



# A data analysis procedure for phase identification in nanoindentation results of cementitious materials

Fabien Bernachy-Barbe 

Received: 15 April 2019 / Accepted: 26 August 2019 / Published online: 30 August 2019  
© The Author(s) 2020, corrected publication 2020

**Abstract** Measuring accurately phase properties is essential for a realistic mesoscale modeling of materials, and nanoindentation is a popular technique regarding mechanical properties. Given the statistical nature of the grid indentation method, where large arrays of indents are performed blindly, the identification of phases from the distributions of measured properties is an essential step. Many biases can be introduced at that stage when the phases do not have very distinct properties as is often the case for cementitious materials, since many indentation tests may also be in effectively heterogeneous areas. It is proposed in the present work to analyze statistical indentation results on cementitious materials with a hierarchical clustering algorithm making use of enriched information, including the spatial coordinates of the indent. It is shown that it allows to reduce potential biases of the method by eliminating tests in potentially heterogeneous areas and performing model independent identification of the different phases.

**Keywords** Nanoindentation · Micromechanics · Cementitious materials · Unsupervised clustering

## 1 Introduction

Nanoindentation provides an efficient method to probe the mechanical properties of materials at microscopic length scales. By applying force using an indenter of known geometry and properties onto the surface of a sample, the obtained force-depth response is analyzed through contact mechanics in order to extract the local elastic modulus and hardness. In the context of cementitious materials, nanoindentation is being widely used, for example, to study Interfacial Transition Zones [1, 2], chemical degradation [3], carbonation [4], or time-dependent (creep) properties [5–8]. By directly measuring the elementary phases properties [9–13] nanoindentation allows for a direct experimental input to upscaling methods [14] that aim to increase their predictive capabilities in deriving effective mechanical properties.

The currently standard phase identification procedure from nanoindentation data relies on the identification of a Gaussian Mixture Model (GMM) on the 2D distribution of indentation mechanical properties [15, 16]. More descriptors such as chemical information can be added in these classification methods to identify phases more reliably, coupling quantitative SEM-EDS and nanoindentation, where one can use in the procedure atomic ratios combined with indentation parameters [17–19]. This method remains however complex in

---

F. Bernachy-Barbe (✉)  
Den-Service d'Etude du Comportement des  
Radionucléides (SECR), CEA, Université Paris-Saclay,  
F-91191 Gif-sur-Yvette, France  
e-mail: fabien.bernachy-barbe@cea.fr



its implementation since quantitative analysis of the elemental composition using EDS has to be performed at the same locations where shallow indentation testing has been performed, with micrometer scale precision. For the standard approach making use only of micromechanical measurements, some controversy has arisen on the capabilities of the statistical indentation method to determine properties of single phases in cement paste [20, 21], since the fitting of GMM presents many local minima [22]. More generally, it has been notably argued that for cementitious materials, probed volumes smaller than the heterogeneity length scale (and possibly in pure phases) are too small relatively to the minimal obtainable roughness given by the porosity. This hard limit is in the order of a few hundreds of nanometers (where adequate contact conditions are fulfilled), as shown in [23] (for a high quality surface preparation using a focused ion beam).

In view of these issues, it is proposed in the present work to take advantage of the spatialized nature of the data provided by nanoindentation maps in order to reduce these biases commonly introduced in the post-processing stage of statistical nanoindentation. Indeed, nanoindentation maps at high resolution (tight indent spacing) allows to resolve details of the microstructure and adjacent indents have a probability to belong to the same phase higher than purely random sampling. It is first argued that the standard analysis of nanoindentation data based on Gaussian mixture models inevitably introduces biases when phases' mechanical properties do partially overlap. A post-processing method using more fully the obtained experimental data is proposed: after the removal of indentation tests where a local homogeneity criterion is violated, a hierarchical clustering algorithm making use of enriched and spatialized data is used. An application to indentation results on a high water-to-cement ratio cement paste is presented.

## 2 Standard indentation methods

### 2.1 Analysis of load–displacement data

An instrumented indentation test consists in applying a force load path to the surface of a sample using an indenter with precisely characterized properties, and recording the corresponding displacement response of

the indenter tip. This force–displacement data (assuming proper preliminary correction of the indenter's compliance and detection of the contact point) is then subject to an inverse analysis using contact mechanics, often with the Oliver and Pharr method [24]. The hardness  $H$  follows its usual definition:

$$H = \frac{F_{\max}}{A_c}$$

where  $F_{\max}$  is the peak load and  $A_c$  the contact area under the indenter, obtained through  $f$ , the indenter area function with  $A_c = f(h_c)$ . The contact depth  $h_c$  is obtained by correcting the measured depth  $h$  for the displacement of the contact perimeter (sink-in), given by the solution of Sneddon for the indenter elastic contact [25]:

$$h_c = h - \varepsilon \frac{F_{\max}}{S}$$

where  $S$  is the elastic unloading stiffness. The  $\varepsilon$  coefficient takes values 0.72 for conical indenters and 0.75 for spheres at small depths. The contact stiffness  $S$  is the measured initial slope of the unloading force–displacement curve; it follows from the elastic modulus through the equation:

$$S = 2\beta \sqrt{\frac{A}{\pi}} E_r$$

with  $\beta$  a correction coefficient equal to 1 for axisymmetric indenters and  $E_r$  the “reduced” modulus obtained from the indented material moduli  $E$  and  $\nu$ , and the indenter tip moduli  $E_i$  and  $\nu_i$  (usually diamond):

$$\frac{1}{E_r} = \frac{1 - \nu^2}{E} + \frac{1 - \nu_i^2}{E_i}$$

Although sharp pyramidal indenters are widely used in this context, spherical indentation allows for primarily elastic contact for shallow indents [26] and is preferred here for to perform property mapping at high spatial resolution and with therefore small probed volumes. The scale dependency of the spherical indentation (due to the non-self-similarity of the geometry, which possesses an intrinsic length scale corresponding to the radius of curvature) also allows studying the amount of irreversible strain relatively to elasticity for a given loading path.



## 2.2 Gaussian mixture property decomposition and issues

Given the random nature of the probed areas under the indenter (heterogeneity, residual roughness) the statistical indentation technique is the usual way to analyze nanoindentation data on cementitious materials [15]. The technique essentially consists in fitting the mechanical properties experimental statistical distribution with a weighted sum of Gaussian functions, known as a Gaussian Mixture Model (GMM). This fit is a highly dimensional optimization problem (a two-component GMM in two-dimensions has 7 parameters) commonly solved using the Expectancy-Maximization (EM) algorithm [27].

