



Applying grid nanoindentation and maximum likelihood estimation for N-A-S-H gel in geopolymer paste: Investigation and discussion

Zhiyu Lu^a, Wengui Li^{a,*}, Yixiang Gan^b, Kavya Mendu^c, Surendra P. Shah^c

^a School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia

^b School of Civil Engineering, The University of Sydney, NSW 2006, Australia

^c Department of Civil & Environmental Engineering, Northwestern University, Evanston, IL 60208, USA



ARTICLE INFO

Keywords:

Statistical nanoindentation
Maximum likelihood estimation (MLE)
Sodium aluminosilicate hydrate (N-A-S-H)
Nano/micromechanical properties
Alkali activated fly ash geopolymers

ABSTRACT

Static nanoindentation and Maximum Likelihood Estimation (MLE) were applied for the nano/micromechanical properties investigation of alkali-activated fly ash (AAFA) in this study. Some critical issues of statistical nanoindentation were fully discussed, including properties of pure gel phase, influence of bin size when using least-square estimation (LSE), and suitable number of components for deconvolution. Results indicate that the model estimated by MLE method can effectively reflect the micromechanical distribution of AAFA. The number of components needed to separate sodium aluminosilicate hydrate (N-A-S-H) gels is sometimes more than the normally used 3 or 4, depending on the sample and testing factors. The gel phase does not always display as a prominent peak in the histogram and is easy to be mixed with other adjacent peaks even if the bin size is small, indicating the challenges of employing the LSE method to investigate the gel phase in highly heterogeneous materials, such as geopolymers.

1. Introduction

Nano/micromechanical properties are one of the essential aspects for the understanding of materials. The advances in micro and nanoscale mechanical characterization techniques and theory can enable various measurement and open new doors for material analysis [1–4]. In the field of cement-based materials, they have been typically used to test micro- and nanomechanical properties of reacted phases [5–7], measure the width and properties of interface transition zone (ITZ) between aggregate and cement matrix [8–11], confirm the existence of phases [12], etc. Among these techniques, nanoindentation has been most widely used and usually combined with deconvolution techniques, which is known as statistical nanoindentation technique (SNT). In the process of applying statistical nanoindentation in cement-based materials (mainly for Portland cement-based materials), the validity of this kind of method was sometimes questioned, and one of the key issues was the robustness of the normally used LSE method for deconvolution [13]: the global optimum is hard to be found and significantly different results would be generated from different local optimum; the fewer number of phases of 2 and 3 are even found to better fit the frequency distribution histogram than 4 in Portland cement paste. There is another kind of deconvolution technique, the MLE method (Expectation-Maximization algorithm), to the best of knowledge, applied

in cement paste for the first time by Ulm et al. [14]. The MLE method gains much less attention than the LSE method. Only limited studies used this method which can be typically found in reference [14–17]. Besides, based on the MLE method, one study found that the results obtained were not a pure phase but a mixture of phases, particularly calcium silicate hydrate (C-S-H) and (calcium hydroxide) CH [15]. Currently, the LSE method is still the mainstream approach for the deconvolution of nanoindentation data.

Alkali-activated materials are well known for their benefits of reusing industrial by-products and reduced carbon dioxide emission, which are promising supplement or even substitute of Portland cement paste and have attracted extensive relevant studies [18–25]. Although there is great progress in the understanding of alkali-activated materials, resorting micro and nanoscale new model as well as advanced characterization techniques such as Fourier transform infrared (FTIR), nuclear magnetic resonance (NMR) spectroscopy and X-ray microtomography [26,27], there is still no sufficient research to investigate the nano/micromechanical properties of them. Fly ash based geopolymers is one of the most representative alkali-activated materials. The micromechanical properties of alkali-activated fly ash (AAFA) were investigated by previous studies [28–36]. However, inconsistent results were presented by different studies. Typically, the modulus of N-A-S-H gels vary from 4.44 to 16.78 GPa (28 days of curing) in [30], 11 GPa to

* Corresponding author: School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia

E-mail address: wengui.li@uts.edu.au (W. Li).

25 GPa (28 days of curing) in [31], and achieve the similar value of about 17–18 GPa (6 months of curing) in [28], presenting significantly different results. It should be noticed that different testing and analysis methods and parameters were adopted in the above studies. For the statistical analysis of AAFA, the LSE method is still normally adopted with different bin sizes such as 1.0 GPa [28,29] and 2.0 GPa [31] and different phase numbers of 3 [31] or 4 [28,30] were used. This again raised questions regarding the influence of the different choice of parameters in the deconvolution process on the final deconvolution results and also the accuracy and reliability of the LSE method itself.

Based on the AAFA sample, via investigating the effect of bin size, number of phases and grid number on the results, the validity of the LSE method and the MLE method are examined in this study. The MLE method is found to have great advantages over the LSE method. Besides, the gel phase with significant inclusions removed is obtained by MLE in this study by increasing the number of phases for deconvolution. The advantages of the MLE method can be displayed based on a premise that the “compromise approach” is adopted.

2. Experimental and analysis methods

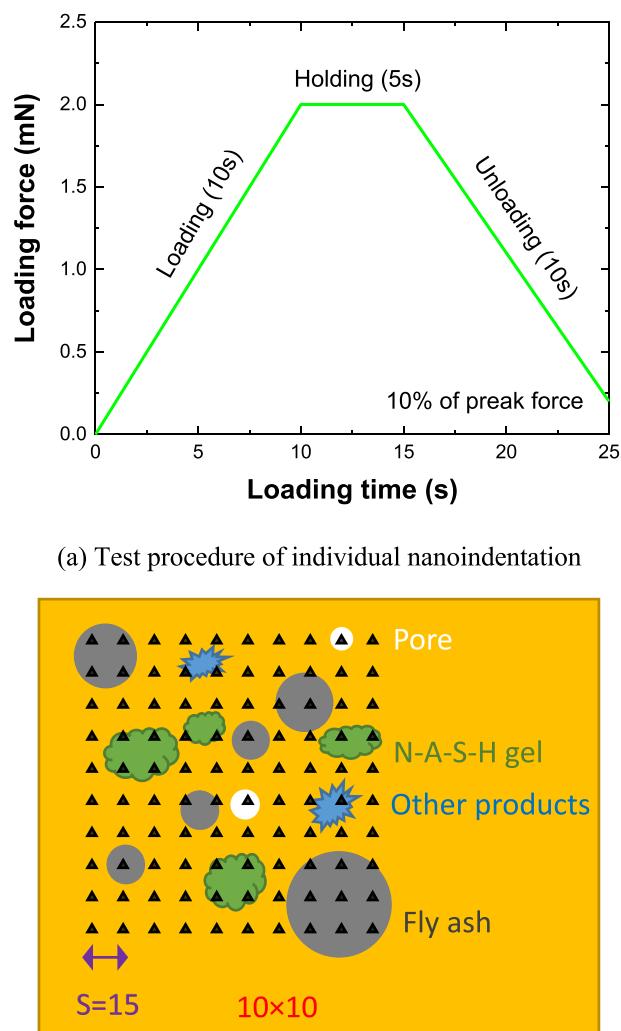
2.1. Materials and sample preparation

Fly ash used was low calcium fly ash with the chemical composition shown in Table 1. Sodium hydroxide pellets were dissolved in water and cooled down. Then, it was mixed with sodium silicate to adjust silica modulus. The original modulus of the commercial sodium silicate is 2.07 with 14.7% of Na₂O and 29.4% of SiO₂. Different samples with Na₂O to fly ash weight ratio of 6%, 8%, and 10% were prepared. For these samples, constant silica modulus of 1.0 in alkali solution, as well as constant water to solid ratio of 0.338 were adopted to keep other factors similar for each kind of sample. Fly ash was mixed with corresponding alkali solution for 5 min and cast into cubic plastic moulds with the size of 50 mm × 50 mm × 50 mm. The samples were vibrated for 3 min and sealed by plastic film. Then, they were put into oven with heat treatment of 70 °C for 24 h. After that, samples were taken out of the oven and put into a standard curing cabinet for further curing (temperature of 20 °C and RH of 95%).

2.2. Nanoindentation

The centre part of cubic specimens was taken out and cut into slices with the size of around 10 mm × 10 mm × 5 mm. Then, these samples were embedded in cold-mounting epoxy resin with one surface exposed to air. After the epoxy solidified, these samples were successively ground on 320, 600 and 1200 grits abrasive papers with each grade lasted for 10 min and then polished with a small force for 40 min by 0.3 µm alumina and 40 min by 0.05 µm cerium oxide slurry to obtain a smooth surface. After each grade of polishing, samples were cleaned in an ultrasonic bath with isopropanol for 3 min to remove particles on their surface. Samples were put in a vacuum oven with the temperature of 50 °C for 12 h to dry and then stored in a vacuum desiccator until testing at 28 days.

After 28 days, Agilent G200 Nano Indenter with a Berkovich tip in The Sydney Centre in Géomechanics and Mining Materials (SciGEM) within The University of Sydney was used to conduct nanoindentation to investigate the micromechanical properties of AAFA. Nine 10 × 10 grid nanoindentation were performed with a grid spacing of 15 µm. The nanoindentation depth can be determined by satisfying both the scale



(b) Schematic of grid nanoindentation

Fig. 1. Nanoindentation testing on geopolymers.

separability condition and roughly one-tenth rule of thumb as the below Eq. (1) [37], where d is the largest heterogeneity of geopolymers phases and D is the characteristic size of microstructure. Besides, enough depth is also necessary to avoid interference from surface roughness [38,39]. For the investigation of N-A-S-H gel, it was reported that the N-A-S-H gel consists of particles with the diameter of about 5 nm [40]. As for characteristic microstructure size, more than about 4 µm can be found even just 7% Na₂O (Ms = 1) is used [41]. After several attempts, trapezoid loading with the peak force of 2 mN was adopted as shown in Fig. 1, with an average penetration depth of 228.21 nm. The loading procedure adopted in this test would be suitable for achieving balance to satisfy the scale separability condition and avoid the effect of surface roughness and multiple phases' responses. Similar loading procedures have also been adopted by other studies for Portland cement paste and AAFA [16,33,34,38,42], making it easy for comparison.

In this study, the constant Poisson's ratio of 0.2 was set for testing.

Table 1

Chemical composition of fly ash for geopolymers.

Oxide	Al ₂ O ₃	CaO	SiO ₂	Fe ₂ O ₃	K ₂ O	MgO	Na ₂ O	MnO	P ₂ O ₅	TiO ₂	LOI
Weight (%)	25.21	1.73	64.55	2.85	1.47	0.41	0.48	0.07	0.19	0.91	1.54

Table 2

Deconvolution results for AAFA-10% (clustered blue points for $K = 2, 4, 5$ and 12).

k	M [GPa]	H [GPa]	f	BIC	C
2	20.80	1.26	44.64%	9033.55	53.86 3.33 0.27
3	20.26	1.20	40.29%	8825.91	49.13 3.01 0.23
4	19.83	1.16	38.04%	8762.12	46.91 2.85 0.22
5	15.63	0.77	10.30%	8733.89	12.13 0.72 0.06
6	15.73	0.77	10.41%	8715.35	12.29 0.73 0.06
7	15.46	0.73	8.67%	8704.53	11.62 0.67 0.05
8	15.53	0.74	8.77%	8690.73	11.77 0.68 0.05
9	15.39	0.71	8.95%	8683.30	12.49 0.72 0.05
10	15.43	0.74	9.06%	8679.30	11.72 0.68 0.05
11	15.32	0.70	7.38%	8671.01	12.71 0.71 0.05
12	15.21	0.70	7.57%	8662.68	11.44 0.66 0.05
Ave-grid5-12	15.46	0.73	8.89%	–	12.02 0.70 0.05

According to nanoindentation load-penetration curves and some basic parameters, nanoindentation modulus E and hardness H for material at each testing point can be calculated by Eqs. (2) to (4), based on the initial unloading stage and peak load of the load-depth curve, respectively [1,43]. All testing points with abnormal load-penetration curves were deleted before the subsequent deconvolution analysis. The following equations summarise the fundaments of nanoindentation tests used in this study:

$$d \ll h_{\max} < D/10 \quad (1)$$

$$H = \frac{P_{\max}}{A} \quad (2)$$

$$S = \frac{dP}{dh} \Big|_{h=h_{\max}} = \frac{2}{\sqrt{\pi}} E_r \sqrt{A} \quad (3)$$

$$\frac{1}{E_r} = \frac{1 - v^2}{E} + \frac{1 - v_i^2}{E_i} \quad (4)$$

where P and h are the indentation load and indentation depth, respectively; S is the initial unloading stiffness; E_r is the reduced elastic modulus, which contains the elastic response of both indenter tip and tested material; E_i and v_i are the elastic modulus and Poisson's ratio of the indenter, respectively; E and v are the corresponding parameters for sample; the A is the projected contact area.

