17 May 2024

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Quantifying API polymorphs in formulations using X-ray powder diffraction and multivariate standard addition method combined with net analyte signal analysis

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*
Quantifying API polymorphs in formulations using X-ray powder diffraction and multivariate standard addition method combined with net analyte signal analysis / Zappi, Alessandro; Maini, Lucia; Galimberti, Giuliano; Caliandro, Rocco; Melucci, Dora*. - In: EUROPEAN JOURNAL OF PHARMACEUTICAL SCIENCES. - ISSN 0928-0987. - STAMPA. - 130:(2019), pp. 36-43. [10.1016/j.ejps.2019.01.014]

This version is available at: https://hdl.handle.net/11585/661828 since: 2022-02-15

*Published:*

DOI: http://doi.org/10.1016/j.ejps.2019.01.014

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

(Article begins on next page)

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

Accepted Manuscript


# Quantifying API polymorphs in formulations using X-ray powder diffraction and multivariate standard addition method combined with net analyte signal analysis

Alessandro Zappi, Lucia Maini, Giuliano Galimberti, Rocco Caliandro, Dora Melucci

PII: S0928-0987(19)30022-3
DOI: https://doi.org/10.1016/j.ejps.2019.01.014
Reference: PHASCI 4805

To appear in: *European Journal of Pharmaceutical Sciences*

Received date: 5 October 2018
Revised date: 21 December 2018
Accepted date: 9 January 2019

Please cite this article as: Alessandro Zappi, Lucia Maini, Giuliano Galimberti, Rocco Caliandro, Dora Melucci , Quantifying API polymorphs in formulations using X-ray powder diffraction and multivariate standard addition method combined with net analyte signal analysis. Phasci (2019), https://doi.org/10.1016/j.ejps.2019.01.014

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

# Quantifying API polymorphs in formulations using X-ray powder diffraction and multivariate standard addition method combined with net analyte signal analysis

***Article type:*** Original research

***Authors:*** Alessandro Zappi[1], Lucia Maini[1], Giuliano Galimberti[2], Rocco Caliandro[3], Dora Melucci[1,*]

[1] *Department of Chemistry "G. Ciamician", University of Bologna, Bologna, IT*

[2] *Department of Statistical Sciences, University of Bologna, Bologna, IT*

[3] *Institute of Crystallography, CNR, Bari, 70126, IT*

***Corresponding author:*** Dora Melucci

*e-mail address:* dora.melucci@unibo.it

*mailing address:* Via Selmi 2, 40126 Bologna (ITALY)

*telephone number:* +39 340 7268273 (mobile), +39 051 2099530 (office)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

**Abstract**

The direct quantification of Active Pharmaceutical Ingredients in solid formulations is a challenging open issue. A consolidated analytical technique based on X-ray Powder Diffraction is available, being the definitive test for the identification of polymorphs and crystal phases. However, its application for quantitative analysis is hindered by matrix effects: refinement methods (*e.g.* Rietveld method) require a complete knowledge of samples' composition, while univariate calibration methods require the matrix effect to be studied and severely suffer from the co-presence of crystalline and amorphous phases in the sample. Multivariate analysis is the only way to bypass problems affecting refinements procedures and univariate calibration. In particular, the multivariate standard addition method (SAM) is promising; however, it is straightforward only when the analytical blank (matrix devoid of analyte) is available: in that case SAM is applied by simply extrapolating the SAM model to the matrix experimental signal. In this work, the quantitative analysis of polymorphic forms of Active Pharmaceutical Ingredients based on X-ray Powder Diffraction is performed for the first time by a method based on multivariate standard addition method combined with net analyte signal procedure; it allows for reliable quantification of polymorphs of active principles in solid formulations, which are rapidly analyzed without any sample pre-treatment. Two test cases are presented: quantification of two polymorphs of piracetam in binary mixtures (forms II and III), and quantification of paracetamol (form I) in Tachifludec®.

**Keywords:** direct analysis, chemometrics, NAS, XRPD, SAM, RootProf

ACCEPTED MANUSCRIPT

## 1. Introduction

There is an urgent need of highly performing analytical methods to analyze samples without pre-treatment. In the case of Active Pharmaceutical Ingredients (APIs) in solid formulations, the importance of this analytical target is enhanced by the need to protect patents, especially when polymorphs are involved (Hilfiker, 2006). The relative amount of a crystalline form compared to the amorphous fraction (Siddiqui et al., 2014) and the direct analysis of co-amorphous (Beyer et al., 2015) and co-crystals (Soares and Carneiro, 2014) are also very important, being related to solubility and consequent efficacy of solid drug formulations (Caliandro et al., 2013). The elective analytical technique to face all these requirements is X-Ray Diffraction applied to powder samples (XRPD).

In fact, many other solid-state analytical techniques have been applied and compared with XRPD for direct APIs determination in solids: Raman spectroscopy (Chieng et al., 2009), Near Infra-Red spectroscopy (NIR) (Xu et al., 2015), Attenuated Total Reflectance-infrared spectroscopy (ATR) (Hu et al., 2010), Differential Scanning Calorimetry (DSC) (Soares and Carneiro, 2014). However, XRPD has a particular potential for its suitability to study many aspects in a single run: identification and quantification of polymorphs, co-crystals and amorphous.

All the above cited analytical techniques take advantages from multivariate data processing: chemometrics can extract all the useful analytical information from huge amount of data acquired by many spectra or diffractograms. Multivariate analysis is also needed when neither refinement methods nor univariate regression models (Classical Least-Squares, CLS) can solve a strong issue affecting quantitative analysis: the contribution of components co-present with the analyte (matrix) to the analytical signal. This is just the case of XRPD: the consolidated Rietveld method (Madsen and Scarlett,

ACCEPTED MANUSCRIPT

2008) requires the full sample composition to be known, and its accuracy is adversely influenced by the amorphous fraction; when univariate calibration is attempted (Rahman et al., 2015), the analyst must choose one specific diffraction angle to measure the response (peak-area or peak-height): at least one specific diffraction angle is required (no interfering species). Even when specific angles are available, the effect of the matrix on the signal should be controlled: this is not easy when polymorphic and amorphous forms are involved.

Multivariate analysis offers tools for various purposes (Brereton, 2007): exploring data to easily visualize samples and variables; creating and validating multivariate models, relating dependent variables (responses) to independent variables; predicting unknown samples. Responses may be both qualitative (classification models) or quantitative (calibration models).

The main chemometric tools recently applied to direct analysis of APIs in solid formulates are: Principal Components Analysis (PCA), Cluster Analysis (CA), Principal Components Regression (PCR), Partial Least Squares (PLS) regression, Multivariate Curve Resolution (MCR). Examples from recent literature are the following. One of the first scientists who explored this issue (Jørgensen et al., 2006) applied PCA for monitoring dehydration phenomena and phase conversion of crystalline and amorphous lactose. De Oliveira et al. (De Oliveira et al., 2008) compared the Rietveld method with multivariate calibration based on PLS, to perform quantitative phase analysis of ternary inorganic mixtures. Xie et al. (Xie et al., 2008) applied MCR and PLS to Raman, NIR and XRPD for the quantitative determination of three solid-state forms of a pharmaceutical compound (one amorphous, two crystalline). Moore et al. (Moore et al.,

ACCEPTED MANUSCRIPT

2009) compared CLS, PCR and PLS in analyzing quaternary mixtures (containing both crystalline and amorphous components) by XRPD.

A very appealing and powerful development of chemometric tools for quantitative analysis using XRPD is offered by Net Analyte Signal (NAS) analysis (Lorber, 1986). This procedure is here applied in a mode based on the pure-analyte diffractogram: by mathematically combining the pure-analyte signal with diffractograms of standard samples, the matrix contribution to the signal can be separated from analytes signal, and quantification is straightforward. The NAS algorithm consists in projecting the pure-analyte signal onto a new mathematic space, while the interfering signal (due to the matrix) is projected onto a perpendicular independent space. Two recent papers (Moore et al., 2008; Palermo et al., 2012) applied NAS to XRPD data to quantify crystal and amorphous forms in poly-component mixtures; in this case, the total amount of analyte in standard samples was taken as independent variable (interpolation mode).

