analytical chemistry

pubs.acs.org/ac Article

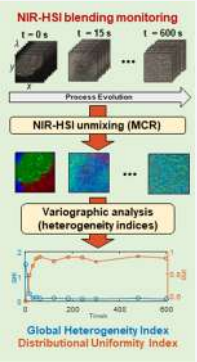
Design of Heterogeneity Indices for Blending Quality Assessment Based on Hyperspectral Images and Variographic Analysis

Rodrigo Rocha de Oliveira* and Anna de Juan*

Cite This: *Anal. Chem.* 2020, 92, 15880−15889 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Heterogeneity characterization is crucial to define the quality of end products and to describe the evolution of processes that involve blending of compounds. The heterogeneity concept describes both the diversity of physicochemical characteristics of sample fragments (constitutional heterogeneity) and the diversity of spatial distribution of the materials/compounds in the sample (distributional heterogeneity, DH). Hyperspectral images (HSIs) are unique analytical measurements that provide physicochemical and spatial information on samples and, hence, are ideal to perform heterogeneity studies. This work proposes a new methodology combining HSI and variographic analysis to obtain a good qualitative and quantitative description of global heterogeneity (GH) and DH for samples and blending processes. An initial step of image unmixing provides a set of pure distribution maps of the blending constituents as a function of time that allows a qualitative visualization of the heterogeneity variation along the blending process. These maps are used as seeding information for a subsequent variographic analysis that furnishes the newly designed quantitative global heterogeneity index (GHI) and distributional uniformity index (DUI), related to GH and DH indices, respectively. GHI and DUI indices can be described at a sample level and per component within the sample. GHI and DUI curves of blending processes are easily interpretable and adaptable for blending monitoring and control and provide invaluable information to understand the sources of the abnormal blending behavior.



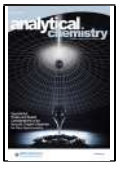
Blending process monitoring and control is an essential operation in many industrial processes. Indeed, a good blend is the necessary ground to ensure many other quality attributes linked to physical and compositional properties of manufactured products. Understanding blending means understanding heterogeneity, with all the complex aspects encompassed by this concept. The theory of sampling (TOS) by Gy provided an excellent and renewed definition of heterogeneity.[1,2] TOS distinguishes between constitutional heterogeneity (CH) and distributional heterogeneity (DH). Whereas CH focuses on the diversity of physical and chemical properties that present individual fragments of the materials in a sample, DH is focused on the quality of spatial distribution of the different materials/compounds in the sample, that is, on how far they are of presenting an even distribution. Because the DH concept is very linked to spatial correlation, studying this heterogeneity side requires looking at the properties of neighboring fragments (increments).

Traditionally, blending was controlled by off-line analysis of material increments taken every certain time or, more recently, by on-line spectroscopic monitoring using diverse sensor typologies that provide a single spectrum (or few spectra) per sample.[3−6] In most of these studies, a good blend implies that a reference composition is achieved and gets stabilized in time. Bulk sample properties are thus controlled, but the spatial distribution side linked to a good blend is overlooked. Nowadays, hyperspectral imaging (HSI) techniques work attributing a spectrum to every individual pixel in the image and, thus, connect chemical and spatial information of samples. Hence, HSI are excellent measurements for a deeper study of heterogeneity.[7,8]

The heterogeneity concept in TOS can be used to interpret this kind of information in HSI. Indeed, different heterogeneity aspects can be addressed focusing on the study of properties of individual pixels or drawing the attention to properties of neighboring pixels or neighboring pixel areas. It is very tempting associating the first approach with the concept of CH and the latter to the definition of DH. However, whereas looking at pixel areas or neighboring pixels will provide a good indication of DH, CH cannot be derived from the study of properties of individual pixels because every pixel in an HSI may offer information on one or more fragments of the material scanned. From now on and to be accurate, we will use the term global heterogeneity (GH) to design the heterogeneity information issued from the independent exam of individual pixel properties, which reflects both CH and DH, and the term DH to express the information coming from the analysis of neighboring pixels or pixel areas.

Even ignoring the TOS formulation, some attempts to use images to define the different heterogeneity aspects mentioned can be found. Thus, GH has often been defined using histograms

Received: July 30, 2020
Accepted: November 12, 2020
Published: November 25, 2020



https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889

derived from pixel image intensities or pixel concentration values issued from multivariate calibration models.[9,10] Or, approaches such as macropixel analysis, were connected to the definition of DH by studying properties of pixel neighborhood areas of different increasing sizes covering all scanned images.[11−14]

Within the TOS context, variographic analysis has been proposed to statistically study the influence of spatial correlation in heterogeneity.[1,2,15] A recent work has been published and monitors blending by using variographic analysis based on large field of view single spectroscopic measurements acquired as a function of time on the material circulating on a conveyor belt.[16] An attempt of using variographic analysis on HSI can be encountered but is limited to extract and interpret variogram parameters obtained after fitting the experimental variogram with models inspired in geostatistical theory.[17]

In our work, we have designed quantitative indices of GH and DH directly derived from the experimental HSI variograms and easy to be interpreted. The full data analysis pipeline incorporates the use of the multivariate curve resolution−alternating least squares (MCR−ALS) method on the raw image to compress HSI information and obtain the distribution maps of the pure compounds in the sample analyzed.[18−20] This step allows defining GH and DH per sample and also individually per compound, thus completing the heterogeneity description. The distribution maps are the seeding information to obtain the so-called GHI (global heterogeneity index) and the DUI (distributional uniformity index), related to GH and DH, respectively. When images are collected as a function of blending time, GHI and DUI curves provide a very good tool to understand the evolution of GH and DH along the blending process and can be potentially used for end-point blending detection or for blending control of end-products.

The indices designed are tested on simulated data and real in-house blending runs of pharmaceutical products monitored by NIR imaging. GHI and DUI curves have provided insight on the quality of the blending evolution and on the detection and characterization of blending faults at a sample and at a compound level. Although the blending runs mimic a batch process, the same methodology would apply to blending control of continuous processes.

EXPERIMENTAL SECTION

A process mimicking the blending of a solid pharmaceutical formulation was carried out using caffeine (CAF) and acetylsalicylic acid (ASA) as active pharmaceutical ingredients (APIs), both purchased at Sigma-Aldrich (a.r.), and sodium starch glycolate, Explotab (EXP) as an excipient, donated by JRS Pharma. Three batches were performed with API mass proportions of 10:1, 1:1, and 1:10 (ASA/CAF), named B1, B2, and B3, respectively. The mass fraction of EXP was kept at 15% in all batches. An approximate total mass of 0.8 g of the formulation was weighed in a 2-halves cylindrical capsule (23 mm diameter × 5 mm height). Before starting the blending process, an initial NIR HSI was collected at time = 0 s (t_0) from the capsule containing the three segregated ingredients. The closed capsule was placed in a rotating device for mixing and a total of 11 NIR images at cumulative blending times of 15, 30, 45, 60, 120, 180, 240, 300, 480, and 600 s, were recorded per batch.

