

## Optimization of Experimental Parameters in Analysis of Pharmaceutical Pellets by Near Infrared-Chemical Imaging (NIR-CI) and Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS)

José Augusto Da-Col<sup>\*,a,#</sup> and Ronei J. Poppi<sup>a</sup>

<sup>a</sup>Laboratório de Quimiometria em Química Analítica (LAQQA), Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas-SP, Brazil

The quality control of pellets homogeneity cannot be assessed by conventional techniques and near infrared-chemical imaging combined with multivariate curve resolution with alternating least squares is an attractive alternative. In this study, composition and spatial distribution of pellets components were determined after assessment of experimental parameters. The use of a 25  $\mu\text{m}$  intermediate pixel size, an initial estimation matrix with instrumental signals for pure substances and individual matrices provided a model with explained variance of more than 99% and a value of 0.00263 for percentage of lack of fit. In addition, the similarity between the pure substances spectra and those recovered by the model were 0.9501 for sucrose, 0.9480 for starch, 0.9910 for ketoprofen and 0.5941 for  $\text{SiO}_2$ . Chemical images were generated and show that the pellet is composed of an inert nucleus of starch and cellulose, surrounded by a ketoprofen layer. All this information was obtained quickly, in minutes, being an excellent alternative for pellets analysis.

**Keywords:** NIR chemical imaging (NIR-CI), hyperspectral image, pharmaceutical pellets, MCR-ALS, chemometrics

### Introduction

One of the great challenges of pharmaceutical industry is the ability to have tools to determine product properties. These properties can define whether the drug is suitable for sale and consumption, or to check if the active pharmaceutical ingredient (API) and excipients are available. An analytical technique should be chosen based on its ability to detect and quantify different properties in a fast and easy way.<sup>1</sup>

Solid oral preparations are widely used among the forms of drugs presentation. Examples are pellets, small spherical structures with controlled particle size.<sup>2</sup> Their production is usually made by agglomerating different substances as fine powders, but may also receive a coating film, and may incorporate drugs that are incompatible. This film coating may also contribute to aesthetic characteristics or for a uniform API distribution, avoiding concentration peaks on plasma, and intestinal mucosa aggression.<sup>3</sup> Such formulations may further be packaged in gelatin capsules or compacted as tablets.<sup>4,5</sup> In addition, API

dispersion in the medicine may affect its solubilization and bioavailability and, consequently, the drug efficacy.<sup>6</sup> This sample heterogeneity cannot be assessed by conventional techniques such as high-performance liquid chromatography (HPLC) or ultraviolet (UV), which determine only the overall concentrations of different species. Solid dispersions morphology can be evaluated by optical microscopy or scanning electron microscopy (SEM), however, without providing sample chemical information.<sup>7</sup>

The knowledge of the constituents distribution of these materials is very important and the hyperspectral images may help to determine the API properties in tablets, as spatial distribution and homogeneity. Hyperspectral images are, in a simplified way, spectral sets of a sample delimited region, obtained with an analytical technique capable of responding to the sample physicochemical properties. For image generation, sample area must be subdivided into several pixels and, for each one of these pixels, an absorption, emission or reflectance spectrum must be obtained, for example.<sup>8</sup>

A very useful technique for hyperspectral images generation is the infrared spectroscopy, especially in the near infrared region (NIR), between 13,000 to 4,000  $\text{cm}^{-1}$ .<sup>9</sup> In this regard, the near infrared-chemical imaging (NIR-CI)

\*e-mail: jose.col@udesc.br

<sup>#</sup>Present address: Departamento de Química, Centro de Ciências Tecnológicas, Universidade do Estado de Santa Catarina (UDESC), 89219-710 Joinville-SC, Brazil (<http://www.cct.udesc.br>).

has as one of its characteristics operating with diffuse reflectance, directly on the sample surface.<sup>10</sup> Because it is a non-destructive technique and requires little or no sample preparation,<sup>9</sup> it has been recommended by US Food and Drug Administration (FDA)<sup>11</sup> for pharmaceutical processes control.

Images generated by NIR-CI contain an immense amount of information in these spectra sets and their interpretation can be done with multivariate analysis methods. These methods can be either pattern recognition (such as principal component analysis, PCA), classification methods (such as linear discriminant analysis, LDA), or calibration methods (such as multivariate curve resolution, MCR).<sup>12</sup> There are some examples of application of NIR-CI<sup>13-17</sup> in the pharmaceutical area, with very promising results. Acetylsalicylic acid (ASA) concentration maps on commercial tablet samples were obtained by MCR without a previous calibration step.<sup>18</sup> Multivariate curve resolution with alternating least squares (MCR-ALS) works iteratively by optimizing concentration profiles and spectra (**C** and **S<sup>T</sup>** matrices) with previous constraints application. These restrictions are based on the chemical or mathematical knowledge about the data characteristics.<sup>19,20</sup> Instrumental data matrix (**D**) is decomposed by alternating least squares (ALS) algorithm in a relative concentration matrix of each pure substance (**C**) and another that refers to its spectra (**S**). The MCR-ALS expression can be described as  $\mathbf{D} = \mathbf{CS}^T + \mathbf{E}$ , where **E** is the residue matrix. Thus, MCR-ALS method can predict compounds concentration in samples, being suitable for industrial processes use or even in drug development steps.<sup>21-27</sup>

A valuable information in pharmaceutical industry is knowing halves of a fractionated tablet have the same API amount, as demonstrated by Franch-Lage *et al.*<sup>22</sup> Lorazepam tablets, an important antidepressant, usually present reduced size and only 1% in API mass. The verification of API and other 4 excipients spatial distribution was possible with hyperspectral spectroscopy and MCR, and the chemometric model proved no significant differences between the tablets halves.

In NIR-CI technique, there are three crucial steps that can affect analysis result: (i) data acquisition, with the correct sample preparation and instrumental irradiation conditions; (ii) careful choice of wavelengths and spectra preprocessing and (iii) with images to use for extracting studied system qualitative and quantitative information, such as API total concentration and distribution. If one of these steps is not properly performed, an erroneous system observation may result.

In this sense, Sabin *et al.*<sup>28</sup> compared several chemometric methods (multivariate curve resolution,

multiple linear regression, classical least squares and partial least squares) to evaluate the similarity of concentration maps obtained on carbamazepine tablets. Models were constructed and similarity was measured for the 6 studied compounds predicted values. CLS, MCR and PLS models gave equivalent results, probably because these methods consider all variables in the calculations. Only the MLR method did not present equivalent results with the other methods studied. Another way to evaluate methods similarity are chemical images histograms, where concentrations frequencies in the pixels are arranged in a graph. These histograms show the number of relative concentration values separated by ranges and the similarity of their profiles indicates the image similarity.

Excipients and diclofenac API spatial distribution in a medicament was achieved with a data treatment step to remove low reflectance values, typical of regions outside the sample, and to convert the reflectance values to pseudo-absorbance. Then, chemical images were generated by CLS, discriminating the presence and the concentrations of the constituents in the layers that constitute the pellet.<sup>29</sup>

However, for these analyses to be performed, some conditions must be optimized, such as spatial resolution. In NIR-CI, spatial resolution can be correlated with pixel size and should consider sample characteristics, such as particle size, and equipment characteristics, such as the ability to produce an intense signal to be detected.<sup>30</sup> Spectra type used to initialize the MCR-ALS model must also be considered to avoid that a probable local minimum is chosen as the system solution. Normally, initial estimations are constructed with an instrumental signal of pure components or with an estimate generated from the sample signal by chemometric tools capable of detecting the purest variables.<sup>24,31-33</sup> Among these chemometric tools, we highlight the simple-to-use interactive self-modeling mixture analysis (SIMPLISMA)<sup>31,34</sup> and the PURE function.<sup>19</sup> Another possibility is the use of individual matrices, for one sample, or augmented matrices, with the signal of more than one sample. These augmented matrices can provide a greater amount of information, as well as solve problems related to rank-deficient matrices.<sup>33,35-37</sup>

Taking into account these precautions, we can verify that NIR-CI and chemometrics combination is an excellent alternative in pharmaceutical area, but several parameters must be optimized to obtain realistic and trustworthy chemical images. In this paper, we intend to study the effects of pixel size variation, initial estimation type and augmented matrix in pharmaceutical pellets analysis optimization by near infrared-chemical imaging (NIR-CI) and multivariate curve resolution with alternating least squares (MCR-ALS).

