



HAL
open science

Representative elementary area of shale at the mesoscopic scale

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

► **To cite this version:**

Philippe Cosenza, Dimitri Prêt, Anne-Laure Fauchille, Stephen Hedan. Representative elementary area of shale at the mesoscopic scale. *International Journal of Coal Geology*, 2019, 216, pp.103316 -. 10.1016/j.coal.2019.103316 . hal-03488609

HAL Id: hal-03488609

<https://hal.science/hal-03488609>

Submitted on 20 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Representative Elementary Area of Shale at the Mesoscopic Scale**

2

3 Philippe COSENZA^{(1)*}, Dimitri PRET⁽¹⁾, Anne-Laure FAUCHILLE⁽²⁾ and Stephen

4 HEDAN⁽¹⁾

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

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

8

9

10

11 ***Corresponding author**

12 Philippe COSENZA

13 Ecole Nationale Supérieure d'Ingénieurs de Poitiers

14 Université de Poitiers-CNRS

15 1 rue Marcel Doré, Bat B1

16 TSA 41105

17 86073 Poitiers cedex 09

18 France

19 Email: philippe.cosenza@univ-poitiers.fr

20

21

22

Intended for publication in International Journal of Coal Geology

23 **Abstract**

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

43

44

45 **Keywords:** representative elementary area, microstructure, shale.

46

47 **1. Introduction**

48 Clay rocks, often called shales, are considered potential host rocks for high-level
49 radioactive waste repositories in several industrialized countries (Callovo-Oxfordian (France),
50 Opalinus Clay (Switzerland), Boom Clay (Belgium)). This interest is mainly explained by the
51 following properties (e.g. Pusch, 2006): (1) because of their high-specific surface area, shales
52 can absorb a significant amount of ions; (2) they have low-hydraulic conductivity values.
53 These particular physicochemical properties are mainly controlled by the type and amount of
54 clay minerals that are present in shale but also by its complex multiscale microstructure (Fig.
55 1). A major part of the literature agrees that the following microstructural levels must be
56 considered to describe the shale texture or microstructure (e.g. Bennett et al., 1991; Ulm et al.,
57 2005; Loucks et al., 2012; Chalmers et al., 2012; Curtis et al., 2012; Han et al., 2017; Ma et
58 al., 2017) (Fig. 1):

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

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

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

77 In the following work, we will focus on the mesoscopic scale for which numerous
78 petrographic studies using advanced imaging techniques allowed to obtain morphological,
79 structural and topological information on shales (e.g., Robinet et al., 2012; Houben et al.,
80 2014; Klaver et al., 2015; Keller, 2015; Fauchille, 2015; Ma et al., 2016; Fauchille et al.,
81 2018). In order to capture all the microstructural features of interest, these imaging techniques
82 produce high resolution images, usually obtained on limited volumes or areas. However, the
83 small size of the field of view reached (typically in the range of a few dozen to a few
84 hundreds of micrometers) may question the representativeness of the petrographic
85 observations. In particular datasets obtained at this scale can be uncertain: are the structural
86 characteristics and petrophysical/numerical properties determined on these small 3D volumes
87 or 2D images representative of the characteristics and properties at the upper scale? In
88 practice, the smallest representative volume or area is identified as the so-called
89 Representative Elementary Volume (REV) (or Representative Elementary Area -REA in 2D),
90 which is required to “separate” the two following space scales: firstly, the scale of
91 heterogeneity, i.e., the distribution of nonclayey grains/clay matrix in our case, and secondly,
92 the scale for which shale is viewed as an “equivalent” continuum medium, i.e., an effective
93 medium. Thus, REV is usually considered a volume of the heterogeneous material that is
94 sufficiently large to be statistically representative of the rock, i.e., to include a relevant
95 sampling of all structural heterogeneities present in the rock at the scale of interest (here the

96 mesoscopic scale). Simultaneously, the REV has to be sufficiently small compared with the
97 scale of the macroscopic geological system of interest, “so that it may be considered as
98 infinitesimal in the mathematical treatment” (Biot, 1941). Another mathematical requirement
99 for the definition of REV is the statistical homogeneity of the volume or the image under
100 study: the REV must be independent of the point of calculation (localization of the volume in
101 3D or 2D space) (e.g., Rozenbaum and Rolland, 2014).

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

113 Other methods use more sophisticated statistical information: the covariance or the
114 two-point probability function (Rolland et al., 2007), the lineal path function (Łydźba and
115 Róžański, 2014) and the percolation length based on a percolation analysis of 2D or 3D
116 microstructures (Hilfer 1991, 1996; Boger et al., 1992; Keller et al., 2013; Cosenza et al.,
117 2015a,b). The variogram, a geostatistical function that was recently used to quantify the
118 microscopic heterogeneity of shale (Gaboreau et al., 2016; Semmani and Borja, 2017), can
119 also provide an interesting tool to infer L_{REV} (L_{REA} in 2D). Most of these statistical descriptors
120 are used to obtain statistical representations or reconstructions of porous media (e.g. Torquato

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

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

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

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

147

148 **2. Materials and Methods**

149 *2.1 Geological setting and mineral maps*

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

152 The first mineral map, hereafter called the COx map, was developed from a sample
153 obtained from Callovo-Oxfordian (COx) claystone, which is extensively studied in the
154 Meuse/Haute-Marne Underground Research Laboratory (MHM-URL) (Eastern France). The
155 thickness of this formation is 130 m, and its age is 150-160 My. The formation is located 420-
156 550 m below the surface, in the eastern part of the Paris Basin (Andra, 2005). The Callovo-
157 Oxfordian formation contains mainly 25 to 65 wt.% clay minerals, with 20-42 wt.% carbonate
158 (calcite, dolomite, ankerite) and 15-31 wt.% tectosilicate (quartz and feldspars) (Andra,
159 2005). This mineral map was prepared from a drill-core, denoted as EST05-709 (-492.2 m)
160 and extracted from the Andra EST205 borehole (Jorand, 2006). It was obtained at micrometer
161 spatial resolution from an advanced image processing of a chemical elements map that was
162 acquired through the use of a Cameca SX100 electron probe microanalyzer (Prêt, 2003). This
163 electron microanalyzer provides quantitative concentration maps of 14 chemical elements (Al,
164 Na, K, Ca, Si, Mg, Ti, Fe, S, Ba, Zr, P, Zn, Sr) on a 3 x 0.5 mm² area with a spatial resolution
165 of 2 µm/pixel. The image processing of these maps is based on mineral identification methods
166 that accommodate mixtures and solid solutions and that are implemented in the in-house
167 µPhaseMap software (Prêt, 2003). For details, the reader is referred to Prêt (2003), Prêt et al.
168 (2010a,b) and Gaboreau et al. (2017). In our case, this methodology allows the spatial
169 localization of all 16 different rock forming minerals, including different clay minerals: illite-

170 smectite mixed layers, kaolinite, mica (glaucanie and muscovite) and chlorite (i.e.,
171 chamosite).

172 The geometrical and mineralogical features of the COx map are presented in Table 1.
173 The surface fractions of clay matrix, carbonates and tectosilicates are 50.4%, 30.7% and
174 15.1%, respectively. Note that the greatest side (1563 pixels, 3072 μm) and smallest side (250
175 pixels, 500 μm) are perpendicular and parallel to the bedding, respectively.

176 The second mineral map, hereafter called the Toar map, has been extracted from a
177 large mosaic of SEM images acquired from a nonimpregnated and dried Tournemire clay rock
178 sample (Fauchille et al., 2014; Fauchille, 2015). The studied Tournemire clay rock sample has
179 been sampled in the horizontal and cylindrical borehole FD90 in the 1996 East gallery of the
180 Tournemire Underground Research Laboratory (URL) of the French Institute for
181 Radioprotection and Nuclear Safety (IRSN). The sample was located at a depth between 4.20
182 to 4.40 meters far from the gallery wall, outside the so-called Excavation Damaged Zone. The
183 Tournemire URL is located in a Mesozoic basin on the southern border of the Massif Central
184 (Aveyron, France), in the subhorizontal consolidated argillaceous Toarcian (Toar) formation
185 (200 meters thick) and marly layers of the Domerian age (50 meters thick). The sample comes
186 from the upper Toarcian formation, the mineralogical composition of which shows that clay
187 minerals represent nearly 25-50 wt% of the rock with illite (10-40 wt%) and illite/smectite
188 mixed-layer minerals (5-25 wt%), kaolinite (10-35 wt%) and chlorite (1-5 wt%). The
189 Tournemire shale also contains 10-40 wt% of carbonates, 10-30 wt% of quartz, 2-7 wt% of
190 sulfides and less than 2 wt% of feldspars (Cabrera et al., 2001).