One may consider a simple numerical experiment based on a simple statistical model for the indentation modulus on a two-phase material. Let us consider that the effective indentation modulus measured in a biphasic area is given by the Voigt (uniform strain) bound [28], to which is added some error term. For a uniform Poisson's ratio, it yields the simple rule of mixtures:

$$E_{\text{eff}} = fE_1 + (1 - f)E_2 + e$$

where  $E_1$  is the indentation modulus of phase "1",  $f$  its volume fraction in the probed volume and  $e$  a measurement error. One may check that using the Reuss (uniform stress) bound does not yield significantly different conclusions for the range of parameters chosen here. This random variable is obtained from the random variables  $E_1$ ,  $E_2$  and  $e$ , which will assumed to follow a normal distribution and  $f$  which is assumed to follow a beta distribution, with probability density:

$$\beta(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$$

with  $B$  the beta function which is the proper normalization constant. Parameters  $a$  and  $b$  define the shape of this distribution. The beta distribution allows for general modelling of a continuous random variable on  $[0, 1]$  and is therefore well adapted to the statistical behavior of proportions. In our case, we assume that indents in nearly homogenous areas are highly probable ( $f$  very likely to be close to 0 or 1) and therefore  $a \ll 1$  and  $b \ll 1$ . Phase "1" is also assumed to be of

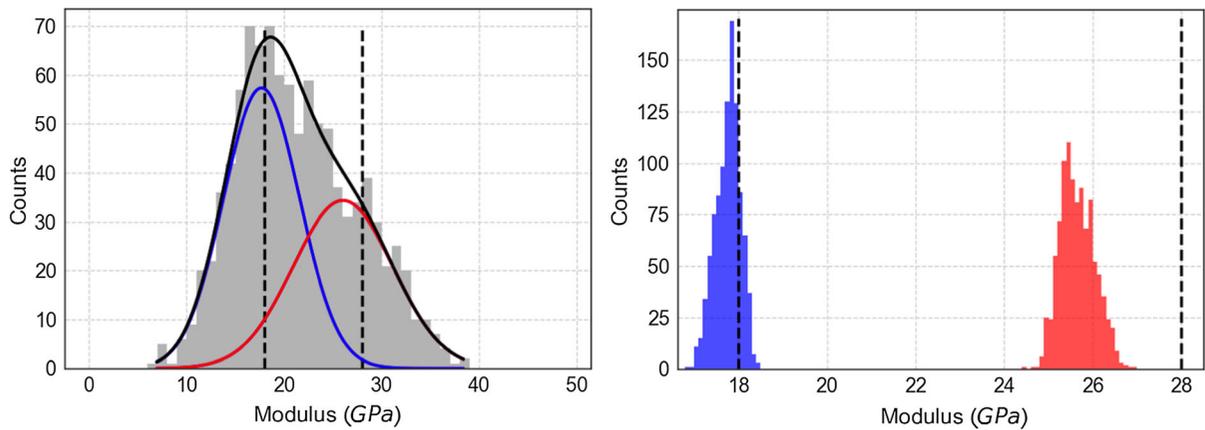
higher volume fraction than phase "2" (hence,  $f$  is more likely to be close to 1 than 0) and therefore  $a < b$ .

As a numerical application, we assume phases with mechanical properties similar to the two usually considered types of C-S-H [29, 30] and chose  $E_1$  and  $E_2$  to follow a Gaussian distribution with means 18 and 28 GPa and standard deviation 4 GPa. This assumption matters since we assume phases have an intrinsic dispersion in mechanical properties and that the final statistical dispersion is only partially due to heterogeneous probed volumes. The error  $e$  is taken from a centered Gaussian distribution with standard deviation 1 GPa. Finally, coefficients for the beta distribution are taken arbitrarily as  $a = 0,017$  and  $b = 0,035$  which results in a probability of  $f$  to be in  $[0, 0.01]$  of approximately 62% and 28% between  $[0.99, 1]$ . Having 10% of indents in somewhat heterogeneous areas is voluntarily an extremely optimistic model that assumes that the experiment is properly designed [22].

Datasets with 1000 modulus measurements are then generated as they appear empirically to be a reasonable amount of data to differentiate phases in cementitious materials. A two-component Gaussian mixture model is then fitted to this data with the usual EM algorithm, as it can be observed for a dataset in Fig. 1. One expects to recover the correct means of the  $E_1$  and  $E_2$  distributions as the two means of the Gaussian mixture components, at least as a limit for a large number of repeated experiments.

For 1000 experiments each containing 1000 indents, the fit of a Gaussian mixture model with two components is performed. The distribution of the means of these Gaussian curves is represented in Fig. 1: it can be observed in particular that these means generally underestimate the means of the original phase distributions; finding the correct average value for the stiff phase modulus has actually almost zero probability. The most likely value is around 25.4 GPa, to be compared to the expected 28 GPa; the estimated weight (volume fraction) of the soft phase is approximately 55% as compared to the expected 67% (as defined as the value of the cumulative distribution function of  $f$  at 50%). However, one may check that constraining either the weights (phase volume fractions) or means of the Gaussian components would yield correct values for the free parameters, which is characteristic of an underdetermined problem. As a conclusion, with very optimistic





**Fig. 1** One realization of the  $E_{\text{eff}}$  distribution for 1000 indents and associated best Gaussian mixture fit (left), distributions of average values of the Gaussian components fitted for 1000 experiments with 1000 indents each (right)

assumptions about the statistics of nanoindentation results (two phases, separated with moderate overlap, moderate heterogeneity, low experimental errors), it can be observed that results yielded by Gaussian mixture phase decomposition do not generally converge to the correct average values for the phases properties (both mechanical properties and phase fractions), even for an unrealistic large number of experiments. This type of method may converge to many local minima as shown by [22], and the model parameters at the global minimum do not generally coincide with the parameters of the underlying distributions. The potential bias of the method in the case where more than two phases partially overlap (for example in cementitious matrices with portlandite CH around 40 GPa and stiffer anhydrous phases), or have non-Gaussian distributions, may be significantly high.

### 3 Alternative method: hierarchical clustering applied to nanoindentation

In view of the previous results, it is useful to attribute phases to each indent based on the properties of the force–displacement curves themselves instead of identifying phase probability distributions on the whole dataset. With no a priori knowledge about the mechanical properties of each phase, one may use unsupervised clustering algorithms, a class of algorithms aiming to regroup automatically data vectors into classes based on some similarity measure. Moreover, the usual decomposition method makes no use of

spatial correlations of the dataset that arises when the experiment is properly designed (indent spacing adequately small—lower than the characteristic heterogeneity size). In particular, indents in heterogeneous areas can be detected as small scale variations in measured mechanical properties, and spatially close areas with similar mechanical properties are likely to be from the same phase. The proposed algorithm attempts to take into account these remarks.