The images from the pharmaceutical mixture at each blending time have been acquired with a pushbroom NIR image acquisition system Specim FX17 by Spectral Imaging Ltd., Oulu, Finland, for industrial and laboratory use. The imaging system consists of a hyperspectral camera and a 20 cm × 40 cm scanning bed. From the raw signal provided by the camera, reflectance and related absorbance spectra were calculated as explained in Section 1 of the Supporting Information.

The camera frame rate was set to 35 Hz and the scanning bed speed to 3.2 mm/s to keep an adequate aspect ratio of the image. The FX17 sensor exposure time was set to 2 ms according to the signal provided by the "white" reference to avoid saturated signals. Spectra were recorded in the 900−1700 nm NIR spectral range with a spectral resolution of 3.5 nm. The pixel size in all images is approximately 0.1 × 0.1 mm^2.

To study the reproducibility of the proposed heterogeneity indices, several images were collected from the same sample with different sensor exposure times in different days, see Section 4 of the Supporting Information for detailed experimental description.

DATA TREATMENT

The data treatment is oriented to monitor the evolution of the sample heterogeneity during a blending process. Below, the step related to MCR−ALS analysis of blending images to obtain the distribution maps of the pure ingredients of the formulation and the subsequent use of these maps to obtain heterogeneity indices based on variographic analysis per component and per sample is described.

HSI Unmixing. *Image Preprocessing.* Before data analysis, a squared area (150 × 150 pixels) from the center of each image was cropped for further analysis, which represents a sample area of ca. 15 × 15 mm. The NIR spectra of the image were preprocessed using Savitzky−Golay first derivative (second order polynomial and window size of five points)[21] for baseline correction. See Figure S1 in the Supporting Information.

Multivariate Curve Resolution−Alternating Least Squares. Image unmixing was performed with MCR−ALS, which provides the iterative decomposition of the preprocessed hyperspectral data ($\mathbf{D}$) into concentration profiles, from which distribution maps can be derived (matrix $\mathbf{C}$), and pure spectra ($\mathbf{S}^T$) of the sample constituents. Although a HSI dataset can be visualized as a three-dimensional (3D) data cube, where two dimensions (x and y) are the pixel coordinates and the third is the spectral dimension (λ), the data cube is unfolded into a two-dimensional (2D) matrix $\mathbf{D}$ with rows ($x \times y$ pixels) and columns (λ) that is decomposed according to the bilinear model in eq 1[18,20,22]

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (1)$$

where $\mathbf{D}$ is the data matrix containing the preprocessed NIR pixel spectra and $\mathbf{C}$ and $\mathbf{S}^T$ are the matrices with the concentration and spectral profiles of the pure components in the samples, respectively. $\mathbf{E}$ contains the variance not explained by the bilinear model, related to the experimental error. After the MCR−ALS resolution of the HSI dataset, the pure distribution maps of the image constituents can be obtained by folding back the stretched concentration profiles in $\mathbf{C}$ to recover the original 2D spatial structure of the image (see Figure 1).

The same bilinear model of MCR−ALS holds for multiset analysis, which consists of the simultaneous analysis of multiple images.[19,20,23,24] In this case, multiset structures $\mathbf{D}$ are built appending the submatrices $\mathbf{D}_i$ linked to the pixel spectra of the images collected in the different blending steps and three additional matrices with spectra coming from images of the pure ingredients to help in the unmixing analysis, as shown in Figure

https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889

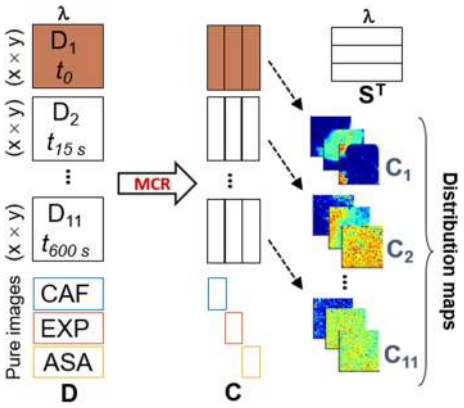


Figure 1. MCR−ALS analysis of an image multiset, where x and y are spatial pixels and λ represents the spectra wavelengths.

1. The decomposition of the multiset structure using eq 1 provides a single matrix $\mathbf{S}^T$ of pure spectra, valid for all the images analyzed, and a matrix $\mathbf{C}$, formed by as many $\mathbf{C}_i$ submatrices as images in the data set. The profiles in each of these $\mathbf{C}_i$ submatrices can be appropriately folded back to recover the related distribution maps of the images recorded at the different blending times, see Figure 1.

The MCR−ALS algorithm requires an initial estimate of either $\mathbf{C}$ or $\mathbf{S}^T$ matrices to start the iterative optimization. In this work, $\mathbf{S}^T$ was estimated based on the selection of the purest pixel spectra[25] from the matrix $\mathbf{D}$. The constrained ALS calculation of $\mathbf{C}$ and $\mathbf{S}^T$ was performed until convergence was reached.[18−20] The constraints used were normalization of pure spectra in $\mathbf{S}^T$ and non-negativity in the concentration profiles in $\mathbf{C}$.

The correspondence among species constraint, which sets presence/absence of components in the different images, was applied to the pure component images in order to decrease ambiguity in the MCR solutions and provide more accurate results[19] (see Figure 1). The MCR analysis was carried out using an in-house GUI developed under MATLAB and related routines.[26]

Design of Heterogeneity Indices Based on Variographic Analysis of Images. Heterogeneity can be very well-studied with variograms. A variogram displays the evolution of the variance as a function of a lag (expressed in time or distance units). In a variogram, the variance values are estimated by comparing pairs of observations separated at different lags.[1,2,15] Variograms can be easily adapted to explore correlation phenomena in 2D images (or 2D derived maps from 3D HSI) and, if needed, in 3D images formed by three spatial coordinates. For 2D image maps, experimental variograms are calculated comparing properties of pixel pairs separated a certain lag using the following equation

$$V(h) = \frac{1}{2}\cdot\frac{1}{N(h)}\sum_{i=1}^{N(h)}[c(x_i + h) - c(x_i)]^2 \quad (2)$$

where $V(h)$ is the variance associated with the lag (h), which is found as half of the average of the squared differences of all $N(h)$ pairs of measured pixel values $c(x_i + h)$ and $c(x_i)$ separated by a lag distance (h). Note that eq 2 expresses variance in absolute units. If the results need to be expressed in relative terms, the expression must be divided by the square of the average c value for all pixels in the image analyzed.