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

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

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

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

217 *2.2 Calculations of L_{REA}*

218 *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 A_i
231 ($i=1,\dots,6$ for COx map; $i=1,\dots,4$ for Toar map), are associated with starting domains that are
232 defined and discriminated by the coordinates of their center C_i ($i=1,\dots,6$ for COx map; $i=1,\dots,4$
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

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

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

254 *c Percolation length and connectivity*

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

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

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

$$268 \quad p(L) = \int_0^1 \mu(\phi, L) \lambda(\phi, L) d\phi \quad (1)$$

269 where ϕ is the local clay fraction measured in each box of size L and $\mu(\phi, L)$ and
 270 $\lambda(\phi, L)$ are the local clay fraction distribution and the local percolation probability,
 271 respectively. In practice, the local clay fraction distribution $\mu(\phi, L)$ is the frequency
 272 histogram of boxes of size L , having a local clay fraction ϕ . The local percolation probability
 273 $\lambda(\phi, L)$ is the fraction of boxes with a local clay fraction ϕ and side length L that allow
 274 percolation in the x and z directions. A box percolates in the x- (y-, z-, resp.) direction if there
 275 exists a path inside the clay phase connecting two faces of the measurement cell that are
 276 perpendicular to the x- (y-, z-, resp.) axis. This box is called a percolating box in the x- (y-, z-,
 277 resp.) direction (see appendix, Hilfer, 1991, 1996, Keller et al., 2013, Cosenza et al., 2015a,b,
 278 for the details of the calculations).

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

$$283 \quad \left(\frac{d^2 p}{dL^2} \right)_{L=L_p} = 0 \quad (2)$$

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

293 *d Two-point probability function*

294 The two-point probability function, often named the covariance function, is usually is
 295 defined from the autocorrelation function of a phase i given by (e.g. Yeong and Torquato,
 296 1998):

$$297 \quad S_i(\mathbf{r}_1, \mathbf{r}_2) = \langle I_i(\mathbf{r}_1), I_i(\mathbf{r}_2) \rangle \quad (3)$$

298 where \mathbf{r}_1 and \mathbf{r}_2 are two vectors associated with two arbitrary points in the system,
 299 angular brackets denote an ensemble average, and the characteristic function $1_i(\mathbf{r})$ is defined
 300 as

$$301 \quad 1_i(\mathbf{r}) = \begin{cases} 1, & \text{when } \mathbf{r} \text{ is in phase } i \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

302 The function $S_i(\mathbf{r}_1, \mathbf{r}_2)$ is interpreted as the probability of finding two points at
 303 positions \mathbf{r}_1 and \mathbf{r}_2 both in phase i . When the microstructure is spatially stationary or
 304 statistically homogeneous, the two-point probability function depends only on the distance
 305 $r = |\mathbf{r}_1 - \mathbf{r}_2|$ between two points and, therefore, can be simply expressed as $S_i(r)$ of phase i .
 306 $S_i(r)$ can also reach its maximal value of $\bar{\phi}_i$ (volume fraction of phase i of the whole map) at
 307 $r=0$ and decays with $r \rightarrow \infty$ to the asymptotic value of $\bar{\phi}_i^2$:

$$308 \quad \lim_{r \rightarrow 0} S_i(r) = \bar{\phi}_i \quad (5)$$

$$309 \quad \lim_{r \rightarrow \infty} S_i(r) = \bar{\phi}_i^2 \quad (6)$$

310 If the latter limit in equation (6) is reached before $r \rightarrow \infty$, for instance, for a value $r=R$,
 311 the points of the phase i separated from a distance larger than R are not correlated (Kanit et
 312 al., 2003). This parameter R , often called the covariance range, defines a ‘‘correlation length’’
 313 or a ‘‘characteristic size’’ of heterogeneity (e.g. Rolland et al., 2017). Following this definition,
 314 parameter R can be considered an estimate of the minimum L_{REA} (Łydźba and Róžański,
 315 2014; Fauchille et al., 2018). Note that phase i corresponds herein to the clay or the nonclay
 316 phase.

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

$$327 \quad S_i(r) = \frac{1}{N_x N_y} \sum_{n=1}^{N_y} \sum_{m=1}^{N_x} \frac{M[m,n]M[m,n+r] + M[m,n]M[m+r,n]}{2} \quad r$$

$$328 \quad = 1, 2, \dots \quad (7)$$

329 where the product $N_x N_y$ is the total number of pixels in the image, and r is expressed in the
330 pixels. Here, three remarks have to be formulated. First, the dimensions of the matrix M are
331 $(N_x+r)(N_y+r)$. The elements of $M[m,n]$ for $m>N_x$ and $n>N_y$ are taken equal to zero and do
332 not contribute to the summations in equation (7). Second, it should be noted that function $S_i(r)$
333 can be calculated following a given direction if all r directions are parallel to this direction.
334 Third, L_{REA} estimates from the two-point probability function have been calculated on whole
335 maps and not following the map partitioning defined in Figure 4.

336 *e Lineal path function*

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

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

$$343 \quad L^{(i)}(0) = \bar{\phi}_i \quad (8)$$

$$344 \quad \lim_{r \rightarrow \infty} L^{(i)}(r) = 0 \quad (9)$$

345 Regarding the two-point probability function, phase i corresponds herein to the clay or
 346 the nonclay phase. Note that for statistically homogeneous but anisotropic media, function
 347 $L^{(i)}(r)$ will depend not only on the amplitude of r but also on the orientation of the
 348 corresponding vector \mathbf{r} . The function $L^{(i)}(r)$ can be calculated by using the same procedure
 349 used to infer chord length distributions of a given phase (e.g. Cousin et al., 1996).

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

$$352 \quad \forall r \in [t^*, +\infty[\Rightarrow L^{(i)}(r) \leq t \quad (10)$$

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

357 *f Variogram range*

358 An additional microstructural descriptor can be obtained from the concept of the
 359 variogram that is widely used in geostatistics and recently in analyses of microtomographic
 360 images acquired for shale (Gaboreau et al., 2016; Semmani and Borja, 2017). This
 361 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:

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

366 where r is the lag distance between two measures of Z obtained 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

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

378 where $Var(Z)$ 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.

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

$$389 \quad \gamma(r) = S_i(0) - S_i(r) \quad (13)$$

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

395 In our study, variograms have been calculated using commercial Surfer© software. To
396 calculate a variogram, Surfer© introduces a variogram grid approach instead of the classical
397 pair comparison files. The variogram grid is a polar grid in which Surfer© places and stores
398 all the pairs introduced by the equation (11). In practice, the user defines (a) the angular
399 divisions, i.e., the number of angular division in the polar grid, (b) the radial divisions, i.e.,
400 the number of concentric circles in the grid and (c) the largest separation distance contained in
401 the variogram grid. In our calculations, the angular division and the radial division have been
402 fixed for both maps at 180° and 100, respectively. The largest separation division of areas
403 associated with the COx map and Toar map has been taken equal to 110 pixels (220 μm) and
404 200 pixels (125 μm), respectively. These choices are justified *a posteriori* since these
405 parameters allow us to obtain the variogram ranges, as shown below.

406 **3. Results and Discussion**

407 *3.1 Comparison between COx and Toar maps*

408 *a Box-counting method*

409 Figures 5 and 6 display the evolutions of the clay fraction calculated for increasing
410 box sizes L considering the different starting domains of the COx map and Toar map,
411 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 $\bar{\phi}_{COx}$ 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 $\bar{\phi}_{COx}$ 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, $\bar{\phi}_{Toar}$ 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 $\bar{\phi}_{Toar}$ 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: $\varepsilon = 0.1$
436 (10%) and $\varepsilon = 0.05$ (5%). For each area, the L_{REA} parameter has been identified as the lowest

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

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

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

453 *b Statistical method*

454 The results obtained by the S approach are displayed in Figures 7 and 8. The REA
455 size, L_{REA} following the S approach, has been estimated in Figure 8 from the slope of the
456 standard deviation values displayed in Figure 7. In fact, L_{REA} has been calculated as the
457 intersection of a smooth curve fit (bold line) of the slope standard deviation curve and a
458 horizontal line (dashed line) symbolizing a constant evolution with the box size (see white
459 arrows in Fig. 8). Following Figure 8, the L_{REA} estimates of the COx map and the Toar map
460 are $100 \pm 10 \mu\text{m}$ and $140 \pm 10 \mu\text{m}$, 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

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

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

$$506 \quad \langle d \rangle = \sum_{i=1}^N \left(d_i \frac{f_i S_i}{\sum_{j=1}^N f_j S_j} \right) \quad (14)$$

507 where d_i is the equivalent spherical diameter of grains having area S_i ($d_i = \sqrt{\frac{4}{\pi} S_i}$), N
 508 is the total number of classes of GSD, and f_i is the fraction in number of grains having area S_i .
 509 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