## Experimental

### Samples

The samples used in this study were three commercial pellets (P1, P2 and P3), with same composition, containing the active ingredient ketoprofen, a non-steroidal anti-inflammatory drug with anti-inflammatory and analgesic properties. Pellets consist of an inert core and layers that cover the central region, having a mean diameter of 900  $\mu\text{m}$ .

For the sampling, pellets were cut in half with a razor blade in a magnifying glass (Carl Zeiss, Stemi DV4 model, 8 to 32 $\times$  magnification). For measurements, they were affixed in a glass plate using transparent double-sided tape to be monitored on PerkinElmer Spotlight 400N FT-NIR Imaging System.

### Materials

Analytical grade reagents ketoprofen (Sigma-Aldrich, USA), starch (Synth, Brazil), sucrose (Synth, Brazil) and silicon dioxide (Sigma-Aldrich, USA) were also used for comparison of original and recovered signals by MCR-ALS model.

### Operating conditions

For the acquisition of pellet and pure substances spectra, the employed PerkinElmer Spotlight 400 N FT-NIR Imaging System spectrometer operating conditions are described in Table 1.

### Experimental data treatment

Spectra were obtained in reflectance mode (R) and converted to absorbance ( $\log 1/R$ ), then the standard

normal variation (SNV)<sup>38</sup> was applied as pre-processing, to minimize possible radiation scattering effects and changes caused by particle size. Data were also smoothed by Savitzky and Golay<sup>39</sup> with a 19-point window for experimental noise removal.

To remove glass plate and adhesive tape spectra from the data set, a selection step was performed with PCA scores values of the mean centered data. Thus, a range of values of scores corresponding only to pellets spectra was selected. This process could be performed graphically with the Hyper-Tools package.<sup>40</sup>

With this new data set, MCR-ALS models were built, always with 4 components. Components number choice considered the results of a PCA with the mean-centered data, which explained more than 99% of the variance.

As an initial estimate to start MCR-ALS model, matrices were constructed with pure substances spectra obtained experimentally and also with purest variables estimated with chemometric tools, such as SIMPLISMA algorithm.<sup>34</sup> In all calculations, constraints were applied, such as 500 iterations, 0.1% maximum error, and no negativity for concentration.

All calculations were performed in Matlab R2016b<sup>41</sup> environment, with PLS Toolbox v. 8.1.1<sup>42</sup> and HYPER-Tools package.<sup>40</sup>

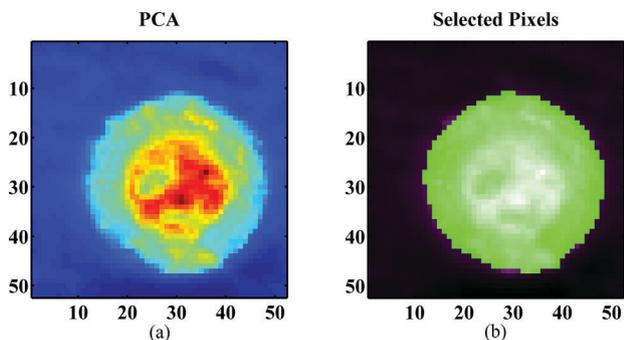
## Results and Discussion

Comparison of pixel size (6.25, 25 and 50  $\mu\text{m}$ ) effects in the MCR-ALS models results

After removal of non-pellet spectra, using the routine provided by Amigo,<sup>40</sup> the data set spectra number was reduced from 43264 to 16137 (6.25  $\mu\text{m}$ ), 2704 to 1021 (25  $\mu\text{m}$ ) and 676 to 284 (50  $\mu\text{m}$ ). In Figure 1, a graphic example of this selection step can be seen, for a 25  $\mu\text{m}$  pixel size.

**Table 1.** Operating conditions of Spotlight 400N FT-NIR Imaging System spectrometer for pellets and pure substances analysis

Operating condition	6.25 $\mu\text{m}$	25 $\mu\text{m}$	50 $\mu\text{m}$
Resolution / $\text{cm}^{-1}$	16	16	16
Scans number	32	32	32
Spectral range / $\text{cm}^{-1}$	7800-4000	7800-4000	7800-4000
Pixel size / $\mu\text{m}$	6.25	25	50
Image area / $\mu\text{m}$	1300 $\times$ 1300	1300 $\times$ 1300	1300 $\times$ 1300
Image area in pixel number	208 $\times$ 208	52 $\times$ 52	26 $\times$ 26
Pure substances area / $\mu\text{m}$	75 $\times$ 75	300 $\times$ 300	600 $\times$ 600
Pure substances area in pixel number	12 $\times$ 12	12 $\times$ 12	12 $\times$ 12
Spectra number	43264	2704	676
Variables number	476	476	476

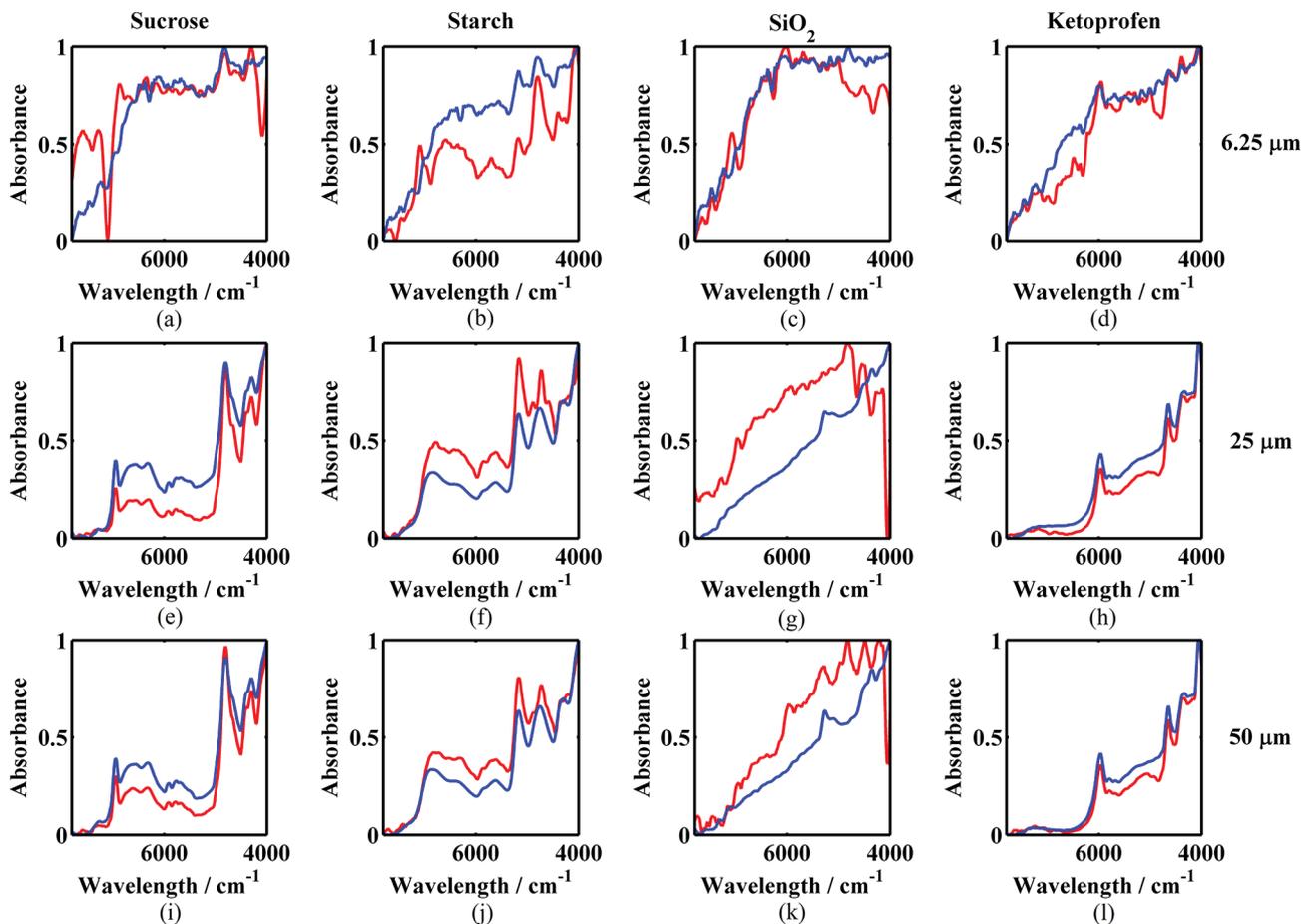


**Figure 1.** Generated image by PCA of (a) all pixel spectra and (b) after selection of only sample P1 pixels, in green, by using HYPER-Tools routine.