2.3. Deconvolution

Maximum likelihood estimation [44] was adopted for the result analysis. The micromechanical properties distribution of AAFA was assumed as a Gaussian mixture model (GMM) as in Eq. (5). Each phase was treated as two-dimensional Gaussian distribution with probability density function shown as in Eq. (7), where $x = (M, H)^T$ is a column vector.

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x | \mu_k, \Sigma_k) \quad (5)$$

$$\sum_{k=1}^K \pi_k = 1 \quad (6)$$

$$\mathcal{N}(x | \mu, \Sigma) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right) \quad (7)$$

where π_k , μ_k and Σ_k are the corresponding weighting coefficient, mean value and covariance of the k^{th} component, respectively.

The log-likelihood function is given by Eq. (8) and the mean value (μ), weighting coefficient (π) and covariance matrix (Σ) that make the likelihood function achieve the maximum value for the data set of observations are the estimated parameter values.

$$\ln L = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k) \right\} \quad (8)$$

EM algorithm was used to achieve this purpose. The E step is shown in Eq. (9) and M step shown in Eqs. (10) to (12), where $x_n = (M_n, H_n)$ is a row vector with the modulus and hardness data from the n^{th} nanoindentation point. In the E step, $\gamma(z_{nk})$ is a posterior probability that can be regarded as the responsibility that component k takes for 'explaining' the observation x_n [44], or understood as the probability of the given n^{th} nanoindentation point belongs to the k^{th} component in this case. In the M step, the parameters were re-estimated by the current responsibilities. Iteration calculation of E and M steps was made until the convergence of the log-likelihood. K-means algorithm was combined used to find the suitable initial values for parameters to decrease the time for iteration calculation. At least 1000 times of repeated calculation with random initial input values were made to find global maximum log-likelihood function value. Then the final estimated parameter values and GMM model were determined.

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)} \quad (9)$$

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n \quad (10)$$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k) (x_n - \mu_k)^T \quad (11)$$

$$\pi_k = \frac{N_k}{N} \quad (12)$$

$$N_k = \sum_{n=1}^N \gamma(z_{nk}) \quad (13)$$

$$\text{BIC} = k \ln n - 2 \ln L \quad (14)$$

where π_k , μ_k and Σ_k are the corresponding weighting coefficient, mean value and covariance of the k^{th} component, respectively. N is the total number of observed data, corresponding to nanoindentation testing data here.

In theory, GMM model can fit any type of probability distribution and higher order of model tends to fit given distribution better. In order to penalize the overfitting errors, Bayesian Information Criterion (BIC) was adopted [44,45], which is known for more heavily penalizing the model complexity than the Akaike Information Criterion especially when the amount of data N is huge. The model with minimum BIC value is the target model that has a suitable number of components. After the above calculation, the raw testing data was clustered, which was based on the estimated parameters and model by MLE. For a given nanoindentation data point, which belongs to the component where achieves the maximum posterior probability. Confidence ellipses under different confidence levels of 95%, 80% and 75% for each estimated component model were also plotted.

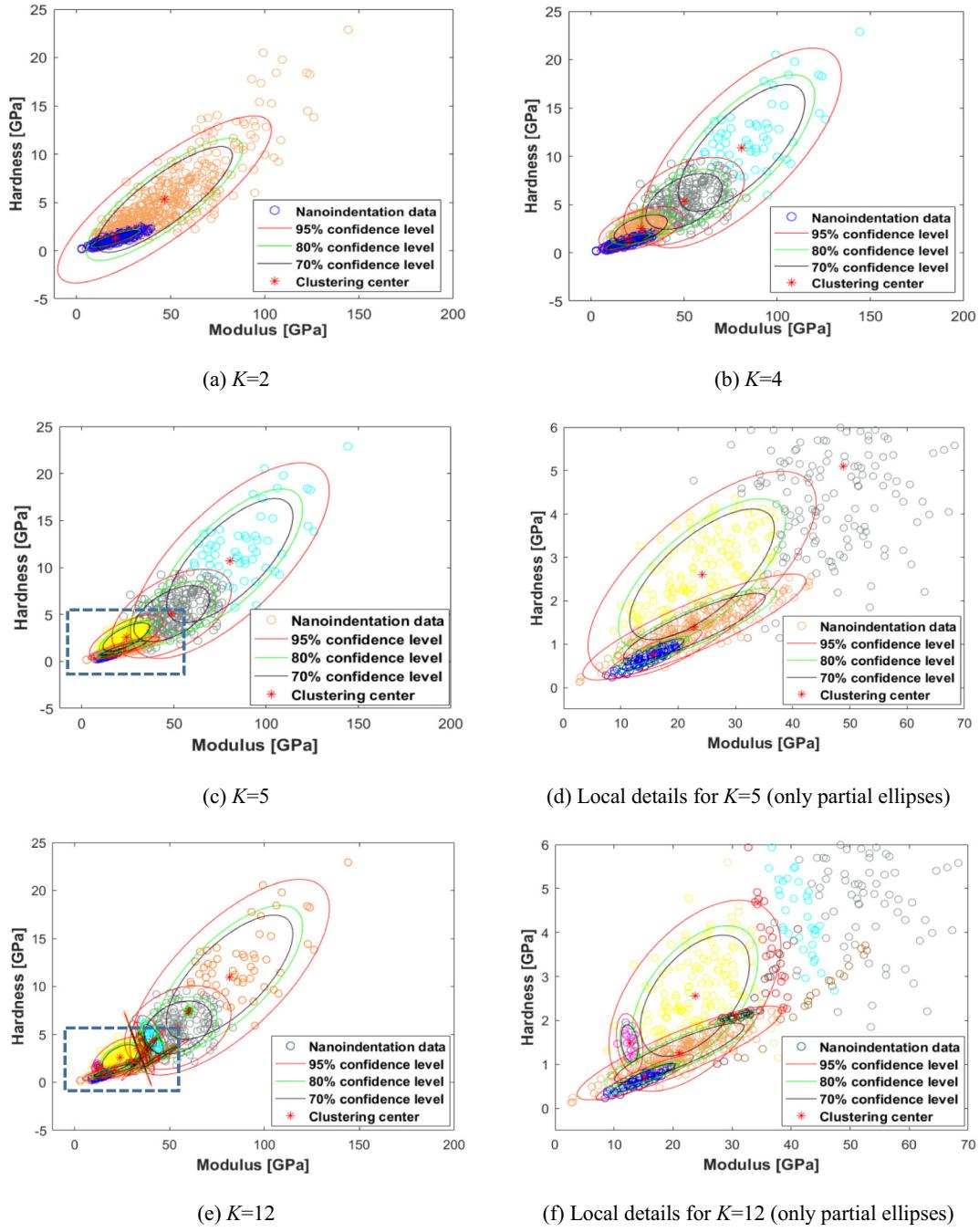


Fig. 2. Deconvolution of nanoindentation data for AAFA-10%.

3. Deconvolution results

3.1. Deconvolution for AAFA-10%

The deconvolution results for AAFA-10% were summarized in Table 2. In Fig. 2, some critical deconvolution processes were presented to reveal the variation of phases and corresponding micromechanical properties with the increase of components. It is clear that there is a good consistency between clustered data and the GMM model determined by the EM algorithm. The tilt of the axis of the confidence ellipses means that the covariance of modulus and hardness for phases are not zero, namely the corresponding correlation coefficients are not zero and there is a linear relationship between these two properties.

As a gel phase, N-A-S-H would have lower elastic modulus and hardness than unreacted fly ash particles. This feature acted as a

criterion in the initial stage to judge whether a new phase presented in the deconvolution process is the possible N-A-S-H phase. When the number of components in the GMM model was set as 2, there is a component with the modulus of 20.80 GPa and hardness of 1.26 GPa (clustered blue points). The properties of this phase change just slightly when the number of components increases from 2 to 4. However, a new phase (clustered blue points) with a lower modulus of 15.63 GPa and hardness of 0.77 GPa emerges when the number of components in the model reaches 5. More details can be found by the magnified area in Fig. 2(d). The new phase even presents when the number of components for the GMM model increased to 12 as shown in Fig. 2(f) and Table 2. For this study, nine 10×10 grids were set for AAFA-10%. The total number of test points here is more than those usually used in the nanoindentation test for Portland cement paste, AAFA and alkali-activated slag paste. However, it should be noted that the BIC still has not

Table 3Deconvolution results for AAFA-6% (clustered blue points for $K = 2$ and 4).

k	M [GPa]	H [GPa]	f	BIC	C
2	31.74	3.55	65.78%	7315.92	234.90 31.32
					5.98 1.17
3	13.42	0.68	11.66%	7158.23	21.23 0.09
					1.17
4	12.75	0.63	10.83%	7086.09	16.98 0.87
					0.87 0.06
5	12.74	0.63	10.83%	7052.31	16.95 0.87
					0.87 0.06
6	12.71	0.62	10.76%	7035.45	16.77 0.86
					0.86 0.06
7	12.73	0.62	10.81%	7014.24	16.85 0.86
					0.86 0.06
8	12.55	0.62	10.14%	7001.98	15.88 0.82
					0.82 0.06
9	12.71	0.62	10.20%	6993.97	15.95 0.84
					0.84 0.06
10	12.72	0.63	10.54%	6984.93	15.59 0.84
					0.84 0.06
11	12.54	0.61	10.60%	6979.63	15.93 0.80
					0.80 0.06
12	12.88	0.63	10.97%	6938.78	17.43 0.92
					0.92 0.07
Ave-grid 4-12	12.70	0.62	10.67%	–	16.48 0.85 0.85 0.06

Table 4Deconvolution results for AAFA-8% (clustered blue points for $K = 8$ and 12).

k	M [GPa]	H [GPa]	f	BIC	C
2	31.66	2.94	72.15%	11,191.64	178.21 17.28
					3.03
3	23.39	1.66	45.36%	11,004.61	65.51 3.66
					0.54
4	24.73	1.45	37.08%	10,930.86	85.58 4.82
					0.39
5	24.62	1.44	36.82%	10,893.79	84.00 4.72
					0.38
6	23.95	1.39	36.83%	10,870.81	71.05 3.80
					0.33
7	24.09	1.40	37.34%	10,864.59	73.57 3.95
					0.34
8	22.95	1.31	30.71%	10,850.75	61.36 3.32
					0.27
9	23.20	1.33	33.42%	10,831.49	67.43 3.53
					0.29
10	22.73	1.30	30.57%	10,823.63	60.10 3.22
					0.27
11	22.96	1.31	30.84%	10,819.10	60.60 3.16
					0.26
12	22.69	1.29	30.25%	10,811.48	60.32 3.27
					0.27
Ave-grid 8-12	22.90	1.31	31.16%	–	61.96 3.30 3.30 0.27

reached the optimum value even 12 phases were assumed in the reacted AAFA system (for $K = 13$, BIC is 8647.16). Further calculation with more components becomes more and more difficult because it takes a longer time and hard to be calculated successfully for many random initial input values. It leads to more difficult to find the real global optimal solution. What more important is when increasing the number of components, the insufficient test data for some phases leads to greater error for estimated results. Therefore, the calculation was stopped at 12 components. For the constant phase existing from 5 components to 12 components, the average modulus and hardness are 15.46 GPa and 0.73 GPa, and the corresponding standard deviation of them are 3.47 and 0.22, respectively, which are all close to the properties of low-density (LD) C-S-H reported [2,42]. It is therefore empirically regarded as the possible N-A-S-H gel phase, and subjected to

further discuss in the next section for whether it is a pure phase, why it should be an individual phase instead of spurious phase, and the possible reasons of the small proportion of this phase.

3.2. Deconvolution for AAFA-6%

The deconvolution results of AAFA-6% are summarized in Table 3. For this sample, the deconvolution process is similar to AAFA-10% and the possible N-A-S-H phase appears when K is 3 earlier than that of AAFA-10%. It becomes a stable phase when K is 4 and shows almost the same value even K reaches 12. This stable phase has average modulus, hardness and fraction of (12.70, 0.62, and 10.67%). The average standard deviation for modulus and hardness are 4.06 and 0.24, respectively, which are similar to that of AAFA-10%.