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

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

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

548 *3.2 Comparison with the literature – Towards a practical methodology*

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

558 In addition, Table 4 confirms that the classical BC method is largely used to calculate
559 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 minimize 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 $100 \pm 10 \mu\text{m}$
619 and $140 \pm 10 \mu\text{m}$, 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

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

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

659

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

666 **Appendix. Mathematical definition of $p(L)$**

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

$$669 \quad p(L) = \int_0^1 \mu(\phi, L) \lambda(\phi, L) d\phi \quad (\text{A1})$$

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

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

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

$$687 \quad \phi(K_j) = \frac{V(C \cap K_j)}{V(K_j)} = \frac{1}{M_j} \sum_{\mathbf{r}_i \in K_j} 1_C(\mathbf{r}_i) \quad (\text{A2})$$

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

$$691 \quad 1(\mathbf{r}_i) = \begin{cases} 1 & \text{if } \mathbf{r}_i \in C \\ 0 & \text{otherwise} \end{cases} \quad (\text{A3})$$

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

$$694 \quad \mu(\phi, K) = \frac{1}{M \Delta\phi} \sum_{j=1}^M \sum_{i=1}^k \left(1_{I_i}(\phi) 1_{I_i}(\phi(K_j)) \right) \quad (\text{A4})$$

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

$$698 \quad 1_{I_i}(\phi) = \begin{cases} 1 & \text{if } \phi \in I_i \\ 0 & \text{otherwise} \end{cases} \quad (\text{A5})$$

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

$$702 \quad \mu(\phi, L) = \frac{1}{M \Delta\phi} \sum_{j=1}^M \sum_{i=1}^k \left(1_{I_i}(\phi) 1_{I_i}(\phi(\mathbf{r}_j, L)) \right) \quad (\text{A6})$$

703 The local clay fraction distribution $\mu(\phi, L)$ also has the following physical meaning: it
704 measures the probability of finding the local clay fraction ϕ between ϕ and $\phi + d\phi$ in a
705 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:

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

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

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

712 A measurement cell $K(\mathbf{r}_j, L)$ 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 $\Lambda(\mathbf{r}_j, L)$ 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 corrélation 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 from 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.

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

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

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

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

812 Łydźba, D., Róžański, A., 2014. Microstructure measures and the minimum size of a
813 representative volume element: 2D numerical study. Acta Geophysica 62(5), 1060-
814 1086.

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

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

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

823 Ortega, J. A., Ulm, F. J., Abousleiman, Y., 2007. The effect of the nanogranular nature of
824 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).

850 Robinet, J. C., Sardini, P., Coelho, D., Parneix, J.C., Prêt, D., Sammartino, S., Altmann, S.,
851 2012. Effects of mineral distribution at mesoscopic scale on solute diffusion in a clay-
852 rich rock: Example of the Callovo-Oxfordian mudstone (Bure, France). *Water Resour.*
853 *Res.* 48(5).

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

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

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

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

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

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

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

872 Torquato S., Stell, G., 1982. Microstructure of two-phase random media. I. The n -point
873 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

891 **Table captions**

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

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

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

897 Table 4. REA estimates from the literature.

898 Table 5. Recapitulation and comparison between methods.
899

900 **Figure captions**

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

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

903 Figure 3. Extraction of the mineral Toar map used in this work (modified from Fauchille,
904 2015).

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

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

915 Figure 6. Estimation and comparison of the Representative Elementary Area (REA) sizes of
916 both mineral maps using the box-counting (BC) method. Evolution of the clay fraction
917 with an increasing box size and for different starting domains. The coordinates of the
918 starting domains are given in the caption box (see the origin of the system of Cartesian
919 coordinates in Figure 3). The horizontal dashed lines indicate the range [62.9-76.9%]
920 corresponding to $(1 \pm \varepsilon)\bar{\phi}_{COx}$ or $(1 \pm \varepsilon)\bar{\phi}_{Toar}$ with $\bar{\phi}_{COx}=0.504$ (50.4%),
921 $\bar{\phi}_{Toar}=0.699$ (69.9%) and $\varepsilon=0.1$ (10%).

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

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

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

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

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

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

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

950
951
952

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

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

955

956

957

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

959

Map	Area (A_i)	Center of area (C_i)	Asymptotic value (ϕ_∞) (%)	REA estimate	
				(L_{REA}) (μm) $\varepsilon=0.1$ (10%)	(L_{REA}) (μm) $\varepsilon=0.05$ (5%)
COx	A1	C1 (x=250 μm , z=250 μm)	49.0	262	314
	A2	C2 (x=250 μm , z=750 μm)	48.6	322	438
	A3	C3 (x=250 μm , z=1250 μm)	49.4	128	170
	A4	C4 (x=250 μm , z=1750 μm)	51.6	72	86
	A5	C5 (x=250 μm , z=2250 μm)	52.4	202	310
	A6	C6 (x=250 μm , z=2500 μm)	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=625 μm , z=625 μm)	71.4	163	179
	A2	C2 (x=1250 μm , z=625 μm)	68.5	71	749
	A3	C3 (x=625 μm , z=1250 μm)	70.2	214	260
	A4	C4 (x=1250 μm , z=1250 μm)	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

960

961

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

Map	Statistical (S) approach	Two-point probability function		Lineal path function ($\varepsilon=5\%$)		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

964

965

966
967
968
969
970

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 Fine-grained facies: 64-94 μm Coarse-grained facies: 287-453 μm	2D	BC method	Porosity Non-clay phase	N/A	Hemes et al. (2013)
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)