The variogram represents the variance estimated by the comparison of pixel pairs along the image as a function of the lag distance among the pixels compared. In this work, the $V(h)$ values of image variograms were calculated using concentration values (c) extracted from distribution maps obtained from MCR−ALS and taking the lag distance in both vertical and horizontal directions of the square image as depicted in Figure 2.

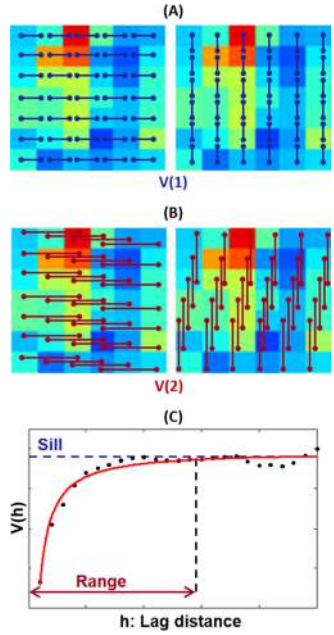


Figure 2. Representation of the pixel pairs used for the variance calculation in lag distance (a) $h = 1$ and (b) $h = 2$. (c) Representative variogram showing the extension of the correlation part, range, and the sill, linked conceptually to the variance in the absence of correlation.

Figure 2A,B displays the pairs of pixels compared to calculate $V(1)$ and $V(2)$, variances associated with a lag $h = 1$ and $h = 2$, respectively. Figure 2C shows the complete variogram obtained once $V(h)$ values are calculated for all lags from 1 until the maximum lag distance, which is set to half the number of pixels of the squared image side, that is, 75 pixels in this study.

A representative shape for a variogram obtained from a 2D image is shown in Figure 2C. When a moderate level of GH exists, typical from pharmaceutical or alimentary mixtures, neighboring pixel pairs, with small lag h, are expected to present more similar properties than pixel pairs far away from each other; therefore, variance values will be smaller for low lag distances and will increase as the lag does, until a stabilization is reached, which indicates that correlation among pixel pairs does not exist anymore. The extension of the increasing part of the variogram is called the range and represents the lag distance in which there is correlation between the pixel pairs compared. Beyond that distance, there is no correlation anymore and the variance values get very similar to each other. The range defines the extension of the spatial correlation within the image and, therefore, relates to the DH. The sill is the maximum variance in a variogram, although technically in increasing variograms is often computed as the average of variance values.[2] The sill can be used as an estimate of the GH of the sample material.

In this work, the 2D distribution maps from MCR−ALS have been chosen for variogram calculations because heterogeneity can be estimated at individual component and at a sample level. However, the same approach could use as initial information predicted pixel (c) values obtained from multivariate calibration models, pixel intensities from a specific spectral band, global

https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889

pixel intensities of all spectral range scanned, or any other type of measurement that represents a 2D map of the specific material or property to be analyzed.

To illustrate how variographic analysis can be used to study heterogeneity, Figure 3 shows the procedure followed to extract

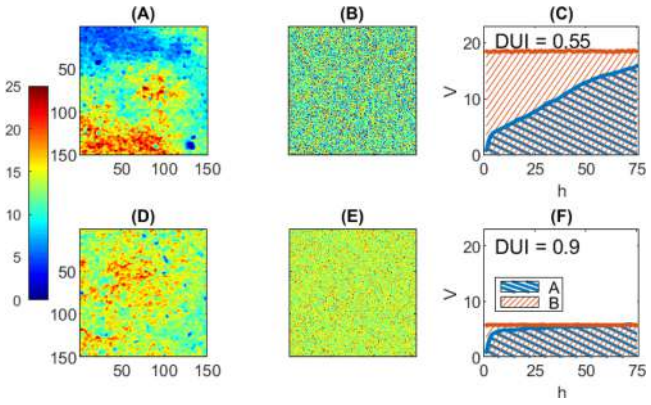


Figure 3. Distribution map of EXP at the beginning (A) and in the middle (D) of a blending process (the colorbar refers to the MCR-derived pixel concentration values in the maps). (B,E) are the randomized maps from pixels in (A,D), respectively. (C) Overlapped variograms from maps in (A), blue curve, and (B), red curve. (F) Overlapped variograms from maps in (D), blue curve, and (E), red curve. Striped blue and striped red areas are the areas under the real map variogram curves and the randomized map variogram curves, respectively.

heterogeneity indices from two distribution maps related to real images at different blending times obtained in this work. Thus, Figure 3A,D show the distribution maps of the excipient, Explotab, at the beginning and in the middle of a blending batch, respectively. In Figure 3A, the excipient was not yet uniformly mixed, as seen by large clumps with high (in red) or very low concentration (in blue) of the substance. In Figure 3D, the excipient was found to be better mixed with the other ingredients, as shown by the small clumps present in the distribution map and the narrower range in the pixel concentration values. Figure 3C and 3F show the variograms (in blue) related to maps in Figure 3A and 3D, respectively. In Figure 3C a continuously increasing variogram is obtained, with a range beyond the maximum lag distance used. On the other hand, the variogram in Figure 3F shows a shorter range, around 35 pixel distance. There is also a clear difference in the sill of both variograms, with the largest values found for the variogram related to the map in Figure 3A, at the beginning of the blending, where the variance among pixel concentration values is larger.

From a qualitative point of view, it can be concluded that both GH, linked to the sill of the variogram, and DH, linked to the range, are higher for the map in Figure 3A than for that in Figure 3D. However, there is a need for a quantitative reference indicating how far from perfect mixing, that is, minimum DH, the material of each map is. To set this ideally mixed reference, the pixels of each map were randomized, as seen in Figure 3B and 3E for the maps in Figure 3A and 3D, respectively. These randomized maps have a double advantage: (a) the GH of the real material is preserved, that is, the pixel concentration values are the same as for the real map, and (b) there is a complete lack of correlation among pixel concentration values, that is, the situation that would happen when perfect mixing is achieved and no DH is present. Figure 3C and 3F show the variograms of the randomized maps (in red) in Figure 3B and 3E, respectively. As expected, flat variograms with steady variance values are obtained for all lag distances h because the lack of correlation makes that neighboring pixels show concentration values as similar as those shown by pairs of pixels very distant from each other.

Looking at the variograms of the randomized maps, the sill of the flat variogram obtained from the randomized map in Figure 3B is higher than the map in Figure 3E because the variation in pixel concentration values is higher at the beginning of the blending process, that is, many pixels have very high or very low concentration values, whereas when the ingredients are better mixed, the pixel concentration values get more similar. This fact connects in a straightforward way with the expected decrease of GH during blending.

In addition, comparing the variograms of the original maps (in blue) with the related variograms of the randomized maps (in red), it can be observed that the shape of the real variogram gets closer to the shape of the flat variogram as blending progresses, that is, the two variograms are more similar at the middle of the process (Figure 3F), when the range for the real variogram gets shorter and variance stabilizes at earlier lags, than at the beginning of the blending (Figure 3C). This observation connects with the expected decrease of DH during blending.

