Received: 29 August 2022 | Revised: 0 xxxx 0000 | Accepted: 7 October 2022

DOI: 10.1111/jmi.13150

Journal of Microscopy RMS

ORIGINAL ARTICLE

Segmentation of SEM images of multiphase materials: When Gaussian mixture models are accurate?

Manolis Chatzigeorgiou[1,2] | **Michalis Vrigkas**[3] | **Margarita Beazi-Katsioti**[2] | **Marios Katsiotis**[4] | **Nikos Boukos**[1] | **Vassilios Constantoudis**[1]

[1]Institute of Nanoscience and Nanotechnology, National Center for Scientific Research 'Demokritos, Patr. Gregoriou E & 27 Neapoleos Str, Athens, Agia Paraskevi 15341, Greece

[2]School of Chemical Engineering, National Technical University of Athens, Iroon Polytechneiou 9 str, Athens, Zografou 157 72, Greece

[3]Department of Communication and Digital Media, University of Western Macedonia, Kastoria, Area Fourka 52100, Greece

[4]Group Innovation & Technology, TITAN Cement S.A., TITAN Cement S.A., 22A Halkidos Street, Athens, 111 43, Greece

Correspondence
Manolis Chatzigeorgiou, Institute of Nanoscience and Nanotechnology, N.C.S.R. Demokritos, P.O. Box 60037, 15310 Agia Paraskevi, GREECE.
Email: e.chatzigeorgiou@inn.demokritos.gr

Funding information
Stavros Niarchos Foundation, Grant/Award Number: E.E.-12149

Abstract

Scanning electron microscopy has been a powerful technique to investigate the structural and chemical properties of multiphase materials on micro and nanoscale due to its high-resolution capabilities. One of the main outcomes of the SEM-based analysis is the calculation of the fractions of material components constituting the multiphase material by means of the segmentation of their back scattered electron SEM images. In order to segment multiphase images, Gaussian mixture models (GMMs) are commonly used based on the deconvolution of the image pixel histogram. Despite its extensive use, the accuracy of GMM predictions has not been validated yet. In this paper, we proceed to a systematic study of the evaluation of the accuracy and the limitations of the GMM method when applied to the segmentation of a four-phase material. To this end, first, we build a modelling framework and propose an index to quantify the accuracy of GMM predictions for all phases. Then we apply this framework to calculate the impact of collective parameters of image histogram on the accuracy of GMM predictions. Finally, some rules of thumb are concluded to guide SEM users about the suitability of using GMM for the segmentation of their SEM images based only on the inspection of the image histogram. A suitable histogram for GMM is a histogram with number of peaks equal to the number of Gaussian components, and if that is not the case, kurtosis and skewness should be smaller than 2.35 and 0.1, respectively.

KEYWORDS
accuracy prediction, Gaussian mixture models, scanning electron microscopy, segmentation

1 | INTRODUCTION

Material characterisation has provided very important tools enabling advancements in material science and technology. The study of micro- and nanostructures and the plethora of morphologies they may exhibit has led to the discovery of new composite materials and, at the same time, allows the precise quality control of material industry products.

A large number of composite materials fall into the category of multiphase materials consisting of more than one phase. The coexistence of phases in these materials causes the emergence of new physicochemical properties which can be exploited in a large variety of applications. An indicative example of a multiphase crystalline material is the ordinary Portland cement (OPC) clinker.[1] The presence and proportion of several crystallographic phases within the OPC clinker are responsible for the early and

later hydraulic hardening properties of the cement. This is also true in most multiphase materials and therefore, the accurate quantification of phase proportions is of vital importance to optimise their performance.

A widely used imaging technique in material characterisation is Scanning Electron Microscopy (SEM). SEM is capable of extracting multiple information about materials such as porosity,[2] spatial distribution of crystallographic phases, and grain size to name just a few. This is achieved by the utilisation of the multiple detectors of SEM such as secondary electron, backscattered, or energy dispersive X-ray spectroscopy detectors. Despite their usefulness, some of the above techniques are very demanding concerning sample preparation and measurement time. A fast SEM technique that is widely used to capture information about the elemental composition of the sample with a high spatial resolution is backscattered electron imaging (BSE). The greyscale intensity of BSE images is affected by the Z-contrast of surface elements and the topography of the analysed sample.[3] The extraction of quantitative results from the acquired BSE images requires the application of a supplementary image analysis method in order to separate segments of the image that are attributed either to different crystallographic phases (multiphase materials) or to pores (porous materials). There are numerous segmentation methods for SEM images proposed in the literature. These methods may be either supervised[4–6] or unsupervised.[7,8] Supervised segmentation models are trained in a specific data set and therefore have limited applicability taking into account the large spectrum of used acquisition parameters and the differences in the composition of analysed samples. In addition, supervised segmentation requires annotated data sets for their training which are not easily available, and their collection increases the workload of the segmentation process. Therefore, unsupervised segmentation methods may be more suitable to apply in BSE images. There are a lot of unsupervised segmentation methods proposed for image segmentation such as GMM, Otsu multi-threshold,[9] watershed methods,[10] etc. GMM is one of the most widely used segmentation methods which has also been applied in BSE images.[11]

In this paper, we consider the application of GMMs only on greyscale intensities without taking into account any other spatial image aspects. The main question addressed here concerns the limitations of the accurate application of GMM segmentation in SEM images of multiphase materials. To this end, an extensive parameter study is performed in synthesised SEM images endowed with a developed neural network trained to accelerate the evaluation process. The practical outcome of this study is to propose a fast and easily implemented method for the prediction of the GMM accuracy based on the analysis of the real image histogram alone so that an SEM user could take rational decisions about the optimum segmentation strategy which should be applied in SEM image analysis.

The contributions of this paper are the proposal of a new metric for the accuracy of GMM deconvolution, a thorough study on the effects of different image histogram parameters on GMM accuracy and finally a rule of thumb for the prediction of GMM accuracy based on histogram shape.

The paper starts with the validation of the basic assumption of GMM, namely the Gaussian character of single-phase greyscale histograms (Section 2). In the same section, a short description of the GMM method is given along with a reference on its limitations. In Section 3, we present a novel method for the evaluation of GMM segmentation accuracy while Section 4 describes the algorithms for the generation of synthesised SEM images and histograms with predetermined phase proportions which will be used in the validation of GMM. The results of our work concerning the prediction of the accuracy of GMM-based segmentation for a large spectrum of involved parameters are presented and discussed in Section 5. The study of marginal histogram parameters such as skewness and kurtosis is the subject of Section 6. The paper summarises the main findings and draws conclusions in the last Section 7.

2 | GMM: ASSUMPTIONS, LIMITATIONS AND DESCRIPTION

BSE images can be thought of as a set of measurements of pixel intensities neglecting any spatial information. Under the assumption that every phase contributes with a Gaussian distribution to the overall greyscale distribution of the image, GMM may be a useful tool for the accurate deconvolution of image greyscale distribution. A result of image histogram deconvolution can be the quantification of the proportion[12] or can be an initialisation tool for a probabilistic segmentation method.[13] The basic assumption of GMM is that the Gaussian distribution describes sufficiently the greyscale intensities of all single-phase material BSE images. To justify this assumption, a typical example of a histogram of a BSE image of a single-phase material is shown in Figure 1. In particular, we display the histogram of the pixel intensities of a BSE image of a SiO_2 layer with a thickness of 2 μm grown on a Si wafer. As usual, before measurement, the SiO_2 layer has been sputtered by gold to avoid charging. As we can see with the red line in Figure 1, the experimental histogram is nicely fitted with a Gaussian distribution. A similar success of Gaussian fitting has also been found for histograms of images acquired using other SEM settings and sample materials.

