Cement and Concrete Research 141 (2021) 106327

Contents lists available at ScienceDirect

Cement and Concrete Research

journal homepage: www.elsevier.com/locate/cemconres

edxia: Microstructure characterisation from quantified SEM-EDS hypermaps

Fabien Georget *, William Wilson, Karen L. Scrivener

Laboratory of Construction Materials, LMC, EPFL-STI-IMX, Station 12, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Keywords:
SEM-EDS
Microstructure
Image analysis
Quantitative analysis
Data analysis

ABSTRACT

The characterisation of cement paste microstructure is an important step towards understanding durability mechanisms in cementitious materials. Scanning electron microscopy (SEM) coupled with energy dispersive spectroscopy (EDS) is a widely used technique to analyse the microstructure at the micron-scale. However, it is challenging, notably because the characteristic size of many phases is found on a scale smaller than the EDS interaction volume. This work presents a new image analysis framework to identify phases and quantify the microstructure of cementitious materials from SEM-EDS hypermaps. By leveraging domain knowledge, representative points are attributed to phases and mixtures of phases based on ratio plots. Then, quantitative analysis of the microstructure can be carried out (chemical composition, particle size distributions, volume fractions, …). We demonstrate the abilities of the framework, and we present possible applications and extensions of the method. The framework is available as both a graphical interface and a Python code.

1. Introduction

Scanning electron microscopy (SEM) coupled with electron dispersive spectroscopy (EDS) has many applications in cementitious materials. For example, it can be used to obtain the C-A-S-H composition [1], to analyse the composition of complex SCMs [2,3], to calculate the degree of reaction [4,5], or to interpret the results of other methods such as nanoindentation [6,7]. A particular use of SEM-EDS is the acquisition of multispectral or hyperspectral mappings. In these mappings, each pixel contains a complete EDS spectrum which can be quantified to obtain the chemical composition, and then combined with spatial information to provide mappings of all quantified elements. In combination with backscattered electron (BSE) micrographs, such mappings can be used to study hydration [2,3,8], and/or durability issues [2,9,10]. The amount of information in these hyperspectral images is large but, unfortunately, it is not easily accessible due to intrinsic difficulties of displaying and interpreting this information. The challenge is twofold: the phases must first be identified [2,3,8,11], and a meaningful quantification must then be obtained [3,11,12].

As an example, Fig. 1 shows an annotated version of a BSE image of a limestone calcined clay cement paste (LC^3) sample that will be studied in this article. An expert can already identify phases in this image, especially by using the information from EDS (either point analysis or qualitative mappings). However, this approach has limitations. First, the particles can only be identified one by one, limiting the understanding of the distribution of phases, as well as the estimation of their volume fractions. Secondly, representative quantitative information about the chemical composition may not be easily retrieved. Thirdly, from XRD analysis, we know that the ettringite content should be around 12% (in mass), although this phase cannot be identified on the BSE image! We are looking for a method which is able to solve all of these limitations.

The identification of phases is a clustering problem [6,7,13–16], which can be solved with the use of automatic clustering algorithms, such as the decision-tree [14], the SVM [15], the k-means [16], or the Gaussian mixture clustering [6,7] algorithms. Although good results can be obtained, these techniques are not widely used in practice, even if some tools (e.g. [16]) are available in well-established software, such as ImageJ [17,18]. In the opinion of the authors, this is due to the intrinsic complexity of both the algorithms and the cement paste microstructure, compared to the usefulness of the information that can be directly obtained. The clustering algorithms have been developed to analyse a large volume of data with a lower signal to noise ratio than those available in practice.

For SEM-EDS mappings, time is the main limiting factors. Studies based on a large number of hypermaps are rare. For example, the work presented here is based on just 19 hypermaps. More maps would have

* Corresponding author.
E-mail address: fabien.georget@epfl.ch (F. Georget).

https://doi.org/10.1016/j.cemconres.2020.106327
Received 1 June 2020; Received in revised form 2 October 2020; Accepted 30 November 2020
Available online 16 December 2020

improved the statistical significance of our results, but this is the condition of many studies using SEM-EDX as a tool for a mean. The main limitation is the time required to prepare the sample, analyse the sample in the microscope (a couple of hours to a day for a high-quality map) and quantify the maps (typically 10 h). Therefore, it is important to use a reproducible and reliable analysis approach (that can be checked), rather than a fully automated analysis that can treat a large amount of data. The importance of constraining the algorithms by domain-knowledge was previously recognized by Munch et al. [16], who fully integrated inputs from the expert into their algorithm. A second requirement for the adoption of these tools is the use of a graphical interface [12,16], which can provide a direct feedback to the user. Many of these algorithms require a large number of steps and parameters to be adjusted by the user. A contrast can be made with the Rietveld method for quantitative X-Ray analysis: it is similar in complexity, but the ease-of-use of common software makes it a tool available to researchers without a programming background [19]. In addition, the clustering problem and the separation of phases is just a first step. It is important to provide to the user analysis tools [11] that can be used to answer specific scientific questions. For example, below is a non-exhaustive list of research questions (from different fields of the cement material science) that could benefit from a powerful hypermap analysis tool:

- Where are the minor elements in polyclinker phases?
- Where is the ettringite in the microstructure?
- Where does strätlingite precipitate when it forms?
- Does exposure to chlorides change the morphology of AFm phases?
- What is the composition of AFm phases after exposure to chlorides?
- Can we directly quantify chloride binding in C-S-H?
- What are the characteristic features of a carbonation front?
- What is the impact of sulfates on the hydration process?
- How is the microstructure impacted by sulfate attack?

All these questions are currently under study in our lab using the tool presented here as a step in the data analysis. The importance of these scientific questions over the clustering problem, has led to the combined use of multispectral images (generally qualitative) to study the spatial distribution of phases and, EDS point analyses to study their composition [5,7,9]. However, to obtain quantitative measures of the spatial distribution, a bridge between a robust phase separation and quantitative chemical analyses is required.

In this article, we propose a new approach that allows phase separation, chemical analysis, and visualization on the BSE image. It addresses both the physical and the practical challenges of analysing SEM-EDS hypermaps of complex cement paste microstructures. The proposed method is based on a combination of (1) a denoising of quantified maps, (2) a user defined phase clustering based on domain knowledge and (3) a visual tool to interact with the user (i.e., with the domain-knowledge and the research problem). First, using a segmentation of the composite image [3,7,11], we extract representative points from the quantitative EDS hypermap. The emulation of the point analysis method using ratio plots [1,19] allows the composition of the phases to be studied, but also identifies mixtures of these phases from representative points. In addition, phase masks can be defined by translating the point clusters onto the maps. Then, particle shapes and spatial distribution can be analysed from these phase masks. This method was implemented into a new framework called *edxia*, and added into the Glue software [20,21] to make this tool accessible to all researchers. This accessibility as well as the flexibility of our method represent an important advantage compared to previous approaches. Thanks to the versatility of the approach, the scientific questions presented above were all studied with our software as part of the development process (although they could not be described in detail here).

In this paper, we present the experimental protocols for sample preparation, SEM-EDS hypermaps acquisition and quantification. Section 3 presents in details the image analysis methods and its justifications. Section 4 gives some examples of microstructural investigations enabled by quantitative phase maps, chemical analysis of phases,

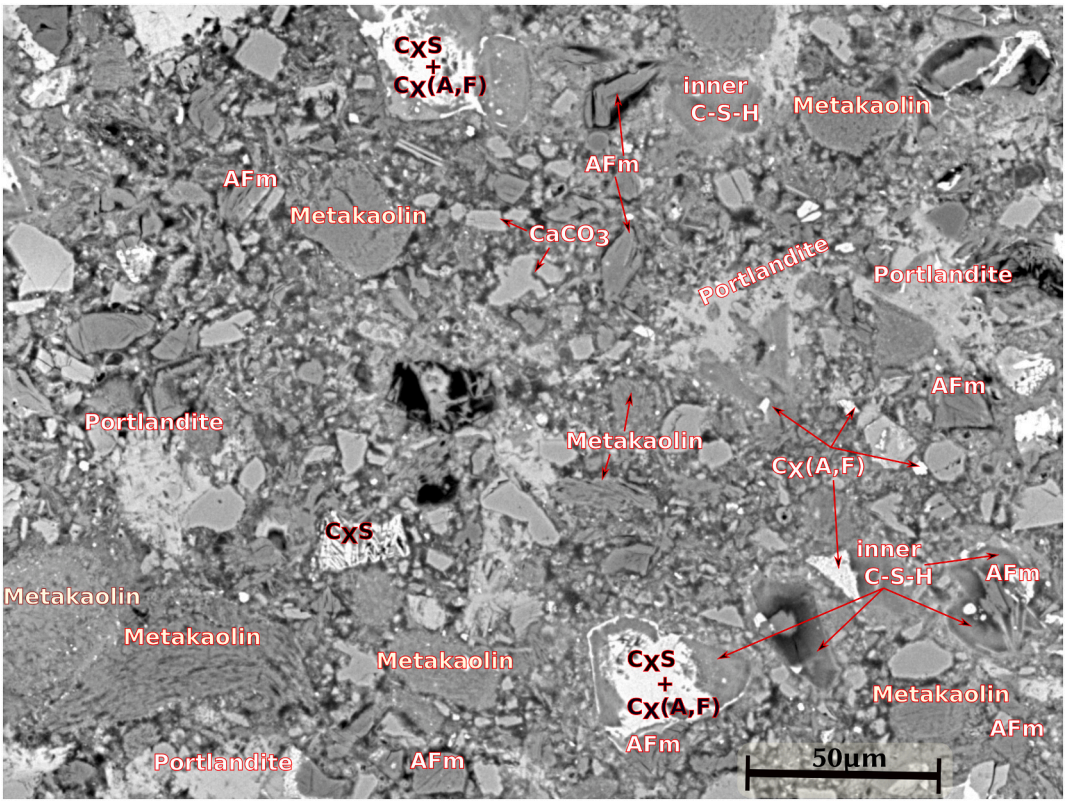

Fig. 1. Phase annotation from BSE and elemental maps of an LC^3 sample.

quantification of phase volume fractions and estimation of microstructure descriptors. Section 5 presents some additional applications for our tool: investigation of profiles, coping with lower-efficiency EDS detectors, filtering data at the spectrum level, and incorporating machine learning. Overall, this paper not only describes the method and some current applications, but also aims at setting the stage for further developments in an open-source framework.

2. Materials and methods

2.1. Materials

To illustrate the robustness of the edxia approach, a CEM I Ordinary Portland Cement (OPC) and blended system were chosen as examples of cementitious microstructures with different levels of complexity. More specifically, the blended system is a limestone calcined clay (LC3) cement paste [22]. LC^3-50 paste incorporates 53% of OPC (CEM I), 30% of calcined clay (fired and ground in India, with a calcined kaolinite content of ~45%), 15% fine commercial limestone powder and 2% chemical-grade gypsum. The chemical composition of the cement and calcined clay obtained by XRF are presented in Table 1. The LC^3 cement paste was mixed with deionized water.

2.2. Methods

2.2.1. Sample preparation

To ensure the homogeneity of the samples and the representativeness of the areas investigated, the cement pastes were mixed for 2 min with a 5 cm axial flow impeller at 1600 rotations/min. This high-shear mixing was followed by 1 min of vacuum mixing at 450 rotations/min to remove entrapped air. A typical water-to-binder ratio of 0.4 was chosen and a polycarboxylate-based superplasticizer was employed to obtain sufficient workability for casting. The paste was cast in cylindrical polypropylene molds of 33 mm diameter, which remained sealed for the first 20 h before demolding. They were then water cured for 28 days in slightly larger containers. In addition to limiting the amount of water with the container size, any possible leaching was further avoided by adding to the container a sacrificial finely-ground specimen acting as a leaching buffer protecting the sample.

At 28 days, 2 mm thick slices were cut and dried by isopropanol solvent exchange for 7 days (with renewal of the isopropanol after 1 h, 1 day and 3 days) followed by isopropanol evaporation under slight vacuum for more than 7 days (with CO2 and H20 absorbing agents). After drying, smaller pieces of the slices were polished with a silicon carbide paper #1200 (to remove any possible carbonation and provide a flat surface). The pieces were then impregnated under vacuum using the EPO-TEK 301 epoxy into 25 mm diameter molds, to support the delicate microstructure of the cement paste hydrates. The embedded samples were then polished using a Struers RotoPol-25 automatic polishing machine with Struers MD Largo discs and DP-Spray M diamond suspensions, with deodorized petroleum as lubricant. A high-quality mirror-like surface was obtained by successive polishing for 15 min with a 9 μm diamond suspension under a pressure of 15 N, followed by 1 h with a 3 μm diamond suspension at 20 N and 2 h with a 1 μm diamond suspension at 20 N. Between each step (and after each hour of polishing), the samples were cleaned in an isopropanol ultrasonic bath for 2 min. After 2 days under slight vacuum to evaporate the remaining isopropanol, the samples were carbon coated with a thickness of about 15 nm (to make the surface conductive and enable SEM analyses at high vacuum).

2.2.2. SEM parameters

The majority of SEM-EDS analyses were performed with a Zeiss Schottky SEM equipped with a Gemini®2 column. The accelerating voltage was 15 kV to enable quantification of iron (to access K lines of iron). Previous work showed that with such a voltage, 95% of the characteristic X-rays escaping cementitious materials were generated within a depth of ~900–1700 nm (depending on the phase, it's density and on the measured element) [7]. The beam current was adjusted to 2 nA (using a Faraday cup) to obtain the maximum signal without saturating the EDS detector. Although this value is higher than the recommendations by Rossen et al. [1], the damage to the sample was limited by the ultra-rapid scanning of the surface enabled by an Oxford UltimMax 170 EDS detector having a silicon drift detector surface of 170 mm^2.

