Surface fractal characteristics of preferential flow patterns in field soils: evaluation and effect of image processing

Susumu Ogawa a, Philippe Baveye a,b,*, Charles W. Boast b,1, Jean-Yves Parlange a, Tammo Steenhuis a

a Department of Agricultural and Biological Engineering, Riley-Robb Hall, Cornell University, Ithaca, NY 14853, USA
b Laboratory of Environmental Geophysics, Bradfield Hall, Cornell University, Ithaca, NY 14853, USA

Received 5 December 1997; accepted 28 September 1998

Abstract

In the last few decades, preferential flow has become recognized as a process of great practical significance for the transport of water and contaminants in field soils. Dyes are frequently used to visualize preferential flow pathways, and fractal geometry is increasingly applied to the characterization of these pathways via image analysis, leading to the determination of ‘mass’ and ‘surface’ fractal dimensions. Recent work by the authors has shown the first of these dimensions to be strongly dependent on operator choices related to image resolution, thresholding algorithm, and fractal dimension definition, and to tend asymptotically to 2.0 for decreasing pixel size. A similar analysis is carried out in the present article in the case of the surface fractal dimension of the same stained preferential flow pathway, observed in an orchard soil. The results indicate that when the box-counting, information, and correlation dimensions of the stain pattern are evaluated via non-linear regression, they vary anywhere between 1.31 and 1.64, depending on choices made at different stages in the evaluation. Among the parameters subject to choice, image resolution does not appear to exert a significant influence on dimension estimates. A similar lack of dependency on image resolution is found in the case of a textbook surface fractal, the quadratic von Koch island. These parallel observations suggest that the observed stain pattern exhibits characteristics similar to those of a surface fractal. The high statistical significance ($R > 0.99$) associated with

* Corresponding author. Department of Agricultural and Biological Engineering, Riley-Robb Hall, Cornell University, Ithaca, NY 14853, USA. E-mail: pcb2@cornell.edu
1 Permanent address: Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign, 1102 South Goodwin Avenue, Urbana, IL 61801, USA.

0016-7061/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0016-7061(98)00101-3
each dimension estimate lends further credence to the fractality of the stain pattern. However, when proper attention is given to the fact that the theoretical definition of the surface ‘fractal’ dimension, in any one of its embodiments, involves the passage to a limit, the fractal character of the stain pattern appears more doubtful. Depending on the relative weight given to the available pieces of evidence, one may conclude that the stain pattern is or is not a surface fractal. However, this conundrum may or may not have practical significance. Indeed, whether or not the stain pattern is a surface fractal, the averaging method proposed in the present article to calculate surface dimensions yields relatively robust estimates, in the sense that they are independent of image resolution. These dimensions, even if they are not ‘fractal’, may eventually play an important role in future dynamical theories of preferential flow in field soils. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: mass fractal dimensions; surface fractal dimensions; preferential flow patterns; field soils; quadratic von Koch island

1. Introduction

Preferential flow involves the transport of water and solutes via preferred pathways through a porous medium (Helling and Gish, 1991; Steenhuis et al., 1995). During preferential flow, local wetting fronts may propagate to considerable depths in a soil profile, essentially bypassing the matrix pore space (Beven, 1991). Although the term preferential flow does not imply any particular mechanism, it usually refers to one (or more) of three physically distinct processes: macropore flow, fingering (unstable flow), and funnelled flow.

Macropore flow involves transport through non-capillary cracks or channels within a profile, reflecting soil structure, root decay, or the presence of wormholes, and of ant or termite tunnels. A well-structured soil, for example, has at least two more or less interconnected flow regions for liquids applied at the surface: (1) through the cracks between blocks (interpedal transport), and (2) through the finer pore sequences inside the blocks (inrapedal, or matrix transport). Fingering occurs as a result of wetting front instability. Fingering may cause water and solutes to move in columnar structures through the vadose zone at velocities approaching the saturated hydraulic conductivity (Glass et al., 1988). Fingering may occur for a number of reasons, including changes in hydraulic conductivity with depth and compression of air ahead of the wetting front (Helling and Gish, 1991). Funnelled flow, finally, occurs when sloping geological layers cause pore water to move laterally, accumulating in a low region (Kung, 1990). If the underlying region is coarser-textured than the material above, finger flow may develop.

In a number of studies, the occurrence of preferential flow has been deduced indirectly from the inability of traditional transport equations (e.g., the Richards equation) to predict the outcome of breakthrough experiments in undisturbed soil columns, lysimeters or tile-drained field plots (e.g., Radulovich and Sollins, 1987; Radulovich et al., 1992; McCoy et al., 1994). Various experimental techniques have been used to gain insight into the processes that control...
preferential flow and in particular to identify the soil characteristics (e.g.,
macropores, cracks, etc.) that cause it. Examples of such experimental tech-
niques include X-ray computed tomography (Grevers et al., 1989; Peyton et al.,
1994) or micromorphological analysis of soil thin sections (Grevers et al., 1989;
Aguilar et al., 1990). Most of the studies on preferential flow, however, have
relied on the use of dyes to visualize the preferential flow of water and solutes in
soils, in laboratory experiments or under field conditions (e.g., Bouma and
dekker, 1978; Hatano et al., 1983; Ghodrati and Jury, 1990; Flury et al., 1994;
Flury and Flühler, 1995; Natsch et al., 1996).