The only exception that undermines the Gaussian assumption and therefore the application of GMM comes from the cases where the histograms lie very close to the

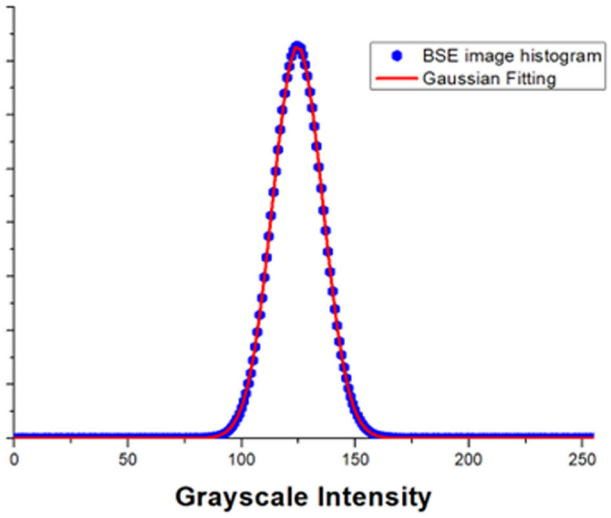

FIGURE 1 The greyscale histogram of a representative BSE image depicted a Gold sputtered layer deposited on a layer of SiO_2 (2 μm) developed on a silicon wafer (blue dotted line). Also, the optimum Gaussian fit is shown revealing that greyscale intensities of BSE images of flat samples are described by Gaussian distributions.

extremes of pixel intensity range (in the case of an 8 bit depth image 0 or 255) posed by the limited sensitivity of BSE detector. In such cases, a large fraction of pixels possess the extreme intensity values of 0 or 255 causing an artificial sharp peak in the histogram which cannot be captured by a Gaussian fitting (see Figure 2A). This issue imposes the first limitation in the application of GMM segmentation method since the basic assumption of single-phase Gaussian distributions is violated.

A second limitation occurs when the distributions of different phases are strongly overlapped to compose a unimodal distribution. This happens when the mean distribution values are close in respect to the variances or some distribution has much less weight than others and is totally overlapped within a major one. These limitations have been extensively studied in the literature.[14,15,16]. An example of this limitation is schematically shown in Figure 2B where the GMM method has been applied in the segmentation of two strongly overlapped distributions with poor outcomes.

The implementation of the GMM in this work is realised with the expectation-maximisation (EM) algorithm. Histograms are univariate data and hence the brief description that follows concerns univariate EM-GMM. In general, GMMs are probabilistic models that assume that every point of the analysed data set is generated from a set of weighted Gaussian distributions. Let $X = \{x^n\}_{n=1}^N$ be the set of pixel intensities, or in general pixel feature vectors, corresponding to a single image. Viewing the required segmentation as a clustering problem on X, we can assume that the x^n are independent, identically distributed and that they are generated by a finite mixture model. This probability $p(x^n)$ is the probability of x^n pixel intensity to appear in the image and it is given by Equation (1) for fixed parameters of the model:

$$p\left(x^n\right) = \sum_{i=1}^{K} \pi_i^n \mathcal{N}\left(x^n | \mu_i^n, \sigma_i^n\right), \tag{1}$$

where π_i^n are the Gaussian coefficients (or weights) that represent the probability the nth pixel to belong to the $i$thcluster and satisfy the constraints: $0 \le \pi_i^n \le 1$, and $\sum_{i=1}^{K} \pi_i^n = 1$. Also, the parameters μ_i^n and σ_i^n represent the mean value and the variance of the ith Gaussian distribution $\mathcal{N}$, respectively, given by the following formula:

$$\mathcal{N}\left(x^n | \mu_i^n, \sigma_i^n\right) = \frac{1}{\sigma_i^n \sqrt{2\pi}} e^{-\frac{\left(x^n - \mu_i^n\right)^2}{2\sigma_i^2}}. \tag{2}$$

From now on for simplicity, we omit superscript n. Under the assumption that histogram is a sum of K Gaussian distributions and parameters of all distributions are described by $\Theta = \{\pi_i, \mu_i, \sigma_i\}_{i=1}^K$, then the probability of any pixel to take on intensity value X can be given by Equation (3):

$$P\left(X = x | \Theta\right) = \sum_{j=1}^{K} \pi_j \mathcal{N}\left(X = x | \mu_j, \sigma_j\right). \tag{3}$$

The best model of Gaussian distributions that can describe the analysed data set is given by the model that maximises the log-likelihood:

$$\max_{\Theta} \log P\left(X | \Theta\right) = \max_{\Theta} \sum_{j=1}^{N} \log\left(P\left(x_j | \Theta\right)\right)$$

$$= \max_{\pi,\mu,\sigma} \sum_{j=1}^{N} \log\left(\sum_{i=1}^{K} \pi_i \mathcal{N}\left(x_j | \mu_i, \sigma_i\right)\right), \tag{4}$$

where N is the number of pixels.

Consequently, the expectation and the maximisation steps are optimisation steps in order to find the optimum model parameters.

According to Xu and Jordan[17] in the so-called E-step, a quantity called responsibility γ is computed given by the equation:

$$\gamma_{ij} = \frac{\pi_i \mathcal{N}\left(xj | \mu_i, \sigma_i\right)}{\sum_{i=1}^{K} \pi_i \mathcal{N}\left(xj | \mu_i, \sigma_i\right)}. \tag{5}$$

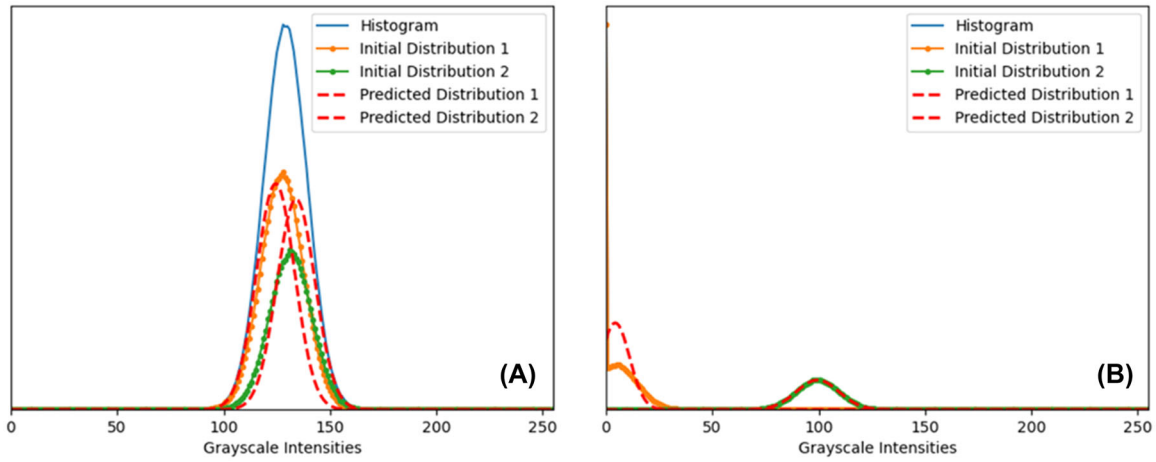

FIGURE 2 (A) A synthetic BSE image histogram is shown (blue line) composed of two Gaussian distributions (discontinuous lines) and GMM predicted distributions (dotted lines). (B) The inconsistency of GMM predicted distributions and the real distributions could be caused by overlapping of distributions (A) or by the loss of normality (B).

In the maximisation step, an update of the Gaussian mixture parameters occurs:

$$\mu_j^{new} = \frac{1}{N_k}\sum_{i=1}^{N}\gamma_{ij}x_i, \tag{6}$$

$$\sigma_j^{new} = \frac{1}{N_k}\sum_{i=1}^{N}\gamma_{ij}\left(x_i - \mu_j^{new}\right)^2, \tag{7}$$

$$\pi_j^{new} = \frac{N_j}{N}, \tag{8}$$

where $N_j = \sum_{i=1}^{N}\gamma_{ij}$.