Unless otherwise specified, the EDS hypermaps were acquired using the Aztec software at a resolution of 1024 × 768 pixels over regions of 280 μm by 210 μm. To avoid image quality losses due to possible drift of the surface, the Autolock drift correction was employed with a measurement every 30 s and predictive correction every 10 s. The hypermap acquisition parameters were set as a compromise to obtain sufficient counts at each pixel, while limiting both surface damage and the required machine time. The combination of 256 s dwell time per pixel with an averaging over 12 frames was found to provide good results (although other satisfactory combinations could be possible). In this condition, the acquisition time of a map is around 1 h.

To ensure the consistency of EDS analyses, the beam intensity was calibrated on a copper tape before each analysis (even if Schottky SEMs are extremely stable). The deadtime was kept under 30–35% for the different types of polished sections investigated. To achieve quantitative EDS analyses, certified standards were used for all elements of interest: $CaSiO_3$ for O and Ca, Jadeite for Na, Al_2O_3 for Al, SiO_2 for Si, $CaSO_4$ for S, KCl for K and Cl, MgO for Mg and Fe_2O_3 for Fe (standards from Micro-Analysis Consultants Ltd). After acquisition, each hypermap was quantified using the Aztec built-in QuantMap function with the processing of all elements, no normalisation, corrections for the window artefact and pulse pile up. Although the oxygen may not be perfectly measured, it was still directly quantified from the spectrum which was deemed the best option for cement pastes: stoichiometric oxygen from oxides would lead to large errors of matrix correction, because of the important non-quantified hydrogen and carbon contents. Nevertheless, the content of water and carbonates was still estimated using the sum of oxides (SOX), which sums the mass of all measured elements and their stoichiometric oxygens (considering the oxides expected to occur in the cement paste, e.g., CaO, SiO_2, Al_2O_3, Fe_2O_3, etc.)

In addition to the main analyses carried out with approach, the choice of parameters and the possibility of using lower-efficiency detectors was explored with a Quanta 200 tungsten filament SEM equipped with a 30 mm^2 EDS detector (Brucker XFlash 4030). The same standards were employed with a similar calibration and quantification approach. Different sets of parameters were used, with two types of instruments as described in Table 2.

2.2.3. Additional characterisation of the cement paste

In addition to SEM-EDS analyses, X-ray diffraction (XRD) and thermogravimetric analyses (TGA) were also done on the same LC^3 paste to provide reference phase fractions. XRD analyses were performed on fresh discs of 33 mm diameter, using a PANalytical X'pert Pro Bragg–Brentano diffractometer. The parameters were chosen according to the

Table 1
Chemical composition (mass%) of the cement and the calcined clay used in this study. LOI = loss of ignition.

Material	Na_2O	MgO	Al_2O_3	SiO_2	SO_3	K_2O	CaO	TiO_2	Fe_2O_3	LOI
CEM I	0.4	1.1	5.4	20.7	2.9	0.7	65.0	0.2	1.9	1.8
Calcined clay	0.1	0.0	36.7	56.1	0.1	0.2	0.2	2.4	3.4	0.5

Table 2
EDS detectors and hypermaps acquisition parameters used in this study.

Set	EDS detector	Resolution	Beam current	Dwell time/ frames	Total time
A	Brucker 30	1024 × 768	~6 nA	512 μs/10	~1 h
B	Brucker 30	1600 × 1200	~0.8 nA	512 μs/30	~8 h
C[a]	Oxford 170	1024 × 768	2 nA	256 μs/12	~1 h
D	Oxford 170	2048 × 1536	2 nA	256 μs/12	~4 h
E	Oxford 170	1024 × 768	4 nA	256 μs/6	~30 min
F	Oxford 170	512 × 384	2 nA	256 μs/12	~15 min

[a] Set C is the default parameter set used unless otherwise stated.

equipment available and to the recommendations of [19]: a copper tube operated at 45 kV and 40 mA; Soller slits of 0.04 rad for incident and diffracted beam; incident anti-scatter and divergence slits fixed at 0.5°; beam knife; X'Celerator linear position-sensitive X-ray detector with a length of 2.122° 2θ; scanning from 5 to 70° with a 0.017° 2θ step; spinning stage at 15 rotation per minute; 30 s per integrated step, for a scan duration of about 15 min. The patterns were analysed using the Rietveld method with an external standard [23], along with the iterative approach proposed by Li et al. [24]: the cement phases were first refined from the pattern of the anhydrous cement and then transferred to the refinement of the cement paste; this latter quantification was then normalised with the K-factor obtained using a rutile standard and the mass attenuation coefficient of the sample. This approach provides absolute mass fractions for crystalline phases and total amorphous content. TGA analyses were made with a Mettler Toledo TGA/SDTA851 instrument, considering the approach described in [19]. A sample of about 50 mg was placed in an alumina crucible and tested with a 10 °C/min temperature ramp from 30 °C to 1000 °C, in a nitrogen flow of 30 mL/min. The recorded curves were evaluated using the tangent method, mainly to estimate the portlandite $Ca(OH)_2$ content from the water loss centered around 450 °C.

2.3. Implementation of the edxia method

The new framework developed during this work, *edxia*, is coded in python [25], using the well-established scipy [26,27], matplotlib [28] and scikit-image [29] libraries. The Python scientific stack was chosen for its flexibility and extensibility. The framework is developed as an Application Programming Interface (API). It means that the user can develop their own analysis tools, and work-flows adapted to their problems. The open-source framework is freely available [30].

To make the framework available to researchers without coding experience, it was included as a plugin to Glue [20,21]. This python software was developed to easily visualize and create links between datasets. For example, Glue was initially developed to identify stars in telescope pictures from star catalogs. We used it here to identify the phases on the BSE map, using the chemical information provided by the EDS analysis. Documentation of the interface is also available [31].

3. The edxia method

To answer the challenges identified in the introduction, our method follows a sequential approach summarized as follows:

1. Maps are denoised and combined
2. Representative points are extracted from the maps
3. Ratio plots are used to identify phases by the user
4. Masks are created from the regions identified in the ratio plots, to identify phases in the microstructure.

The novelty of the method is (1) to propose a reproducible combination of methods to obtain a ratio plot than can be analysed from a SEM-EDX hypermaps, and (2) relate this ratio plots to the BSE image, including the analysis of the mixed phases. After these processing steps, different types of analyses can be run on these masks depending on the specific scientific problem. The flowchart of the complete approach is presented in Fig. 2, and each step is further described in a corresponding section of this paper.

3.1. Denoising of chemical maps

Unless the user has access to a high-end EDX detector for a long period of time, the number of counts per pixel is usually below the threshold required for a reliable point analysis [1], i.e., the maps are noisy. The noise manifests itself as brightness variations on the maps (i.e., a high-frequency texture on homogeneous regions). As an example, this texture is visible on the raw calcium map presented in Fig. 3. This texture significantly lowers the output quality of the image analysis algorithms. Therefore, it must be removed before further analysis. Although some pixels may not be reliable on their own, we can use the information distributed over the large number of pixels to refine the maps.

Many denoising algorithms exist. As a guideline, a good denoising algorithm should remove the high frequency texture from the map while keeping sharp edges. The smearing of the edges leads to an apparent increase of the interaction volume effect. In addition to the raw map, Fig. 3 shows the results from three denoising algorithms on the Ca map: the total variation algorithm [32], a Gaussian filter, and the joint bilateral filter [33,34]. The joint bilateral filter is an edge-preserving filter using a different (but related) image for edge detection. The reference image for edge detection in our workflow is the BSE image. The maps are displayed in the left column of Fig. 3, while the histograms corresponding to these maps are shown in the right column. Each algorithm depends on parameters controlling the extent of the denoising. The total variation algorithm depends on a weight parameter (here equal to 0.1), the Gaussian filter depends on a scale parameter (here equal to 0.5) controlling the width of the Gaussian window. The joint bilateral filter depends on a space parameter (here 10.0) and an intensity parameter (here 2.5). For typical SEM-EDX hypermaps of cementitious materials, we observed that the total variation algorithm provides a better denoising than the standard medium filter, that is robust with respect to the parameters. On the other hand, the Gaussian filter is a difficult trade-off between the removal of the noise and the smearing of the information with no clear optimal points. Because the bilateral filter was developed to reduce the smearing across edges, it is a good choice to limit the impact of the denoising. Therefore, it can reduce the need for a compromise, but it is quite sensitive to parameters, which should be adapted to the resolution, the field of view, and also the contrast in the BSE image.

Another method to judge the effect of denoising is to analyse the histogram of the intensity of the images. The denoising results in a refinement of the main peak, and the appearance of secondary peaks. The main central peak corresponds to the main hydrates phases such as the C-S-H and AFm phases (layer double hydroxides of the hydrocalumite family) [35]. The peak at high Ca corresponds to the clinker phases. The shoulder on the C-S-H/AFm peak is the portlandite. This shoulder is more visible for plain cement (CEM I) where the portlandite content is higher. The peaks at low Ca are the metakaolin and the mixes between metakaolin and hydration products. Even if the map is more refined after denoising, the peaks are still overlapping. Therefore, they cannot be used to quantitatively select phases on their own. It should be noted that the denoising cannot be optimized based on just the width of the peaks in the histogram. Because, it would lead to over-smoothing, where mixes of phases cannot be resolved anymore.

Denoising is not an operation without compromise, as the denoising might introduces a bias. However, ratios of composition (Al/Ca,Si/Ca,S/Al, …) were found to be quantitative on average, as mentioned by Harrisson et al. [36]. This point is developed further in Section 5.2. Since

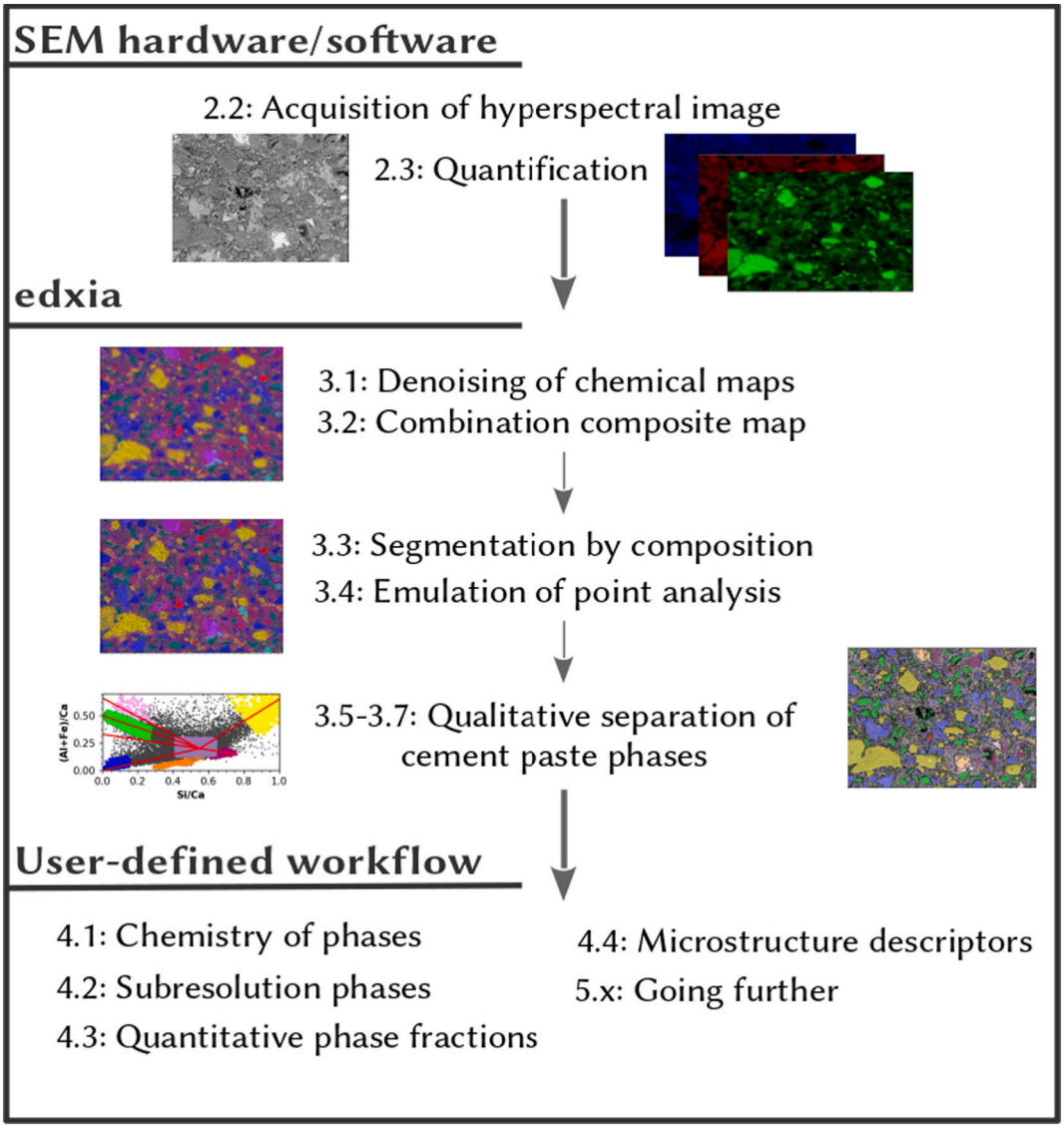

Fig. 2. Flowchart of the edxia method. Numbers indicate the section of this paper.

the total-variation algorithm provides the best results in terms of robustness and final aspect of the maps, it was used in the remainder of the examples presented in this paper. It is also the default option in the implementation of our method.

