Uncertainty of the box-counting method for estimating the fractal dimension of river networks

MIAO ZHOU, FUQIANG TIAN & HEPING HU

Department of Hydraulic Engineering, State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing 100084, China

tianfq@tsinghua.edu.cn

Abstract The fractal dimension of river networks has become a significant geomorphologic parameter for identifying hydrological patterns. Several definitions for fractal dimension are used in hydrological literature, among which the box-counting dimension via Digital Elevation Model (DEM) is widely applied. To explore the uncertainty of the box-counting method along with the mathematical implementing processing, three factors, i.e. threshold area for river networks extracted from DEM, range and increment ratio of box sizes, were selected to identify their influence on the value of the box-dimension. One real watershed, the Chabagou watershed, and one ideal watershed generated by an iterative algorithm, were selected for numerical experiments. The results suggest the following properties of the box-dimension of river networks: it shows an asymptotic power law behaviour for larger threshold areas; it has a logarithmic dependence for smaller threshold areas; and it is monotonic to the range of calculated box sizes while insensitive to the increment ratio of calculated box sizes. These relationships reveal the complexity of the box-dimension of fractal river networks.

Key words fractal river network; box-counting method; digital elevation model (DEM); uncertainty

INTRODUCTION

Natural channel networks show a fascinating variety of shapes and forms, which present a deep sense of regularity in the relationship among the parts. This regularity is presented as self-similarity by Horton (1945). Mandelbrot (1983) proved that natural channel networks have a fractal character. La Barbera & Rosso (1989) promoted a fractal dimension in terms of Horton's laws. Thereafter, several approaches to calculate the fractal dimension of networks have been applied. Rodriguez-Iturbe & Rinaldo (1997) mentioned the Hausdorff dimension, box-counting dimension, similar dimension, correlation dimension, mass dimension, information dimension, etc. Using three different approaches, Veltri *et al.* (1996) counted the fractal dimensions of 17 Calbrian drainage basins.

Over the last two decades, with the help of Digital Elevation Models (DEMs), the fractal dimension of some river basins has been counted. La Barbera & Rosso (1989) claimed that different networks, or channel networks in different physiographic locations, may display different values of fractal dimension. As fractal dimensions of many river basins were counted, the underlying connection between the fractal dimension and hydrological character were also studied. For example, Kusumayudha *et al.* (2000) showed that the box-counting dimension of river networks is related to the erosion of rock, as well as to the fractal dimension of underground river pattern.

Consequently, for further investigation on the relationship between the fractal character of rivers and their hydrological properties, it is of great importance to determine the fractal dimension accurately and subtly. However, it has been pointed out that the fractal dimension calculated by different approaches will probably be different. Therefore, to compare the applicability of those approaches for estimating the fractal dimension of river networks, further investigations are indispensable. Moreover, even for a particular method, there are some other issues that can bring uncertainty to the results, e.g. Helmlinger *et al.* (1993) found that the range of box sizes affects the fitting of box-counting dimensions. Particularly, Ichoku *et al.* (1996) showed that those four kinds of fractal dimensions decrease as the threshold of confluence area increases, albeit with only five threshold areas. Radziejewski *et al.* (1997) counted the box-counting dimension for excursion of flow process at a threshold level, and observed the same trend between dimension and threshold areas; however, the study lacks adequate data for more detailed discussion. Yang *et al.* (1997, 2001) proposed that the multifractal spectra of width function and area function were also sensitive to the threshold area. Huang (1996) studied the factors affecting the reliability of box-counting dimension from a mathematical aspect. Up until now, the sensitivity of the box-counting dimension of river networks to the threshold area has not yet been discussed comprehensively, nor has the uncertainty coming from the factors of chosen method; further study of the relationship needs more and adequate data.

In this paper, we investigate the uncertainty of fractal characteristics in river networks along with the mathematical dealing processes, with the box-counting dimension as an example due to its simplicity, and thus popularity in the hydrological literature. Previous work has discussed this method itself and the effects of threshold area and box size in its calculation. However, there is still a lot of work that needs to be done in this area. The paper first presents a comparative study using a theoretical fractal to validate the result. Then it takes the Chabagou watershed as an example, and focuses on the investigation of the factors determining the box-counting dimension.