3.3. Deconvolution for AAFA-8%

For the deconvolution of AAFA-8%, the results are different from those of AAFA-10% and AAFA-6%. As shown in Table 4 and Fig. 4(a) to (d), with the increase of the number of components, a phase presents when K is 3 and shows as a stable phase in the GMM model with 8 components to the model with 12 components. However, based on the tested results for AAFA-6% and AAFA-10%, this phase is not accepted as a possible N-A-S-H gel phase. The modulus and hardness of this phase are more than normal properties found for the possible gel phase in this study. Besides, the high variance of 61.96 and 0.27 for modulus and hardness also indicate that this phase may not be a pure N-A-S-H gel phase. Typically, when comparing with the deconvolution results for AAFA-10%, there is a phase with very similar properties. This similar phase can be seen in Table 2 and Fig. 2(a) and (b) (clustered blue points) when K is 2, 3 and 4. Actually, as can be seen from Figs. 2(b) to (f), the possible N-A-S-H phase is separated from this phase. In order to acquire the corresponding N-A-S-H phase as AAFA-10%, further deconvolution for AAFA-8% is necessary. However, further increase of components in this model is not efficient to deal with this problem as illustrated previously, and attempts for the GMM models with 13, 14 and 15 components also failed to separate the possible N-A-S-H gel phase.

In theory, upper-level data would be better to be used as the original data of the second deconvolution. Namely, choosing the clustered data of the phase which corresponds to the lowest modulus and hardness in the GMM model with $1 \leq K < m$. m is the number of components where the stable phase with the average modulus of 22.90 GPa and hardness of 1.31 GPa is observed for the first time. This would reduce the risk of information loss of meaningful data but didn't achieve efficient results in this case. In order to simplify the deconvolution process, the nanoindentation data that have been clustered to the phase with the average modulus of 22.90 GPa and hardness of 1.31 GPa were used as the data directly and subjected to the second deconvolution. The clustered data obtained at 12 components are adopted for this deconvolution process.

The results from the second deconvolution display in Fig. 4(e) to (g) and Table 5. Fig. 4(e) shows the clustered data that belong to the stable phase when K is 12 in the first deconvolution, and subjected to the second deconvolution with one component. As the results listed in Table 4 ($K = 12$) and Table 5 ($K = 12-1$), there is a good agreement between the previous Gaussian model for this stable phase and the second time's Gaussian model estimated based on the clustered data. This means it is reasonable to use this data to represent the original stable phase. Similar to the first deconvolution, there is also a stable phase in the second deconvolution that starts at the model with 3 components and lasts until model with 7 components. When comparing with the results of AAFA-10% and AAFA-6%, this phase was regarded as the possible N-A-S-H phase as they show similar modulus, hardness and covariance matrix value. When K is higher than 7, although smaller BIC value still can be obtained, there will be no more stable phase, and

Table 5

Deconvolution for clustered data belong to stable phase when K is 12 (AAFA-8%, clustered yellow points for $K = 12\text{-}3$ and $12\text{-}7$).

k	M [GPa]	H [GPa]	f	BIC	C
12-1	22.71	1.29	100.00% (30.25%)	1473.28	58.54 3.14 0.25
12-2	22.20	1.24	93.54% (28.30%)	1460.87	58.45 2.96 0.22
12-3	14.03	0.56	19.90% (6.02%)	1446.38	16.84 0.75 0.75 0.04
12-4	14.19	0.58	21.06% (6.37%)	1435.05	16.90 0.79 0.05
12-5	14.67	0.58	20.17% (6.10%)	1425.61	19.83 0.84 0.84 0.04
12-6	15.26	0.63	19.99% (6.05%)	1419.11	19.36 0.81 0.81 0.04
12-7	14.79	0.60	22.74% (6.88%)	1416.02	19.68 0.91 0.91 0.05
Ave-grid 3-7	14.59	0.59	20.77% (6.28%)	-	18.53 0.82 0.82 0.05

Note: the fraction without brackets is the proportion of the component to all components in the second deconvolution, and the fraction within brackets is the proportion of the component to all components in original AAFA reaction system.

phases with smaller modulus, hardness and very low percentage occur. As the limited number of data here, more components for deconvolution were regarded as an excessive division.

4. Results analysis and discussion

4.1. Pure N-A-S-H phase properties

When conducting the nanoindentation test for C-S-H or N-A-S-H in cement paste material, there are several critical issues that determine whether the micromechanical properties of gel detected are based on pure individual phase [1]. Typically, intermix pores or other small phases in the gel phase are suspected to influence the accurate measurement of modulus and hardness of the gel phase. Choosing the suitable indentation depth and deleting the abnormal indentation test points help partially avoid these problems but may not be all.

In order to verify these issues, Davydov et al. [15] conducted comparative research and pointed that part of the CH is mixed with C-S-H because the percentage of CH detected by nanoindentation is lower than that by Back-scattered Electron Detector (BSE) and X-ray powder diffraction (XRD). Besides, even test points precisely conducted on the same target phase C-S-H_{HD}, big scatter properties results were obtained and the peak corresponding to CH can even sometimes be observed directly on the frequency densify figure. The number of component K was set as 4 to keep it the same as the K usually used in the nanoindentation test [15]. The total proportion of low-density (LD) C-S-H and high-density (HD) C-S-H reached a high level of 85% with only 10% CH. This proportion is not reasonable and means the existence of mixed phase as put forward by Davydov et al. [15]. Actually, these facts also hint that when the nanoindentation data is sufficient, increasing the number of components may be able to separate purer gel phase since there is a difference between the micromechanical properties of gel and gel with significant inclusions. At the same time, it should be noted that the modulus and hardness of low-density (LD) C-S-H and high-density C-S-H reached a high value of (30.10, 1.25) and (36.23, 1.55), respectively in that study [15]. For the other studies [42,46] that also identified and verified a mixed phase, that phase also has high modulus and hardness and is called ultra-high density (UHD) phase. The UHD phase is considered to be an intimate nanocomposite where nanoscale CH reinforces C-S-H by partially filling the latter's gel pores. The modulus and hardness of this phase achieved a high value of 47.2 GPa and 1.6 GPa, respectively. Therefore, mixed phase is

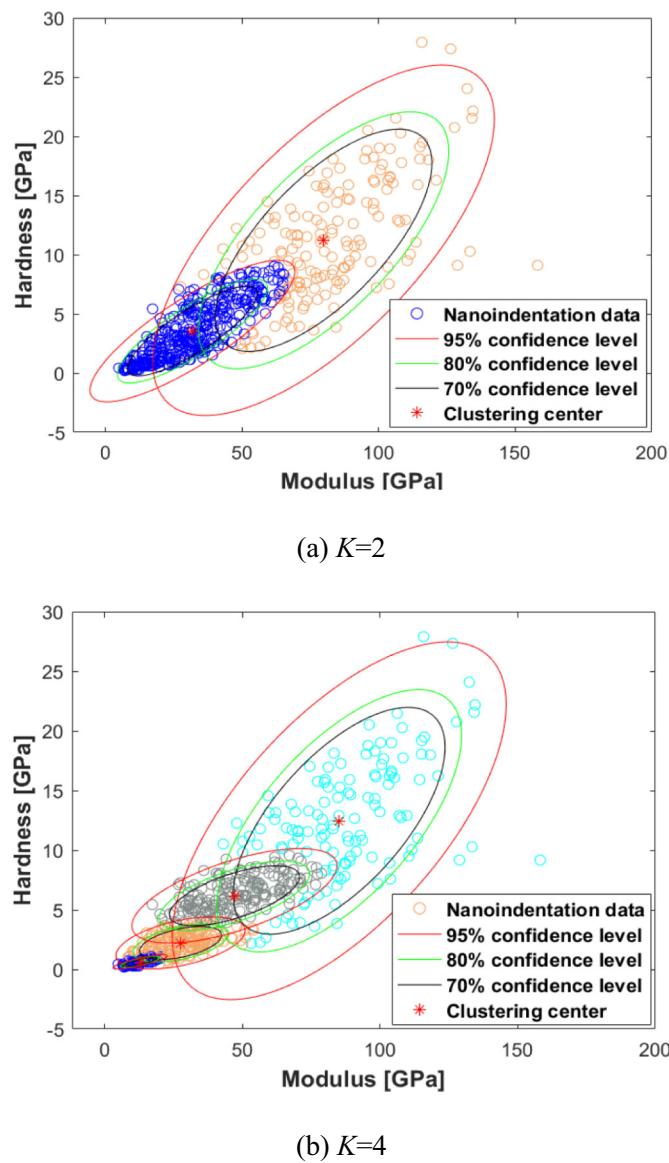


Fig. 3. Deconvolution of nanoindentation data for AAFA-6%.

characterized by high mechanical properties (e.g. H reflects strength) as the reinforcing effect by inclusions. As typically shown in Fig. 2, the clustered sandy brown points span a very large range of modulus and hardness value. Some of the points reach high modulus and hardness of almost 40 GPa and 2 GPa, respectively. Besides, the similar phase does not exist in AAFA-6%. These features indicate that the phase shows high average modulus and hardness of (22.90, 1.31) in AAFA-8% as shown in Table 4 and exists in AAFA-10% when K is 2, 3 or 4 as shown in Table 2 is one of the mixed phases instead of a single gel phase. The final stable phases separated from mixed phase with small modulus, hardness and proportion in the range of 12.7 to 15.46 GPa, 0.59 to 0.73 GPa and 6.28% to 10.67% in this study were therefore regarded as the purer gel phases. The separation process can be seen in Figs. 2 to 4, and the variation of modulus and hardness in key separation steps is given in Fig. 5.

In order to verify if the possible pure phase obtained can be further divided into purer N-A-S-H phase, deconvolution analysis was conducted on the clustered data that belongs to the 'possible N-A-S-H phase', as typically for AAFA-10% with 12 components shown in Fig. 6 and Table 6. This process is conducted for a maximum of 5 components and this number is considered as more than necessary. For 2 and 3

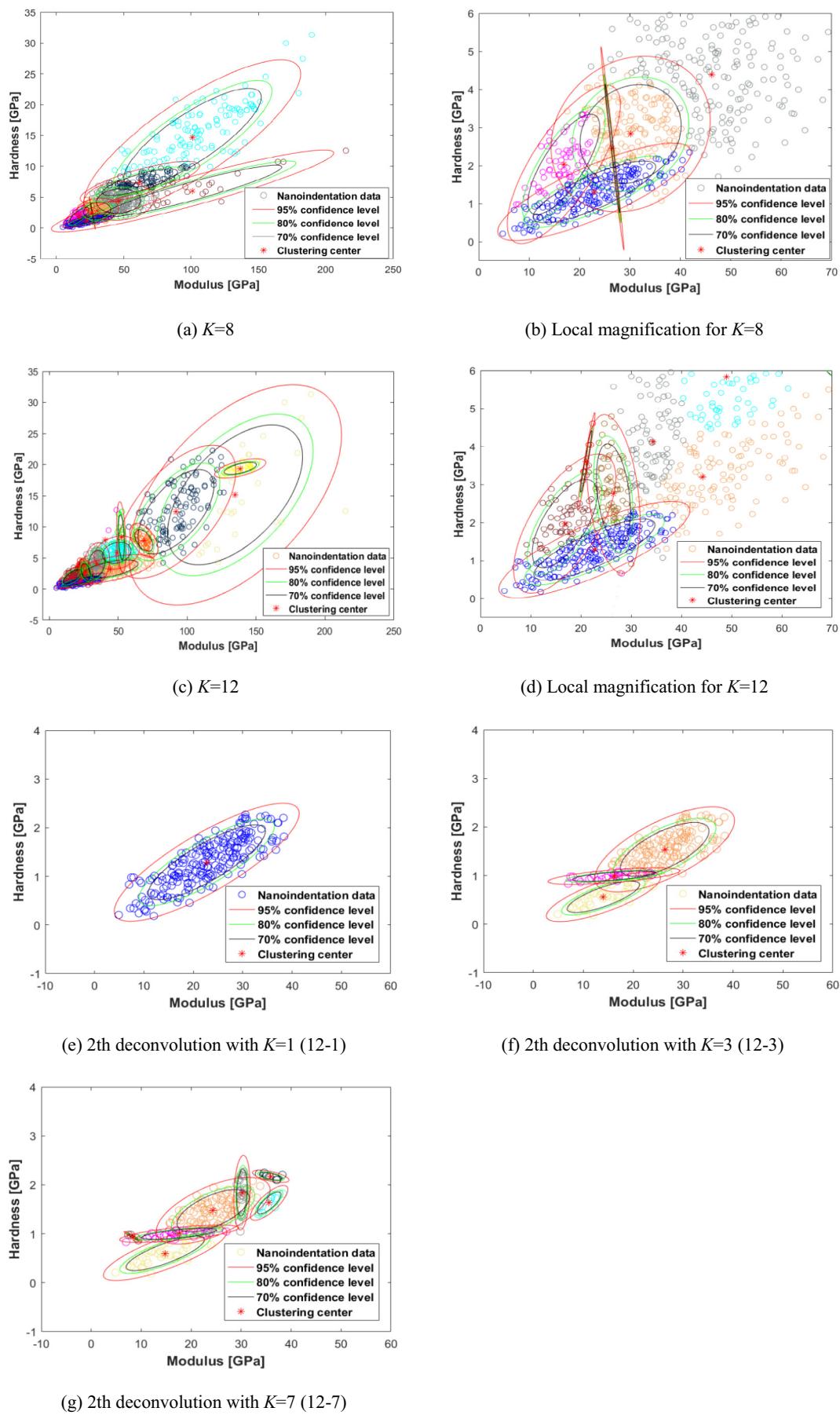


Fig. 4. Deconvolution of nanoindentation data for AAFA-8%.