A four-component model was created for each pixel size studied, and the four components of the model are the API and the excipients sucrose, starch and silicon dioxide ( $\text{SiO}_2$ ). In this specific study, a matrix was constructed containing the pure spectra of the 4 substances as initial estimate. Pure spectra were obtained under the same sample image operating conditions, with only a smaller sample area of  $12 \times 12$  pixels (144 spectra).

MCR-ALS model basically decomposes original signal into an array containing the relative concentrations of the components (matrix **C**) and another with the spectral signals of those components (matrix **S**). When these signals are plotted next to the signals obtained by the analysis of their pure components, it is possible to compare if the profiles are coincident. The higher the match, the greater the model ability to recognize these less random signals in instrumental signals mixture. This comparison can be seen in Figure 2.

According to the Figure 2, we can initially verify that the signals obtained by the  $6.25 \mu\text{m}$  pixel size are noisier and with less defined bands than for larger pixel size. This is due to the way in which the equipment selects the pixel size, by physical light passage restriction. When we use a smaller pixel size, less radiation reaches the sample and, consequently, less radiation reaches the detector. This situation causes a decrease in signal to noise ratio, when less system information can be modeled by the MCR-ALS model, which implies a less complete model. To avoid this, a larger number of scans should be used, which would make



**Figure 2.** Comparison for each pixel size of normalized pure substances spectra (blue) and recovered spectra by the MCR-ALS model (red) for (a, e, i) sucrose, (b, f, j) starch, (c, g, k)  $\text{SiO}_2$  and (d, h, l) ketoprofen.

the analysis very time consuming, rendering it unattractive for practical purposes.

Signals observation for larger pixels indicates that there are no differences of spectral profiles between 25 and 50  $\mu\text{m}$  sizes. In these cases, bands are better defined and coincidental between the pure substances spectra and the recovered spectra, only with variations in intensities. These variations may be due, for example, to interactions between species in the pellet, which do not occur when analyzing the substances separately. For the  $\text{SiO}_2$  this does not occur, since it has a spectral profile without many characteristic bands. This is due to its low density, which generates a large radiation diffusion in its interior, not allowing an appreciable amount of signal to reach the detector.

A more elegant way of conferring similarity is by calculating the correlation coefficients ( $r^2$ ) between spectra (Table 2), which is related to the covariance matrix of each spectrum. In these cases, the compared spectra will be more similar when the values are close to unity. This can be observed for ketoprofen in the three pixel sizes, especially for larger pixels, with values higher than 0.9903, against 0.9597 of the smaller pixel. For sucrose and starch the same behavior is observed, with values higher than 0.94, while at the 6.25  $\mu\text{m}$  pixel the values are lower than 0.8005. Finally, for  $\text{SiO}_2$  the values do not follow a direct trend with the pixel size variation, with the best value for

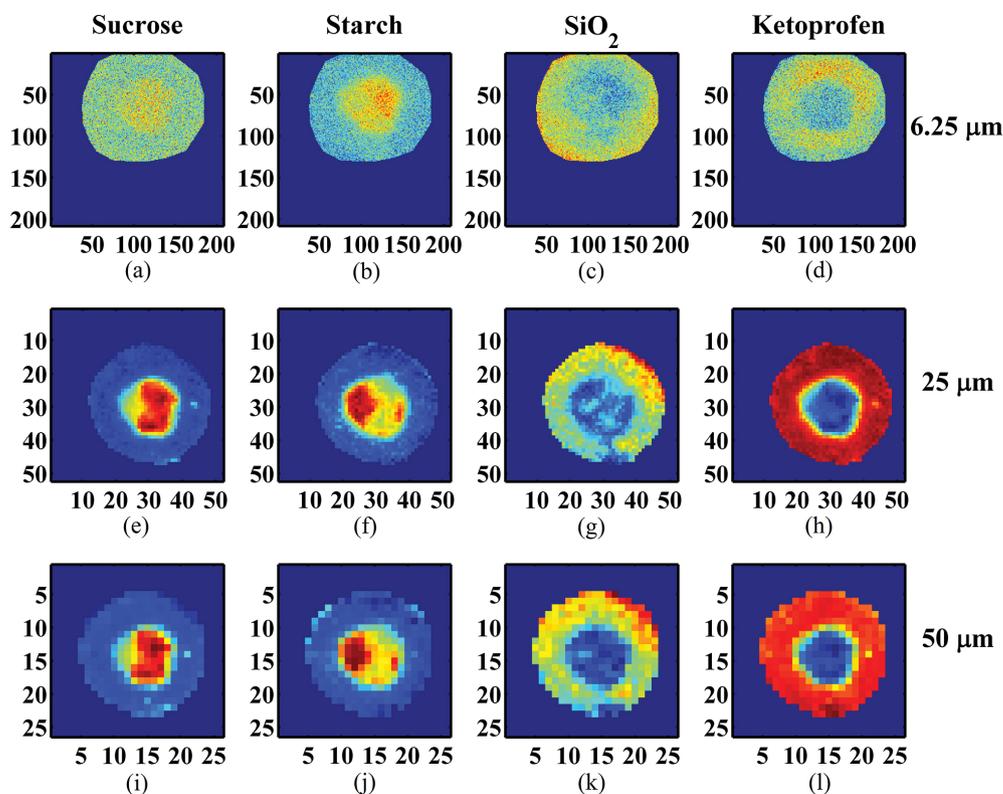
the smallest pixel (0.9464) and worst value (0.6112) for the intermediate pixel size.

**Table 2.** Calculated correlation coefficients values ( $r^2$ ) between obtained spectra by the MCR-ALS model and those obtained for the pure substances at 6.25, 25 and 50  $\mu\text{m}$  pixel sizes

Pixel size / $\mu\text{m}$	Correlation coefficient			
	Sucrose	Starch	$\text{SiO}_2$	Ketoprofen
6.25	0.8005	0.8705	0.9464	0.9597
25	0.9607	0.9485	0.6112	0.9903
50	0.9836	0.9766	0.8751	0.9922

Another MCR-ALS result is the generation of a matrix with the relative concentrations of the system components (matrix **C**), which will provide the chemical images construction, as indicated in Figure 3.

Initially, the images at 6.25  $\mu\text{m}$  pixel size show a projection to the upper image region, different from that obtained visually at the moment of the sample positioning in the equipment. This may indicate a possible equipment limitation, more sensitive to a sample angulation variation relative to the sample port axis, in situations where the signal magnitude approaches the instrumental limit of detection. Regarding the type of signal obtained, these results are consistent with those obtained by Ma and



**Figure 3.** Distribution maps generated by MCR-ALS models with different pixel sizes (6.25, 25 and 50  $\mu\text{m}$ ) for the 4 components studied (warm colors indicate the presence and cold colors indicate components absence).