Since the maximum of log-likelihood is not known, the algorithm stops when log-likelihood converges when:

$$\left|\log p\left(X|\Theta'\right) - \log p\left(X|\Theta\right)\right| < tol, \tag{9}$$

where *tol* is a tolerance parameter of the algorithm. If Equation (9) is not satisfied, E- and M-steps are repeated. After convergence of the EM, the data are assigned to the component with the highest degree of responsibility based on the maximum value of the posterior probabilities using the maximum a posteriori (MAP) principle. It must be mentioned that the initialisation of the GMM parameters in this work is realised with the k-means algorithm.[18] The choice of this initialisation method is realised taking into consideration that is one of the most widely-used algorithms for this task.[19]

3 | EVALUATION OF GMM ACCURACY

The accuracy of GMM deconvolution can be measured as the convergence of the predicted distributions with respect to the initial ones. Measures such Kullbach–Leibler divergence[20] or Mahalanobis distance could be utilised to this end.[21] In this work though requirements such as the bounded nature of such quantities and fast computation time lead to the development of a novel index. The proposed index is named distribution similarity index (DS) and is able to quantify the convergence of two distributions. This is realised by quantifying the overlapping of the predicted distribution (PD) with the real one (RD). In more detail, this ratio is calculated as the minimum between the two ratios of a) the overlapping area divided by the real distribution and b) the overlapping area divided by the predicted distribution.

$$DS = \min\left(\frac{overlapping\ area}{predicted\ distribution\ area}, \frac{overlapping\ area}{real\ distribution\ area}\right). \tag{10}$$

In the case of a perfect convergence of two distributions (full similarity), the overlapping area coincides with the areas of the predicted, and the real distributions and therefore $DS = 1$. Whereas when the convergence deteriorates, the overlapping of distributions is reduced and DS gets values lower than 1 (Figure 3A).

We propose to use the minimum of the two ratios of the overlapping area over the predicted and real distribution area respectively, in order to avoid misleading characterisation of distribution similarity when a full overlapping is not

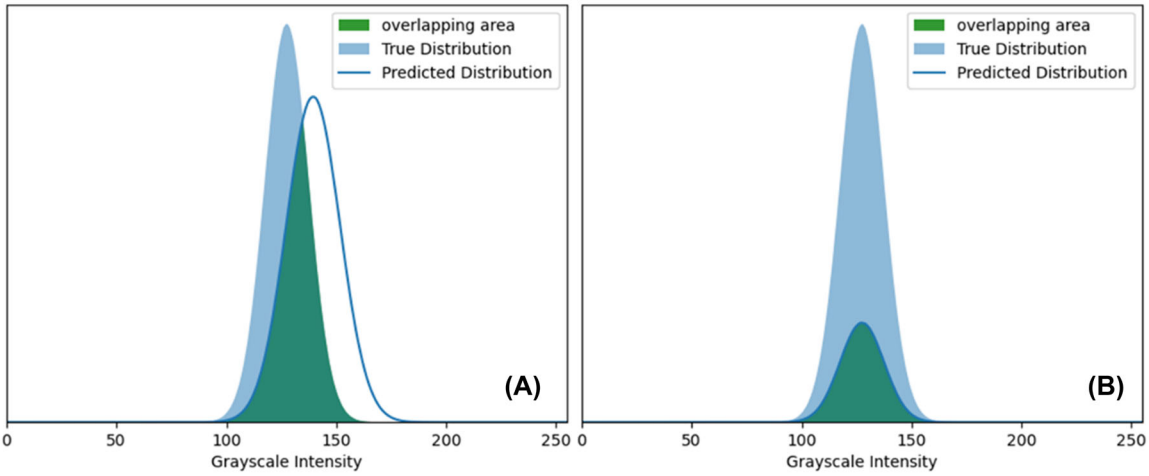

FIGURE 3 (A) Two overlapping Gaussian distributions in which areas are very similar. (B) Two overlapping Gaussian distributions in which areas differ. In the first case (A) ratios of the overlapping area with respect to the distribution area would be invariant under the choice of distribution true or predicted. However, in (B) this choice would affect the value of the ratio and hence could be a deceptive measure of distribution convergence without calculation and the choice of minimum between the two ratios.

accompanied by area similarity too as shown in Figure 3B. In such cases, the use of a single ratio in the definition of DS instead of the minimum could lead to the deceptive value $DS = 1$ although the distributions are apparently different. Hence the use of the minimum ratio (i.e. the denominator with the maximum area) in Equation (10) is critical to provide a correct quantification of the similarity of the GMM predicted distribution with the real one.

TABLE 1 Gold layer thickness (1st column) and backscattered coefficient of the corresponding sample (2nd column)

Au (nm)	Backscattered coefficient
0	0.098542
5	0.121918
12	0.147930
20	0.176482
25	0.193902

4 | GENERATION OF SYNTHETIC AND HYBRID-SYNTHETIC DATA

The evaluation of the accuracy of the GMM-based deconvolution of a multiphase histogram requires its application in cases where the parameters of the contributing single-phase distributions are known before. This knowledge is hard to be justified in real BSE images and therefore, the use of synthetic images with predefined single-phase histograms can provide a means for GMM evaluation. For this reason, we generate synthetic BSE image histograms by using generators of random integers normally distributed provided in the NumPy library.[22] These algorithms allow the independent control of mean value, variance and the number of integers generated for each single-phase distribution.

To generate realistic BSE image histograms, we have performed a thorough investigation of the dependence of histogram parameters such as mean values and variances on BSE image acquisition parameters and their correlations. To this end, a silicon wafer sample has been coated with a 2 μm layer of SiO_2 and on top of that, gold layers of different thicknesses were sputtered to change at will the mean atomic numbers, in more detail, the thickness of the gold layers are approximately 0, 5, 12, 20 and 25 nm. These stacks can be used to obtain BSE images with different backscattered coefficients. For the calculation of backscattered coefficients of measured stacks, the Monte Carlo simulation method was employed on the trajectory of electrons using the CASINO software.[23] In Table 1 the calculated backscattered coefficients are shown with respect to Au layer thickness.

These images were obtained with an FEI Quanta INSPECT SEM with beam current measured with a Faraday cage and kept constant at 2 nA, while beam acceleration voltage was 15 kV. The working distance was set at 10 mm while a defocus of +5 mm was used in order to exclude topology contrast. The resolution of BSE images is 4096 × 3535. In Table 2, representative BSE image statistical moments are presented for indicative contrast and brightness settings.

First, the Gaussian character of BSE images, except the oversaturated ones, is evident since the high statistical moments' skewness and kurtosis are close to 0 and 3 respectively. A thorough study of image acquisition conditions, such as contrast, brightness and dwell time,

TABLE 2 Statistical moments of different thickness layers of gold spattered on a 2 μm of SiO_2 on a Si-wafer acquired with 2 different contrast and brightness settings

	Contrast = 55 Brightness = 72					Contrast = 60 Brightness = 65				
Gold thickness (nm)	**0**	**5**	**12**	**20**	**25**	**0**	**5**	**12**	**20**	**25**
Mean	84.43	99.68	118.99	144.63	160.22	20.17	46.62	80.18	124.57	151.42
Variance	2.68	3.23	3.93	4.838	5.41	7.56	9.52	11.60	14.29	15.94
Skewness	0.01	0.01	0.01	0.01	0.01	0.05	0.01	0.01	0.01	0.01
Kurtosis	3.00	3.00	3.00	3.00	3.00	−3.02	3.00	3.00	3.00	3.00