3.2. *Combination into composite maps*

A common approach to analysing chemical information, e.g. [3,11,37], is the creation of a composite map. The composite map is also used as a pre-processing step in our workflow to increase the signal to noise ratio. This composite map is obtained by merging scaled elemental EDS maps assigned to different color channels. A convention for creating composite maps is the use of Ca, Si and Al elemental maps, which allows the best separation between the main cement phases [3,38].

The individual colored maps, for the LC^3 cement paste are shown in Fig. 4. By assigning these maps into the corresponding channels of a RGB image (red = Si, green = Al, blue = Ca), a composite map can be obtained. Each individual map can further be scaled to obtain an adjusted range of colors. The composite map can also be displayed as a transparent layer over the BSE map to highlight the boundaries of the particles, or mixed with another gray-level map (e.g. Fe, Mg) to increase the visual separation of specific phases.

The resulting composite image is shown in Fig. 5a. This image can be analysed to separate the phases by color. The main phases are identified in the legend of that figure. To better identify phases by colors, the background color of the legend is set to an average color of the outer C-S-H matrix.

Phase masks could be obtained from this composite image, as usually performed by transforming the RGB channels into hue, saturation and brightness (HSB) channels [37]. The phases are then defined by thresholds on the hue. Additional thresholds might be defined on the saturation, brightness, and/or on the BSE value, to separate between anhydrous phases and hydrates. The analysis of the composite image is useful to quickly analyse a map. However, the main limitation is that it is limited to 3 components, or 4 if a gray channel is added instead of the BSE map. Another limitation is that the threshold is based on a non-physical parameter, the hue rather than the composition. Although it is sufficient to distinguish C-S-H from portlandite, it is not directly possible to separate hemicarboaluminate from the monosulfoaluminate, for example. An analysis using all the available elements is necessary [16].

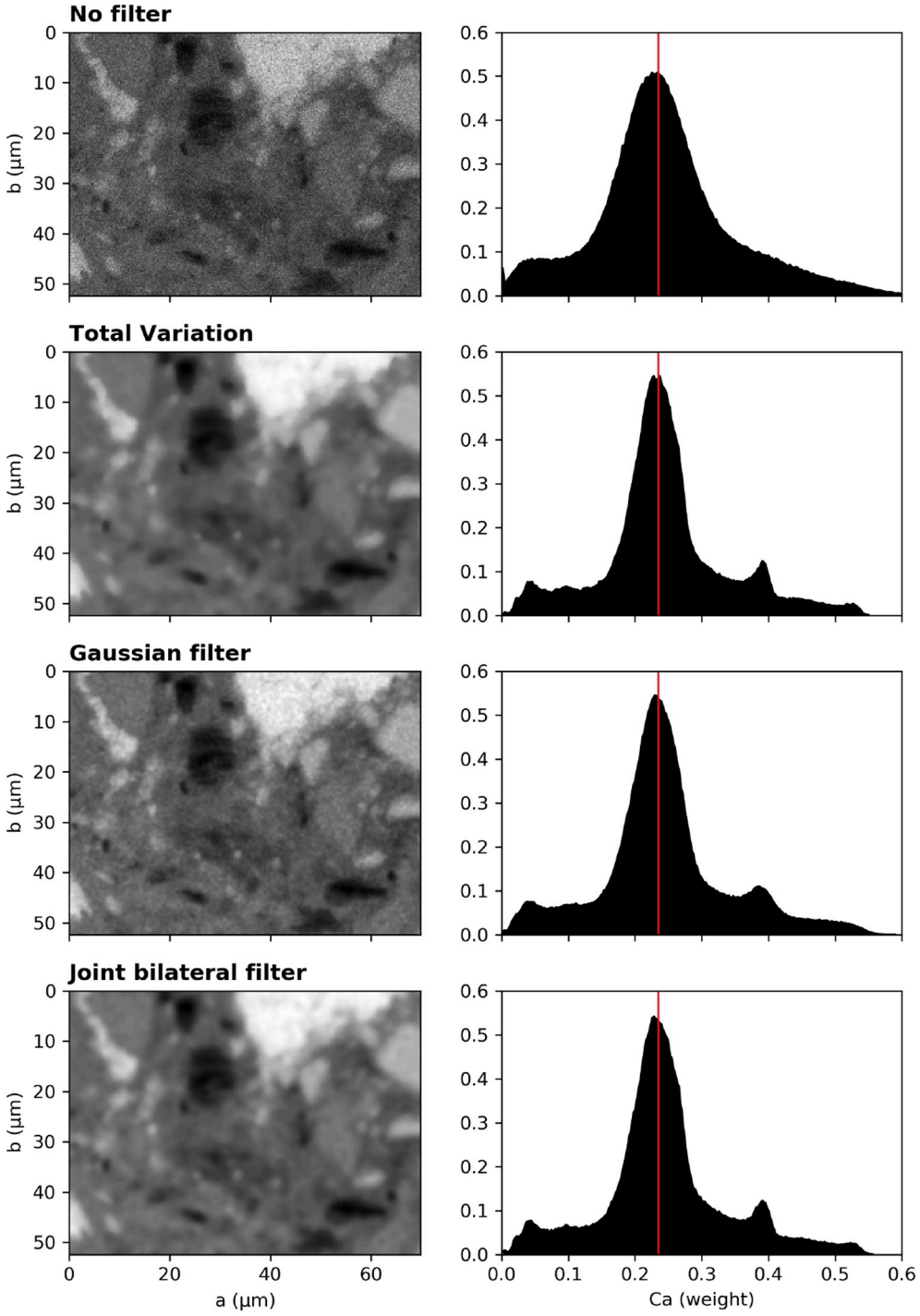

Fig. 3. Effect of the denoising algorithm on the Ca map. The left column presents the calcium maps, the right columns are the corresponding histograms. The red line was positioned at the maximum value of the histogram corresponding to the Ca map denoised with the total variation algorithm. Parameters for the denoising algorithms are described in the text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Segmentation of regions with similar compositions

To facilitate the identification of the phases, the composite map can be separated into regions using a superpixel segmentation algorithm [39]. This segmentation identifies adjacent pixels of similar color (i.e. similar composition for a composite map) and group them into continuous regions.

A segmentation algorithm could be applied to the gray BSE image in order to separate the regions without information about the composition. These regions would correspond to particles, grains, agglomerate of particles, or homogeneous sections of particles or gel. Standard algorithms can provide an adequate result for the anhydrous clinker grains. However, they usually fail for the hydrated phases as their characteristic sizes are smaller (sometimes even smaller than the resolution), and the gray levels are similar and dependent on the sample and the microscope. Therefore, it is preferable to carry out the segmentation on the composite map. The segmentation algorithm creates region of similar composition. The degree of similarity considered here is usually defined by the number of regions created by the algorithm. In most algorithms, the number of regions is a function of a user-defined

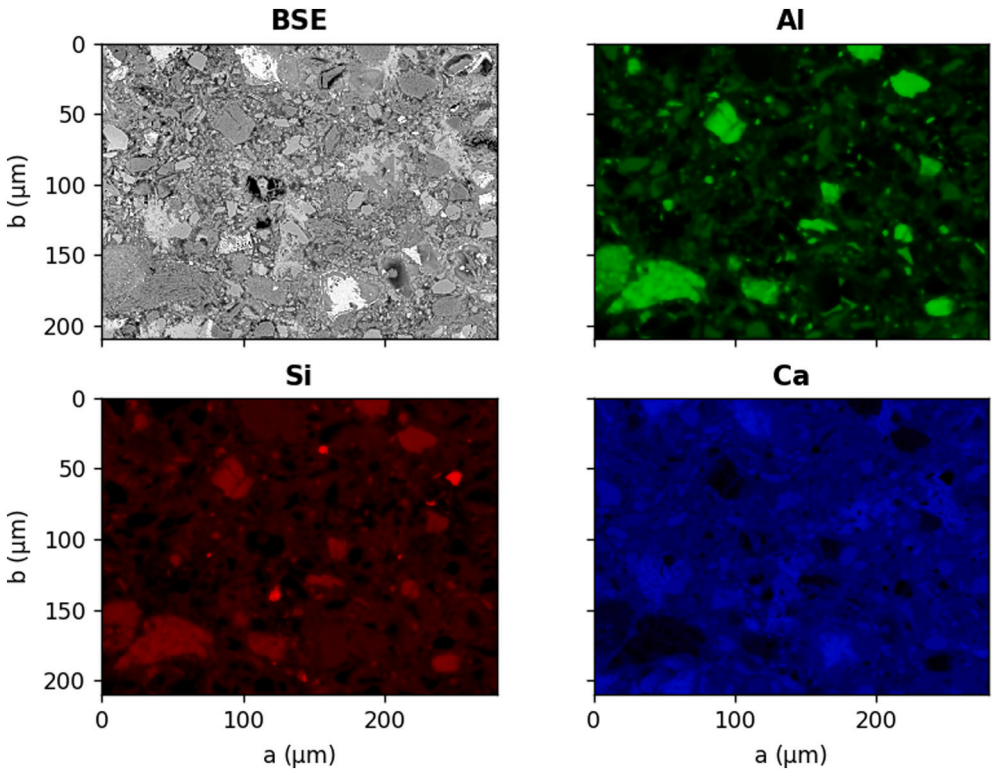

Fig. 4. Components of a composite image. These denoised maps are each assigned to the channels of an RGB image to create the composite image.

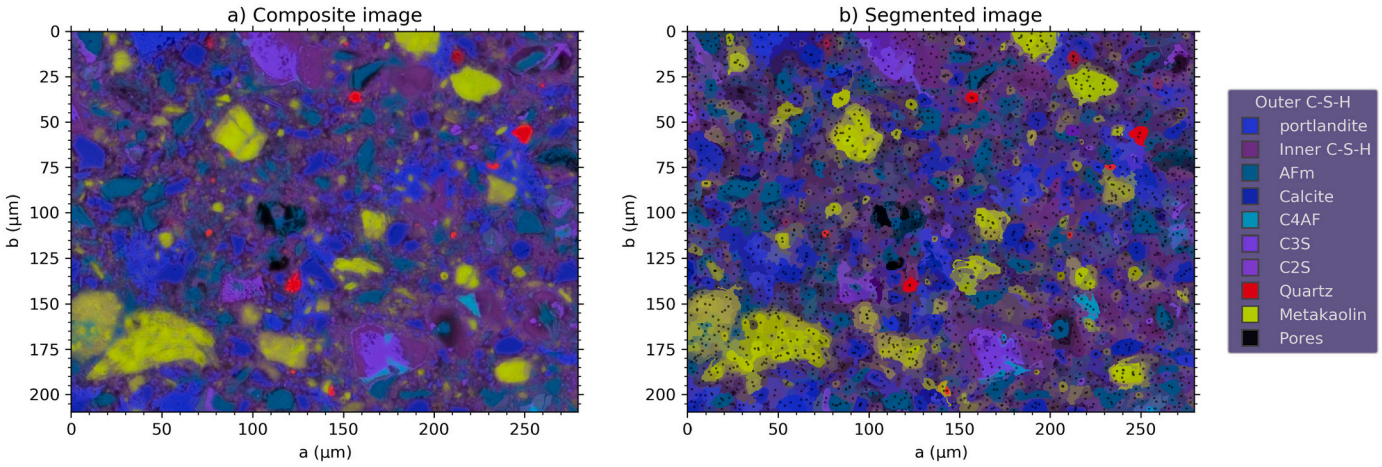

Fig. 5. a) Composite image formed from the combination of the individual channels presented in Fig. 4, and, b) corresponding segmentation using the SLIC algorithm. The dots represent the geometrical center of each segmented region.

parameter.

In edxia, the SLIC segmentation algorithm [39] is used by default. Its effect is shown in Fig. 5b. As observed in this figure, the main features of the composite map are kept after the segmentation. The main advantage of this method is that we can transform a large number of pixels (e.g., $1024 \times 768 = 786,432$) into a much smaller number of regions (e.g., $<20,000$). Although it is not mandatory, this step is important to simplify further analysis. It should be noted that the loss of information during this step is minimal, because neighboring points will have a similar composition (due to the overlapping interaction volumes, and the denoising). On the other hand, the signal-to-noise ratio is increased because many points at the interface between two phases, or in pores are not considered.

3.4. Emulation of point analysis

Chemical analysis with SEM-EDS is usually done with point measurements where the results are plotted as ratio plots [19,36]. For example, the composition of C-A-S-H is usually determined using an Al/Ca vs. Si/Ca scatter plot [1]. Our framework emulates that approach. The main challenge is to select representative points in the map using a reliable and reproducible method. For ideal systems made of particles in a gel matrix, the most representative points are at the center of the particles, because the effect of neighboring phases in the interaction volume are minimized. Unfortunately, as described in the previous section, it is not possible to identify perfectly each particle that can be detected visually in the backscatter micrograph. One of the main reasons is that superpixel algorithms constrain the superpixel sizes to be of

similar sizes, even if the particles sizes in the systems vary by a few orders of magnitude. However, we have identified regions of similar composition by segmentation. As an approximation, each of these regions is assimilated to a particle, a section of a particle, an agglomerate of particles, or a gel of relatively homogeneous composition. Therefore, the representative points are taken at the geometric center of these segmented regions. The reasoning is that for perfectly circular regions in a matrix of outer C-S-H, only the pure phases are sampled, and not the mixture of phases. Of course, real cementitious systems are not made of circular regions. However, this approach has been shown to provide an adequate filtering as long as the number of points is adjusted to the resolution and the field of view of the map. The filtering can be observed in the ratio plots presented in this article where most points are clustered between the pure phases (e.g. Fig. 6). An ideal situation would be to have one point per particle, or per agglomerate of particles. However, due to the large heterogeneities in cementitious materials, it is not possible. In practice, values between 5000 and 20,000 segmented regions were found to be a good compromise for standard maps.