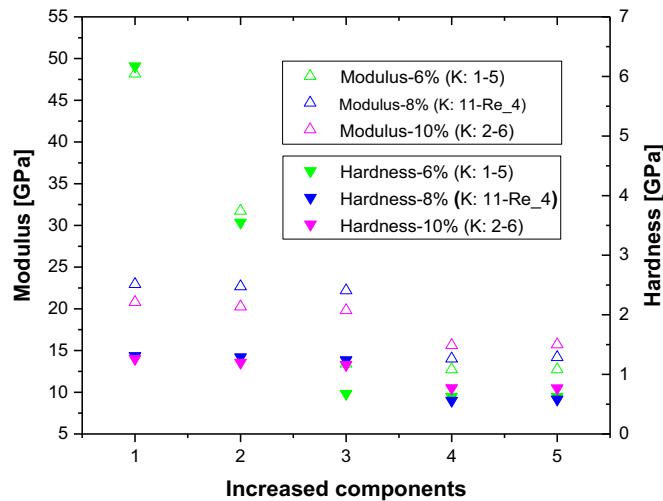


Fig. 5. Separate N-A-S-H from mixed phase in AAFA.

components, the ‘possible N-A-S-H phase’ ($K = 1$) is divided into a major phase and one or two phases with a small proportion. The major phases still show similar modulus and hardness as ‘possible N-A-S-H phase’. When K reached 4 and 5, different results present and there are two phases with considerable proportion, the clustered blue points and the clustered sandy brown points.

For this study here, even if up to nine 10×10 nanoindentation grids were tested, due to the small proportion of the ‘possible N-A-S-H phase’, nanoindentation testing points for this deconvolution are still relatively sparse, especially when points away from the centre of the Normal model. The sparse data would undoubtedly influence the clustered results and result in unreliable new phases. Besides, in the deconvolution process, some new phases are generated based on only 3 tested points (e.g. $K = 2$) or 2 tested points (e.g. $K = 3, 4, 5$). That would not be able to be identified as a phase in a test that has near 800 valid tested points due to the negligible proportion. These phases consist of 2 or 3 points also cannot be regarded as testing errors as they are at varied locations shown in Fig. 6 in the deconvolution process. Although there is a phase with modulus and hardness similar to the average results in AAFA-6% when K is 4 or 5, this phase does not show as a stable phase, and the fraction even increased with K . Besides, this phase shows much smaller proportion and variance than that in AAFA-6%. When conducting deconvolution for ‘possible N-A-S-H phase’ in AAFA-6%, the phase with the average modulus of 12.70 GPa and hardness of 0.62 GPa no more exists when reaching similar small proportion and variance, and also no stable phase appears. Therefore, the further deconvolution for the ‘possible N-A-S-H phase’ is regarded as an excessive division of this phase, at least for the number of available data here.

The second round of verification is conducted by deconvolution of clustered data that belongs to mixed phase for AAFA-10% with 3 components. The clustered data contain all possible data for N-A-S-H but do not have lots of data that belong to unreacted fly ash, which are upper-level data for N-A-S-H gel and treated as a mixed phase. This deconvolution process can more strictly separate the N-A-S-H than deconvolution of all tested data, but less strict than the deconvolution of ‘possible N-A-S-H phase’ directly. The results in Fig. 6(f) show that the ‘possible N-A-S-H phase’ presents when K is 2 and is always there even the number of components set as 10, although some data within the ellipses of the ‘possible N-A-S-H phase’ are divided into individual phases when K is 8, 9 and 10 due to excessive division. This process still proved that the N-A-S-H phase is still a stable phase at a relatively stricter extent and verified the feasibility of the way of separating N-A-S-H phase from the mixed phase for AAFA-8%.

In summary, there is a clear separation of the ‘possible N-A-S-H

phase’ from the mixed phase. This phase behaves as a relatively stable phase in the deconvolution process and is not suitable to be separated anymore based on the available data. Therefore, it is accepted as the pure N-A-S-H phase in this study. However, this is the pure phase that can be obtained by the deconvolution technique, not necessarily the real pure phase defined by the chemical composition. For the N-A-S-H phases obtained, they show similar modulus, hardness and standard deviation as previous research for the LD C-S-H gel phase in Portland cement paste.

4.2. Number of phases, bin size and feasibility of PDF by MLE

For the deconvolution of nanoindentation data, the number of phases is usually determined by empirical estimation or features of frequency distribution histogram. In this study, Bayesian Information Criterion is used for choosing the number of phases. It should be noted that the Bayesian Information Criterion aims at the whole model, thus it is not always ideal for the investigation of the micromechanical properties of pure N-A-S-H, and it can just be used for the preliminary estimation of the number of phases.

For alkali-activated fly ash, the raw material fly ash is a highly heterogeneous material composed of amorphous phases SiO₂ glass, Si-Al glass, Si-Al [Na, K, Ca] glass, etc. and also many crystalline phases such as mullite, quartz, and iron oxides hematite, with size ranges from 0.5 μm to 300 μm [47–49]. It makes the activated system more complex than the Portland cement paste. Besides, there are a considerable number of pores on the surface of fly ash particles and new pores can be generated in the dissolution process, which further increases the divergence when testing micromechanical properties. This can be observed directly from Figs. 2 to 4 for the unreacted fly ash phases, the biggest ellipse means the biggest deviation of data. All these factors make the investigation of nanomechanical properties of pure N-A-S-H more difficult due to the presence of more kinds of partly-activated and unreacted particles and mixed phases.

For most of the cases, the number of components set for current research of modulus of Portland cement paste and AAFA is 3 or 4 and bin size is 1 or sometimes 2.0 GPa, based on the LSE method for deconvolution. In this study, when adopting the MLE method, the number of phases for AAFA-8% and AAFA-10% are more than the typically used number. Besides, The LSE method obtains the mechanical properties of different phases by fitting the probability density functions (PDF), or sometimes the cumulative distribution function (CDF). The former is the most commonly used one. It can display the distribution of modulus and hardness of different phases intuitively based on frequency distribution histogram, but is also questioned for the artificial choice of the bin size. In order to make some comparisons, the PDF obtained in this study by MLE is plotted with frequency distribution histograms as shown in Figs. 7 and 8.

PDF of GMM models obtained based on a different number of phases is given in Figs. 7 and 8. As listed in Table 2, the N-A-S-H gel appears when K is 5, which can also be observed by modulus frequency density histogram in Figs. 7(a) to 7(c) for the necessity of treating this phase as an individual phase. When K is 3 or 4, the model failed to include an evident phase (gel phase) displayed by ‘green’ colour in Fig. 7(c), which leads to worse matching between the two models and histograms in the adjacent area of this phase than that for 5 components. When increased the number of components in the GMM model to 12, the corresponding PDF in Fig. 7(d) can reflect information in the modulus histogram better than the GMM model with 5 components. For hardness frequency density histogram, as revealed in Figs. 7(e) and (f), the GMM model with 12 components is even less match histogram than the GMM model with 5 components. For this phenomenon, before questioning the estimated model, the more important concerns should be the different bin sizes needed for modulus histogram and hardness histogram.

The effect of bin size was studied. Firstly, the normally used bin size of 1.0 or 2.0 GPa in deconvolution of modulus data was investigated as

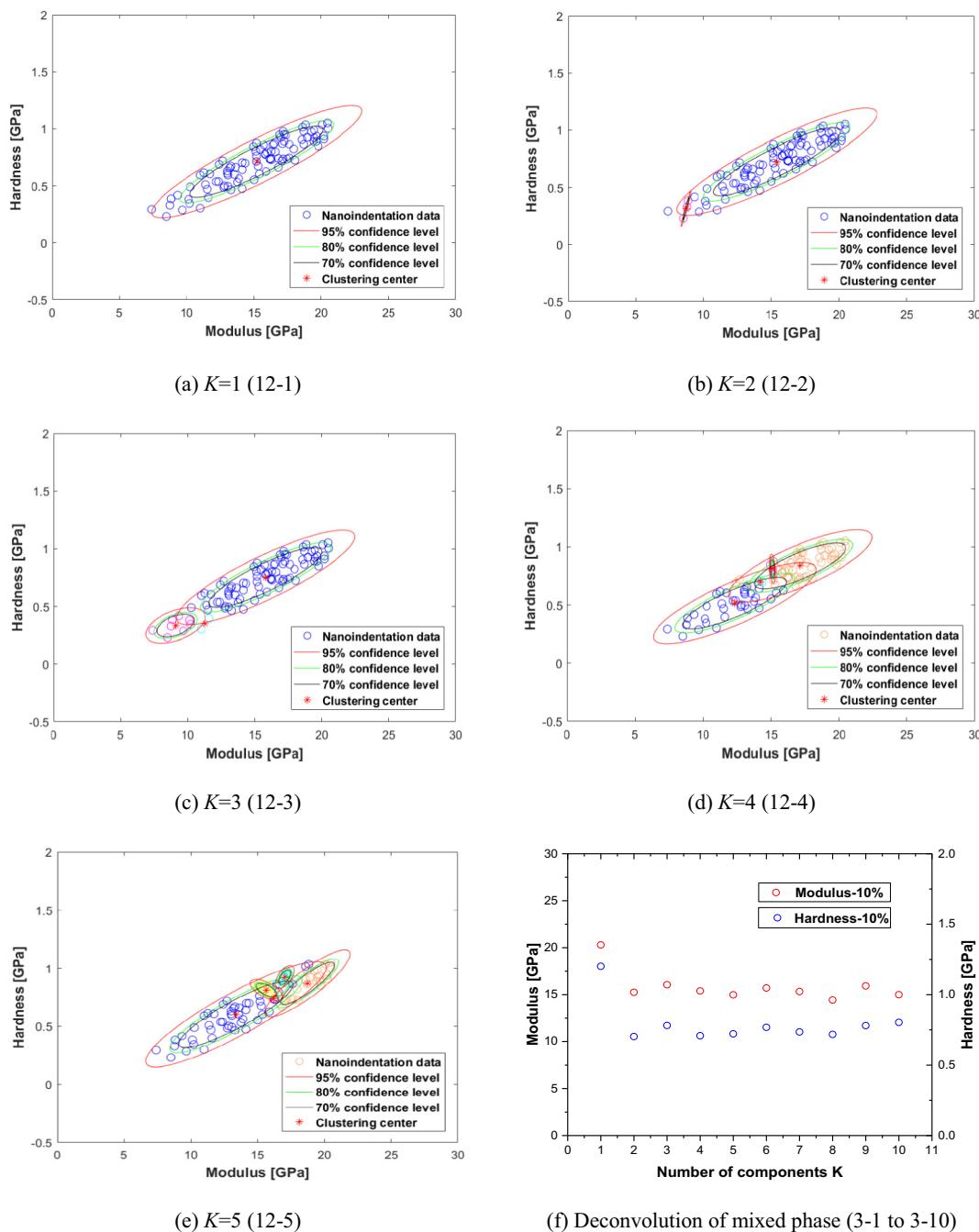


Fig. 6. Deconvolution of clustered data belongs to possible N-A-S-H (AAFA-10%, $K = 12$) and mixed phases (AAFA-10%, $K = 3$).

shown in Figs. 7(d) and 8(a). When the number of bin size only increased to 2.0 GPa, there is no distinct difference for the frequency distribution histogram. In order to reveal more information from the histogram, bin size was decreased to 0.7 GPa. Due to the large range of modulus value and corresponding low frequency density and especially the limited number of test points, the decrease of bin size leads to more discrete data instead of detailed frequency density information in Fig. 8(b). Due to the significantly smaller distribution range and higher frequency density, the change of bin size in hardness frequency histogram leads to a different phenomenon. For large bin size of 2.0 GPa, as shown in Fig. 8(c), it is difficult to identify hardness distribution information of individual phases from the histogram, and the peaks of the histogram clearly deviate from the peaks in the histogram with a bin size of 1 as shown in Fig. 7(f). When decreased the bin size from 1 to 0.7 GPa as shown in Fig. 8(d), the histogram is also markedly changed,

which means the bin size adopted would influence the LSE results for hardness. Further decrease of bin size to 0.4 and 0.3 GPa is accompanied by a better match of the PDFs estimated by the MLE method with histograms. It is clear that some extra information (peaks) didn't show in the histogram in Fig. 7(f) has appeared. Namely, for the model that can match modulus histogram with the bin size of 1, it matches the hardness histogram well only when the bin size decreased to at least 0.4. For a given test data set, it is very difficult to specify the bin size reasonably for the LSE analysis. The needed bin size is at least determined by factors such as the nature of samples, the types of the histogram and the number of the test data. The inappropriate bin size may either lose the real information or lead to spurious peaks and irregular histogram.