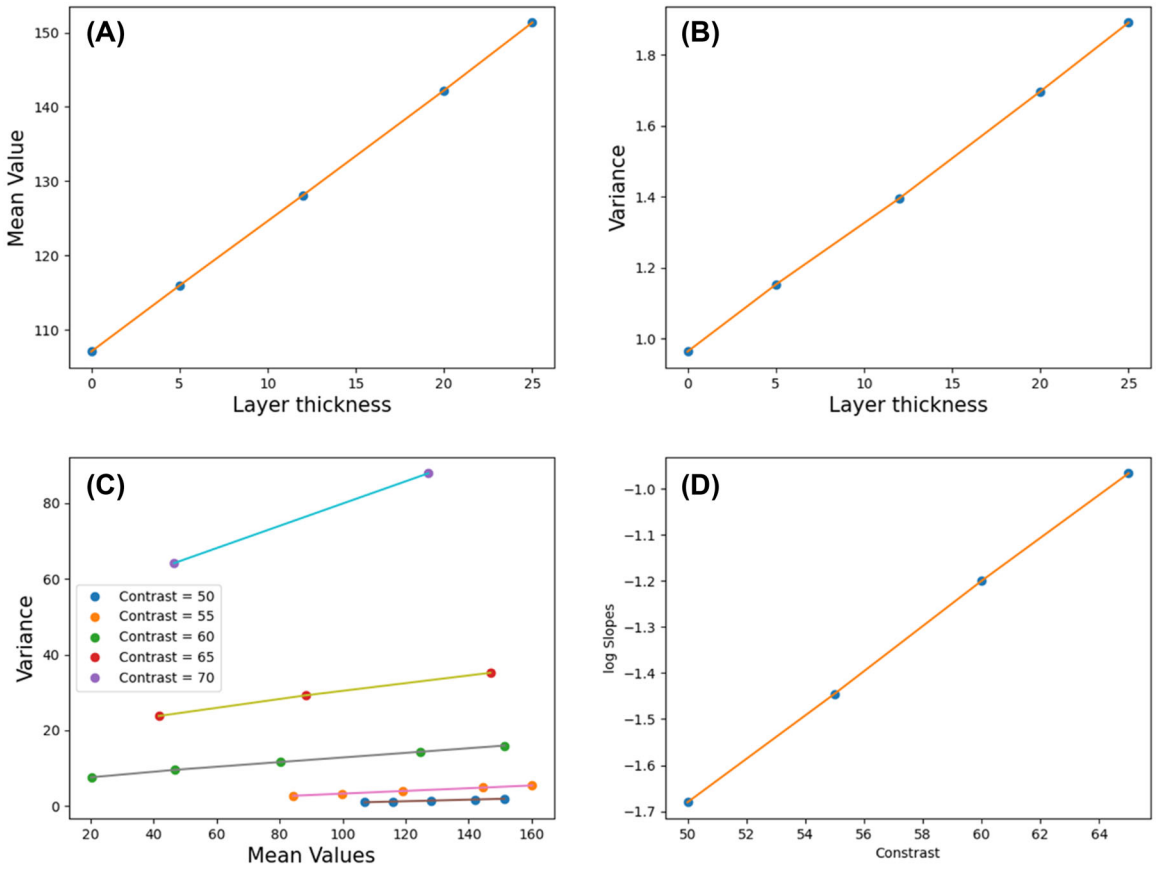

FIGURE 4 (A) Mean greyscale intensity values versus the thickness of gold layer sputtered for contrast value 55. (B) Variance of greyscale intensity values versus the thickness of gold layer sputtered contrast value 55. (C) *Variance* vs. *Mean* intensity values for several gold layer thicknesses and contrasts 50, 55, 60, 65 and 70. It must be mentioned that in high contrast values the oversaturation of images deters the valid computation of mean and variance values and hence are not shown in the graph. (D) Slopes of *Variances* vs. *Mean* values in logarithmic scale vs. contrast values.

reveals their correlation to the Gaussian parameters. Decreasing dwell time is expected to have an impact on the noise level of the greyscale image, resulting in an increase in the variance of its intensity distribution. Another anticipated effect is that the increase of contrast also increases the variance. On the other hand, while an increase of sample Z effective is expected to result in an increase in the mean greyscale distribution value (Figure 4A), the variance is also increased (Figure 4B). In particular, as the gold overlayer thickness increases corresponding to a higher Z effective of the sample, an increase in the greyscale distribution variance is noticed. Such behaviour could be attributed to the Poisson distribution (shot noise) of the SEM thermionic electron gun emission. As more electrons are backscattered due to a higher value of sample Z effective, any variations in the number or the energy of these electrons are depicted on the number of the detected electrons. Since the main source of noise in BSE is shot noise,[24] this behaviour is justified. However, shot noise is not the only source of noise that contribute to the total

distribution of the depicted greyscale.[25] The overall noise can be calculated with the application of the central limit theorem, resulting in a Gaussian intensity distribution. Still, as the above-described Poisson process is the main noise component, similarities with a Poisson distribution are observed in the greyscale intensity. More specifically the variance of the intensity distribution depends linearly on the distribution mean value, as can be seen in Figure 4C, while the linear slope is controlled by the experimental contrast setting (Figure 4D).

Additionally, linear correlation emerges between mean value and variance in Figure 3C. The slope of variances and mean values with changing contrast, obey a semi-logarithmic relation with contrast values of the microscope. Those two correlations were taken into account in the generation of hybrid synthetic BSE image histograms.

5 | RESULTS: PARAMETRIC STUDY

The histogram of the BSE images of a multiphase material with N_p single phases will be determined by $3N_p - 1$ parameters given that each phase contributes a Gaussian distribution characterised by three parameters (mean intensity value, variance and weight) and the sum of weights should be fixed. In this work, we are limited to multiphase materials with four phases ($N_p = 4$) and therefore the image histograms are determined by 11 parameters. The high dimensionality of the parameter space challenges the comprehensive sampling of parameters and dictates more targeted choices. To this end, we reduce the impact of four mean distribution values with the definition of a collective parameter called *Range* which defines the intensity difference between the higher and the lower mean distribution value. The rest mean values of the other two phases are assumed to be equally distributed within *Range* (see Figure 5). In addition, we assume that all distribution variances are equal and described by the parameter *Var*. In real-synthetic images, *Var* is calculated utilising the linear relation of mean values and variances extracted in the BSE measurements mentioned above (see Table 1). Finally, for the weight parameters of distributions, we consider two alternatives. (a) First, we set all weights equal to each other ($w_i = 0.25$, with $i = 1, .., 4$), and (b) one of the external weight distributions varies from 0 to 1 while the other three distribution weights are left unchanged and determined by the restriction that in every configuration the sum of all weights should be equal to one. A schematic representation of the greyscale distributions of four phases along with the meaning of the collective parameters *Range*, *Variance* and *Weight* used in our analysis is shown in Figure 5.

