Representative elementary area of shale at the mesoscopic scale

Philippe Cosenza, Dimitri Prêt, Anne-Laure Fauchille, Stephen Hedan

Representative Elementary Area of Shale at the Mesoscopic Scale

Philippe COSENZA[(1)*], Dimitri PRET[(1)], Anne-Laure FAUCHILLE[(2)] and Stephen HEDAN[(1)]

(1) University of Poitiers, CNRS, UMR 7285 IC2MP- HydrASA, ENSI Poitiers, France.

(2) Institut de Recherche en Génie Civil et Mécanique (GeM), Centrale de Nantes, Université de Nantes, UMR 6183 CNRS, France

***Corresponding author**

Philippe COSENZA

Ecole Nationale Supérieure d'Ingénieurs de Poitiers
Université de Poitiers-CNRS
1 rue Marcel Doré, Bat B1
TSA 41105
86073 Poitiers cedex 09
France
Email: philippe.cosenza@univ-poitiers.fr

Intended for publication in International Journal of Coal Geology

Abstract

The Representative Elementary Area (REA) of two shales (Callovo-Oxfordian claystone and Tournemire argillite) that are actively studied in the framework of the deep disposal of radioactive waste have been estimated from two mineral maps by classical methods (i.e., the box-counting and statistical approaches) and by different microstructural descriptors (i.e., the two-point probability function, lineal path function, percolation length, and variogram). The classical box-counting method provides estimates of the REA size of the clay fraction in the range from 129 μm to 441 μm, consistent with estimates obtained from the literature on other shales. However, these estimates show an extreme sensitivity to the chosen ε threshold or error and a wide scatter, thereby bringing the statistical homogeneity of both maps into question. Although the two-point probability function and lineal path function infer lower bounds of the REA size, these microstructural descriptors are relevant to demonstrate the microstructural anisotropy of both shales due to the alignment of nonclay grains parallel to bedding at the study scale. The results from the two-point probability function and variogram undoubtedly confirm that the Tournemire mineral map is not statistically homogeneous with regard to its mineral composition. This aspect it makes difficult to interpret the results and even questions the REA size determination of this particular map. Finally, our set of results allow us to recommend the use of the two-point probability function and variogram to preliminarily validate the statistical homogeneity of maps under study before calculating the REA size using conventional methods, e.g. the box-counting and statistical approaches.

Keywords: representative elementary area, microstructure, shale.

1. Introduction

Clay rocks, often called shales, are considered potential host rocks for high-level radioactive waste repositories in several industrialized countries (Callovo-Oxfordian (France), Opalinus Clay (Switzerland), Boom Clay (Belgium)). This interest is mainly explained by the following properties (e.g. Pusch, 2006): (1) because of their high-specific surface area, shales can absorb a significant amount of ions; (2) they have low-hydraulic conductivity values. These particular physicochemical properties are mainly controlled by the type and amount of clay minerals that are present in shale but also by its complex multiscale microstructure (Fig. 1). A major part of the literature agrees that the following microstructural levels must be considered to describe the shale texture or microstructure (e.g. Bennett et al., 1991; Ulm et al., 2005; Loucks et al., 2012; Chalmers et al., 2012; Curtis et al., 2012; Han et al., 2017; Ma et al., 2017) (Fig. 1):

- Level 0 is the scale of elementary clay layers.
- Level 1 is the scale for which the elementary clay layers are packed together to form clay particles.
- Level 2 is the submicrometer scale, often called the “microscopic” scale of porous clay matrix based on an assemblage of clay particles or aggregates. For organic rich shales, small patches of solid organic matter are also closely associated to the clay particles.
- Level 3, often called the “mesoscopic” scale in geosciences, is the scale where the characteristic size is in the submillimeter range. At this scale, the rock is considered to be a porous clay matrix mixed with a population of nonclayey grains (quartz, carbonates and pyrite). In organic rich shales, isolated and often porous organic patches are mixed with nonclayey grains. Both are distributed

in a clayey and organic matrix. This scale has been chosen to establish the mineral maps used in this paper.

- Level 4 is a lamina type that is associated with an alternation of clay-rich layers and other layers that are richer in non–clayey materials (mainly quartz and carbonates). This scale usually corresponds to the bulk samples and cores used for laboratory experiments for measuring physical properties.

In the following work, we will focus on the mesoscopic scale for which numerous petrographic studies using advanced imaging techniques allowed to obtain morphological, structural and topological information on shales (e.g., Robinet et al., 2012; Houben et al., 2014; Klaver et al., 2015; Keller, 2015; Fauchille, 2015; Ma et al., 2016; Fauchille et al., 2018). In order to capture all the microstructural features of interest, these imaging techniques produce high resolution images, usually obtained on limited volumes or areas. However, the small size of the field of view reached (typically in the range of a few dozen to a few hundreds of micrometers) may question the representativeness of the petrographic observations. In particular datasets obtained at this scale can be uncertain: are the structural characteristics and petrophysical/numerical properties determined on these small 3D volumes or 2D images representative of the characteristics and properties at the upper scale? In practice, the smallest representative volume or area is identified as the so-called Representative Elementary Volume (REV) (or Representative Elementary Area -REA in 2D), which is required to "separate" the two following space scales: firstly, the scale of heterogeneity, i.e., the distribution of nonclayey grains/clay matrix in our case, and secondly, the scale for which shale is viewed as an "equivalent" continuum medium, i.e., an effective medium. Thus, REV is usually considered a volume of the heterogeneous material that is sufficiently large to be statistically representative of the rock, i.e., to include a relevant sampling of all structural heterogeneities present in the rock at the scale of interest (here the

mesoscopic scale). Simultaneously, the REV has to be sufficiently small compared with the scale of the macroscopic geological system of interest, "so that it may be considered as infinitesimal in the mathematical treatment" (Biot, 1941). Another mathematical requirement for the definition of REV is the statistical homogeneity of the volume or the image under study: the REV must be independent of the point of calculation (localization of the volume in 3D or 2D space) (e.g., Rozenbaum and Rolland, 2014).

A number of approaches have been considered to estimate the size of REV and REA, hereafter named L_{REV} and L_{REA}, respectively. The "box-counting" method is likely the most popular and has recently been applied to 2D images or 3D volumes acquired in shales that have been actively studied in recent years as potential gas and reservoirs or for geological disposal (i.e., Pasidonia shale, Germany, Klaver et al., 2015; Callovo-Oxfordian argillites, France, Song et al., 2015; Opalinus clay, Switzerland, Keller et al., 2013; Houben et al. 2014; Bakken shale, United States of America, Liu and Ostadhassan, 2017; and Bowland shale, United Kingdom, Ma et al., 2016; Fauchille et al., 2018). Following this approach, REV is defined as the elementary volume below which the mean and/or standard deviation of a given property (e.g. porosity and volume fraction of a given mineralogical phase) vary significantly with scale.

Other methods use more sophisticated statistical information: the covariance or the two-point probability function (Rolland et al., 2007), the lineal path function (Łydżba and Różański, 2014) and the percolation length based on a percolation analysis of 2D or 3D microstructures (Hilfer 1991, 1996; Boger et al., 1992; Keller et al., 2013; Cosenza et al., 2015a,b). The variogram, a geostatistical function that was recently used to quantify the microscopic heterogeneity of shale (Gaboreau et al., 2016; Semmani and Borja, 2017), can also provide an interesting tool to infer L_{REV} (L_{REA} in 2D). Most of these statistical descriptors are used to obtain statistical representations or reconstructions of porous media (e.g. Torquato

and Stell, 1982; Singh et al., 2008) and are known to provide estimates of the REV minimum size of random media (Łydżba and Różański, 2014). However, they are scarcely used in practice to infer the L_{REV} or L_{REA} of shale, and one may wonder if all the aforementioned methods would provide similar L_{REV} or L_{REA} estimates in cases in which they would have been calculated for the same shale microstructure.

The objective of this paper is thus threefold. We would like to provide

- New L_{REA} estimates obtained from two mineral maps (Jorand, 2006; Fauchille, 2015) acquired from two shales that are actively studied in the framework of the deep disposal of radioactive waste: the Callovo-Oxfordian (COx) claystone from the Meuse/Haute-Marne underground research laboratory (Eastern France) and the Toarcian argillite from the experimental station of Tournemire (Southern France). These L_{REA} estimates have been calculated not only using classical box-counting methods but also statistical and geostatistical descriptors that are usually used to quantitatively describe microstructures (two-point probability function, lineal path function, percolation length and variogram).
- A review of the different values of L_{REA} estimates provided by the literature and by our study, all acquired for shales that have been actively studied in the last decade. This review accounts for all types of shales, in terms of clay-rich rocks and whatever their organic matter content. Indeed, in a practical viewpoint, the methods used to estimate values of L_{REA} at the mesoscopic scale (level 3) as a function of the spatial distribution of non-clay/organic matter grains or patches, are similar for organic-rich and organic-poor shales, whatever their organic matter content. This is mainly due to the methodologies used in these studies which focus on the sole clay phase; the others phases, mineral or organic, being considered as a whole (e.g., Ma et al., 2016) or embedded in the pore phase (e.g., Klaver et al., 2015; Fauchille et al., 2018).

- A discussion of our different L_{REA} estimates and their associated methodologies, aiming to identify the most suitable method to infer in a practical viewpoint L_{REA} of shale.

2. Materials and Methods

2.1 Geological setting and mineral maps

The L_{REA} estimates calculated in this study were obtained from two mineral maps that were acquired following the methodologies described below.

The first mineral map, hereafter called the COx map, was developed from a sample obtained from Callovo-Oxfordian (COx) claystone, which is extensively studied in the Meuse/Haute-Marne Underground Research Laboratory (MHM-URL) (Eastern France). The thickness of this formation is 130 m, and its age is 150-160 My. The formation is located 420-550 m below the surface, in the eastern part of the Paris Basin (Andra, 2005). The Callovo-Oxfordian formation contains mainly 25 to 65 wt.% clay minerals, with 20-42 wt.% carbonate (calcite, dolomite, ankerite) and 15-31 wt.% tectosilicate (quartz and feldspars) (Andra, 2005). This mineral map was prepared from a drill-core, denoted as EST05-709 (-492.2 m) and extracted from the Andra EST205 borehole (Jorand, 2006). It was obtained at micrometer spatial resolution from an advanced image processing of a chemical elements map that was acquired through the use of a Cameca SX100 electron probe microanalyzer (Prêt, 2003). This electron microanalyzer provides quantitative concentration maps of 14 chemical elements (Al, Na, K, Ca, Si, Mg, Ti, Fe, S, Ba, Zr, P, Zn, Sr) on a 3 x 0.5 mm^2 area with a spatial resolution of 2 µm/pixel. The image processing of these maps is based on mineral identification methods that accommodate mixtures and solid solutions and that are implemented in the in-house µPhaseMap software (Prêt, 2003). For details, the reader is referred to Prêt (2003), Prêt et al. (2010a,b) and Gaboreau et al. (2017). In our case, this methodology allows the spatial localization of all 16 different rock forming minerals, including different clay minerals: illite-

smectite mixed layers, kaolinite, mica (glauconie and muscovite) and chlorite (i.e., chamosite).

The geometrical and mineralogical features of the COx map are presented in Table 1. The surface fractions of clay matrix, carbonates and tectosilicates are 50.4%, 30.7% and 15.1%, respectively. Note that the greatest side (1563 pixels, 3072 µm) and smallest side (250 pixels, 500 µm) are perpendicular and parallel to the bedding, respectively.

The second mineral map, hereafter called the Toar map, has been extracted from a large mosaic of SEM images acquired from a nonimpregnated and dried Tournemire clay rock sample (Fauchille et al., 2014; Fauchille, 2015). The studied Tournemire clay rock sample has been sampled in the horizontal and cylindrical borehole FD90 in the 1996 East gallery of the Tournemire Underground Research Laboratory (URL) of the French Institute for Radioprotection and Nuclear Safety (IRSN). The sample was located at a depth between 4.20 to 4.40 meters far from the gallery wall, outside the so-called Excavation Damaged Zone. The Tournemire URL is located in a Mesozoic basin on the southern border of the Massif Central (Aveyron, France), in the subhorizontal consolidated argillaceous Toarcian (Toar) formation (200 meters thick) and marly layers of the Domerian age (50 meters thick). The sample comes from the upper Toarcian formation, the mineralogical composition of which shows that clay minerals represent nearly 25-50 wt% of the rock with illite (10-40 wt%) and illite/smectite mixed-layer minerals (5-25 wt%), kaolinite (10-35 wt%) and chlorite (1-5 wt%). The Tournemire shale also contains 10-40 wt% of carbonates, 10-30 wt% of quartz, 2-7 wt% of sulfides and less than 2 wt% of feldspars (Cabrera et al., 2001).

A mosaic (7.1 x 5.2 mm², 11302 x 8355 pixels) has been built from one hundred and fifty three back-scattered electron images (spatial resolution of 0.625 µm.pixel^{-1}) acquired by scanning electron microscopy (SEM, JEOL JSM 56000LV with an acceleration voltage of 15 KV, a probe current of 5 nA, a working distance of 16.3 mm, a magnification of x200, and a

dwell time of 128 µs per pixel). On the mosaic of images, clay-matrix, carbonates, tectosilicates, heavy minerals (e.g. pyrite) and macropores were discriminated by in house algorithms implemented in the µPhaseMap© software developed in the IC2MP laboratory in Poitiers, allowing a threshold for clay matrix and nonclay grains (Prêt et al., 2010a,b). The mineral map used in this study is an extraction (4000x4000 pixels) of the most homogeneous part of the mosaic to respect at best the statistical homogeneity assumed in all further L_{REA} calculations and to exclude laminae that imply additional heterogeneities at the macroscopic scale (i.e., level 4 in Fig. 1). Statistical homogeneity means herein that the statistical properties of interest (e.g. surface fraction and corresponding variance of a given mineral and microstructural descriptors under consideration) do not depend on the absolute positions where they are calculated. The geometrical and mineralogical features of the Toar map are also presented in Table 1.

An extraction of both maps is displayed in Figures 2 and 3. Note that both maps were prepared from a polished section in a plane perpendicular to the stratigraphic plane. The x-direction indicated in Figures 2 and 3 is parallel to the bedding planes, whereas the z-direction is perpendicular to the bedding.

The results of these image analysis are given in numerical table files in which the location and mineral code of each pixel of the mineral maps are indicated. However, in the following, only two phases will be considered in these numerical files: the clay phase corresponding to all clay minerals, and the nonclay phase gathering all nonclay minerals. These two-phase numerical files constitute the input files for calculations of L_{REA} estimates, which are presented in the next section.

2.2 Calculations of L_{REA}

a Box-counting method

219 The box-counting method, which is likely the most popular to infer L_{REA}, starts from a
220 given domain or box in the digitalized image. Then, the mean of a surface property (surface
221 mineral contents, surface porosity, physical property etc.) is calculated within increasing
222 domains or boxes until reaching the actual image size (Houben et al., 2014; Klaver et al.,
223 2015; Wang et al., 2016). The characteristic size of the L_{REA} is considered to be reached when
224 the mean of the considered property does not evolve significantly with the increasing size of
225 the boxes. This procedure can be repeated for several starting domains to ensure that the
226 inferred L_{REA} is statistically representative of the whole image. This first method will be
227 named hereafter the classical box-counting (BC) method.

228 Regarding the BC method, which uses square domains, the COx map and the Toar
229 map have been divided into five and four nonoverlapping square areas, respectively,
230 following the partitioning presented in Figure 4. These nonoverlapping areas, named Ai
231 (i=1,..,6 for COx map; i=1,..,4 for Toar map), are associated with starting domains that are
232 defined and discriminated by the coordinates of their center Ci (i=1,..,6 for COx map; i=1,..,4
233 for Toar map) (Fig. 4). Note that the origin, i.e., x=0, z=0 of system of coordinates, is located
234 in the top left corner of both maps (Fig. 2, 3 and 4).

