DIGITAL MODELS AND ALGORITHMS FOR FRACTAL ANALYSIS OF GEOLOGICAL DATA STREAMS

Uralova Iroda Abduvali qizi
Computer engineering Tashkent university of information technologies named after Muhammad al-Khwarizmi, assistant
Tashkent, Uzbekistan
E-mail: irodaabduvaliyevna1995@gmail.com
Malikova Nodira Turg‘unovna
Computer engineering Tashkent university of information technologies named after Muhammad al-Khwarizmi, associate Professor
Tashkent, Uzbekistan
E-mail: Malikova.nodira2020@gmail.com

Abstract. This article analyzes the fractal properties of geological systems and their measurement methods. The necessary conditions for accurate fractal dimension determination, including data quantity and scale range issues, are thoroughly examined. The research employs the box-counting algorithm with computer software to enhance the reliability of fractal analysis, with results statistically validated. The authors specialized software improves measurement accuracy and enables analysis of methodological errors found in the literature. The findings can serve as a basis for applying fractal analysis in practical geological research.

Keywords: Fractal analysis, geological data stream, digital models, self-similarity, Box-counting algorithm, Linear regression, digital images.

Annotatsiya. Ushbu maqolada geologik tizimlarning fraktal xususiyatlari va ularning o‘lchov usullari tahli qilindi. Fraktal o‘lchovni aniqlashda talab qilinadigan mezonlar, jumladan ma’lumotlar hajmi va o‘lchamlar oralig‘i masalalari chuqur tahli qilindi. Tadqiqotda fraktal analizning ishonchliligi uchun kompyuter dasturlari yordamida quti hisoblash (box-counting) algoritmi qo‘llanildi va natijalar statistik jihatdan tasdiqlandi. Mualliflar tomonidan ishlab chiqilgan maxsus dasturiy ta’minot orqali fraktal o‘lchovning aniqligi oshirildi va soha adabiyotlarida uchraydigan metodik xatolar tahlil qilingan. Ishning natijalari fraktal tahlilni amaliy geologik tadqiqotlarda qo‘llash uchun asos bo‘lishi mumkin.

Kalit so‘zlar: Fraktal tahlil, geologik ma’lumotlar oqimi, raqamli modellar, o‘ziga-o‘xshashlik, Quti hisoblash algoritmi, Chiziqli regressiya, raqamli tasvirlar.

Аннотация. В данной статье анализируются фрактальные свойства геологических систем и методы их измерения. Исследованы необходимые условия для точного определения фрактальной размерности, включая вопросы объема данных и диапазона масштабов. В исследовании для повышения надежности фрактального анализа использован алгоритм box-

ISSN 3030-3907

counting с применением компьютерных программ, и результаты статистически подтверждены. Разработанное авторами специальное программное обеспечение повысило точность фрактальных измерений и позволило проанализировать методологические ошибки, встречающиеся в литературе. Результаты работы могут служить основой для применения фрактального анализа в практических геологических исследованиях.

Ключевые слова: Фрактальный анализ, поток геологических данных, цифровые модели, самоподобие, Алгоритм подсчета ячеек, Линейная регрессия, цифровые изображения.

Introduction

One of the most relevant problems of modern geology is the study of complex shapes observed in natural geological objects and their scale-invariant properties [1]. Fractal geometry provides the means to quantify and systematically analyze this complexity [2]. However, practical investigations often encounter serious difficulties in estimating fractal dimensions, such as small data volume, application of inappropriate methodologies, and lack of statistical reliability [3]. The presented article deeply studies the conditions for increasing the reliability of fractal analysis and gives concrete, practical recommendations [4]. The novelty of the work is not only an analysis of methods of fractal dimension calculation but a demonstration of the way methodological errors can be minimized by statistical validation and use of computer software [5]. In such a way, the work is directed to the development of an approach that will guarantee high accuracy and reliability of the fractal properties studies of geological systems, and the results obtained will be a good tool for scientific research and practical geological estimations.

A striking feature of geology is that many natural objects are self-similar, meaning that they appear the same over a wide range of scales. For instance, if a picture of some geological outcrop is taken without including an object of known dimensions such as a rock hammer or scale bar then it is impossible to tell how big the features in the picture really are. Another characteristic that many natural shapes share is that they are rugged and highly irregular. One of the cardinal difficulties with natural sciences has long been the inability of classical Euclidean geometry to adequately describe real shapes, which makes one of the fundamental tasks of natural sciences the systematic classification of objects complicated. Fractal geometry introduces this notion of scale invariance directly into its language, and one of the main reasons for its wide acceptance lies in its capability to model such complexity.