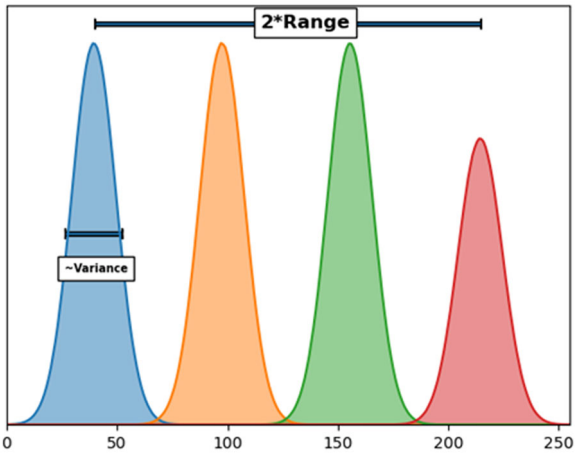

FIGURE 5 Schematic representation of the Gaussian distributions of four phases indicated with different colours. The meaning of the parameters *Range* and *Variance* (*Var*) are straightforwardly shown while the *Weight* of each phase is quantified by the area under the corresponding Gaussian distribution. In our analysis, these parameters define sufficiently the phase Gaussian distributions since we consider equidistant mean values, identical variances, and weights which are equal for 3 out of 4 distributions

5.1 | First case study: equiprobable distributions

A complete map of the dependence of the DSs of all distributions on Range and Variance for the case of equiprobable distributions ($w_i = 0.25$) is shown in the contour graphs of Figure 6. In all distributions, the map is dominated by a large area that corresponds to high DS values (larger than 0.95) shown in yellow colour. In this area, the Range and Variance parameters define total histograms characterised by the presence of four well-distinguished peaks (see insets in Figure 6) which can be deconvoluted into distributions almost identical to the real ones ($DS > 0.95$). This area of GMM success extends in ranges from about 40 to 100 for large variances while it enlarges when Variance tends to zero. As we are moving towards smaller (<40) or larger (>100) Ranges, DSs of all distributions get lower values quantifying the accuracy limits of GMM predictions. At small Ranges, the degradation of GMM predictions is due to the strong overlapping of distributions which has been described as the second limitation of GMM accuracy in Section 2. Indeed, DS starts to lower at values smaller than 0.95 when the total histogram becomes unimodal, and the four peaks are merged into a broad peak of a unimodal histogram (see the inset histograms in Figure 6). On the other side, at large Ranges, the external distributions touch the extremes of the greyscale spectrum of values (0 and 255) and their pixels degenerate on these creating

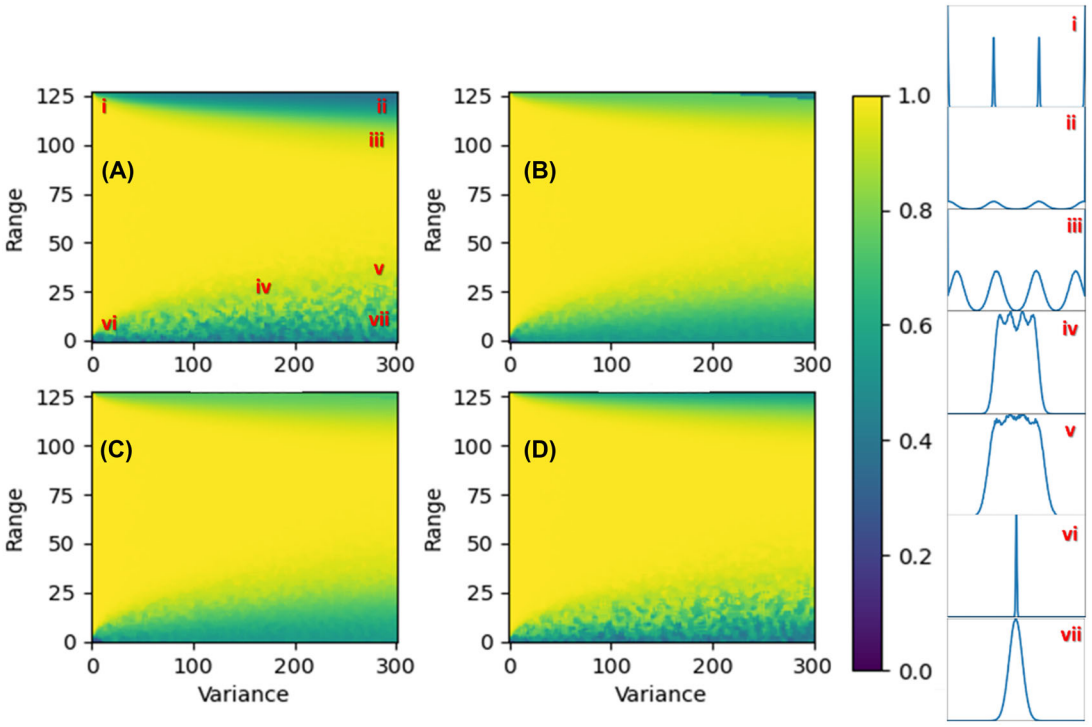

FIGURE 6 *DS* versus *Range* and *Variance* for all four distributions as calculated for a constant and equal weight of all distributions. (A, B, C, D) For a better understanding of this figure, a set of histograms are shown along with their positions in the *Range* and *Variance* parametric space.

gradually steep high peaks (first limitation in Section 2). These peaks undermine the Gaussian hypothesis of distributions and therefore the successful application of GMM.

Besides the overall similarity of DS maps in all distributions, one should notice some remarkable differentiations. Not surprisingly, the GMM predictions for the external distributions (1 and 4 in Figure 6) are more vulnerable to the degradation effects of pixel degeneration at the extreme values 0 and 255 since these distributions reach first the extremes of the greyscale spectrum and deviate from Gaussian shape. This difference explains the decrease of DS at lower Ranges in external distributions as shown in Figure 6 (cf. Figure 6A,D with Figure 6B,C). Concerning the small Ranges, again the DSs of the external distributions 1,4 exhibit different behaviour. Here, they exhibit strong fluctuations versus both parameters Range and Variance contrary to the smooth reduction of DS of internal distributions 2,3. This difference can be attributed to the fact that GMM starts the deconvolution process with the fitting of the central distributions to capture the main peak of the total histogram and then proceeds to the external distributions to reproduce the much weaker tails to the histogram. Therefore, it is expected that the innate stochasticity of GMM implementation has a stronger impact on the external distributions with respect to the internal ones.

5.2 | Second case study: one distribution with varying weight

In our second case study, the phase distributions are not all equiprobable since the weight of one external or internal distribution can vary from 0 to 0.25 while the other three distributions have equal probability defined by the condition that the total sum should be fixed to unity. When the varying weight distribution is external, a visualisation of the effects of weight and *Range* for constant variance $Var = 100$ is shown in Figure 7.

The results are in similar lines to those of Figure 6. The stronger deviations of GMM-based predictions from the true ones are observed at small and high ranges due to the overlap and degeneration effects explained in the first case above. Also, the predictions for external distributions are less successful since both deterioration effects (overlapping and degeneration) have a stronger impact. The additional effect revealed by the contour graphs of Figure 7 concerns the distribution with the varied weight (see Figure 7D). Not surprisingly, it is found that at small ranges and weights the DS of this distribution gets much smaller values than the other ones (cf. Figure 7D with Figure 7A–C) since it is more vulnerable to overlapping effects due to its smaller contribution to the total histogram.