#### 3.1 Data preprocessing

Each indentation force–displacement curve is summarized as a vector that describes as completely as possible its properties with minimal data. To each point (“observation”) from an indent is associated the vector  $X$ :

$$X = [x, y, E, H, h_r, R]$$

where  $(x, y)$  is the spatial position,  $E$  and  $H$  the usual indentation modulus and hardness,  $h_r$  the residual depth and  $R$  the ratio of the elastic strain energy to the total strain energy, that is linked to the compared curvatures of the loading and unloading curves.  $R$  is calculated as the ratio of the integrals of the force–displacement unloading branch and the loading branch. Although large (non-linear) correlations exist between these quantities, they are not entirely redundant. Moreover, using an enriched dataset relatively to the usual  $[E, H]$  is expected to provide additional robustness into the method when dealing with noisy or imperfectly corrected data:  $R$  is notably invariant

through a shift in displacement and therefore insensitive to errors in zero-point correction.

The local relative error of the indentation hardness is computed as the ratio of the standard deviation  $S$  to the mean  $\mu$  of hardness values in a neighborhood (called  $H^x$ ):

$$s(H, x) = \frac{S(H^x)}{\mu(H^x)}$$

where the neighborhood is chosen here as the indent itself and its four nearest neighbors. This parameter characterizes the local variability in mechanical properties. The physical size of the neighborhood should be smaller than the characteristic heterogeneity length scale. In order to eliminate probable heterogeneous regions, indents are eliminated from the analysis if  $s(H)$  is higher than some arbitrarily defined value. Finally, since  $X$  components are of different magnitudes and are to be compared, each measured quantity is normalized in a robust way such that each variable has identical 25% and 75% percentiles.

### 3.2 Data reduction and decomposition

Although data remains of reasonable dimensionality, unlike datasets with many descriptors resulting for example of Acoustic Emission signals [31], large non-linear correlations between the chosen descriptors are observed and might tend to give excessive importance to some underlying features of the dataset. Several methods are classically used to perform dimensionality reduction and build uncorrelated variables such as Principal Component Analysis (PCA), but are based on the assumption that the initial variables are linearly correlated. It is proposed to apply here the ISOMAP algorithm [32], of the family of manifold learning algorithms [33]. The objective of this generic dimensionality reduction method is to discover the underlying non-linear degrees of freedom in the dataset while preserving the intrinsic geometry of data. In practice, it consists of three steps: first, determining the neighborhood of each point by constructing a graph of data points weighted with distances; second, computing the distances between pairs of points—as distances of the shortest path in the graph—and finally building a lower representation of the data from this distance matrix. This method is used here to obtain from the four nanoindentation parameters  $E, H, h_r, R$

two independent descriptors,  $V_1, V_2$ , of the nanoindentation data (of decreasing magnitude), reducing the data vector to:

$$X = [x, y, V_1, V_2]$$

These new variables are left unscaled to preserve the relative ponderation of descriptors.

### 3.3 Clustering algorithm

The selected clustering algorithm is of agglomerative hierarchical type [34]: with initially each observation belonging to its own cluster (“singlet”), clusters are successively merged until the final number of required clusters is reached. At each step, merging is performed according to some criterion which aims at globally achieving maximal similarity of elements inside each cluster and maximal dissimilarity between clusters (or, geometrically, maximal intra-cluster compacity and inter-cluster separation). We use the weighted Euclidean distance as a similarity measure in the now 4D space of data vectors:

$$d(X^a, X^b) = \sqrt{\sum_{i=1}^4 w_i^2 (X_i^a - X_i^b)^2}$$

The weights  $w_i$  are introduced in order to adjust the influence of the different measured quantities on the similarity measure. Spatial coordinates will be chosen to have a reduced influence when selecting the clusters to merge. The selected algorithm is Ward’s method [35]; at each merging step, one selects the couple of clusters to merge such that the increase in intra-cluster “inertia” is minimal, defined as:

$$I_{\text{intra}} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{N_i} d(X^j - \chi^i)^2$$

where we have at the current step  $N$  observations in  $k$  clusters with each  $N_i$  elements, and center of gravity  $\chi^i$  (such as defined with the aforementioned distance). For a given dataset and distance, the final cluster hierarchy is unique: it does not depend on any initialization of the algorithm or on the number of clusters to be found. The final number of clusters  $k$  (where to “stop” the merging process) in the dataset must however be selected or deduced from some clustering quality evaluation parameter (such as [36]). The interest of the method is that one takes into



account the fact that indents in the same phases should be similar in mechanical properties but also spatially close; using different wording, two distant points, to be affected to the same cluster, have to exhibit very similar mechanical properties; the practical implementation of this compromise however requires the addition of the weights as adjustable parameters.

## 4 Application

### 4.1 Materials and methods

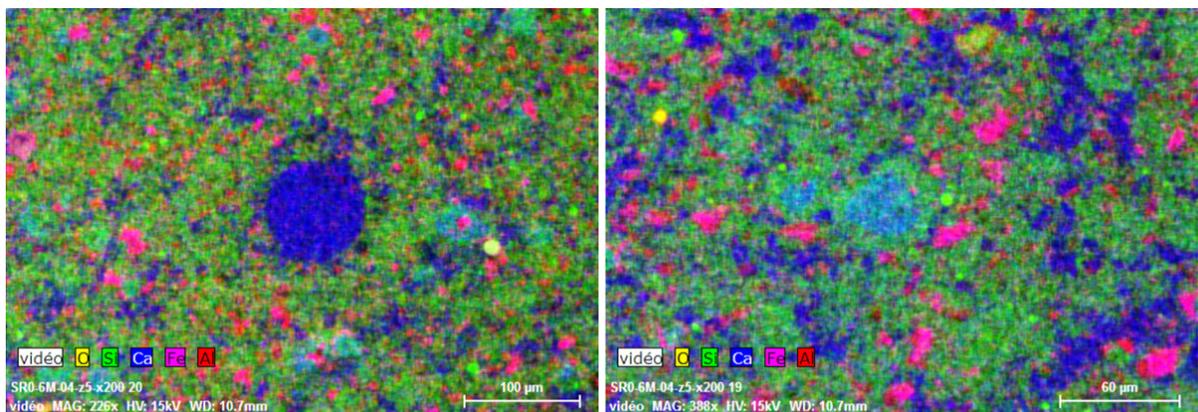
The characterized material was a CEM I cement paste with water-to-cement ratio of 0.52 (see, for example, [37]) (Fig. 2). A small sample was extracted from a large mature sample and was mounted in epoxy resin, grinded and polished with a final step during 1 h in an alcohol based diamond suspension with average grain size 0.25  $\mu\text{m}$ , in order to reach adequate surface quality.