Color or black-and-white pictures of dye-stained soil profiles may be ana-
lyzed to provide the percentage of stained areas in vertical or horizontal cuts in
the soil (e.g., Natsch et al., 1996). As useful as the information contained in
these percentages may be to predict the extent and the kinetics of preferential
flow in soils, one would undoubtedly want a more detailed description of the
geometry of stained patterns and some way to relate this geometry to known
morphological features of the soils. In this respect, the close similarity that is
often apparent between these stained patterns and the very intricate details
exhibited by fractals has encouraged a number of researchers to apply the
concepts of fractal geometry to characterize preferential flow pathways. This
approach was pioneered by Hatano et al. (1992) and Hatano and Booltink
(1992). These authors found that the geometry of stained patterns in 2-D images
of soil profiles may be characterized very accurately with two numbers; a
‘surface’ fractal dimension associated with the perimeter of the stained patterns
and a ‘mass’ fractal dimension, relative to the area. The first fractal dimension
varied little among, or with depth within, the five soils tested by Hatano et al.
(1992). However, the mass fractal dimension varied appreciably both among
soils and with depth for a given soil, with a total range extending from 0.59 to
2.0.

The wide range of values assumed by the ‘mass fractal dimension’ in the
work of Hatano et al. (1992) suggests that this parameter may serve as a far
better basis for comparison among soils than the virtually constant surface
fractal dimension. Baveye et al. (1998) have shown, however, that estimates of
the mass fractal dimensions of stained preferential flow patterns in field soils
depend strongly on various subjective (operator-dependent) choices made in the
estimation, in particular on the resolution (pixel size) of the pictures of the
stained soil profiles. When picture resolution is taken into account, the dimen-
sions of stain patterns converge to a value of 2.0 for pixel sizes tending to zero,
indicating that the stain patterns are not mass fractals, and that apparent ‘mass
fractal’ dimensions lower than 2.0 are artefactitious.

At present, no information is available concerning the effects that subjective,
operator-dependent choices that must be made in the analysis of digitized
images may have on the surface ‘fractal’ dimensions of preferential flow
pathways in field soils. Furthermore, the theoretical framework necessary to
interpret the influence of some of these choices is lacking. In the following, (a) we attempt to provide such a conceptual framework, in part through a detailed analysis of the surface fractal characteristics of images of a textbook surface fractal, the quadratic von Koch island, (b) we assess in detail the practical consequences of several of the choices that are made in the evaluation of fractal dimensions of preferential flow pathways, using the same images of a dye-stained soil facies described by Baveye et al. (1998), and (c) we suggest a practical approach for making the estimation of surface fractal dimensions more robust.

2. Theory

2.1. The quadratic von Koch island

The theoretical framework needed for the interpretation of the results reported in this article is probably best described on the basis of a ‘textbook’ surface fractal, the quadratic von Koch island. The iterative algorithm that generates this geometrical fractal is presented in detail in many publications (e.g., in the work of Baveye et al. (1998)), and is illustrated in Fig. 1. By definition, the quadratic

![Figure 1: Illustration of the first steps in the iterative construction of the quadratic von Koch island. The (square) initiator is in the upper left. A first application of the generator leads to the structure in the upper right. Steps two and three correspond to the structures at the bottom left and right, respectively.](image-url)
von Koch island is obtained when the iterative algorithm in Fig. 1 is carried out ad infinitum. Geometrical structures obtained at finite steps in the iteration, like those illustrated in Fig. 1, are termed prefractals.

It is straightforward to show that the quadratic von Koch island has the same area as each one of its prefractals, including the starting square initiator. Also, the perimeter of the $i$-level prefractal is equal to $L_i = 4 \times 8^i \times (1/4)^i$, which diverges as $i \to \infty$. Therefore, it is clear that the perimeter, or coastline, of the quadratic von Koch island is infinitely long.

As a result of these features, the quadratic von Koch island is a surface fractal (we shall adhere to this terminology here, even though it is clear that ‘perimeter fractal’ would be more appropriate). The fractal character of the quadratic von Koch island may be quantified by calculating its similarity dimension, $D_s$. Since the generator of the island consists of eight line segments of length $r = 1/4$, $D_s$ is given by the ratio $-\ln 8/\ln(1/4) = 1.5$, which turns out to be identical to the Hausdorff dimension, $D_H$, of the island (e.g., Feder, 1988; Baveye and Boast, 1998). Both $D_s$ and $D_H$ are often used indiscriminately to denote the surface ‘fractal dimension’ of the island.

Another way to evaluate the surface ‘fractal’ dimension of the quadratic von Koch island involves so-called ‘edge’ squares. The $i$th prefractal of the island may be viewed as an assemblage of squares of side $s_i = (1/4)^i$, with a fraction of squares touching or intersecting with the perimeter (Fig. 2). In selecting the squares that are at the edge, one may decide to include or not to include diagonal squares that touch the perimeter at only one point. This choice can sizably influence the slope of the line representing the number of squares vs. the square side length, in a log–log plot like that of Fig. 3a (where only the first 10 prefractals are considered). Indeed, under these conditions, the absolute value of the slope is higher when diagonal squares are counted (1.549) than when they are not (1.531). However, calculations show that the slope $\frac{\log_{10} N}{\log_{10} (s)}$ in these two cases converges rapidly to a value of 1.5 as $s$ is decreased, whether or not one includes diagonal squares (cf. Fig. 3b).

![Fig. 2. Boundary of a prefractal of the quadratic von Koch island when one excludes (a) or includes (b) diagonal squares that touch the prefractal’s perimeter at only one point.](image-url)
Another method to characterize the dimensionality of random or non-exactly-self-similar sets of points involves the so-called box-counting dimension (e.g., Falconer, 1992), commonly denoted by $D_{BC}$ and defined as

$$D_{BC} = \lim_{\varepsilon \to 0} \frac{\log_{10} N_\varepsilon(F)}{-\log_{10} \varepsilon}$$

(1)

for an arbitrary set $F \subseteq \mathbb{R}^n$. In this definition, $N_\varepsilon(F)$ is the number of ‘boxes’ (i.e., squares, in the present situation) of size $\varepsilon$ needed to completely cover $F$. 