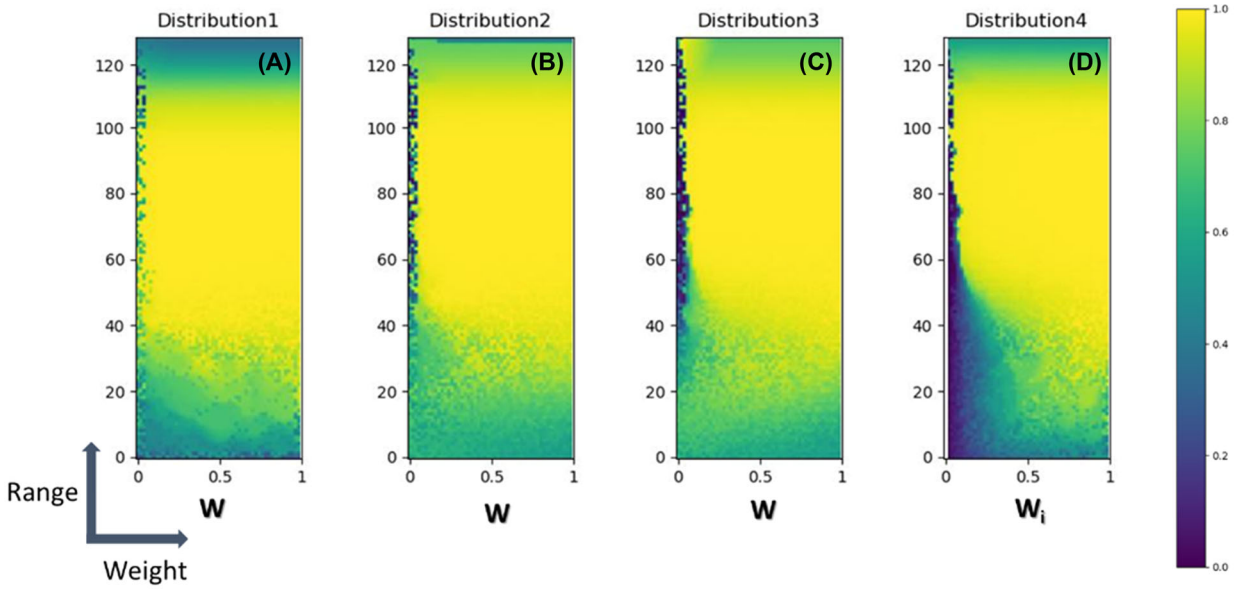

FIGURE 7 *DS* versus the collective parameter *Range* and the normalised *Weight* (w_i) of one external distribution (D). The weight of the latter (D) changes from 0 to 0.25 while w (A, B, C) depends on w_i according to $w = (1 - w_i)/3$.

The results of Figures 6 and 7 indicate that the GMM-based deconvolution of an image histogram does not have the same accuracy for all distributions. When the distribution is external or has a smaller weight than the others then GMM predictions are less accurate and characterised by more narrow windows of success in the parametric space of Range, Variance, and weight. To identify which one of the above-mentioned factors (position or weight) is more critical in GMM accuracy, we repeated the calculations of Figure 7 in the case that the distribution with varied and smaller weights is internal and the results are displayed in Figure 8. The overall pattern of contour graphs are quite similar to those shown in Figure 7. However, now the more vulnerable distribution to the overlapping and degeneration effects is the distribution 3 with the varied and smaller weight though it is internal. This means that the reduced weight of a distribution undermines more GMM accuracy than its external position.

5.3 | Third case study: histograms of real-synthetic BSE images

In hybrid-synthetic BSE histograms, distribution variances are correlated with distribution mean values. For the third case study, the equation that connects variance with mean distribution value emerges from the measurements shown in Section 4. In more detail, the contrast setting chosen for this case is Contrast = 60 where the correlations of mean values and variance are given by the linear equation:

$$Var = 0.06312 \times mu + 6.44227, \tag{11}$$

where *mu* is the mean distribution values. One would expect that distribution with large mean values is more susceptible to lower DS values. However, this is not observed in this case. External distributions are attributed with lower DS values, indicating the effect of different Z-contrast has less impact on the GMM predictability than the relative distribution position (i.e. internal, or external). In other respects, the changes of DS with respect to Range values as presented in Figure 9 exhibit similar overall behaviour to the other test cases presented in this work, where in small Range values overlapping decrease DS values and increase fluctuations of DSs and in large Range values saturation diminish DS values.

5.4 | Dependence of GMM accuracy on tolerance values

In the previous calculations and graphs, the implementation of EM-GMM uses a tolerance 10^{-5} to compromise the accuracy of the GMM algorithm and the time the EM algorithm needs to converge.

One could claim that a lower tolerance in the GMM algorithm could lead to more accurate histogram deconvolution and therefore smaller DS values. This claim is

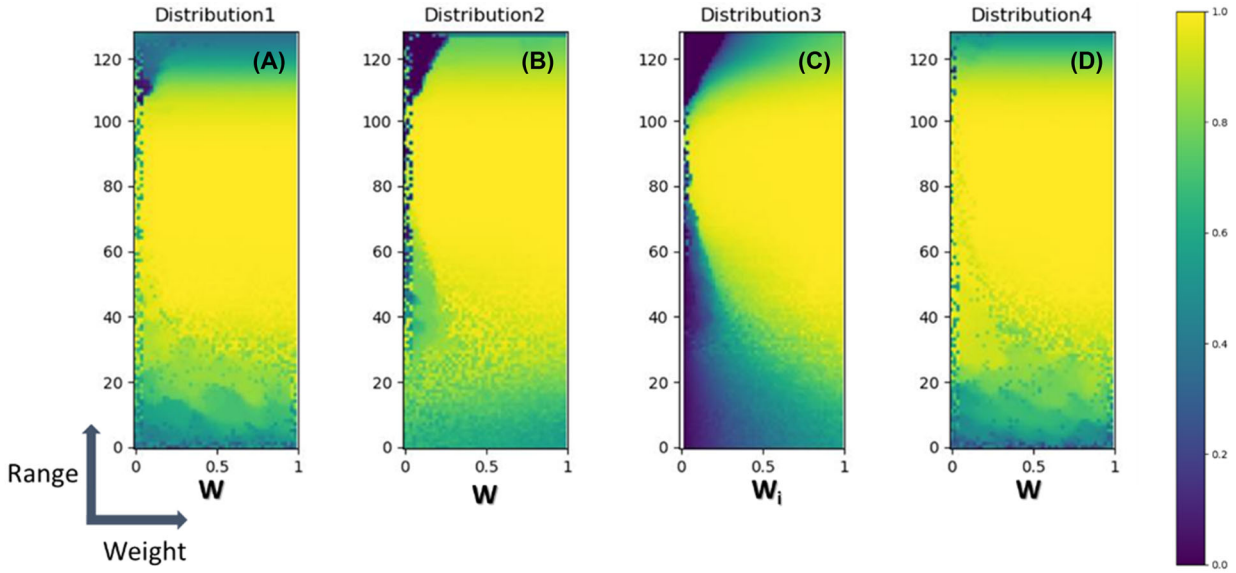

FIGURE 8 *DS* versus *Range* and *Weight* of one internal distribution (C) varying from 0 to 0.25 (equal to the weight of the other distributions). Again, the variable w (A, B, D) is extracted from w_i according to the equation $w = (1 - w_i)/3$.

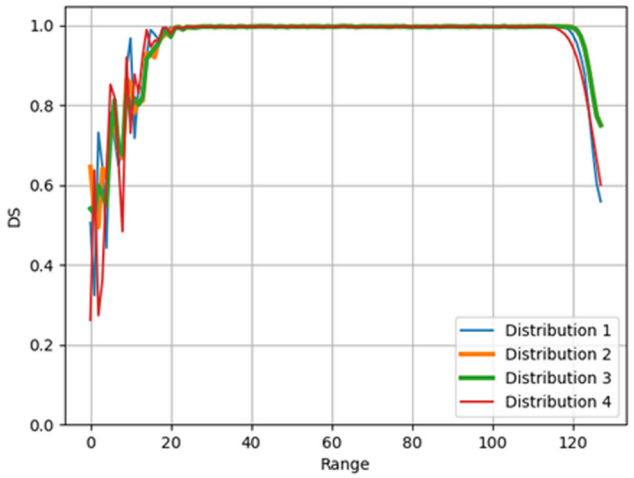

FIGURE 9 *DS* versus *Range* and *Variance* for all four distributions as calculated for constant and equal weights of all distributions where the *Variance* of distributions is correlated with distribution *Mean* values inspired by BSE images.

partially true. As shown in Figure 10, a decrease of tolerance from 10^{-5} to 10^{-8} leads only to a marginal increase in the average DS values at both small and high Ranges where the overlapping (unimodality) and degeneration issues deteriorate GMM predictions respectively. Additionally, the application of the EM-GMM algorithm with tolerance equal to 10^{-8}, despite the significant effect on the computational time required, does not contribute to the reduction of fluctuations in low Range values. An explanation of these variations may be again the strong overlapping of distributions leading to unimodal histograms which GMM fails to deconvolute with accuracy.