A dataset made from the chemical composition at all of these points can then be created. These points can also be filtered based on their BSE values, for example to remove points corresponding to pores. The composition of these points can then be analysed using the standard ratio plots. An example of the points obtained with this method is presented in Fig. 6 for the OPC and the LC^3 samples. The challenge is now to understand how these points are related to the cement phases as identified by XRD or by thermodynamic modelling.

3.5. Characterisation of mixtures

Due to the interaction volume of the electron beam, the nanocrystalline phases and the many solid solutions typical of cement pastes [40], as well as the inherent low signal-to-noise ratio of EDS hypermaps, most pixels usually have to be considered as mixtures of phases rather than attributed to single pure phases. The main example corresponds to the C-S-H matrix, the glue of the cement paste which incorporates many small particles (unhydrated clinker grains, quartz, limestone, AFms, Ettringite, SCMs, etc.). Therefore, an automatic identification of each pixel to a particular phase is not possible without strong assumptions. Instead, the characterisation of these mixtures is a feature of this framework.

To achieve this goal, we adopt a manual approach using ratio plots. In a ratio plot, the points from the EDS analysis are plotted on axes of elemental ratios. The most common is the Al/Ca vs. Si/Ca plot, which can be used to identify the main phases of cementitious materials. An idealized version of this plot is displayed in Fig. 7. The colored circles represent the common end points, representing "pure phases" at the microvolume level, and the gray points represent the main mixture lines. In practice, the gray points were obtained by using random samples from a triangular distribution centered along the pure phases in the direction of the binary mixtures. It should be noted that these mixture lines are generally not well defined and lines from different mixtures can overlap.

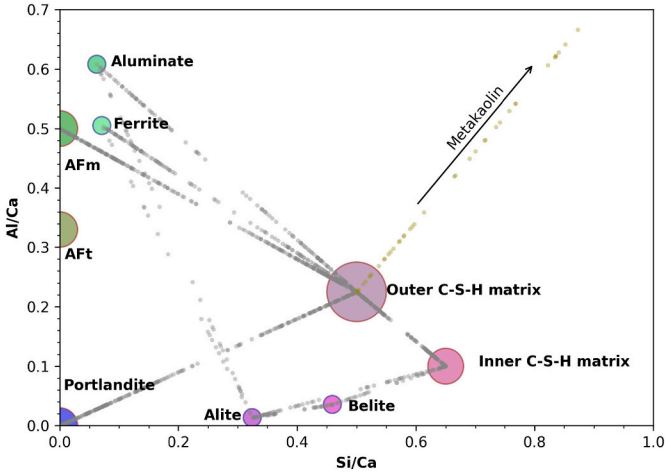

Fig. 7. Ideal phases and average anhydrous phases in the Al/Ca vs. Si/Ca. Examples of binary mixtures were created from triangular distributions around the position of the pure phases. Real mixtures points would include noise and ternary mixtures.

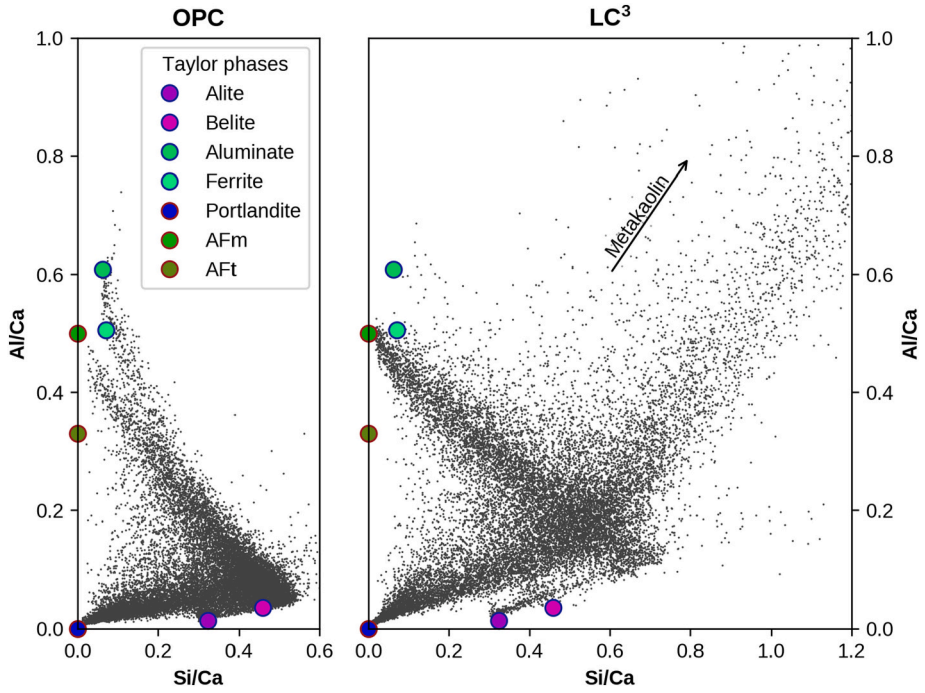

Fig. 6. Representative points extracted from the EDS hypermaps and plotted in the Si/Ca-Al/Ca ratio plot for the OPC and the LC^3 samples. Taylor phases correspond to the stoichiometries provided by Taylor [35], Table 1.3, p. 10.

For example, a mixture of C3A and C3S (in a polyphase clinker grain) can be found at the same position in this plot as a mixture of AFm and outer C-S-H. Therefore, it is important to develop combinations of filters by using several representations. For example, these two points can be separated by adding an additional condition on the BSE intensity, or on the sum of oxides to distinguish between anhydrous and hydrous phases. The crude representation of Fig. 7 is sufficient to understand Fig. 6. However additional effects, such as random noise, background and ternary mixtures must be considered to obtain the clouds of points of Fig. 6.

One of the main consequences of the intermixing is that pure phases are hard to isolate. It is especially true for the C-S-H which is finely intermixed with other phases such as ettringite, hydrotalcite, or even fine particles of metakaolin. As such, the properties of pure C-S-H are hard to obtain through this type of analysis [1]. To make this point clearer in the rest of the paper, we call this fine mix of C-S-H and other phases the C-S-H matrix. The properties of this matrix are an average of the true C-S-H gel composition as well as a proportion of the other mixed phases at a scale lower than the interaction volume of the electron beam.

In Fig. 6 it can be observed that the density of points and the known

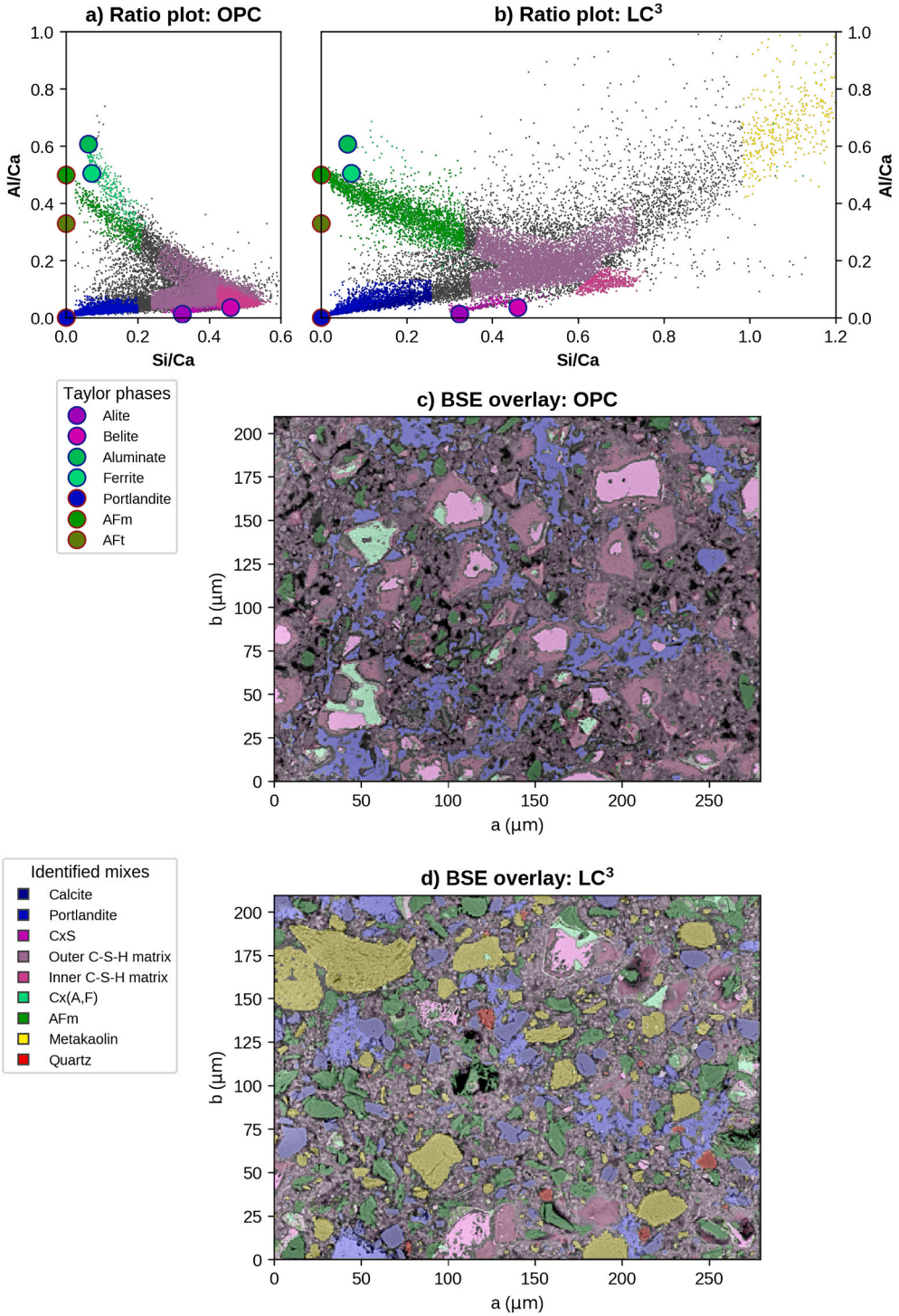

Fig. 8. Phase separation in 28 day cured samples of an OPC and a LC^3 samples. a) and b) are the Al/Ca vs. Si/Ca ratio plots for the OPC and LC^3 samples. c) and d) are the phase masks overlay on top of their respective BSE micrograph. Taylor phases correspond to the stoichiometries provided by Taylor [35], Table 1.3, p. 10.

position of the main cement phases is sufficient to understand the main mixture lines. By defining thresholds along these lines, it is possible to relate each point to a particular phase. The selection of points can be defined by a series of inequalities (e.g. Si/Ca < 0.2 and Al/Ca < 0.2 for portlandite), or graphically by using the Glue interface [20,31], to draw a surface around the region of interest (ROI). These inequalities are then used to identify the corresponding regions in the maps. Specific inequalities for OPC-based cementitious systems are described in the next section.

3.6. Qualitative phase separation

As explained in the previous section, it is possible to identify most of the cement paste phases visible at the BSE micrograph scale using a manual selection of points. This section applies this approach to the two cement pastes as examples. The separation of phases is presented in Fig. 8 for the OPC and the LC^3 samples, in the Si/Ca and the Al/Ca ratio plot as well as overlays of the phase masks on the BSE image.

The separation was made according the mixtures identified in Fig. 7 and guidelines detailed in Table 3. Table 3 summarizes the main inequalities used to separate phases based on the Si/Ca, Al/Ca, BSE, Sum of OXides (SOX) and others. The BSE histogram is roughly separated into two main peaks (clinker and hydrates). The actual values are dependent on the sample, the microscope and the microscope operator. This table does not provide strict guidelines, but it gives rules of thumb for thresholds to be adapted to each sample, similar to the decision tree presented by Bentz et al. [13].

The Al/Ca vs. Si/Ca plot is the most useful representation to separate phases, but it is not sufficient. As an example, the calcite and portlandite were separated using the BSE vs. SOX plot: both phases are very similar in backscattered contrast and chemical ratios, but they exhibit different absolute contents of calcium oxide (i.e., different SOX). Therefore, they can still be separated. This separation can be verified on the BSE image due to the morphology difference between calcite (rounded particles) and portlandite (elongated clusters).

Fig. 8c and d demonstrates the main microstructural differences between the OPC paste and the LC^3 paste. The inner C-S-H is a main hydrate phase in the OPC paste with the portlandite linked by the outer C-S-H matrix. The AFm grains are small and they have precipitated in spaces left by dissolved clinker grains. The LC^3 paste is more heterogeneous and compact at this scale. One visual difference is the dominance of the yellow and green hue compared to the blue and purple of the OPC paste. This is due to the metakaolin (Al, Si) present in large and small clusters as well as the AFm phase which is present in larger masses and, in higher amounts. These observations are consistent with our current understanding of the LC^3 microstructure.