Anderson,<sup>30</sup> who studied the increase of the magnification in obtaining chemical images with NIR and realized that the spatial resolution increase did not provide additional information. Despite this restriction, images could be obtained, but without great agreement with larger pixels images. Thus, it is evident that there are quite distinct regions in the pellet. In the center, there is an inert region containing sucrose and starch. Inert pellet use as starting material for the interest compounds deposition is a common practice in the pharmaceutical industry.<sup>23,43,44</sup> Immediately after this inert core, there is a layer containing the API, which appears in the image with an intense red color, demonstrating the great contribution of this species in this region. Finally, SiO<sub>2</sub> appears to be more externally distributed in the sample, perhaps as a coating layer or to protect the pellet. However, as the conditions for obtaining an instrumental signal for this compound is not as good as for the others, some care must be taken with this component results.

To evaluate the MCR-ALS models quality, the percentage of variance explained ( $R^2$ ) and the percentage of lack of fit (LOF) were calculated. Those figures of merit were calculated as described by Jaumot *et al.*,<sup>20</sup> and can be seen in Table 3.

**Table 3.** Percentage of variance explained ( $R^2$ ) and percentage of lack of fit (LOF) values of the MCR-ALS model for each pixel size (6.25, 25 and 50  $\mu\text{m}$ )

Figure of merit	Pixel size		
	6.25 $\mu\text{m}$	25 $\mu\text{m}$	50 $\mu\text{m}$
Percentage of variance explained ( $R^2$ ) / %	91.19	99.82	99.88
Percentage of lack of fit (LOF) / %	0.0381	0.00263	0.00000725

The percentage of variance explained indicates how much of the original information, in the form of instrumental signal, is being described by the MCR-ALS model. When the calculated value is closer to 100%, the model created will be more representative of the original signals. Analysis of the obtained values for this parameter (Table 3) shows that both the model for the 25 and 50  $\mu\text{m}$  pixel size can express more than 99% of variance explained. MCR-ALS model for 6.25  $\mu\text{m}$  pixel size describes only 91% of variance explained, which indicates that the use of a larger spatial resolution does not necessarily imply an increase in the MCR-ALS model capacity to describe the original information. This same trend is verified by the percentage of lack of fit values, which expresses the difference between the original data matrix (**D**) and the

matrix reproduced by the MCR-ALS model (**CS<sup>T</sup>** matrix).<sup>20</sup> As these values must be the smallest possible, the increase in spatial resolution cannot provide an improvement in the models according to this figure of merit. Percentage of lack of fit magnitude is  $10^{-6}$  for the 50  $\mu\text{m}$  pixel size, while for 25 and 6.25  $\mu\text{m}$  is  $10^{-3}$  and  $10^{-2}$ , respectively.

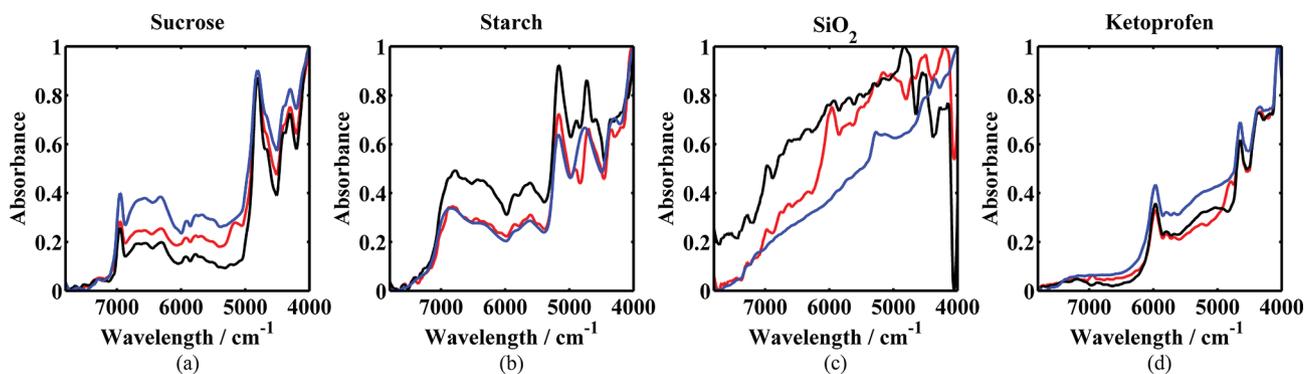
According to the results analysis, the use of 6.25  $\mu\text{m}$  pixel size should be discarded because it requires a much longer analysis time than for 25 and 50  $\mu\text{m}$  pixels. Thus, it would be more appropriate to use the maximum pixel size, 50  $\mu\text{m}$ , with shorter analysis times, since the spectral profile recovery trends, chemical images and figures of merit indicate its viability. However, it is necessary to balance between the analysis time and the generated image resolution. Thus, the use of an intermediate pixel size, 25  $\mu\text{m}$ , can provide adequate spectral recoveries and figures of merit with more detailed chemical images.

Comparison of the use or not of an initial estimation with pure substances instrumental data in the MCR-ALS results

As discussed in the previous section, the use of intermediate pixel provides adequate information and upcoming studies will be performed under this condition. One of these studies is about the type of initial estimation used to initialize the MCR-ALS model. When using the ALS algorithm, this step is critical to obtaining a solution that best describes the system. One of the implications of choosing a good initial estimate is to avoid a possible local minimum, which could lead to a solution that is not representative of the system.<sup>45,46</sup>

There are two possibilities of constructing an initial estimation matrix and one of them is with some previously known information, as pure substances signals obtained experimentally, under the same conditions in which the sample signal will be obtained. In this specific case, matrices containing 144 spectra were generated for each pure substance considered (sucrose, starch, silicon dioxide and ketoprofen). Other way to generate an initial estimate for ALS is to use some tool that can detect the purest variables within the entire sample signal. An example is the simple-to-use interactive self-modeling mixture analysis (SIMPLISMA)<sup>34</sup> or the PURE function, provided by Tauler,<sup>19,47</sup> which are widely used for the purest variables detection. Thus, the MCR-ALS can be started and the iterations can be developed, taking a model with smaller errors.

With these estimates, MCR-ALS models were constructed considering each of them and the spectra provided by the models can be seen in Figure 4. In this Figure, the spectra recovered by the model can be compared



**Figure 4.** Comparison of the normalized spectra of the pure substances (blue) and reconstructed by the MCR model with instrumental estimation of the pure substances (black) and initial estimation obtained by the PURE function (red) for (a) sucrose; (b) starch; (c) SiO<sub>2</sub>; (d) ketoprofen.

to each other and to the pure substances spectra profiles obtained experimentally.

The obtained profiles for the substances sucrose, starch and ketoprofen present a fair similarity tendency, with small variations of intensity for normalized signals and few non-coincident bands. This is shown in the values of the correlation coefficients between the spectra recovered by the MCR-ALS models and the spectra obtained experimentally from the pure substances, which are between 0.9485 and 0.9903. Comparison of the correlation coefficients values for spectra recovered between the models with distinct types of initial estimation also indicates a good concordance in the three substances, with values of  $r^2$  in the range of 0.9609 and 0.9936. These values are shown in Table 4.

By observing the profiles for SiO<sub>2</sub> (Figure 4c), there is only a trend of similarity in the spectra slope, without coincident bands. Correlation coefficient values are lower than for other substances, 0.6112 for the model with the experimental estimate of pure substances signal, 0.9184 for the model that employed pure function and 0.8466 between the models. As discussed previously, a likely explanation for these results is the large diffusion of SiO<sub>2</sub> radiation, which limits that an intense signal reaches the detector. Another possibility is that this component is present in low concentration, reducing the ability of technique to discriminate the analyte signal from instrumental noise.

In the chemical images analysis, we can see that there are no significant differences between the images generated for the two approaches to construct the estimates that will start the models (Figure 5). Trends are the same, from an inert central layer containing starch and sucrose, an attached layer with high API concentration and a non-uniform distribution of SiO<sub>2</sub>.