What is really interesting, however, is to apply NAS in extrapolation mode, by combining NAS with the multivariate standard addition method (MSAM): MSAM allows ignoring the matrix composition, and indeed this is crucial when the matrix is completely unknown or not available. The combination of NAS with multivariate SAM (NASSAM) has been proposes by Hemmateenejad et al. (Hemmateenejad and Yousefinejad, 2009), who showed how Euclidean norms of NAS vectors are plotted against the added concentration to calibrate the quantitative method. NASSAM has also recently been applied in studies about co-crystal and co-amorphous properties to quantify co-formers in solution by UV-Vis Spectroscopy (Keramatnia et al., 2015; Shayanfar et al., 2013; Shayanfar and Jouyban, 2013). NASSAM has been also used for the simultaneous determination of two

ACCEPTED MANUSCRIPT

different drugs in the same solution by spectrofluorimetry and spectrophotometry (Asadpour-Zeynali and Bastami, 2010).

In this work, NASSAM is applied to accurately quantify polymorphs of APIs in formulates by XRPD. At the Authors' knowledge, it is the first time that NAS procedure is used combined to SAM for quantification purposes in solid samples. Results are compared with those obtained by applying an alternative SAM algorithm which does not implement NAS and, for reference, a univariate procedure.

The rationale of the present work is the following.

First, NAS is applied to XRPD diffractograms of binary mixtures containing two polymorphs of piracetam, a drug acting on cognitive diseases (dementia, Alzheimer, depression, *etc*.). Five polymorphs of piracetam are known and their stabilities have been already reported (Fabbiani et al., 2007; Maher et al., 2012; Picciochi et al., 2011) and two of them can be easily prepared and stored at room conditions: the thermodynamic stable form FIII (monoclinic), and the metastable form FII form (triclinic). This API was chosen because it has already been analyzed by a multivariate calibration method based on PLS in interpolation mode (the response is total concentration) (Croker et al., 2012), and its pure polymorphs FIII and FII are easy to be synthetized. The NAS method was applied in extrapolation mode (the response is added concentration) aiming to quantify traces of FII in a FIII matrix (Asadpour-Zeynali and Bastami, 2010; Hemmateenejad and Yousefinejad, 2009). NAS results were compared with those obtained on the same samples by a SAM multivariate and a univariate procedure, both implemented in the program RootProf (Caliandro and Belviso, 2014). Trueness was estimated by comparison to the known amount of standard samples prepared in the laboratory.

ACCEPTED MANUSCRIPT

The good results obtained with piracetam encouraged to face the target of this work: to quantify APIs in a solid commercial formulation, for which the matrix is not completely known. A simple case was chosen: only one polymorph in a commercial drug (Tachifludec®), whose active principle (paracetamol in polymorphic Form I) has been already studied by a multivariate explorative approach (CA and PCA) (Khanmohammadi et al., 2010). Accuracy was estimated both with respect to the amount declared on the label, the one obtained by HPLC analysis, and by comparison to results obtained by the RootProf procedures.

## 2. Materials and methods

### 2.1. Samples

#### 2.1.1 Piracetam

Piracetam FIII was purchased from Sigma Aldrich. Pure FII was prepared by recrystallization from saturated solution in ethanol (300 mg of FIII in 2 ml), on heated plate at 60°C with agitation and successive precipitation at 0°C. The precipitated crystals were separated from the solvent by vacuum filtration, heating at 130°C for 24 h in oven and storing under ambient conditions for 3 days. Purity of FIII and FII were also verified by DSC, to determine the temperature and enthalpies of fusion and solid-solid phase transition.

Pure FIII and FII were analyzed using a Perkin Elmer Pyris Diamond differential scanning calorimeter equipped with a model ULSP 90 intra-cooler, operated as a conventional DSC. Baseline, obtained running an empty sealed aluminum pan, was subtracted from the sample signal. Ground samples were weighed with microbalance (weight range 2-7 mg, ± 0.001 mg), in sealed aluminum pans. FIII and FII were analyzed at heating and

ACCEPTED MANUSCRIPT

cooling rate of 5°C $min^{-1}$, heated from 40 to 160°C, cooled from 160 to 50°C and then heated again from 50 to 160°. Temperature and enthalpy calibration was performed by using high-purity standards n-decane ($T_m$= 243.3 K, $\Delta_m H$= 28.7 kJ $mol^{-1}$), benzene ($T_m$= 278.64 K, $\Delta_m H$= 9.9 kJ $mol^{-1}$), and indium ($T_m$= 429.75 K, $\Delta_m H$= 3.267 kJ $mol^{-1}$). To prepare standard samples, the pure forms FII and FIII were ground individually in an agate mortar with a pestle for 5 min each. The purity of the polymorphs was checked by XRPD after grinding, to assure that no transformation occurred during grinding. Samples containing 300 mg of binary calibration mixtures were prepared with respectively 2, 3, 5, 10, 15, 20 and 25% of FII, with the remaining mass balance given by FIII. Pure FII was also employed as standard sample (NAS calculation). Hence, eight standard samples were used (Table I). The binary samples were prepared by weighing out the appropriate amounts of both polymorphs, and then mixing them with a Vortex mixer (geometric dilution method). An analytical balance (± 0.1 mg) was employed. These ranges of FII concentrations were chosen with the aim to determine low levels of polymorphs contamination in API formulations using conventional XRPD sampling system. To ensure homogeneity of the powder samples before XRPD measurements, each mixture was mixed thoroughly using a Vortex mixer.

To perform SAM using the samples in Table I, the one with lower concentration of FII (2%) has been used as the 0% added concentration sample and the added concentration has been consequently calculated (Table I). Although everything is known about these samples, the aim was to simulate a case where quantification by SAM of an impurity of FII in a FIII formulation is sought.

### 2.1.2 Paracetamol

ACCEPTED MANUSCRIPT

Tachifludec® is an antipyretic containing paracetamol, according to the following formulation for each sachet (as reported in the package leaflet): Suchrose (2000 mg), Paracetamol (600 mg), Ascorbic acid (40 mg), Phenylephrine chloridrate (10 mg), excipients (Citric acid, Sodium citrate, Cornstarch, and 8 other excipients; amounts not declared).

Powder Tachifludec® was purchased from Angelini (Ancona, Italy). The whole content of a single drug-sachet was accurately weighted, and the result was 4.1379 g ± 0.0001 g. This value corresponds to 14.5%$_{w/w}$ of paracetamol. Samples containing 300 mg of calibration mixtures were prepared with respectively 0, 15, 25, and 35% of added paracetamol (Sigma Aldrich), with the remaining mass balance given by Tachifludec® and 10% of silicon (used as internal standard) (Sigma Aldrich). Each component was firstly grinded separately with a ball mill Retsch MM 400 (Geass, Turin, Italy) for 30 min, and the purity of paracetamol and silicon was checked by XRPD. Then, samples were prepared by weighing the appropriate amounts of each component, and then mixing them with a Vortex mixer (geometric dilution method). An analytical balance with accuracy of ± 0.1 mg was employed. Samples were grinded again before analysis, using a ball mill.

Standard samples for the determination of piracetam in binary mixtures of Form II-Form III and for the determination of paracetamol in Tachifludec® are reported in Table I and Table II, respectively.

### 2.2. XRPD measurements

The polymorphic standards were collected with a Philips PANalytical X' Pert MPD Pro. This instrument is equipped with PixcelTM detector (active length 3.347°), a Cu Kα

ACCEPTED MANUSCRIPT

source (λ= 1.5418 Å, no monochromator), nickel filter, fixed divergent slit of 1/4°, rad soller 0.04 and accelerating voltage and anode current set as 40 kV and 40 mA, respectively. 100 mg of the same samples were placed on zero background sample holder and sample surface was smoothed with a glass slide. The measurement was carried by spinning the sample holder with 1 s revolution time, step size of 0.026 2θ, range angle 12.8-47.0 2θ, and time step 77 s for Piracetam and range angle 5.0-60.0 2θ, and time step 36 s for Tachifludec. The samples were measured three times; for each measure, the powder was removed, blended and refilled on the sample holder. Reference XRPD patterns for the polymorphs were generated with the CIF files present in the CSD (Groom et al., 2016). with refcode BISMEV01, BISMEV and BISMEV03 for FIII, FII (Admiraal, G., Eikelenboom, J. C., Vos, 1982) and FI (Louër et al., 1995), respectively.

**2.3 HPLC analysis**

In order to confirm the quantity of paracetamol in Tachifludec® and to have a comparison value for the chemometric quantifications, a HPLC analysis was carried out. The quantity of paracetamol reported in the label for each sachet is 600 mg. This was used as the nominal value. An entire sachet of Tachifludec® was solved in 200 mL of water (MilliQ grade). The solution was then filtered and diluted to reach a nominal concentration of paracetamol of 300 ppm. Five standards were prepared at different concentrations (100 to 500 ppm) solving pure paracetamol in MilliQ water. All samples were analyzed three times by HPLC (Agilent Technologies). The stationary phase was a C18 column (Agilent Technologies) and the mobile phase was a mixture 80:20 (v/v) of MilliQ water (buffered with formic acid at pH 3.4) and acetonitrile HPLC grade (Sigma Aldrich), flowing at 1

ACCEPTED MANUSCRIPT

mL $min^{-1}$. The detector was a UV-Vis diode array (Agilent Technologies) set to a wavelength of 244 nm.