235 The particular partitioning displayed in Figure 4 is due to two reasons. First, the shape
236 of the COx map is clearly elongated following the z-axis and the center of the initial box
237 could not be located only at the center of this map to investigate the whole map. Second, it
238 was of interest from a statistical viewpoint to compare L_{REA} estimates calculated on different
239 areas with a comparable surface and thus to check the statistical homogeneity of each map
240 regardless of potential macroscopic heterogeneities (i.e., level 4 in Fig. 1).

241 *b Statistical method*

242 In this second method, the digitalized image is divided into nonoverlapping square
243 domains of a given size *L*. For each domain of size *L*, the surface property of interest (surface

mineral contents, etc.) is calculated. The mean and the standard deviation of this set of surface property values is then inferred. The size L_{REA} is considered to be reached when both the mean and the standard do not evolve significantly with the increasing size of the boxes L (e.g. VandenBygaart and Protz, R. 1999; Zhang et al. 2000; Song et al., 2015). This method assumes *a priori* that the statistical homogeneity of the considered property in the image is satisfied. This approach is often used to present a direct visualization of microstructural variability at the image scale. It will be referred to hereafter as the statistical (S) approach (Zhang et al. 2000).

The S approach has been carried out on whole maps and not by using the partitioning indicated in Figure 4.

c Percolation length and connectivity

The method described below has been introduced to determine a characteristic length that defines the size of domain of a digitalized image in which effective medium approaches can be used to explicitly account for percolation (e.g. Hilfer, 1991; 1996). Indeed, a quantitative characterization of percolation and connectivity is crucial to model effective transport properties, e.g., hydraulic conductivity (e.g. Keller et al., 2013), diffusion coefficient or electrical conductivity (e.g. Cosenza et al., 2015a).

This method looks like the previous statistical (S) approach. It is also based on spatial partitions of the mineral map: the digitalized image is divided into independent square domains (boxes) of a given size L on which the surface property of interest is calculated. However, this approach differs from the S approach in two aspects.

First, the connectivity of the clay fraction is determined in each box. This property allows the calculation of the total fraction of percolating boxes of size L, named $p(L)$, which characterizes the overall connectivity of the image at length scale L:

$$p(L) = \int_0^1 \mu(\phi, L)\lambda(\phi, L)d\phi \quad (1)$$

where ϕ is the local clay fraction measured in each box of size L and $\mu(\phi, L)$ and $\lambda(\phi, L)$ are the local clay fraction distribution and the local percolation probability, respectively. In practice, the local clay fraction distribution $\mu(\phi, L)$ is the frequency histogram of boxes of size L, having a local clay fraction ϕ. The local percolation probability $\lambda(\phi, L)$ is the fraction of boxes with a local clay fraction ϕ and side length L that allow percolation in the x and z directions. A box percolates in the x- (y-, z-, resp.) direction if there exists a path inside the clay phase connecting two faces of the measurement cell that are perpendicular to the x- (y-, z-, resp.) axis. This box is called a percolating box in the x- (y-,z-, resp.) direction (see appendix, Hilfer, 1991, 1996, Keller et al., 2013, Cosenza et al., 2015a,b, for the details of the calculations).

The second aspect of this approach deals with the explicit calculation of a characteristic length L_p, named the percolation length, which is assumed to be an estimation of the minimum value of L_{REA} (Biswal et al., 1998; Widjajakusuma et al., 2003). It is obtained using the following criterion:

$$\left(\frac{d^2p}{dL^2}\right)_{L=Lp} = 0 \qquad (2)$$

In practice, in the following criterion (2), parameter L_p corresponds to the inflexion point of the $p(L)$ curve. Following Widjajakusuma et al. (2003), the percolation length L_p is the length around which $p(L)$, which is often sigmoidal in shape, rapidly changes from a low value at small L to its trivial value $p(L \rightarrow \infty) = 1$ (if clay space is connected at the scale of the entire image). In other words, Equation (2) defines the domain of the transition between local connectivity (at small L) and global connectivity (at large L). Widjajakusuma et al. (2003) demonstrated that a reasonable estimate of the effective permittivity and conductivity can be obtained at length L_p. This approach has been carried out on whole maps and not by using the partitioning indicated in Figure 4.

d Two-point probability function

The two-point probability function, often named the covariance function, is usually is defined from the autocorrelation function of a phase *i* given by (e.g. Yeong and Torquato, 1998):

$$S_i(\boldsymbol{r_1}, \boldsymbol{r_2}) = \langle I_i(\boldsymbol{r_1}), I_i(\boldsymbol{r_2}) \rangle \quad (3)$$

where $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are two vectors associated with two arbitrary points in the system, angular brackets denote an ensemble average, and the characteristic function $1_i(\boldsymbol{r})$ is defined as

$$1_i(\boldsymbol{r}) = \begin{cases} 1, & when\ \boldsymbol{r}\ is\ in\ phase\ i \\ 0, & otherwise \end{cases} \quad (4)$$

The function $S_i(\boldsymbol{r_1}, \boldsymbol{r_2})$ is interpreted as the probability of finding two points at positions $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ both in phase i. When the microstructure is spatially stationary or statistically homogeneous, the two-point probability function depends only on the distance $r = |\boldsymbol{r_1} - \boldsymbol{r_2}|$ between two points and, therefore, can be simply expressed as $S_i(r)$ of phase *i*. $S_i(r)$ can also reach its maximal value of $\bar{\phi}_i$ (volume fraction of phase *i* of the whole map) at $r=0$ and decays with $r \to \infty$ to the asymptotic value of ϕ_i^2:

$$\lim_{r \to 0} S_i(r) = \bar{\phi}_i \quad (5)$$

$$\lim_{r \to \infty} S_i(r) = \bar{\phi}_i{}^2 \quad (6)$$

If the latter limit in equation (6) is reached before $r \to \infty$, for instance, for a value $r=R$, the points of the phase *i* separated from a distance larger than *R* are not correlated (Kanit et al., 2003). This parameter *R,* often called the covariance range, defines a "correlation length" or a "characteristic size" of heterogeneity (e.g. Rolland et al., 2017). Following this definition, parameter *R* can be considered an estimate of the minimum L_{REA} (Łydżba and Różański, 2014; Fauchille et al., 2018). Note that phase *i* corresponds herein to the clay or the nonclay phase.

In the following, function $S_i(r)$ will be calculated following the simple and efficient procedure described by Yeong and Torquato (1998) and Łydżba and Różański (2014). Consider a random microstructure from which a binary image constituting two phases named *i* and *j* (here clay and nonclay phase) has been obtained. In this binary and digitalized image, each pixel is attributed to a value: *0* or *1* depending on whether the pixel is phase *i* or phase *j*. Consequently, this image can be represented by a ~~square~~ matrix, named *[M]*, in which each element is associated with a pixel through indexes *m* and *n*, both defining the pixel location in the image. The element *M[m,n]* is equal to 1 or 0, for instance, if the corresponding pixel *(m,n)* is in phase *i* or in phase *j*, respectively. Therefore, following this procedure and equations (3) and (4), function $S_i(r)$ for phase *i* is simply calculated as follows:

$$\boldsymbol{S_i(r) = \frac{1}{N_x N_y} \sum_{n=1}^{N_y} \sum_{m=1}^{N_x} \frac{M[m,n]M[m,n+r] + M[m,n]M[m+r,n]}{2} \quad r = 1, 2, ..} \qquad (7)$$

where the product $N_x N_y$ is the total number of pixels in the image, and *r* is expressed in ~~the~~ pixels. Here, three remarks have to be formulated. First, the dimensions of the matrix *M* are $(N_x+r)(N_y+r)$. The elements of *M[m,n]* for m>Nx and n>Ny are taken equal to zero and do not contribute to the summations in equation (7). Second, it should be noted that function $S_i(r)$ can be calculated following a given direction if all *r* directions are parallel to this direction. Third, L_{REA} estimates from the two-point probability function have been calculated on whole maps and not following the map partitioning defined in Figure 4.

e Lineal path function

As the two-point probability function, the lineal path function is also a microstructural descriptor that is used to estimate a minimum of L_{REA} (e.g. Łydżba and Różański, 2014). In the case of statistically isotropic media, the lineal path function for phase I, named $L^{(i)}(r)$, is

defined as the probability that a line segment r lies wholly in phase i (here the clay or nonclay phase) when randomly "thrown" into the sample (Torquato, 2002). The lineal path function is a monotonically decreasing function of r and obeys the following conditions:

$$L^{(i)}(0) = \bar{\phi}_i \tag{8}$$

$$\lim_{r\to\infty} L^{(i)}(r) = 0 \tag{9}$$

Regarding the two-point probability function, phase i corresponds herein to the clay or the nonclay phase. Note that for statistically homogeneous but anisotropic media, function $L^{(i)}(r)$ will depend not only on the amplitude of r but also on the orientation of the corresponding vector $\boldsymbol{r}$. The function $L^{(i)}(r)$ can be calculated by using the same procedure used to infer chord length distributions of a given phase (e.g. Cousin et al., 1996).

By assuming a reasonable value of threshold t (e.g. 1-10%), the parameter L_{REA} is estimated at the lineal path t^* such as (Łydżba and Różański, 2014):

$$\forall\, r \in [t^*, +\infty[\Rightarrow\ L^{(i)}(r)\ \leq\ t \tag{10}$$

Equation (10) means that L_{REA} (equal to r^*) corresponds to the line segment r, which gives $L^{(i)}$ values lower (or equal) than the given threshold t. Note that L_{REA} estimates from the lineal path function have been calculated on whole maps and not following the map partitioning defined in Figure 4.

f Variogram range

An additional microstructural descriptor can be obtained from the concept of the variogram that is widely used in geostatistics and recently in analyses of microtomographic images acquired for shale (Gaboreau et al., 2016; Semmani and Borja, 2017). This microstructural descriptor is the variogram range defined in this section.

362 The variogram, commonly called the semivariogram, is often defined as a measure of
363 the spatial continuity of data acquired in heterogeneous media. Considering a property Z, the
364 associated variogram is defined as follows:

365 $$\gamma(\mathrm{r}) = \frac{1}{2} E[(\mathrm{Z}(x) - \mathrm{Z}(x + \mathrm{r}))^2] \quad (11)$$

366 where *r* is the lag distance between two measures of Zobtained at two locations, i.e., *x*
367 and *x*+*r* (note that *x* and *r* are scalars in 1D or vectors in 2D and 3D). *E[X]* is the expected
368 value of property *X*. In our case, property *Z* is a pixel value associated with two mineral
369 phases, i.e., the clay phase and the nonclay phase. Property *Z* is equal to 1 or 0 if the pixel of
370 interest is a clay phase or a nonclay phase.

371 In the case of statistically isotropic media, the variogram increases with increasing lag
372 distance until a certain distance is reached at which it becomes constant. The lag distance at
373 which the variogram becomes constant defines a correlation length or a range of influence,
374 and the value at of this point is called the range (e.g. Peters, 2012). This variogram range is
375 considered hereafter as an estimate of L_{REA}. It should be noted that if the medium is
376 statistically homogeneous, then

377 $$\lim_{r \to \infty} \gamma(\mathrm{r}) = Var(\mathrm{Z}) \quad (12)$$

378 where $Var(\mathrm{Z})$ corresponds to the variance of Z.

379 At this stage, three remarks have to be formulated. First, this is the first time, to our
380 knowledge, that the concept of variogram range is used to infer a correlation length of a shale
381 at the mesoscopic scale.

382 Second, note that if the images have *N* pixels, equation (11) introduces *N(N-1)/2* pairs
383 of pixel values associated with locations *x* and *x*+*r*. Thus, even an image of moderate size can
384 generate a very large number of pairs inducing large numerical files and a large computation
385 time; this is why the variogram calculations have been restricted for both maps to areas
386 corresponding to the map partitioning defined in Figure 4.

Third, if the medium under study is statistically homogeneous, the variogram is linked to the two-point probability function as follows (e.g. Matheron, 1971):

$$\gamma(\mathrm{r}) = S_i(0) - S_i(r) \quad (13)$$

In equation (13), the variogram is calculated obviously for the phase i of interest. Despite the theoretical link between $S_i(r)$ and $\gamma(\mathrm{r})$, the latter is much easier to calculate in practice because there exist numerous commercial software packages in geostatistics. In other words, this additional microstructural descriptor was introduced, i.e., the variogarm range, because it is likely easier to determine it in practice than the two-point probability function.

In our study, variograms have been calculated using commercial Surfer© software. To calculate a variogram, Surfer© introduces a variogram grid approach instead of the classical pair comparison files. The variogram grid is a polar grid in which Surfer© places and stores all the pairs introduced by the equation (11). In practice, the user defines (a) the angular divisions, i.e., the number of angular division in the polar grid, (b) the radial divisions, i.e., the number of concentric circles in the grid and (c) the largest separation distance contained in the variogram grid. In our calculations, the angular division and the radial division have been fixed for both maps at 180° and 100, respectively. The largest separation division of areas associated with the COx map and Toar map has been taken equal to 110 pixels (220 µm) and 200 pixels (125 µm), respectively. These choices are justified *a posteriori* since these parameters allow us to obtain the variogram ranges, as shown below.

3. Results and Discussion

3.1 Comparison between COx and Toar maps

a Box-counting method

Figures 5 and 6 display the evolutions of the clay fraction calculated for increasing box sizes L considering the different starting domains of the COx map and Toar map, respectively. Both figures confirm that the calculated mean clay fraction converges to a value,

412 referred hereafter to as parameter ϕ_∞, which is very close to the mean clay fraction obtained
413 for the whole map.

414 Considering the COx map (Fig. 5), the six curves associated with the six starting
415 domains converge to ϕ_∞values in the range [48.6-52.4%], including the mean clay fraction,
416 $\overline{\phi}_{COx}$ of 50.4% calculated for the whole map (Table 1). The difference between these
417 asymptotic values, ϕ_∞, for a *L* box of 500 µm and mean clay fraction $\overline{\phi}_{COx}$ is less than 4%. In
418 the same way in Figure 6, the four curves associated with the Toar map converge to
419 ϕ_∞ values in the range [68.5-71.4%], including also the mean clay fraction, $\overline{\phi}_{Toar}$ of 69.9%
420 calculated for the whole map (Table 1). The difference between the ϕ_∞values for a *L* box of
421 500 µm and $\overline{\phi}_{Toar}$ is less than 2.5%.

422 Below 100 µm and 200 µm for the COx and Toar maps, respectively, the curves
423 obtained for the different subdomains present non-correlated and high frequency evolutions.
424 Such behavior is associated with the occurrence of a few grains with a large size (Robinet,
425 2008; Fauchille, 2015), and a sufficiently large box size including several grains should be
426 used to estimate a meaningful REA (Gaboreau et al 2016). For a box size larger than 100 µm
427 and 200 µm for COx and Toar maps, respectively, the gap between curves decreases
428 progressively with low frequency variations. A meaningful REA corresponding to the
429 mesoscopic scale (level 3 in Fig. 1) can be estimated with improved accuracy when the box
430 size increases. Careful observation of both maps reveals that the grain size is larger for Toar
431 than for COx (compare the calcite grains in Fig. 2 and carbonates grains in Fig. 3), explaining
432 why a larger box size is needed for the Toar map to reach REA, as illustrated below (Fig. 6).