The dataset on cement paste was obtained from a  $32 \times 32$  indentation grid of spacing 2  $\mu\text{m}$  acquired with a 10  $\mu\text{m}$  radius spherical diamond indenter and a Zwick-Roell ZHN/SEM<sup>®</sup> (Ulm, Germany) nanoindenter in the laboratory conditions to avoid vacuum related issues (this SEM compatible device is used here at air in a vibration dampening chamber), with maximum force 2 mN. The quality of these points was assessed by detecting tests with large displacement jumps characteristic of surface damage, or failing the different curve fitting procedures. In the presented

datasets, no such issues were met due to the smooth indenter geometry. Moreover, due to the tight spacing of the indentation grids, interaction between adjacent indents is likely in soft porous regions and should be taken with caution. In the following cases it is estimated that less than 6% of indents are carried out on partially overlapping areas, though it should mainly be of concern for largely porous areas, of limited interest. Finally, the use of spherical indentation is motivated at small indent spacing by the reported lower plastic deformation and damage in brittle materials [38, 39]; however, direct comparisons of hardness results between pyramidal and spherical indenters is not straightforward, even at identical loadings since it depends on the material properties and type of behavior [40]. The 2  $\mu\text{m}$  spacing is assumed to be able to resolve pure phase areas of the material, in particular in the light of latest quantitative 3D observations of cement pastes' phases by ptychographic X-ray tomography [41, 42]. This technique has been used to show the 3D arrangement of single phase agglomerates in cement paste with resolution of order 100  $\mu\text{m}$ .

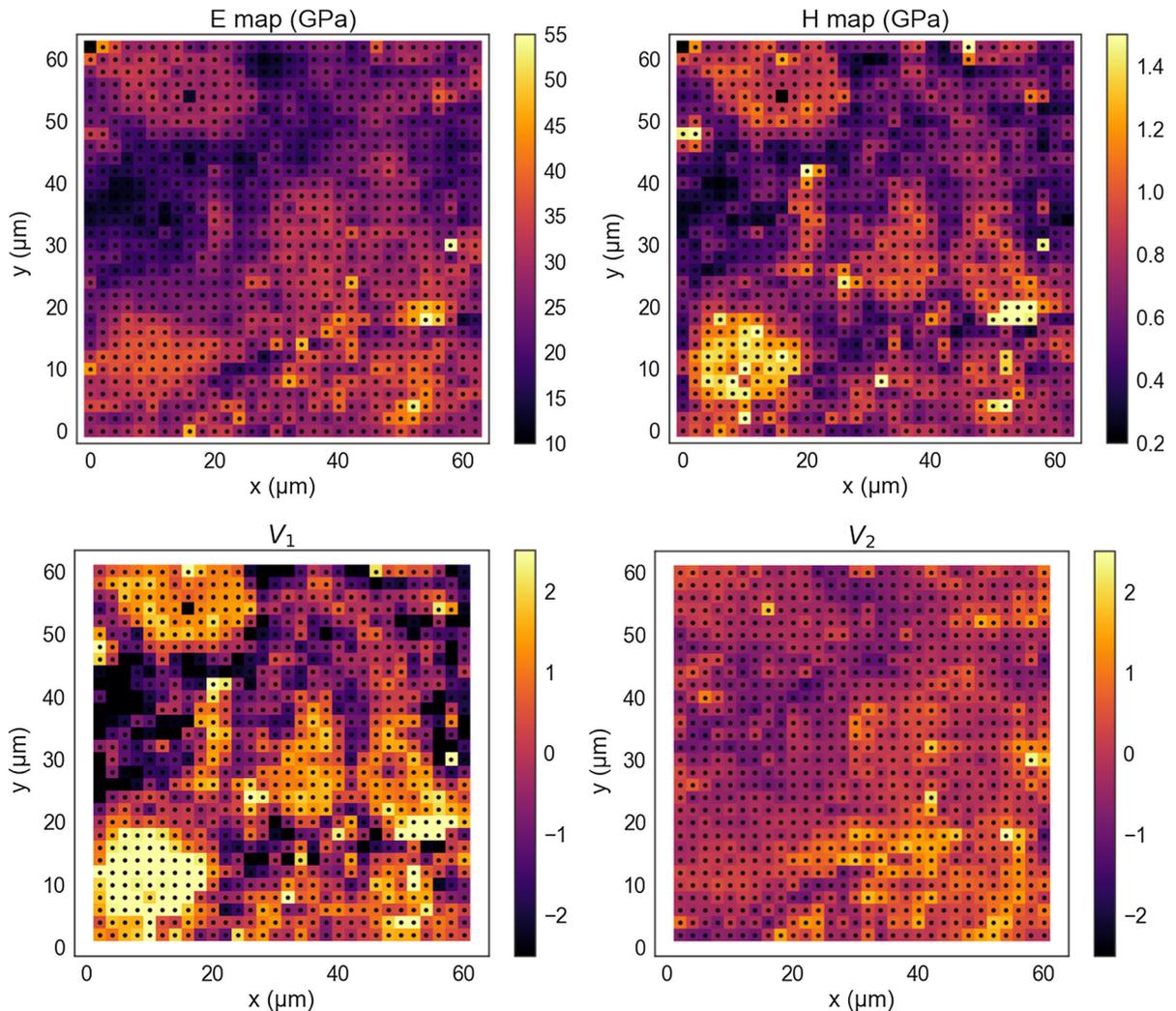
### 4.2 Results on cement paste

The resulting maps of the indentation modulus and hardness are presented in Fig. 3, with residual depth and elastic-to-total strain energy ratio strongly correlated to them. The resulting mapping of a  $64 \mu\text{m} \times 64 \mu\text{m}$  area shows relatively homogeneous subdomains of approximately 10–20  $\mu\text{m}$  that can visually be



**Fig. 2** Composite Si + Ca + Fe + Al (+O) Energy Dispersive X-ray Spectroscopy elemental maps of the studied sample at  $\times 226$  and  $\times 388$ . Usual phases are observed in SEM–EDS

maps: C–S–H, (green), portlandite (CH, blue), anhydrous residual cement grains (cyan), and other minor hydration products (red and pink). (Color figure online)