1. RELATED WORKS

The fractal dimension, central to fractal geometry, is a concept that emerges from very basic theoretical considerations rooted in an object's self-similarity [6].

ISSN 3030-3907

Consider, for simplicity, a straight line segment. The segment can be divided into $N=b$ equal parts by selecting an arbitrary integer b. Each part is a scaled-down version of the original segment, related by a similarity ratio $u(N)=1/b=1/N$. A rectangle can be subdivided into $N=b^2$ identical smaller rectangles, while the similarity ratio in this case is given by $u(N)=1/b=1/N^{1/2}$. Using the same procedure in the case of a three-dimensional parallelepiped, one gets $u(N)=1/N^{1/3}$. For any arbitrary dimension D, the similarity ratio depends on the form $u(N)=1/N^{1/D}$.

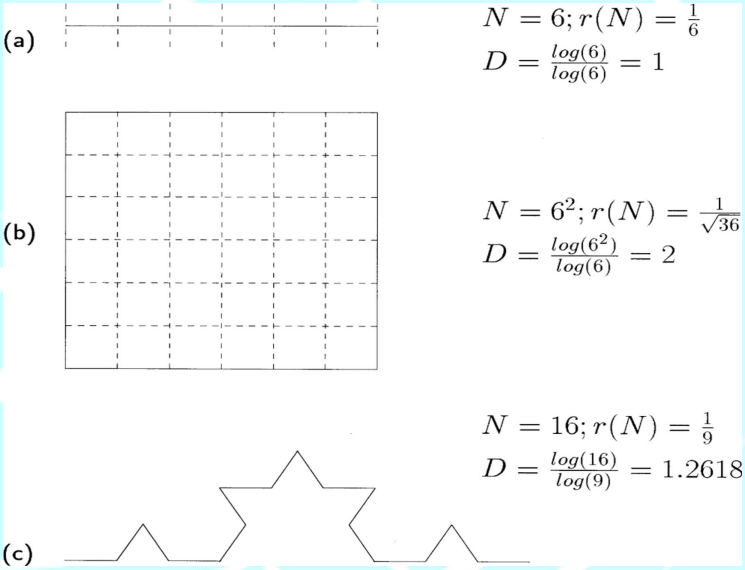


Figure 1. Self-similarity demonstrated for: (a) a line segment, (b) a square, and (c) a Koch curve. The dimension of the object is computed from its similarity ratio as discussed in the main text; these produce integer dimensions for the two standard geometric figures, (a) and (b) whereas for the fractal, (c)

Fractal geometry operates with non-integer values of D and focuses on characterizing self-similar structures, whether they form a connected path (such as the Koch curve) or a disconnected set (like Cantor dust) [7]. The wide variety of artificially created fractal lines highlights two defining properties of fractal objects: self-similarity and the absence of differentiability [8].

For fractal lines generated by a geometric construction process, a practical formula can be used to determine D directly from the parameters of the iterative construction [9]. At construction level l, let Nl represent the number of segments and Ll their individual length.

In fact, he took each scene from that Greek play and copied it into his own play with such ease that if we read the two plays, we could easily find the matching scenes of both plays. One finds that

$$D = \frac{\log\left(\frac{N_l}{N_{l+1}}\right)}{\log\left(\frac{L_{l+1}}{L_l}\right)} \qquad (1)$$

So, considering the Koch curve that is obtained from the original segment through

ISSN 3030-3907

the similarity ratio 1/3, one has $N_l / N_{l+1}=1/4$; $L_{l+1}/L_l =1/3$, whence $D=1.2618$ (Fig. 1c).

Measuring the fractal dimension of real objects. Unlike synthetic fractal curves, whose fractal dimension D whereas D can be computed exactly using analytical formulae, the fractal dimension of natural or empirical objects must be estimated through specialized algorithms - a number of which are described below. Lying at the heart of such methods, and therefore at the heart of practical fractal analysis, is a versatile yet frequently overlooked mathematical instrument: linear regression.