433 The REA size, i.e., L_{REA}, of both maps has been estimated in two steps (Table 2). In a
434 first step, REA sizes have been calculated for each nonoverlapping area of both maps (6 areas
435 for the COx map and 4 areas for the Toar map) and for two errors or threshold values: ε= 0.1
436 (10%) and ε= 0.05 (5%). For each area, the L_{REA} parameter has been identified as the lowest

box size L for which the calculated mean clay fraction was significantly similar to that of the whole map ($\overline{\phi}_{COx}$ or $\overline{\phi}_{Toar}$) with a maximum error of ε. In a second step, the average over all L_{REA} estimates of all nonoverlapping areas has been calculated for each map. The calculated mean L_{REA} for a given map has been considered as the L_{REA} of the latter.

Our results presented in Table 2 indicate the L_{REA} values are dispersed and decreasing functions of the chosen error or threshold values ε. Indeed, in the first case (ε=0.1), the mean value of the L_{REA} estimates of the COx map and Toar map are 173 µm and 129 µm, respectively (Table 2), whereas in the second case (ε=0.05), the mean values of L_{REA} estimates of the COx map and Toar map are much higher: 234 µm and 441 µm, respectively.

The scatter of the mean values of L_{REA} estimates and the associated standard deviation values (Table 2) question the statistical homogeneity of both maps and can be explained by a small but significant evolution of the microstructure in the x direction and/or in the z direction (i.e., with depth), as shown by (a) the ϕ_∞ estimates of the COx map, which globally increase with depth, and by (b) the ϕ_∞ estimates of the Toar map, which roughly decrease in the x direction (Table 2). This aspect will be discussed further with regard to the results obtained with the microstructural descriptors under consideration in this work.

b Statistical method

The results obtained by the S approach are displayed in Figures 7 and 8. The REA size, L_{REA} following the S approach, has been estimated in Figure 8 from the slope of the standard deviation values displayed in Figure 7. In fact, L_{REA} has been calculated as the intersection of a smooth curve fit (bold line) of the slope standard deviation curve and a horizontal line (dashed line) symbolizing a constant evolution with the box size (see white arrows in Fig. 8). Following Figure 8, the L_{REA} estimates of the COx map and the Toar map are 100 ± 10 µm and 140 ± 10 µm, respectively. These L_{REA} estimates are clearly less

461 dispersed than those provided by the BC method. In addition, and again in contrast to the
462 results obtained using the BC method, the S approach clearly indicates that the L_{REA} of the
463 COx map is lower than that of the Toar map.

464 Note that the maximum box sizes of the COx map and Toar map in Figure 7 have been
465 restricted to values of 140 μm and 250 μm, respectively. This choice is justified *a posteriori*
466 by Figure 8, which shows that the standard deviations do not evolve significantly after the
467 L_{REA} estimates indicated by white arrows.

468 *c Microstructural descriptors*

469 The results obtained using the following three microstructural descriptors, i.e., the
470 two-point probability function, lineal path function and variogram range given in Figures 9,
471 10, 11 and 12 (also see the recapitulation in Table 3) confirm two outcomes provided by the
472 previous BC and S approaches.

473 First, COx L_{REA} is lower than Toar L_{REA}. If the two-point probability function and the
474 lineal path function are considered and irrespective of the mineral phase (clay or nonclay
475 phase) under consideration, L_{REA} estimates from the COx map are always significantly lower
476 than those obtained from the Toar map (Table 3). A factor almost equal to 2 exists between
477 both groups of L_{REA} estimates.

478 Second, the two-point probability function (Fig. 9 and 10) and variogram (Fig. 12a and
479 12b) exhibit features that again question the statistical homogeneity of the maps, especially
480 the Toar map. Indeed, regardless of the direction and mineral phase (clay or nonclay phase)
481 under consideration, the two-point probability function $S_i(r)$ of both maps does not converge
482 to the expected asymptotic values (i.e., the square of the phase fraction of the whole map, see
483 equation (6)) (Fig. 9 and 10). This nonconvergence underlines the existence of a "long-range
484 order", i.e., the existence of a macrostructure with a characteristic size exceeding the size of
485 the investigated subdomains (i.e., 100 μm). A gradient of the property under consideration at

the map scale is a typical expression of a “long-range order”. The existence of this “long-range order” or macrostructure is confirmed at least on the Toar map by our geostatistical calculations. Four variograms calculated at different locations on the Toar map do not converge to the same plateau, i.e., to the same value of clay fraction variance (see equation (9)) (Fig. 12b). Indeed, if these four variograms taken independently suggest a statistical homogeneity at the area scale of interest, the four plateau values towards which they converge are clearly different.

Moreover, these three microstructural descriptors exhibit two additional features that have not been evidenced using previous BC and S approaches. First, the microstructural L_{REA} values are all much lower than those obtained using previous BC methods. The L_{REA} estimates of the COx map and Toar map are in range [16-53 µm] and [27-103 µm] (Table 3), whereas the L_{REA} estimates obtained from the classical BC are in range [86-438 µm] and [179-576 µm] (Table 2), respectively. This feature confirms the statements of Łydżba and Różański (2014) and Fauchille et al. (2018) that the two-point probability function and the lineal path function predict the lower bounds of L_{REA}. These functions define a characteristic heterogeneity size in the image, here the clay matrix and grain domains, and not a size that would be sufficiently large to be statistically representative of all heterogeneities present in the same image. In practice, this characteristic size of heterogeneity can be linked to a surface weighted mean grain diameter <d> estimated from the frequency distribution of the grain area or grain size distribution (GSD) as follows:

$$< d >= \sum_{i=1}^{N} \left(d_i \frac{f_i S_i}{\sum_{j=1}^{N} f_j S_j} \right) \quad (14)$$

where d_i is the equivalent spherical diameter of grains having area S_i ($d_i = \sqrt{\frac{4}{\pi} S_i}$), N is the total number of classes of GSD, and f_i is the fraction in number of grains having area S_i. By considering the GSD of both shales under study in Robinet (2008) and Fauchille (2011),

510 the calculated mean grain diameter of COx claystone and Tournemire argillite is 19 μm and
511 28 μm, respectively. These values are close to the L_{REA} estimates derived from microstructural
512 descriptors (Table 3).

513 Second, these microstructural descriptors evidence an intrinsic and well-known
514 property of shale: its structural anisotropy. Shale structural anisotropy is reflected herein by
515 the dependence on the two-point probability function $S_i(r)$ and the lineal path function
516 $L^{(i)}(r)$ with respect to the direction of calculation. As indicated in Figures 8 and 9, the
517 horizontal $S_i(r)$ and $L^{(i)}(r)$ of both shales are different from the vertical $S_i(r)$ and $L^{(i)}(r)$. In
518 particular, $L^{(i)}(r)$ in the x-direction is clearly larger than in the z-one. In our context and
519 following the methods used in this work, this anisotropy is not due to the alignment of clay
520 particles and aggregates but rather to the alignment of elongated and oriented nonclay grains
521 parallel to the bedding planes. This result confirms previous petrographical and petrophysical
522 studies (e.g. Robinet et al., 2007, 2012; David et al., 2007; Cosenza et al., 2015a). Moreover,
523 it should be emphasized that such a structural anisotropy could not be evidenced using BC
524 and S methods, which intrinsically do not depend on a given direction.

525 *d Effect of the connectivity of the clay phase*

526 The effect of the connectivity of the clay phase on the L_{REA} estimate, which cannot be
527 accounted for by previous approaches, may be discussed using the concept of the percolation
528 length L_p. It can be recalled that parameter L_p is a correlation length that is calculated using (i)
529 the function $p(L)$ expressing the total fraction of percolating boxes of size L for a given map
530 and (ii) the condition given by the equation (2). This method is conceptually close to the S
531 approach, since both methods use a map partitioning by nonoverlapping boxes of size L over
532 which statistical parameters are calculated.

533 Figure 13 displays the $p(L)$ curves of both maps. The COx $p(L)$ curve indicates an L_p
534 estimate in range [110-130] μm, which is significantly lower than the L_{REA} estimates provided

by the classical BC method (Table 2). Thus, it may be tempting to conclude that consideration of the clay phase connectivity of the mineral map would lead to lower estimates of the L_{REA} parameter. However, if this L_p value is now compared with L_{REA} estimates given by the S approach (Fig. 7), it is difficult to reach a clear conclusion. Indeed, the COx L_{REA} estimate obtained by the S approach is 100 µm, which is close to the L_p range indicated in Figure 13.

Considering the Toar map, its *p(L)* curve does not exhibit a clear inflexion point, and thus, an L_p estimate cannot be unambiguously achieved. This difficulty can be linked to the other microstructural descriptors, which have shown that the Toar map is not homogeneous from a view statistical viewpoint.

Consequently, we think that our results do not permit a clear conclusion regarding the impact of the clay phase connectivity on the REA size estimate. Thus, we recommend the calculation of L_p parameter for mineral maps of other shales to better assess the impact of the clay phase connectivity.

3.2 Comparison with the literature – Towards a practical methodology

Table 4 displays REA estimates obtained from different shales that have been extensively studied in the last decade. It illustrates the variety of investigation techniques and target phases used to infer L_{REA}. Moreover, it shows the following main result: regardless of the shale under consideration and investigation techniques and corresponding resolution, L_{REA} values are always on the order of a few hundred microns. If only shaley facies and clay phase targets are considered, L_{REA} estimates are restricted in the range [50-200 µm]. Our COx and Toar estimates do not escape this range of values. Our estimates obtained with a BC method and with an error of 10% are close to those calculated from other shales and using other investigation techniques.

In addition, Table 4 confirms that the classical BC method is largely used to calculate L_{REA} from acquired images that are rarely larger than one millimeter. In our opinion, however,

560 this method conceals a few drawbacks and biases that may ultimately question the accuracy of
561 L_{REA} estimates shown in table 4. First, as mentioned previously, L_{REA} estimates provided by
562 the BC method are often very dispersed and thus suffer from high uncertainty (see the
563 standard deviation values in Table 2), especially in comparison to the S method. Moreover, it
564 should be recalled that the L_{REA} values found using the BC method depend considerably on
565 the threshold ε, which is not always indicated by the authors. In other words, some L_{REA}
566 estimates in Table 4 have been obtained with ε thresholds that likely differ from the classical
567 10% value and thus should be cautiously compared with the other estimates.

568 Third, the BC method assumes a statistical homogeneity at the image scale that may
569 not be satisfied in all cases, as especially illustrated with the Toar map in this study. The
570 failure of this assumption may again question the accuracy of L_{REA} values provided by the
571 literature. Moreover, all the shales under study in table 4 are, to a certain extent, anisotropic,
572 and one may wonder if their structural anisotropy significantly influences L_{REA} determination
573 using the classical BC approach, which intrinsically does not account for this property.

574 These drawbacks and biases finally pose a methodological challenge that can be
575 summarized using the following questions: what is the most suitable and more practical
576 methodology to determine L_{REA} regarding shale properties? Is there a simple and robust
577 methodology that would minimizes the biases associated with the classical BC method? To
578 answer to these questions, the advantages and drawbacks of each method used in this work
579 are listed in Table 5.

580 Table 5 shows three main features. First, the S method easily provides L_{REA} estimates
581 that are close to those obtained from the BC method and much less dispersed (compare the
582 results presented in Tables 2 and 3). Second, the two-point probability function is a simple
583 approach that is easy to implement (see equation (7)) and that can clearly evidence the
584 statistical heterogeneity and anisotropy of mineral maps under study (Fig. 10). Third, despite

585 their interest in characterizing the structural anisotropy and heterogeneity at the mesoscopic
586 scale, the lineal path function and variogram range are more difficult to implement, and their
587 estimates are clearly L_{REA} lower bounds, which are rather far from the BC and S estimates.

588 This set of features finally leads to the proposal of a two-step methodology to infer
589 L_{REA}. In a first step, the two-point probability function and variogram should be used to check
590 the statistical homogeneity of the microstructure under study. The two-point probability
591 function can be calculated following the simple algorithm described in this paper (e.g.
592 equation (7)), and the variogram can be achieved using commercial software packages for
593 geostatistical analysis. If statistical homogeneity is verified, in a second step, L_{REA} can be thus
594 estimated using the S method and eventually using the classical BC method for verification.
595 This methodology offers three benefits: (i) a simplicity in the estimation of L_{REA}, since simple
596 and conventional algorithms and software are combined and used; (ii) a better estimation
597 accuracy through the use of the S approach; and above all (iii) a validation of the statistical
598 homogeneity of the studied maps and images, which is rarely ensured in practice. Regarding
599 the last point, such a methodology is well in line with the occurrence of additional
600 heterogeneities on a larger scale, i.e., the macroscopic scale of sedimentary laminae, which is
601 also defined as level 4 in Figure 1.

602 **4. Conclusion**

603 The main objective of this work was to estimate REA sizes of two shales that are
604 actively studied in the framework of the deep disposal of radioactive waste: Callovo-
605 Oxfordian (COx) claystone from the Meuse/Haute-Marne underground research laboratory
606 (Eastern France) and Toarcian argillite from the experimental station of Tournemire
607 (Southern France). The L_{REA} estimates obtained from two mineral maps on a mesoscopic
608 scale were calculated using classical methods (i.e., box-counting and statistical approaches)
609 and different microstructural descriptors (i.e., two-point probability function, lineal path

610 function, percolation length based on a percolation analysis of 2D or 3D microstructures,
611 variogram range). These calculations provided the following results:

612 • The classical box-counting method provides L_{REA} estimates ranging 129 μm up to 441
613 μm. These estimates are consistent with those obtained from other shales in the literature
614 but are very sensitive to the chosen ε threshold. Moreover, they show a wide scatter,
615 which questions the statistical homogeneity of the mineral maps.

616 • In comparison to the box-counting method, the statistical method provides L_{REA} estimates
617 on the same order of magnitude but with a much lower scatter (approximately 10% of the
618 inferred value). The L_{REA} estimates of the COx map and the Toar map are 100 ± 10 μm
619 and 140 ± 10 μm, respectively.
620 • Although the two-point probability function and lineal path function infer L_{REA} lower
621 bounds, they are able to evidence the microstructural anisotropy of both shales of interest
622 and, by considering the former, to assess the statistical homogeneity of the maps.
623 • In particular, the two-point probability function and variogram have both confirmed that
624 the Tournemire mineral map is clearly not statistically homogeneous with regard to its
625 mineral composition. This aspect makes it difficult and even questionable to determine
626 the L_{REA} of this particular map.
627 • The calculations of percolation length L_p on both maps and their comparison with L_{REA}
628 estimates do not permit a clear conclusion regarding the impact of the clay phase
629 connectivity on the REA size. Thus, we suggest the calculation of the L_p parameter for
630 mineral maps and images of other shales to better assess the impact of the clay phase
631 connectivity.

632 This set of results ultimately leads to the recommendation of a two-step methodology
633 to infer L_{REA} from a practical viewpoint. In a first step, the two-point probability function and

variogram should be used to check the statistical homogeneity of the microstructure under study. The two-point probability function can be calculated following the simple algorithm described in this paper (e.g. equation (7)), and the variogram can be achieved simply using commercial software packages for geostatistical analysis. If the statistical homogeneity is satisfied, in a second step, L_{REA} can be estimated by the statistical method and eventually by the classical box-counting method for verification. Moreover, it should be mentioned that this methodological recommendation is also valid for organic-rich shales since the methods used in this work, focus on the sole clay phase; the others phases, mineral or organic are considered as a whole set.

One perspective of this work is to support our results and recommendations concerning the use of more sophisticated and stochastic methods to determine shale L_{REA}. For instance, the approach proposed by Kanit et al. (2003) introduces the concept of the "statistical" REV or REA, which is related not only to the microstructure and properties of each component, but above all, to a given precision in the estimation of the effective property, depending on the number of realizations "that one is ready to generate" (Jeulin et al., 2004). The probabilistic concept of realization herein is any representation or observation of the microstructure considered with a given size and fraction of heterogeneities, i.e., in our case, a subdomain of a given size or a set of subdomains of a mineral map under study. Consequently, compared with our previous "deterministic" approaches, this "statistical" L_{REA} depends explicitly on an additional parameter: the precision desired for the estimate of the effective property (for instance, the mean clay fraction in our case) and reached for a given number of realizations. As a consequence, the estimate of "statistical" L_{REA} cannot be unique. This approach has been recently applied to Opalinus clay samples at Mont Terri rock laboratory in Switzerland (Houben et al., 2014; Keller, 2015) and could be applied fruitfully to our mineral maps.