Based on the previous observations of the variographic analysis of images with different degree of mixing, two heterogeneity indices are proposed using information that can be extracted from the real and randomized variograms obtained from a component distribution map, namely:

a) The GHI: estimated from the sill, that is, the average of the variance for all lags of the flat variogram from randomized maps. Actually, the sill of the flat variogram is an approximate estimation of the global variance of all pixel concentration values in the image.[27] Hence, GHI can be easily interpreted as the variance (absolute or relative) of pixel concentration values in the image.

b) The DUI: estimated by calculating the ratio of the area of the variogram obtained from the real distribution map to the area of the flat variogram derived from the related randomized map

$$\mathrm{DUI} = \frac{A}{B} \tag{3}$$

where A is the blue striped area under the variogram for the real map and B is the red striped area under the variogram for the randomized map (see Figure 3C,F).

The DUI can vary between 0 and 1 and allows quantifying the variation of DH based on variographic analysis. Experimental variograms far from their related flat horizontal variogram, as in Figure 3C, will give low DUI values, indicating high DH. Instead, experimental variograms close to its randomized map variogram provide DUI values close to 1 indicating that the mixture has low DH. In this case, the DUI values for the maps in Figure 3A and 3D are 0.55 and 0.9, respectively, meaning that 55 and 90% of ideal mixing is reached, respectively. It is relevant to note that the DUI value changes depending on the extent of the lag scale, that is, looking at Figure 3D,F, it is easy to see that if the lag scale had a limit lower than 75, the DUI values would be lower and if the images and related variograms had extended until a longer lag scale limit, the DUI values would be smaller. This means that the lag scale should adjust to the spatial scale level of heterogeneity that needs to be studied. Far from being a disadvantage, this

https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889

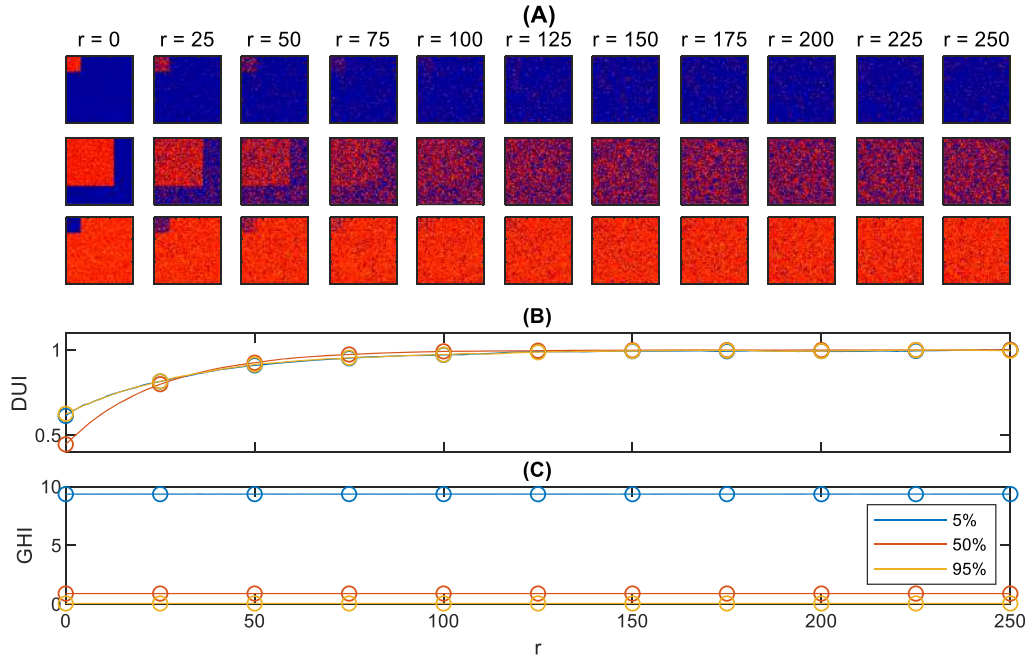


Figure 4. Simulated system I. (a) Simulated maps before mixing, $r = 0$ and at different blending steps until $r = 250$ for a bulk abundance of compound of interest equal to 5% (top), 50% (middle), and 95% (bottom); (b) DUI and (c) GHI curves calculated using the proposed variographic analysis of the generated maps for all steps. Circles represent the heterogeneity indices related to the maps shown above.

means that, if needed, the DH index can be studied at different spatial scale levels.

Both GHI and DUI indices can be obtained from variograms of individual components, but also total indices can be calculated for several selected components or all components of the formulation together. In this case, the indices are estimated from the total variogram obtained by averaging the variograms of the individual components using the following equation

$$V_T(h) = \frac{1}{nc}\sum_{j=1}^{nc} V_j(h) \tag{4}$$

where $V_T(h)$ is the variance at lag h for the total variogram, $V_j(h)$ the variance at lag h for the variogram of component j and nc is the total number of components considered in the calculation. In this work, the proposed heterogeneity indices have been calculated using the distribution maps obtained at different times of each blending process per component and per total formulation.

The evolution of these indices has been used to follow the progress of the blending process and understand faults related to it.

RESULTS AND DISCUSSION

Simulated Systems. For a better understanding of the heterogeneity indices proposed, the approach has been tested in two systems that mimic the evolution of the distribution map of an individual compound from a binary mixture during blending. In system I (see maps in Figure 4A), it is assumed that the pixel size is equal to the fragment size of the compound. The blending simulation is carried out so that the fragments of the compound of interest only change position, but the pixel concentration keeps invariant (meaning that every pixel contains the compound or not). Top, middle, and bottom lines in Figure 4A indicate situations where the compound of interest have a bulk abundance of 5, 50, or 95% in the sample. The color code in the maps is red for pixels with 100% abundance of the compound and dark blue when the abundance is 0%. A certain amount of noise has been added to the concentration values (see Section 2 of the Supporting Information for more details on the simulation).

Figure 4B shows the evolution of the DUI curves for the maps in Figure 4A. Note that the increase of the DUI index matches perfectly well the change in the spatial distribution of the compound of interest (in red). The closer the maps to a uniform distribution, the highest the DUI index. In all three systems, DUI values very close to one (ideal mixing) are found when the map is shown at $r = 125$ or higher. It is also interesting to note that DUI curves when the compound of interest is at a 5% of abundance or 95% of abundance are almost identical. This clearly proves that the DUI index, as mentioned in the data treatment section, only relates to variations in the distributional pattern of compounds, not to their concentration level (it is a concentration-independent index). Bearing this in mind, the initial maps at 5% of abundance and 95% of abundance show an identical spatial pattern, with a big 95% area with similar concentrations (either low at the 5% abundance map or high at the 95% abundance map) and a small 5% zone different from the rest. From a spatial point of view, the initial situation when the abundance of the compound of interest is 50% is worse than the previous ones because two big different regions in the map are present; hence, the lower initial DUI value.