A fundamental problem in searching for fractal behaviour in a dataset is the definition of the range of scales over which scale invariance obtains. This reduces to a problem of fitting a straight line over a preselected interval in a log-log plot. As will be demonstrated in Section 3, it is almost always possible to carry out such a fit on any given set of observations; however, this does not necessarily imply that the outcome represents a meaningful self-similarity. In practice, cautious use of linear regression is of paramount importance in determining whether a morphology exhibits fractal properties, and hence the process needs to be scrutinized in detail.

Common Fractal Algorithms. Two of the most commonly used algorithms for fractal dimension estimates are the Richardson method [10] and the box-counting method [11]. In the following, they will be referred to with the abbreviations RM and BC, respectively.

2. METHODS

The Richardson (Divider) Method. The RM is the algorithm applied by Richardson [10] in his seminal study on how the measured length of a boundary changes as a function of the step length (ruler size). For reliable results, two conditions should be satisfied: the curve must be connected, and the range of step lengths should span as wide an interval as possible. The curve need not be differentiable, but continuity is required.

To begin with (Fig. 2), one employs a wide-range map, say 1:100 000. A map divider is then set to, say, 10 cm, fixed on one end of the line, then rotated n_1 times until the other end is reached; the residual, if any, is ignored. (This is an inherent cause of error in the algorithm.) The first value of L is given by $n_1|10$ km. The divider is then reduced to a length of 5 cm and rotated n_2 times: now $L=n_2|5$ km. This step is repeated M times, using increasingly shorter spreads and more accurate
maps (1:50 000, 1:10 000, .. .).

ISSN 3030-3907

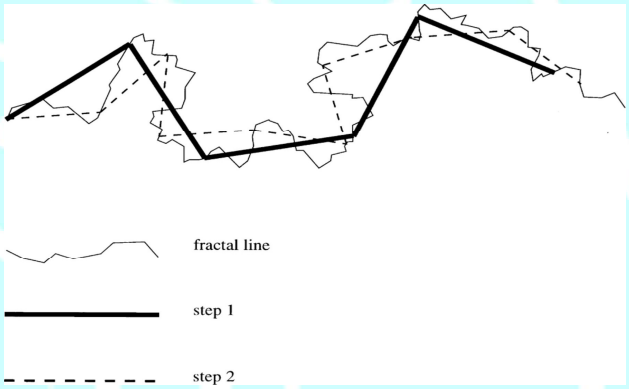


Figure 2. The divider, or walking-rule, method. The apparent length of a curvilinear feature is estimated by approximating it with a series of straight segments, which are set using a fixed span on a pair of map dividers. The smaller the setting of the divider, the larger the measured length will be, and the more precise an approximation it will be.

At this stage, one has calculated M couples of $L(\eta_i)$ and η_i. These values are related by the power law

$$L(\eta)=\eta^{1-D} \qquad (2)$$

Which shows how L increases with decreasing η. It is con- venient to rewrite relation (2) by applying a logarithmic transformation to both sides:

$$\log (L(\eta))=a+(1-D) \log (\eta) . \qquad (3)$$

Now one has to calculate the coefficients of the regression line between the dependent variable $\log (L(\eta_i))$ and the independent variable $\log (\eta_i)$, where $i=1, \ldots, M$, and M is the number of η considered. The fractal dimension D is given by 1— line slope.

One of the problems with the WR method is that it has a finite remainder. This is due to the fact that for each ruler length L, part of the object is bound to be left uncovered by the last step. There is no way to solve this inconvenience, which can cause considerable errors in the estimate of D [13].

The WR method can only be employed for measuring the D of connected self-similar objects. This limitation, together with the remainder error and the difficulty of objective com- puter implementation, makes this algorithm inferior to the BC method, which is more powerful and of more general use. We therefore restrict our subsequent analysis to BC as the general algorithm for fractal analysis.

The box counting algorithm. This procedure is based on the notion of 'covering' discussed above and comprehends different definitions of 'dimension' [12]. It is more general than the WR, since it can be used on both connected and non-connected forms. One takes eq. (2) and lets $L(\eta)/\eta=N\ \eta$, where $N\ \eta$ is the number of objects whose linear dimension exceeds η, to obtain

$$N(\eta)a\eta^{-D}. \qquad (4)$$

ISSN 3030-3907

Again, it is convenient to turn eq. (4) into

$$\log (N(\eta))=a—D \log (\eta). \quad (5)$$

(N(η))=a—D log (η).