### 2.4. Software

The NASSAM procedure was performed by *The Unscrambler* version 10.4 (Camo, Norway) for pre-processing of diffractograms and *R* version 3.2.3 (R Core Team, Vienna, Austria) for NAS calculations.

Alternative quantification procedures were performed by the program RootProf (Caliandro and Belviso, 2014), which is based on the Root framework (Rademakers, 1998). Here a pre-processing step was followed by an original multivariate procedure implementing the SAM protocol, and by a univariate procedure carried out uniquely on samples without any standard addition.

### 2.5. Chemometrics

Data pre-processing strongly affects NAS results, and the best choice, made with the criterion of optimizing the figures of merit ($R^2$ of the standard addition line and limit of detection), was to use the Standard Normal Variate (SNV) transformation, i.e. by subtracting each point of the diffractorgram to its average value and dividing by its standard deviation. NAS calculations were implemented in the *R* environment, according to the literature (Hemmateenejad and Yousefinejad, 2009). The NAS procedure starts from original (or pre-processed) diffractograms, which are the rows ($s_j$) of a matrix *S*. Firstly, a PLS regression is computed, from which scores and loadings (*P*) matrices are obtained. These matrices are used to rebuild the original dataset ($S_{reb}$), by multiplying each other. In this way, by selecting the optimal number (*A*) of principal components (by

ACCEPTED MANUSCRIPT

minimizing Root Mean Square Error), some noise is discarded from the dataset. $S_{reb}$ is then used to calculate that part of the signal due to all components except the *k*-th analyte of interest:

$$S_{-k} = S_{reb} - \alpha c_k s^T \quad (1)$$

where superscript *T* indicates matrix transpose, *s* is the diffractogram of the pure *k*th analyte, $c_k$ is the vector of standard added concentration ($c_{add}$) projected onto the PLS sub-space by $c_k = S_{reb} S_{reb}^{+} c_{add}$ (where + indicates pseudo-inverse), and $\alpha$ is a scalar calculated as $\alpha = 1/(s^T S_{reb} c_k)$.

The NAS vector of the *k*th analyte can be now computed for each sample by finding the orthogonal part of its diffractogram with respect to $S_{-k}$:

$$s_j^* = (I - S_{-k} S_{-k}^{+}) s_j \quad (2)$$

Once obtained $s_j^*$ for all samples, their Euclidean norms can be used as pseudo-univariate signals, used as dependent variable against $c_{add}$ to compute a SAM regression line, from which the concentration of interest can be extrapolated.

The standard deviations of the NAS-extrapolated values were obtained by the jackknife method (Stute, 1996). The NAS procedure was repeated as many times as the number of samples, keeping each time out from the computation one sample and extrapolating the corresponding concentration. The standard deviation of all the jackknife extrapolated values was considered as a good estimation of the overall standard deviation for the extrapolated concentrations. The limit of detection (LoD) for the NAS standard addition line was calculated with reference to literature (Bro and Andersen, 2003; Hemmateenejad and Yousefinejad, 2009). For this computation, three replicates of a blank sample (empty sample holder) were registered, as an estimation of the instrumental noise. These signals were pre-treated by SNV, as it was made for the corresponding samples, and their NAS

ACCEPTED MANUSCRIPT

vector ($\varepsilon$) were computed by eq. (2), using the pre-treated signal as $s_j$. The mean vector ($\bar{\varepsilon}$) of the three ε was calculated and then LoD was estimated by:

$$LoD = 3\frac{\|\bar{\varepsilon}\|}{\|b_A\|} \quad (3)$$

Where $b_A$ is the vector of the regression coefficents for the *A*-th principal component of the starting PLS model and ||•|| indicates the Euclidean norm.

The ratio:

$$Sn = \frac{1}{\|b_A\|} \quad (4)$$

is an alternative way to compute the sensitivity (*Sn*) of a NAS calibration line (Bro and Andersen, 2003), and it is used as a further figure of merit, which has to be similar to the slope of the SAM regression line. The R code for NAS computation is reported as supplementary material.

The approach of RootProf is instead based on whole profile least square fitting. The XRPD profile of a mixture of *N* crystal phases can be written as:

$$y = \sum_{i=1}^{N} c_i f_i + y_b \quad (5)$$

where $y_b$ is the background (comprising the amorphous content), $f_i$ and $c_i$ are the profile and weight fraction of the *i*-th pure crystal phase. The weight fractions should be corrected for the absorption coefficient for X-ray scattering of the corresponding crystal phases, but in first approximation all the coefficients are supposed similar.

The quantitative analysis protocol implemented in RootProf allows estimating the weight fractions of the pure phase components in a mixture, through a whole-profile fitting procedure. The experimental background-subtracted profiles of single pure crystal phases $f_i$ are fitted on the background-subtracted experimental profile of the mixture $y - y_b$, to

ACCEPTED MANUSCRIPT

get the estimates $\hat{c}_i$ of the true $c_i$ weight fractions, according to the following fitting model:

$$y - y_b = \sum_{i=1}^{N} \hat{c}_i f_i \quad (6)$$

The background-subtraction step is performed as an automatic pre-processing of the powder diffraction profiles through the SNIP algorithm (Ryan et al., 1988). This procedure has been adapted to solve the scientific case where a single pure component profile $f_x$ is available and an estimate $\hat{c}_x$ of that component in a complex mixture is wanted (RootProf-SAM). In this case eq. (5) can be written as:

$$y = y_M + c_x f_x + y_b \quad (7)$$

where $f_x$ is the profile of the specific pure crystal phase and the term $y_M = \sum_{i=1}^{N-1} c_i f_i$ describes the matrix effect. According to the SAM method, known quantities $c_{add}$ of pure phase *x* can be added to the mixture, by creating new samples. Their corresponding XRPD profiles will be:

$$y_j = y_M + (c_x + c_{add,j}) f_x + y_b \quad (8)$$

As a first step, a pre-processing constituted by a background subtraction and a SNV rescaling is performed on the pure phase $f_x$ and the mixture $y_j$ experimental profiles.

As a second step, the MultiFit procedure is applied on the rescaled profiles $y'_j - y_b$, taken each one independently, and by considering $f_x$ as unique pure phase. Thus $y'_j = \hat{y}_M + \hat{c}_{add,j} f_x + y_b$ are the best estimates of profiles in eq. (8).

As a third step, the estimated weight fractions $\hat{c}_{add,j}$ are combined to extrapolate the best estimate $\hat{c}_x$ for $c_x$ by a standard addition regression line. Errors on $\hat{c}_{add,j}$ values calculated by the fitting procedure are propagated through the regression procedure to

ACCEPTED MANUSCRIPT

obtain the error on $\hat{c}_x$. For the specific case of RootProf, the SAM calibration plot contains the values $c_{add,j}$ along the *x* axis and their estimates $\hat{c}_{add,j}$ along the *y* axis. Thus two possible estimates $\hat{c}_x$ are available: the intercept of the regression line on the *y* axis and the absolute value of its intercept on the *x* axis. A weighted mean of the two estimates is taken as RootProf best estimate of $c_x$. It is worth noting that the RootProf fitting procedure is also able to obtain a rough estimate $\hat{c}_x$ even without standard additions from eq. (6) (RootProf-No SAM). This univariate determination is expected to be heavily affected by matrix errors, which is reduced by using combined information from SAM. The LoD was obtained by applying the MultiFit procedure to the three replicates of a blank sample and calculating

$$LoD = 3\frac{\sigma}{m}, \qquad (9)$$

where $\sigma$ is the standard deviation of the $\hat{c}_{add,j}$ estimates obtained for the blank samples and $m$ is the slope of the regression line. The sensitivity was estimated as the slope of the regression line, i.e. $Sn = m$. Flow diagrams of RootProf processing and details of RootProf analysis of the two test cases are reported as supplementary material. The pre-processing procedure and 2θ data selection for RootProf has been optimized on the bases of sensitivity and LOD parameter values, as well as by considering the average fit residuals and weight fraction fluctuations around the regression line (see Figure S2 and S4).

## 3. Results and discussion

### 3.1 Piracetam

The samples listed in Table I were processed by XRPD.

