetic phases as the coefficient determination ranged between 0.93–0.99 and with both RPD and RER showing high values. This is an indication that the calibration model used for the study is fit for purpose based on the figures of merit and as recommended by other researchers^[27, 38, 39].

This study demonstrates the prospect of Raman spectroscopy to identify and quantify inclusion types of calcium aluminate, (CA, C12A7 and C3A), CaS and $\text{MgO.Al}_2\text{O}_3$ in steel samples since the steel matrix is Raman-inactive. Consequently, combining previous studies^[14, 17, 19, 23] and this study presents a potential reference database and is applicable for future inclusion characterisation of steel samples using Raman spectroscopy. Furthermore, future studies could consider the application of Raman spectroscopy for characterising both laboratory and industrial steel samples, since typical phases in non-metallic inclusions have been demonstrated to be Raman active materials.

4.0. Conclusion

This work demonstrates Raman spectroscopy, combined with PLS regression, as a suitable technique to quantify five different sets of ternary samples consisting of synthetic phases of CA, C12A7, C3A, CaS and $\text{MgO.Al}_2\text{O}_3$. The findings of this study are summarised as follows:

- Pre-processed Raman spectral data using the SNV method gave the best figures of merit when estimating the phase contents compared to the raw Raman spectra data.

- The calibration model can be used to predict individual phases with good statistical performance regarding the RMSECV, RMSEP, R^2 , RPD and RER values.
- This study, combined with previous studies presents the prospect of characterising $x\text{CaO}\cdot y\text{Al}_2\text{O}_3\text{--MgO}\cdot\text{Al}_2\text{O}_3\text{--CaS}$ inclusions in steel samples.

Acknowledgement

The authors for this work wish to acknowledge the support of the I4Future doctoral programme funded by the European Union's H2020 under the Marie Skłodowska-Curie grant agreement No. 713606 and the Academy of Finland Proliferation project (Academy of Finland, No. 311934).

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Z. Y. Xu, J. H. Liu, Z. J. He, and Q. H. Pang, *Metalurgija*, **2018**, 57, 79.
- [2] T. Lis: *Metalugija*, **2009**, 48, 95.
- [3] J. C. S. Pires and A. Garcia, *Rem Rev. Esc. Minas*, **2004**, 57, 183.
- [4] V. Gollapalli, M. B. Venkata Rao, P. S. Karamched, C. R. Borra, G. G. Roy and P. Srirangam, *Ironmak. Steelmak*, **2018**, 46, 663.
- [5] W. Yang, L. Zhang, X. Wang, Y. Ren, X. Liu and Q. Shan, *ISIJ Int.*, **2013**, 53, 1401.
- [6] Z. Deng and M. Zhu, *ISIJ Int.*, **2013**, 53, 450.
- [7] M. Jiang, X. Wang, B. Chen, and W. Wang, *ISIJ Int.*, **2010** 50, 95.
- [8] H. Y. Tang, Y. Wang, T. Wu, J. S. Li, and S. F. Yang, *Ironmak. Steelmak*, **2017**, 44, 377.
- [9] T. Zhang, Y. Min, C. Liu, and M. Jiang, *ISIJ Int.*, **2015**, 55, 1541.
- [10] S-f. Yang, J-s. Li, Z-f. Wang, J. Li, and L. Lin, *Int. Min. Met. & Mat*, **2011**, 18, 18.

- [11] M. Thunman and D. Sichen *Steel Res. Int.*, **2008**, 79, 124.
- [12] D. Zhao, H. Li, C. Bao and J. Yang, *ISIJ Int.*, **2015**, 55, 2115.
- [13] L. Zhang and B. G. Thomas, *ISIJ Int.* **2003**, 43, 271–291.
- [14] F. Gyakwaa, M. Aula, T. Alatarvas, T. Vuolio, M. Huttula, and T. Fabritius, *ISIJ Int.*, **2019**, 59, 1846.
- [15] D. R. Neuville, G. S. Henderson, L. Cormier, and D. Massiot, *Am. Mineral.* **2010**, 95, 1580.
- [16] D. Torrns-Martn, L. Fernndez-Carrasco, S. Martnez-Ramrez, J. Ibez, L. Arts, and T. Matschei, *J. Am. Ceram. Soc.* **2013**, 3589.
- [17] F. Gyakwaa, M. Aula, T. Alatarvas, T. Vuolio, Q. Shu, M. Huttula, and T. Fabritius *ISIJ Int.*, **2020**, 60 988.
- [18] C. Avril et al., C. Avril, V. Malavergne, R. Caracas, B. Zanda, B. Reynard, E. Charon, E. Bobocioiu, F. Brunet, S. Borensztajn, S. Pont, M. Tarrida, and F. Guyot *Meteor Planet Sci.* **2013**, 48, 1415.
- [19] F. Gyakwaa, M. Aula, T. Alatarvas, T. Vuolio, Q. Shu, M. Huttula, and T. Fabritius, *Appl. Sci.*, **2020**, 10, 6.
- [20] S. Li and L. H. Hihara, *J. Raman Spectrosc.* **2017**, 48, 137.
- [21] G. Jing, C. Shu-sen, and C. Zi-jian, *ISIJ Int.*, **2013**, 53, 2142.
- [22] J. M. A. Geldenhuis and P. C. Pistorius, *Ironmak. Steelmak.* **2000**, 27, 442.
- [23] F. Gyakwaa, M. Aula, T. Vuolio, T. Alatarvas, Q. Shu, M. Huttula, and T. Fabritius, *Steel Res. Int.*, **2020** 2000084,
- [24] J. R. Ferraro, K. Nakamoto and C. W. Brown: *Introductory Raman Spectroscopy*, 2nd ed., Elsevier, Cambridge, **2003**, 15.
- [25] H. Park, J. Y. Park, G. H. Kim, and I. Sohn, *Steel Res. Int.*, **2012**, 83, 150.
- [26] Y. Mohassab and H. Y. Sohn, *Steel Res. Int.*, **2015**, 86, 740.