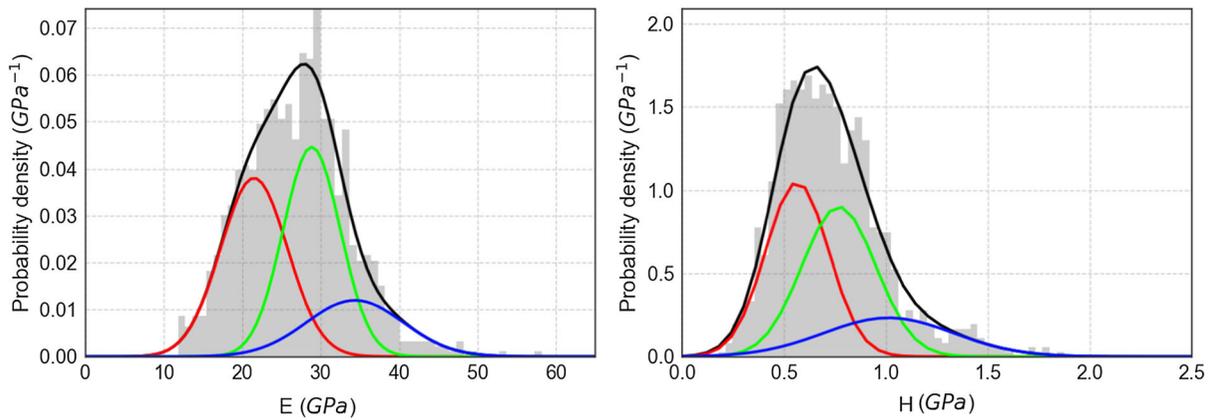
**Fig. 3** Maps of the indentation modulus, indentation hardness, and of the two resulting descriptors provided by the ISOMAP algorithm, on a cement paste sample using spherical nanoindentation

attributed to 4 different phases of distinct mechanical properties. The probed area may be too small to constitute a representative sample of the cement paste mechanical properties, especially in terms of phase volume fractions, although microstructure data yields comparable orders of magnitude [43, 44]. The elastic-to-total strain energy ratio ( $R$ ) map allows to distinguish phases with properties around the values 0.3, 0.5, 0.65 and 0.8. The standard analysis of statistical nanoindentation has been applied to this 1024 point dataset and a Gaussian mixture model with three components fitted (with full covariance matrix) to the  $(E, H)$  results using the implementation of Python module scikit-learn [45]. The distribution components

and  $(E, H)$  data are presented in Fig. 4 and numerical parameters in Table 1.

It can be observed for this sample that phases are not clearly separated in the data histograms and that issues of the type presented in 2.b may be expected, as a large number of models may yield fits as adequate as the global optimum presented here. In particular, one may check that using a four component Gaussian mixture yields optima with components of comparable weights and largely overlapping, which constitutes an even worse decomposition of the material phase's properties.

The maps of the  $V_1$  and  $V_2$  variables are presented in Fig. 3 and can be seen to be much less correlated



**Fig. 4** Obtained 2D Gaussian mixture using the standard method, fit projected onto the  $E$  and  $H$  axes

**Table 1** Properties of the three component optimal Gaussian mixture model fitted to the data, with standard deviations given as errors

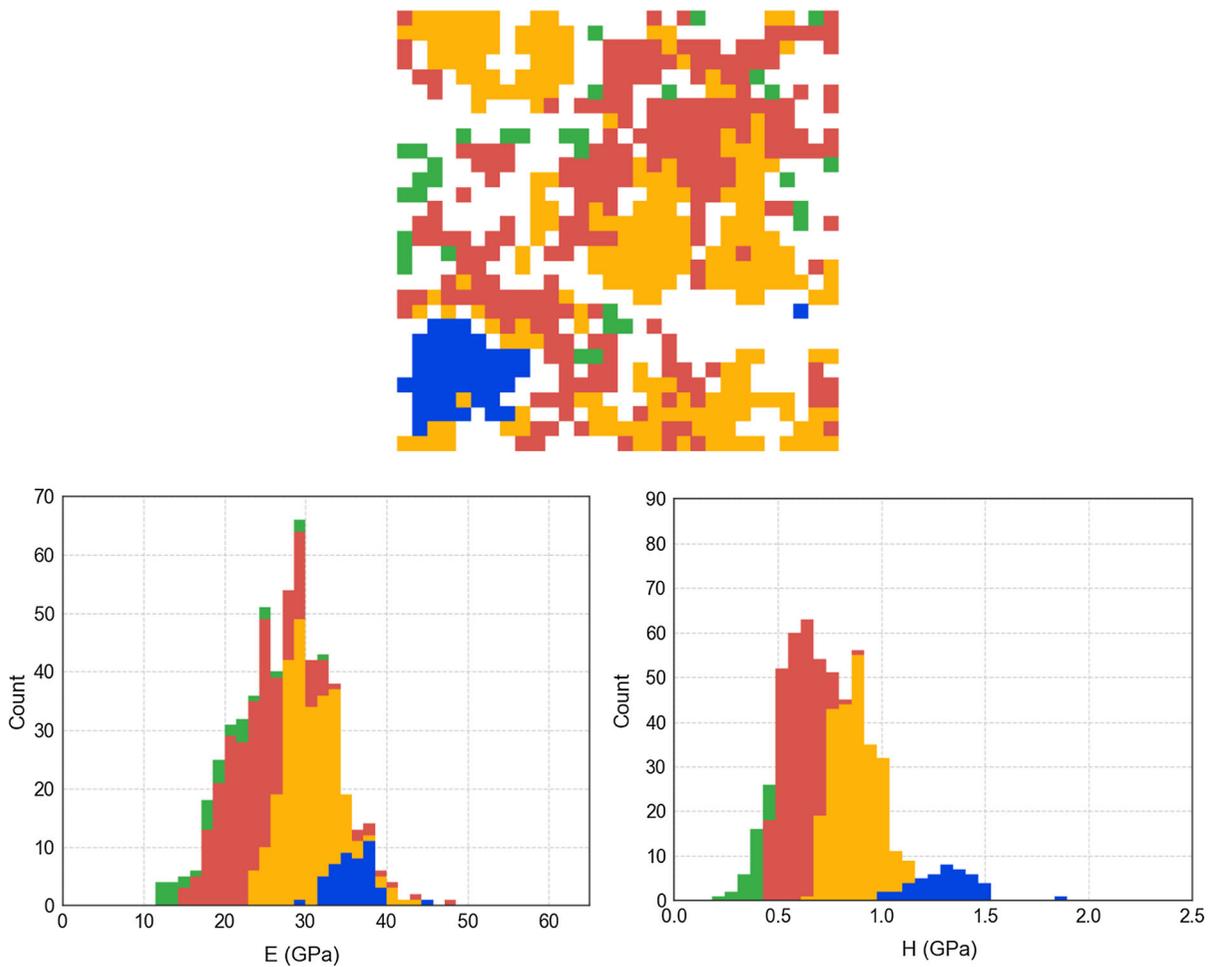
	Weight (%)	$E$ (GPa)	$H$ (GPa)
Component 1	41	$21 \pm 4.3$	$0.56 \pm 0.15$
Component 2	41	$29 \pm 3.6$	$0.77 \pm 0.18$
Component 3	19	$34 \pm 6.2$	$1.01 \pm 0.32$