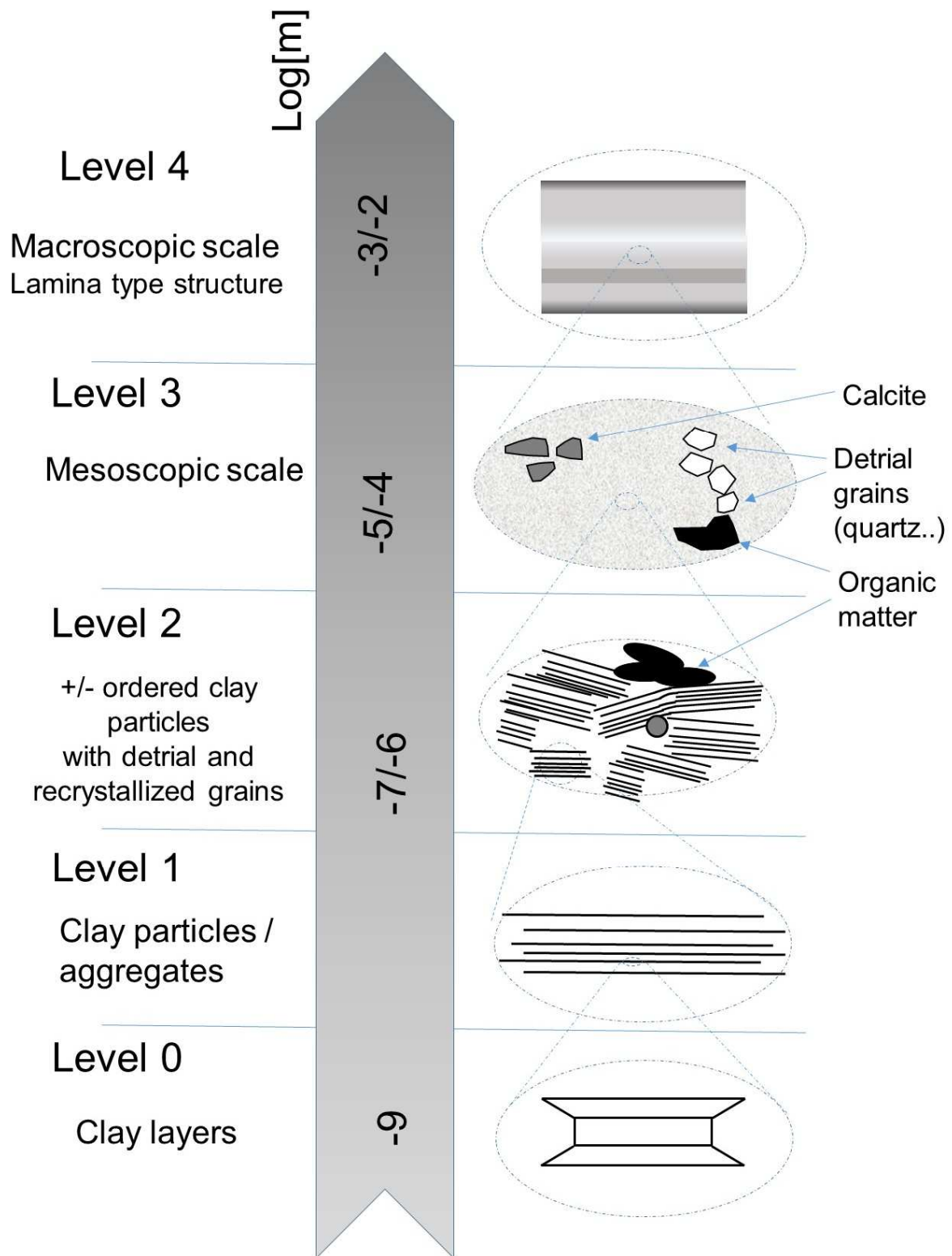
971

972 Table 5. Recapitulation and comparison between methods.
 973

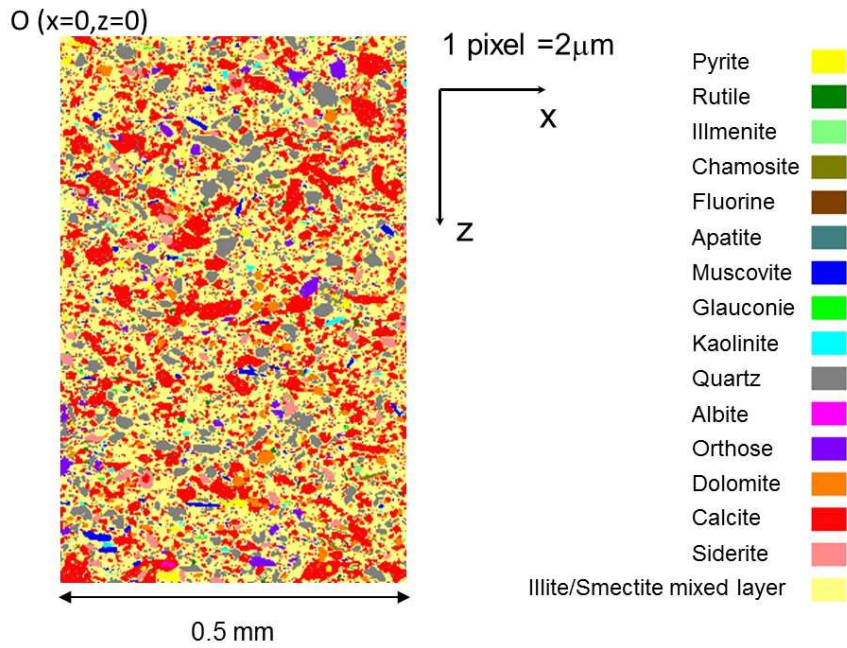
Methods	L_{REA} estimate		Advantages	Limitations
	COx map	Toar map		
Box-counting (CB) method	173 μm ($\epsilon=10\%$)	129 μm ($\epsilon=10\%$)	- Easy to implement	- L_{REA} estimate sensitive to the chosen ϵ threshold - Statistical homogeneity and isotropy often assumed a priori
	234 μm ($\epsilon=5\%$)	441 μm ($\epsilon=5\%$)		
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) ($\epsilon=5\%$)	31-38 μm (clay phase) ($\epsilon=5\%$)	- 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