D is calculated by covering the object with a grid of squares, (or, in the case of three-dimensional surfaces, of cubes) firstly of side η_1, then counts the number $N(\eta_1)$ of squares which include part of the object. The measurement is then repeated using side η_2, obtaining $N(\eta_2)$ squares. This step is repeated S times, with squares of ever shorter side. Finally, one calculates the regression line between the independent variable $\log (\eta_i)$ and the dependent variable $\log (N(\eta_i))$, where $i=1, \dots, S$. D is given by the absolute value of the line slope.

The BC algorithm is also at the root of all the main definitions of fractal dimension. In fact, high-order dimen- sions, such as the capacity, information and correlation dimensions [ ], can all be calculated through the BC algorithm according to the moments formula:

$$M_q(\eta) = \sum_{j=1}^{N(\eta)} P_j^q, \quad (6)$$

Where P_j is proportional to the sum of the objects in a box of side η. The relation $M_q(\eta)\approx\eta^{(q—1)Dq}$ Dq allows the calculation of any of the above high-order dimensions, according respectively to values of q: 0, 1, 2.

Also in this case, self-similarity can be inferred experimentally from the goodness of fit of line (5) to the data. As discussed above, to have a realistic, and practically useful, estimate of D, it is necessary to use the widest possible range of values of the square side. Another condition is that, at each step, a large number of boxes should contain part of the object: in other words, the object to be measured should fit the whole map, not a small part of it, otherwise the result will be biased by the outline of the object, a common problem with incomplete geological exposure [13].

Linear regression: the key to fractal dimension. All the methods for measuring scale invariance in a range and its characterizing parameter, fractal dimension D, are based on a linear regression procedure. Such a technique tries to assess if there is a linear relationship among data, in other words, whether it is possible to describe data by using a linear model. We have seen that power laws are equations governing the main algorithms for fractal analysis and can be transformed into linear laws by logarithmic transformations.

There appears to be general ignorance of the underlying assumptions of regression analysis. Probably the most common error is the uncritical use of R^2 as an indicator of the suitability of the linear model. This is incorrect. The key tool in studying the adequacy of linear regression is the analysis of uesiduals. For example, doing an R^2 calculation for the parabola $y=x^2$ in the interval $0\leq x\leq 1$, one obtains a value of R^2 as large as 0.936, whereas it is clear from the residuals that the linear model is poor.

ISSN 3030-3907

What marks an adequate model is, in fact, the way the residuals scatter around the regression line. First, the variance - the scatter of the residuals around the line - has to be constant across the whole range of X_i. Fortunately, this requirement is not a strict one, and moderate deviations are unlikely to give real problems. The second requirement, and this one is crucial, is that there must not be any 'clumps' of residuals on either side of the zero level. An easy way to check the presence of clumps is through the Runs Test. Runs are defined as uninterrupted sequences of the same state. The parabola $y=x^2$ is not fitted appropriately by a line in $10 \leq x \leq 1$ because the signs of the residuals are (see Fig. 3): 10+, 29 and 11+, that is the number of runs is three. The probability that such a low number of runs could occur by chance is less than 0.01, and thus the linear model must be rejected despite its high regression coefficient. Having too few runs is the most common situation in fractal analysis.

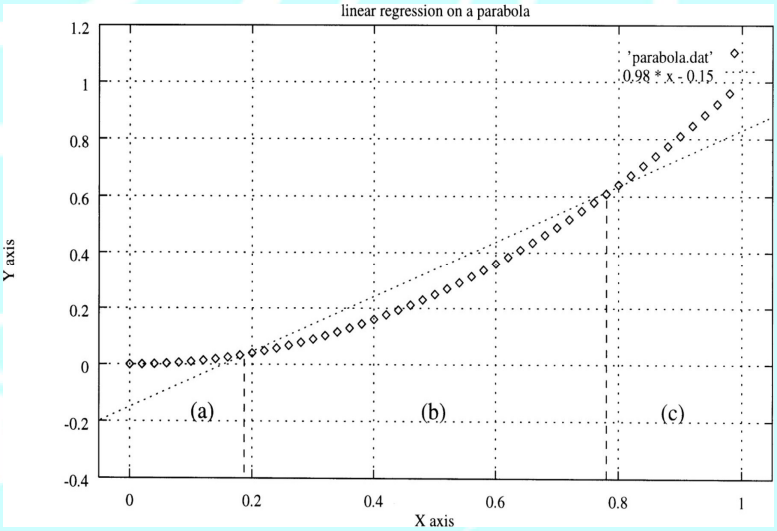


Figure 3. The parabola $y=x^2$ plotted against a linear regression line of the form $y=b_1x+b_0$, with $b_1=0.98$ and $b_0= -0.15$. There are three systematic clusters of residuals above [regions (a) and (c)] and [region (b)] below the line, showing that the linear model fits the parabola poorly despite the high regression coefficient ($R^2=0.936$, see main text).