Based on these results, we can affirm that there are no significant differences in using any approach to the generation of the initial estimates. Table 5 shows the results for the explained variance and the percentage of lack of fit for both models. Models can explain more than 99% of the

**Table 4.** Calculated values of the correlation coefficients ( $r^2$ ) between the pure substances spectra and those obtained by the MCR-ALS model with and without initial estimation with instrumental signals of the pure substances, and among the models

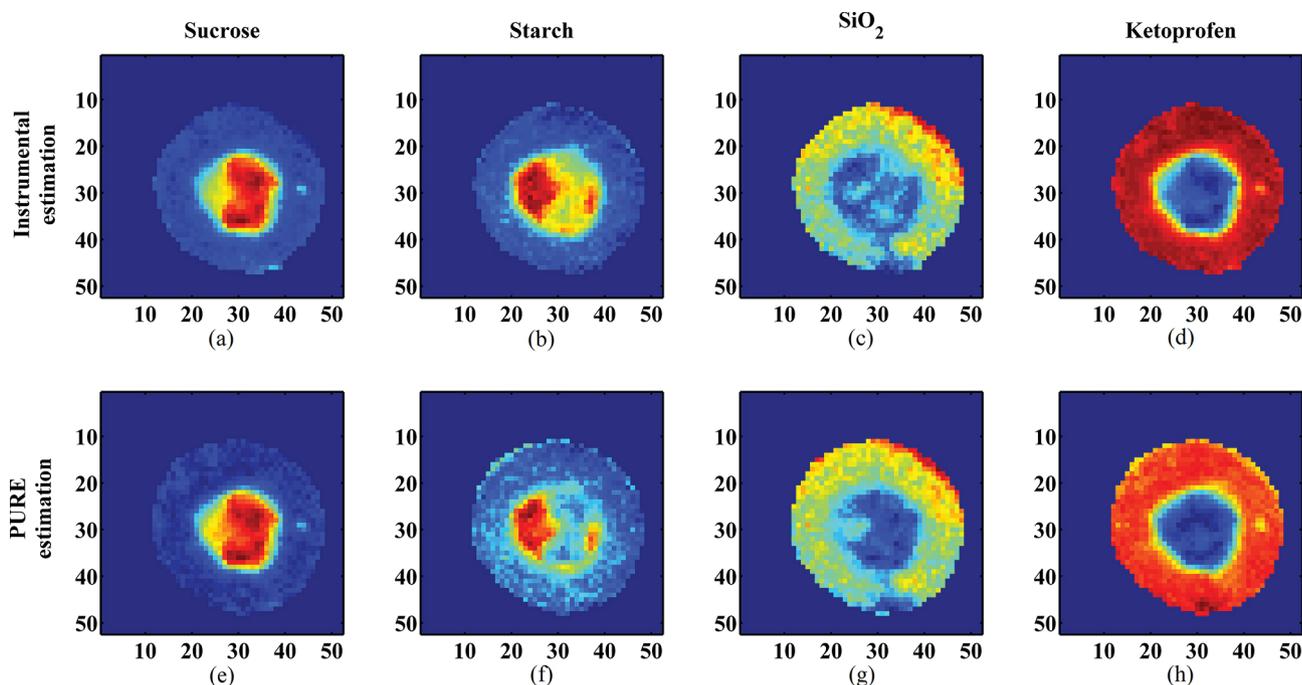
Component	Pure substances spectra		Between models
	With instrumental signal	Using PURE function	
Sucrose	0.9607	0.9866	0.9860
Starch	0.9485	0.9849	0.9609
SiO <sub>2</sub>	0.6112	0.9184	0.8466
Ketoprofen	0.9903	0.9831	0.9936

original system information, with values of percentage of lack of fit less than zero, from 0.00263 for the model that uses instrumental signals of pure substances and 0.00940 for the model that uses pure function.

Finally, if the choice is practical, the PURE routine provides a quicker and faster analysis, since it eliminates the need to analyze pure substances individually. However, it is necessary to compare the signal generated by the model with some characteristic signal of the components that allows their characterization. In drugs, these signals are more easily obtained because the formulations are simpler, and the pure substances are available. Although the use of this type of instrumental signal causes this approach to take a little longer, it is preferable to facilitate the characterization of the signals obtained.

Comparison of the use of individual or augmented matrices in the results of MCR-ALS

Now that the pixel size has been chosen (25  $\mu\text{m}$ ), as well as the type of signal used to generate the initial estimation for the ALS algorithm, a final consideration can be made. For those situations where more than one sample replica is available, one may choose to create an individual model for each of these replicas or a single



**Figure 5.** Distribution maps images generated by the MCR-ALS models (a-d) with the use of initial signal estimation of pure substances and (e-h) without the instrumental signal of pure substances for the 4 components studied (warm colors indicate presence and cold colors indicate the component absence).

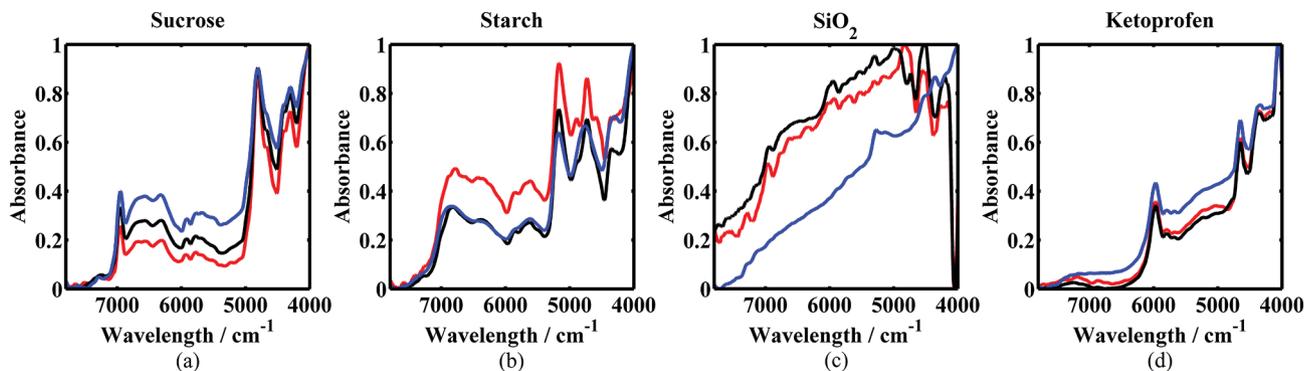
**Table 5.** Values of percentage of variance explained ( $R^2$ ) and percentage of lack of fit (LOF) of the MCR-ALS models with initial estimation with instrumental signal and with signals obtained by PURE function

Figure of merit	Initial estimation	
	With instrumental signal	Using PURE function
Variance explained ( $R^2$ ) / %	99.82	99.79
Lack of fit (LOF) / %	0.00263	0.00940

model containing the instrumental signals of all such replicas. In this second case, an augmented column-wise data matrix is created, and after the model is made, this information is separated so that the images of each one of these samples can be generated. With this approach, the expected advantage would be a greater amount of

information associated to the system, which could result in a model with improved characteristics.<sup>48</sup>

In this study, 3 different pellets were analyzed and for each one was constructed a matrix containing 1021 spectra for pellet P1, 971 for P2 and 1305 for P3. The augmented matrix containing the 3 pellet signals had the size of 3297 rows and 476 columns, with lines 1 through 1021 containing the information from sample P1, lines 1022 to 1992 from sample P2 and 1993 to 3297 from sample P3. After the models were made, the recovered signals profiles could be compared in Figure 6. As in the other cases, the results are very similar, with variations in signal intensity after normalization. However, several bands that are in the spectrum of pure substances are present in those recovered spectra, allowing identification of the components.



**Figure 6.** Comparison of the normalized spectra of pure substances (blue) and reconstructed by the MCR-ALS model with individual image for the sample P1 (red) and with increased matrix (black) for (a) sucrose; (b) starch; (c)  $\text{SiO}_2$  and (d) ketoprofen.