Acknowledgments. We thank NEEDS-MIPOR – VARAPE (VARiabilité spatiale des propriétés de l'Argilite au travers d'une approche numérique-expérimentale à Plusieurs Echelles) program for its financial support. A part of this study was also performed in the framework of the research project "ExCiTING" funded by the French National Research Agency (grant agreement ANR-17-CE06-0012). The authors also acknowledge financial support from the European Union (ERDF) and "Région Nouvelle Aquitaine".

Appendix. Mathematical definition of p(L)

The total fraction of percolating boxes of size *L*, named *p(L),* which characterizes the overall connectivity of the image at length scale *L,* is defined as follows:

$$p(L) = \int_0^1 \mu(\phi, L)\lambda(\phi, L)d\phi \quad \text{(A1)}$$

where ϕ is the clay fraction measured in each box of size *L*, and $\mu(\phi, L)$ and $\lambda(\phi, L)$ are the local clay fraction distribution and the local percolation probability, respectively. These two functions, $\mu(\phi, L)$ and $\lambda(\phi, L)$, are assumed to constitute an approximate but reasonable geometric characterization of the clay distribution in shales (i.e., the local simplicity assumption introduced by Hilfer 1991, 1996). By definition, these functions can be calculated from photographs or numerical images of 2D thin sections in a fairly straightforward manner as explained below.

Let us consider a shale sample *S* (here a mineral map), constituting a clay space *C* and solid nonclay space *NC* (i.e., $S = C \bigcup NC$). We choose a partitioning $K=\{K_1, .., K_j, .., K_M\}$ of the sample space *S* into *M* mutually disjoint subsets, called boxes. As a result, $\bigcup_{j=1}^{M} K_j = S$ and $K_i \bigcap K_j = \emptyset$ if $i \neq j$. Each box K_j constitutes itself in M_j elementary volume elements. An elementary volume element is the elementary voxel in a 3D sample or the elementary pixel in a 2D sample or map.

A particular and simple partitioning K is a cubic lattice for a 3D sample or a square lattice for a 2D sample. This choice conveniently features K_j cells that are translated copies of one another and the same set (they all have the same shape). The local clay fraction $\phi(K_j)$ inside a box K_j can be defined as

$$\phi(K_j) = \frac{V(C \cap K_j)}{V(K_j)} = \frac{1}{M_j} \sum_{\boldsymbol{r}_i \in K_j} 1_c(\boldsymbol{r}_i) \quad \text{(A2)}$$

where $V(K_j)$ is the volume of a subset, M_j denotes the number of volume elements (voxels or pixels) in K_j and 1 is the characteristic function (indicator function) of the clay space C:

$$1(\boldsymbol{r}_i) = \begin{cases} 1 & if\ \boldsymbol{r}_i \in C \\ 0 & otherwise \end{cases} \quad \text{(A3)}$$

From this definition of the local clay fraction $\phi(K_j)$, the histogram called the local clay fraction distribution $\mu(\phi, K)$ can be introduced as follows:

$$\mu(\phi, \boldsymbol{K}) = \frac{1}{M\,\Delta\phi} \sum_{j=1}^{M} \sum_{i=1}^{k} \left(1_{I_i}(\phi)\ 1_{I_i}\left(\phi(K_j)\right) \right) \quad \text{(A4)}$$

where k is the number of classes of the histogram, and I_1,., I_i,., I_k are the classes of the histogram, $\Delta\phi$ defines the interval width of each class (all classes have the same width), 1_{I_i} is the indicator function:

$$1_{I_i}(\phi) = \begin{cases} 1 & if\ \phi \in I_i \\ 0 & otherwise \end{cases} \quad \text{(A5)}$$

In the practical case of a cubical box $K_j = K(\boldsymbol{r}_j, L)$ of side-length L centered at the lattice vector $\boldsymbol{r}_j$ (i.e., typically a Bravais lattice), the local clay fraction distribution can be rewritten as follows:

$$\mu(\phi, L) = \frac{1}{M\,\Delta\phi} \sum_{j=1}^{M} \sum_{i=1}^{k} \left(1_{I_i}(\phi)\ 1_{I_i}\left(\phi(\boldsymbol{r}_j, L)\right) \right) \quad \text{(A6)}$$

The local clay fraction distribution $\mu(\phi, L)$ also has the following physical meaning: it measures the probability of finding the local clay fraction ϕ between ϕ and $\phi + d\phi$ in a measurement cell of linear dimension L.

706 The second geometrical property to characterize the local geometry of shales is $\lambda(\phi,L)$,
707 the fraction of percolating box of side-length L with local clay fraction ϕ. The local function
708 $\lambda(\phi,L)$, also called the "local percolation probability", is defined as follows:

709 $$\lambda(\phi, L) = \frac{\sum_{j=1}^{M} \Lambda(\boldsymbol{r}_j,L) \sum_{i=1}^{k} \left(1_{I_i}(\phi)\, 1_{I_i}(\phi(\boldsymbol{r}_j,L)) \right)}{\sum_{i=1}^{k} \left(1_{I_i}(\phi)\, 1_{I_i}(\phi(\boldsymbol{r}_j,L)) \right)} \quad \text{(A7)}$$

710 where the indicator function $\Lambda(\boldsymbol{r}_j,L)$ for the percolation of cell $K(\boldsymbol{r}_j, L)$ is given by

711 $$\Lambda((\boldsymbol{r}_i, L) = \begin{cases} 1 & \text{if box at } \boldsymbol{r}_i \text{ percolates in x and z direction} \\ 0 & \text{otherwise} \end{cases} \quad \text{(A8)}$$

712 A measurement cell $K(\boldsymbol{r}_j, L)$ percolates in the x- (y-, z-, resp.) direction if a path inside
713 the clay phase exists connecting two faces of the measurement cell that are perpendicular to
714 the x- (y-, z-, resp.) axis. In practice, the function $\Lambda(\boldsymbol{r}_j,L)$ can be calculated using the Hoshen-
715 Kopelman algorithm (Hoshen and Kopelman, 1976).

716 **References**

717 Andra 2005, Dossier 2005 Argile: Synthesis. Evaluation of the feasibility of a geological
718 repository in an argillaceous formation. Andra, France (available at www.andra.fr).

719 Bennett, R.H., O'Brien, N.R., Hulbert, M.H., 1991. Determinants of clay and shale
720 microfabric signatures: Processes and mechanisms. In: Bennett, R., Bryant, W.,
721 Hulbert, M. (Eds), 1991, Microstructure of fine grained sediments: from mud to shale,
722 Springer-Verlag, New York, Chapter 2, pp. 5-32.

723 Biot, M.A., 1941. General theory of three-dimensional consolidation. J. Appl. Phys. 12(2),
724 155-164.

725 Boger, F., Feder, J., Jossang, T., Hilfer, R., 1992. Microstructural sensitivity of local porosity
726 distributions, Physica A. 187, 55-70.

727 Biswal, B. Manwart, C., Hilfer, R., 1998. Threedimensional local porosity analysis of porous
728 media, Physica A. 255, 221.

729 Cabrera, J., Beaucaire, C., Bruno, G., De Windt, L., Genty, A., Ramambasoa, N., Rejeb, A.,
730 Savoye, S., Volant, P., 2001. Projet Tournemire : Synthèse des Résultats des
731 Programmes de Recherche 1995/1999. IRSN Report 2001.
732 Chalmers, G. R., Bustin, R. M., Power, I. M. (2012). Characterization of gas shale pore
733 systems by porosimetry, pycnometry, surface area, and field emission scanning
734 electron microscopy/transmission electron microscopy image analyses: Examples
735 from the Barnett, Woodford, Haynesville, Marcellus, and Doig units. Am. Assoc. Pet.
736 Geol. Bull. 96(6), 1099-1119.
737 Cosenza, P., Prêt, D., Zamora, M., 2015a. Effect of the local clay distribution on the effective
738 electrical conductivity of clay rocks. J. Geophys. Res. 120(1), 145-168.
739 Cosenza, P., Prêt, D., Giraud, A., Hedan, S., 2015b. Effect of the local clay distribution on the
740 effective elastic properties of shales. Mech. Mater. 84, 55-74.
741 Cousin, I., Levitz, P., Bruand, A., 1996. Three-dimensional analysis of a loamy-clay soil
742 using pore and solid chord distributions. Eur. J. Soil Sci. 47(4), 439-452.
743 Curtis, M. E., Cardott, B. J., Sondergeld, C. H., Rai, C. S., 2012. Development of organic
744 porosity in the Woodford Shale with increasing thermal maturity. Intern. J. Coal Geol.
745 103, 26-31.
746 David, C., Robion, P., Menendez, B., 2007. Anisotropy of elastic, magnetic and
747 microstructural properties of the Callovo-Oxfordian argillite, Phys. Chem. Earth. 32,
748 145-153.
749 Fauchille, A.L., Hedan, S., Prêt, D., Valle, V., Cabrera, J., Cosenza, P., 2014. Relationships
750 between desiccation cracking behavior and microstructure of the Tournemire clay rock
751 by coupling DIC and SEM methods. Proceedings of IS on Geomechanics from Micro
752 to Macro, Cambridge, UK. CRC Press/Balkema, Leiden, The Netherlands, 1421-1424.

753 Fauchille, A.L., 2015. Déterminismes microstructuraux et minéralogiques de la fissuration
754 hydrique dans les argilites de Tournemire : apports couples de la pétrographie
755 quantitative et de la correlation d'images numériques, PhD thesis, University of
756 Poitiers (France) (in French).
757 Fauchille, AL van den Eijnden, A.P., Ma L., Chandler, M., Taylor, K.G., Madi, K., Lee, P.D,
758 Rutter, E., 2018. Variability in spatial distribution of mineral phases in the Lower
759 Bowland Shale, UK, from the mm- to μm-scale: Quantitative characterization and
760 Modeling, Mar. Petrol. Geol. 109-127.
761 Gaboreau, S., Robinet, J.C., Prêt, D., 2016. Optimization of pore-network characterization of
762 a compacted clay material by TEM and FIB/SEM imaging. Micro. Meso. Mater. 224,
763 116-128.
764 Gaboreau, S., Prêt, D., Montouillout, V., Henocq, P., Robinet, J.C., Tournassat C., 2017.
765 Quantitative mineralogical mapping of hydrated low pH concrete, Cem. Con.Comp.
766 83, 360-373.
767 Giraud, A., Gruescu, C., Do, D.P., Homand, F., Kondo, D., 2007. Effective thermal
768 conductivity of transversely isotropic media with arbitrary oriented ellipsoidal
769 inhomogeneities. Intern. J. Sol. Struct. 44(9), 2627-2647.
770 Han, Y., Horsfield, B., Wirth, R., Mahlstedt, N., Bernard, S., 2017. Oil retention and porosity
771 evolution in organic-rich shales. Am. Assoc. Pet. Geol. Bull. 101(6), 807-827.
772 Hemes, S., Desbois, G., Urai, J.L., De Craen, M., Honty, M., 2013. Variations in the
773 morphology of porosity in the Boom Clay Formation: insights from 2D high resolution
774 BIB-SEM imaging and Mercury injection Porosimetry. Nether. J. Geos. 92(4), 275-
775 300.
776 Hilfer, R., 1991. Geometric and dielectric characterization of porous media, Phys. Rev. B.
777 44(1), 60-75.

778 Hilfer, R., 1996. Transport and relaxation phenomena in porous media, Advances in Chemical
779 Physics, Ed. By I. Prigogine and S.T. Rice, 299-424.
780 Hoshen, J., Kopelman, R., 1976. Percolation and cluster distribution. I. Cluster multiple
781 labeling technique and critical concentration algorithm. Phys. Rev. B. 14(8), 3438.
782 Houben, M. E., Desbois, G., Urai, J.L., 2014. A comparative study of representative 2D
783 microstructures in Shaly and Sandy facies of Opalinus Clay (Mont Terri, Switzerland)
784 inferred form BIB-SEM and MIP methods. Mar. Petrol. Geol. 49, 143-161.
785 Houben, M. E., Barnhoorn, A., Wasch, L., Trabucho-Alexandre, J., Peach, C. J., Drury, M.R.,
786 2016. Microstructures of early jurassic (Toarcian) shales of Northern Europe. Intern. J.
787 Coal Geol. 165, 76-89.
788 Jeulin, D., Kanit, T., Forest, S., 2004. Representative volume element: a statistical point of
789 view. In Continuum Models and Discrete Systems (pp. 21-27). Springer, Dordrecht.
790 Jorand R., 2006, Etude expérimentale de la conductivité thermique: application au forage
791 EST205 du site de Meuse/Haute Marne (Andra), Ph-D Thesis, University of Denis
792 Diderot (Paris) (in French).
793 Kanit, T., Forest, S., Galliet, I., Mounoury, V., Jeulin, D., 2003. Determination of the size of
794 the representative volume element for random composites: statistical and numerical
795 approach. Intern. J. Sol. Struct. 40(13-14), 3647-3679.
796 Keller, L. M., Holzer, L., Schuetz, P., Gasser, P., 2013. Pore space relevant for gas
797 permeability in Opalinus clay: Statistical analysis of homogeneity, percolation, and
798 representative volume element. J. Geophys. Res., 118(6), 2799-2812.
799 Keller, L.M., 2015. On the representative elementary volumes of clay rocks at the
800 mesoscale. J. Geol. Min. Res. 7(6), 58-64.

Klaver, J., Desbois, G., Urai, J. L., Littke, R., 2012. BIB-SEM study of the pore space morphology in early mature Posidonia Shale from the Hils area, Germany. Intern. J. Coal Geol. 103, 12-25.

Klaver, J., Desbois, G., Littke, R., Urai, J.L., 2015. BIB-SEM characterization of pore space morphology and distribution in postmature to overmature samples from the Haynesville and Bossier Shales. Mar. Petrol. Geol. 59, 451-466.

Liu, K., Ostadhassan, M., 2017. Quantification of the microstructures of Bakken shale reservoirs using multi-fractal and lacunarity analysis. J. Nat. Gas Sci. Eng. 39, 62-71.

Loucks, R.G., Reed, R.M., Ruppel, S. C., Hammes, U., 2012. Spectrum of pore types and networks in mudrocks and a descriptive classification for matrix-related mudrock pores. Am. Assoc. Pet. Geol. Bull. 96(6), 1071-1098.

Łydżba, D., Różański, A., 2014. Microstructure measures and the minimum size of a representative volume element: 2D numerical study. Acta Geophysica 62(5), 1060-1086.

Ma, L., Taylor, K.G., Lee, P. D., Dobson, K. J., Dowey, P.J., Courtois, L., 2016. Novel 3D centimetre-to nano-scale quantification of an organic-rich mudstone: The Carboniferous Bowland Shale, Northern England. Mar. Petrol. Geol. 72, 193-205.

Ma, L., Fauchille, A.L., Dowey, P.J., Pilz, F.F., Courtois, L., Taylor, K.G., Lee, P.D., 2017. Correlative multi-scale imaging of shales: a review and future perspectives. Geological Society, London, Special Publications, 454, SP454-11.

Matheron, G. (1971). The theory of regionalized variables and their applications. *Centre de Geostatistique, Fontainebleau, Paris*, 211.

Ortega, J. A., Ulm, F. J., Abousleiman, Y., 2007. The effect of the nanogranular nature of shale on their poroelastic behavior. Acta Geotechnica 2(3), 155-182.