It is relevant to note that it is always possible to find a scale range in a dataset, where a linear relationship seems to exist. In practice, running the algorithms always yields at least two data points, and it is always possible to draw a straight line through them. This is illustrated in Figure 4, showing the result of a box-counting analysis applied to a tectonic map. The source tectonic map is shown in Figure 4a.

ISSN 3030-3907

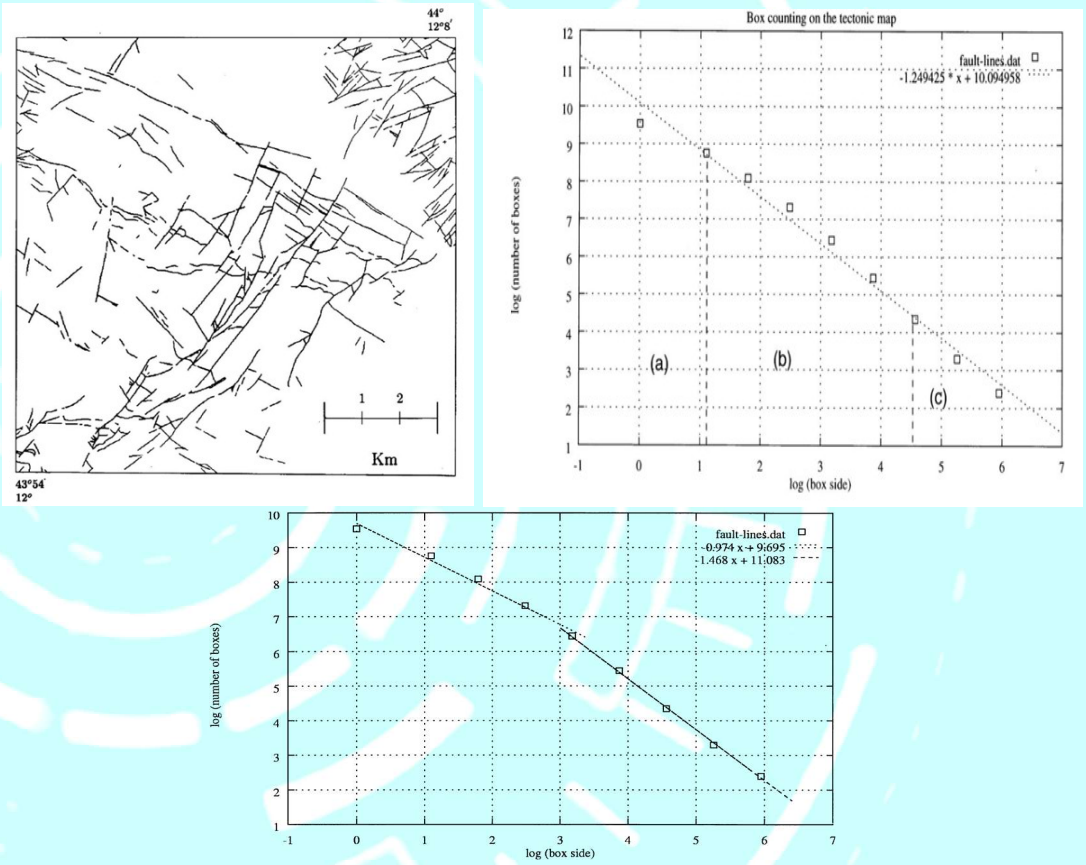


Figure 4. (Top) Tectonic map showing part of the Aral Sea. Surface fault traces are drawn as linear features. (Middle) Fractal analysis on the map shown in (a). Again, the residual values fall systematically into three categories: (a) below, (b) above, and (c) below the best-fit regression line. (Bottom) Same data plotted with two different regression lines fitted to different portions of the scale range.