Fig. 3. (a) Logarithmic graph, vs. $\varepsilon$, of the number $N$ of ‘inland’ squares of side length $\varepsilon$ associated with the perimeter of prefractals of the quadratic von Koch island when diagonal squares are included (crosses and solid line) or excluded (open circles and dashed line). The power–law equations in the boxes correspond to the straight lines obtained via non-linear regression. (b) Illustration of the asymptotic behavior of the slope $d \log_{10}(N)/d \log_{10}(\varepsilon)$ vs. $\varepsilon$. The meaning of the symbols is the same as in (a).
The calculation of the number of squares of a given side length \( \varepsilon \) needed to completely cover the quadratic von Koch island is greatly simplified if one takes as values of \( \varepsilon \) the segment lengths of the island’s prefractals, i.e., \( \varepsilon = (1/4)^i \) for \( i = 0, \ldots, \infty \), and if one uses a square grid, four of whose corners coincide with the four corners of the initiator. Such a coverage is illustrated in Fig. 4. When one includes ‘diagonal’ boxes (touching the edge at one and only one point) both inside and outside the island, one finds a value of \( D_{BC} \) equal to 1.488 (\( R = 1 \)). Removal of the diagonal boxes outside the island leads to \( D_{BC} = 1.507 (R = 1) \), whereas removal of all diagonal boxes yields \( D_{BC} = 1.497 (R = 1) \). These different values of \( D_{BC} \) are obtained by starting with a box equal in size to the island’s initiator and by halving the side length 10 times. Regardless of which diagonal boxes are included or excluded, the slope \( d\log_{10}(N)/d\log_{10}(\varepsilon) \) of the \( N(\varepsilon) \) relationship always converges rapidly to a value of 1.5 as \( \varepsilon \) is decreased, a value that is identically equal to \( D_s \) and \( D_H \) (Fig. 5).

2.2. Images, resolution and thresholding

The discussion in Section 2.1 presumed that the evaluation of the ‘fractal’ dimension, using any particular definition, could be carried out on prefractals of a fractal or on the geometric fractal itself. This assumption is valid in the case of

![Fig. 4. Schematic illustration of the coverage of the edge of the quadratic von Koch island with boxes of size \( \varepsilon \) equal to 1/16th the side length of the island’s initiator. Boxes that are in contact with the island’s perimeter at only one point, either inside or outside the island, are not included.](image-url)
Fig. 5. Asymptotic behavior of the slope $d \log_{10}(N)/d \log_{10}(\epsilon)$ vs. $\epsilon$, where $N$ represents the number of boxes of relative side length $\epsilon$ needed to cover the edge of the quadratic von Koch island. Open circles correspond to the case where diagonal boxes are counted both inside and outside. Crosses are for cases where diagonal boxes outside the island are not counted. Finally, full diamonds correspond to the situation of Fig. 4, where diagonal boxes are included neither inside nor outside the island.

mathematical fractals like the quadratic von Koch island, but is never met for real, or natural, fractals such as clouds, river networks or soil samples. These objects do not have ‘prefractals’ and, for physical reasons, it is impossible to obtain a representation of these systems with an infinite level of detail. Any attempt to depict them via, e.g., digitized photographs, radar traces, or tomographic 3-D reconstructions, unavoidably involves some coarse-graining of the features of the original systems.

Usually, at least one aspect of this coarse-graining process is analogous to the application of the box-counting method illustrated in Fig. 4 in that a regular square grid is superimposed on the system. In the application of the box-counting method to the perimeter of the quadratic von Koch island, for example, a given square is tallied if it intersects the perimeter of the island. In a digitized image of the island, contrastedly, a square grid defines the image pixels and the degree of overlap of each pixel with the island determines, via a proportionality rule, the grayscale level associated with the pixel. Grayscale levels customarily range from 0 to 255 (i.e., there are 256 grayscale values in total). In the following, it will be assumed that white and black correspond to grayscale values of 0 and 255, respectively.

A digitized image of the quadratic von Koch island with square pixels of size $\lambda = 1$, and, for convenience, positioned such that one pixel coincides with the initiator of the fractal, has three grayscale levels (Fig. 6a). The outer diagonal pixels touch the island at only one point, and therefore remain completely white (grayscale level = 0). In comparison, the four outer non-diagonal pixels overlap.
Fig. 6. Grayscale images of the quadratic von Koch island with pixels of size (a) $\lambda = 1$, (b) $1/2$ and (c) $1/4$. As with the box size $\varepsilon$ in Figs. 3–5, $\lambda$ is relative to the length of the sides of the square initiator in Fig. 1.

significantly with the von Koch island. The extent of this coverage may be calculated exactly by considering the overlap for successive prefractals of the
von Koch island. The overlap converges to $1/14$ as $i \to \infty$, i.e., for the island itself. This percentage translates into a grayscale level equal to 18 ($255/14$, truncated to an integer). Since the area of the island is one, the central pixel in Fig. 6a must have a grayscale level of $255(1 - 1/14) = 182$. In digitized images at higher resolution (i.e., with smaller pixel size), the number of grayscale levels increases (cf. Fig. 6b and c).

In practical applications of fractal geometry to real systems, images like those of Fig. 6 constitute the starting point of analyses. Available methods for the evaluation of fractal dimensions however require the number of grayscale levels to be reduced to just two; black and white. In other words, the images need to be thresholded. Various automatic algorithms, such as those described briefly in Section 3, could be used to this end. However, given the intrinsic symmetry of the histograms of images of the von Koch island obtained by coarse-graining and the lack of dispersion of grayscale levels outside of certain narrow ranges of values, it is sufficient for the purposes of the present analysis to consider three special cases of thresholding. The first, or high-threshold, consists of considering that any pixel with a grayscale value $< 255$ should become white (grayscale level $= 0$). Alternatively, adopting a low-threshold, one could consider that any pixel with a grayscale level $> 0$ should become black. Between those two extremes, one may take as a medium-threshold the value that splits the grayscale into two equal parts; pixels with grayscale value $\leq 127$ become white, and pixels with grayscale value $> 127$, black. Application of these three approaches to the quadratic von Koch island, coarse-grained with pixels of size $\lambda = 1/256$ (Fig. 7) shows that the resulting binary images have significantly different appearances. The high-threshold (Fig. 7a) yields an image composed of the pixels that are entirely contained within the island. Implicitly, this is the threshold adopted by De Cola and Lam (1993) in their analysis of photographs at different resolutions. The medium-threshold (Fig. 7b) yields a prefractal of the quadratic von Koch island. Finally, the low-threshold (Fig. 7c) produces a fattened version of the island, with a total number of pixels (of size $e$) equal to the number of boxes of size $e$ needed to cover the island. The edge in the high-threshold case appears very prickly, whereas that for the low-threshold is very rounded.