825 Peters, E. J., 2012. Advanced Petrophysics: Volume 1: Geology, Porosity, Absolute
826 Permeability, Heterogeneity and Geostatistics. Live Oak Book Co, Austin, Texas.
827 Prêt, D., 2003. Nouvelles méthodes quantitatives de cartographie de la minéralogie et de la
828 porosité dans les matériaux argileux : application aux bentonites compactées des
829 barrières ouvragées. PhD-Thesis, University of Poitiers, (France) (in French).
830 Prêt, D., Sammartino, S., Beaufort, D., Meunier, A., Fialin, M., Michot, L., 2010a. A new
831 method for quantitative petrography based on image processing of chemical elements
832 maps : Part I. Mineral mapping applied to compacted bentonites. American Min. 95,
833 1379-1388.
834 Prêt, D., Sammartino, S., Beaufort, D., Fialin, M., Sardini, P., Cosenza, P., Meunier, A.,
835 2010b, A new method for quantitative petrography based on image processing of
836 chemical elements maps : Part II. Semi-quantitative porosity maps superimposed on
837 mineral maps. American Min. 95, 1389-1398.
838 Pusch, R., 2006. Clays and nuclear waste management. In: Handbook of Clay Science,
839 Editors: Bergaya, F., Theng, B.K.G., Lagaly, G., Developments in Clay Science,
840 Elsevier, 703-716.
841 Revil, A., Cathles, L. M., 1999. Permeability of shaly sands. Water Resour. Res. 35(3), 651-
842 662.
843 Robinet, J.C., Prêt, D., Sardini, P. Coelho, D., 2007. Solute diffusion in Bure argillite at
844 millimeter to micrometer scales: the role of mineral and microstructural
845 heterogeneities, 3rd. Annual Workshop Proceedings 6th EC FP – FUNMIG IP,
846 Edinburgh 26th-29th November 2007.
847 Robinet, J. C. 2008. Minéralogie, porosité et diffusion dans l'argilite du Callovo-Oxfordien de
848 Bure (Meuse/Haute-Marne, France) de l'échelle centimétriquea micrométrique. Ph-D
849 Thesis, University of Poitiers, (France) (in French).

Robinet, J. C., Sardini, P., Coelho, D., Parneix, J.C., Prêt, D., Sammartino, S., Altmann, S., 2012. Effects of mineral distribution at mesoscopic scale on solute diffusion in a clay-rich rock: Example of the Callovo-Oxfordian mudstone (Bure, France). Water Resour. Res. 48(5).

Rolland du Roscoat, S.R., Decain, M., Thibault, X., Geindreau, C., Bloch, J.F., 2007. Estimation of microstructural properties from synchrotron X-ray microtomography and determination of the REV in paper materials. Acta Materialia 55(8), 2841-2850.

Rozenbaum, O., & Rolland du Roscoat, S.R., 2014. Representative elementary volume assessment of three-dimensional x-ray microtomography images of heterogeneous materials: Application to limestones. Physical Review E, 89(5), 053304.

Saraji, S., Piri, M., 2015. The representative sample size in shale oil rocks and nano-scale characterization of transport properties. Intern. J. Coal Geol. 146, 42-54.

Semnani, S.J., Borja, R.I., 2017. Quantifying the heterogeneity of shale through statistical combination of imaging across scales. Acta Geotechnica 12(6), 1193-1205.

Singh, H., Gokhale, A. M., Lieberman, S.I., Tamirisakandala, S., 2008. Image based computations of lineal path probability distributions for microstructure representation. Mater. Sci. Eng. A. 474(1-2), 104-111.

Stineman, R.W., 1980. A consistently well-behaved method of interpolation. Creative Computing.

Song, Y., Davy, C. A., Troadec, D., Blanchenet, A. M., Skoczylas, F., Talandier, J., Robinet, J. C., 2015. Multi-scale pore structure of COx claystone: Towards the prediction of fluid transport. Mar. Petrol. Geol. 65, 63-82.

Torquato S., Stell, G., 1982. Microstructure of two-phase random media. I. The *n*-point probability functions, J. Chem. Phys. 77, 2071.

874 Torquato, S., 2002. Statistical description of microstructures. Ann. Rev. Mater. Res. 32(1),
875 77-111.
876 Ulm, F. J., Delafargue, A., Constantinides, G., 2005. Experimental microporomechanics. In
877 Applied micromechanics of porous materials (pp. 207-288). Springer Vienna.
878 VandenBygaart, A. J., Protz, R., 1999. The representative elementary area (REA) in studies of
879 quantitative soil micromorphology. Geoderma, 89(3-4), 333-346.
880 Wang, P., Jiang, Z., Ji, W., Zhang, C., Yuan, Y., Chen, L., Yin, L., 2016. Heterogeneity of
881 intergranular, intraparticle and organic pores in Longmaxi shale in Sichuan Basin,
882 South China: Evidence from SEM digital images and fractal and multifractal
883 geometries. Mar. Petrol. Geol., 72, 122-138.
884 Widjajakusuma, J., Biswal, B., Hilfer, R., 2003. Quantitative comparison of mean field
885 mixing laws for conductivity and dielectric constants of porous media, Physica A. 318,
886 319-333.
887 Yeong, C.L.Y., Torquato, S., 1998. Reconstructing random media. Phys. Rev. E. 57(1), 495.
888 Zhang, D., Zhang, R., Chen, S., Soll, W.E., 2000. Pore scale study of flow in porous media:
889 Scale dependency, REV, and statistical REV. Geophys. Res. Lett., 27(8), 1195-1198.
890

Table captions

Table 1. Geometrical and mineralogical features of both mineral maps used in this work. Note that mineral contents are the surface content and not the gravimetric content.

Table 2. REA estimates obtained using the classical box-counting (BC) method.

Table 3. Ranges of REA estimates obtained using the statistical (S) approach and microstructural descriptors.

Table 4. REA estimates from the literature.

Table 5. Recapitulation and comparison between methods.

Figure captions

Figure 1. Microstructure of clay rocks at various scales (modified from Ulm et al. 2005).

Figure 2. Extraction of the mineral COx map used in this work (modified from Jorand, 2006).

Figure 3. Extraction of the mineral Toar map used in this work (modified from Fauchille, 2015).

Figure 4. A. Partitioning of the COx map used for the classical box-counting (BC) method. B. Partitioning of the Toar map used for the classical box-counting (BC) method. In both cases, the direction of bedding is indicated.

Figure 5. Estimation of the Representative Elementary Area (REA) size of the COx mineral map using the classical box-counting (BC) method. Evolution of the clay fraction with increasing box size and for six different starting domains. The x-coordinates of the starting domains is 250 μm. The z-coordinates of the starting domains are given in the captions at the top right of figure (see the origin of the system of Cartesian coordinates in Figure 2). The horizontal dashed lines indicate the range [45.4-55.4%] corresponding to $(1 \pm \varepsilon)\overline{\phi}_{COx}$ with $\overline{\phi}_{COx}$=0.504 (50.4%) and ε=0.1(10%).

Figure 6. Estimation and comparison of the Representative Elementary Area (REA) sizes of both mineral maps using the box-counting (BC) method. Evolution of the clay fraction with an increasing box size and for different starting domains. The coordinates of the starting domains are given in the caption box (see the origin of the system of Cartesian coordinates in Figure 3). The horizontal dashed lines indicate the range [62.9-76.9%] corresponding to $(1 \pm \varepsilon)\overline{\phi}_{COx}$ or $(1 \pm \varepsilon)\overline{\phi}_{Toar}$ with $\overline{\phi}_{COx}$=0.504 (50.4%), $\overline{\phi}_{Toar}$=0.699 (69.9%) and ε=0.1(10%).

Figure 7. Statistical (S) approach: evolutions of the mean clay fraction and standard deviation with increasing box size for both maps.

Figure 8. Estimation of the Representative Elementary Area (REA) size of both mineral maps by the S approach. Evolution of the standard deviation of the slope with increasing box size for both maps. The REA size is estimated as the intersection of a smooth curve fit (bold line) and a horizontal line (dashed line) symbolizing a constant evolution. The smooth curve fit corresponds to a Stineman function implemented in the Kaleidagraph software (Stineman, 1980).

Figure 9. Two-point probability function of both mineral maps calculated along the z direction, x-direction and following both directions. The mineral phase considered is the clay fraction. The horizontal dashed lines correspond to the asymptotic values for both maps. The vertical arrows indicate the locations of the L_{REA} estimate.

Figure 10. Two-point probability function of both mineral maps calculated along the z direction, x-direction and following both directions. The mineral phase considered is the nonclay fraction (quartz, carbonates, etc.). The horizontal dashed lines correspond to the asymptotic values for both maps. The vertical arrows indicate the locations of the L_{REA} estimate.

Figure 11. Lineal path function of both mineral maps calculated along the z direction and x-direction. Both mineral phases (clay and nonclay phase) are considered. The horizontal dashed lines correspond to the threshold values of 5% with respect to the total clay fraction. The vertical arrows indicate the locations of the range of L_{REA} estimate with respect to clay.

Figure 12 a. Variograms of six areas extracted from the COx map. b. Variograms of four areas extracted from the Toar map. In both figures, the horizontal dashed lines indicate the clay fraction variance of each extracted areas.

Figure 13. The total fraction of percolating boxes of size L, $p(L)$. For clarity, a polynomial fit is indicated. Considering the COx map, the tangent crossing the $p(L)$ curve at the inflexion point is drawn to use the criterion (2).

Table 1. Geometrical and mineralogical features of both mineral maps used in this work. Note that mineral contents are the surface content and not the gravimetric content.

Map	Resolution (μm)	Total number of pixels	Dimensions (pixels)	Dimensions (μm)	Clay minerals (%)	Tectosilitates (%)	Carbonates (%)
COx	2	384 000	250 x 1536	500 x 3072	50.4	15.1	30.7
Toar	0.625	$16\ 10^6$	4000 x 4000	2500 x 2500	69.9	13.2	14.1

Table 2. REA estimates obtained using the classical box-counting (BC) method.

Map	Area (Ai)	Center of area (Ci)	Asymptotic value (ϕ_∞) (%)	REA estimate (L_{REA}) (µm) ε=0.1 (10%)	REA estimate (L_{REA}) (µm) ε=0.05 (5%)
COx	A1	C1 (x=250µm, z=250µm)	49.0	262	314
	A2	C2 (x=250µm, z=750µm)	48.6	322	438
	A3	C3 (x=250µm, z=1250µm)	49.4	128	170
	A4	C4 (x=250µm, z=1750µm)	51.6	72	86
	A5	C5 (x=250µm, z=2250µm)	52.4	202	310
	A6	C6 (x=250µm, z=2500µm)	51.5	50	86
			Mean: 51.5	Mean: 172.7	Mean: 234.0
			St. Dev.: 1.6	St. Dev.: 108.0	St. Dev.: 142.6
Toar	A1	C1 (x=625µm, z=625µm)	71.4	163	179
	A2	C2 (x=1250µm, z=625µm)	68.5	71	749
	A3	C3 (x=625µm, z=1250µm)	70.2	214	260
	A4	C4 (x=1250µm, z=1250µm)	69.1	68	576
			Mean: 69.8	Mean: 129.0	Mean: 441.0
			St. Dev.: 1.3	St. Dev.: 71.8	St. Dev.: 267.4

Table 3. Ranges of REA estimates obtained using the statistical (S) approach and microstructural descriptors.

Map	Statistical (S) approach	Two-point probability function		Lineal path function (ε=5%)		Variogram range (clay phase)
		Clay phase	Non-clay phase	Clay phase	Non-clay phase	
COx	90-110 µm	16-20 µm	16-19 µm	16-19 µm	16-20 µm	25-53 µm
Toar	130-150 µm	35-43 µm	70-103 µm	31-38 µm	27-33 µm	28-50 µm

Table 4. REA estimates from the literature.

BIB= Broad ion beam; BC= Box-Counting method; EPMA=electron probe microanalyzer; FIB= Focused ion beam; SEM= Scanning electron microscopy; STEM: Scanning transmission electron microscopy XCT= X-ray computed tomography; XRD= X-ray diffraction; XRT= X-ray tomography; 3D-EM= three-dimensional electron microscopy.

Geological setting		Investigation technique	Resolution	REA estimates	REA/REV Methodology				Reference
Formation(s)	Age				2D/ 3D	Method	Phase(s) Target	Error ε (%)	
Posidonia shale	Toarcian formation	Combination of BIB polishing and SEM	10 nm	> 140 µm	2D	BC method	5 mineral phases: calcite, OM, clay matrix, pyrite and others	N/A	Klaver et al., (2012)
Boom clay	Oligocene formation	Combination of BIB polishing, FIB, XCT and SEM	~10 nm	Fine-grained facies: 61-90 µm Coarse-grained facies: 125-295 µm	2D	BC method	Porosity	N/A	Hemes et al. (2013)
				Fine-grained facies: 64-94 µm Coarse-grained facies: 287-453 µm			Non-clay phase		
Opalinus clay	Mesozoic formation	Combination of tomographic methods: FIB and STEM	2-20 nm	~100 µm	3D	Stochastic approach (Kanit et al., 2003)	Porosity	10	Keller et al., (2013)
Opalinus clay	Mesozoic formation	Combination of BIB polishing and SEM	< 5 nm	Sandy facies: 250 µm Shaly facies: 180 µm	2D	BC method + Stochastic approach (Kanit et al., 2003)	8 mineral phases: pyrite, mica, siderite, calcite, quartz, feldspar, fossil shell and clay matrix	10	Houben et al., (2014)
Opalinus clay	Mesozoic formation	Synchrotron XCT	N/A	> 200 µm	3D	Stochastic approach (Kanit et al., 2003)	Clay phase	10	Keller et al., (2015)
Haynesville	Jurassic	Combination of	< 5 nm	> 200 µm	2D	BC method	4 mineral phases:	N/A	Klaver et al.,