ACCEPTED MANUSCRIPT

As an example, a diffractogram obtained with sample 7 is reported in Figure 1a, overlying to the diffractograms of pure-FII and pure-FIII. Figure 1b, instead, shows the evolution of some peaks of FII through the increasing of its concentration in samples.

The diffractometric signals were pre-processed by SNV and they were then processed by NAS. The plot of NAS Euclidean norm vs. added concentration reported in Figure 2 was obtained. The extrapolated concentration with the corresponding standard deviation are reported in table III, while the relevant figures of merit are reported in Table IV. The analytical performance of the regression model is very good: the determination coefficient $R^2$ is close to the ideal unit value, and the root mean square error (RMSE) of regression is low compared to the concentration of FII in the most diluted standard sample ($2\%_{w/w}$). The predictive ability of the model is also good: the comparison between the extrapolated value ($1.73\%_{w/w}$), also considering the relevant standard deviation ($0.241\%_{w/w}$), indicates that good trueness was achieved.

For comparison, the alternative SAM implementation of RootProf was applied on the same data. The SAM calibration plot (Figure 3) shows weight fraction estimates with their errors and the corresponding regression line. The extrapolated concentration is reported in Table III, figures of merit are reported in Table IV. Excellent agreement with the expected FII concentration of piracetam in the reference-standard and with the NASSAM estimate is shown (although the RootProf procedure estimates a larger error). The average concentration estimated by the univariate RootProf procedure, applied on diffraction profiles with no standard addition, is in agreement with SAM determinations, but it has very large errors. It is worth noting that reliable RootProf results could only be achieved by excluding from analysis data in the 2θ range [21°, 22°], where a huge mixture peak in overlap with a pure FII peak strongly bias the whole-profile fitting (Figure 1a).

ACCEPTED MANUSCRIPT

The 2θ selection has been conceived and optimized by considering the plot of average fit residuals, calculated for all processed samples (Figure S4).

### 3.2 Paracetamol

The samples listed in Table II were processed by XRPD.

As an example, a diffractogram obtained with sample 3 is reported in Figure 4a, overlying the pure-paracetamol and some excipients diffractograms. It is evident from Figure 4a that finding peaks specific for the API was a harder task than the case of binary mixtures of piracetam Forms II and III described in the previous section, because most of the paracetamol's peaks overlap with at least one excipient's peak. Moreover, while in the case of piracetam the standard samples for NAS calibration were completely known, being the samples prepared in the laboratory, in the case of Tachifludec® a real sample is managed whose complete composition is unknown: the matrix effect can be bypassed by SAM combined with NAS analysis. Figure 4b, instead, shows the evolution of some peaks of paracetamol through the increasing of its concentration in samples. From figure 4b, it is also interesting to note that the intensity of the excipient's peaks decrease with the increasing of paracetamol concentration (because their relative amount decrease).

The diffractograms were pre-processed by SNV, then the NAS algorithm was applied. Figure 5 reports the calibration plot, which is NAS Euclidean norm vs. added concentration. The extrapolated concentration with the corresponding standard deviation are reported in table III, while the relevant figures of merit are reported in Table IV. The general analytical performance of the calibration model is very good: the determination coefficient $R^2$ is close to unit, and RMSE is low compared to the concentration of paracetamol in the most diluted standard sample (no addition). Indeed, the standard

ACCEPTED MANUSCRIPT

sample to which no pure paracetamol was added had the expected paracetamol concentration reported on the leaflet (14.5%$_{w/w}$). The predictive ability of the model is also good, because the comparison between the extrapolated value (14.8%$_{w/w}$), also considering the relevant standard deviation (1.55%$_{w/w}$), and the expected value indicates that good trueness was achieved.

For comparison with NAS-alternative procedures, the RootProf algorithm was also applied (Figure 6). Despite the great overlap of peaks between paracetamol form I and Tachifludec® diffractograms, in this case no data selection was necessary to carry out reliable whole-profile fittings. The RootProf SAM algorithm finds out nearly the same concentration as the NASSAM method, but with a lower error; the RootProf univariate algorithm (no SAM) gives lower, though compatible, weight fractions (Table III). Converging results obtained by NASSAM and RootProf alghoritms indicate a slight overestimation with respect to the amount declared on the label of the commercial product.

Both the results of NAS and RootProf are also in agreement with the paracetamol concentration obtained by HPLC analysis of Tachifludec®: (13 ± 2)%$_{w/w}$. Although HPLC slightly underestimates the paracetamol concentration, this value is not significantly different from the one reported in Tachifludec® label and from the ones obtained by NASSAM and RootProf procedures. This is a further demonstration of the reliability of the two methods.

## 4. Conclusions

The suitability of the NASSAM method to accurately quantify two polymorphs of piracetam in reference samples prepared in the laboratory by XRPD was demonstrated,

ACCEPTED MANUSCRIPT

thus confirming that NASSAM is a good tool for processing diffractograms and to study solid samples.

The NASSAM method was then successfully applied to a commercial drug (Tachfludec®), in which the API (paracetamol) was quantified by XRPD and NASSAM, without physical or chemical sample pre-treatment. The matrix effect was by-passed by chemometrics, using the pure-API signal in calculating the NAS specifically related to the analyte, regardless of the chemical species co-existing with it in the sample.

The comparison of NASSAM results with those obtained by an alternative procedure based on whole-profile fitting of SAM against pure-phase diffractograms highlighted the advantages of the chemometric NAS method: it overcomes matrix effects, even in cases of strong peak overlap; weight fraction estimates with very small errors are achieved. On the other hand, a univariate RootProf approach can supply fast rough weight fraction estimates even without standard addition, thus minimizing experimental efforts.

This method can be used in the pharmaceutical field for the quantification of any ingredient or any impurity (as was simulated in the piracetam case) in a batch, with a low consumption of the product.

**Acknowledgements**

Authors would like to thank Rossella Gelsi, Eleonora Ciuti, and Marta Colagrande for performing experimental measurements on piracetam, paracetamol, and HPLC respectively. Investigation supported by University of Bologna (Funds for Selected Research Topics).

ACCEPTED MANUSCRIPT

**References**

Admiraal, G., Eikelenboom, J. C., Vos, A., 1982. Structures of the Triclinic and Monoclinic Modifications of. Acta Crystallogr. Sect. B 85, 2600–2605.

Asadpour-Zeynali, K., Bastami, M., 2010. Net analyte signal standard addition method (NASSAM) as a novel spectrofluorimetric and spectrophotometric technique for simultaneous determination, application to assay of melatonin and pyridoxine. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 75, 589–597. https://doi.org/10.1016/j.saa.2009.11.023

Beyer, A., Grohganz, H., L??bmann, K., Rades, T., Leopold, C.S., 2015. Multivariate quantification of the solid state phase composition of co-amorphous naproxen-indomethacin. Molecules 20, 19571–19587. https://doi.org/10.3390/molecules201019571

Brereton, R.G., 2007. Applied Chemometrics for Scientists, Applied Chemometrics for Scientists. https://doi.org/10.1002/9780470057780

Bro, R., Andersen, C.M., 2003. Theory of net analyte signal vectors in inverse regression. J. Chemom. 17, 646–652. https://doi.org/10.1002/cem.832

Caliandro, R., Belviso, D.B., 2014. RootProf: Software for multivariate analysis of unidimensional profiles. J. Appl. Crystallogr. 47, 1087–1096. https://doi.org/10.1107/S1600576714005895

Caliandro, R., Di Profio, G., Nicolotti, O., 2013. Multivariate analysis of quaternary carbamazepine-saccharin mixtures by X-ray diffraction and infrared spectroscopy. J. Pharm. Biomed. Anal. 78–79, 269–279. https://doi.org/10.1016/j.jpba.2013.01.042

ACCEPTED MANUSCRIPT

Chieng, N., Rehder, S., Saville, D., Rades, T., Aaltonen, J., 2009. Quantitative solid-state analysis of three solid forms of ranitidine hydrochloride in ternary mixtures using Raman spectroscopy and X-ray powder diffraction. J Pharm Biomed Anal 49, 18–25. https://doi.org/10.1016/j.jpba.2008.09.054

Croker, D.M., Hennigan, M.C., Maher, A., Hu, Y., Ryder, A.G., Hodnett, B.K., 2012. A comparative study of the use of powder X-ray diffraction, Raman and near infrared spectroscopy for quantification of binary polymorphic mixtures of piracetam. J. Pharm. Biomed. Anal. 63, 80–86. https://doi.org/10.1016/j.jpba.2012.01.013