For each of the images of the quadratic von Koch island, with different resolutions and obtained with different thresholds, the relationship between the number of covering boxes and the box size $e$ suggests very convincingly that the image’s representation of the quadratic von Koch island is a surface fractal. For example, the log–log plots associated with the three highest-resolution images are presented in Fig. 8a. For practical reasons, the regression lines in this figure are restricted to the range of box sizes considered by the computer code fd3 (cf. Section 3). In all three cases in Fig. 8a, the $R$ values are remarkably high, which is typical of values found in the present research; none of the ‘fractal’ dimensions reported in this article had an associated $R$ value lower...
than 0.99. The extremely high statistical significance associated with the regression lines in Fig. 8a suggests that, in the three images to which this figure relates, the representation of the quadratic von Koch island is a surface fractal. In other words, the coarse-graining associated with the images and the further approximation caused by thresholding apparently manage to preserve the fractal character of the initial object. This point of view rests implicitly on the premise that the slope of the regression line in log–log plots like that of Fig. 8a is a good estimate of the true box-counting dimension.

By looking at the data points in Fig. 8a from a different perspective, more in line with the definition of the box-counting dimension in Eq. (1), one finds a number of perplexing features, which clearly point in another direction. Instead of fitting a power–law relationship to the observed $N(\varepsilon)$ values, yielding a straight line in a $\log_{10}(N)$ vs. $\log_{10}(\varepsilon)$ plot, one can compute differences between the logarithms of neighboring values of $N$, and divide these differences by $\log_{10}(2)$ to obtain a piecewise approximation of the slope. This approach, for the three cases considered in Fig. 8a, leads to the points presented in Fig. 8b. All three sets of points show a pronounced decrease, from 2.0 at the largest box sizes to 0.0 when the box size equals the pixel size of the image. Between these two extremes, the points that are based on data included in the regressions in Fig. 8a are connected, pairwise, in Fig. 8b. In the left half of this intermediate
range, the high- and low-threshold curves decrease strongly. In the same range of box sizes, the middle-threshold curves exhibits a form of limit behavior, with an apparent plateau value near to 1.5.

The slopes shown in Fig. 8b do not correspond to the theoretical limit embodied in the definition of the box-counting dimension in Eq. (1), since they do not result from taking the limit of \(|\text{slope}|\) as \(\varepsilon \to 0\). Since the limit of vanishing \(\varepsilon\) does not make sense for finite-resolution images of the von Koch island, one might be tempted, for the low- and middle-threshold data in Fig. 8b,
to discern a tiny (two-point) plateau for the left hand of the data connected by lines. The values of 1.285 and 1.445, respectively, associated with these plateaux could then be viewed as the best available estimates of the box-counting dimension of the geometrical structures in Fig. 7f and e. A similar approach applied to the data points for the high-threshold curve in Fig. 8b leads however to no such conclusion. This approach is probably meaningless, that is, it probably does not make sense to examine the left end of the curves in Fig. 8b too closely.

Another way to look at Fig. 8b is to consider that the region of the graph in which ‘fractal’ behavior is exhibited is roughly in the range of log_{10}(box size) between -1.2 and -0.5. Indeed, if one considers the five rightmost points in Fig. 8b, one is struck by the clear similarity that exists between the pattern of these points and that of the data points in Fig. 3b. After a sharp drop of |slope| as ε is decreased, |slope| stabilizes around a stable plateau value. The three data points in the right half of the curves in Fig. 8b exhibit a slight wiggle, but one may consider this to be merely a random oscillation, due perhaps to the position of the boxes relative to the images. At smaller box sizes (below log_{10}(box size) = -1.2), the finite resolution of the images reduces more and more the number of boxes needed for full coverage, compared with the number required to cover the true quadratic von Koch island. As a result, |slope| decreases sharply until it vanishes. From this viewpoint, only in a narrow range of log_{10}(box size) do the images of the quadratic von Koch island exhibit a ‘fractal’ behavior, in spite of the fact that the initial structure, the quadratic von Koch island, is an acknowledged surface fractal. The box-counting dimensions associated with the plateaus in this range are equal to 1.523 (±0.063), 1.553 (±0.044) and 1.533 (±0.045) for the high-, middle- and low-threshold cases, respectively. These mean values are higher than 1.5, but not significantly so at P = 0.05 level.

The three approaches just described (i.e., regression analysis in Fig. 8a, search for a plateau in the left or in the right part of the curves in Fig. 8b) will be discussed again, later in this article, in connection with the evaluation of the surface fractal dimensions of stain patterns in a field soil. To simplify the rest of the analysis of images of the quadratic von Koch island, only the average ‘fractal’ dimensions, obtained via regression, will be addressed here. Patterns that this analysis reveals are qualitatively similar when dimensions based on the ‘right’ plateau approach are used instead.

In Fig. 9a, the variability due to thresholding appears commensurate with that due to image resolution. A similar observation pertains when one considers alternative dimensions (Fig. 9b); the discrepancies among values obtained for the box-counting, information, and correlation dimensions (for definitions of these dimensions see, e.g., Korvin, 1992; Baveye et al., 1998) are of the same order of magnitude as the variability due to image resolution. At small pixel sizes, the theoretical inequalities $D_{bc} \geq D_I \geq D_C$ (e.g., Korvin, 1992; Baveye and Boast, 1998) are verified. However, at the largest pixel size, the dimensions
follow a reverse ordering, as do the corresponding mass fractal dimensions (see the work of Baveye et al. (1998)).