than the initial descriptors. The proposed method is then applied to the analysis of this dataset using empirically defined weights  $w_1^2 = w_2^2 = 0.3\sigma(V_1)$  where  $\sigma(V_1)$  is the standard deviation of the first nanoindentation descriptor; the other weights are fixed to 1. This algorithm is implemented from the Python module scikit-learn [45]. In order to calculate in each point the local variability criterion, chosen to be  $s(H, x) < 20\%$ , the outer indents are excluded from the analysis and therefore the studied dataset makes use of 900 indents; moreover the criterion excludes 344 points from the subsequent steps (38% of the dataset). The remaining indents are classified using the hierarchical algorithm described above and results for the  $k = 4$  cluster number are represented in Fig. 5.

The obtained per phase property distributions are “irregular” and unsymmetrical and cannot be accurately fitted with Gaussian distributions. The clustering results for  $k = 4$  shows that areas excluded from the analysis are mostly of “weaker” mechanical properties that may correspond to roughness due to local porosity and/or damage, as demonstrated for example for surface defects in [46]. Their sizes are

consistent with the porosity sizes observed in the electron microscopy images (few micrometers, Fig. 2). Also excluded are isolated indents with high mechanical properties, which may be attributed to minor hydration products or anhydrous phases. The procedure yields necessarily well separated clusters that optimize the prescribed criterion of separation in the space of mechanical properties mixed with some preference for local space compacity. The properties of the four phases are given in Table 2.

At the studied high water-to-cement ratio, the need to separate very high porosity areas from the bulk of the cementitious matrix is confirmed [29] although most of it is already eliminated when considering our criterion of local homogeneity. These very high porosity areas are clustered in Cluster 0. Cluster 1, 2 and 3 may therefore be classically attributed to respectively C–S–H (in majority), mixed C–S–H and CH, and mostly homogeneous portlandite (CH), with consistent mechanical properties with those reported in the literature [29, 30]. The large anisotropy in elastic modulus of CH both predicted theoretically [47] and measured experimentally [46] could play a role in the large overlap between these values and those of C–S–H, since one can assume that we sample a random crystal agglomerate [41], with possible defects. Therefore the separation is mostly carried out by the other variables. The size of the studied example area being relatively small, the representativity of the sample may be too small for definitive results. Moreover, the given weights are not supposed to be equal to volume fractions: areas studied here are of arguable representativity and many points with



**Fig. 5** Obtained clustering in physical space (map), and  $E$  and  $H$  distributions per phase as stacked histograms for cement paste

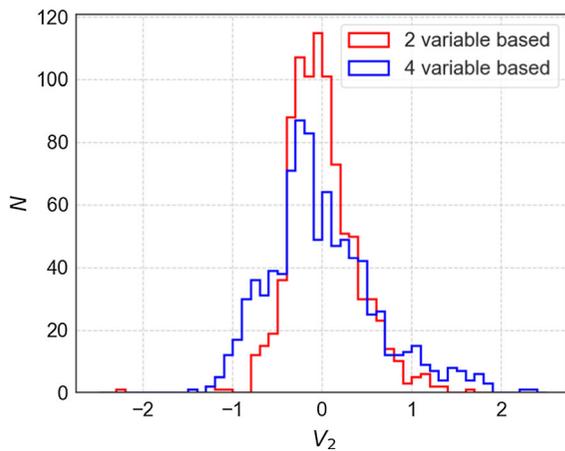
**Table 2** Properties of the four clusters found using the proposed algorithm, with standard deviations given as errors

	Weight (%)	$E$ (GPa)	$H$ (GPa)	Proposed attribution
Cluster 0 (green)	5.9	$19 \pm 5.3$	$0.38 \pm 0.05$	High porosity area
Cluster 1 (red)	42.6	$24 \pm 4.8$	$0.60 \pm 0.08$	Majority C–S–H
Cluster 2 (yellow)	43.3	$30 \pm 3.3$	$0.87 \pm 0.10$	Mixed C–S–H + CH
Cluster 3 (blue)	8.1	$36 \pm 2.5$	$1.31 \pm 0.15$	Majority CH

heterogeneous microvolumes are removed by the procedure, explaining discrepancies with the expected phase assemblage (for example [43]).

Finally, it is interesting to demonstrate that the added variables  $h_r$  and  $R$  add non-redundant information to the independent variables constructed using the ISOMAP algorithm. To this end, the histograms for

the second variable  $V_2$  when using only  $E$  and  $H$  mechanical data and for the full four-variable dataset are shown in Fig. 6. It can be observed that using more variables allows to change the distribution to almost Gaussian-like to a more complex distribution that can be assumed to result from the mixing of differentiated phases. Maps of the variable  $V_2$  can also



**Fig. 6** Distributions of the variables  $V_2$  from the ISOMAP algorithm applied to a two-variable dataset ( $E, H$ ) and to the full four-variable dataset ( $E, H, h_r, R$ )

be shown to have a greater information content and demonstrate the interest of these additional variables.