De Oliveira, T.F., Epprecht, E.K., Ziolli, R.L., Saraiva, A.C.F., De Avillez, R.R., 2008. Multivariate calibration by partial least squares for the quantitative phase analysis using Rietveld method. J. Chemom. 22, 141–148. https://doi.org/10.1002/cem.1128

Fabbiani, F.P.A., Allan, D.R., David, W.I.F., Davidson, A.J., Lennie, A.R., Parsons, S., Pulham, C.R., Warren, J.E., 2007. High-pressure studies of pharmaceuticals: An exploration of the behavior of piracetam. Cryst. Growth Des. 7, 1115–1124. https://doi.org/10.1021/cg0607710

Groom, C.R., Bruno, I.J., Lightfoot, M.P., Ward, S.C., 2016. The Cambridge structural database. Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater. 72, 171–179. https://doi.org/10.1107/S2052520616003954

Hemmateenejad, B., Yousefinejad, S., 2009. Multivariate standard addition method solved by net analyte signal calculation and rank annihilation factor analysis. Anal. Bioanal. Chem. 394, 1965–1975. https://doi.org/10.1007/s00216-009-2870-1

Hilfiker, R., 2006. Polymorphism: In the Pharmaceutical Industry, Polymorphism: In the Pharmaceutical Industry. https://doi.org/10.1002/3527607889

Hu, Y., Erxleben, A., Ryder, A.G., McArdle, P., 2010. Quantitative analysis of

ACCEPTED MANUSCRIPT

sulfathiazole polymorphs in ternary mixtures by attenuated total reflectance infrared, near-infrared and Raman spectroscopy. J. Pharm. Biomed. Anal. 53, 412–420. https://doi.org/10.1016/j.jpba.2010.05.002

Jørgensen, A.C., Miroshnyk, I., Karjalainen, M., Jouppila, K., Siiriä, S., Antikainen, O., Rantanen, J., 2006. Multivariate data analysis as a fast tool in evaluation of solid state phenomena. J. Pharm. Sci. 95, 906–916. https://doi.org/10.1002/jps.20573

Keramatnia, F., Shayanfar, A., Jouyban, A., 2015. Thermodynamic Solubility Profile of Carbamazepine-Cinnamic Acid Cocrystal at Different pH. J. Pharm. Sci. 104, 2559–2565. https://doi.org/10.1002/jps.24525

Khanmohammadi, M., Garmarudi, A.B., Moazzen, N., Ghasemi, K., 2010. Qualitative Discrimination Between Paracetamol Tablets Made by Near Infrared Spectroscopy and Chemometrics With Regard to Polymorphism. J. Struct. Chem. 51, 663–669. https://doi.org/10.1007/s10947-010-0097-9

Lorber, A., 1986. Error Propagation and Figures of Merit for Quantification by Solving Matrix Equations. Anal. Chem. 58, 1167–1172. https://doi.org/10.1021/ac00297a042

Louër, D., Louër, M., Dzyabchenko, V.A., Agafonov, V., Ceolin, R., 1995. Structure of a metastable phase of piracetami from X- ray powder diffraction using the atom–atom potential method. Acta Crystallogr. Sect. B 51, 182–187. https://doi.org/10.1107/S0108768194008840

Madsen, I.C., Scarlett, N.V.Y., 2008. Chapter 11 Quantitative Phase Analysis, in: Powder Diffraction: Theory and Practice. pp. 298–331. https://doi.org/10.1039/9781847558237-00298

Maher, A., Seaton, C.C., Hudson, S., Croker, D.M., Rasmuson, Å.C., Hodnett, B.K.,

ACCEPTED MANUSCRIPT

2012. Investigation of the solid-state polymorphic transformations of piracetam. Cryst. Growth Des. 12, 6223–6233. https://doi.org/10.1021/cg301418c

Moore, M.D., Cogdill, R.P., Short, S.M., Hair, C.R., Wildfong, P.L.D., 2008. The use of net analyte signal orthogonalization in the separation of multi-component diffraction patterns obtained from X-ray powder diffraction of intact compacts. J. Pharm. Biomed. Anal. 47, 238–247. https://doi.org/10.1016/j.jpba.2007.12.042

Moore, M.D., Cogdill, R.P., Wildfong, P.L.D., 2009. Evaluation of chemometric algorithms in quantitative X-ray powder diffraction (XRPD) of intact multi-component consolidated samples. J. Pharm. Biomed. Anal. 49, 619–626. https://doi.org/10.1016/j.jpba.2008.12.007

Palermo, R.N., Short, S.M., Anderson, C.A., Tian, H., Drennen, J.K., 2012. Determination of figures of merit for near-infrared, Raman and powder X-ray diffraction by net analyte signal analysis for a compacted amorphous dispersion with spiked crystallinity. J. Pharm. Innov. 7, 56–68. https://doi.org/10.1007/s12247-012-9127-9

Picciochi, R., Diogo, H.P., Minas Da Piedade, M.E., 2011. Thermodynamic characterization of three polymorphic forms of piracetam. J. Pharm. Sci. 100, 594–603. https://doi.org/10.1002/jps.22294

Rademakers, F., 1998. ROOT: An object-oriented data analysis framework. Linux J. 1–19.

Rahman, Z., Mohammad, A., Siddiqui, A., Khan, M.A., 2015. Comparison of Univariate and Multivariate Models of 13C SSNMR and XRPD Techniques for Quantification of Nimodipine Polymorphs. AAPS PharmSciTech 16, 1368–1376. https://doi.org/10.1208/s12249-015-0327-8

ACCEPTED MANUSCRIPT

Ryan, C.G., Clayton, E., Griffin, W.L., Sie, S.H., Cousens, D.R., 1988. SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. Nucl. Inst. Methods Phys. Res. B 34, 396–402. https://doi.org/10.1016/0168-583X(88)90063-8

Shayanfar, A., Asadpour-Zeynali, K., Jouyban, A., 2013. Solubility and dissolution rate of a carbamazepine-cinnamic acid cocrystal. J. Mol. Liq. 187, 171–176. https://doi.org/10.1016/j.molliq.2013.06.015

Shayanfar, A., Jouyban, A., 2013. Drug-drug coamorphous systems: Characterization and physicochemical properties of coamorphous atorvastatin with carvedilol and glibenclamide. J. Pharm. Innov. 8, 218–228. https://doi.org/10.1007/s12247-013-9162-1

Siddiqui, A., Rahman, Z., Bykadi, S., Khan, M.A., 2014. Chemometric methods for the quantification of crystalline tacrolimus in solid dispersion by powder X-ray diffractrometry. J. Pharm. Sci. 103, 2819–2828. https://doi.org/10.1002/jps.23912

Soares, F.L.F., Carneiro, R.L., 2014. Evaluation of analytical tools and multivariate methods for quantification of co-former crystals in ibuprofen-nicotinamide co-crystals. J. Pharm. Biomed. Anal. 89, 166–175. https://doi.org/10.1016/j.jpba.2013.11.005

Stute, W., 1996. The jackknife estimate of variance of a Kaplan-Meier integral. Ann. Stat. 24, 2679–2704. https://doi.org/10.1093/biomet/81.3.602

Xie, Y., Tao, W., Morrison, H., Chiu, R., Jona, J., Fang, J., Cauchon, N., 2008. Quantitative determination of solid-state forms of a pharmaceutical development compound in drug substance and tablets. Int. J. Pharm. 362, 29–36. https://doi.org/10.1016/j.ijpharm.2008.05.038

ACCEPTED MANUSCRIPT

Xu, K., Zheng, S., Zhai, Y., Guo, L., Tang, P., Yan, J., Wu, D., Li, H., 2015. Two solid forms of tauroursodeoxycholic acid and the effects of milling and storage temperature on solid-state transformations. Int. J. Pharm. 486, 185–194. https://doi.org/10.1016/j.ijpharm.2015.03.072

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

**Table I**. Standard samples for piracetam analysis. Three replicates per sample.

| Sample | FII total concentration (%w/w) | FII added concentration (%w/w) |
|---|---|---|
| 1 | 2 | 0 |
| 2 | 3 | 1 |
| 3 | 5 | 3 |
| 4 | 10 | 8 |
| 5 | 15 | 13 |
| 6 | 20 | 18 |
| 7 | 25 | 23 |
| 8 | 100 | Pure Piracetam FII |

**Table II**. Standard samples for paracetamol analysis. Three replicates per sample.

| Sample | Added Concentration (%w/w) |
|---|---|
| 1 | 0 |
| 2 | 15 |
| 3 | 25 |
| 4 | 35 |
| 5 | Pure Paracetamol |