STUDY AREA AND NUMERICAL ANALYSIS METHOD

The analysis is based on the case study in the Chabagou watershed, a branch of the Yellow River basin. The Chabagou watershed is located at longitude 109°47′E and latitude 37°31′N, with a total area of 187 km^2. Figure 1 shows the DEM map of the study area with the resolution of 50m.

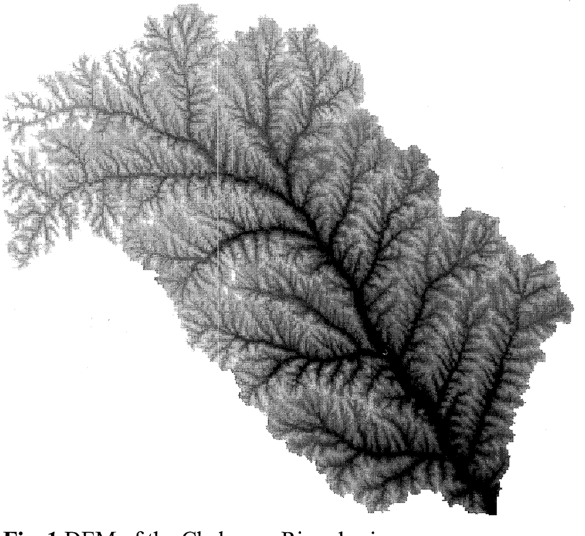

Fig. 1 DEM of the Chabagou River basin.

To calculate the box-counting dimension for a fractal F, we first assume this fractal lies on an evenly-spaced grid, and count the number of boxes in the grid required to cover the set. The box-counting dimension is then calculated by measuring how this number changes as we make the grid finer. Let r be a positive number and F a non empty and bounded subset of a metric space (X, d). Suppose that $N(r)$ is the number of boxes with side length r required to cover the set F. As $r \to 0$, if $\log N(r)/\log(1/r)$ converges to a finite value, then the box-counting dimension is defined as:

$$D_{box} = \lim_{r \to 0} \frac{\log N(r)}{\log(1/r)} \tag{1}$$

The box-counting dimension is used in application for its compatibility in a numerical experiment by the definition. A sketch of such a numerical algorithm, described by Lovejoy *et al.* (1987), is as follows. First, put set F into a grid, which is divided into boxes of size r. Second, count the number of boxes $N(r)$ that actually cover F. In practice, r is chosen from a finite range in a geometric series, and $N(r)$ is obtained correspondingly. Finally, make a log $N \sim$ log $1/r$ plot and the slope is the box-counting dimension.

According to this method, we set up the following steps to approach the box-counting dimension of the Chabagou River. (1) Extract the river networks from the DEM map of the Chabagou watershed, with different thresholds of confluence areas; the threshold of confluence area is a critical factor in defining a river network. The smaller the threshold area is, the denser the river network will be. Consequently it strongly affects the characteristics of river networks, including the fractal dimension. (2) Generate boxes with different sizes, and use them to cover the river networks. Count the number of boxes $N(r)$ which contain the river segments. The choice of sizes, e.g. the range and the increment ratio of r, will impact the final result. In principle, an infinitesimal increment ratio of r and an infinitely large range are necessary to get the true fractal dimension. In order to reduce errors resulting from imperfectness of the range and increment ratio of r, we used small increment ratio and large ranges. Meanwhile, we also investigate the effects of their variation on the fractal dimension. (3) Generate a logarithmic plot, with the slope showing the box-counting dimension of river networks.

Before the detailed calculation of the box-dimension of the Chabagou River, it is necessary to verify that our numerical analysis method is technically correct. Hence as a control example, a treelike model, which is an ideal fractal and an approximate representation of natural river networks (mentioned by Nilora, 1993), is taken to estimate the box-counting dimension using this method. This ideal fractal network is generated using a simple iterative algorithm (cf. Fig. 2). According to the iterative algorithm, it shows that the real box-counting dimension is log5/log3 = 1.46. Then we calculate its box-counting dimension by the same method, which will be applied in the Chabagou River networks.