Again, the profiles for SiO<sub>2</sub> are the only ones different than expected, which is explained by the difficulties discussed above. Correlation coefficients values for the component are 0.6112 and 0.5848 for the individual and augmented matrix, respectively (Table 6). We can verify that the spectra recovered by both models are very similar, with  $r^2$  values higher than 0.9711.

**Table 6.** Calculated correlation coefficients values ( $r^2$ ) between the pure substances spectra and those obtained by the MCR-ALS model with the use of individual matrix and augmented matrix

Component	Pure substances spectra		Between models
	Individual matrix	Augmented matrix	
Sucrose	0.9607	0.9846	0.9931
Starch	0.9485	0.9813	0.9711
SiO <sub>2</sub>	0.6112	0.5848	0.9812
Ketoprofen	0.9903	0.9895	0.9994

Chemical images analysis for each component can be seen in Figure 7 and indicate the same distribution trend of components in the pellet, regardless of how the model was constructed.

The same type of information can be extracted from Table 7, which presents the values of percentage of variance explained and percentage of lack of fit, which are very close. This indicates that the option for a larger number of samples, in this specific case, does not imply an increment in the model that justifies the longer time that should be

**Table 7.** Values of percentage of variance explained ( $R^2$ ) and percentage of lack of fit (LOF) of the MCR-ALS models with individual matrix and with augmented matrix

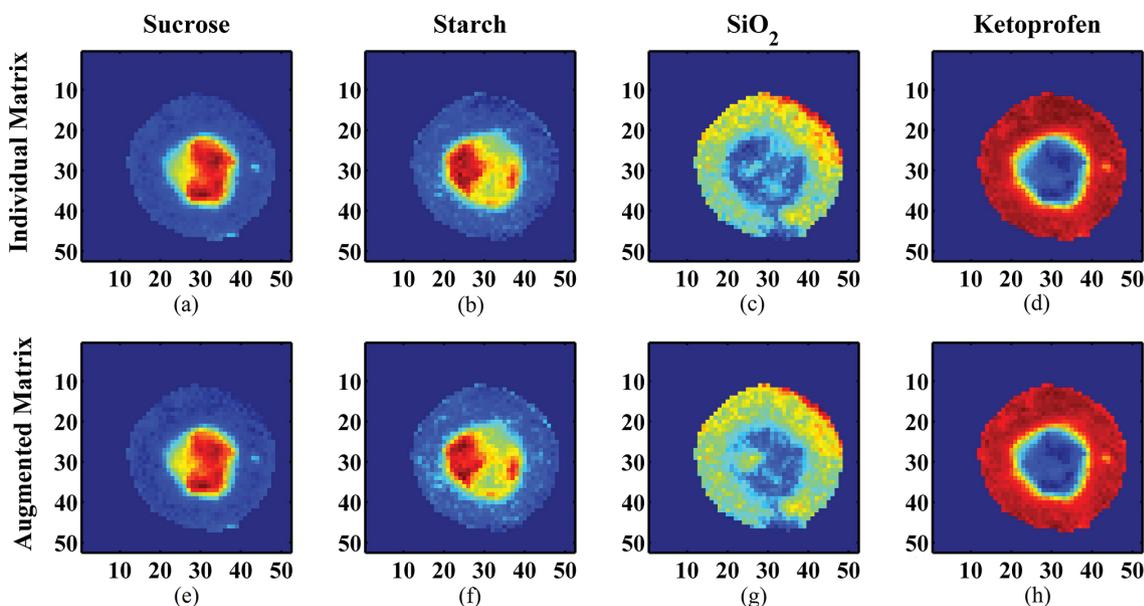
Figure of merit	Individual matrix	Augmented matrix
Variance explained ( $R^2$ ) / %	99.82	99.83
Lack of fit (LOF) / %	0.00263	0.00632

used for its execution. Therefore, it is recommended to use a model with individual images.

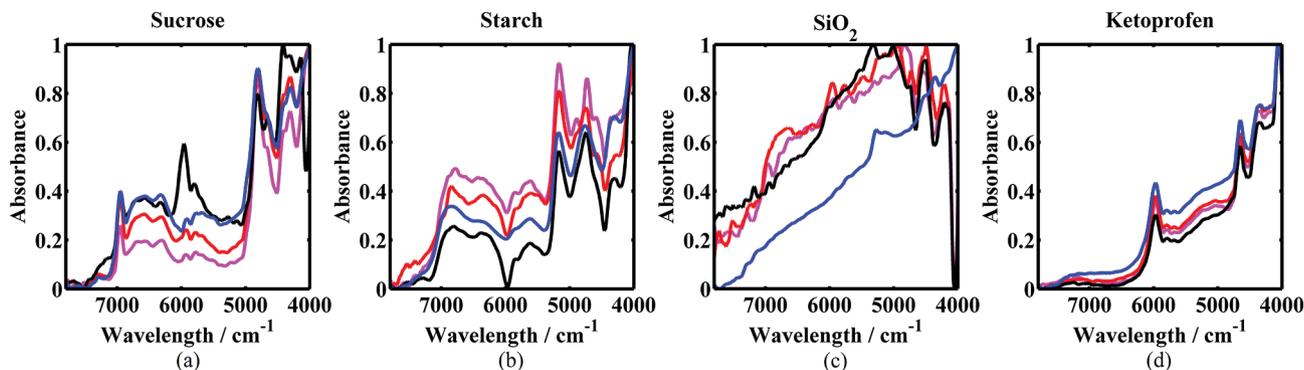
### Results for samples P1, P2 and P3 under optimized experimental conditions

The last step of this study was the verification of the feasibility of use in other pellet replicates. For this, two new samples were submitted to the same treatment and the signals are compared in Figure 8. We can observe a great agreement between the signals obtained for the 3 samples (P1, P2, and P3) and between those signals and those of the pure substances. The only exception is the sucrose spectrum for sample P3 (Figure 8a), which has a different band near 6000 cm<sup>-1</sup>, associated with ketoprofen.

This trend is confirmed by the analysis of the correlation coefficients values (Table 8). All values are very close, except for sucrose in sample P3, which is 0.9076. The low values for the SiO<sub>2</sub> component, already expected by their characteristics, demonstrate the difficulty of determining the ingredient in a formulation by NIR-CI and MCR-ALS.



**Figure 7.** Distribution maps generated by MCR-ALS models (a-d) with the use of individual matrix for the sample P1 and (e-h) with matrix augmented for the 4 components studied (the warm colors indicate the presence and the cold colors indicate the absence of the components).



**Figure 8.** Comparison of the normalized spectra of pure substances (blue) and reconstructed by the MCR-ALS model for replicates P1 (magenta), P2 (red) and P3 (black) for (a) sucrose; (b) starch; (c) SiO<sub>2</sub> and (d) ketoprofen.