## 5 Conclusions

A classification method based on a classical hierarchical clustering algorithm has been applied to nanoindentation results, using enriched information: to the usual indentation hardness and modulus is added additional mechanical parameters—that are synthesized using a data reduction algorithm—and spatial position of the indent. This method has been shown to:

1. Allow an identification of phases on nanoindentation results that is independent of a selected statistical distribution (no need for Gaussian properties distribution or widely separated phases), identify each point unambiguously in the case of micrometer scale heterogeneity, or at least providing an efficient classification (instead of deriving a probability distribution for each phase),
2. Allow eliminating points in areas that may be largely porous or heterogeneous.
3. As a perspective, this type of method can be extended in a straightforward manner to “multi-channel” or “hyperspectral” maps obtained using other experimental techniques. It will require the definition of suitable components to the data vector and an updated distance function. It may

include qualitative or quantitative compositional information such as elemental ratios (SEM–EDS), density/average atomic number (SEM-backscattered imaging gray level) or structure (Raman microspectroscopy) that would allow an increased reliability of the phase separation as well as establishing correlations between the local measured quantities. If the properties of Gaussian Mixture Models are desirable (smooth distributions of properties) the hierarchical clustering method presented here may also be used to initialize or constrain fit parameters (such as the means) when many local minima are expected.

**Acknowledgements** This work has been carried out in the framework of the CEA-EDF-Framatome agreement. The author thanks S. Poyet (CEA) for discussions regarding the manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Mondal P, Shah SP, Marks LD (2009) Nanomechanical properties of interfacial transition zone in concrete. In: Nanotechnology in construction, vol 3, Springer, Berlin, pp 315–320. [https://doi.org/10.1007/978-3-642-00980-8\\_42](https://doi.org/10.1007/978-3-642-00980-8_42)
2. Zhu W, Bartos PJM (2000) Application of depth-sensing microindentation testing to study of interfacial transition zone in reinforced concrete. *Cem Concr Res* 30:1299–1304. [https://doi.org/10.1016/S0008-8846\(00\)00322-7](https://doi.org/10.1016/S0008-8846(00)00322-7)
3. Constantinides G, Ulm F-J (2004) The effect of two types of C–S–H on the elasticity of cement-based materials: results from nanoindentation and micromechanical modeling. *Cem Concr Res* 34:67–80. [https://doi.org/10.1016/S0008-8846\(03\)00230-8](https://doi.org/10.1016/S0008-8846(03)00230-8)
4. Han J, Pan G, Sun W, Wang C, Cui D (2012) Application of nanoindentation to investigate chemomechanical properties change of cement paste in the carbonation reaction. *Sci China Technol Sci* 55:616–622. <https://doi.org/10.1007/s11431-011-4571-1>

5. Frech-Baronet J, Sorelli L, Charron J-P (2017) New evidences on the effect of the internal relative humidity on the creep and relaxation behaviour of a cement paste by micro-indentation techniques. *Cem Concr Res* 91:39–51. <https://doi.org/10.1016/j.cemconres.2016.10.005>
6. Pichler Ch, Lackner R (2009) Identification of logarithmic-type creep of calcium-silicate-hydrates by means of nanoindentation. *Strain* 45:17–25. <https://doi.org/10.1111/j.1475-1305.2008.00429.x>
7. Vandamme M, Ulm F-J (2013) Nanoindentation investigation of creep properties of calcium silicate hydrates. *Cem Concr Res* 52:38–52. <https://doi.org/10.1016/j.cemconres.2013.05.006>
8. Zhang Q, Le Roy R, Vandamme M, Zuber B (2014) Long-term creep properties of cementitious materials: comparing microindentation testing with macroscopic uniaxial compressive testing. *Cem Concr Res* 58:89–98. <https://doi.org/10.1016/j.cemconres.2014.01.004>
9. Constantinides G, Ulm F-J, Vliet KV (2003) On the use of nanoindentation for cementitious materials. *Mater Struct* 36:191–196. <https://doi.org/10.1007/BF02479557>
10. Hughes JJ, Trtik P (2004) Micro-mechanical properties of cement paste measured by depth-sensing nanoindentation: a preliminary correlation of physical properties with phase type. *Mater Charact* 53:223–231. <https://doi.org/10.1016/j.matchar.2004.08.014>
11. Mondal P, Shah SP, Marks L (2007) A reliable technique to determine the local mechanical properties at the nanoscale for cementitious materials. *Cem Concr Res* 37:1440–1444. <https://doi.org/10.1016/j.cemconres.2007.07.001>
12. Velez K, Maximilien S, Damidot D, Fantozzi G, Sorrentino F (2001) Determination by nanoindentation of elastic modulus and hardness of pure constituents of Portland cement clinker. *Cem Concr Res* 31:555–561. [https://doi.org/10.1016/S0008-8846\(00\)00505-6](https://doi.org/10.1016/S0008-8846(00)00505-6)
13. Wei Y, Liang S, Gao X (2017) Phase quantification in cementitious materials by dynamic modulus mapping. *Mater Charact* 127:348–356. <https://doi.org/10.1016/j.matchar.2017.02.029>
14. Bernard O, Ulm F-J, Lemarchand E (2003) A multiscale micromechanics-hydration model for the early-age elastic properties of cement-based materials. *Cem Concr Res* 33:1293–1309. [https://doi.org/10.1016/S0008-8846\(03\)00039-5](https://doi.org/10.1016/S0008-8846(03)00039-5)
15. Constantinides G, Ravichandran KS, Ulm FJ, Vanvliet KJ (2006) Grid indentation analysis of composite microstructure and mechanics: principles and validation. *Mater Sci Eng, A* 430:189–202. <https://doi.org/10.1016/j.msea.2006.05.125>
16. Ulm F-J, Vandamme M, Bobko C, Alberto Ortega J, Tai K, Ortiz C (2007) Statistical indentation techniques for hydrated nanocomposites: concrete, bone, and shale. *J Am Ceram Soc* 90:2677–2692. <https://doi.org/10.1111/j.1551-2916.2007.02012.x>
17. Chen JJ, Sorelli L, Vandamme M, Ulm F-J, Chanvillard G (2010) A coupled nanoindentation/SEM-EDS study on low water/cement ratio portland cement paste: evidence for C–S–H/Ca(OH)<sub>2</sub> nanocomposites. *J Am Ceram Soc* 93:1484–1493. <https://doi.org/10.1111/j.1551-2916.2009.03599.x>
18. Krakowiak KJ, Wilson W, James S, Musso S, Ulm F-J (2015) Inference of the phase-to-mechanical property link via coupled X-ray spectrometry and indentation analysis: application to cement-based materials. *Cem Concr Res* 67:271–285. <https://doi.org/10.1016/j.cemconres.2014.09.001>
19. Wilson W, Sorelli L, Tagnit-Hamou A (2018) Automated coupling of nanoindentation and quantitative energy-dispersive spectroscopy (NI-QEDS): a comprehensive method to disclose the micro-chemo-mechanical properties of cement pastes. *Cem Concr Res* 103:49–65. <https://doi.org/10.1016/j.cemconres.2017.08.016>
20. Trtik P, Münch B, Lura P (2009) A critical examination of statistical nanoindentation on model materials and hardened cement pastes based on virtual experiments. *Cem Concr Compos* 31:705–714. <https://doi.org/10.1016/j.cemconcomp.2009.07.001>
21. Ulm F-J, Vandamme M, Jennings HM, Vanzo J, Bentivegna M, Krakowiak KJ, Constantinides G, Bobko CP, Van Vliet KJ (2010) Does microstructure matter for statistical nanoindentation techniques? *Cem Concr Compos* 32:92–99. <https://doi.org/10.1016/j.cemconcomp.2009.08.007>
22. Lura P, Trtik P, Münch B (2011) Validity of recent approaches for statistical nanoindentation of cement pastes. *Cem Concr Compos* 33:457–465. <https://doi.org/10.1016/j.cemconcomp.2011.01.006>
23. Trtik P, Dual J, Muench B, Holzer L (2008) Limitation in obtainable surface roughness of hardened cement paste: ‘virtual’ topographic experiment based on focussed ion beam nanotomography datasets. *J Microsc* 232:200–206. <https://doi.org/10.1111/j.1365-2818.2008.02090.x>
24. Oliver WC, Pharr GM (1992) An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J Mater Res* 7:1564–1583. <https://doi.org/10.1557/JMR.1992.1564>
25. Sneddon IN (1965) The relation between load and penetration in the axisymmetric Boussinesq problem for a punch of arbitrary profile. *Int J Eng Sci* 3:47–57. [https://doi.org/10.1016/0020-7225\(65\)90019-4](https://doi.org/10.1016/0020-7225(65)90019-4)
26. Bushby AJ (2001) Nano-indentation using spherical indenters. *Nondestruct Test Eval* 17:213–234. <https://doi.org/10.1080/10589750108953112>
27. Dempster AP, Laird NM, Rubin DB (1976) Maximum likelihood from incomplete data via the EM algorithm. <https://dash.harvard.edu/handle/1/3426318>. Accessed 1 Feb 2018
28. Randall NX, Vandamme M, Ulm F-J (2009) Nanoindentation analysis as a two-dimensional tool for mapping the mechanical properties of complex surfaces. *J Mater Res* 24:679–690. <https://doi.org/10.1557/jmr.2009.0149>
29. Hu C, Han Y, Gao Y, Zhang Y, Li Z (2014) Property investigation of calcium–silicate–hydrate (C–S–H) gel in cementitious composites. *Mater Charact* 95:129–139. <https://doi.org/10.1016/j.matchar.2014.06.012>
30. Hu C, Li Z (2015) A review on the mechanical properties of cement-based materials measured by nanoindentation. *Constr Build Mater* 90:80–90. <https://doi.org/10.1016/j.conbuildmat.2015.05.008>
31. Moeuv M, Godin N, R’Mili M, Rouby D, Reynaud P, Fantozzi G, Farizy G (2008) Analysis of damage