Figure 4C shows GHI indices calculated in relative scale. As mentioned in the data treatment section, this index is calculated from the sill of the variogram of the randomized map of the image. The reason why the GHI index remains invariant during all blending processes in Figure 4A is due to the nature of the blending simulation. Remember that in this case, mixing was simulated by changing the pixel positions in the map, but not their concentration values. Hence, the variance of the pixel concentration values in the randomized maps at all blending

https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889

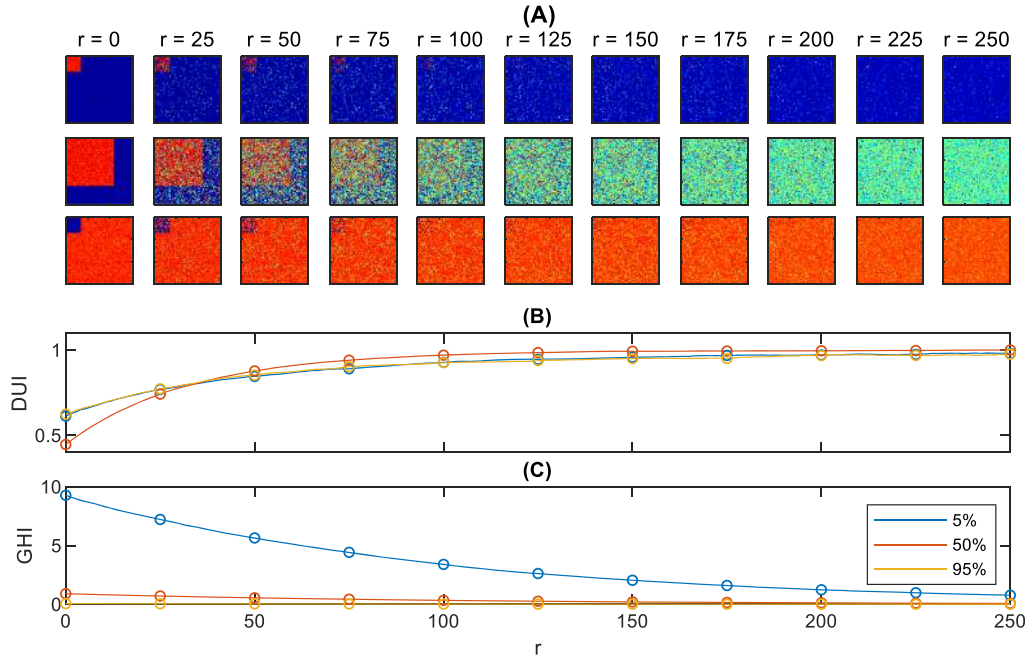


Figure 5. Simulated system II. (A) Simulated maps before mixing, $r = 0$ and after every 25 mixing steps interval until $r = 250$ for a bulk abundance of compound of interest equal to 5% (top), 50% (middle), and 95% (bottom); (B) DUI and (C) GHI curves calculated using the proposed variographic analysis of the generated maps for all steps. Circles represents the heterogeneity indices related to the maps shown above.

stages, represented by the sill of the variogram, is identical because the pixel concentration values are the same (see the invariance of the histograms of the distribution maps in Figure S4 of the Supporting Information). Note that the GHI value derived from the variogram sill is practically identical to the variance of the pixel concentration values of the distribution map studied, as shown in Figure S4 of the Supporting Information. Because GHI is given here in relative terms, the GHI is higher when the compound of interest is minor and decreases as its abundance increases.

System II shows the scenario where the pixel size can enclose several fragments of the material. As a consequence, when blending progresses, the fragments of the compound of interest change position, but a single pixel can contain fragments of the two different compounds in the binary mixture and, hence, the pixel concentration of the compound of interest can acquire different values from 0 to 100% depending on the proportion of fragments present in the pixel. As in system I, top, middle, and bottom lines in Figure 5A indicate situations where the compound of interest have a bulk abundance of 5, 50, or 95% in the sample (see Section 2 of the Supporting Information for more details in the simulation).

It is important to note in maps of Figure 5A that not only the distributional pattern gets more uniform as blending progresses, but also the pixel concentration range is reduced (see the evolution of the histograms of the distribution maps in Figure S5 from the Supporting Information). These two phenomena reflect in the DUI and GHI curves, respectively.

Figure 5B shows the DUI curves for the blendings in 5A and they are very similar to those shown in system I (Figure 4B) because the modification in distributional pattern of the blending has been done in the same manner. Again, curves for 5 and 95% compound abundance are very similar and differ from the 50% compound abundance blending.

GHI curves in Figure 5C show a clear change with respect to those in Figure 4C. Because blending causes that the pixel concentration range narrows, the variance associated with pixel concentration values, reflected by the sill of the variograms of the randomized maps, decreases and so does the GHI index. Note again that the GHI values derived from the variogram sill agree with the variances of the pixel concentration values of the distribution maps studied, as shown in Figure S5 of the Supporting Information. The decrease in these relative GHI indices happens for the three blendings studied, but the decay can be more clearly seen when the compound of interest is in a minor proportion.

The increase of the DUI curve and the decrease of the GHI curve is the behavior expected for real blending processes monitored by imaging when mixing proceeds in a correct way. Deviations from this behavior are indications of blending problems in the formulation studied or in individual compounds.

Real Blending Processes. The real blending processes studied correspond to the scenario simulated in system II, where the pixel size is clearly bigger than the fragment size of the different materials in the formulation.

Note that each of the images recorded provides a number of pixel spectra large enough to derive reliable statistical indicators and, besides, covers a sample area slightly higher than a pill size (15×15) mm^2. This means that the lag scale in the variograms will adjust to the spatial level of heterogeneity that needs to be studied.

As mentioned in the Data Treatment section, the study of the real blending processes first requires an unmixing step to obtain the pure distribution maps of the compounds in the formulation, followed by the computation of the GHI and DUI curves associated with the formulation and with each of their individual compounds.