If all data points are included and a simple regression line is fitted, a fractal dimension of D=1.249 is obtained (Figure 4b). If the fit is instead applied to two By fitting two separate regression lines over the ranges $0 \leq x \leq 2.5$ and $3 \leq x \leq 6$ each of which seems more linear on its own, the resulting dimensions become D=0.974 and D=1.468 D=1.468 (Figure 4c). If only the first and last points are used, the value D=1.199 is obtained. Which of these is correct? In truth, none can be reliably chosen and the reliability of none can be checked because each of the cases mentioned above depends on too few data points, a number so small that statistical tests like the runs test cannot function. In fact, no statistical test works well on sparse data: less than roughly 15 points, for instance, render the runs test statistically meaningless, even if the data span a large range of scales six orders of magnitude of length, here, as Figure 4b shows.

The fundamental and most serious challenge in fractal analysis is therefore insufficient data. A major practical obstacle is that many algorithms are still carried out manually an extremely laborious and time-consuming process. Consequently, researchers rarely complete all the steps required to collect an adequate number of

ISSN 3030-3907

observations. Using specialized computer software is the only realistic solution. We have developed a computational package that works directly with digitized images, employs a high-resolution virtual screen (default 2040×2048 pixels, expandable), and includes recursive zoom functionality. These features enable consistent and reliable fractal analysis by ensuring a sufficient number of data points are always used.

3.1 PRACTICAL GUIDELINES FOR CALCULATING *D*

In other words, a meaningful fractal analysis needs an adequate number of data points to reliably support linear regression while having a broad enough range of scales such that the self-similarity becomes physically interpretable. As useful heuristics: In the log–log plot at least about 20 data points are needed to reasonably assess linearity. We would need the self-similarity to extend over at least two to three orders of magnitude before it should be considered physically significant. What would be meant by identifying self-similarity over half an order of magnitude? Not very much. There is no universal rule, but common practice is that self-similarity extending for less than two to three orders of magnitude is unlikely to reflect meaningful physical processes and can usually be ignored.

First and foremost, one must check whether data are indeed following a linear dependence. This task is further complicated by the fact that both WR and BC techniques share a common feature: by their very nature, they yield the most extreme points that correspond to the largest and smallest step lengths (box sizes). The curve that passes through such points will always have a part that will look approximately linear. The linearity and, hence, self-similarity should be carefully checked for being statistically significant. If one of the aforementioned conditions is not fulfilled, the estimated *D* cannot be statistically justified. That is, a fractal analysis which only uses a few data points has an inherent methodological flaw in that, even though scale invariance may indeed exist, sparse data make it impossible to distinguish this reliably from cases where scale invariance does not exist.

Consequently, the use of an inadequate number of observations in analyses carries a double jeopardy: the computations become deceptively easier, while the findings may, through an illusion, emerge as significant. The one sure thing is a strong possibility of invalid or noninterpretable results. Applications of such flawed analyses, selected from the literature, are considered in the next section.

3. RESULTS

The foregoing examples emphasize one point that repeatedly comes up in fractal analysis: the root problems are generally related to a lack of data. The inadequacy might stem directly from the source of the data (e.g., a scant earthquake catalog for the region of interest) or from the measuring technique itself (e.g., counting boxes or manually applying the walking rule). As we see it, fractal

ISSN 3030-3907

analysis should not be attempted if there is no way to get a large enough number of valid observations to support a statistically valid analysis.

When the experimental data are not inherently sparse nor burdened by substantial noise, the remedy lies in stepping away from the manual implementation of the algorithm and taking a computational approach instead. Accordingly, we developed a software package which performs a box-counting algorithm with the corresponding significance tests. Taking all precautions discussed hitherto into consideration, we now re-analyze the Koch curve as a benchmark.

A digital map of the Koch curve at full resolution is created. The program starts with a box size of 2048 pixels and decreases the box size progressively in each step. The data obtained are given in Table 1, from which we calculate the regression parameters and the residuals. With the help of equation we evaluate the probability P_N for observing the sequence of residual runs obtained which is, in this case, 0.425.

All the methodological requirements are met. We measure the fractal dimension reliably as $D = 1.2712 \pm 0.0112$. Since this relationship holds over three orders of magnitude, the self-similarity is physically meaningful. We can therefore confidently conclude that the Koch curve is fractal - a result that is, of course, already well established. On real-world objects, in fact, data availability would probably remain the principal challenge for this kind of analysis. We believe that fractal analysis can only make sense when the investigated phenomenon is capable of providing enough data to meet the criteria developed above [14].