When using the LSE method, enough test points are needed to allow reasonably small bin size to show the details of distribution and avoid

Table 6

Deconvolution for clustered data belong to possible N-A-S-H phase (AAFA-10%, $K = 12$).

k	M [GPa]	H [GPa]	f	BIC	C
12-1	15.21	0.71	100.00%	288.04	10.19 0.60 0.04
12-2	15.41	0.72	96.88%	280.26	8.76 0.47 0.03
12-3	15.80	0.75	90.38%	218.08	7.40 0.39 0.03
12-4-1	17.16	0.84	56.75%	209.82	4.66 0.22 0.02
12-4-2	12.33	0.52	37.65%	209.82	6.01 0.29 0.02
12-5-1	13.38	0.60	58.26%	178.59	7.65 0.46 0.03
12-5-2	18.74	0.87	24.59%	178.59	1.70 0.13 0.01

the spurious peaks and irregular histogram. This point can be observed by the histograms from 5 grids in Fig. 9, where 5 grids-1st means 5 grids randomly chosen from 9 grids and 5 grids-2nd consists of the remaining 4 grids and 1 random grid from the 5 grids-1st. It is obvious that when the bin size is 1, the shape of the histogram of 5 grids-1st and especially the 5 grids-2nd is more discrete and irregular than the corresponding histogram of 9 grids in Fig. 7(c), increasing the difficulty of fitting by LSE. Besides, as shown in Fig. 10(a) and (b) and Table 7, the deconvolution results of the 5 grids-1st based on 5 components are quite close to the results from 9 grids. It means that the PDF determined by the MLE method from 5 grids-1st still can reflect the micromechanical properties distribution of AAFA and is reliable. The mismatch of PDF and histogram in Fig. 9(a) implies that when the test data are not enough, specifying of small bin size forcibly for a histogram and then fitting it by the LSE method would cause deviation in results. In Fig. 9(b), in the result of 5 grids-2nd, the corresponding deconvolution result with 5 components doesn't generate the gel phase. However, the gel phase occurs when the number of components specified for deconvolution is raised to 6 as can be observed in Fig. 10(c) and (d) and Table 7. Decrease of grid number from 9 to 5 significant increases the difficulty for LSE method as it is dependent on the histogram, while the MLE method helps to reveal the properties of gel phase and other key phases from all the three kinds of data sources and the results from 9 grids, 5 grids-1st, and 5 grids-2nd are similar to each other. It clearly displays the reliability and stability of the MLE method. In addition, the deconvolution process again indicates that the number of components needed to generate the gel phase in the model should not be a constant value of 3 or 4 and would change with both sample and test factors.

In fact, even the suitable bin size based on enough tested points is specified, it is still difficult to identify the N-A-S-H phase from a histogram. It may be ignored or be combined into a big peak becoming properties of a mixed phase when using the LSE method. Typically, as shown in Figs. 7(c), (d) and 8(b), the 'green' Gaussian distribution is the N-A-S-H phase, but it does not correspond to any recognizable peaks in histogram or PDF of GMM model. At the same time, the 'blue' Gaussian distribution in Figs. 7(d) and 8(b) under a prominent peak is easy to be considered as the gel phase. In this study, a two-dimensional GMM model is adopted, which can reveal both modulus and hardness of each phase at the same time. The 'blue' Gaussian distribution phase corresponds to clustered purple points in Fig. 2(f) and has modulus and hardness of 12.62 GPa and 1.48 GPa, respectively. This phase shows high hardness, small proportion (2.06%) and negative correlation coefficient between modulus and hardness. Therefore, it is not accepted as a gel phase. As for hardness histogram, even the bin size decreased to 0.4 or 0.3 GPa, in Fig. 8(e) and (f), it is still easy to include the N-A-S-H ('green' Gaussian distribution) into the big peak in the dotted ellipses rather than a single phase especially when a small number of phases is

set. The results indicate that there are challenges for using the LSE method to investigate the N-A-S-H phase.

For this study, it is highly suspected that the BIC cannot reach optimal value even for a model with 12 components is caused by disturbance from limited and discrete data. However, what undoubtedly is that at least 5 components (for AAFA-10%) are needed to obtain the pure N-A-S-H phase and more components are needed when describing the micromechanical distribution of the whole model. These are caused by the highly heterogeneous components of fly ash and the complex reacted system, and also the limited data. It should be mentioned that for the estimation method itself, better fitting of the model from the mathematical perspective should be at the cost of introducing spurious phases even if the BIC is used. Part of the analysis based on 12 components presented above aims to illustrate that the model estimated by MLE and BIC can reflect the overall micromechanical distribution of AAFA but does not mean that 12 components conform to the real situation in AAFA. The most reasonable model that can both reflect the real nano/micromechanical distribution and also conforms to the real number of components in AAFA may have the number of components > 5 and < 12 , but that is not the focus of this study.

4.3. Small proportion of N-A-S-H gel

The deconvolution results for AAFA with different alkali concentrations are listed in Table 8 and shown in Fig. 11. All the fraction, average modulus and hardness vary in a small range. For the fraction of N-A-S-H gel, it is in the range of 6.28% to 10.67%. Even for AAFA-6%, which achieved the maximum value, the proportion of N-A-S-H gel of it is still far less than the normally reported range for C-S-H gel. Easy to mix or interact of the gel phase with other phases is considered as the main reason that largely decreases the proportion of pure N-A-S-H gel detected. The "other phases" include partly-activated or un-activated small particles in the highly heterogeneous AAFA system. Besides, mixing of C-S-H gel with nanoscale CH crystal is the most common form of the mixed phase detected in the research of Portland cement paste [15,46]. CH crystal does not exist in the AAFA and nanoscale crystal is not considered in any current nanoindentation investigation of AAFA. However, the AAFA has crystal phases and some micron-sized crystals can be observed directly by SEM as in Fig. 12.

Some of those crystals would also exit in smaller sizes like nano and sub-micron scale. In addition, based on an analysis of extensive existing results, Provis et al. [40] proposed that the geopolymer binders are formed by agglomerating of nanocrystalline zeolites and amorphous gel phase, and the nanocrystalline zeolites phase cannot be detected by XRD due to the small size. The nanoindentation results in this study agree well with the phenomenon of coexistence of crystal and amorphous gel phase in binders. As given in Tables 2 and 4, the upper-level mixed phases before generating the gel phase shows average modulus and hardness of (22.90, 1.31) in AAFA-8% and (19.83, 1.16) in AAFA-10% may be the mixture of crystal and gel phase, as the mechanical properties of it is similar and slightly lower than the phase mixed by low-density (LD) C-S-H with nanoscale CH crystal [15]. Due to the low concentration of AAFA-6%, it may be difficult to form the super-saturation solution and the mixed phases mainly exist in this sample are the mixture of gel and partly-activated or un-activated small particles, thus no similar mixed phase with low mechanical properties ($1 < H < 2$) as in AAFA-8% and AAFA-10% can be found. Therefore, for AAFA-8% and AAFA-10%, the nutrition consumed for forming of crystals and especially the interaction of gel with crystals in nanoindentation test are possibly important factors leading to the low proportion of pure gel detected, which are even lower than that in AAFA-6% where the fly ash may be harder to be activated due to the low alkali concentration.

The presence of crystals can then help to more reasonably explain why the N-A-S-H gel detected in this study is less than results from other techniques. Typically, when segmenting of phases based on grey value

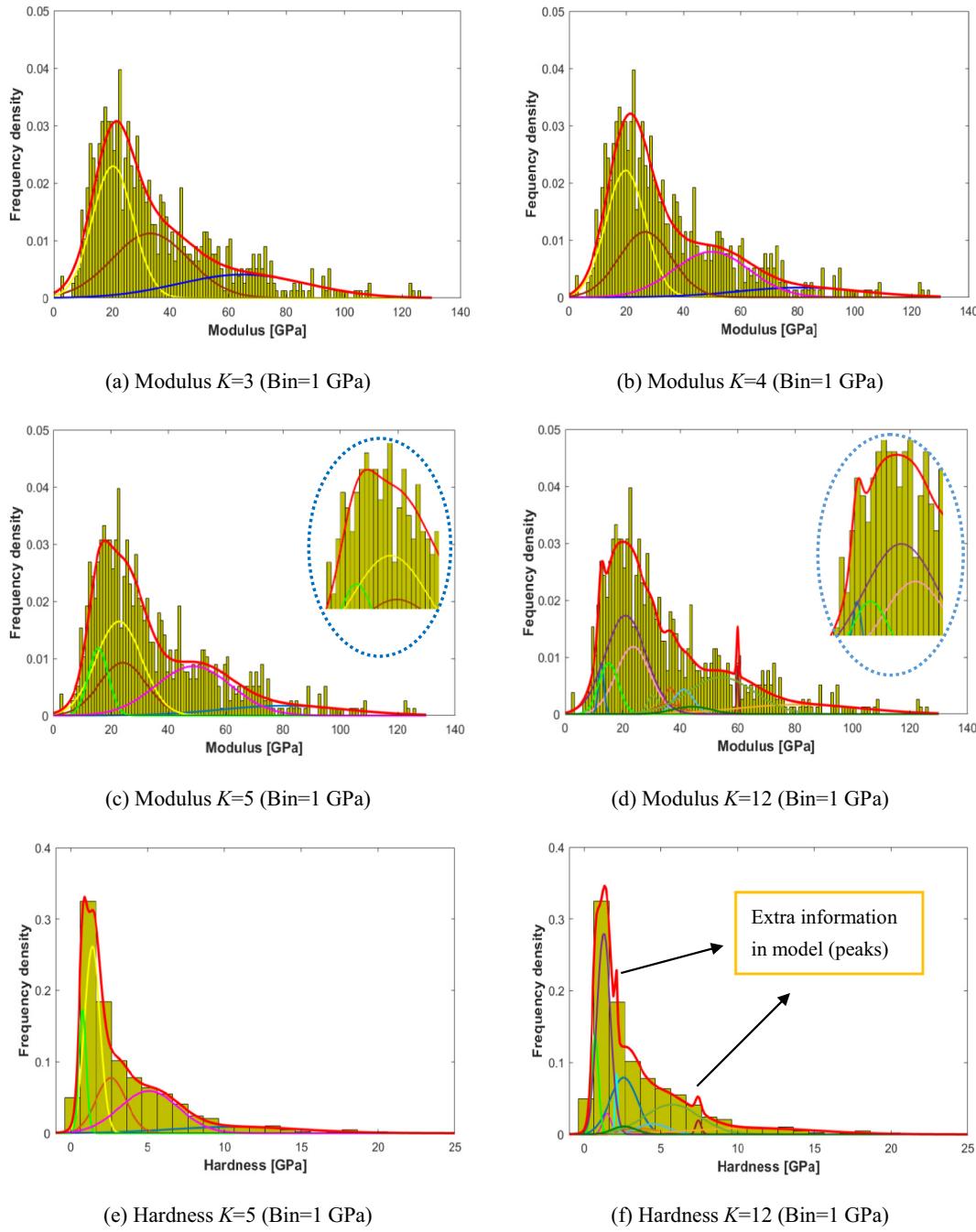


Fig. 7. Comparison on PDF obtained by MEL with frequency histogram: effect of number of components.