Unmixing of NIR-HSI Data and Qualitative Evaluation of Blending Evolution. NIR HSI unmixing by MCR−ALS was carried out on a multiset structure containing a total of 36 preprocessed images, structured as a column-wise augmented

https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889

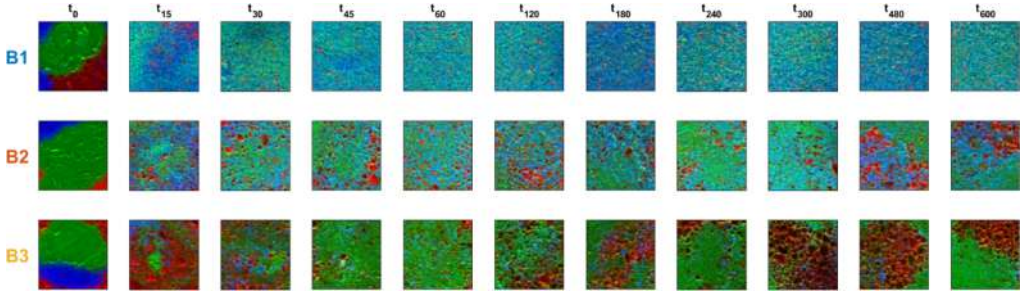


Figure 6. Combined RGB maps with overlaid pure component distribution maps obtained with MCR−ALS for batches B1 (top row), B2 (middle row), and B3 (bottom row). Red—CAF, green—EXP, and blue—ASA.

matrix. The multiset structure analyzed was formed by the preprocessed images of the three batches B1, B2, and B3, at 11 blending times each, and the three images of the pure ingredients (ASA, CAF, and starch, EXP). The pure resolved spectral profiles of the mixture formulation ingredients are shown in Figure S6. The evolution of the blending process can be qualitatively assessed by observing the MCR−ALS pure component distribution maps from the NIR HSI. Figures S7−S9 in the Supporting Information show the evolution of the distribution maps of the three ingredients, CAF, ASA, and EXP, at the 11 blending times for batch B1 (10:1—ASA/CAF), B2, and B3, respectively.

Figure 6 shows combined RGB maps overlaying the information of the three pure component maps (red for CAF, green for EXP, and blue for ASA) for batches B1, B2, and B3 in the first, second, and third row of the figure, respectively. In the blending evolution of batch B1 (10:1—ASA/CAF), the RGB map at t_0 shows the segregated ingredients before blending started. The succeeding maps, t_{15} and t_{30}, show the decrease of the segregation level, but still some clumps of pure ingredients are visible. After consecutive blending steps, from t_{45} onward, all three components were visually more evenly distributed in the imaged area. For batch B2, t_0 shows the segregated ingredients and a decrease of the segregation level is observed in distribution maps from consecutive blending times. However, at long blending times, the ingredients start to segregate, probably due to overmixing, as can be observed by the visible large clumps of different ingredients in the last two maps, at t_{480} and t_{600}. Last, batch B3 (1:10—ASA/CAF) starts with the segregated situation at t_0 and a certain blending improvement in the immediate blending times. However, from blending time t_{120} and beyond, an increase of segregation was observed mainly because of the formation of large granules of pure CAF, the major ingredient of this formulation. It is also interesting to note that the segregation behavior is different in every component. Thus, when visualizing maps from batches B2 and B3, clumps are generally associated with CAF (in red) and starch (in green), whereas ASA (in blue) seem to show a more even distribution (see maps in Figures S3−S5 for more clarity).

From a qualitative point of view, it can be observed that the blending quality decreases from batch B1 to B2, being B3 the worst blended batch. It is also seen that blending quality is compound-dependent. These different situations will be quantitatively confirmed using the heterogeneity indices proposed in this work.

Blending Process Monitoring with Image Variogram-Derived Heterogeneity Indices. In this section, the assessment of blending quality using the proposed indices related to GHI and DUI is presented.

First, a description of the blending quality of the formulation for the three batches is provided. Thus, Figure 7A shows the per sample heterogeneity indices GHI (left plot) and DUI (right plot) obtained from the total variograms (see eq 4) taking into account all formulation ingredients in the blendings. The evolution of the quantitative heterogeneity indices confirms the qualitative interpretation pointing out that batch B1 had a good blending evolution, whereas abnormal blending behaviors were detected in batches B2 and B3.

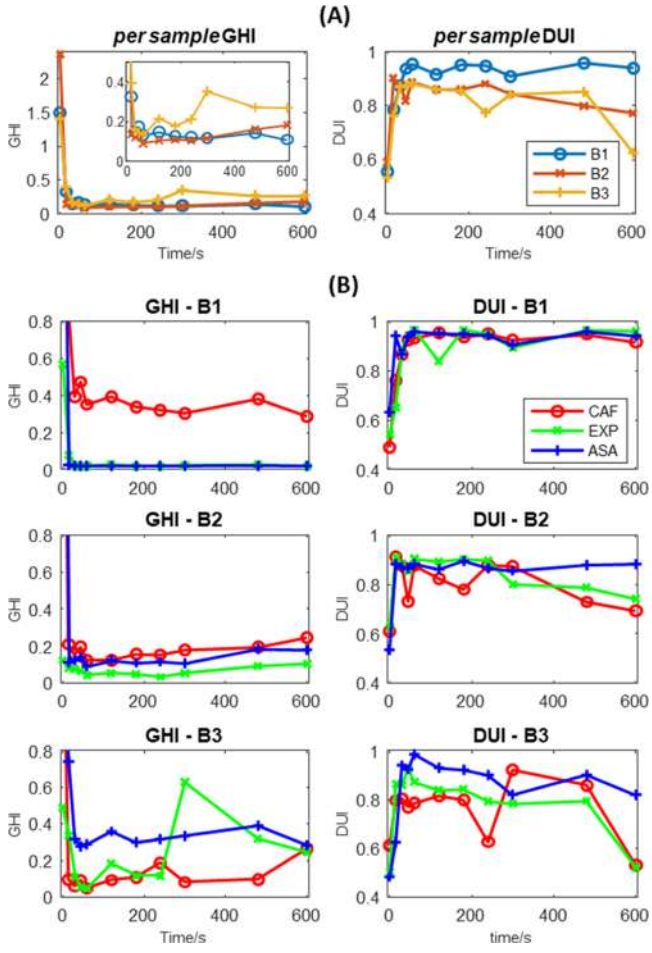


Figure 7. (A) Per sample GHI and DUI curves for blending of batches B1, B2, and B3. Inset plot zooms per sample GHI values after t_0. (B) GHI and DUI curves per component for blending batches B1, B2, and B3. Left plots, GHI curves. Right plots, DUI curves. Note that some per component GHI values are outside *y*-axis scale at the beginning of the process.

https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889

Thus, observing the per sample GHI curves in Figure 7A (left plot), a rapid drop of GHI is observed right after the blending of all batches started. That was an expected behavior considering that the formulation ingredients were completely segregated in the capsule before the start of the blending process. Per sample GHI kept decreasing for all batches until 60 s of blending time, when all batches reached a GHI, expressed in relative variance, below 0.13. After t_{60}, GHI stabilized for batch B1, but batches B2 and B3 showed an increasing trend, more visible and erratic for B3, see inset plot in Figure 7A (left plot). Per sample GHI at the end of each blending were 0.1, 0.18, and 0.26 for batches B1, B2, and B3, respectively.