ACCEPTED MANUSCRIPT

**Table III**. NAS analysis results compared with RootProf results obtained with and without SAM

The expected concentration values are: 2%$_{w/w}$ for piracetam, 14.5 %$_{w/w}$ for paracetamol.

| Case Study | NASSAM (%$_{w/w}$) | RootProf SAM (%$_{w/w}$) | RootProf no SAM (%$_{w/w}$) |
|---|---|---|---|
| **Piracetam** | 1.7 ± 0.2 | 2.1 ± 0.6 | 2.7 ± 2 |
| **Paracetamol** | 15 ± 2 | 15.2 ± 0.1 | 14 ± 1 |

**Table IV.** Figures of merit for NAS and RootProf calibration lines

| | Piracetam | | Paracetamol | |
|---|---|---|---|---|
| | **NAS** | **RootProf** | **NAS** | **RootProf** |
| **Slope** | 0.123 | 1.30 | 0.438 | 0.78 |
| **Slope standard deviation** | 0.002 | 0.08 | 0.0286 | 0.04 |
| **Intercept** | 0.21 | 0.024 | 6.50 | 0.14 |
| **Intercept standard deviation** | 0.04 | 0.01 | 0.604 | 0.01 |
| **$R^2$** | 0.989 | 0.991 | 0.9472 | 0.987 |
| **RMSE** | 0.0966 | 0.0001 | 1.46 | 0.0001 |
| **Sensitivity** | 0.128 **(a)** | 1.30 | 0.675 **(a)** | 0.78 |
| **LoD** | 0.75 | 0.82 | 0.1 | 0.50 |

**(a)** Slight differences between NAS sensitivities and the corresponding line-slopes are due to the different way of calculating them: by eq. (5) for sensitivity and by regression for slope

ACCEPTED MANUSCRIPT

**Captions to figures**

**Figure 1** a) Piracetam diffractograms, relevant to sample 7 of Table I, diffractograms of forms II and III of Piracetam are also reported b) Piracetam diffractograms of one replicate of the samples reported in Table I, zoomed in the 2θ interval 10-24, square root scale used for intensity counts to magnify the variation on the weak peaks; Piracetam form II diffractogram is also reported to show how the corresponding peak intensity increases in samples

**Figure 2** NAS calibration plot for piracetam standards listed in Table I, the vertical red line correspond to the extrapolated concentration

**Figure 3** SAM calibration plot obtained by the RootProf procedure for piracetam standards listed in Table I. Errors calculated from the whole-profile fitting procedure on single mixtures are shown. The regression line is superimposed in red. The RootProf-SAM determination reported in Table III is the weighted mean of the intercept of the regression line on the Y axis and of the absolute value of its intercept on the X axis; the RootProf-no SAM determination is calculated as the weighted mean of the weight fractions lying on the Y axis..

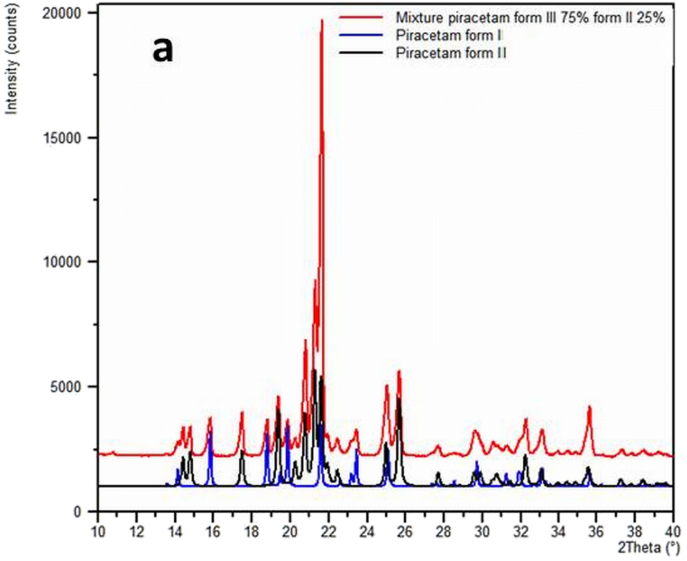
**Figure 4** a) Paracetamol diffractograms, relevant to sample 3 of Table II; diffractograms of some identified excipients of Tachifludec® are also reported b) Paracetamol diffractograms of one replicate of the samples reported in Table II, zoomed in the 2θ interval 10-15, square root scale used for intensity counts to magnify the variation on the weak peaks; Paracetamol form I diffractogram is also reported to show how the corresponding peak intensity increase in samples

**Figure 5** NAS calibration plot for paracetamol standards listed in Table II, the vertical red line correspond to the extrapolated concentration

ACCEPTED MANUSCRIPT

**Figure 6** SAM calibration plot obtained by the RootProf procedure for Piracetam standards listed in Table I. Errors calculated from the whole-profile fitting procedure on single mixtures are shown. The regression line is superimposed in red. The RootProf-SAM determination reported in Table III is the weighted mean of the intercept of the regression line on the Y axis and of the absolute value of its intercept on the X axis; the RootProf-no SAM determination is calculated as the weighted mean of the weight fractions lying on the Y axis.

ACCEPTED MANUSCRIPT


a
Mixture piracetam form III 75% form II 25%
Piracetam form I
Piracetam form II
Intensity (counts)
2Theta (°)

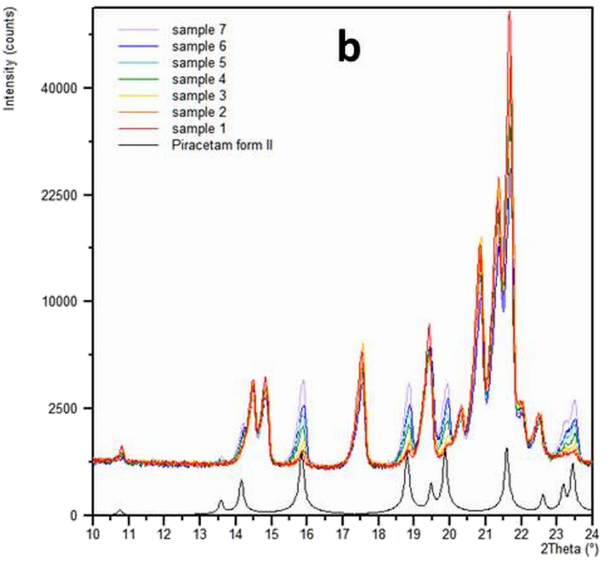

b
sample 7
sample 6
sample 5
sample 4
sample 3
sample 2
sample 1
Piracetam form II
Intensity (counts)
2Theta (°)