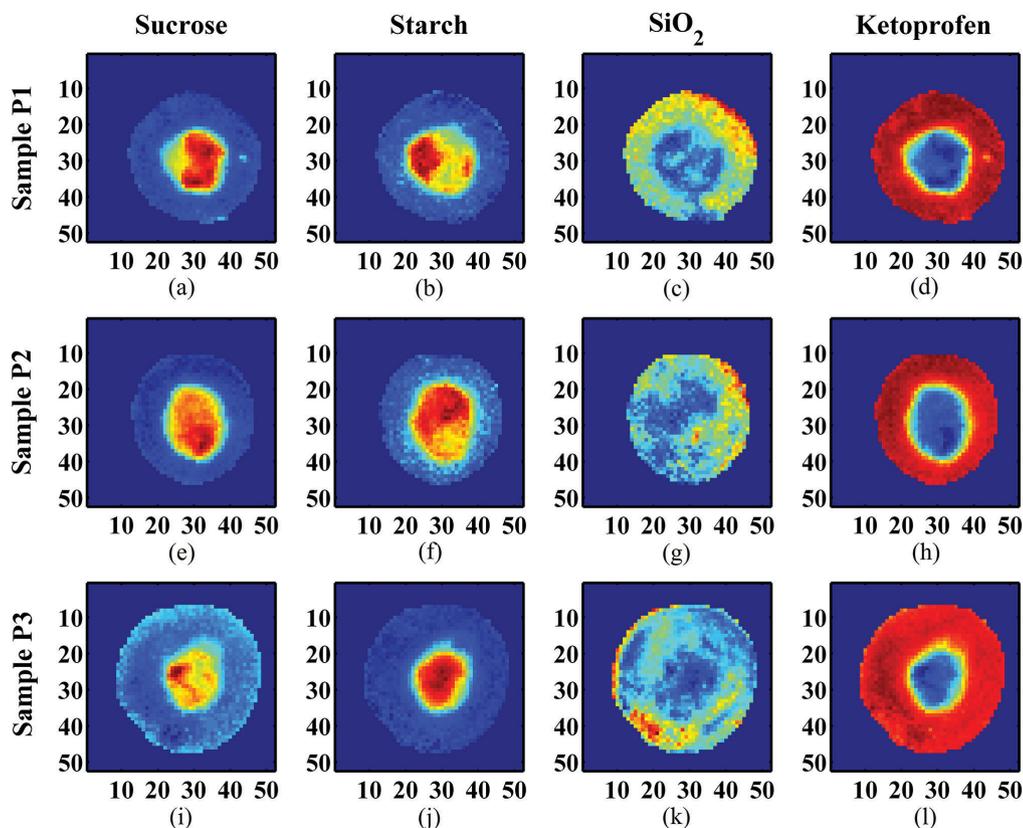
**Table 8.** Calculated correlation coefficients values ( $r^2$ ) between the pure substances spectra and those obtained by the MCR-ALS model for the samples replicates P1, P2 and P3

Component	Sample P1	Sample P2	Sample P3
Sucrose	0.9607	0.9820	0.9076
Starch	0.9485	0.9437	0.9518
SiO <sub>2</sub>	0.6112	0.6066	0.5644
Ketoprofen	0.9903	0.9947	0.9881

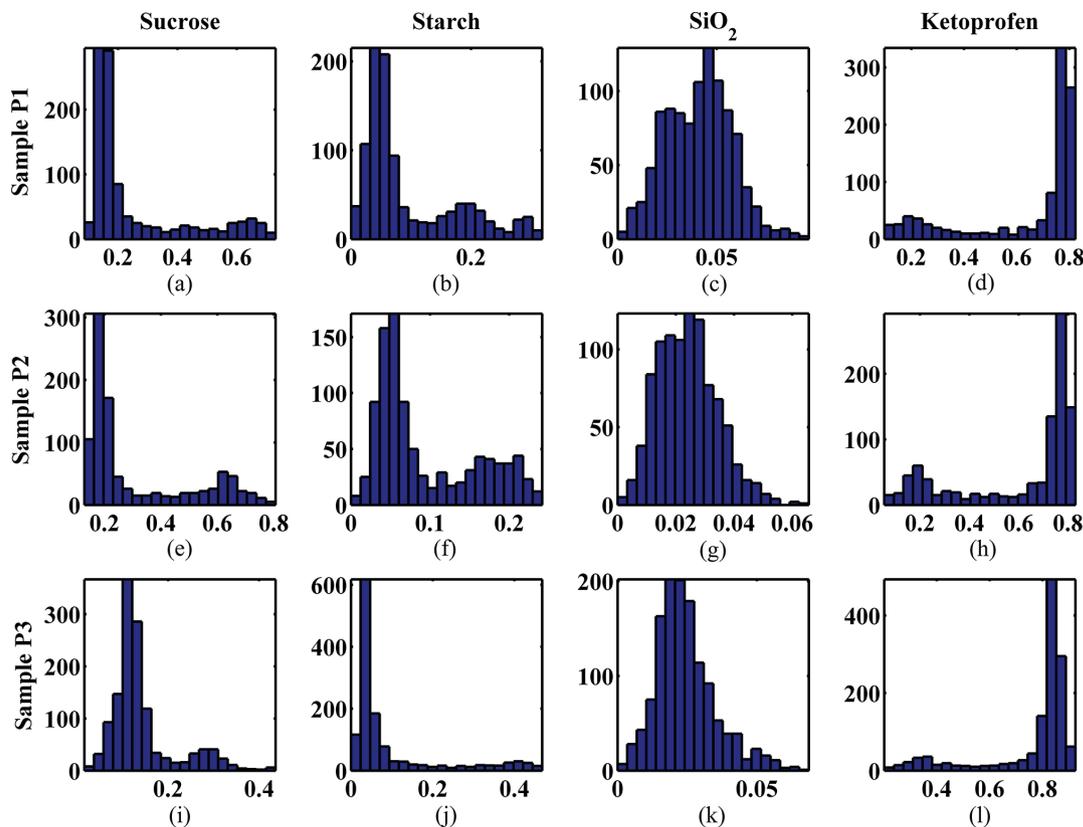
Finally, the comparison of the chemical distribution images of components (Figure 9) reproduces the trend pointed out in the first sample. There is an inert core of

sucrose and starch, a very uniform layer of ketoprofen and a very uneven distribution of silicon dioxide.

Another very interesting way to verify the distribution of these species in drug is with histograms (Figure 10), where the relative concentrations of each component in each pixel are expressed. These histograms contain the counts of each value in each range of relative concentration and the more similar their profiles, the closer the components distributions in each sample should be. Although not identical, the profiles of the histograms are similar, indicating that the images are quite equivalent.



**Figure 9.** Comparison of the distribution map images generated by the MCR-ALS models for the replicates (a-d) P1, (e-h) P2 and (i-l) P3 for the 4 components studied (warm colors indicate presence and cold colors indicate absence of components).



**Figure 10.** Histograms for the relative image concentrations of the 4 components of the replicates (a-d) P1; (e-h) P2 and (i-l) P3. Relative concentrations ranges are expressed on the x-axis and relative concentration counts, on the y-axis.

Although the sample P3 apparently has a cut a little distant from its center, the results are consistent with the others, whose cut is more in the central region. This indicates that the technique is robust to small variations in the sample cut. The comparison of the values of percentage of variance explained ( $R^2$ ) and percentage of lack of fit (LOF) corroborate this information, since the figures of merit do not present significant variations (Table 9), the models being equivalent.

**Table 9.** Percentage of variance explained ( $R^2$ ) and percentage of lack of fit (LOF) values of the MCR-ALS models for the replicates P1, P2 and P3

Figure of merit	Sample P1	Sample P2	Sample P3
Variance explained ( $R^2$ ) / %	99.82	99.86	99.87
Lack of fit (LOF) / %	0.00263	0.00150	0.00434

## Conclusions

Combination of NIR image spectroscopy and multivariate curve resolution is extremely efficient in solving problems involving components identification and their distributions in pharmaceutical samples as used in this study. Using 25  $\mu\text{m}$  intermediate size pixels, this information could be obtained quickly, with models

containing more than 99% of the original information and 0.00263 for a percentage of lack of fit. Species presence could be verified by comparing the signals recovered by the chemometric model with those obtained instrumentally from the pure reagents, with a mean similarity of 0.9501 for sucrose, 0.9480 for starch and 0.9910 for ketoprofen. Even in the case of  $\text{SiO}_2$ , whose mean correlation coefficient was only 0.5941, there is a tendency to associate the compound with the recovered signal.

Chemical distribution images of the species allow the identification of a highly homogeneous layer of API that surrounds an inert nucleus containing the excipients starch and sucrose. Apparently, there is also a heterogeneous layer of silicon dioxide, which surrounds the pellet, perhaps for its protection. The models constructed for the studied pellets show that the images are obtained equally either if the data were obtained for a single replica (individual matrix) or for more than one (augmented matrix).