The overall effect of resolution, thresholding and choice of a ‘fractal’ dimension on the surface dimensions obtained for the quadratic von Koch island is depicted in Fig. 9c. In contrast with a similar graph obtained earlier by Baveye et al. (1998) for the apparent mass fractal dimensions of the quadratic von Koch island, this figure does not suggest either an upward or a downward trend of the surface dimensions at increasing image resolutions (decreasing pixel sizes).

3. Materials and methods

3.1. Field site

The field site is located in the old Cornell University orchard, in Ithaca (NY). Soils in this orchard are moderately well-drained, were formed on a lacustrine deposit, and have been classified alternatively as a fine, illitic, mesic Glossaquic Hapludalf (Vecchio et al., 1984) or as a mixed, mesic Udic Hapludalf (Merwin and Stiles, 1994). The orchard was originally planted in 1927, but the trees were removed in 1977–1978. Between 1979 and 1983, a variety of test crops, including tobacco, sunflower and vegetables were grown on the plot. In 1985, the site was deep plowed with 12 t/ha of dolomite lime, and ryegrass and red fescue were planted (Merwin and Stiles, 1994). In April 1986, dwarf apple trees were planted 3 m apart in rows spaced 6 m. Sod grass ground cover has since been maintained between the tree rows and has been regularly mowed to a height of 6–10 cm.

3.2. Dye experiment

On July 14, 1995, a 68.6-cm i.d. metal ring was pushed into the surface layer of the soil. A total of 20 l of a 1% solution (10 g/l) of blue food coloring (F&D C1) were poured inside the cylinder and rapidly infiltrated into the soil. A total of 15 min later, a 1.8-m deep trench was dug, tangential to the outer surface of the metal ring. Initial digging was done with a backhoe, followed by carefully removing soil with shovels in order to obtain as vertical as possible a height of 6–10 cm.
soil profile. Color pictures were taken of the exposed soil facies with a hand-held camera. Then a 5-cm thick slice of soil was further excavated toward the center of the ring. This same procedure was repeated three times, at 15-cm intervals, to obtain evidence of dye preferential transport at various points underneath the metal ring.

3.3. Image manipulation

Two different color pictures of a single soil facies were used in the present study. These pictures, labeled ‘16’ and ‘17’, differ slightly with respect to viewing angle and exposure. Color prints and slides were obtained in both cases. The slides were scanned and the resulting \((2048 \times 3072\) pixels) digitized images were stored in RGB (red–green–blue) color-coding format on a Kodak CD-ROM.

The software Adobe Photoshop™ (version 4.0, Adobe Systems) was used to manipulate and analyze soil images. Using features of this software, images 16 and 17 were retrieved at five different resolutions \((2048 \times 3072, 1024 \times 1536, 512 \times 768, 256 \times 384, 128 \times 192\) pixels).

To ensure that all digitized images would receive identical treatments, precisely the same field of view was cropped (i.e., delineated and cut) in each case. In addition, to maximize the contrast between stained and background soil material, the storage format of the cropped images was changed from RGB to CYMK (cyan–yellow–magenta–black), and the cyan channel was retained for further analysis. This channel corresponds very closely with the color of the dye used in the field experiment, a feature that makes the stain patterns much more sharply contrasted than for any of the other channels available in Adobe Photoshop™. For the remainder of the work, the cyan channel of each image was converted to a grayscale image.

3.4. Thresholding algorithms

To threshold or ‘segment’ a digitized image, one could in principle proceed by trial and error until one achieves a thresholding that appears reasonable, i.e., coincides with some a priori idea one may have about the two categories of pixels one attempts to separate. Unfortunately, this procedure is very subjective and may lead to biases when one is trying to compare images, or in the analysis of time sequences of images of a given object (e.g., under evolving lighting conditions). To palliate these difficulties, numerous automatic, non-subjective thresholding algorithms have been developed (e.g., Glasbey and Horgan, 1995). Two of the most commonly used were adopted in the research described in the present article. Both are iterative.

The intermeans algorithm is initiated with a starting ‘guess’ for the threshold. Then the mean pixel value of the set of pixels with grayscale level greater than
the initial threshold is calculated, and likewise for the set of pixels with grayscale level less than or equal to the initial threshold. The average of these two means is calculated, and truncated to an integer, to give the next ‘guess’ for the threshold. This process is continued, iteratively, until it converges, i.e., until there is no change in the threshold from one iteration to the next.

In the minimum-error algorithm, the histogram is visualized as consisting of two (usually overlapping) Gaussian distributions. As with the intermeans algorithm, a starting ‘guess’ for the threshold is made. The fraction of the pixels in each of the two sets of pixels defined by this threshold is calculated, as are the mean and variance of each of the sets. Then, in effect, a composite histogram is formulated, which is a weighted sum of two Gaussian distributions, each with mean and variance as just calculated, and weighted by the calculated fraction. The (not necessarily integer) grayscale level at which these two Gaussian distributions are equal is calculated (involving solution of a quadratic equation). This grayscale level, truncated to an integer, gives the next ‘guess’ for the threshold. Again, the process is continued, iteratively, until it converges.

Both thresholding algorithms suffer from the fact that the choice of the starting guess used to initiate the iterative calculations influences the convergence to a final threshold value. The resulting indeterminacy was avoided by using an objective approach developed by Boast and Baveye (submitted).

3.5. Removal of islands and lakes

After thresholding the images of soil profiles with one of the algorithms described above, the resulting geometrical structure is generally very disconnected; besides two or three large ‘continents’ that extend downward from the soil surface, there is a myriad of ‘islands’ of various sizes and shapes. Some of these islands may in fact be peninsulas, artificially separated from the continents by the coarse-graining associated with the generation of images at a specified resolution. Some of the islands, however, may be truly disconnected from the continents, and may be manifestations of 3-D flow, not strictly in the plane of the images.