Figure 1

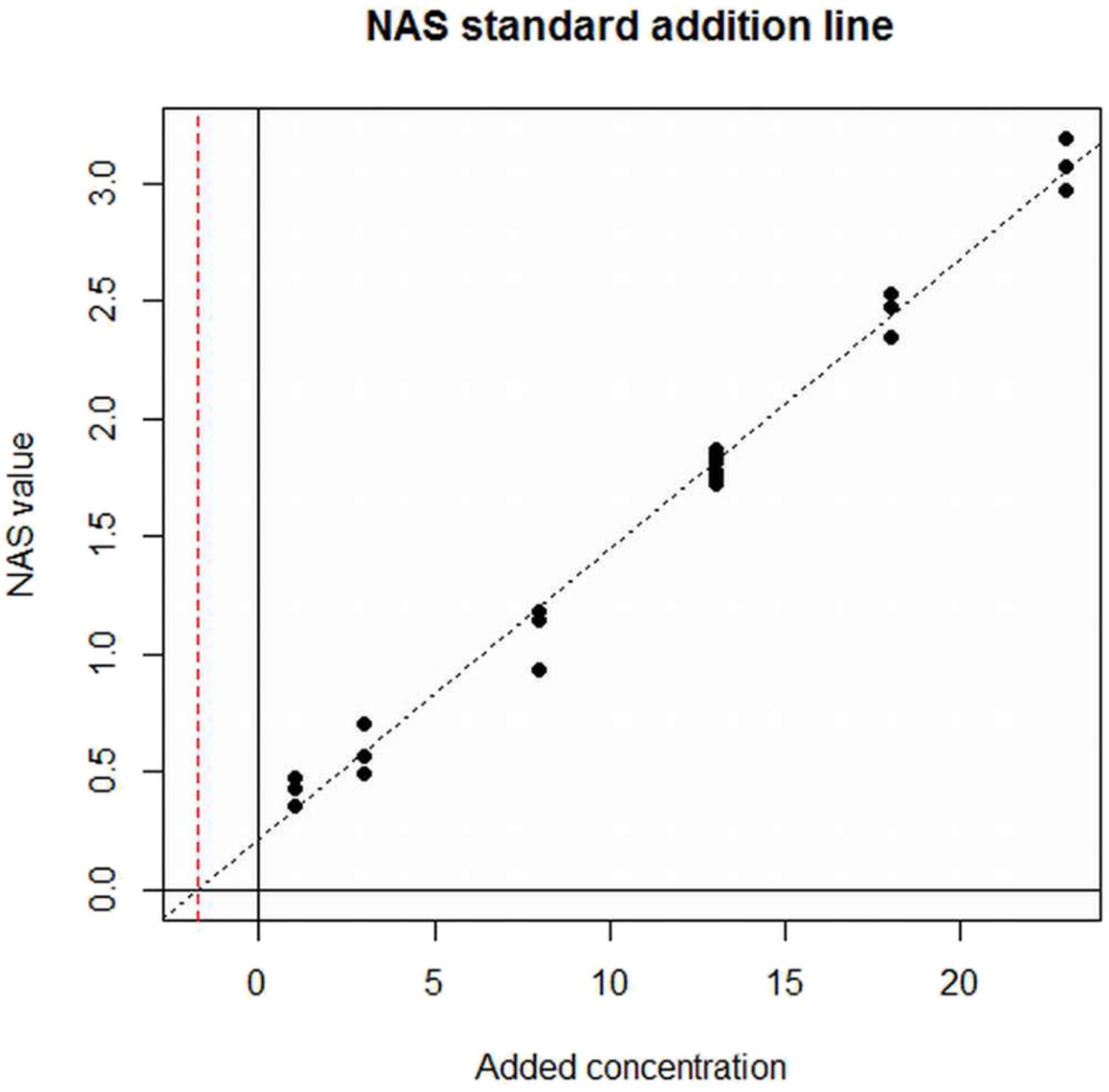

NAS standard addition line
NAS value
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
5
10
15
20
Added concentration

Figure 2

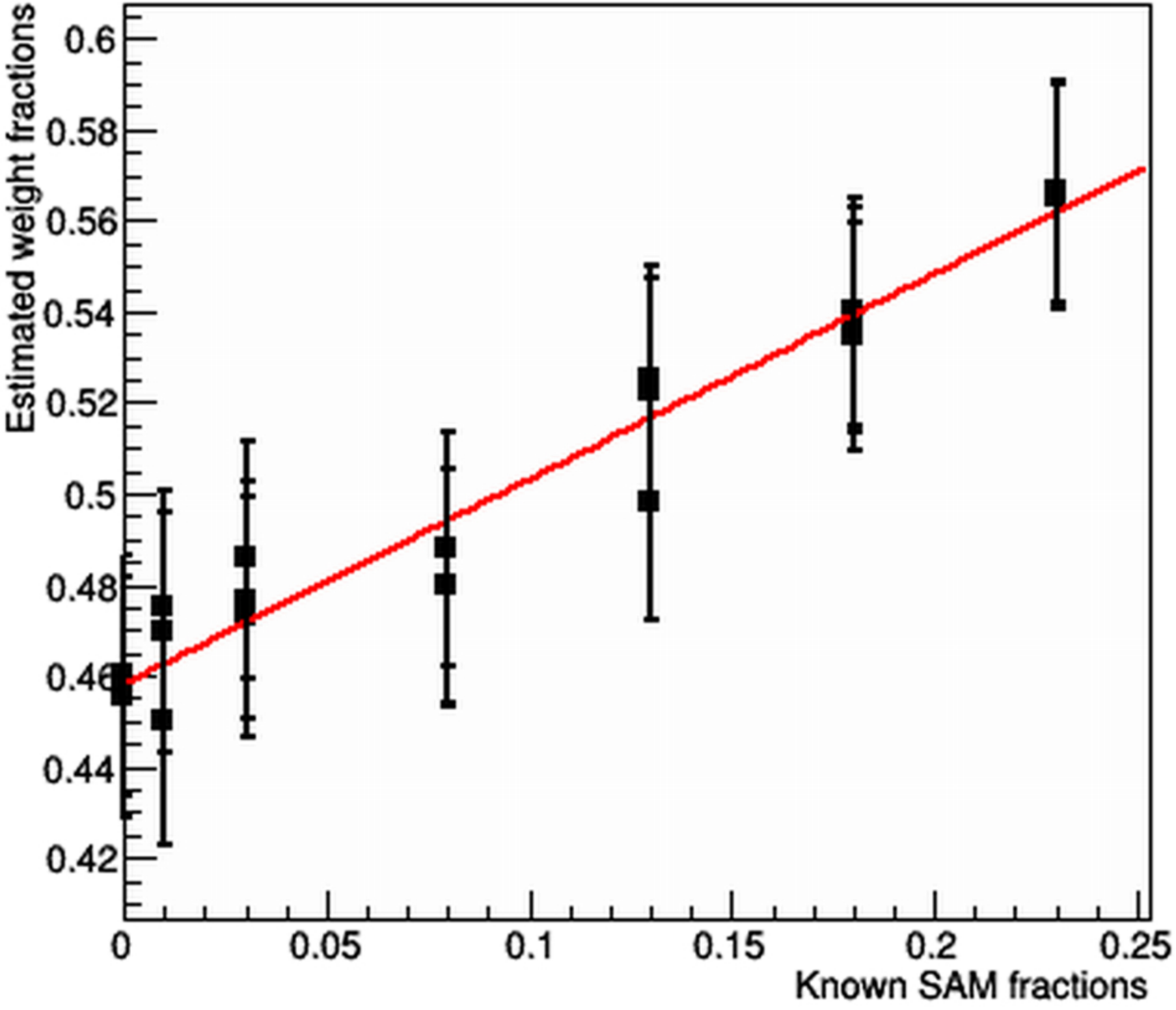
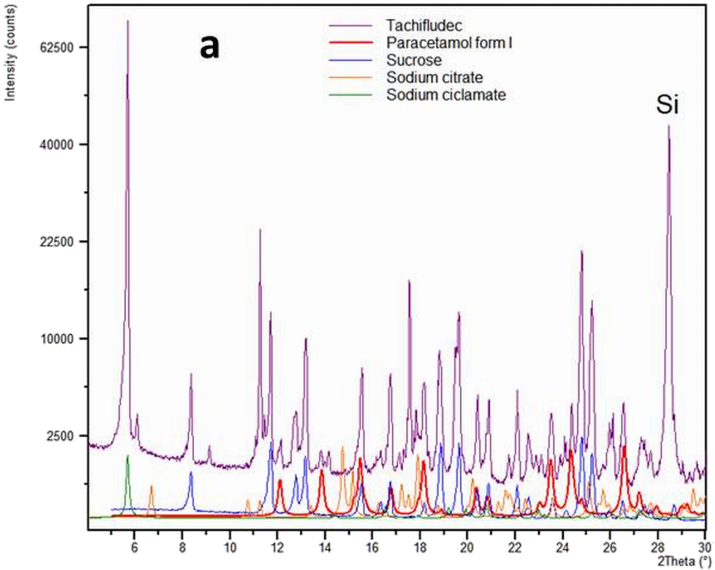

Estimated weight fractions
0.6
0.58
0.56
0.54
0.52
0.5
0.48
0.46
0.44
0.42
0
0.05
0.1
0.15
0.2
0.25
Known SAM fractions

Figure 3


a
Tachifludec
Paracetamol form I
Sucrose
Sodium citrate
Sodium ciclamate
Si
Intensity (counts)
62500
40000
22500
10000
2500
6
8
10
12
14
16
18
20
22
24
26
28
30
2Theta (°)

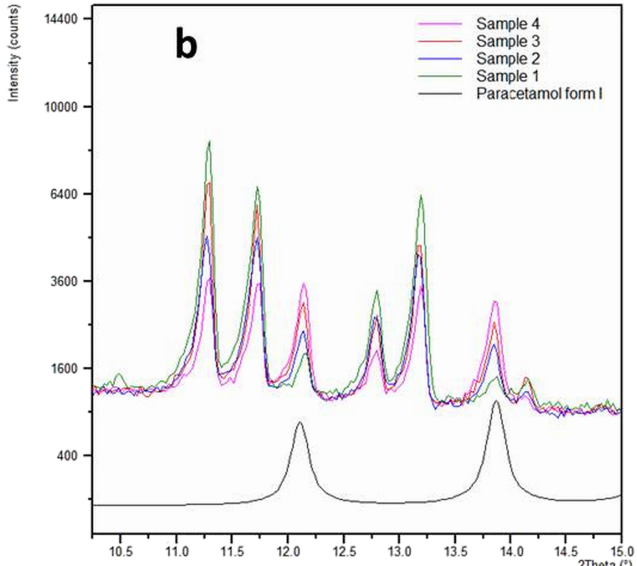

b
Sample 4
Sample 3
Sample 2
Sample 1
Paracetamol form I
Intensity (counts)
14400
10000
6400
3600
1600
400
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
2Theta (°)