974
 975
 976

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

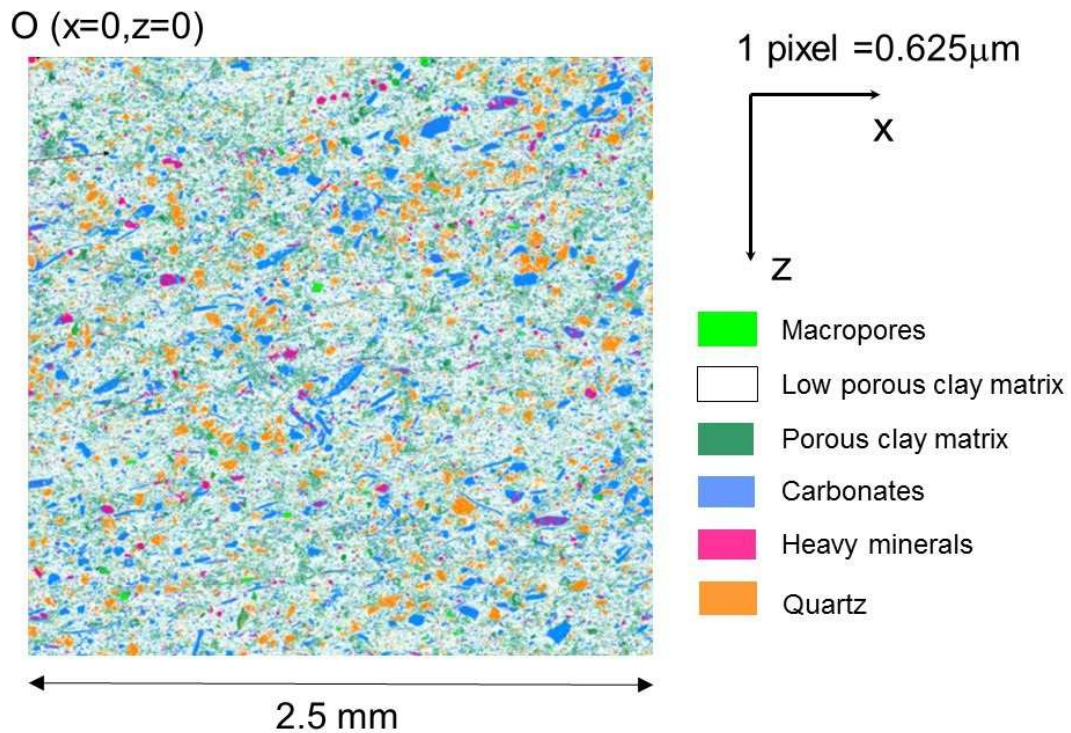


978
979
980



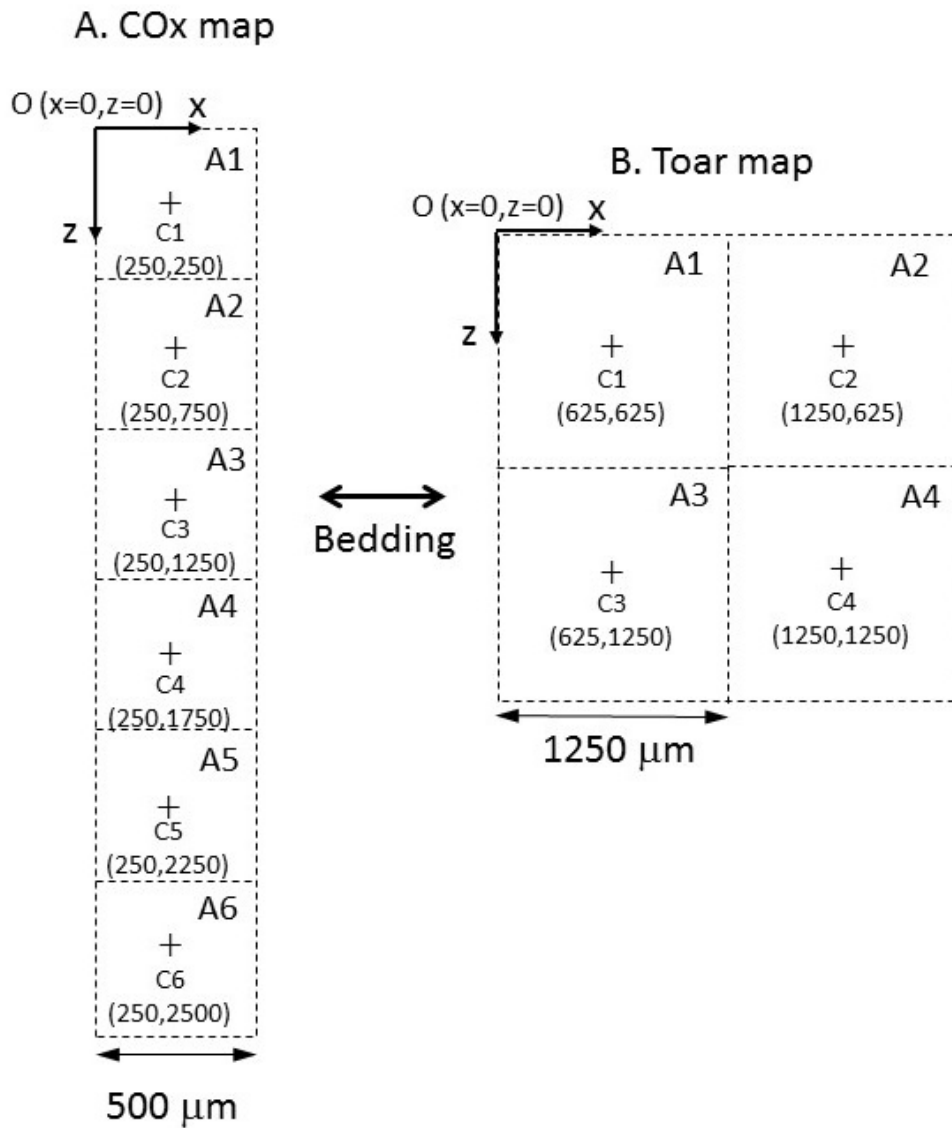
981
982
983

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



984
985
986

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

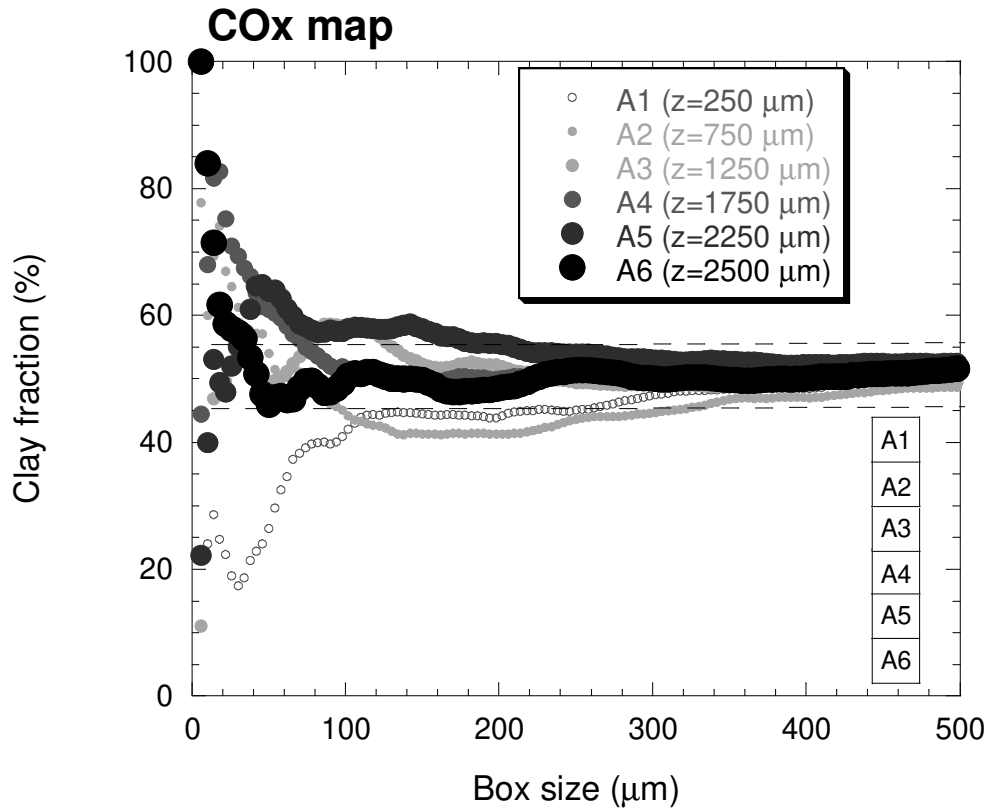


988

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

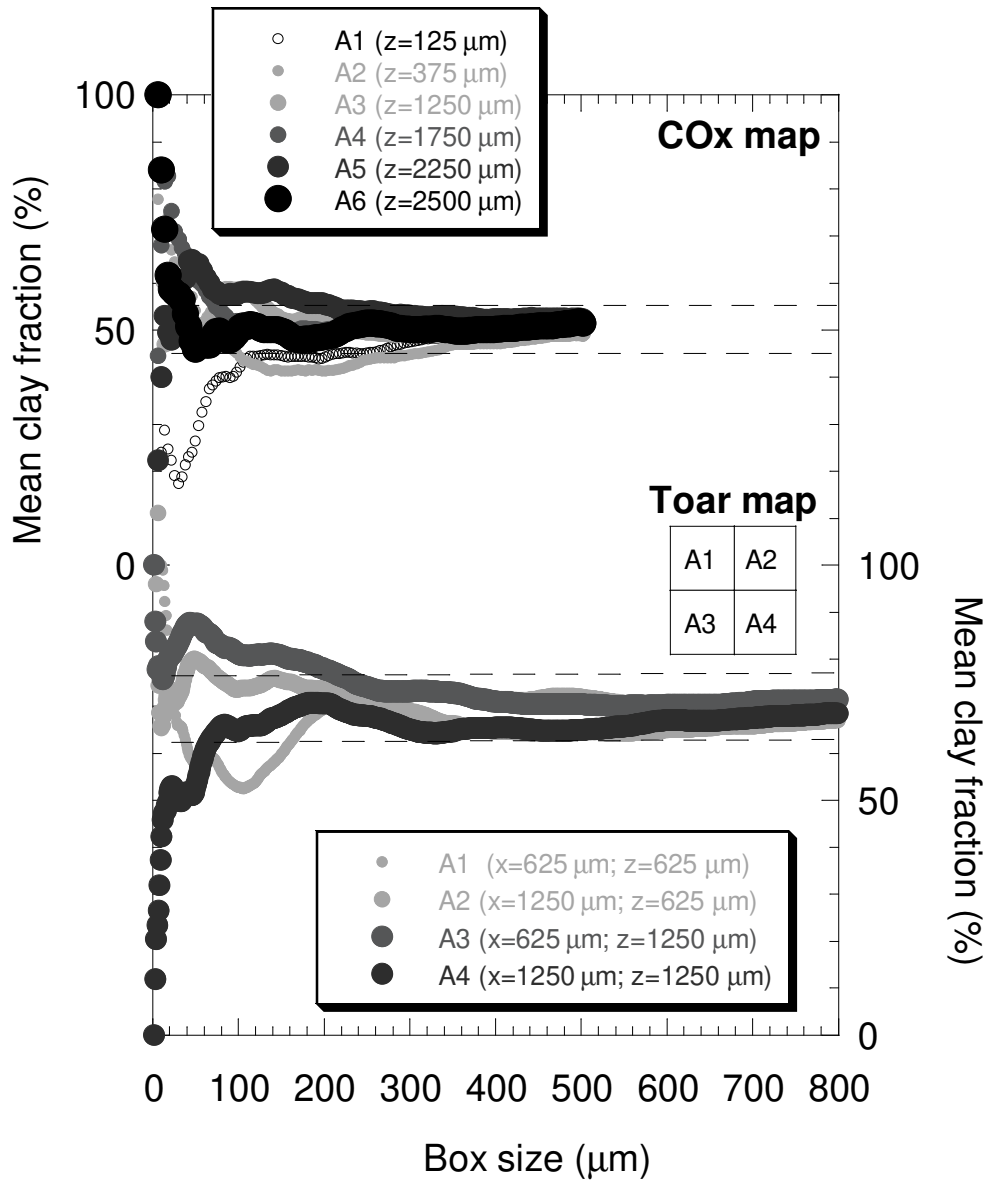
992

993



994
 995
 996
 997
 998
 999
 1000
 1001
 1002
 1003
 1004

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 $(1 \pm \varepsilon)\bar{\phi}_{COx}$ with $\bar{\phi}_{COx}=0.504$ (50.4%) and $\varepsilon=0.1$ (10%).



1005

1006

1007 Figure 6. Estimation and comparison of the Representative Elementary Area (REA) sizes of

1008 both mineral maps using the box-counting (BC) method. Evolution of the clay fraction

1009 with an increasing box size and for different starting domains. The coordinates of the

1010 starting domains are given in the caption box (see the origin of the system of Cartesian