[16,50,51] as shown in Fig. 13, the statistical amount of reacted products in AAFA-8% is 50.38%. In fact, when referring to Table 4, for the upper-level mixed phases starting from 2 components identified by statistical nanoindentation, they all have relatively large proportions. When considering factors such as different resolutions between techniques and easy to encounter others phases (particles & crystals) in interaction zone even if the surface is identified as N-A-S-H by BSE, the final small proportion of pure N-A-S-H obtained by statistical nanoindentation is reasonable.

For micromechanical properties results in this study, when the alkali concentration increased from 6% to 10%, the results show that there is a trend of a slight increase of modulus and hardness. The micromechanical properties probability distribution of N-A-S-H was plotted in Fig. 11 for more intuitive observation. It shows that although there is an increasing trend it is in a small range and distributions are similar for

AAFA with different concentrations. For the study of Portland cement paste with different water to cement ratios, regarding the modulus and hardness as intrinsic material properties is also not based on fully consistent results, but on a small range of variation, such as smaller mean properties difference than the standard deviation, and so on [42]. This is a reasonable consideration as errors are unavoidable in any real experiments, but it also cannot eliminate the possibility that the micromechanical properties themselves just change in a very small range and the increase trend also seems plausible. Therefore, further studies are needed for understanding whether the modulus and hardness are intrinsic material properties of N-A-S-H.

4.4. Classical questions on statistical nanoindentation

As mentioned in the introduction, the LSE method has been

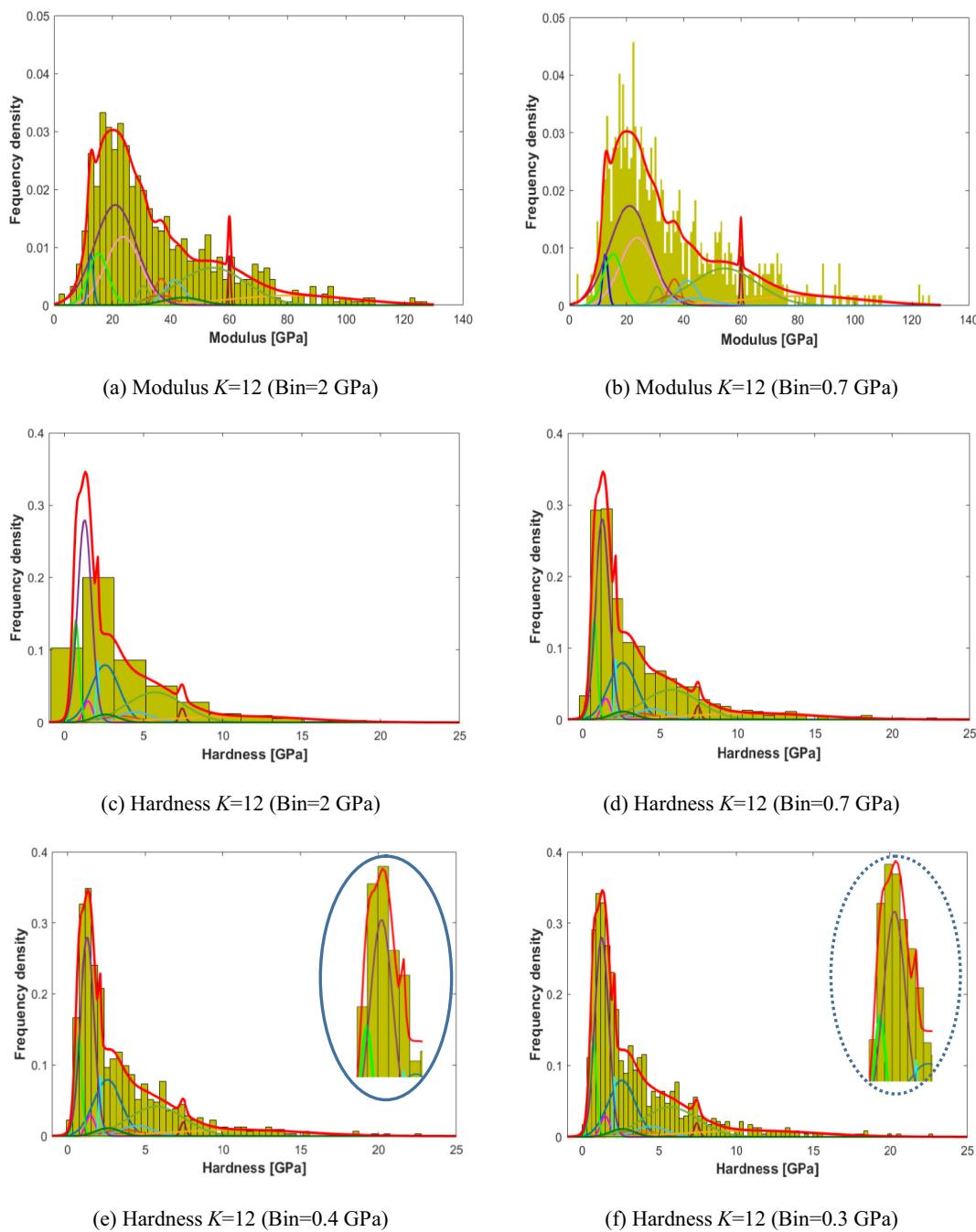


Fig. 8. Compare of PDF obtained by MLE with frequency histogram: effect of different bin size.

questioned for several aspects [13]. For the research of AAFA in this paper by MLE and BIC, the notion of MLE itself includes the method of searching global optimum. The constant gel properties in a model with a different number of components and varied grid numbers indicate the robustness of the results for gel phase. Besides, the statistic process does not depend on the distribution histogram, while the model obtained by the MLE method is expected to be able to match the histograms very well if there are infinite test data and then infinitely small bin size. These points indicate the reliability of the MLE method, which can avoid aspects that are questioned for the LSE method. Another key point for the validity of the statistical nanoindentation is the interaction volumes. The interaction volume of about $1 \mu\text{m}^3$ was thought to be too large for Portland cement paste because the collected test data are not able to present a clear peak for all known phases [13]. The two aspects of interaction volumes and statistical method constitute the main

questions for using statistical nanoindentation as an experimental approach to investigate the Portland cement paste, and combining of indentation with microstructure investigation is considered as a promising alternative way for research [13].

Regarding the issues about interaction volumes as a restriction for using statistical nanoindentation, in fact, in some aspects, this concern is modestly exacerbated in the virtual nanoindentation. The amount of clinker is just 13.8% [13]. The local sample adopted in the FIB-nt study ($14 \mu\text{m}^3$) contains much less linker than the real sample in the SNT test. Besides, the involved range of clinker should be much $< 1 \mu\text{m}^3$ cube, since it has larger mechanical properties and smaller indentation depth than the gel phase. Moreover, the ternary plot in the virtual nanoindentation is for different phases and not for mechanical properties as in SNT. The significant mechanical properties differences would magnify the peak. These factors would affect the appearance of the

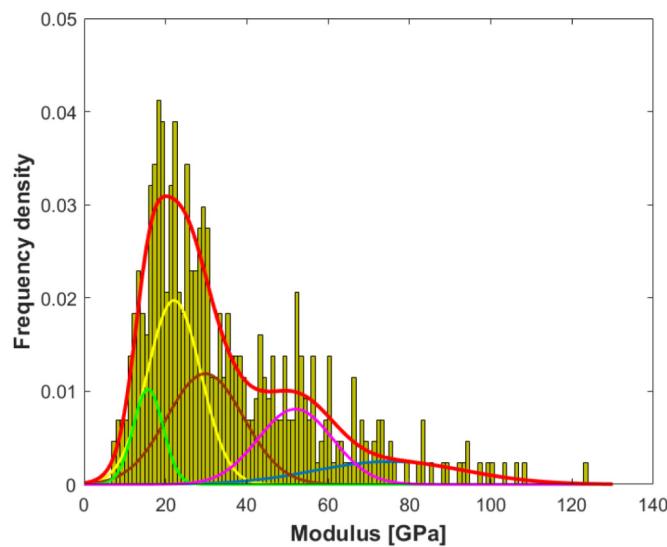
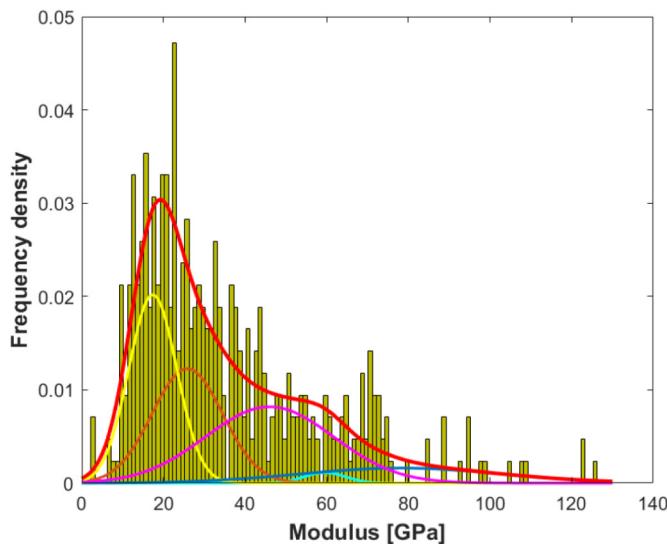
(a) 5 grids-1st ($K=5$, Bin=1 GPa)(b) 5 grids-2nd ($K=5$, Bin=1 GPa)

Fig. 9. Compare of Modulus PDF obtained by MLE with frequency histogram: 5 grids data.

peak for clinker. Even for the gel phase, the effective involved volume of the gel phase would be smaller than the $1 \mu\text{m}^3$. The $1 \mu\text{m}^3$ homogeneous cubes used is based on 3–4 times of the indentation depth, while the strain energy density field decreases with distance significantly in this range [14]. Even so, it is still difficult to obtain the mechanical properties of all phases by SNT in the complex AAFA system with limited test points.

In this study, SNT is used in a compromise way, namely, focusing on the gel phase only. The mechanical difference (combining of modulus and hardness) between gel and other phases such as mixture phases and unreacted phases is the foundation to separate them. It is a fact that the data points are density in the location around the gel phase but sparse in the locations with larger mechanical properties. Before reaching the limit of mechanical property difference to separate the gel phase from the mixture phase, spurious phases may sometimes be introduced from discrete data location, sacrificing the accuracy of other phases. The

increase of number of components from 5 for 9 grids to 6 for 5 grids-2nd is possible an evidence. However, it should be reminded that it is the gel phase that most important and also most hard to be detected while unreacted fly ash can be investigated by target indentation under BSE or even microscope equipped for nanoindentation system with high accuracy. This idea may also be used for the investigation of Portland cement paste. Even if $1 \mu\text{m}^3$ is used, there is still a clear peak for the gel phase in [13].

For the research of AAFA in this study, it is known that the Si/Al of N-A-S-H gel itself is in a range of around 1–3 [40]. Then, considering the chemical formula of mullite, quartz and possible zeolite, it is hard to identify and exclude if there are such crystals or their combination in N-A-S-H by EDS. Besides, the involved volume of EDS (Monte Carlo simulation) and nanoindentation (finite element modelling) can be matched approximately, but currently is just based on the pure phase. It may be changed for different mixed phases. Considering the above factors and possible test error, it is almost impossible to judge whether the test points are on the real pure N-A-S-H gel. In fact, the pure phase determined by deconvolution that based on mechanical properties' differences may not be 100% pure in chemical composition. In a real test, it is highly possible that, sometimes, there are small inclusions within the $1 \mu\text{m}$ strain energy density field range of the gel phase, but do not significantly change the mechanical properties detected by nanoindentation and still clustered as the gel phase. However, the mechanical properties of the gel phases obtained by deconvolution in this study would be close to the real pure gel phase for several reasons. The content of the crystal phase is much less than the gel phase in cementitious materials and the involved volume is not too large. There must be test points hitting on the area without or with just a very small content of inclusions that not change the mechanical properties significantly. When the test point is enough, this phase can be separated in deconvolution. Similarly, pore is the phase with the smallest proportion. Large pores can be removed according to the abnormal nanoindentation curves. The very small proportion of remaining pores would not change the properties of gel significantly, as the result obtained is the average value of many test points.