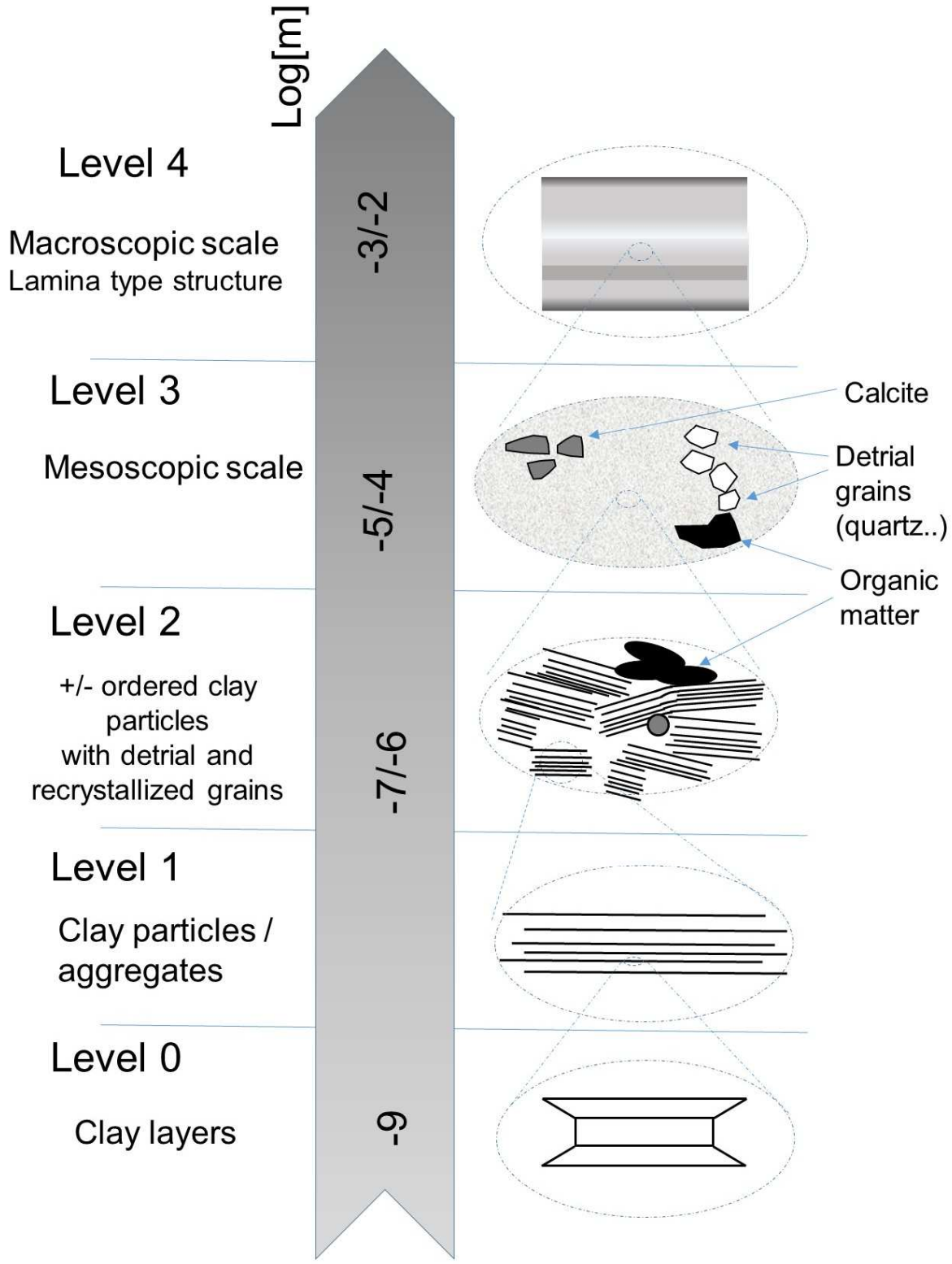
shale / Bossier shale	formation	BIB polishing and SEM					OM, clay matrix, carbonates, pyrite		(2015)
Bakken shale	Mississippian/ Devonian formation	Combination of FIB milling and SEM	10-40 nm	"A few hundred μm"	2D	BC method	Porosity	N/A	Saraji and Piri (2015)
Posidonia shale Whitby mudstone	Toarcian formation	Combination of Ion Beam polishing, SEM and XRD	300 nm	~ 200 μm	2D	BC method	All mineral phases	2	Houben et al., (2016)
Bowland shale	Carboniferous, Namurian formation	Combination XRT imaging techniques and 3D-EM	0,5 μm	380 μm	3D	BC method	Non-clay phase	N/A	Ma et al.,(2016)
Bakken shale	Mississippian/ Devonian formation	SEM imaging	N/A	176 μm	2D	BC method	Porosity	N/A	Liu and Ostadhassan (2017)
Bowland shale	Carboniferous, Namurian formation	SEM imaging	0,5 μm	Organic-rich lamina: 50 μm Gradual organic-poor lamina: 120 μm 600 μm	2D	BC method	Clay phase All mineral phases (except organics and fracture)	10	Fauchille et al. (2018)

971

Table 5. Recapitulation and comparison between methods.

Methods	L_{REA} estimate		Advantages	Limitations
	COx map	Toar map		
Box-counting (CB) method	173 µm (ε=10%) 234 µm (ε=5%)	129 µm (ε=10%) 441 µm (ε=5%)	- Easy to implement	- L_{REA} estimate sensitive to the chosen ε threshold - Statistical homogeneity and isotropy often assumed a priori
Statistical (S) method	90-110 µm	130-150 µm	- Easy to implement - L_{REA} estimate less scattered than that obtained by CB method	- Statistical homogeneity and isotropy assumed a priori
Two-point probability function	16-20 µm (clay phase)	35-43 µm (clay phase)	- Easy to implement - Statistical heterogeneity and anisotropy easily evidenced	- Lower bound of L_{REA}
Lineal path function	16-19 µm (clay phase) (ε=5%)	31-38 µm (clay phase) (ε=5%)	- Anisotropy easily evidenced	- Lower bound of L_{REA} - L_{REA} estimate sensitive to the chosen ε threshold
Variogram	25-53 µm (clay phase)	28-50 µm (clay phase)	- Commercial software packages available	- Difficult to implement - Large numerical files to process

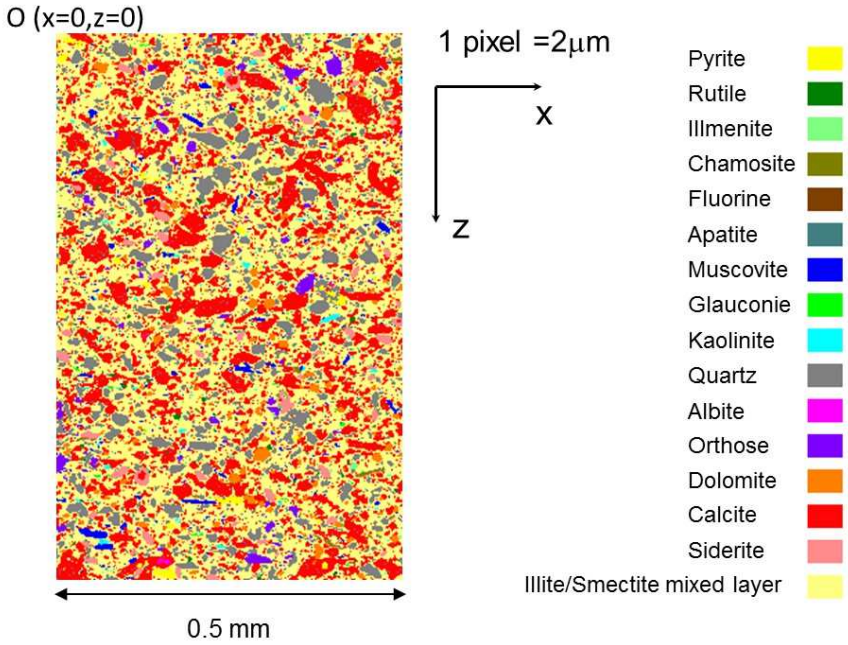
977 Figure 1. Microstructure of shales at various scales (modified from Ulm et al. 2005).



978

979

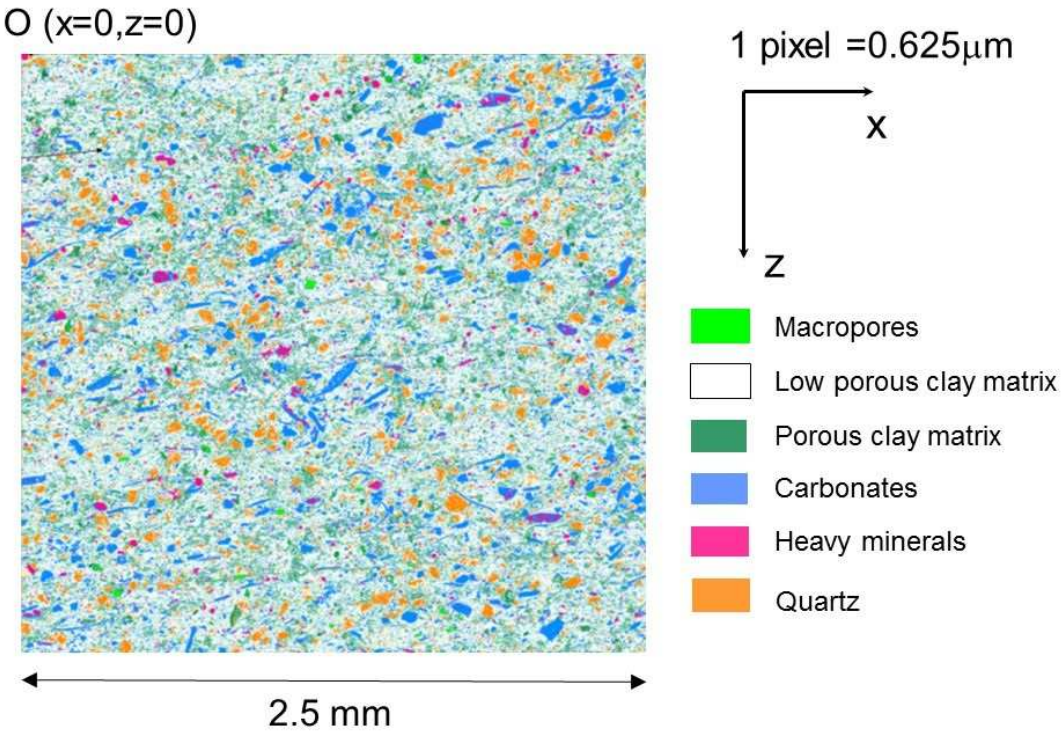
980



981

982 Figure 2. Extraction of the mineral COx map used in this work (modified from Jorand, 2006).

983



984

985 Figure 3. The mineral Toar map used in this work (modified from Fauchille, 2015).

986

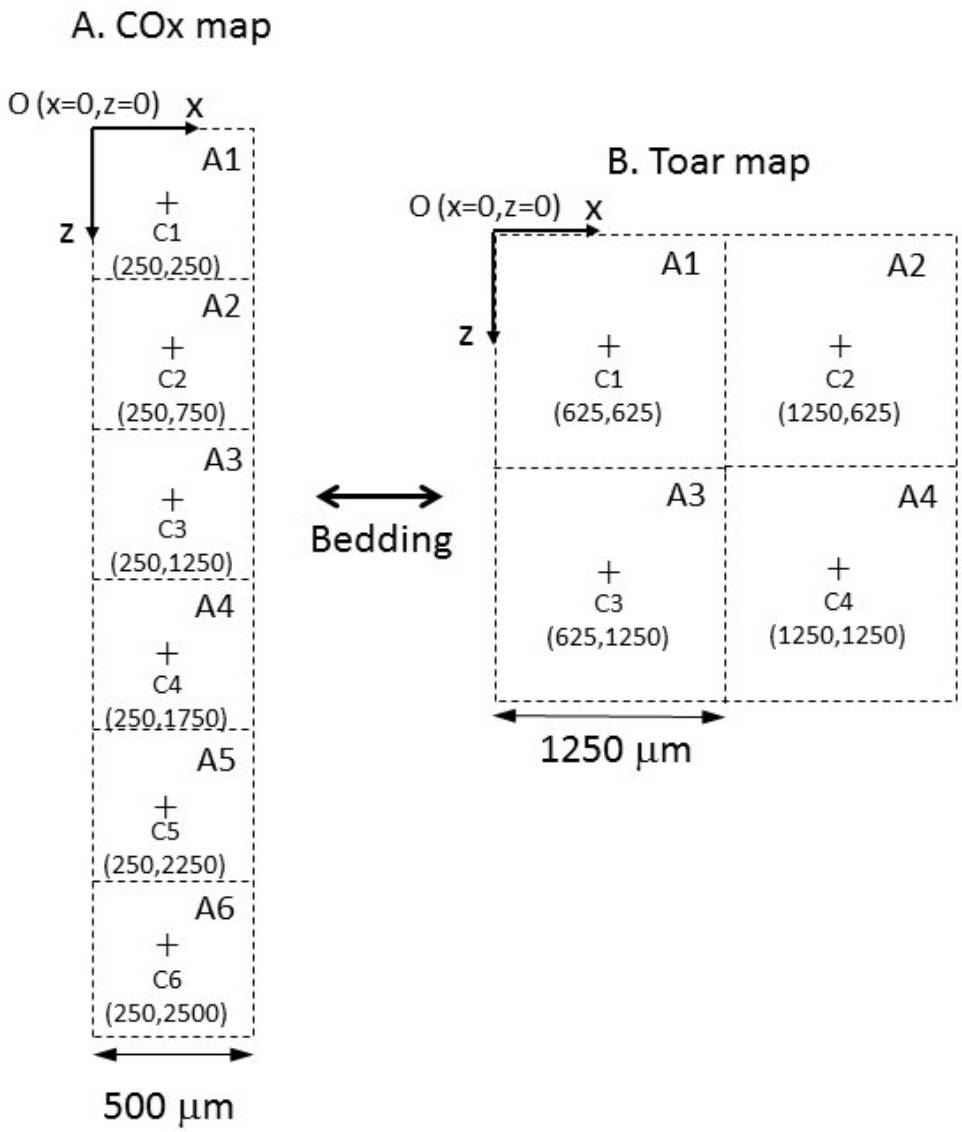
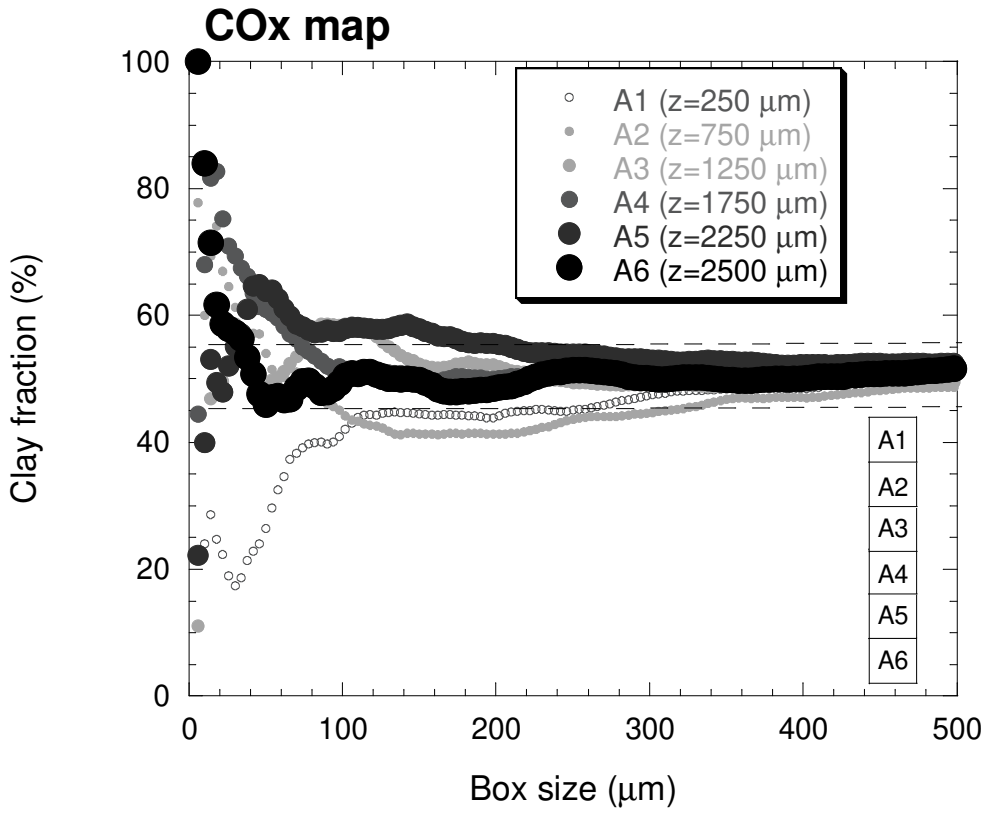


Figure 4. A. Partitioning of the COx map used for box-counting (BC) method. B. Partitioning of the Toar map used for counting box (CB) method. In both cases, the direction of bedding is indicated.



994
995

996 Figure 5. Estimation of the Representative Elementary Area (REA) size of the COx mineral
997 map using the classical box-counting (BC) method. Evolution of the clay fraction with
998 increasing box size and for six different starting domains. The x-coordinates of the
999 starting domains is 250 µm. The z-coordinates of the starting domains are given in the
1000 captions at the top right of figure (see the origin of the system of Cartesian coordinates
1001 in Figure 2). The horizontal dashed lines indicate the range [45.4-55.4%]
1002 corresponding to $(1 \pm \varepsilon)\overline{\phi}_{COx}$ with $\overline{\phi}_{COx}$=0.504 (50.4%) and ε=0.1(10%).