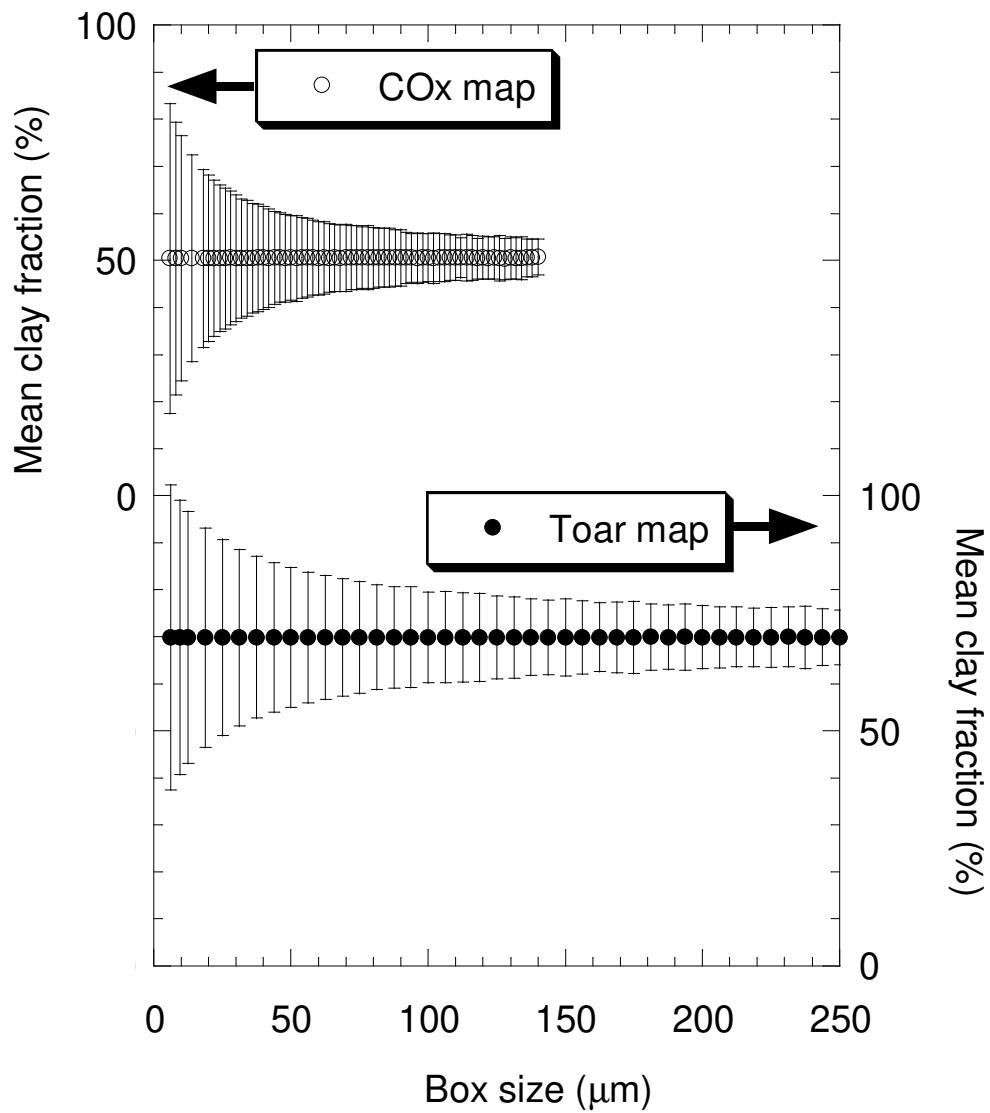
1011 coordinates in Figure 3). The horizontal dashed lines indicate the range [62.9-76.9%]

1012 corresponding to $(1 \pm \varepsilon)\bar{\phi}_{COx}$ or $(1 \pm \varepsilon)\bar{\phi}_{Toar}$ with $\bar{\phi}_{COx}=0.504$ (50.4%),

1013 $\bar{\phi}_{Toar}=0.699$ (69.9%) and $\varepsilon=0.1$ (10%).

1014

1015



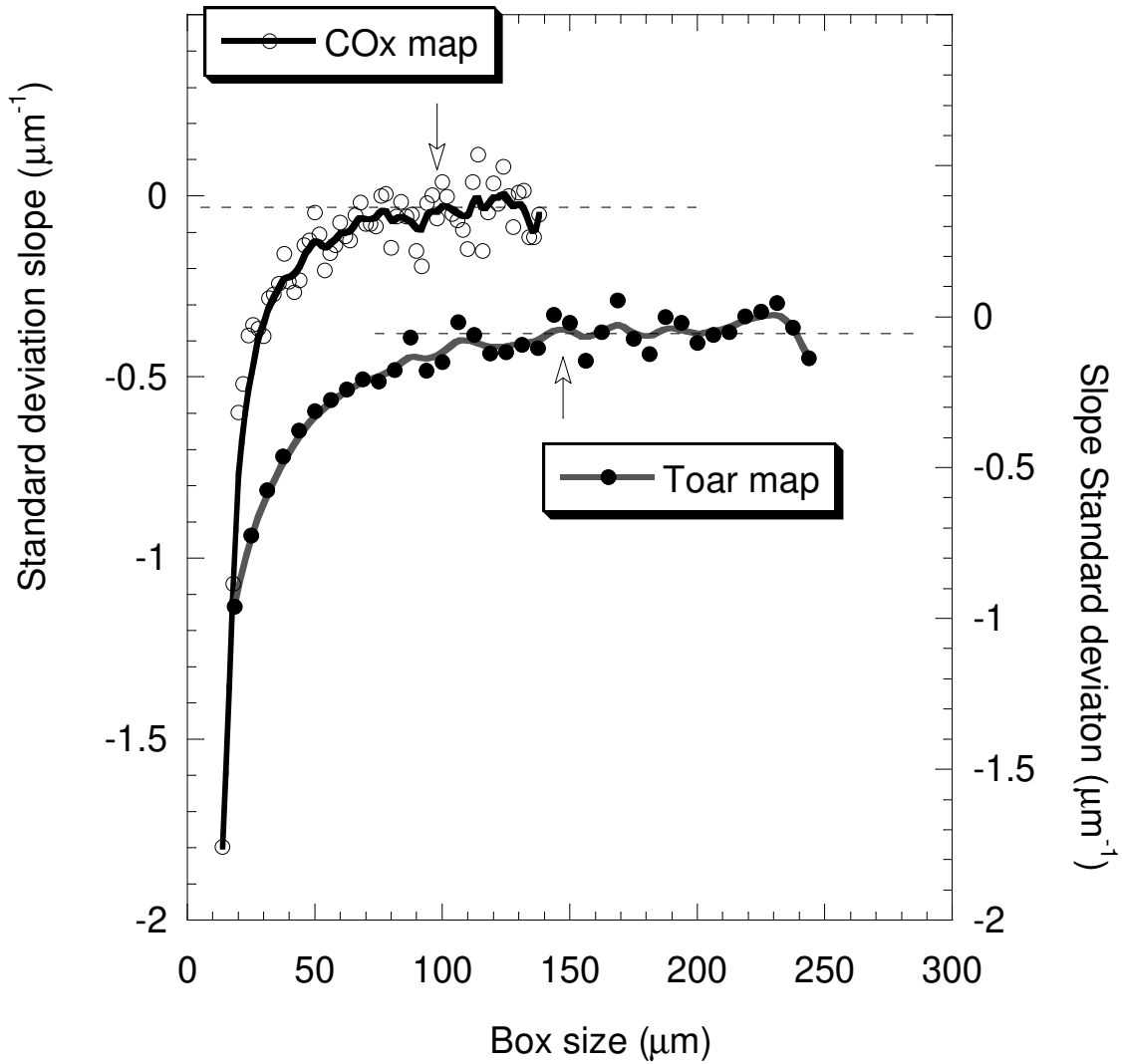
1016

1017 Figure 7. Statistical (S) approach: evolutions of the mean clay fraction and standard deviation

1018

with increasing box size for both maps.