For the purpose of describing 1-D preferential flow in field soils, one may want to restrict application of fractal geometry to the part of a stain pattern that is connected to the inlet surface. This can be achieved with Adobe Photoshop™ by selecting the continents with the magic wand tool, inverting the selection (i.e., selecting everything but the continents), and making the latter selection uniformly white by adjusting its contrast and brightness. This procedure effectively eliminates islands.

In a similar manner, even though a physical justification is less obvious in this case, it is possible to remove the ‘lakes’, or patches of unstained soil within the continents.
3.6. Edge delineation

After thresholding and removal of islands/lakes (in case they are removed),
the filtering capability of Adobe Photoshop™ is used to isolate the edge pixels
of the stain patterns, using the Adobe Photoshop™ ‘Laplacian’ filter, which
isolates all inside edge pixels, including diagonal ones, and requires a subse-
quent inversion of grayscale values to produce a final image where the edge
pixels, in black, contrast with the white background.

This procedure is applied to digitized images without any prior change in
pixel size, which in Adobe Photoshop™ may be modified arbitrarily. To isolate
the edge of stain patterns, however, it may be tempting to decrease substantially
the pixel size. Every pixel would then be replaced by, for example, 4, 9, 16, 25,
36, . . . smaller pixels. In this manner, the ‘edge’ pixels would approximate more
closely the perimeter of the stained patterns. Preliminary tests have shown that
this approach significantly affects the final values found for the surface fractal
dimensions of the stain patterns. However, in the range of pixel sizes analyzed,
there was no obvious optimal pixel size reduction factor, common to all images.
Therefore, until this question is better understood, it was decided not to modify
the pixel size in the present research. Also, since in the relevant literature
previous fractal analyses based on digitized images make no mention of the role
of pixel size in the delineation of edges, it was decided not to involve it as one
of the possible ‘subjective’ choices analyzed in the present work.

3.7. Calculation of fractal dimensions

The box-counting, information and correlation dimensions were calculated
using the C++ code ‘fd3’ written by John Saraille and Peter Di Falco
(California State University at Stanislaus). This code, widely available on the
Internet, e.g., via anonymous ftp at ftp.cs.csustan.edu, is based on an algorithm
originally proposed by Liebovitch and Toth (1989). As do virtually all other
algorithms that are meant to evaluate fractal dimensions of geometrical struc-
tures in a plane, fd3 only considers the centroids of the various pixels constitut-
ing the images of these structures. The side of the smallest square that fully
covers the given set of points is successively halved 32 times, yielding box
coverages with progressively smaller boxes. The two largest box sizes are
considered too coarse and are therefore not taken into account in the calculation
of the box-counting, information, and correlation dimensions. Similarly, at the
low end of the range of box sizes, the data points for which the number of boxes
is equal to the total number of points (= number of pixels in the image) are
ignored.

The box-counting dimensions of several of the stain patterns were also
calculated using a Pascal code written especially for the present work.
4. Results and discussion

In all the (grayscale) images resulting from selecting the cyan layer in the CMYK representation of pictures 16 and 17 at various resolutions, the preferential pathways appear darker than the background soil, both in the surface and in the deeper horizons (e.g., Fig. 10a).

Application of the thresholding algorithms to these images gives the threshold values reported in Table 1. The systematic differences between the threshold values for the images derived from pictures 16 and 17 correspond to differences in the exposure of the pictures, picture 16 being slightly underexposed compared with picture 17. Beside this influence of picture exposure, it appears that the minimum-error algorithm yields threshold values that are generally, but not always (cf. images 17-1 and 17-5), larger than those obtained with the intermeans algorithm. This discrepancy, when it is large, affects a sizeable portion of the pixels, e.g., in image 16-5, 17.2% of the pixels that are above the intermeans threshold are not above the minimum-error threshold.

After thresholding of a grayscale image with one of the two algorithms, and delineation of the edge using the Laplacian filter, one obtains an image like that of Fig. 10b. Further removal of islands and lakes produces the much ‘cleaner’ image of Fig. 10c. In each case, the solid lines correspond to the inside edge pixels, including diagonal pixels.

Not surprisingly, the differences between threshold values in Table 1 translate into different surface fractal dimensions, as shown in Fig. 11. Without exception, for a given image and a given definition of fractal dimension, the surface fractal dimension is larger in the binary image obtained with the lowest threshold value. In most cases (except in images 17-1 and 17-5), the intermeans threshold is smaller than its minimum-error counterpart, and the surface fractal dimensions determined using the intermeans threshold are higher than those based on the minimum-error threshold (open symbols generally higher than full symbols in Fig. 11). Quantitatively, the influence of the thresholding method on fractal dimensions remains relatively small. The largest difference, 0.029, is found in the case of the box-counting dimension in image 16-4 (circles in the fourth set of points from the right in Fig. 11).

More significant, quantitatively, is the influence of the choice of a ‘fractal’ dimension among the three candidates envisaged in the present research. These dimensions satisfy the inequalities $D_{bc} < D_1 < D_C$, with $D_{bc}$ often appreciably lower than the other two. The largest difference, 0.08, is found in image 17-5 between the box-counting dimension and the other two dimensions.