1003
1004

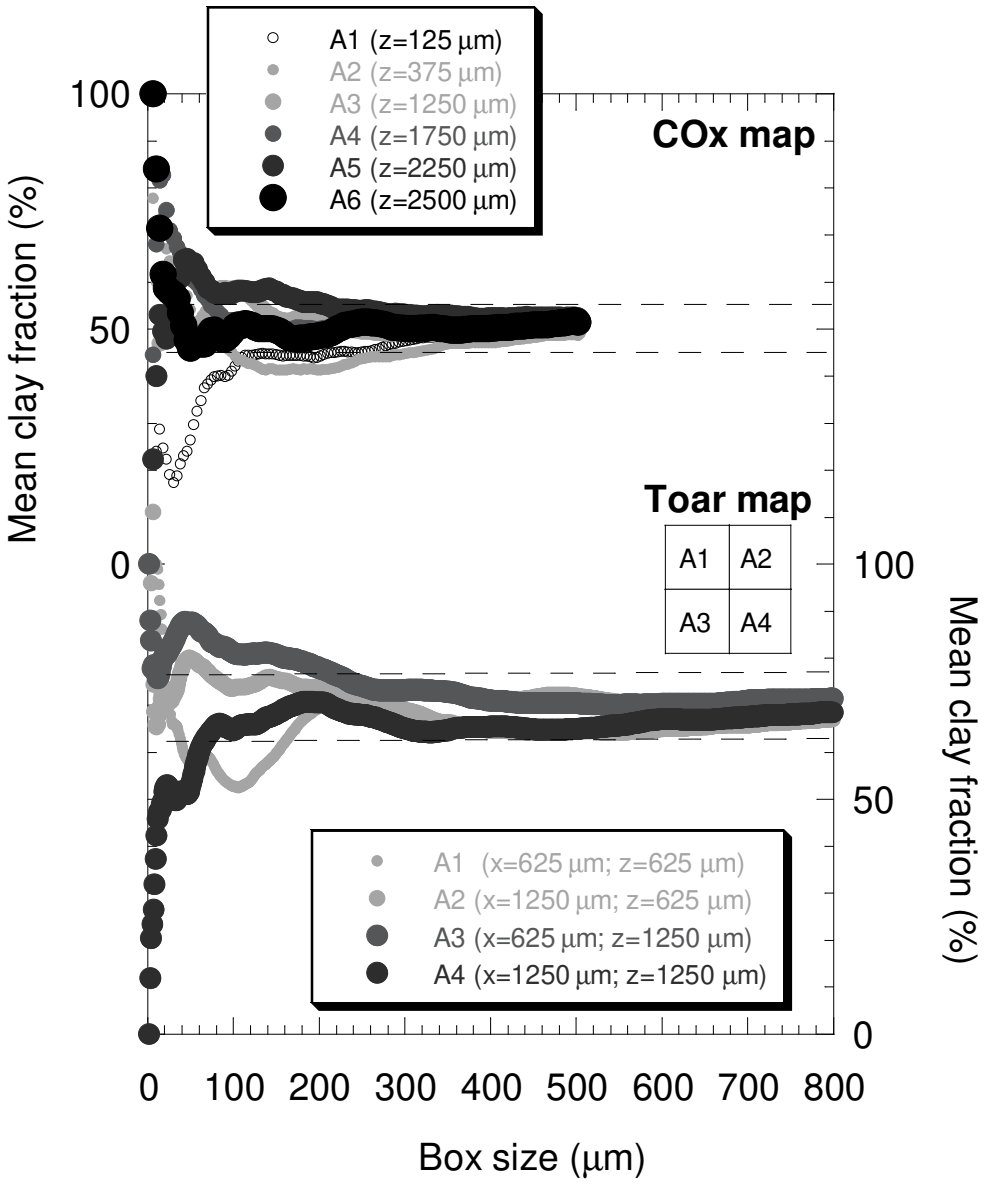
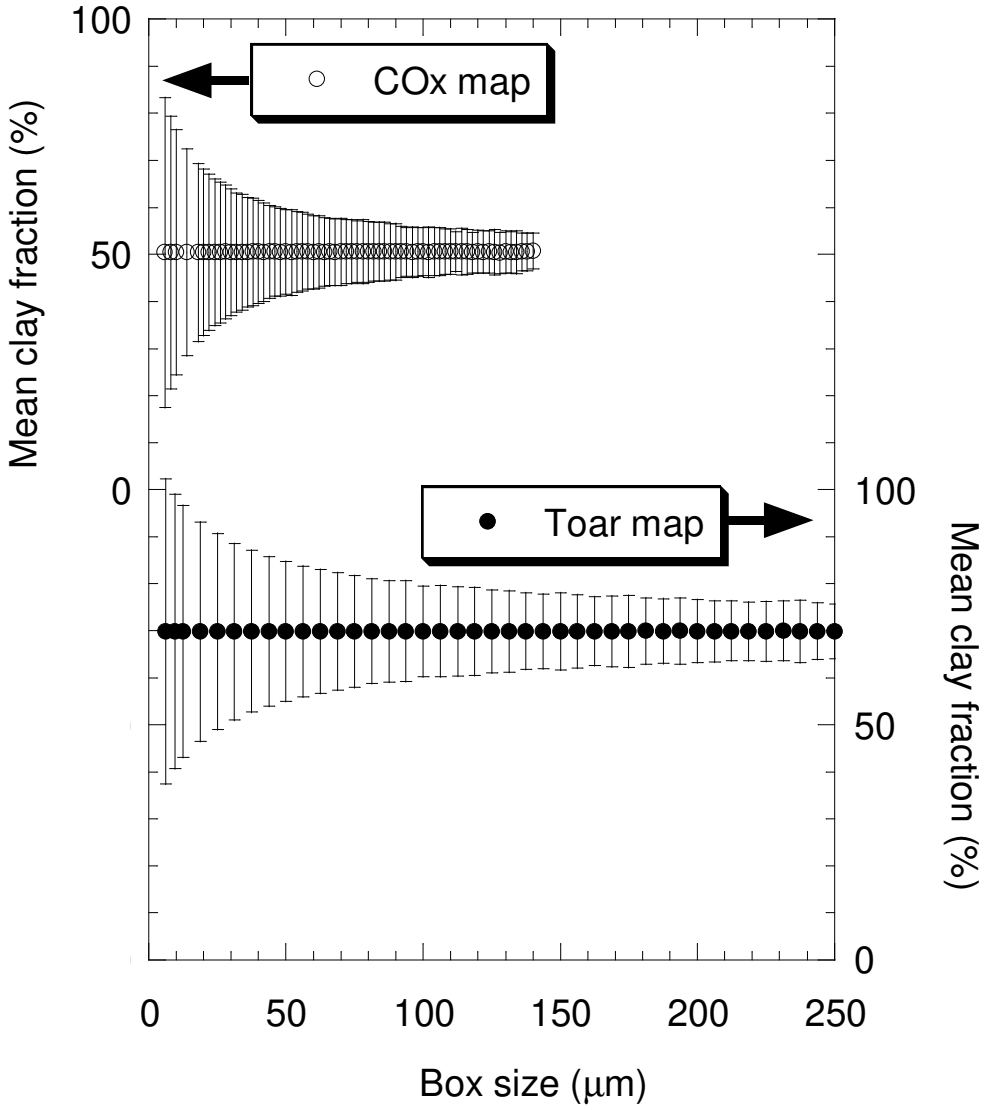


Figure 6. Estimation and comparison of the Representative Elementary Area (REA) sizes of both mineral maps using the box-counting (BC) method. Evolution of the clay fraction with an increasing box size and for different starting domains. The coordinates of the starting domains are given in the caption box (see the origin of the system of Cartesian coordinates in Figure 3). The horizontal dashed lines indicate the range [62.9-76.9%] corresponding to $(1 \pm \varepsilon)\overline{\phi}_{COx}$ or $(1 \pm \varepsilon)\overline{\phi}_{Toar}$ with $\overline{\phi}_{COx}$=0.504 (50.4%), $\overline{\phi}_{Toar}$=0.699 (69.9%) and ε=0.1(10%).

1015



1016

1017 Figure 7. Statistical (S) approach: evolutions of the mean clay fraction and standard deviation
1018 with increasing box size for both maps.

1019

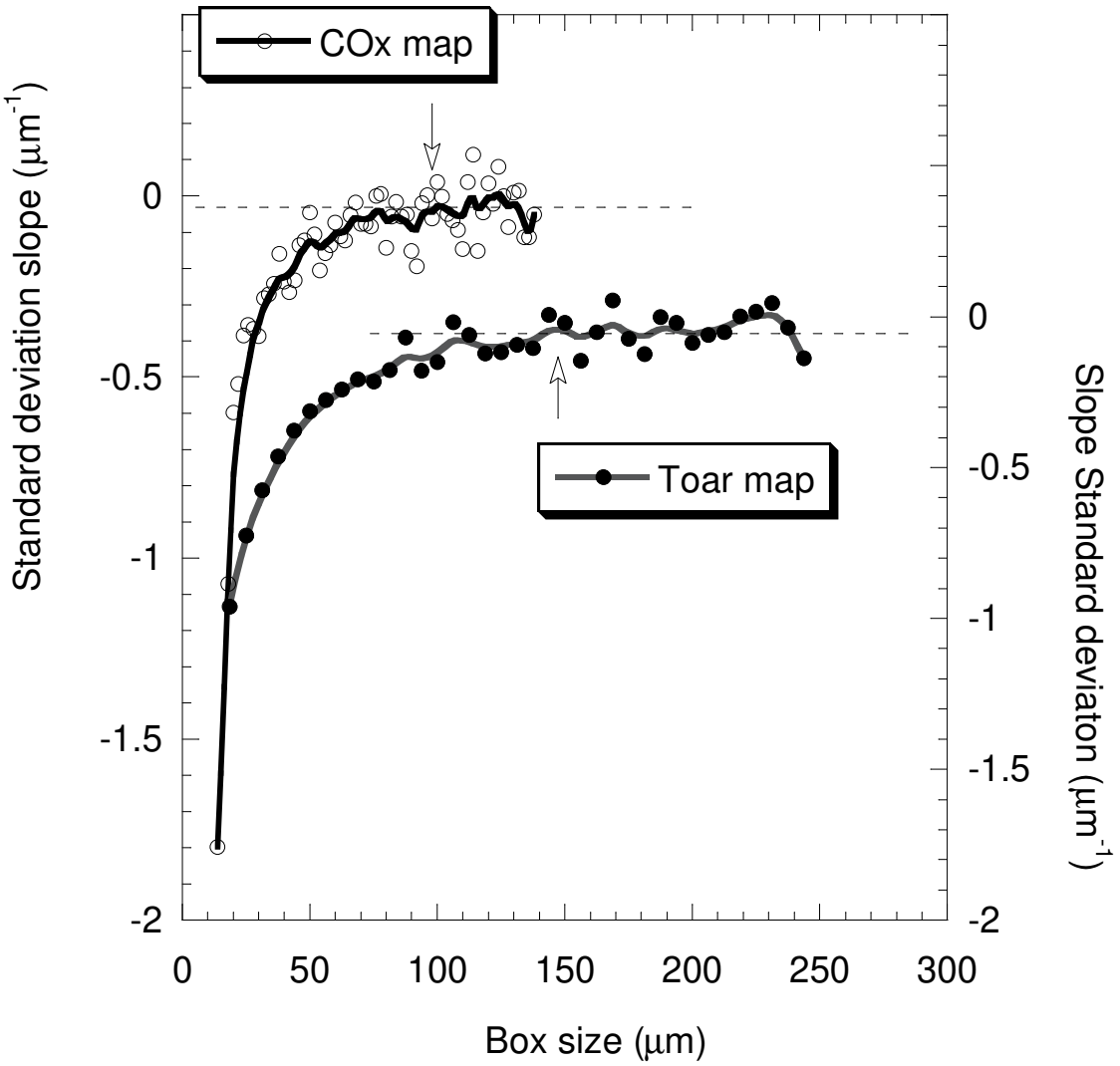


Figure 8. Estimation of the Representative Elementary Area (REA) size of both mineral maps by the S approach. Evolution of the standard deviation of the slope with increasing box size for both maps. The REA size is estimated as the intersection of a smooth curve fit (bold line) and a horizontal line (dashed line) symbolizing a constant evolution. The smooth curve fit corresponds to a Stineman function implemented in the Kaleidagraph software (Stineman, 1980).

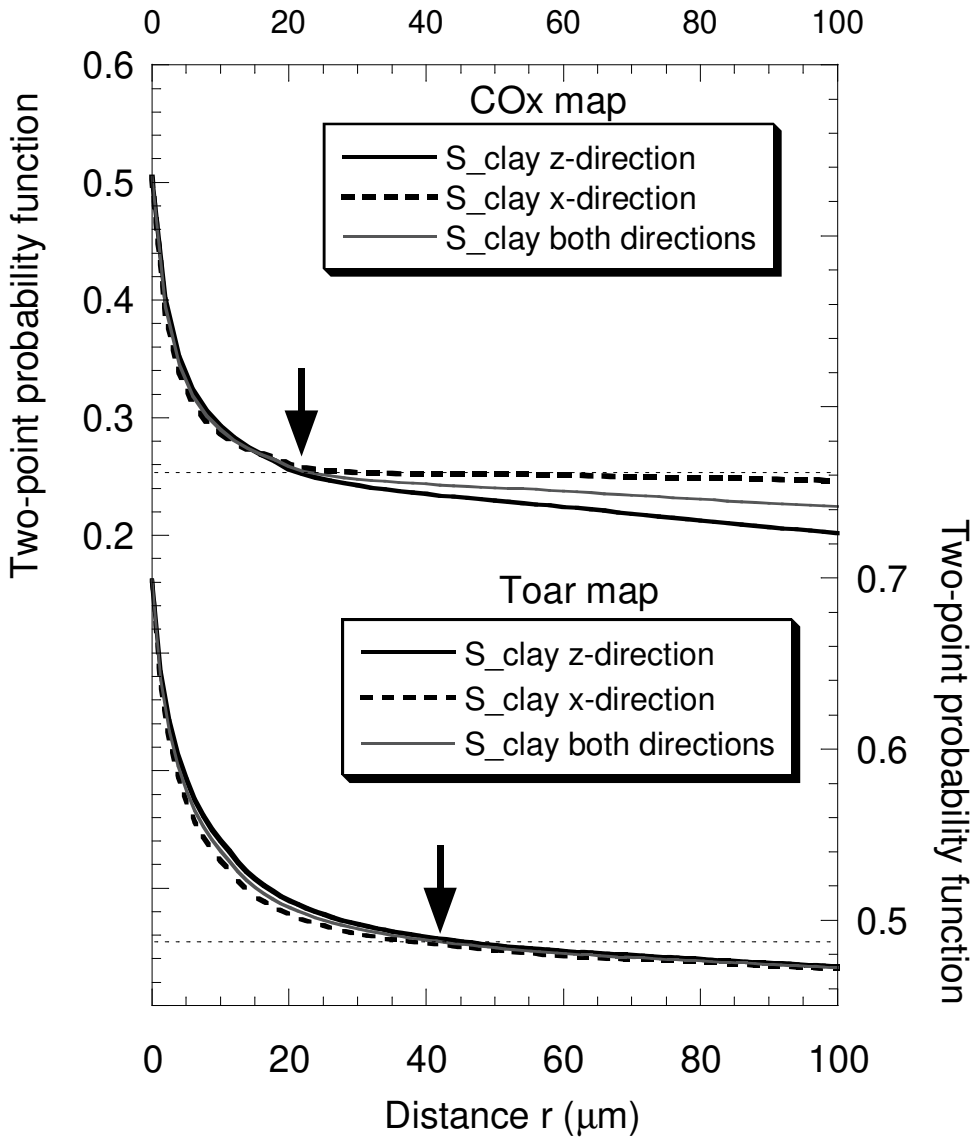


Figure 9. Two-point probability function of both mineral maps calculated along the z direction, x-direction and following both directions. The mineral phase considered is the clay fraction. The horizontal dashed lines correspond to the asymptotic values for both maps. The vertical arrows indicate the locations of the L_{REA} estimate.