After optimization, the best conditions were the use of a 25  $\mu\text{m}$  intermediate pixel, an initial estimation matrix with instrumental signals for pure substances and individual matrices, that provides a MCR-ALS model with possibility to explain the composition and spatial distribution of pellets components for each pellet sample, in few minutes and without waste generation.

## Acknowledgments

The authors thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) for supporting this study, André F. P. Biajoli for English evaluation and valuable critical analysis and Marco A. Z. Arruda for lending the magnifying glass.

## References

- Amigo, J. M.; *Anal. Bioanal. Chem.* **2010**, *398*, 93.
- Yadav, N.; Verma, A.; *Indian J. Pharm. Educ. Res.* **2016**, *50*, S146.
- Dukic-Ott, A.; Thommes, M.; Remon, J. P.; Kleinebudde, P.; Vervaet, C.; *Eur. J. Pharm. Biopharm.* **2009**, *71*, 38.
- Santos, H. M. M.; Veiga, F. J. B.; de Pina, M. E. T.; de Sousa, J. J. M. S.; *Rev. Bras. Cienc. Farm.* **2004**, *40*, 455.
- Santos, H. M. M.; Veiga, F. J. B.; de Pina, E. M. S. T.; de Sousa, J. J. M. S.; *Rev. Bras. Cienc. Farm.* **2006**, *42*, 309.
- Leuner, C.; *Eur. J. Pharm. Biopharm.* **2000**, *50*, 47.
- de Juan, A.; Tauler, R.; Dyson, R.; Marcolli, C.; Rault, M.; Maeder, M.; *TrAC, Trends Anal. Chem.* **2004**, *23*, 70.
- Geladi, P.; Burger, J.; Lestander, T.; *Chemom. Intell. Lab. Syst.* **2004**, *72*, 209.
- Treado, P. J.; Priore, R. J.; Nelson, M. P. In *Handbook of Vibrational Spectroscopy*; Chalmers, J. M.; Griffiths, P. R., eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2006.
- Lopes, M. B.; Wolff, J.-C.; Bioucas-Dias, J. M.; Figueiredo, M. A. T.; *Anal. Chim. Acta* **2009**, *641*, 46.
- Ravn, C.; Skibsted, E.; Bro, R.; *J. Pharm. Biomed. Anal.* **2008**, *48*, 554.
- Geladi, P.; Sethson, B.; Nyström, J.; Lillhonga, T.; Lestander, T.; Burger, J.; *Spectrochim. Acta, Part B* **2004**, *59*, 1347.
- Zhou, L.; Xu, M.; Wu, Z.; Shi, X.; Qiao, Y.; *Drug Test. Anal.* **2016**, *8*, 72.
- Khorasani, M.; Amigo, J. M.; Sun, C. C.; Bertelsen, P.; Rantanen, J.; *Eur. J. Pharm. Biopharm.* **2015**, *93*, 293.
- Yekpe, K.; Abatzoglou, N.; Bataille, B.; Gosselin, R.; Sharkawi, T.; Simard, J.; Cournoyer, A.; *Int. J. Pharm.* **2015**, *486*, 242.
- Koide, T.; Yamamoto, Y.; Fukami, T.; Katori, N.; Okuda, H.; Hiyama, Y.; *Chem. Pharm. Bull. (Tokyo)* **2015**, *63*, 663.
- Khorasani, M.; Amigo, J. M.; Bertelsen, P.; Sun, C. C.; Rantanen, J.; *Powder Technol.* **2015**, *300*, 120.
- Cruz, J.; Blanco, M.; *J. Pharm. Biomed. Anal.* **2011**, *56*, 408.
- Tauler, R.; *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133.
- Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R.; *Chemom. Intell. Lab. Syst.* **2005**, *76*, 101.
- de Juan, A.; Tauler, R.; *Anal. Chim. Acta* **2003**, *500*, 195.
- Franch-Lage, F.; Amigo, J. M.; Skibsted, E.; Maspocho, S.; Coello, J.; *Int. J. Pharm.* **2011**, *411*, 27.
- Rowe, R. C.; Sheskey, P. J.; Quinn, M. E. In *Handbook of Pharmaceutical Excipients*, 6<sup>th</sup> ed.; Rowe, R. C.; Sheskey, P. J.; Quinn, M. E., eds.; Pharmaceutical Press: London, 2009.
- Piqueras, S.; Burger, J.; Tauler, R.; de Juan, A.; *Chemom. Intell. Lab. Syst.* **2012**, *117*, 169.
- Terra, L. A.; Poppi, R. J.; *Chemom. Intell. Lab. Syst.* **2014**, *130*, 91.
- Monteyne, T.; Adriaensens, P.; Brouckaert, D.; Remon, J.-P.; Vervaet, C.; de Beer, T.; *Int. J. Pharm.* **2016**, *512*, 158.
- Grangeiro Junior, S.; França, L. M.; Pimentel, M. F.; Albuquerque, M. M.; de Santana, D. P.; Santana, A. K. M.; Souza, J. A. L.; Simões, S. S.; *Microchem. J.* **2015**, *118*, 252.
- Sabin, G. P.; Rocha, W. F. C.; Poppi, R. J.; *Microchem. J.* **2011**, *99*, 542.
- Sabin, G. P.; Breitzkreitz, M. C.; de Souza, A. M.; da Fonseca, P.; Calefe, L.; Moffa, M.; Poppi, R. J.; *Anal. Chim. Acta* **2011**, *706*, 113.
- Ma, H.; Anderson, C.; *J. Near Infrared Spectrosc.* **2007**, *15*, 137.
- Windig, W.; Stephenson, D. A.; *Anal. Chem.* **1992**, *64*, 2735.
- Felten, J.; Hall, H.; Jaumot, J.; Tauler, R.; de Juan, A.; Gorzás, A.; *Nat. Protoc.* **2015**, *10*, 217.
- Amigo, J. M.; Ravn, C.; *Eur. J. Pharm. Sci.* **2009**, *37*, 76.
- Windig, W.; Guilment, J.; *Anal. Chem.* **1991**, *63*, 1425.
- de Juan, A.; Tauler, R.; *Crit. Rev. Anal. Chem.* **2006**, *36*, 163.
- Alexandrino, G. L.; Khorasani, M. R.; Amigo, J. M.; Rantanen, J.; Poppi, R. J.; *Eur. J. Pharm. Biopharm.* **2015**, *93*, 224.
- Brondi, A.; Terra, L.; Sabin, G.; Garcia, J.; Poppi, R.; Trevisan, M.; *J. Near Infrared Spectrosc.* **2014**, *22*, 211.
- Barnes, R. J.; Dhanoa, M. S.; Lister, S. J.; *Appl. Spectrosc.* **1989**, *43*, 772.
- Savitzky, A.; Golay, M. J. E.; *Anal. Chem.* **1964**, *36*, 1627.
- <https://www.hypertools.org>, accessed in November 2017.
- Matlab R2016b*; MathWorks: Natick, MA, USA, 2016.
- PLS Toolbox*, v. 8.1.1; Eigenvector Research Inc.: Wenatchee, WA, USA, 2016.
- Avalle, P.; Pollitt, M. J.; Bradley, K.; Cooper, B.; Pearce, G.; Djemai, A.; Fitzpatrick, S.; Capece, M.; Barrows, J.; Davé, R. N.; *Eur. J. Pharm. Biopharm.* **2014**, *87*, 244.
- Sinchaipanid, N.; Chitropas, P.; Mitrevej, A.; *Pharm. Dev. Technol.* **2004**, *9*, 163.
- Saurina, J.; Hernández-Cassou, S.; Tauler, R.; Izquierdo-Ridora, A.; *J. Chemom.* **1998**, *12*, 183.
- de Juan, A.; Jaumot, J.; Tauler, R.; *Anal. Methods* **2014**, *6*, 4964.
- <http://www.mcrals.info/>, accessed in November 2017.
- Cruz, J.; Bautista, M.; Amigo, J. M.; Blanco, M.; *Talanta* **2009**, *80*, 473.

Submitted: November 10, 2017

Published online: March 22, 2018