- [27] C. A. F. de Oliveira Penido, M. T. T. Pacheco, E. H. Novotny, I. K. Lednev, and L. Silveira, *J. Raman Spectrosc.*, **2017**, 48, 1732.
- [28] P. Asadena, C. Alifornia: *Appl. Spectrosc.*, **2003**, 57, 20A.
- [29] T. Lipiäinen, J. Pessi, P. Movahedi, J. Koivistoinen, L. Kurki, M. Tenhunen, J. Yliruusi, A. M. Juppo, J. Heikkonen, T. Pahikkala and C. J. Strachan: *Anal. Chem.* **2018**, 90, 4832.
- [30] Grinberg, N., Rodriguez, S. *Ewing's Analytical Instrumentation Handbook*, Fourth Edition. **2019**, Boca Raton: CRC Press.
- [31] S. Wold, A. Ruhe, H. Wold, and W. Dunn, *SIAM J. -Sci. and Stat. Comput*, 1984, 5 735.
- [32] N. Chieng, S. Rehder, D. Saville, T. Rades, and J. Aaltonen, *J. Pharm. Biomed. Anal.*, **2009**, 49, 18.
- [33] M. K. D. Rambo, M. M. C. Ferreira, P. M. de Melo, C. C. Santana Junior, D. A. Bertuol, and M. C. D. Rambo, *Food Sci. Technol.*, **2020**, 40, 444.
- [34] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, *Appl. Spectrosc.*, **1989**, 43, 772.
- [35] M. R. de Almeida, D. N. Correa, W. F. C. Rocha, F. J. O. Scafi, and R. J. Poppi, **2013**, *Microchem. J.*, 109, 170.
- [36] A. C. Teixeira dos Santosa, R. N.M.J. Páscoaa, P. A.L.S. Portob, A. L. Cerdeirab, J.M. González-Sáizc, C. Pizarroc, J. A. Lopes, *Talanta*, **2018**, 186, 306.
- [37] Kachrimanis K., Braun D. E., Griesser U. J. *J. Pharm and Biomed Anal.* **2007** 43, 407–412.
- [38] A. G. Quentin, T. Rodemann, M. F. Doutreleau, M. Moreau, and N. W. Davies, *Tree Physiol.*, **2017**, 37, 131.
- [39] Z. Yang, K. Li, M. Zhang, D. Xin, and J. Zhang, *Biotechnol. Biofuels*, **2016**, 9, 1.
- [40] S. Dash, R. K. Sahoo, A. Das, S. Bajpai, D. Debasish, and S. K. Singh, *J. Alloys Compd.*, **2017**, 726, 1186.

List of Tables

Table 1. The sample matrix used for this study.....	8
Table 2. Measurement Raman bands for the synthetic phase used for this study.	13
Table 3. Parameters and figures of merit for calibration and validation based on the PLS models for i) C12A7–CaS–CA and ii) C12A7–CaS–C3A.....	17
Table 4. Parameters and figures of merit for calibration and validation based on the PLS models for a) MA–C12A7–CaS, b) MA–C3A–CaS and c) MA–CA–CaS.....	20
Table 5. Figures of merit estimated from the calibration model using RDP and RER.	23

Table of Figures

Figure 1. Schematic chart illustrating how this current study is associated with the authors' previous studies and basic inclusion formation classification.	5
Figure 2. Raman spectra for i) C12A7–CaS–C3A and ii) MA–C12A7–CaS ternary samples. The phase composition for C12A7–CaS–C3A samples are 70C12A7–10CaS–20C3A for a), 10C12A7–70CaS–20C3A for b), 20C12A7–10CaS–70C3A for c), and for MA–C12A7–CaS samples; 70MA–20C12A7–10CaS for d, 35MA–45C12A7–20CaS for e, and 45MA–10C12A7–45CaS for f phase composition.....	15
Figure 3. Comparison between weighted and estimated phase contents of ternary mixtures using PLS regression, where i) is the C12A7 content in C12A7–C3A–CaS and iii) is the C3A content in C12A7–C3A–CaS analysed using the raw Raman data, ii) is the C12A7 content in C12A7–C3A–CaS and iv) is the C3A content in C12A7–C3A–CaS using SNV Raman spectra data.	18

Figure 4. Comparison between weighted and estimated phase content of ternary mixtures using PLS regression, where (i) is for analysis based the raw Raman data and (ii) is for analysis using SNV Raman spectra data for MA content in MA–CA–CaS; (iii) is for analysis using the raw Raman data and (iv) is for analysis using SNV Raman spectra data for CaS content in MA–CA–CaS samples..... 21

Figure 5. Graphical illustration for the performance of raw Raman data (black columns) and SNV Raman data (grey columns) using RDP and RER; where i–ii denotes C12A7–CaS–CA, iii–iv for MA–CA–CaS, v–vi are for C12A7–CaS–C3A, vii–viii for MA–C12A7–CaS and ix–x for MA–C3A–CaS samples..... 24