5.5 | The role of skewness and kurtosis of the image histogram

The main question that motivated this work has been to identify the critical features of the histogram of an SEM image which can be used to predict the success of the GMM deconvolution process and therefore the quantification of material phase proportions. The results of our study presented in the previous figures and sub-sections revealed that this critical feature may be the presence of peaks or even knees in the total histogram whose number should be equal to the number of phases comprising the material under study. When these peaks disappear because they merge to form a unimodal smooth histogram due to strong overlapping or they exhibit saturation at the extreme pixel intensity values (0 or 255) then GMM predictions deteriorate, and DS gets values less than 0.95.

However, this transition from success to failure is not abrupt. We can exploit the benefits of the quantification of GMM accuracy with DS to follow the effects of peak merging and of the concomitant formation of a unimodal distribution at small Ranges on the DSs of distributions. Since the increase of overlapping and the loss of multiple peaks in the distribution is followed by the appearance of a platykurtic distribution, we try to use the kurtosis to quantify these changes and calculate its relationship with the worst minimum DS. Additionally, changes in distribution weight affect the symmetry of the overall histogram. To capture such changes in the histogram symmetry, the measure of skewness is also used. The result of this calculation for the first and second case studies are illustrated in Figure 11A and B, respectively, where Range

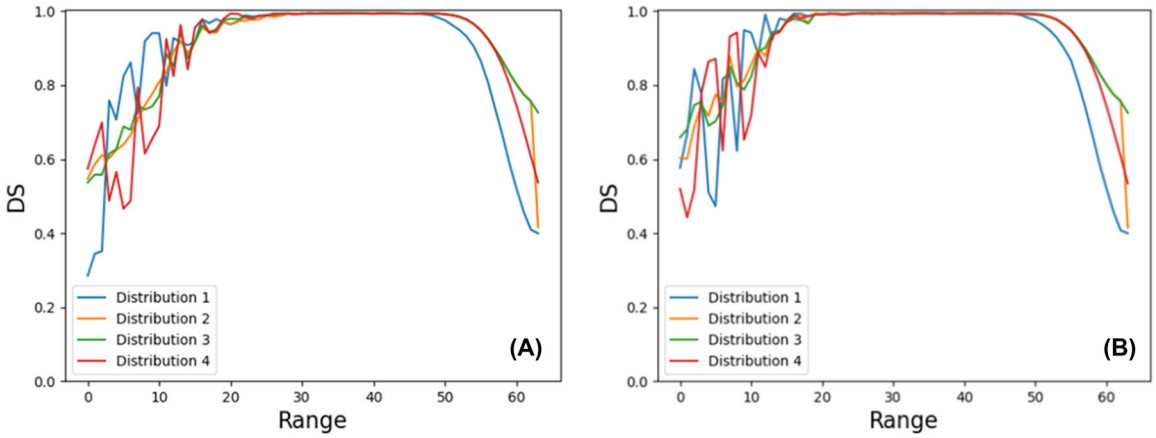

FIGURE 10 *DS* versus *Range* for all four phases for EM-GMM implemented with tolerance 10-5 (A) and 10-8 (B) with *Variance* values constant and equal to 120.

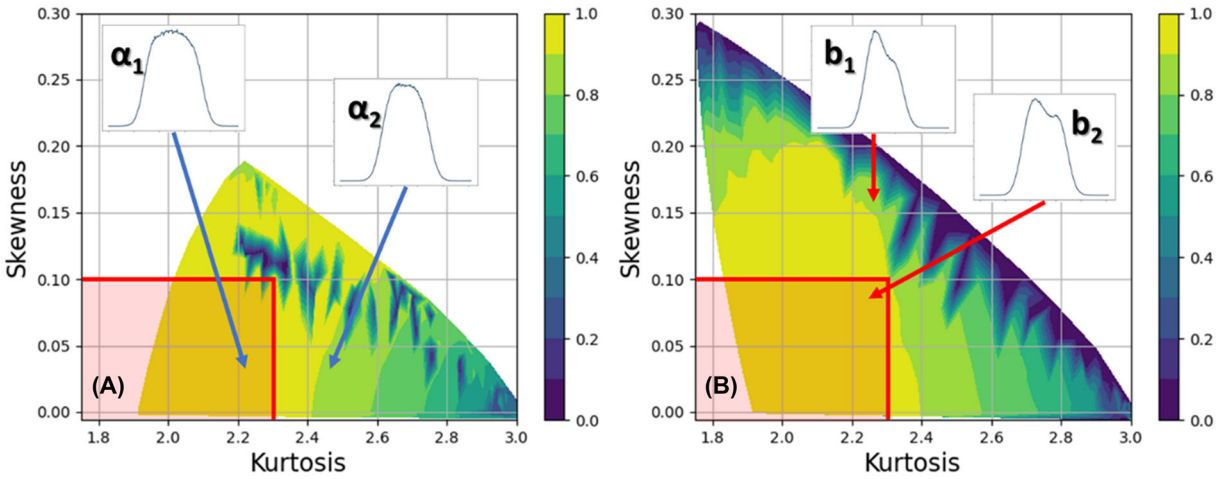

FIGURE 11 Contour graphs of DS minimum values versus Kurtosis and Skewness of the total image histograms when the weight of the internal (A) or external (B) distribution changes. Kurtosis decreases in both cases as the Range value increases from 0 to 30, indicating deviations from the Gaussian shape. The change in skewness comes from the decrease in weight variable in both cases from 1 to 0.1. Variance is set constant and equal to 240 for all distributions. The red rectangles in both diagrams show the areas where the rule of thumb holds and GMM can be applied with accuracy. The inset histograms $\alpha 1$, $\alpha 2$ and b1, b2 reveal the importance of rule of thumb since they show examples of histograms with similar characteristics (unimodal in $\alpha 1$, $\alpha 2$ or bimodal in b1, b2) but with values of kurtosis ($\alpha 1$, $\alpha 2$) or skewness (b1, b2) lying inside ($\alpha 1$, $\alpha 2$) and outside ($\alpha 2$, b1) the red area of rule of thumb dictating the accuracy of GMM results.

values vary from 0 to 30 for fixed Variance equal to 240 and Weight changes from 0 to 0.25 for fixed Range. The effect of parameters on DSs have been shown in Figures 6 and 7. In both cases, one can notice a threshold of Kurtosis equal to 2.35 and skewness 0.1 can safely separate the histograms with minimum DSs larger than 0.9 (considered as GMM accurate) from images with histograms that are not able to be segmented accurately with GMM method:

$$\text{skewness} < 0.1 \text{ and kurtosis} < 2.35$$

One could use this condition as a rule of thumb to predict if an image-histogram, with characteristics close to the test cases presented in this work, can be accurately segmented with GMM.

In Figure 11, the inset histograms $\alpha 1$, $\alpha 2$ and b1, b2 show an example of using this rule of thumb. In Figure 11A, both histograms $\alpha 1$, $\alpha 2$ have a unimodal shape however kurtosis of $\alpha 1$ is smaller than $\alpha 2$ and therefore it belongs to the area (shown with the red rectangle) where the rule of thumb holds and GMM can be used safely. A similar behaviour is shown in histograms b1 and b2 (see Figure 11B) which exhibit multimodal shape but the number of peaks do not comply with the number of distributions in the histogram. In this case, the skewness threshold of the rule of thumb is able to predict the accuracy of GMM algorithm.

6 | CONCLUSIONS

GMM is a versatile tool for data clustering, with applications found in image analysis and in segmentation processes. One of the goals of the GMM segmentation process applied in scanning electron microscopy is the separation of different phases (with differences z-effective) in a multiphase material. In this work, GMM algorithms are implemented for the segmentation of BSE images of multiphase materials. In more detail, factors that reduce the accuracy of the GMM segmentation ability are studied and specified. To this end, a new metric for GMM accuracy is introduced that is called distribution similarity index (DS). In addition, an experimental aware framework for the generation of synthetic BSE image histograms is established.