Table 1. Results of the box-counting algorithm applied to the Koch curve. The computation started with a box side length of 2048 pixels. Box sizes were halved in each successive step, from large to small scales.

$log(\eta_i)$	$log(N(\eta_i))$
7.624619	0.000000
7.277939	0.693147
6.931472	0.693147
6.584791	1.609438
6.238325	1.791759
5.891644	2.197225
5.545177	2.639057
5.198497	3.218876
4.852030	3.465736
4.510860	4.143135
4.158883	4.477337
3.806662	4.912655

ISSN 3030-3907

3.465736	5.288267
3.135494	5.834811
2.772589	6.269096
2.397895	6.824374
2.079442	7.143618
1.791759	7.552237
1.386294	8.046549
1.098612	8.414939
0.693147	8.867287
0.000000	9.560152

4. CONCLUSION

Fractal analysis is a very effective method of studying geological complexity. Research shows that result reliability must be ensured by sufficient data and proper methodology, also with statistical validation. Fully automated software solutions based on the box-counting algorithm reduce errors and improve precision. This reliable methodology can therefore be applied when analyzing geological data streams [15].

We have critically reevaluated standard techniques of fractal dimension measurement and have concluded that the inherent tedium of human counting often results in an insufficient number of data points being taken. This practice increases the risk of obtaining results that are unreliable or uninterpretable. We have therefore developed a suite of computer programs to calculate the fractal dimension by box-counting. These programs operate directly on digitized images, use a virtual screen size limited only by the amount of memory available, and incorporate a zoom function enabling selected areas of a larger image to be examined in detail. Inevitably, like all computational methods, our programs cannot compensate for data which are fundamentally inadequate or too noisy.

REFERENCES

1. Cello, G., & Malamud, B. D. Fractal analysis for natural hazards. Geological Society, London, Special Publications, 2020, 261(1), 1–12

2. Falconer, K. Fractal geometry: mathematical foundations and applications (3rd ed.). John Wiley & Sons. 2014.

3. Zhou, W., & Wang, J. Automatic fractal dimension estimation using digital image processing techniques. Computers & Geosciences, 121, 2018. 1–9.

4. Li, J., Du, Q., & Sun, C. An improved box-counting method for accurate fractal dimension estimation. Signal Processing, 158, 2019. 216–223.

5. Gao, X., & Li, Y. An improved box-counting method for image fractal dimension estimation based on adaptive differential evolution. Chaos, Solitons & Fractals, 143, 2021. 110584.

ISSN 3030-3907

6. Barton, C. C., & La Pointe, P. R. (Eds.). Fractals in the earth sciences. Springer. 2019.
7. Shen, G., & Zhao, Y. Characteristics of fractal dimension and its application in signal processing. Journal of Vibration and Control, 23(5), 2017. 825–833.
8. Xu, T., & Moore, I. D. Fractal analysis of soil surface roughness. Water Resources Research, 41(6). 2005.
9. Foroutan-pour, K., Dutilleul, P., & Smith, D. L. Advances in the implementation of the box-counting method of fractal dimension estimation. Applied Mathematics and Computation, 105(2-3), 1999. 195–210.
10. Roy, A., Perfect, E., Dunne, W. M., & McKay, L. D. Fractal characterization of fracture networks: An improved box-counting technique. Journal of Geophysical Research: Solid Earth, 115(B10). 2010.
11. Taud, H., & Parrot, J. F. Measurement of DEM roughness using the local fractal dimension. Géomorphologie: relief, processus, environnement, 11(4), 2005. 327–338.
12. Cheng, Q. Non-linear theory and power-law models for information integration and mineral resources quantitative assessments. Mathematical Geosciences, 40(5), 2008. 503–532.
13. Liu, X., & Zhang, J. A robust fractal dimension estimator for complex networks with scale-free properties. Physica A: Statistical Mechanics and its Applications, 587, 126507. 2022.
14. Sun, W., Xu, G., Gong, P., & Liang, S. Fractal analysis of remotely sensed images: A review of methods and applications. International Journal of Remote Sensing, 27(22), 2006. 4963–4990.
15. Abugharbieh, K., & Marar, H.-W. Integrating multiple state-of-the-art computer-aided design tools in microelectronics circuit design classes. Computer Applications in Engineering Education, 27(5), 2019. 1156–1167.

ISSN 3030-3907