Although not all the points can be attributed to the phases listed, the main grains and regions are identified, as observed on the BSE overlay in Fig. 8c and d. An advantage, but also a limitation of this approach is that not every pixel is identified to a phase. For example, the inner C-S-H close to aluminate clinker phase is not detected as inner C-S-H because its aluminate content is high due to the interaction volume of the electron beam. Instead, it is defined as the outer C-S-H matrix, which corresponds to a phase mixture. This effect is stronger in the LC^3 sample due to the intermixing with fine metakaolin grains. This can be observed in Fig. 8, where the inner C-S-H matrix identified for the LC^3 sample has higher Al/Ca ratio, and it is moved towards the metakaolin. The choice of the threshold is an advantage for chemical composition identification, but an inconvenience for volume fraction quantification. Thus, the user should choose between assigning all pixels to a phase, or separating phases with relatively pure compositions. Both extremes (very pure phases, or all pixels identified) can be appropriate according to the problem at hand. This choice is intrinsic to the physical limitations of the experiment. As such, is not unique to our method. It is present in all methods, even if it is not recognized.

Only qualitative comparisons have been made so far. The next sections will focus on extracting quantitative data and, on comparison with other measurement methods.

Table 3
Example of boundary definition for common phases in cement pastes. ε represents a tolerance dependent on the noise in the map, phases in the sample, and the interaction volume of the electron beam.

Phase	Si/Ca	Al/Ca	BSE peak	SOX	Others
C_3S	$0.33 \pm \varepsilon$	<0.1	Clinker	$1 \pm \varepsilon$	
C_2S	$0.5 \pm \varepsilon$	<0.1	Clinker	$1 \pm \varepsilon$	
C_3A/C_4AF	<0.2	>0.5	Clinker	$1 \pm \varepsilon$	
AFm	<0.2	$0.5 \pm \varepsilon$	Hydrate	<0.8	S/Ca,Cl/Ca, …
Portlandite	<0.2	<0.2	Hydrate	$0.76 \pm \varepsilon$	
Calcite	<0.2	<0.2	Hydrate	$0.48 \pm \varepsilon$	
Strätlingite	$0.5 \pm \varepsilon$	$0.5 < x < 1$	Hydrate	<0.8	
Metakaolin	$>1.0 \pm \varepsilon$	∝ Si/Ca	Hydrate	$1 \pm \varepsilon$	
Slags	Variable	Variable	Hydrate	$1 \pm \varepsilon$	Mg can help [5]
Slags hydrates	Variable	Variable	Hydrate	<0.8	Mg can help [5]
Quartz	>1.0	<0.2	Hydrate	$1 \pm \varepsilon$	Can filter on Si
Fly ash	Variable	Variable	Hydrate	$1 \pm \varepsilon$	
C-S-H matrix	Variable	Variable	Hydrate	<0.7	Depends on cement blend

4. Microstructure characterisation

4.1. Chemistry of phases

Once a phase has been identified, it is possible to study its composition. As an example, the composition of the AFm solid solution formed during chloride ingress was found to be a very good case study for the edxia approach. The AFm phases are part of the layered double hydroxide family. Their interlayers contain anions that can be exchanged with the pore solution. In particular, in contact with chloride, the sulfate and carbonate- AFms of the cement paste can transform to Friedel's salt [41,42]. The conversion to Friedel's salt is only partial at the commonly employed concentrations (e.g. 0.5 M NaCl) and it is important to measure the chloride stoichiometry in the solid solution to fully characterize the chloride binding.

A method based on SEM-EDS point analysis was recently developed in Sui et al. [43]. It is based on extrapolating the AFm/C-S-H mixture line to Al/Ca $= 0.5$ in the Cl/Ca vs. Al/Ca ratio plot, enabling the estimation of the Cl/Ca ratio of the solid solution between Friedel's salt and another AFm (Cl/Ca $= 0$, but Al/Ca $= 0.5$). Using the edxia approach instead of EDS point analysis for this task has two main advantages. First, it is useful to check that the selected points are indeed AFm based on morphology. Secondly, the mixture line is fully captured, and a higher precision can be obtained. Finally, the map further contains additional information, such as the chloride sorption on C-S-H. As an illustration, Fig. 9 shows the Cl/Ca vs. Al/Ca ratio plot for an LC^3 sample cured 28 days before and after being exposed to a 0.5 M NaCl solution. In this case, the Cl/Ca ratio is around 0.24 ± 0.01 in the AFm phases after exposure, which is a little less than half the maximum binding capacity compared to formation of pure Friedel's salt. Additional details on this study is available in [44,45], such as the estimation of the solid solution by XRD which compare very well to our method.

4.2. Sub-resolution phases

Another added value of the full chemical analysis is the possibility to detect sub-resolution phases. Ettringite is one of the main phases in cement paste. However, it is not identified in Fig. 8. The reason for this is that ettringite crystals are below the resolution of a typical hypermap, as ettringite commonly precipitates in needles of less than a hundred

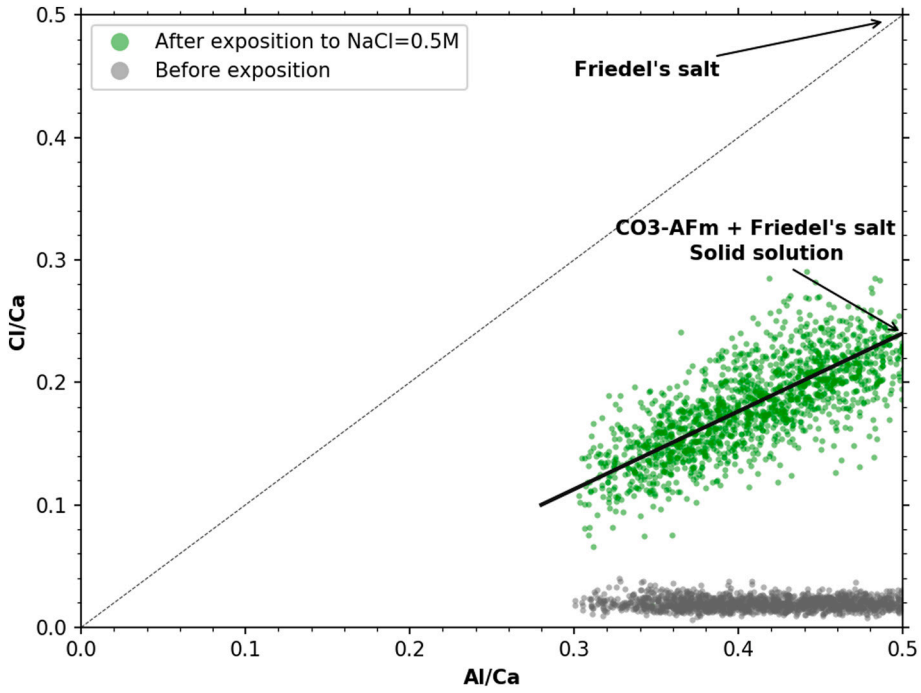

Fig. 9. Composition of the Friedel's salt-hemicarbonate solid solution as measured by SEM Image analysis. Example of a LC^3 sample before and after exposure to a 0.5 M NaCl solution.

nanometers compared to the micrometer-scale interaction volume, and about half a micrometer BSE resolution. Some clusters of ettringite could be identified on the BSE image by a trained expert, especially in the case of OPC. However, this does not tell us about the real distribution of ettringite. Sub-resolution refers here to the fact that no pure ettringite points are detected in the Al/Ca vs. Si/Ca ratio plot, as shown in Fig. 8.

However, the presence of ettringite can still be detected through the EDS chemical composition. The process is presented in Fig. 10. First, the outer C-S-H gel is selected as a reference in the Si/Ca vs. Al/Ca ratio plot. Then, in the S/Ca vs. Al/Ca, the Ettringite/C-S-H mixture line is

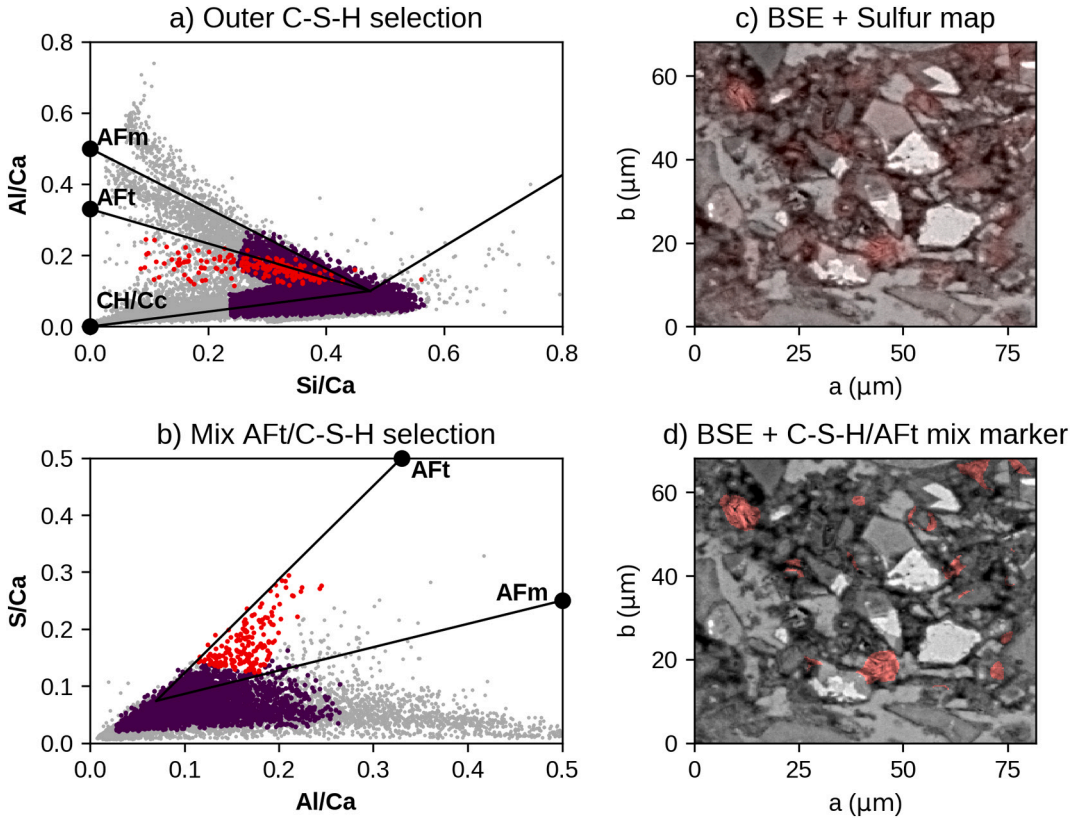

Fig. 10. Subresolution detection of ettringite in the OPC sample at 28 days.

identified. Points along that line, but outside the main C-S-H matrix cluster of points are selected. The selected pixels do not correspond to pure ettringite pixels, but to mixtures of C-S-H and ettringite. This process is of course highly dependent on the threshold chosen. However, it helps to detect the presence of sub-resolution phases. In the case of ettringite, this method can be used to study delayed ettringite formation or sulfate attack to detect the precipitation patterns. Although fractions cannot be computed accurately for these phases intimately intermixed, point densities can be used to compare samples qualitatively. This detection is of course dependent on the level of the intermixing. For example, in the OPC sample, small ettringite grains can be observed as depicted in Fig. 10. However, in the LC^3 samples, the regions detected by this method are not easily identified using only the BSE, as seen in Fig. 11. Detection of mixed phases is a common use of ratio plots. The added value of our framework is that their spatial distribution can also be studied.

4.3. Quantitative phase fractions

One of the simple microstructure descriptors that can be obtained from a phase mask is the volume fraction. The use of image analysis to quantify the volume fraction of anhydrous phases (and the reaction degree) is well documented in the literature as early as 1986 [3,5,46]. The volume fraction of hydrates is less studied. The main reason is the difficulty to separate the hydrates phases easily.

As an illustration of the potential of our framework, we propose a way to quantify the volume and mass fraction of AFm phases in a LC^3 cement paste, which are generally considered difficult to quantify. These phases are important as they fill the pore space in the microstructure, and thus contributes to the compressive strength. They are also important for durability-related issues as they are the main phases able to bind chlorides [41].

A main challenge is to obtain a representative scale to analyse. The measured area should be big enough to obtain a big enough average over the heterogeneous material. However the magnification should be big enough to be able to separate the hydrate phases of interest. The largest source of heterogeneities in the cement paste is the large anhydrous grains. As seen in Fig. 8, the presence or the absence of a clinker grains would make a large difference on the volume fractions. One very time-consuming way to solve this problem would be to acquire many maps at high magnification. A simple method is to choose a phase to normalise the results to a known quantity, and thus to remove the influence of the larger grains. Thus, our approach is the following: (1) first, we separate the big grains, and we compute the surface fractions in the remaining area; (2) and then, a phase present in that volume is used to renormalise the data to include the big grains. This approach requires another experiment to carry out the normalisation but it avoids having to take too many high resolution maps with a very high field-of-view.

The total volume fraction of small particles in the LC^3 system is defined as:

$$\phi^{small} \equiv 1 - \phi_{clinker} - \phi_{metakaolin} - \phi_{quartz} - \phi_{pores}$$

The partial volume fraction of phase X, with small characteristic size, can then be computed by:

$$\phi_X^{small} = \frac{\phi_X}{\phi^{small}}$$

The normalisation factor can be computed for an easily quantifiable mineral Y by XRD or by TGA:

$$\alpha_Y = \frac{\rho_Y \phi_Y^{small}}{w_Y}$$

where ρ_Y is the density of solid phase Y, and w_Y the mass fraction of this phase as measured by another method.