Per sample DUI curves, Figure 7A (right plot), confirmed the visual interpretation of the spatial component distribution seen in maps of Figure 6B. Whereas the DUI curve had stabilized for batch B1 after the first minute of blending, a decreasing trend was observed for batches B2 and B3, with the most erratic behavior linked again to batch B3. The increase of segregation in the formulation ingredients for blends B2 and B3 shown in Figure 6B is reflected by the low DUI values obtained at the end of the blending. Indeed, the DUI value for batch B3 at t_{600} is almost as low as before the blending started at t_0.

The per sample DUI values obtained at the end of each blending at t_{600} were 0.94, 0.77, and 0.62 for batches B1, B2, and B3, respectively, meaning that the batch material reached 94, 77, and 62% of ideal mixing in the three batches.

To complement the heterogeneity description per formulation for all three batches in Figure 7A,B shows the per component heterogeneity indices for all batches studied, estimated as described in the data analysis section. In Figure 7B, the GHI curves for all compounds of batch B1 showed the same behavior as the formulation GHI curve, defined by a decrease and stabilization of GHI values. The higher GHI values obtained for CAF are related to its low concentration level in the B1 formulation, which resulted in a higher relative variance. The DUI curves of the different components of batch B1 showed that as the blending proceeded, the DH decreased and, consequently, DUI values increased for all components. Indeed, after 200 s of blending time, the DUI curves stabilized with a value circa 0.95 for all components of the formulation in B1. Both GHI and DUI curves show the expected evolution for a good blending behavior (as happened in Figure 5 for the simulated system II).

Figure 7B (middle row) shows GHI and DUI curves of batch B2 at the left and right plots, respectively. In this case, although the increasing trend of GHI after t_{60} was observed for all ingredients, slightly higher changes were associated with CAF and ASA. Regarding the DUI curves obtained per component in batch B2, Figure 7B shows that the decreasing trend of the DUI curve for the total formulation seen in Figure 7A was clearly associated only with CAF and EXP (the more even spatial distribution of ASA can be clearly seen in the individual distribution maps of this compound in Figure S8). Thus, while the DH of ASA kept stable and showed steady DUI values around 0.88 during most of the blending process, CAF and EXP decreased from a DUI value roughly equal to 0.9 at t_{60} to a value lower than 0.75 at t_{600}. This increase of DH, quantitatively represented by the decrease of the DUI value, matches the visual qualitative interpretation of CAF and EXP maps of batch B2 in Figure 6B.

Finally, for batch B3, Figure 7B (bottom left) shows that the GHI curve stabilized for the ASA component, although the value of the index remained high because of its low concentration level. Thus, the irregular behavior of the sample GHI curve for batch B3 seen in Figure 7A is mainly due to CAF, with a steady increasing tendency, and EXP, with an erratic evolution reaching a maximum GHI value at t_{300}. Figure 7B (bottom right) shows the DUI curves for the three components in batch B3. Although an even spatial distribution is not fully achieved by any component, the segregation tendency is much more clearly associated with CAF and EXP than with ASA (see separate distribution maps of this compound in Figure S9 for further clarification). Indeed, the lowest DUI values at the end of batch B3 are obtained for CAF and EXP, even though these are the two major ingredients of this formulation. The clear irregular and decreasing tendency of the DUI curves of CAF and EXP matches the emergence of large clumps of these two compounds in their distribution maps at long blending times, particularly visible at t_{600} in red for CAF and in green for EXP, as seen in Figure 6 (bottom).

There are some interesting additional remarks linked to the plots observed. GHI values can be expressed in absolute or relative variance units. When linked to pixel concentration variation, GHI values in absolute variance scale would need maps issued from a calibration-based model, for example, PLS, to obtain a useful interpretation. GHI values in a relative scale allow working with maps derived from calibration-free methodologies, such as MCR. When working with GHI values in a relative scale, it should be reminded that high relative variance values may just appear because a minor compound is studied. This effect is clearly seen in GHI curves of ASA (in blue), the compound that tends to have the best blending in all batches, where the magnitude of GHI values increases from B1 to B2 to B3, matching the decreasing content of this compound in the three batch formulations. Therefore, interpretation should not be focused only on the GHI value, which is concentration scale-dependent, but on the evolution tendency of GHI, that is, whether it gets stabilized during blending or presents an increasing or irregular tendency.

This is not the case for DUI values, which only refer to the spatial distribution pattern of compounds. In this case, compounds present in different concentration levels may reach very similar and equally good DUI values when blending is correct (see the case of the DUI curves of ASA, CAF, and EXP in batch B1, where the ratio ASA/CAF is 10:1). DUI curves, as mentioned before and proven in the simulated blendings, do not suffer from scale-dependency and can be interpreted looking both at the DUI values obtained and at the shape of the curve.

As seen throughout this work, the good performance and easy interpretability of the heterogeneity indices proposed has been proven in simulated and real blending systems. From an analytical point of view, the robustness of the quantitative values of the indices proposed has also been tested and is described in detail in Section 4 of the Supporting Information. To do so, images from the same sample showing a mixture of the same composition as B1, collected at different exposure times and in different days have been acquired. The per sample and per component values of the GHI and DUI indices show a very good reproducibility, as seen in Figures S10 and S11 of Section 4 in the Supporting Information. It is interesting to note that DUI indices are particularly stable because they are obtained from area ratios between variograms of real and randomized maps and all variability contributions other than the distributional pattern of the material are cancelled out. In the case of GHI, satisfactory values are obtained with slightly bigger fluctuations in minor compounds than in major compounds, as expectable in any analytical parameter.

https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889

To conclude, both GHI and DUI indices are needed to describe properly the blending behavior because they focus on global and DH, respectively. When blending evolution is good for all sample constituents, DUI curves reaching high and stable values tend to go with GHI curves that also remain low and stable, as happens in batch B1. However, many other situations can be encountered where there is no synchronicity between the evolution of GHI and DUI curves and strong variations in DUI values do not lead to clear changes in GHI values or vice versa, as seen in batches B2 and B3. Likewise, a complete heterogeneity description should join a per sample and a per component description because the different sample constituents do not necessarily show the same heterogeneity pattern.

CONCLUSIONS

HSI followed by image unmixing and variographic analysis provides an excellent combination to describe the global heterogeneity and DH in samples and the dynamic evolution of these attributes in blending processes. Indeed, a first visual qualitative description of heterogeneity can be extracted from the distribution maps retrieved by MCR−ALS, whereas the quantitative estimation of GH and DH is achieved through the proposed GHI and DUI, respectively.