Empirically, the very small values should also be close to the properties of the pure phase. For Portland cement paste, even for $1 \mu\text{m}^3$ of involved volume, a very significant peak can still be found for the gel phase in the virtual nanoindentation [13]. If there is a significant peak with a large proportion in the micromechanical properties histogram, the LSE method is also expected to be able to obtain the properties of the pure LD C-S-H phase. There must be results close to the pure LD C-S-H in the summary of exiting results [2]. The smaller micromechanical properties of N-A-S-H than that of LD C-S-H [2] and the small proportion of it also indicate it may be close to the pure N-A-S-H phase. Other techniques are expected to be used in the future to obtain the mechanical properties of real pure N-A-S-H phase.

5. Conclusions

Based on the nano/micromechanical properties investigation of N-A-S-H gel in AAFA using statistical nanoindentation, the application of maximum likelihood estimation method for deconvolution of testing data was proposed for discussing critical issues in current statistical nanoindentation study. The related conclusions can be drawn as follows:

- (1) When using the MLE method, the process of separating stable N-A-S-H gel from mixed phases can be intuitively observed. Moreover, nanoindentation data can be clustered to corresponding components with a good match, which allows for further analysis. MLE also shows the advantages in independent on distribution histogram, more reasonable determination of component numbers, and less sensitive to the number of test points than LSE.
- (2) Bayesian Information Criterion is not necessary to achieve the

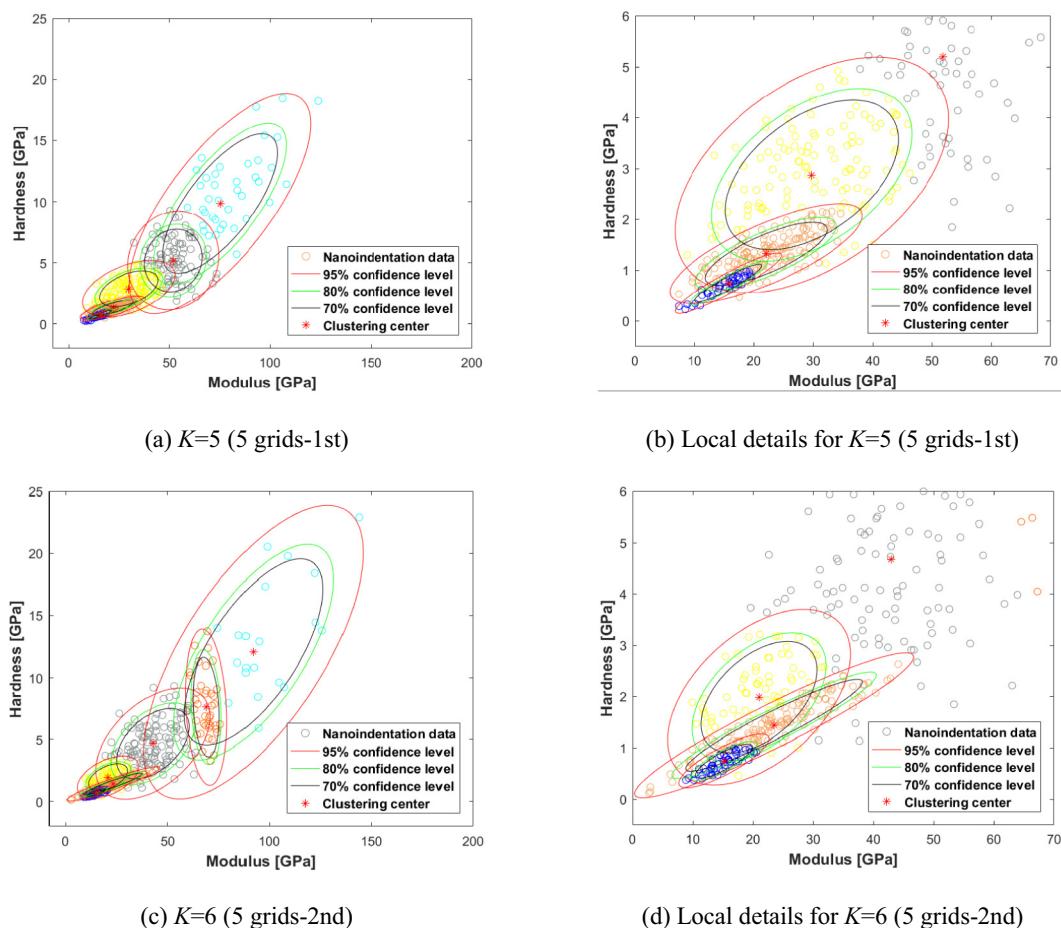


Fig. 10. Deconvolution of different 5 grids nanoindentation data in AAFA-10%.

Table 7
Deconvolution results for AAFA-10% based 5 grids (clustered blue points).

Data source	k	M [GPa]	H [GPa]	f	C
9 grids	5	15.63	0.77	10.30%	12.13 0.72 0.72 0.06
5 grids-1st	5	15.72	0.75	9.38%	13.44 0.83 0.83 0.06
5 grids-2nd	5	17.39	0.99	29.78%	34.66 2.15 2.15 0.16
5 grids-2nd	6	15.18	0.75	11.29%	9.43 0.54 0.54 0.05

Table 8
Properties of N-A-S-H in AAFA with different alkali concentration.

Sample	K	M [GPa]	H [GPa]	f	C
AAFA-6%	≥ 4	12.70	0.62	10.67%	16.48 0.85 0.85 0.06
AAFA-8%	> 12	14.59	0.59	6.28%	18.53 0.82 0.82 0.05
AAFA-10%	≥ 5	15.46	0.73	8.89%	12.02 0.70 0.70 0.05

optimum value when just focusing on the N-A-S-H phase in the highly heterogeneous AAFA reaction system. When the number of components is large, sparse data in some locations leads to mis-identification of new phases, which may also affect reaching the optimum value of BIC and increase the complexity of the model.
(3) The number of components needed to obtain the pure N-A-S-H

phase is usually more than the typically used value of 4, and even the second deconvolution is sometimes needed. Although this is partially caused by different deconvolution method used, it is still necessary that the number of components should be enough to avoid mixed phases.

- (4) GMM model obtained by MLE can match histogram well when bin size is reasonably small. A sufficiently small bin size is needed to present the actual and detailed micromechanical properties distribution of AAFA which calls for enough testing points to construct a regular and ordered histogram when using the LSE method. The N-A-S-H phase does not always show as a distinct peak in the histogram and is easy to be mixed into identifiable large peaks, which means that even if the bin size used is small enough it is still hard to obtain the properties of N-A-S-H phase by the LSE method.
- (5) The proportion, average modulus and average hardness of pure N-A-S-H phase in AAFA with different alkali concentrations vary in small ranges of 6.28% to 10.67%, 12.70 to 15.46 GPa and 0.59 to 0.73 GPa, respectively. The main reason for the small proportion is supposed to be the presence of mixed phases, such as mixing of crystals, partly-activated and un-activated small particles with the gel phase in the highly heterogeneous AAFA system, which largely decreases the volume of pure N-A-S-H gel detected by SNT.

Credit authorship contribution statement

Zhiyu Luo: Formal analysis, Data curation, Methodology, Validation, Writing - original draft, Writing - review & editing.
Wengui Li: Conceptualization, Formal analysis, Data curation, Validation, Investigation, Writing - original draft, Writing - review & editing, Funding acquisition, Supervision, Project administration,

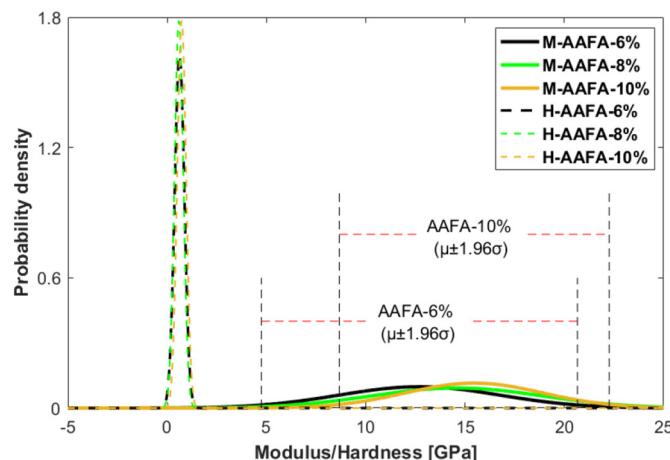
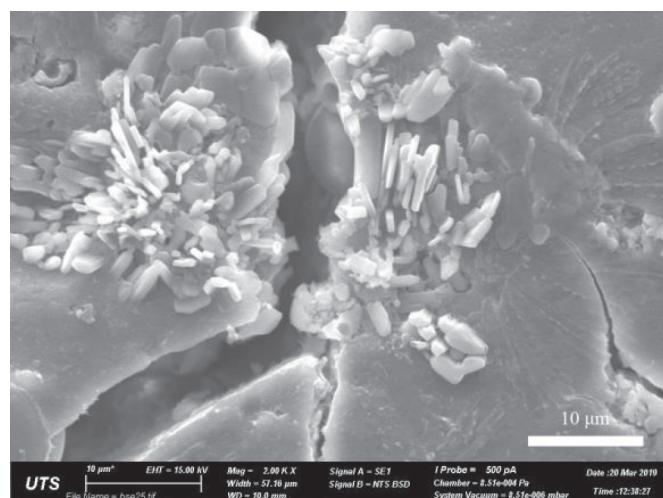
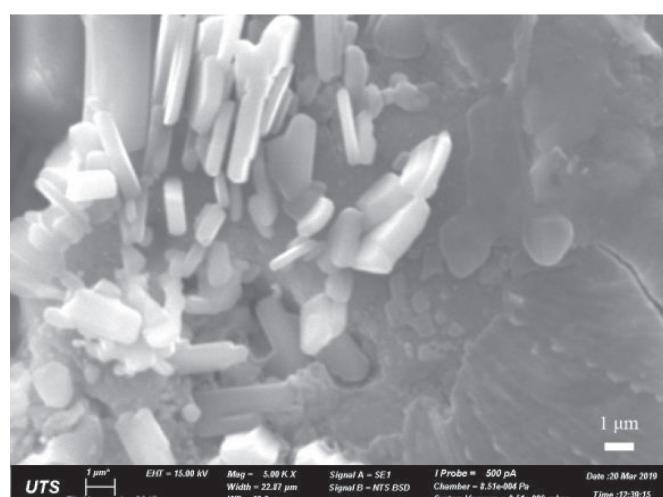


Fig. 11. Modulus and hardness probability density distribution of N-A-S-H with different alkali concentration.

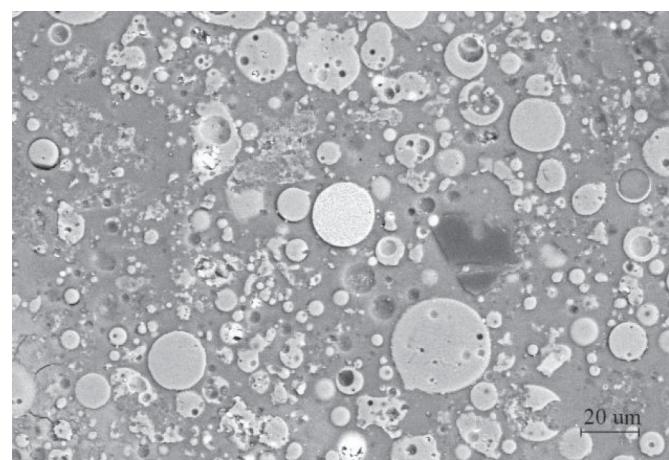


(a) Magnification of 2000 ×



(b) Magnification of 5000 ×

Fig. 12. Morphology of micron-sized crystals in AAFA.



(a) BSE image (500 ×)

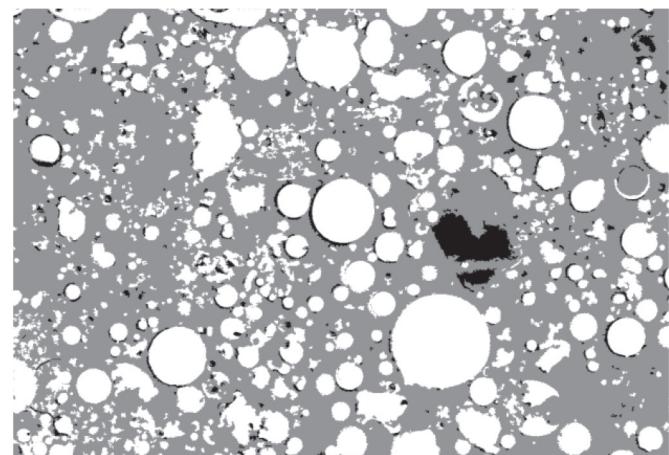


Fig. 13. Microstructures and phase segments of AAFA based on BSE images.