- mechanisms and associated acoustic emission in two SiC<sub>f</sub>/Si-B-C composites exhibiting different tensile behaviours. Part II: unsupervised acoustic emission data clustering. *Compos Sci Technol* 68:1258–1265
32. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323. <https://doi.org/10.1126/science.290.5500.2319>
33. Lee JA, Verleysen M (2007) *Nonlinear dimensionality reduction*. Springer, New York. <http://www.springer.com/us/book/9780387393506>. Accessed 11 Oct 2018
34. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. *ACM Comput Surv* 31:264–323. <https://doi.org/10.1145/331499.331504>
35. Ward J Jr (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244
36. Davies DL, Bouldin DW (1979) A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* 2:224–227. <https://doi.org/10.1109/tpami.1979.4766909>
37. Aili A (2017) Shrinkage and creep of cement-based materials under multiaxial load: poromechanical modeling for application in nuclear industry. PhD Thesis, Université Paris-Est. <https://pastel.archives-ouvertes.fr/tel-01682129/document>. Accessed 13 Feb 2018
38. Marshall DB (1984) Geometrical effects in elastic/plastic indentation. *J Am Ceram Soc* 67:57–60. <https://doi.org/10.1111/j.1151-2916.1984.tb19148.x>
39. Sukanuma M (1999) Spherical and Vickers indentation damage in Ytria-stabilized tetragonal zirconia polycrystals. *J Am Ceram Soc* 82:3113–3120. <https://doi.org/10.1111/j.1151-2916.1999.tb02210.x>
40. Durst K, Göken M, Pharr GM (2008) Indentation size effect in spherical and pyramidal indentations. *J Phys Appl Phys* 41:074005. <https://doi.org/10.1088/0022-3727/41/7/074005>
41. Trtik P, Diaz A, Guizar-Sicairos M, Menzel A, Bunk O (2013) Density mapping of hardened cement paste using ptychographic X-ray computed tomography. *Cem Concr Compos* 36:71–77. <https://doi.org/10.1016/j.cemconcomp.2012.06.001>
42. Cuesta A, De la Torre ÁG, Santacruz I, Diaz A, Trtik P, Holler M, Lothenbach B, Aranda MAG (2019) Quantitative disentanglement of nanocrystalline phases in cement pastes by synchrotron ptychographic X-ray tomography. *IUCrJ* 6:473–491. <https://doi.org/10.1107/s2052252519003774>
43. Ukrainczyk N, Koenders EAB, van Breugel K (2013) Representative volumes for numerical modeling of mass transport in hydrating cement paste. In: *Multi-scale modeling and characterization of infrastructure mater*, Springer, Dordrecht, pp 173–184. [https://doi.org/10.1007/978-94-007-6878-9\\_13](https://doi.org/10.1007/978-94-007-6878-9_13)
44. Yio MHN, Wong HS, Buenfeld NR (2017) Representative elementary volume (REV) of cementitious materials from three-dimensional pore structure analysis. *Cem Concr Res* 102:187–202. <https://doi.org/10.1016/j.cemconres.2017.09.012>
45. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
46. Trtik P, Kaufmann J, Volz U (2012) On the use of peak-force tapping atomic force microscopy for quantification of the local elastic modulus in hardened cement paste. *Cem Concr Res* 42:215–221. <https://doi.org/10.1016/j.cemconres.2011.08.009>
47. Laugesen JL (2005) Density functional calculations of elastic properties of portlandite, Ca(OH)<sub>2</sub>. *Cem Concr Res* 35:199–202. <https://doi.org/10.1016/j.cemconres.2004.07.036>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