Figure 4

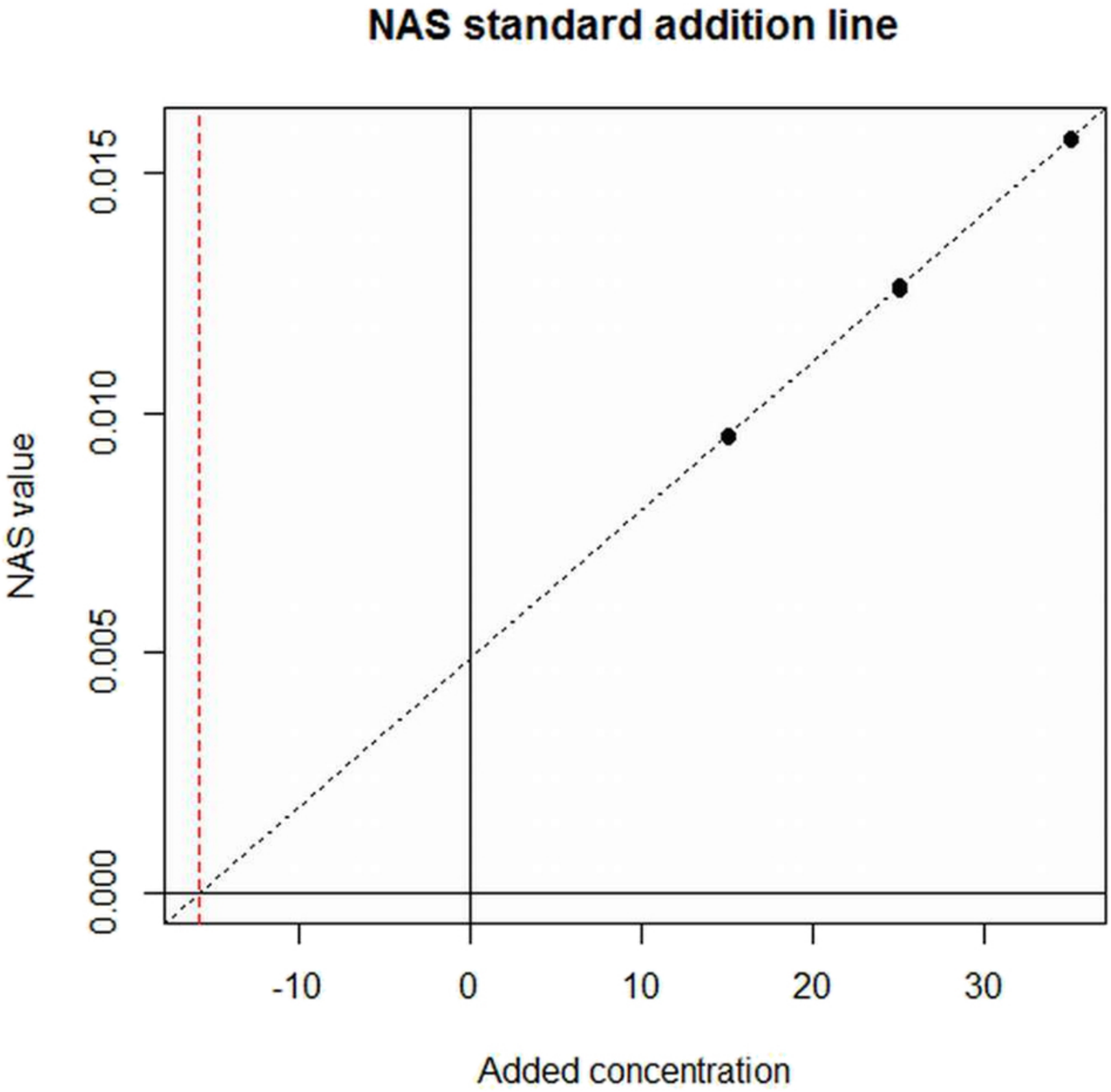

NAS standard addition line
NAS value
0.000
0.005
0.010
0.015
-10
0
10
20
30
Added concentration

Figure 5

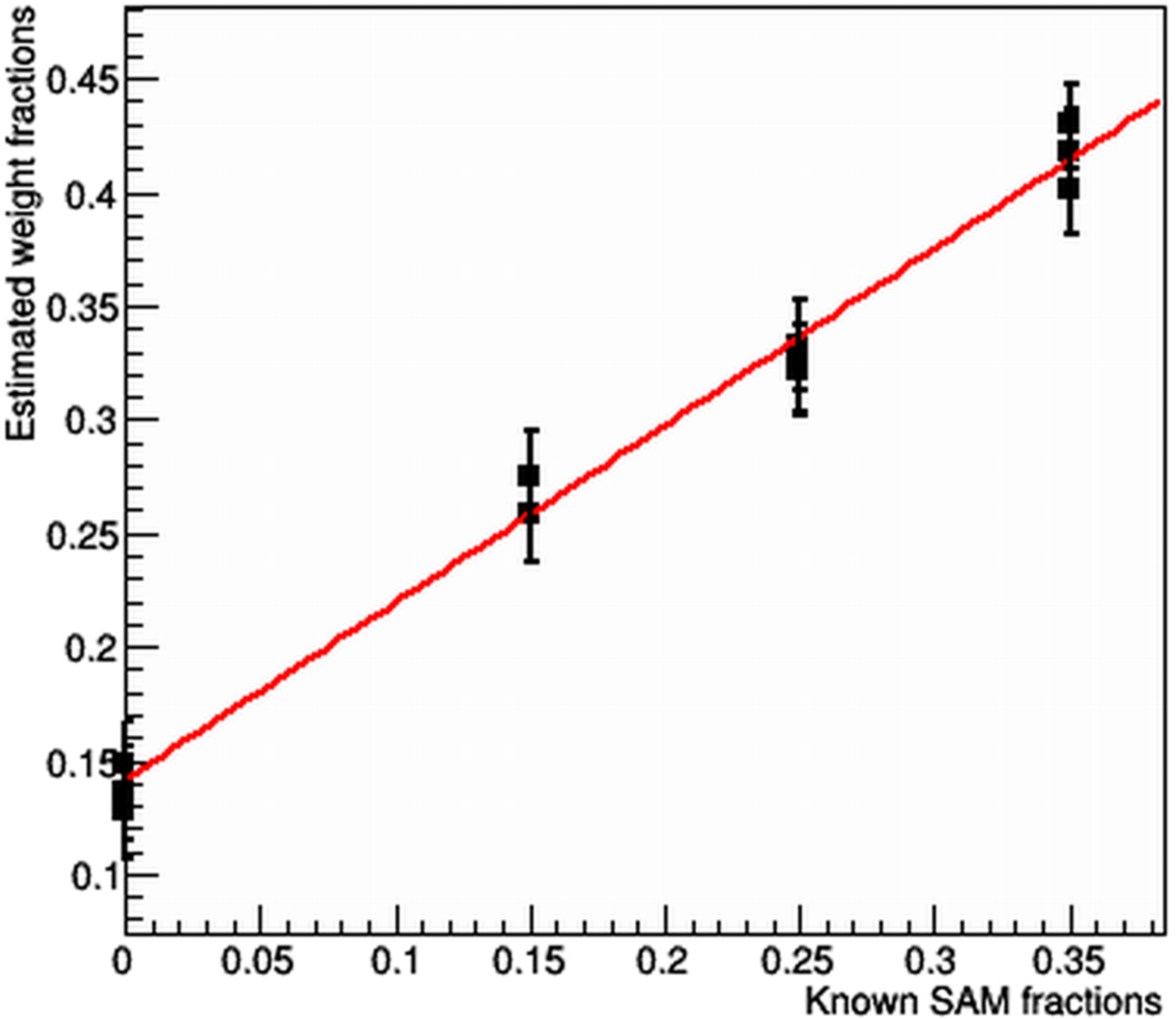

Estimated weight fractions
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Known SAM fractions

Figure 6