Resources. **Yixiang Gan:** Formal analysis, Data curation, Methodology, Validation, Writing - original draft. **Kavya Mendu:** Software, Validation, Visualization, Writing - review & editing. **Surendra P. Shah:** Formal analysis, Investigation, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors appreciate the financial supports from the Australian Research Council (ARC) (DE150101751), University of Technology Sydney Research Academic Program at Tech Lab (UTS RAPT), University of Technology Sydney Tech Lab Blue Sky Research Scheme and the Systematic Projects of Guangxi Key Laboratory of Disaster Prevention and Structural Safety (Guangxi University), China (2019ZDX004) and State Key Laboratory of Subtropical Building Science (South China University of Technology), China (2019ZA06). The first author would like to thank the support by the Australian Government Research Training Program Scholarship.

References

- [1] Z. Luo, W. Li, K. Wang, S.P. Shah, Research progress in advanced nanomechanical characterization of cement-based materials, *Cem. Concr. Compos.* 94 (2018) 277–295.
- [2] C. Hu, Z. Li, A review on the mechanical properties of cement-based materials measured by nanoindentation, *Constr. Build. Mater.* 90 (2015) 80–90.
- [3] W. Li, S. Kawashima, J. Xiao, D.J. Corr, C. Shi, S.P. Shah, Comparative investigation on nanomechanical properties of hardened cement paste, *Mater. Struct.* 49 (2016) 1591–1604.
- [4] W. Li, J. Xiao, S. Kawashima, G.S. Shekhawat, S.P. Shah, Experimental investigation on quantitative nanomechanical properties of cement paste, *ACI Mater. J.* 112 (2015) 229–238.
- [5] L. Sorelli, G. Constantinides, F.-J. Ulm, F. Toutlemonde, The nano-mechanical signature of ultra high performance concrete by statistical nanoindentation techniques, *Cem. Concr. Res.* 38 (2008) 1447–1456.
- [6] W. Zhu, J.J. Hughes, N. Bicanic, C.J. Pearce, Nanoindentation mapping of mechanical properties of cement paste and natural rocks, *Mater. Charact.* 58 (2007) 1189–1198.
- [7] W. Zhu, T. Howind, S. Barbhuiya, R. Stark, K. Bernd, D. Jorge, Nanomechanical study of cement pastes by statistical nanoindentation and peakforce QNM, Proceedings of the 2nd International Conference on Microstructural-related Durability of Cementitious Composites, RILEM Publications SARL, 2012.
- [8] J. Xiao, W. Li, Z. Sun, D.A. Lange, S.P. Shah, Properties of interfacial transition zones in recycled aggregate concrete tested by nanoindentation, *Cem. Concr. Compos.* 37 (2013) 276–292.
- [9] W. Li, J. Xiao, Z. Sun, S. Kawashima, S.P. Shah, Interfacial transition zones in recycled aggregate concrete with different mixing approaches, *Constr. Build. Mater.* 35 (2012) 1045–1055.
- [10] J. Xiao, W. Li, D.J. Corr, S.P. Shah, Effects of interfacial transition zones on the stress-strain behavior of modeled recycled aggregate concrete, *Cem. Concr. Res.* 52 (2013) 82–99.
- [11] J. Xiao, W. Li, D.J. Corr, S.P. Shah, Simulation study on the stress distribution in modeled recycled aggregate concrete under uniaxial compression, *J. Mater. Civ. Eng.* 25 (2012) 504–518.
- [12] G. Constantinides, F.-J. Ulm, The effect of two types of CSH on the elasticity of cement-based materials: results from nanoindentation and micromechanical modeling, *Cem. Concr. Res.* 34 (2004) 67–80.
- [13] P. Lura, P. Trtik, B. Münch, Validity of recent approaches for statistical nanoindentation of cement pastes, *Cem. Concr. Compos.* 33 (2011) 457–465.
- [14] F.-J. Ulm, M. Vandamme, H.M. Jennings, J. Vanzo, M. Bentivegna, K.J. Krakowiak, G. Constantinides, C.P. Bobko, K.J. Van Vliet, Does microstructure matter for statistical nanoindentation techniques? *Cem. Concr. Compos.* 32 (2010) 92–99.
- [15] D. Davydov, M. Jirásek, L. Kopecký, Critical aspects of nano-indentation technique in application to hardened cement paste, *Cem. Concr. Res.* 41 (2011) 20–29.
- [16] C. Hu, Z. Li, Property investigation of individual phases in cementitious composites containing silica fume and fly ash, *Cem. Concr. Compos.* 57 (2015) 17–26.
- [17] R. Thomas, B.S. Gebregziabher, A. Giffin, S. Peethamparan, Micromechanical properties of alkali-activated slag cement binders, *Cem. Concr. Compos.* 90 (2018) 241–256.
- [18] M.S. Reddy, P. Dinakar, B.H. Rao, A review of the influence of source material's oxide composition on the compressive strength of geopolymer concrete, *Microporous Mesoporous Mater.* 234 (2016) 12–23.
- [19] A.M. Rashad, A comprehensive overview about the influence of different additives on the properties of alkali-activated slag—a guide for civil engineer, *Constr. Build. Mater.* 47 (2013) 29–55.
- [20] A.M. Rashad, A comprehensive overview about the influence of different admixtures and additives on the properties of alkali-activated fly ash, *Mater. Des.* 53 (2014) 1005–1025.
- [21] Y. Ling, K. Wang, W. Li, G. Shi, P. Lu, Effect of slag on the mechanical properties and bond strength of fly ash-based engineered geopolymer composites, *Compos. Part B* 164 (2019) 747–757.
- [22] Z. Tang, W. Li, Y. Hu, J.L. Zhou, V.W. Tam, Review on designs and properties of multifunctional alkali-activated materials (AAMs), *Constr. Build. Mater.* 200 (2019) 474–489.
- [23] D. Khale, R. Chaudhary, Mechanism of geopolymerization and factors influencing its development: a review, *J. Mater. Sci.* 42 (2007) 729–746.
- [24] C. Shi, D. Roy, P. Krivenko, Alkali-activated Cements and Concretes, CRC press, 2003.
- [25] J.L. Provis, J.S. Van Deventer, Alkali Activated Materials: State-of-the-art Report, RILEM TC 224-AAM, Springer Science & Business Media, 2013.
- [26] J.L. Provis, A. Palomo, C. Shi, Advances in understanding alkali-activated materials, *Cem. Concr. Res.* 78 (2015) 110–125.
- [27] J.L. Provis, R.J. Myers, C.E. White, V. Rose, J.S. Van Deventer, X-ray micro-tomography shows pore structure and tortuosity in alkali-activated binders, *Cem. Concr. Res.* 42 (2012) 855–864.
- [28] J. Němcéek, V. Šmilauer, L. Kopecký, Nanoindentation characteristics of alkali-activated aluminosilicate materials, *Cem. Concr. Compos.* 33 (2011) 163–170.
- [29] S. Das, P. Yang, S.S. Singh, J.C. Mertens, X. Xiao, N. Chawla, N. Neithalath, Effective properties of a fly ash geopolymers: synergistic application of X-ray synchrotron tomography, nanoindentation, and homogenization models, *Cem. Concr. Res.* 78 (2015) 252–262.
- [30] H. Lee, V. Vimonsatit, P. Chindaprasirt, Mechanical and micromechanical properties of alkali activated fly-ash cement based on nano-indentation, *Constr. Build. Mater.* 107 (2016) 95–102.
- [31] Y. Ma, G. Ye, J. Hu, Micro-mechanical properties of alkali-activated fly ash evaluated by nanoindentation, *Constr. Build. Mater.* 147 (2017) 407–416.
- [32] H. Lee, V. Vimonsatit, P. Chindaprasirt, T. Ngo, P. Mendis, Creep properties of cement and alkali activated fly ash materials using nanoindentation technique, *Constr. Build. Mater.* 168 (2018) 547–555.
- [33] P.G. Allison, C. Weiss Jr., R.D. Moser, A. Diaz, O.G. Rivera, S.S. Holton, Nanoindentation and SEM/EDX characterization of the geopolymers-to-steel interfacial transition zone for a reactive porcelain enamel coating, *Compos. Part B* 78 (2015) 131–137.
- [34] M. Khedmati, Y.-R. Kim, J.A. Turner, H. Alanazi, C. Nguyen, An integrated microstructural-nanomechanical-chemical approach to examine material-specific characteristics of cementitious interphase regions, *Mater. Charact.* 138 (2018) 154–164.
- [35] M. Nedeljković, B. Šavija, Y. Zuo, M. Luković, G. Ye, Effect of natural carbonation on the pore structure and elastic modulus of the alkali-activated fly ash and slag pastes, *Constr. Build. Mater.* 161 (2018) 687–704.
- [36] D.K. Chanda, S.R. Chowdhury, M. Bhattacharya, A.K. Mandal, N. Dey, A.K. Mukhopadhyay, Intelligently designed fly-ash based hybrid composites with very high hardness and Young's modulus, *Constr. Build. Mater.* 158 (2018) 516–534.
- [37] G. Constantinides, F.-J. Ulm, The nanogranular nature of C-S-H, *Journal of the Mechanics and Physics of Solids* 55 (2007) 64–90.
- [38] M. Miller, C. Bobko, M. Vandamme, F.-J. Ulm, Surface roughness criteria for cement paste nanoindentation, *Cem. Concr. Res.* 38 (2008) 467–476.
- [39] C. Zhai, Y. Gan, D. Hanaor, G. Proust, D. Retraint, The role of surface structure in normal contact stiffness, *Exp. Mech.* 56 (2016) 359–368.
- [40] J.L. Provis, G.C. Lukey, J.S. van Deventer, Do geopolymers actually contain nanocrystalline zeolites? A reexamination of existing results, *Chem. Mater.* 17 (2005) 3075–3085.
- [41] R.R. Lloyd, J.L. Provis, K.J. Smeaton, J.S. van Deventer, Spatial distribution of pores in fly ash-based inorganic polymer gels visualised by Wood's metal intrusion, *Microporous Mesoporous Mater.* 126 (2009) 32–39.
- [42] M. Vandamme, F.-J. Ulm, P. Fonollosa, Nanogranular packing of C-S-H at sub-stoichiometric conditions, *Cem. Concr. Res.* 40 (2010) 14–26.
- [43] W.C. Oliver, G.M. Pharr, An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments, *J. Mater. Res.* 7 (1992) 1564–1583.
- [44] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
- [45] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* 6 (1978) 461–464.
- [46] J.J. Chen, L. Sorelli, M. Vandamme, F.J. Ulm, G. Chanvillard, A coupled nanoindentation/SEM-EDS study on low water/cement ratio Portland cement paste: evidence for C-S-H/Ca(OH)₂ nanocomposites, *J. Am. Ceram. Soc.* 93 (2010) 1484–1493.
- [47] P. Zhao, X. Liu, A.G. De La Torre, L. Lu, K. Sobolev, Assessment of the quantitative accuracy of Rietveld/XRD analysis of crystalline and amorphous phases in fly ash, *Anal. Methods* 9 (2017) 2415–2424.
- [48] K. Yan, Y. Guo, Z. Ma, Z. Zhao, F. Cheng, Quantitative analysis of crystalline and amorphous phases in pulverized coal fly ash based on the Rietveld method, *J. Non-Cryst. Solids* 483 (2018) 37–42.
- [49] F. Škvára, L. Kopecký, V. Šmilauer, Z. Bittnar, Material and structural characterization of alkali activated low-calcium brown coal fly ash, *J. Hazard. Mater.* 168 (2009) 711–720.
- [50] K.L. Scrivener, Backscattered electron imaging of cementitious microstructures: understanding and quantification, *Cem. Concr. Compos.* 26 (2004) 935–945.
- [51] A. Brough, A. Atkinson, Automated identification of the aggregate-paste interfacial transition zone in mortars of silica sand with Portland or alkali-activated slag cement paste, *Cem. Concr. Res.* 30 (2000) 849–854.