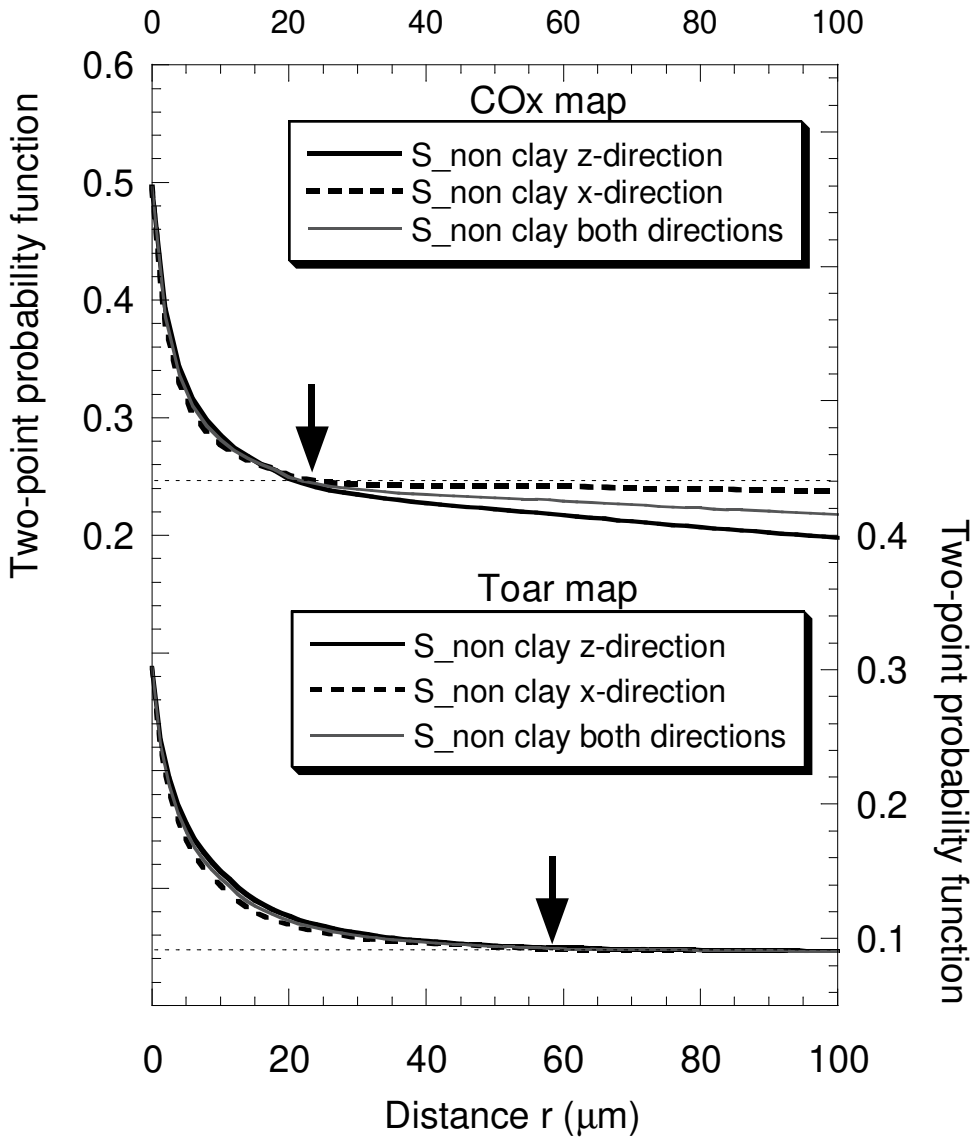


Figure 10. Two-point probability function of both mineral maps calculated along the z direction, x-direction and following both directions. The mineral phase considered is the nonclay fraction (quartz, carbonates, etc.). The horizontal dashed lines correspond to the asymptotic values for both maps. The vertical arrows indicate the locations of the L_{REA} estimate.

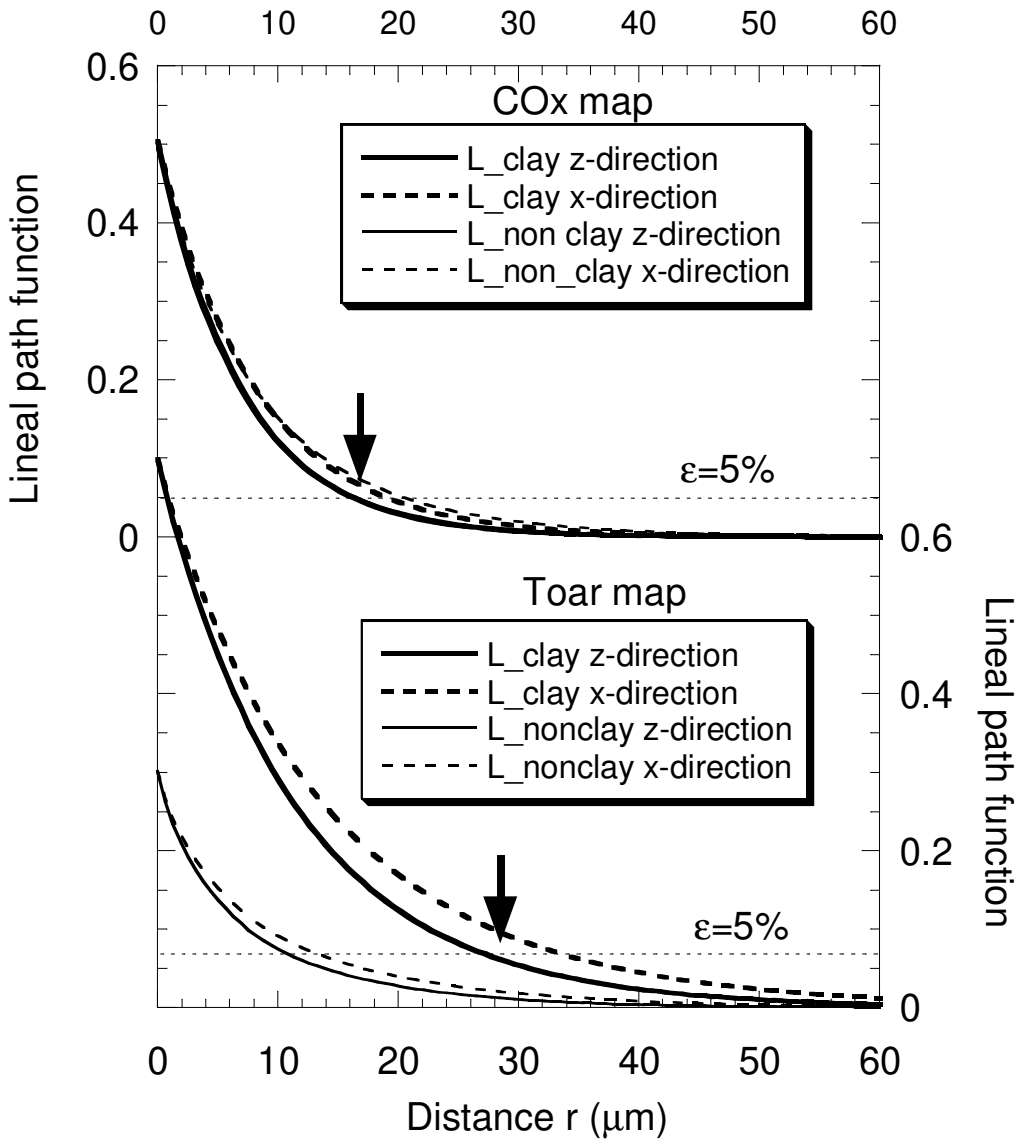
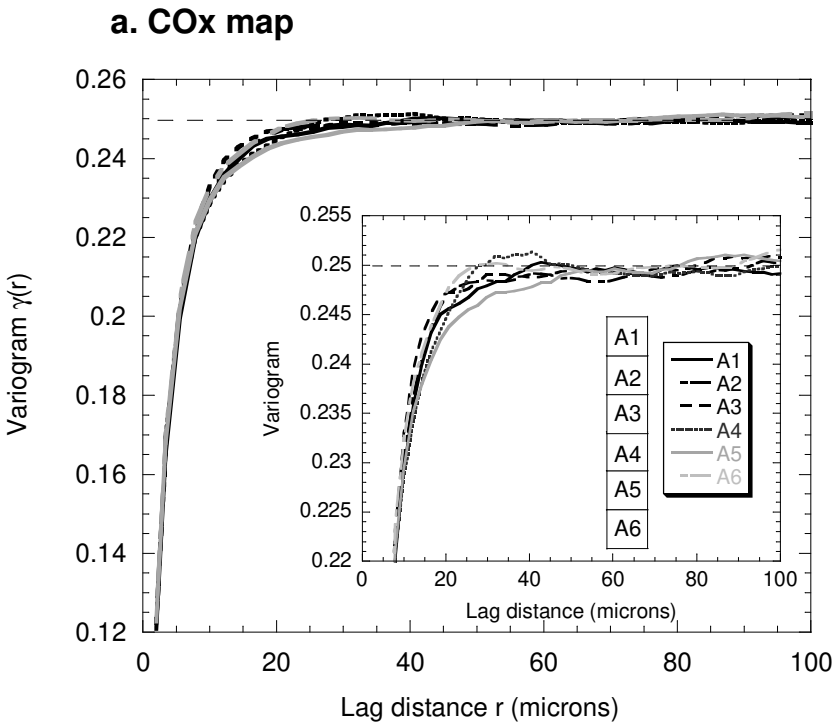


Figure 11. Lineal path function of both mineral maps calculated along the z direction and x-direction. Both mineral phases (clay and nonclay phase) are considered. The horizontal dashed lines correspond to the threshold values of 5% with respect to the total clay fraction. The vertical arrows indicate the locations of the range of L_{REA} estimate with respect to clay.



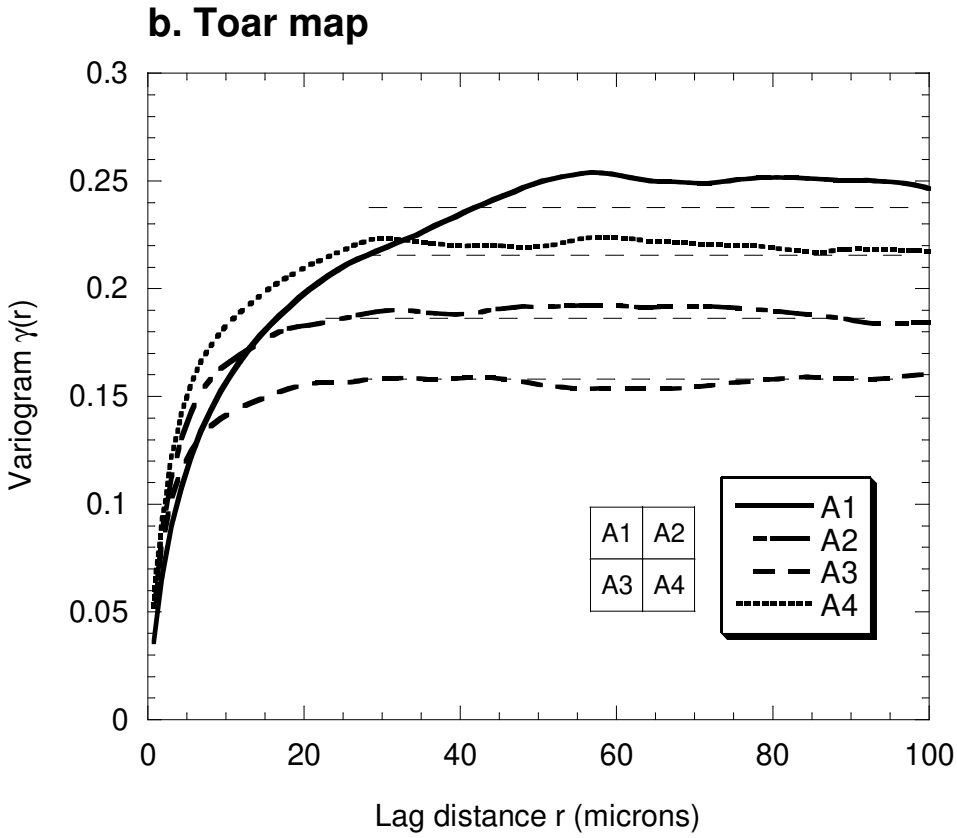


Figure 12 a. Variograms of six areas extracted from the COx map. b. Variograms of four areas extracted from the Toar map. In both figures, the horizontal dashed lines indicate the clay fraction variance of each extracted areas.

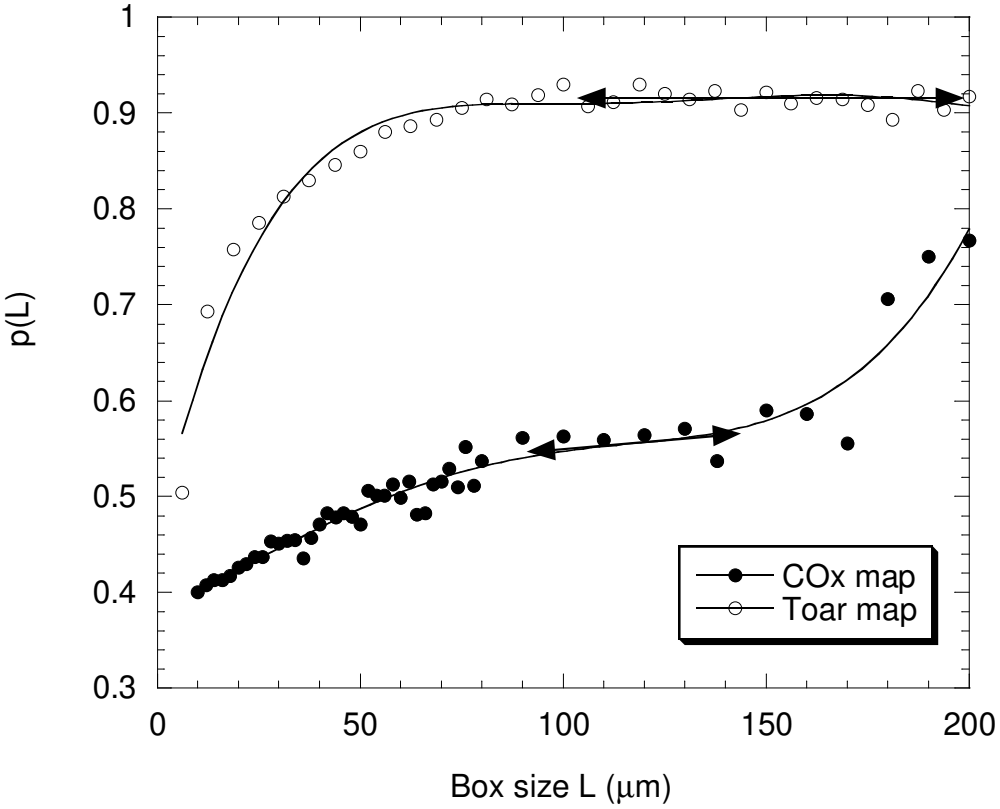


Figure 13. The total fraction of percolating boxes of size *L*, *p(L).* For clarity, a polynomial fit is indicated. Considering the COx map, the tangent crossing the *p(L)* curve at the inflexion point is drawn to use the criterion (2).