The mass fraction of phase X is then given by:

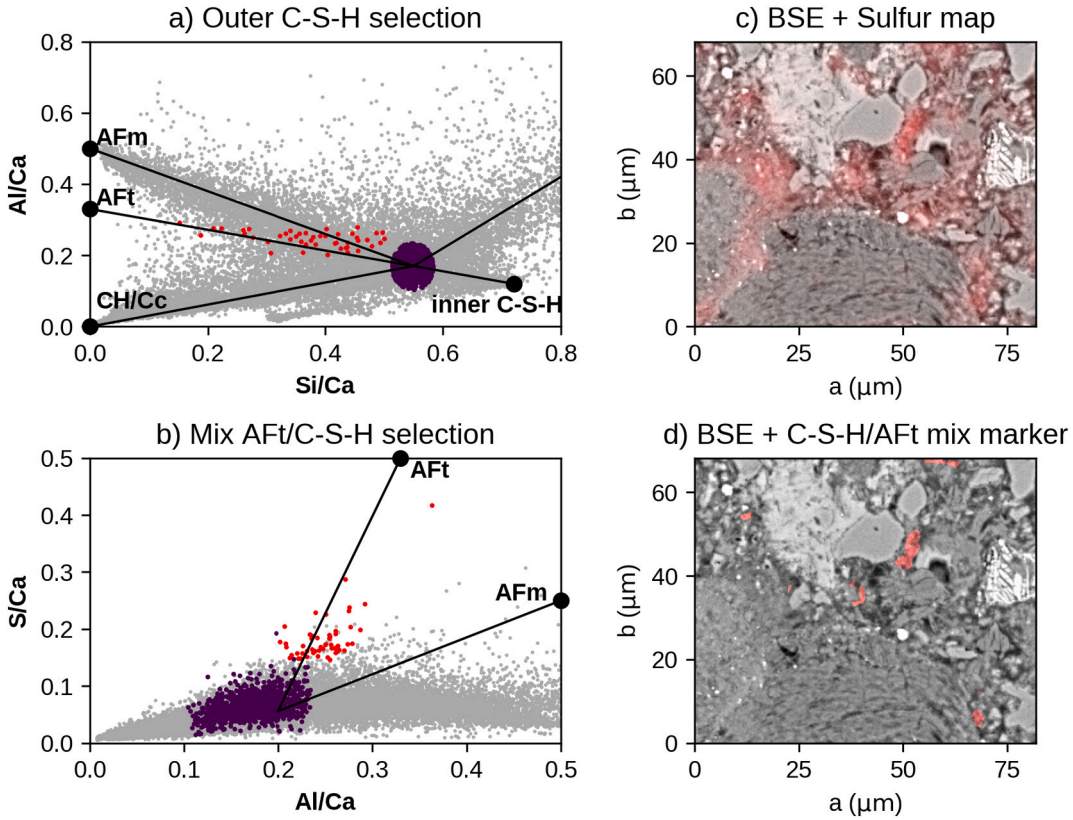

Fig. 11. Subresolution detection of ettringite in the LC^3 sample at 28 days.

$$\overline{w}_X = \frac{\rho_X \phi_x^{small}}{\alpha_Y}$$

In addition to the error of the estimation of the phase fractions from the hypermaps, this method is also sensitive to the error made during the XRD analysis and the estimation of the density. In this example, we choose calcite as the normalisation phase, as it is present in large quantity and it can be measured by XRD or by TGA. Its density is also well known. In order to obtain representative estimates of the volume fractions, the analysis was done over fields of view larger than the feature of interest (here ~300 μm by 200 μm) and with a sufficient number of replicates to cover the local heterogeneity of the material (here 13 hypermaps, set of parameters F in Table 2). Mass densities were obtained from Balonis and Glasser [47].

The mass fractions estimated by this method are presented in Table 4. The mass fractions obtained for calcite and portlandite by XRD and TGA agree very well. The mass fraction of portlandite found by image analysis after normalisation is also very close. Although portlandite and calcite precipitate very close to each other, and their boundaries are not well defined (see Fig. 8), it indicates that our phase separation using the sum of oxide works well.

Due to overlapping peaks, the AFm content cannot be measured accurately from TGA, and only the XRD values can be used as comparison. The mass fractions obtained by image analysis are higher than the XRD values. There are two possible explanations for this. Either the XRD underestimate the amount of AFm phases. The AFm peaks are difficult to fit accurately, and the choices for the Rietveld analysis are not obvious. In addition, some of the AFms might be poorly crystalline. On the other hand, the image analysis can also overestimate the AFm content. First, the AFm precipitates in large pores in what used to be clinker grains. Therefore, they are surrounded by porosity which can be included in part in our analysis. Another source of error is the internal porosity of these grains. Although these grains are smooth in the OPC case, they have a texture in the LC^3 case. The densities used are the solid densities from the crystal structure. The density that should be used is the apparent density, taking into account the internal porosity (due to cracked or non compact grains) and the mixed phases. However, this value is not available. The impact of this density is important as highlighted by the difference between assuming hemicarbonate or monocarbonate as the AFm phase.

As such, this quantification by image analysis is interesting for several reasons: (1) it provides the apparent volume fraction directly, and not a mass fraction to be transformed into volume fraction (the volume fraction is of interest for validating homogenization models to predict the compressive strength); (2) it provides an additional measurement to verify/confirm other methods; and (3) it enables estimation of volume fraction for amorphous phases that cannot be directly measured by XRD-Rietveld, or by TGA due to peak overlapping. If the approach is exemplified here for AFm phases, a similar protocol considering the scale of particles could be adapted for other hydrates and anhydrous phases.

Table 4
Mass and volume fraction obtained from 13 small hypermaps and a normalisation factor computed by XRD or by TGA. Uncertainties for the XRD correspond to the standard deviation on 5 scans using the same Rietveld fitting process.

Solid phase (mass%)	w_{XRD} (mass%)	w_{TGA} (mass%)	$\overline{w}^{IA}_{XRD}$ (mass%)	$\overline{w}^{IA}_{TGA}$ (vol%)	ϕ_X^{small}
Calcite	9.7 ± 0.2	9.9	9.7	9.9	10.9
Portlandite	3.2 ± 0.3	3.3	3.1	3.2	4.2
AFm (as hemicarbo)	9.8 ± 0.5		12.4	12.7	19.1
AFm (as monocarbonate)	9.8 ± 0.5		13.6	13.9	19.1

4.4. Microstructure descriptors

Describing the microstructure of random heterogeneous porous materials is a challenging problem as only a statistical approach can be representative [48]. Our phase separation analysis provides binary masks representing each phase. From these masks, it is possible to compute statistical representations of the microstructure. For example, the point-point correlation function, S_2, [11,48], or the chord-length distribution function [48] can be easily computed in our framework. These functions provide information about the volume fraction, the particle size distribution, the particle shapes and their distribution in the material. This feature is directly available in the graphical interface.

As an example, Fig. 12 presents a comparison of the S2 correlation functions for the portlandite and the AFm phases in the OPC and the LC_3 samples. $S_2(r)$ is the probability of finding a phase i at a distance r from a random location in the sample. From binary phase masks, it is straightforward to compute as described in [48]. By construction, $S_2(0)$ corresponds to the volume fraction (ϕ), and $S_2(\infty) = S_2(0)^2 = \phi^2$. Therefore, to better identify the length scales, Fig. 12 shows $S_2(r) - S_2(0)$ as a function of r. In this view, $S_2(r) - S_2(0) = 0$ corresponds to an absence of spatial correlation, where the probability of finding the phase is only determined by the volume fractions. From these plots, we can see that portlandite have similar characteristic lengths in both systems. However, it's volume fraction is lower in the LC^3 sample due to the pozzolanic reaction. On the other hand, the volume fraction of AFm in the LC^3 system is higher due to the synergetic reaction with limestone [22]. The grains of AFm are also significantly larger in the LC^3 system.

These microstructure descriptors (volume fraction, correlation functions, …) can be used to study the effect of the raw materials processing or curing conditions of the assemblage of phases at different ages to optimize the mix design for compressive strength and/or durability [49].

5. Going further

5.1. Profiles

The field of view of a typical map is usually less than a couple of hundred micrometers. The typical scale of penetration profiles, relevant for are a few centimeters. Therefore, maps taken along the profile of a sample can be considered homogeneous (with respect to the durability issues). As a consequence, all the different analysis shown in the previous section can be done at different depths in a sample. It can be used to study the changes occurring during exposure to environment. For examples, we can obtain the chemical composition of the AFm for a cement paste sample subjected to an electro-migration test [50]. Specific details of the test run are available in [45]. The result of the analysis for an OPC sample of 1 cm is presented in Fig. 13. Additional methods (e.g., microXRF) can be used to obtain the total chloride concentration. Therefore, both the total amount, and the composition of the phases can be obtained as a function of specimen depth. In addition, the method used to obtain the quantitative volume fractions could be used in such a profile to get both the composition and the amount of a specific phase. This combination of information is necessary to correctly validate reactive transport models without overfitting parameters.

5.2. Using lower-efficiency EDS detectors

The maps presented so far were obtained with a high-end EDS detector. In practice, detectors with a lower efficiency might be the only hardware available. By lower efficiency, we mean a detector with lower detection surface, lower sensitivity, higher dead time or any other factor that requires a higher exposition time to obtain the same number of counts. In this section, we discuss the influence of the detector to understand to what extent information can be obtained from lower quality maps. Fig. 15 compares different signals obtained on the same region of

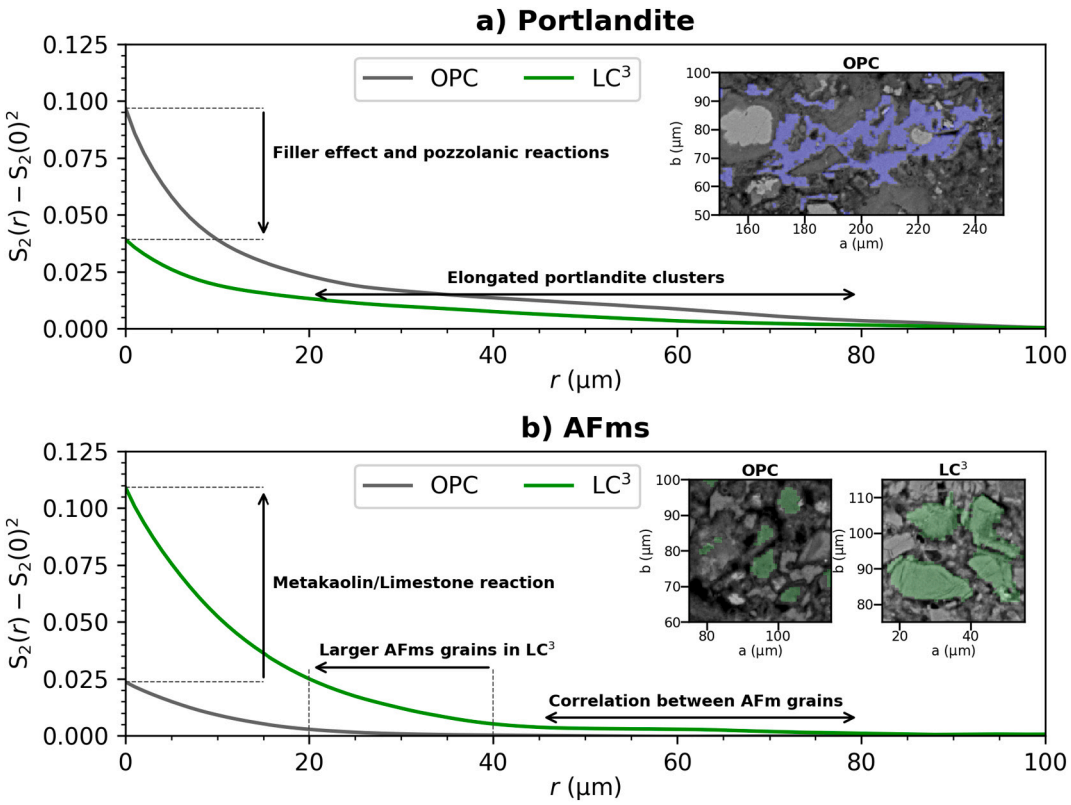

Fig. 12. S_2 correlation function for (a) the portlandite and (b) the AFm phase in a OPC and a LC_3 samples.

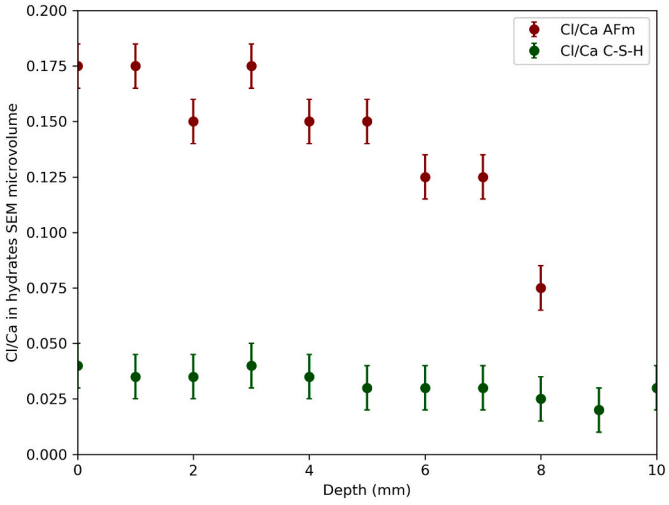

Fig. 13. Chemical composition of the AFm in a OPC sample subjected to an electromigration test.

interest using two detectors and different instrument parameters as described in Table 2 and Section 2.2.2. The set of maps studied until now is the set C. Sets C, D and E have been obtained with the same high-end EDS detector, while sets A and B have been obtained with a detector with a lower surface of detection. Table 5 shows the peak signal-to-noise (PSNR) for these maps. The PSNR of a $M \times N$ EDS map is estimated from the map I, and the denoised map D. The PSNR is defined as:

$$\mathrm{PSNR} = 10 \cdot log_{10}\left(\frac{1}{\mathrm{MSE}}\right)$$

And the mean squared error (MSE) is defined as

Table 5
Peak signal-to-noise ratio (dB) for various maps of the LC^3 sample according the parameters.