The design of the GHI and DUI indices allows heterogeneity descriptions at a sample and component level. The assessment of blending evolution using per sample heterogeneity indices is appropriate to see the overall process evolution and to detect possible abnormal behaviors. Per sample GHI and DUI values can also be adopted as quantitative criteria to define blending quality or blending end-point by setting threshold values that need to be reached to stop a blending process, for example, a desired relative variance level for GHI and/or a preset percentage of ideal mixing for DUI. Although these indices have been tested in batch blending processes, their use can be directly transferred to monitoring and control of continuous blending operations.

The use of unmixing methods on the collected images provides maps that allow a per component description of heterogeneity through GHI and DUI indices that reflect appropriately the individual behavior of the sample or blending constituents. This individual description of heterogeneity offers additional advantages, such as a higher flexibility in blending monitoring and control protocols, for example, if there is only a single or some critical components in a blending process that need to be controlled, and contributes to a better understanding of the sources of abnormal global blending behaviors.

In general, the methodology proposed provides a good qualitative and quantitative description of heterogeneity for any kind of sample and for monitoring and control of processes that involve heterogeneity variations, such as blending operations.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c03241.

NIR imaging preprocessing steps and complete visual output of MCR−ALS analysis of the blending batches studied; thorough description of the simulated systems; and measurements and results associated with the study of the robustness of the quantitative values of GHI and DUI indices (PDF)

AUTHOR INFORMATION

Corresponding Authors

Rodrigo Rocha de Oliveira − *Chemometrics Group, Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona, 08028 Barcelona, Spain;* orcid.org/0000-0002-4309-5236; Email: rodrigo.rocha@ub.edu

Anna de Juan − *Chemometrics Group, Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona, 08028 Barcelona, Spain;* orcid.org/0000-0002-6662-2019; Email: anna.dejuan@ub.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c03241

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

R.R.d.O. acknowledges the research contract linked to the EU Horizon 2020 funds from the ProPAT project, grant no. 637232. Funding from Spanish government under the project PID 2019-1071586B-IOO is also acknowledged. The authors belong to the Catalan excellence research group (2017 SGR 753).

REFERENCES

(1) Gy, P. *Sampling for Analytical Purposes*; Wiley: West Sussex, U.K., 1998.

(2) Esbensen, K. H. *Introduction to the Theory and Practice of Sampling*; IM Publications Open, 2020.

(3) Razuc, M.; Grafia, A.; Gallo, L.; Ramírez-Rigo, M. V.; Romañach, R. J. *Drug Dev. Ind. Pharm.* **2019**, *45*, 1565−1589.

(4) Li, Y.; Anderson, C. A.; Drennen, J. K.; Airiau, C.; Igne, B. *Anal. Chem.* **2018**, *90*, 8436−8444.

(5) Durão, P.; Fauteux-Lefebvre, C.; Guay, J.-M.; Abatzoglou, N.; Gosselin, R. *Talanta* **2017**, *164*, 7−15.

(6) Igne, B.; de Juan, A.; Jaumot, J.; Lallemand, J.; Preys, S.; Drennen, J. K.; Anderson, C. A. *Int. J. Pharm.* **2014**, *473*, 219−231.

(7) Hyperspectral Imaging. In *Data Handling in Science and Technology*; Amigo, J. M., Ed.; Elsevier, 2020; Vol. 32.

(8) Ma, H.; Anderson, C. A. *J. Pharm. Sci.* **2008**, *97*, 3305−3320.

(9) Piqueras, S.; Burger, J.; Tauler, R.; de Juan, A. *Chemom. Intell. Lab. Syst.* **2012**, *117*, 169−182.

(10) Burger, J.; Geladi, P. *Analyst* **2006**, *131*, 1152−1160.

(11) Hamad, M. L.; Ellison, C. D.; Khan, M. A.; Lyon, R. C. *J. Pharm. Sci.* **2007**, *96*, 3390−3401.

(12) Rosas, J. G.; Blanco, M. *J. Pharm. Biomed. Anal.* **2012**, *70*, 680−690.

(13) Sacré, P.-Y.; Lebrun, P.; Chavez, P.-F.; Bleye, C. D.; Netchacovitch, L.; Rozet, E.; Klinkenberg, R.; Streel, B.; Hubert, P.; Ziemons, E. *Anal. Chim. Acta* **2014**, *818*, 7−14.

(14) de Moura França, L.; Amigo, J. M.; Cairós, C.; Bautista, M.; Pimentel, M. F. *Chemometr. Intell. Lab. Syst.* **2017**, *171*, 26−39.

(15) Esbensen, K. H.; Friis-Petersen, H. H.; Petersen, L.; Holm-Nielsen, J. B.; Mortensen, P. P. *Chemom. Intell. Lab. Syst.* **2007**, *88*, 41−59.

(16) Sánchez-Paternina, A.; Sierra-Vega, N. O.; Cárdenas, V.; Méndez, R.; Esbensen, K. H.; Romañach, R. J. *Comput. Chem. Eng.* **2019**, *124*, 109−123.

(17) Herrero-Langreo, A.; Gorretta, N.; Tisseyre, B.; Gowen, A.; Xu, J.-L.; Chaix, G.; Roger, J.-M. *Anal. Chim. Acta* **2019**, *1077*, 116−128.

(18) de Juan, A.; Rutan, S. C.; Tauler, R. Two-Way Data Analysis: Multivariate Curve Resolution, Iterative Methods. In *Comprehensive Chemometrics*; Brown, S., Tauler, R., Walczak, B., Eds.; Elsevier, 2019; pp 153−171.

(19) Tauler, R.; Maeder, M.; de Juan, A. Multiset Data Analysis: Extended Multivariate Curve Resolution. In *Comprehensive Chemo-*

metrics; Brown, S., Tauler, R., Walczak, B., Eds.; Elsevier, 2020; Vol. *2*, pp 305−336.
(20) de Juan, A.; Tauler, R. *Anal. Chim. Acta* **2020**, DOI: 10.1016/j.aca.2020.10.051.
(21) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627−1639.
(22) de Juan, A. Multivariate Curve Resolution for Hyperspectral Image Analysis. In *Hyperspecral Imaging*; Data Handling in Science and Technology; Amigo, J. M., Ed., 2020; Vol. *32*, pp 115−150.
(23) Piqueras, S.; Duponchel, L.; Tauler, R.; de Juan, A. *Anal. Chim. Acta* **2014**, *819*, 15−25.
(24) de Juan, A.; Gowen, A.; Duponchel, L.; Ruckebusch, C. Image Fusion. In *Data Handling in Science and Technology*; Elsevier, 2019; Vol. *31*, pp 311−344.
(25) Windig, W.; Guilment, J. *Anal. Chem.* **1991**, *63*, 1425−1432.
(26) Jaumot, J.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.* **2015**, *140*, 1−12.
(27) Gy, P. *Chemom. Intell. Lab. Syst.* **2004**, *74*, 39−47.

https://dx.doi.org/10.1021/acs.analchem.0c03241
Anal. Chem. 2020, 92, 15880−15889