Removal of the islands and/or lakes introduces another level of subjectivity in the evaluation of the surface fractal dimensions of stained preferential flow patterns. The decision to remove islands and, particularly, that to remove both islands and lakes appear to have a much more significant effect on the final dimension values than did either the choice of a thresholding algorithm or the
Fig. 10. (a) Grayscale image, labeled 16-2-c, of the cyan layer of picture 16, retrieved from disk at the second highest possible resolution, (b) black-and-white (binary) image obtained by thresholding the image with the intermeans algorithm and isolating the edge, and (c) same image as in (b) but with removal of ‘islands’ and filling of ‘lakes’ prior to identifying the edge (see text for details).
Table 1
Values of the physical pixel size (cm), and of intermeans and minimum-error thresholds (grayscale levels) for the various digital versions of the available pictures of the soil profile, sorted in order of pixel size

<table>
<thead>
<tr>
<th>Picture number</th>
<th>Physical pixel size (cm)</th>
<th>Intermeans threshold (grayscale level)</th>
<th>Minimum-error threshold (grayscale level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-1</td>
<td>0.042</td>
<td>117</td>
<td>120</td>
</tr>
<tr>
<td>17-1</td>
<td>0.053</td>
<td>112</td>
<td>111</td>
</tr>
<tr>
<td>16-2</td>
<td>0.083</td>
<td>118</td>
<td>122</td>
</tr>
<tr>
<td>17-2</td>
<td>0.105</td>
<td>113</td>
<td>115</td>
</tr>
<tr>
<td>16-3</td>
<td>0.167</td>
<td>118</td>
<td>122</td>
</tr>
<tr>
<td>17-3</td>
<td>0.211</td>
<td>113</td>
<td>114</td>
</tr>
<tr>
<td>16-4</td>
<td>0.333</td>
<td>118</td>
<td>123</td>
</tr>
<tr>
<td>17-4</td>
<td>0.420</td>
<td>113</td>
<td>115</td>
</tr>
<tr>
<td>16-5</td>
<td>0.667</td>
<td>118</td>
<td>125</td>
</tr>
<tr>
<td>17-5</td>
<td>0.846</td>
<td>109</td>
<td>109</td>
</tr>
</tbody>
</table>

selection of a specific definition of fractal dimension (cf. Fig. 12, similar graphs existing for the information and correlation dimensions). With dimensions for the untransformed state clustered around 1.55, removal of the islands alone makes the dimension values drop to approximately 1.42, whereas after removal of both islands and lakes, the dimension values range between 1.31 and 1.40.

In both Figs. 11 and 12, the resolution of the images used for the evaluation of the surface dimensions does not appear to have a significant influence on the

Fig. 11. Effect of thresholding and of the choice of a fractal dimension on the relationship between surface 'fractal dimensions' and physical pixel size. Open symbols correspond to the intermeans algorithm and full symbols to the minimum-error algorithm. Circles, squares and diamonds are associated with the box-counting, information and correlation dimensions, respectively. The analysis was carried out without prior removal of the islands or lakes.
observed dimensions. In all cases, there is significant scatter in the values found for the different dimensions over the range of image resolutions considered, and there is never any clear overall convergence pattern. In some rare instances, like the ‘no islands/no lakes’ values in Fig. 12, there seems to be a coherent trend at small pixel sizes. Whether or not this is more than a fortuitous occurrence, these trends all suggest a convergence to a non-integer value, in sharp contrast with the behavior identified by Baveye et al. (1998) for the (apparent) mass fractal characteristics of the same stain pattern.

If it is assumed that there are no coherent trends with pixel size and if the scatter in the dimension values is viewed as random, one can calculate a mean and standard deviation for each case. The resulting estimates, provided in Table 2, depend on the choice of a thresholding algorithm, on which fractal dimension is evaluated, and on whether or not islands and/or lakes are removed. However, they present the distinct advantage of being independent of the resolution of images. In this sense, the estimates in Table 2 are somewhat more robust than traditional estimates, oblivious of most of the vicissitudes of image resolution.

The lack of an overall trend towards an integer value over the range of image resolutions considered here is also clearly illustrated by the ‘summary’ graph in Fig. 13. In practical terms, this graph means that different observers, making different choices at various stages in the evaluation of the surface ‘fractal’ dimension of the stain pattern, are likely to end up with dimensions anywhere between 1.32 and 1.64. If these observers used several image resolutions and estimated the means as in Table 2, the range of observed dimension values would narrow somewhat, to between 1.35 and 1.59. Either way, the resulting
Table 2
Means and standard deviations of fractal dimensions obtained by averaging out the effect of image resolution

<table>
<thead>
<tr>
<th></th>
<th>Intermeans threshold</th>
<th>Minimum-error threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Untransformed image</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box-counting dimension</td>
<td>1.55 ± 0.03</td>
<td>1.54 ± 0.04</td>
</tr>
<tr>
<td>Information dimension</td>
<td>1.59 ± 0.03</td>
<td>1.58 ± 0.04</td>
</tr>
<tr>
<td>Correlation dimension</td>
<td>1.59 ± 0.03</td>
<td>1.58 ± 0.03</td>
</tr>
<tr>
<td><strong>No islands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box-counting dimension</td>
<td>1.42 ± 0.03</td>
<td>1.43 ± 0.03</td>
</tr>
<tr>
<td>Information dimension</td>
<td>1.48 ± 0.03</td>
<td>1.48 ± 0.02</td>
</tr>
<tr>
<td>Correlation dimension</td>
<td>1.49 ± 0.03</td>
<td>1.49 ± 0.03</td>
</tr>
<tr>
<td><strong>No islands / no lakes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box-counting dimension</td>
<td>1.35 ± 0.03</td>
<td>1.37 ± 0.04</td>
</tr>
<tr>
<td>Information dimension</td>
<td>1.39 ± 0.03</td>
<td>1.40 ± 0.03</td>
</tr>
<tr>
<td>Correlation dimension</td>
<td>1.40 ± 0.03</td>
<td>1.41 ± 0.04</td>
</tr>
</tbody>
</table>

range of fractal dimensions is a sizable fraction (25% or 33%) of the total permissible range for this fractal dimension (1 to 2).