The parametric space, however, is infinite and hence a compromise on the number of parameters has been made. The subspace studied in this work can be described with three parameters: *Range*, *Variance* and *Weight*, while the number of distributions is kept constant and equal to 4. The parameter *Range* is the interval that is separated equidistantly to set mean distribution values. *Variance* is defined as the variance set equal to all distributions or defined through a linear relation that emerged from experimental data. Finally, the total number of pixels is kept equal in all cases; however, the proportion of all distributions is not equal. The proportion of one distribution (w_i) varies from 0 to 0.25 while other distributions share equally the rest of $\frac{1-w_i}{3}$ fraction of total pixels.

It is shown that one of the major causes of reduction in the GMM accuracy is the overlapping of distributions, where highly overlapped distributions are not able to be deconvoluted correctly with the GMM model. In addition, another reason that GMM fails to deconvolute BSE-image histograms is saturation, where a number of pixels agglomerate at the largest or smallest value of greyscale intensities, changing the Gaussian shape of distributions. Finally, we address the question of the inverse problem of how we can relate the overall histogram shape to GMM accuracy. The outcome of this study indicates that GMM is an accurate method when histograms exhibit a multimodal shape, with the number of peaks equal to the number of phases in our material. In the marginal cases where the histogram has a unimodal shape or less peaks than the number of phases, it has been found that GMM results can be sufficiently accurate if histogram skewness is lower than 0.1 and at the same time kurtosis is smaller than 2.35. This rule of thumb has been tested in histograms similar to the cases studied in this work and complements the main criterion of the number of peaks when we have marginal deviations from it. The study of GMM accuracy in a larger parametric space besides the cases of this work is left as future work. Additionally, a more precise prediction of GMM accuracy based on the total image histogram shape alone could be achieved with neural network approaches trained with modelling data similar to those used in this work. A first attempt towards this direction has been presented in Chatzigeorgiou et al. while a more thorough and in-depth analysis is left as future work.

ACKNOWLEDGEMENTS

This research was supported by NCSR 'Demokritos' Industrial Research Fellowship Program [E.E.-12149], funded by the Stavros Niarchos Foundation and TITAN Cement S.A.

ORCID

Manolis Chatzigeorgiou https://orcid.org/0000-0002-5945-706X
Michalis Vrigkas https://orcid.org/0000-0001-5888-6949
Marios Katsiotis https://orcid.org/0000-0002-9104-584X
Vassilios Constantoudis https://orcid.org/0000-0003-3164-977X

REFERENCES

1. Lea, F. M., & Hewlett, P. C. (2004). *Lea's chemistry of cement and concrete*. New York: Elsevier Butterworth-Heinmann.
2. Edwin, R., Mushthofa, M., Gruyaert, E., & De Belie, N. (2019). Quantitative analysis on porosity of reactive powder concrete based on automated analysis of back-scattered-electron images. *Cement and Concrete Composites*, *96*, 1–10.
3. Goldstein, J. I., Newbury, D. E., Michael, J. R., Ritchie, N. W., Scott, J. H. J., & Joy, D. C. (2017). *Scanning electron microscopy and X-ray microanalysis*, Springer.
4. Qian, H., Li, Y., Yang, J., Xie, L., & Tan, K. H. (2022). Image based microstructure classification of mortar and paste using convolutional neural networks and transfer learning. *Cement and Concrete Composites*, *129*, 104496.
5. Gesho, M., Chaisoontornyotin, W., Elkhatib, O., & Goual, L. (2020). Auto-segmentation technique for SEM images using machine learning: Asphaltene deposition case study. *Ultramicroscopy*, *217*, 113074, 1-15.
6. Teck, P., Snellings, R., & Elsen, J. (2022). Method for quantifying the reaction degree of slag in alkali-activated cements using deep learning-based electron microscopy image analysis. *Journal of Microscopy*, *286*, 174–178.
7. Valmianski, I., Monton, C., & Schuller, I. (2014). Microscopy image segmentation tool: Robust image data analysis. *Review of Scientific Instruments*, *35*, 033701, 1–6.
8. Scrivener, K. L., Patel, H., Pratt, P., & Parrott, L. (1986). *Analysis of phases in cement paste using backscattered electron images, methanol adsorption and thermogravimetric analysis*. MRS Online Proceedings Library (OPL), 85.
9. Liu, W., Shi, H., He, X., Pan, S., Ye, Z., & Wang, Y. (2018). An application of optimized Otsu multi-threshold segmentation based on fireworks algorithm in cement SEM image. *Journal of Algorithms & Computational Technology*, *13*, 1–12.

10. Lee, J. H., & Yoo, S. I. (2008). An effective image segmentation technique for the SEM image. In: *IEEE International Conference on Industrial Technology*.
11. Yousef, A., Duraisamy, P., & Karim, M. (2016). A new clustering algorithm for scanning electron microscope images. SPIE Proceedings Vol. 9845, Optical Pattern Recognition XXVII, 984509.
12. Hu, C., & Ma, H. (2016). Statistical analysis of backscattered electron image of hydrated cement paste. *Advances in Cement Research*, *28*, 469–474.
13. Kodamana, Z. L. H., & Huang, A. A. B. (2019). A GMM-MRF based image segmentation approach for interface level estimation. *Science Direct*, *52*, 28–33.
14. Sun, H., & Wang, S. (2011). Measuring the component overlapping in the Gaussian mixture model. *Data Mining and Knowledge Discovery*, *23*, 479–502.
15. Vuttipittayamongkol, P., Elyan, E., & Petrovski, A. (2021). On the class overlap problem in imbalanced data classification. *Knowledge-Based Systems*, *212*, 106631, 1–17.
16. Vattipittayamongkol, P., & Elyan, E. (2019). Neighbourhoud-based undersampling approach for handling imbalanced and overlapped data. *Information Sciences*, *509*, 47–70.
17. Xu, L., & Jordan, M. I. (1996). On convergence properties of the EM algorithm for Gaussian mixtures, *Neural Computation*, *8*, 129–151.
18. Hu, Z. (2015). *Initalizing the EM Algorithm for data clustering and sub-population detect*. PhD Thesis, Ohio State University.
19. Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
20. Hershey, J. R., & Olsen, P. A. (2007). Approximating the Kullback Leibler divergence between gaussian mixture models. In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP '07*. IEEE.
21. Ekström, J. (2011). Mahalanobis' Distance Beyond Normal Distributions. UCLA: Department of Statistics, UCLA. Retrieved from https://escholarship.org/uc/item/24w7k7m1
22. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*, 357–362.
23. Drouin, D., Couture, A. R., Joly, D., Tastet, X., Aimez, V., & Gauvin, R. (2007). CASINO V2.42—A fast and easy-to-use modeling tool for scanning electron microscopy and microanalysis users. *Scanning*, *29*, 92–101.
24. Reimer, L. (1998). *Scanning electron microscopy*. New York: Springer-Verlag Berlin Heidelberg.
25. Timischil, F., Date, M., & Nemoto, S. (2012). A statistical model of signal-noise in scanning electron microscopy. *Scanning*, *34*, 137–144.
26. Chatzigeorgiou, M., Vrigkas, M., Boukos, N., Beazi-Katsioti, M., Katsiotis, M., & Constantoudis, V. (2022) Machine Learning evaluation of microscopy image segmentation methods: The case of Gaussian Mixture Models. In: *Proceedings of the 12th Hellenic Conference on Artificial Intelligence*.

How to cite this article: Chatzigeorgiou, M., Vrigkas, M., Beazi-Katsioti, M., Katsiotis, M., Boukos, N., & Constantoudis, V. (2023). Segmentation of SEM images of multiphase materials: When Gaussian mixture models are accurate? *Journal of Microscopy*, *289,* 58–70. https://doi.org/10.1111/jmi.13150