1019

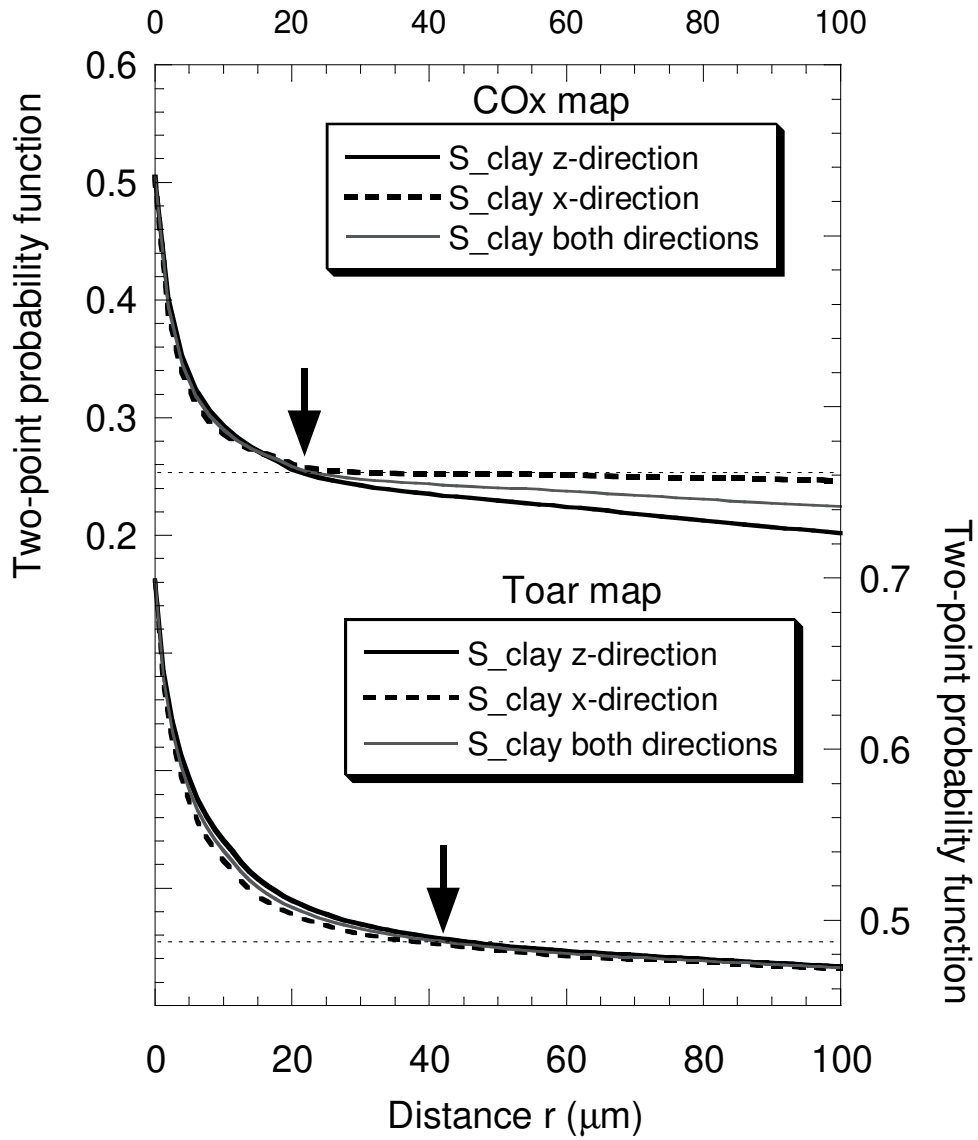


1020

1021 Figure 8. Estimation of the Representative Elementary Area (REA) size of both mineral maps
 1022 by the S approach. Evolution of the standard deviation of the slope with increasing
 1023 box size for both maps. The REA size is estimated as the intersection of a smooth
 1024 curve fit (bold line) and a horizontal line (dashed line) symbolizing a constant
 1025 evolution. The smooth curve fit corresponds to a Stineman function implemented in
 1026 the Kaleidagraph software (Stineman, 1980).

1027

1028

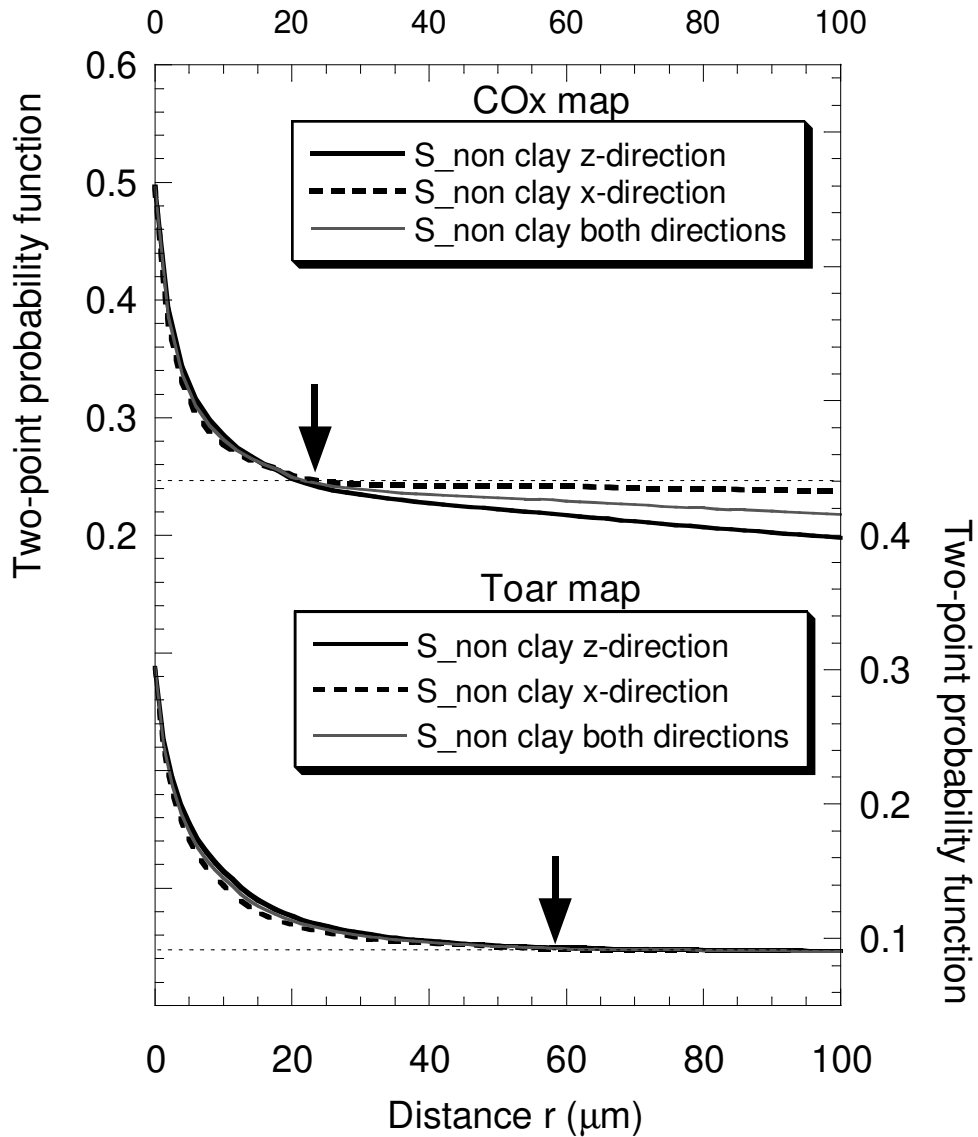


1029

1030 Figure 9. Two-point probability function of both mineral maps calculated along the z
 1031 direction, x-direction and following both directions. The mineral phase considered is
 1032 the clay fraction. The horizontal dashed lines correspond to the asymptotic values for
 1033 both maps. The vertical arrows indicate the locations of the L_{REA} estimate.

1034

1035

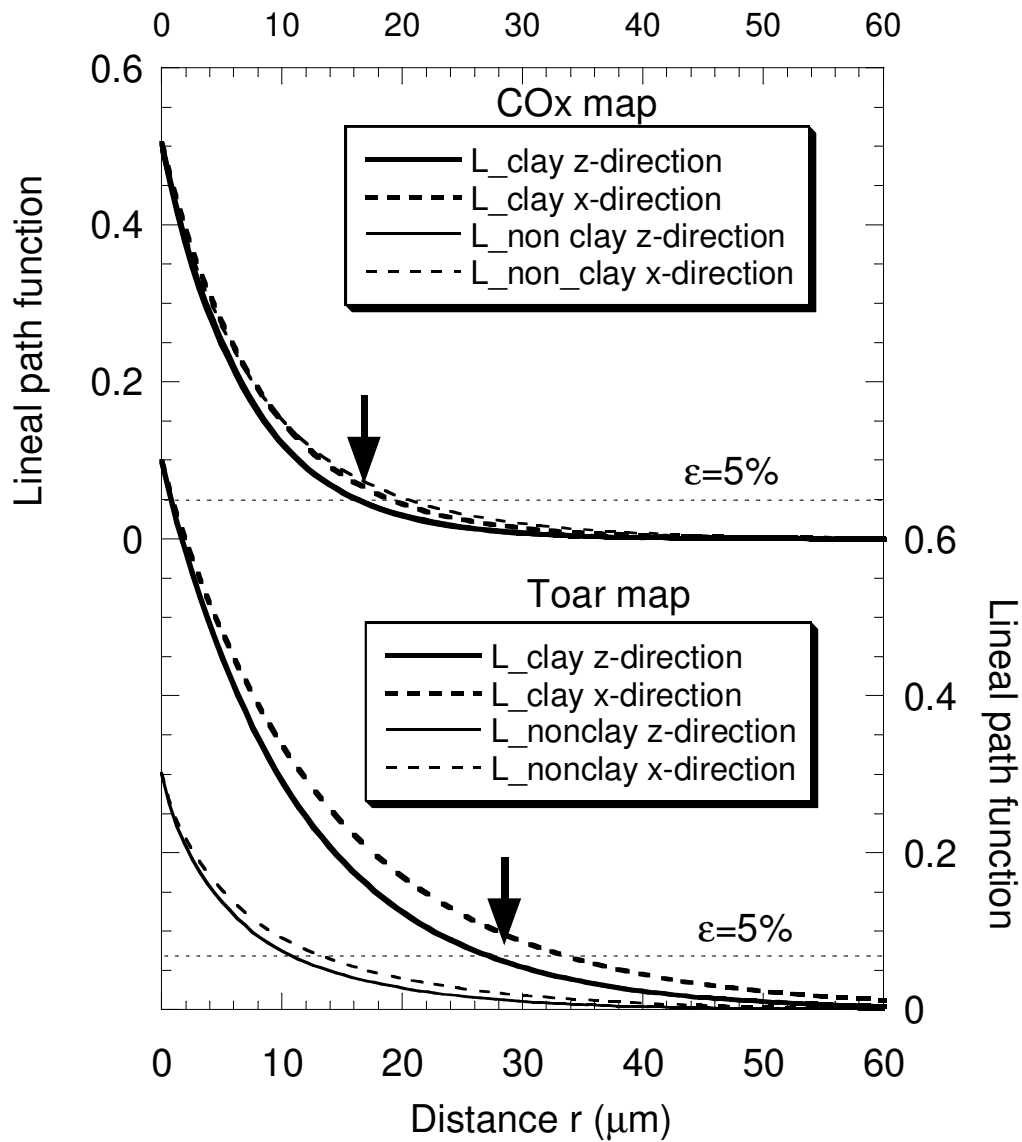


1036

1037 Figure 10. Two-point probability function of both mineral maps calculated along the z
 1038 direction, x-direction and following both directions. The mineral phase considered is
 1039 the nonclay fraction (quartz, carbonates, etc.). The horizontal dashed lines correspond
 1040 to the asymptotic values for both maps. The vertical arrows indicate the locations of
 1041 the L_{REA} estimate.