The graph of Fig. 13 lends strong credence to the contention that the stain pattern is a surface fractal, i.e., that its surface fractal dimension is non-integer. Further support for this viewpoint is provided by the fact that each of the data points in the graph has been obtained (via non-linear regression) with a very high $R$ value, always larger than 0.99. A representative illustration of the estimation procedure is provided in Fig. 14a for image 17-1. The evidence of

![Graph showing "Fractal" Dimension vs. Physical Pixel Size](image-url)

Fig. 13. Compilation of all the surface fractal dimensions obtained in the present study, vs. the physical pixel size. The gray region represents the global envelope of the plotted data, determined by connecting the outermost points with straight lines.
Fig. 14. (a) Relationship between the number of covering boxes and their size for image 17-1, the highest-resolution image derived from picture 17, thresholded with the intermeans algorithm. The average box-counting dimension (i.e., the absolute value of the slope) associated with the regression lines is equal to 1.58 (‘untransformed’, \( R = 0.993 \)), 1.42 (‘no islands’, \( R = 0.997 \)) and 1.32 (‘no islands/no lakes’, \( R = 0.998 \)). (b) Absolute value of the slope \( \frac{d \log_{10}(N)}{d \log_{10}(e)} \) vs. box size, obtained by calculating differences between adjacent points in (a). The straight lines in Fig. 14b connect the points corresponding with the data used to calculate the regression lines in Fig. 14a.

extremely high \( R \) values, combined with the more than two decades of pixel sizes over which power–law behavior is observed, would be considered convincing enough by most authors to conclude that the stain pattern is a surface fractal.
However, as was done earlier with the quadratic von Koch island, the results of Fig. 14a may be analyzed from another perspective, leading in a very different direction (Fig. 14b). Indeed, for all images, the graphs of $|\text{slope}|$ vs. $\log_{10}(\text{actual box size})$ like that in Fig. 14b reveal a largely monotonic decrease of $|\text{slope}|$ as the box size decreases. There is no tendency in the left part of the fd3 range (connected points) for the points to form plateaux, i.e., for $|\text{slope}|$ to be more or less constant for a range of box sizes. To the left of the fd3 range, the predominant behavior seems to be a convergence to unity. In the right part of the fd3 range, one would be hard pressed to identify a plateau for the ‘untransformed’ case (open circles). For the ‘no islands’ and ‘no islands/no lakes’ cases, one might perhaps view the wiggles in the range of $\log_{10}(\text{actual box size})$ between 0 and 1.1 (first four connected points from the right) as a random fluctuation around stable values of $1.50 \pm 0.07$ (‘no islands’) and $1.40 \pm 0.06$ (‘no islands/no lakes’). These box-counting dimensions are somewhat higher than those obtained via regression (Fig. 14a) and leftmost set of points in Fig. 12. This is not surprising, given the fact that the slope of the regression lines in Fig. 14a is in some sense an average over the whole fd3 range, whereas the estimates just calculated are based solely on the subrange of box sizes where $|\text{slope}|$ is largest. Unfortunately, the estimates of the box-counting dimensions obtained by considering plateaux in Fig. 14b suffer from a great deal of subjectivity, associated with the range of box numbers considered in the analysis. Indeed, if for some reason, one were to remove the rightmost connected data point in either of the ‘no islands’ cases, evidence of a plateau would be significantly weakened and one would be tempted to conclude that the corresponding pattern is non-fractal. The decision to include the rightmost point in the graph is linked to the (subjective) choice made in the computer code fd3 to ignore the two largest box sizes when calculating the box-counting dimension (see Section 3). This choice seems very sound, yet one could very well have opted to ignore only the largest, or to be more restrictive and to ignore the three largest box sizes. These alternate choices would affect significantly the decision to consider or not to consider that images of the stain pattern are surface fractals. Of course, as is certainly the case with the quadratic von Koch island, this decision may only pertain to images of the stain pattern and not to the stain pattern itself.

5. Conclusions

The key results of the research reported in the present article are contained in Figs. 13 and 14. The first of these figures shows that operator choices, made during the analysis of images of the stain pattern, cause the final surface ‘fractal’ dimension estimates to vary over a sizable range, approximately from 1.32 to 1.64. This range may be reduced somewhat by averaging out the influence of image resolution (cf. Table 2).
In spite of the subjectivity associated with the evaluation of the surface fractal dimensions, Fig. 13 shows no tendency for the dimensions to tend to an integer value at small pixel sizes. This observation suggests that the stain pattern is a genuine surface fractal. This conclusion is also supported by the very good fit of regression lines to data in Fig. 14a, giving remarkably high $R$ values. Contrastingly, Fig. 14b casts doubt on the fractality of representations of the stain pattern in digitized images. In the range of box sizes considered appropriate for the evaluation of fractal dimensions, there is no appearance of a plateau at the lower end of the range, and although one could conceivably identify plateaux for two of the three curves at the high end of the fd3 range, evidence in this sense is not particularly strong and the choice of where to see a plateau is quite subjective.

The resulting conundrum about the fractality of images of the stain pattern illustrates vividly some of the difficulties associated with the application of fractal geometry to physical systems. To some extent, however, it may be largely academic. The method described above, giving estimates of surface dimensions by averaging out the effect of image resolution, yields numbers that are more robust than traditional estimates, ceteris paribus. Whether or not the observed stain pattern is a surface fractal, these numbers may prove useful in mathematical descriptions of the preferential transport of water and solutes in field soils. Beyond the interest that there may be in characterizing the geometry of a complicated pattern or object with a number, perhaps of even greater interest is the question of what can be done with the number once it is obtained. Further research in the direction explored by Crawford et al. (in press) will be needed to provide answers to this important question.

Acknowledgements

The research reported in the present article was supported in part by grant No. DHR-5600-G-1070-00 PSTC Project Number 11.243 awarded to one of us (P.B.) by the United States Agency for International Development and by a BARD grant awarded to Tammo Steenhuis. Gratitude is expressed to Stokely Boast, who wrote a computer program to generate the prefractals of the quadratic von Koch island, and to an anonymous reviewer for very thoughtful comments.

References