Map	Ca	Al	S
A	22.9	33.0	40.2
B	19.5	28.8	36.2
C	29.6	35.1	41.4
D	26.4	32.8	39.1
E	28.8	34.9	42.6

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\{I(i,j) - D(i,j)\}^2$$

The effect of the parameters is shown in Figs. 14 and 15. A main feature is that the absolute values (%Ca, SOX) are a lot more sensitive than the ratios (Ca/Si) to detector quality. While the absolute values present a clear bias after filtering, the Ca/Si ratio is similar between map A and map C. However, the ratio presents a slight bias for the noisiest set of hypermaps, set B. This is the main a posteriori justification for the selection of phases based on ratios in our general approach: even if the maps are quite noisy, the ratios remain a reliable measure [36]. Nevertheless, the absolute values provide valuable information for microstructure characterisation. For example, the sum of oxide is the most reliable method to separate portlandite from limestone. Selecting low calcium phases (e.g., quartz) on absolute chemical values is also more reliable than selecting it on chemical ratios (e.g., Si/Ca) due to the division by small values.

5.3. *Filtering at the spectrum level*

In this section, we present a method to improve the signal-to-noise ratio. The main source of noise is the low electron counts during the map quantification. The filtering process reduces the error, but it does

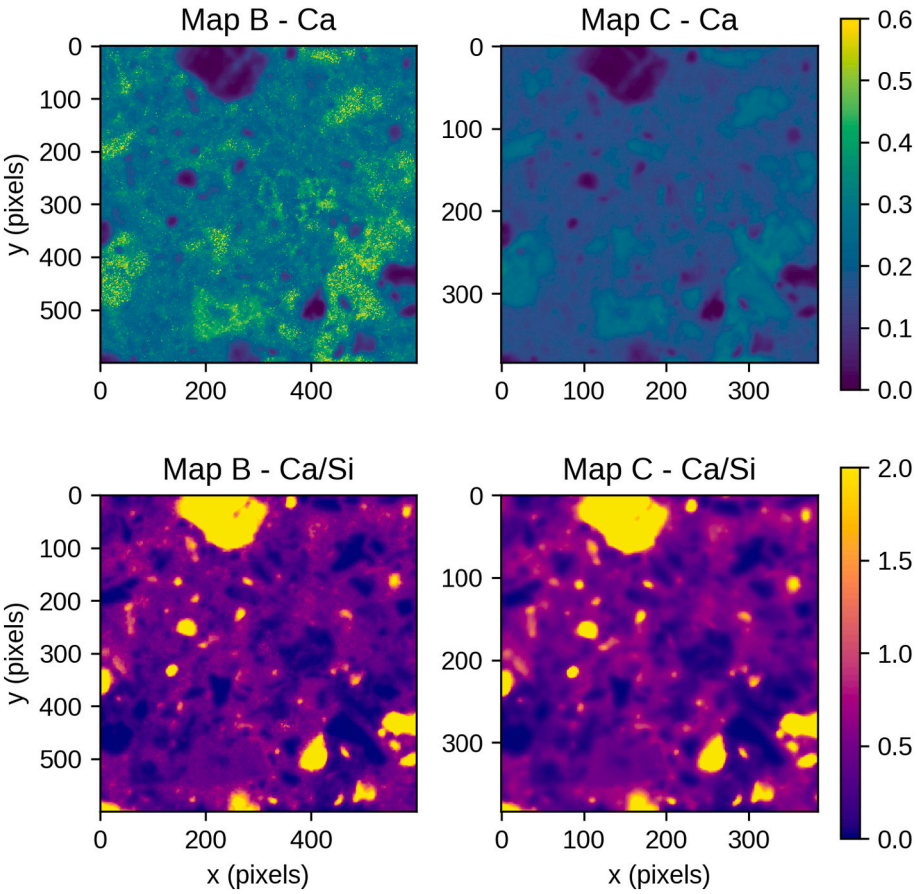

Fig. 14. Comparisons of the Ca, and the Ca/Si maps (after denoising) for the set B (high noise) and the set C (low noise) of hypermaps.

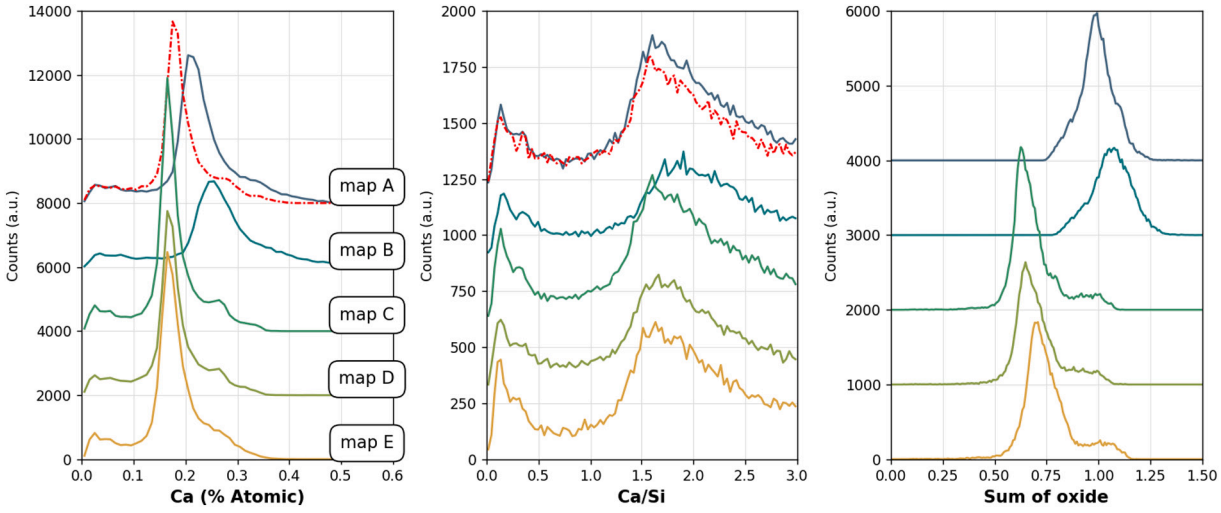

Fig. 15. Comparison of EDS hypermaps of the same region in the same sample with various signal-to-noise ratio. All curves were obtained from the filtered maps. The red dashed curve is obtained from filtered spectrums (see Section 5.3). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

not address the root cause of the problem. A better method would be to filter the hyperspectral image directly. This is unfortunately often restricted by the proprietary format used by the microscope manufacturer, and the difficulty of implementing the quantification algorithm [51]. However, to demonstrate the potential of this method, we used a simpler yet effective approach. For the set A of hypermaps, the coordinates of representative points were first obtained. The spectrum of the pixels of a 5 × 5 square centered at the representative points was then summed. Then, the reconstructed spectrums were quantified using the same quantification algorithm as used for the hypermaps. The results are shown as the red dashed curve in Fig. 15. It can be observed that the Ca/Si ratio is similar, but the bias on the Ca amount has been greatly corrected. The main limitation of such an approach here was the time needed to quantify the spectrums as the analysis could not be fully

automated. Therefore, only one example is presented. The reconstructed spectrums where created with our framework and hyperspy [52] to extract the data from the raw hypermap. With an automated analysis, it would be possible to only average spectrums in the regions found by the segmentation of the composite image, providing a means to obtain high-quality quantitative data from lower-end detectors.

5.4. Possible extension to machine learning

To simplify the analysis, it is natural to think about using one of the many machine learning algorithms developed to solve the classification problem. Our method to extract the representative points can be used as a pre-processing step for these algorithms. A manual classification approach was preferred for the reasons highlighted in the introduction.

As outlined in the article, the challenge is even greater if we realize that mixtures with three or more phases can exist, and especially, mixtures with phases that cannot be observed directly at the scale of the SEM interaction volume (e.g. ettringite). In addition, the main phase of cement paste, C-(A-)S-H, is very sensitive to the formulation. It is not possible to define a boundary that will work for every formulation of cement paste. As a consequence, it means that a black-box classification model would have to be re-calibrated for every formulation, curing conditions, and even, age of the sample. However, even if the chemistry of cement paste is intrinsically complicated, it is now possible to get reasonable predictions. Therefore, by leveraging domain knowledge, it is possible to obtain an a-priori distribution of phases. As emphasized by Munch et al. [16], this domain knowledge is important to obtain good results with machine learning approaches. Rather than adapting machine-learning to include domain knowledge, we chose to by-pass it to only use domain knowledge. The results presented in this paper should highlight the practicality of our choice.

In the opinion of the authors, the next step to improve machine learning methods would be to explicitly handle mixtures and to use the morphology information of the BSE image. For example, particle detection might help define clearer phase boundaries. It is relatively easy for large and clear particles such as the clinker grains, but it is a much more complex challenge for the small hydrate particles. Improving the phase separation is a worthy research topic. As an open-source Python framework, our tool is a good start as it can be easily extended with well-established machine learning framework such as scikit-learn [53].

6. Concluding remarks

A new framework to analyse quantified SEM-EDS hypermaps of cementitious materials is presented in this paper. Using superpixel subsampling of a composite map, representative points are extracted from quantified chemical maps. Ratio plots representations are then used to manually classify the points and create phase masks. The manual approach can be used to adapt the classification to the scientific problem at hand. Pure chemical composition, mixture identification or quantitative volume fractions can all be obtained. The framework is flexible and user-friendly, as it is implemented as a plugin to Glue, a multi-dimensional, linked data exploration graphical interface [20,21]. This interface can be used by the user to select phases in the ratio plots and to observe directly the corresponding phase masks on the BSE image. Quantitative analysis tools are also implemented in the interface. Furthermore, additional analysis can be easily added to our open source Python framework. In particular, other approaches could be added to refine the phase separation.

Several demonstrations of the abilities of our framework are presented. In particular, we analysed the AFm phases. We obtain their composition before and after exposure to chloride. We also measured their volume, and mass fractions, as well as their point to point correlation functions. This analysis can also be done along the depth of a sample to obtain profiles of phase content and/or phase chemical composition. This information is necessary to correctly validate reactive transport models of chloride ingress, as it allows fixing the chemistry model. We also presented how to obtain information of the phases below the resolution of the EDS analysis. Finally, the effect of noisy maps was discussed. Due to our ratio plots approach, good phase separation can still be obtained even if absolute values are not reliable. An additional method to refine the signal-to-noise ratio is proposed.

Future work will focus on using the morphological information to improve further the signal-to-noise ratio as well as extending the quantitative analyses of the phase maps with additional application-specific analysis tools.

CRediT authorship contribution statement

Fabien Georget: Conceptualization, Methodology, Software, Writing - original draft. **William Wilson**: Conceptualization, Methodology, Investigation, Writing - review & editing. **Karen L. Scrivener**: Writing - review & editing, Supervision, Funding acquisition.

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.

Acknowledgments

The authors would like to thank the members of the Laboratory of Construction Materials who provided raw data and feedbacks during the development phase of this project.

Appendix A

A.1. Code download and installation

The code is available in the python package index (https://pypi.org/project/edxia/), the Anaconda distribution (https://anaconda.org/specmicp/edxia), zenodo [30], or openly available git repository (https://bitbucket.org/specmicp/edxia/src/master/). Installation instructions and documentations is available with the code, online (https://edxia.readthedocs.io/en/latest/) or in [31] for the user friendly interface. The main datasets (set C in Table 2) analysed in this study are available [54] to test the interface.

References

[1] J. Rossen, K. Scrivener, Optimization of sem-eds to determine the c–a–s–h composition in matured cement paste samples, Mater. Charact. 123 (2017) 294–306, https://doi.org/10.1016/j.matchar.2016.11.041.
[2] S. Meulenzyer, J. Chanussot, J.J. Chen, S. Crombez, Spectral-spatial image processing strategies for classifying multispectral sem-eds x-ray maps of cementitious materials, in: 14th Euroseminar on Microscopy Applied to Building Materials, 2013 (EMABM), 2013.
[3] P.T. Durdziński, C.F. Dunant, M.B. Haha, K.L. Scrivener, A new quantification method based on sem-eds to assess fly ash composition and study the reaction of its individual components in hydrating cement paste, Cem. Concr. Res. 73 (2015) 111–122, https://doi.org/10.1016/j.cemconres.2015.02.008.
[4] X. Feng, E. Garboczi, D. Bentz, P. Stutzman, T. Mason, Estimation of the degree of hydration of blended cement pastes by a scanning electron microscope point-counting procedure, Cem. Concr. Res. 34 (10) (2004) 1787–1793, https://doi.org/10.1016/j.cemconres.2004.01.014.
[5] V. Kocaba, E. Gallucci, K.L. Scrivener, Methods for determination of degree of reaction of slag in blended cement pastes, Cem. Concr. Res. 42 (3) (2012) 511–525, https://doi.org/10.1016/j.cemconres.2011.11.010.
[6] K.J. Krakowiak, W. Wilson, S. James, S. Musso, F.-J. Ulm, 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 (2015) 271–285, https://doi.org/10.1016/j.cemconres.2014.09.001.
[7] W. Wilson, L. Sorelli, A. Tagnit-Hamou, 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 (2018) 49–65, https://doi.org/10.1016/j.cemconres.2017.08.016.