1042

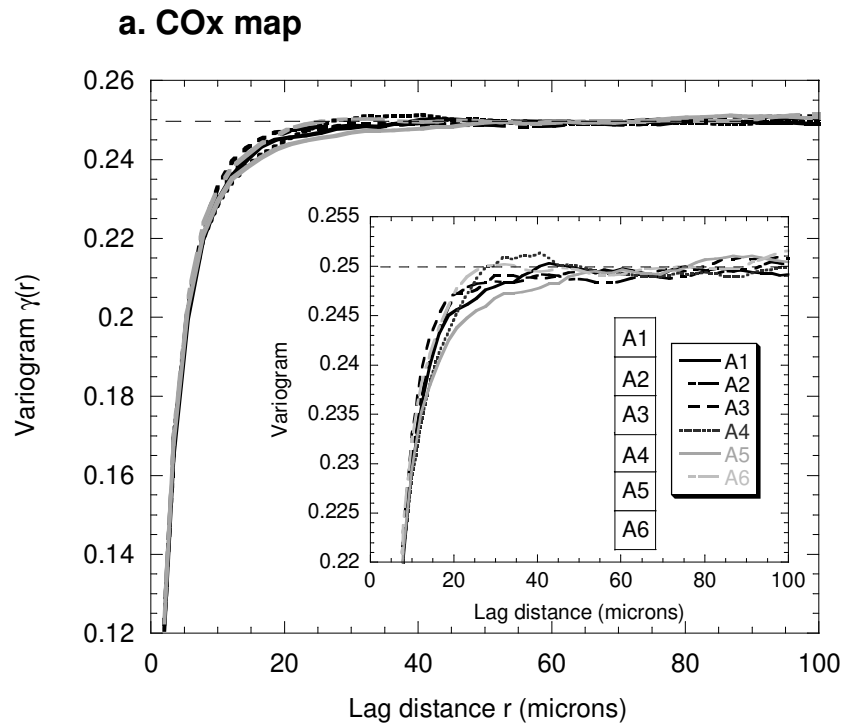
1043



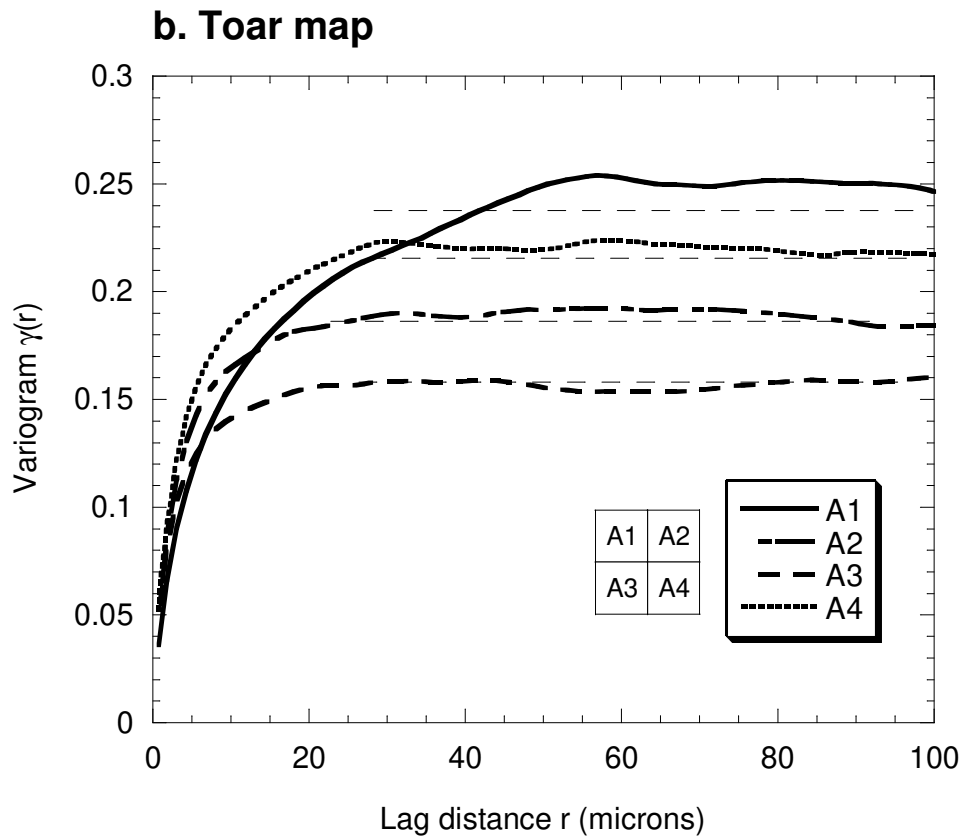
1044

1045 Figure 11. Lineal path function of both mineral maps calculated along the z direction and x-
 1046 direction. Both mineral phases (clay and nonclay phase) are considered. The
 1047 horizontal dashed lines correspond to the threshold values of 5% with respect to the
 1048 total clay fraction. The vertical arrows indicate the locations of the range of L_{REA}
 1049 estimate with respect to clay.

1050

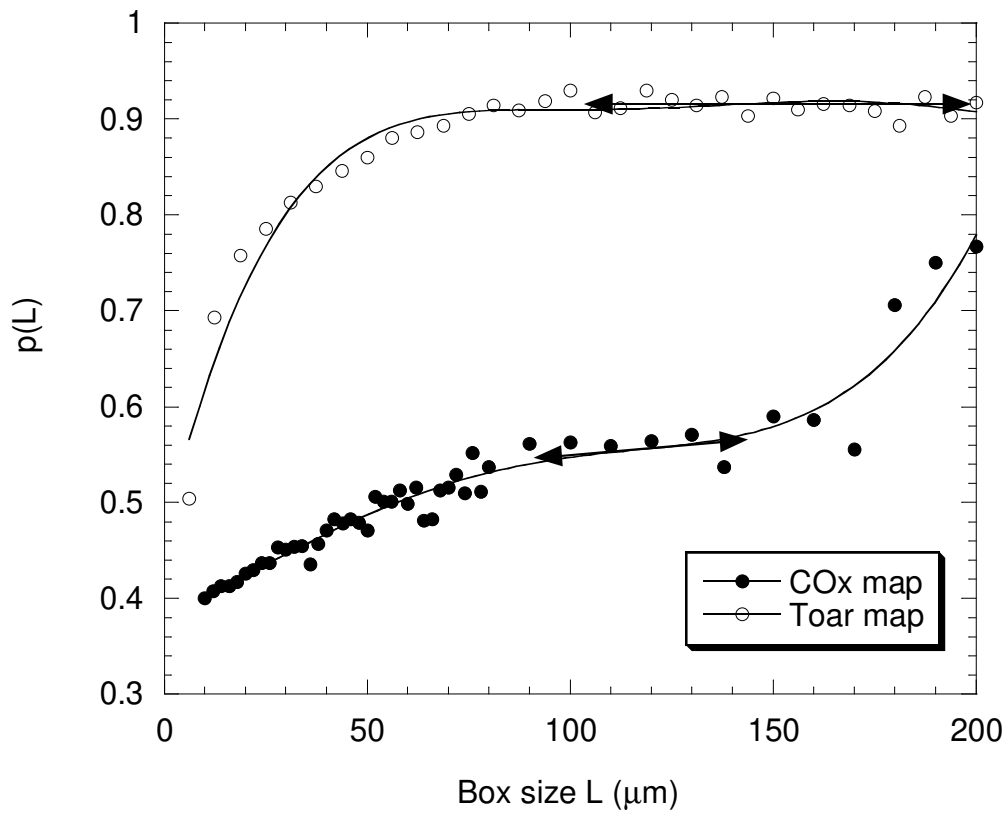


1051
1052



1053
1054
1055
1056
1057
1058

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.



1060

1061

1062 Figure 13. The total fraction of percolating boxes of size L , $p(L)$. For clarity, a polynomial fit
 1063 is indicated. Considering the COx map, the tangent crossing the $p(L)$ curve at the
 1064 inflexion point is drawn to use the criterion (2).

1065

1066