[8] D. P. Bentz, P. E. Stutzman, SEM Analysis and Computer Modelling of Hydration of Portland Cement Particles Petrography of Cementitious Materials, ASTM Int. doi:https://doi.org/10.1520/STP12645S.

[9] U.H. Jakobsen, K.D. Weerdt, M.R. Geiker, Elemental zonation in marine concrete, Cem. Concr. Res. 85 (2016) 12–27, https://doi.org/10.1016/j.cemconres.2016.02.006.

[10] F. Georget, W. Soja, K.L. Scrivener, Characteristic lengths of the carbonation front in naturally carbonated cement pastes: implications for reactive transport models, Cem. Concr. Res. 134 (2020) 106080, https://doi.org/10.1016/j.cemconres.2020.106080.

[11] P. E. Stutzman, J. W. Bullard, P. Feng, Quantitative imaging of clinker and cement microstructure, Tech. Rep. 1877 (April 2015). doi:https://doi.org/10.6028/NIST.TN.1877.

[12] S. Meulenzyer, B. Eric, L. Didier, R. Barbarulo, Sem supervised image analysis contribution on understanding sulfate attack on blended cement mortars, in: 15th Euroseminar on Microscopy Applied to Building Materials, 2015 (EMABM), 2015.

[13] D. P. Bentz, P. E. Stutzman, C. J. Haecker, S. Remond, Sem/x-ray imaging of cement-based materials, Proceedings of the 7th Euroseminar on Microscopy Applied to building Materials (1999) 457–466.

[14] Q. Ding, M. Colpan, Decision tree induction on hyper-spectral cement images, Int. J. Comput. Intell. 2 (3) (2006).

[15] L. Drumetz, M.D. Mura, S. Meulenyzer, S. Lombard, J. Chanussot, Semiautomatic classification of cementitious materials using scanning electron microscope images, J. Electron. Imaging 24 (6) (2015) 1–19, https://doi.org/10.1117/1.JEI.24.6.061109.

[16] B. Münch, L. Martin, A. Leeman, Segmentation of elemental eds maps by means of multiple clustering combined with phase identification, J. Microsc. 260 (3) (2015) 411–426, https://doi.org/10.1111/jmi.12309.

[17] W. Rasband, Imagej, https://imagej.nih.gov/ij/ (1997–2018).

[18] A. Schneider Caroline, S. Rasband Wayne, W. Eliceiri Kevin, Nih image to imagej: 25 years of image analysis, Nat. Methods 9 (7) (2012) 671–675, https://doi.org/10.1038/nmeth.2089.

[19] K. Scrivener, R. Snellings, B. Lothenbach, A Practical Guide to Microstructural Analysis of Cementitious Materials, Taylor & Francis, 2016.

[20] C. Beaumont, A. Goodman, P. Greenfield, Hackable user interfaces in astronomy with glue, in: A. R. Taylor, E. Rosolowsky (Eds.), Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV), Vol. 495 of Astronomical Society of the Pacific Conference Series, 2015, p. 101.

[21] T. Robitaille, C. Beaumont, P. Qian, M. Borkin, A. Goodman, glueviz v0.13.1: multidimensional data exploration (Feb. 2017). doi:https://doi.org/10.5281/zenodo.1237692.

[22] M. Antoni, J. Rossen, F. Martirena, K. Scrivener, Cement substitution by a combination of metakaolin and limestone, Cem. Concr. Res. 42 (12) (2012) 1579–1589, https://doi.org/10.1016/j.cemconres.2012.09.006.

[23] D. Jansen, F. Goetz-Neunhoeffer, C. Stabler, J. Neubauer, A remastered external standard method applied to the quantification of early opc hydration, Cem. Concr. Res. 41 (6) (2011) 602–608, https://doi.org/10.1016/j.cemconres.2011.03.004.

[24] X. Li, R. Snellings, K.L. Scrivener, Quantification of amorphous siliceous fly ash in hydrated blended cement pastes by x-ray powder diffraction, J. Appl. Crystallogr. 52 (6) (2019) 1358–1370, https://doi.org/10.1107/S1600576719013955.

[25] T.E. Oliphant, Python for scientific computing, Comput. Sci. Eng. 9 (3) (2007) 10–20, https://doi.org/10.1109/MCSE.2007.58.

[26] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: open source scientific tools for Python, [online; accessed 11-06-2019] (2001–). URL http://www.scipy.org/.

[27] Virtanen Pauli, Gommers Ralf, Oliphant Travis E., Haberland Matt, Reddy Tyler, Cournapeau David, Burovski Evgeni, Peterson Pearu, Weckesser Warren, Bright Jonathan, W. S. van der J., Brett Matthew, Wilson Joshua, Millman K. Jarrod, Mayorov Nikolay, Nelson Andrew R. J., Jones Eric, Kern Robert, Larson Eric, Carey C. J., Polat İlhan, Feng Yu, Moore Eric W., VanderPlas Jake, Laxalde Denis, Perktold Josef, Cimrman Robert, Henriksen Ian, Quintero E. A., Harris Charles R., Archibald Anne M., Ribeiro Antônio H., Pedregosa Fabian, M. van Paul, Vijaykumar Aditya, Bardelli Alessandro Pietro, Rothberg Alex, Hilboll Andreas, Kloeckner Andreas, Scopatz Anthony, Lee Antony, Rokem Ariel, Woods C. Nathan, Fulton Chad, Masson Charles, Häggström Christian, Fitzgerald Clark, Nicholson David A., Hagen David R., Pasechnik Dmitrii V., Olivetti Emanuele, Martin Eric, Wieser Eric, Silva Fabrice, Lenders Felix, Wilhelm Florian, Young G., Price Gavin A., Ingold Gert-Ludwig, Allen Gregory E., Lee Gregory R., Audren Hervé, Probst Irvin, Dietrich Jörg P., Silterra Jacob, Webber James T., Slavič Janko, Nothman Joel, Buchner Johannes, Kulick Johannes, Schönberger Johannes L., de Vinicius, Miranda Cardoso José, Reimer Joscha, Harrington Joseph, Rodriguez Juan Luis Cano, Nunez-Iglesias Juan, Kuczynski Justin, Tritz Kevin, Thoma Martin, Newville Matthew, Kümmerer Matthias, Bolingbroke Maximilian, Tartre Michael, Pak Mikhail, Smith Nathaniel J., Nowaczyk Nikolai, Shebanov Nikolay, Pavlyk Oleksandr, Brodtkorb Per A., Lee Perry, McGibbon Robert T., Feldbauer Roman, Lewis Sam, Tygier Sam, Sievert Scott, Vigna Sebastiano, Peterson Stefan, More Surhud, Pudlik Tadeusz, Oshima Takuya, Pingel Thomas J., Robitaille Thomas P., Spura Thomas, Jones Thouis R., Cera Tim, Leslie Tim, Zito Tiziano, Krauss Tom, Upadhyay Utkarsh, Halchenko Yaroslav O., Vázquez-Baeza Yoshiki, S. 1.0 Contributors, Scipy 1.0: fundamental algorithms for scientific computing in python, Nature Methods (2020). doi:https://doi.org/10.1038/s41592-019-0686-2.

[28] J.D. Hunter, Matplotlib: a 2d graphics environment, Comput. Sci. Eng. 9 (3) (2007) 90–95, https://doi.org/10.1109/MCSE.2007.55.

[29] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, the scikit-image contributors, scikit-image: image processing in Python, PeerJ 2 (2014) e453. doi:https://doi.org/10.7717/peerj.453.

[30] F. Georget, W. Wilson, edxia (Jan. 2020). doi:https://doi.org/10.5281/zenodo.3492220.

[31] F. Georget, W. Wilson, K. Scrivener, edxia through the glueviz interface (Feb. 2020). doi:https://doi.org/10.5281/zenodo.3686068.

[32] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imaging Vis. 20 (1) (2004) 89–97, https://doi.org/10.1023/B:JMIV.0000011325.36760.1e.

[33] E. Eisemann, F. Durand, Flash photography enhancement via intrinsic relighting, ACM Trans. Graph. 23 (3) (2004) 673–678, https://doi.org/10.1145/1015706.1015778.

[34] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, K. Toyama, Digital photography with flash and no-flash image pairs, ACM Trans. Graph. 23 (3) (2004) 664–672, https://doi.org/10.1145/1015706.1015777.

[35] H. Taylor, Cement Chemistry, Thomas Telford, 1997.

[36] Harrisson A. M., Winter N. B., Taylor H. F. W., X-ray microanalysis of microporous materials, Journal of Materials Science Letters 6 (11) (1987) 1339–1340. doi:https://doi.org/10.1007/BF01794611.

[37] W. Wilson, J. Rivera-Torres, L. Sorelli, A. Durán-Herrera, A. Tagnit-Hamou, The micromechanical signature of high-volume natural pozzolan concrete by combined statistical nanoindentation and SEM-EDS analyses, Cem. Concr. Res. 91 (2017) 1–12, https://doi.org/10.1016/j.cemconres.2016.10.004.

[38] D. Damidot, B. Lothenbach, D. Herfort, F. Glasser, Thermodynamics and cement science, Cement and Concrete Research 41 (7) (2011) 679–695, special Issue: 13th International Congress on the Chemistry of Cement. doi:https://doi.org/10.1016/j.cemconres.2011.03.018.

[39] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282, https://doi.org/10.1109/TPAMI.2012.120.

[40] B. Lothenbach, D.A. Kulik, T. Matschei, M. Balonis, L. Baquerizo, B. Dilnesa, G. D. Miron, R.J. Myers, Cemdata18: a chemical thermodynamic database for hydrated portland cements and alkali-activated materials, Cem. Concr. Res. 115 (2019) 472–506, https://doi.org/10.1016/j.cemconres.2018.04.018.

[41] M. Balonis, B. Lothenbach, G.L. Saout, F.P. Glasser, Impact of chloride on the mineralogy of hydrated portland cement systems, Cem. Concr. Res. 40 (7) (2010) 1009–1022, https://doi.org/10.1016/j.cemconres.2010.03.002.

[42] A. Mesbah, C. Cau-dit Coumes, F. Frizon, F. Leroux, J. Ravaux, G. Renaudin, A new investigation of the Cl-CO3 substitution in afm phases, J. Am. Ceram. Soc. 94 (6) (2011) 1901–1910, https://doi.org/10.1111/j.1551-2916.2010.04305.x.

[43] S. Sui, W. Wilson, F. Georget, H. Maraghechi, H. Kazemi-Kamyab, W. Sun, K. Scrivener, Quantification methods for chloride binding in portland cement and limestone systems, Cem. Concr. Res. 125 (2019) 105864, https://doi.org/10.1016/j.cemconres.2019.105864.

[44] Advanced Characterization of Chloride Binding in OPC and LC3 Pastes, TU Delft, 2020.

[45] W. Wilson, F. Georget, K. Scrivener, The mini-migration approach: a comprehensive characterization of accelerated chloride ingress at the scale of blended-cement pastes, (in preparation).

[46] K.L. Scrivener, H.H. Patel, P.L. Pratt, L.J. Parrott, Analysis of phases in cement paste using backscattered electron images, methanol adsorption and thermogravimetric analysis, MRS Proc. 85 (1986) 67, https://doi.org/10.1557/PROC-85-67.

[47] M. Balonis, F. Glasser, The density of cement phases, Cem. Concr. Res. 39 (9) (2009) 733–739, https://doi.org/10.1016/j.cemconres.2009.06.005.

[48] S. Torquato, Random Heterogeneous Materials: Microstructure and Macroscopic Properties, Springer, New York, 2002.

[49] F. Zunino, K. Scrivener, Assessing the effect of alkanolamine grinding aids in limestone calcined clay cements hydration, Constr. Build. Mater. (2020), 121293, https://doi.org/10.1016/j.conbuildmat.2020.121293.

[50] E. Samson, J. Marchand, K.A. Snyder, Calculation of ionic diffusion coefficients on the basis of migration test results, Mater. Struct. 36 (3) (2003) 156–165, https://doi.org/10.1007/BF02479554.

[51] G.F. Bastin, H.J.M. Heijligers, Quantitative electron probe microanalysis of ultralight elements (boron-oxygen), Scanning 12 (4) (1990) 225–236, https://doi.org/10.1002/sca.4950120408.

[52] F. de la Peña, E. Prestat, V. T. Fauske, P. Burdet, P. Jokubauskas, M. Nord, T. Ostasevicius, K. E. MacArthur, M. Sarahan, D. N. Johnstone, J. Taillon, J. Lähnemann, V. Migunov, A. Eljarrat, J. Caron, T. Aarholt, S. Mazzucco, M. Walls, T. Slater, F. Winkler, pquinn dls, B. Martineau, G. Donval, R. McLeod, E. R. Hoglund, I. Alxneit, D. Lundeby, T. Henninen, L. F. Zagonel, A. Garmannslund, hyperspy/hyperspy: Hyperspy v1.5.2 (Sep. 2019). doi:https://doi.org/10.5281/zenodo.3396791.

[53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[54] F. Georget, W. Wilson, K. Scrivener, edxia examples (May 2020). doi:https://doi.org/10.5281/zenodo.3862958.