In Fig. 3, we find a nearly perfect straight linear fit of data points on the log $N \sim$ log $(1/r)$ plot, with slope 1.43 being consistent with the above theoretical prediction. Even though the coefficient of determination is up to 0.99, the box-counting dimension varies with different ranges of box sizes. This can be explained as follows: when the box size is smaller than the smallest bifurcation of this ideal fractal simulation, its shape is not fractal anymore because its bifurcation ends here. In this case the fractal dimension measured by those boxes will become one. Therefore the slope is flatter when the box size is smaller. For the same reason, the box-counting dimension we estimate is slightly less than the real one.

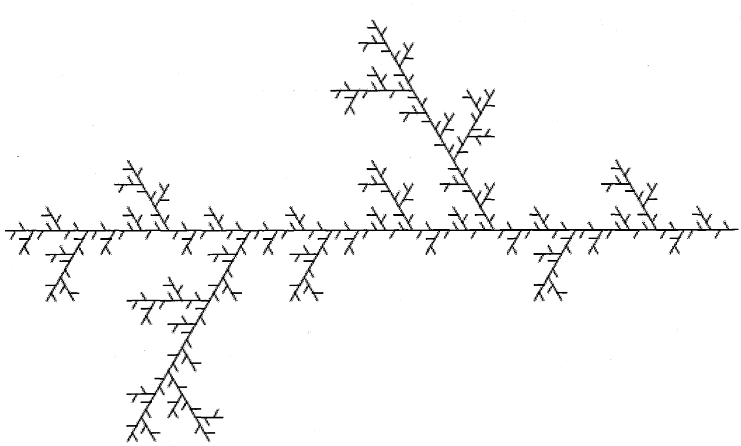

Fig. 2 Ideal fractal network with D = 1.46.

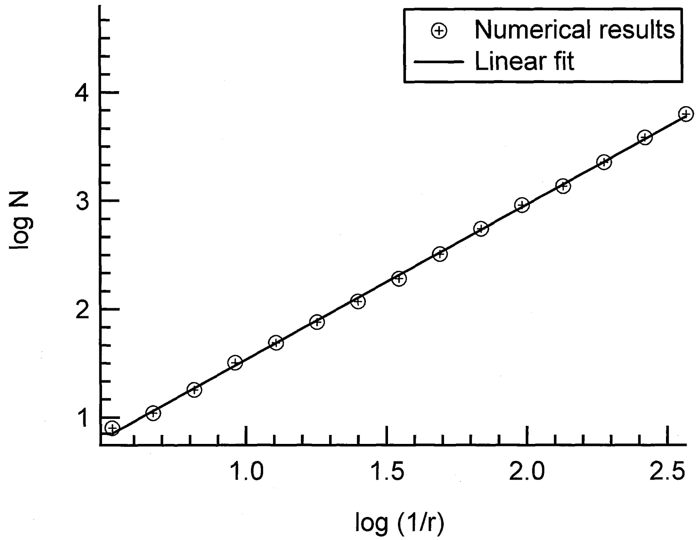

Fig. 3 Box-counting dimension of the ideal fractal with six iterations. The fitted line is log N = 1.43 log(1/r) + 0.10 with R^2 = 0.999.

RESULTS AND DISCUSSIONS

In this method, there are three critical factors in calculating box-dimension of river network. The first one is the threshold of the confluence area, which is determined while generating the river networks. The other two are the increment ratio and range of the boxes sizes. Our work focuses on how each factor influences the calculated box-counting dimension of river networks, and furthermore figuring out how to mitigate the uncertainty imposed by the three factors.

Relationship between the increment ratio of box sizes and the box-counting dimension

According to the definition of box-counting dimension, as the box size r converges to zero, the limitation of the formula requires r to decrease continually, and the increment ratio of box sizes should be as small as possible. However, in numerical analysis, the increment ratio can not be too small. Hence, to study the relationship between the increment ratio and the box-counting dimension, we should try to find out the appropriate increment ratio and exclude the influence of this factor.

From Table 1, it is shown that in a certain range the ratio of box sizes impacts little on the result of the ideal fractal. The result values fluctuate when the increment ratio of box sizes is over 2, but is identical for any ratio smaller than 2. When the ratio is smaller than 1.4, the error is less than 1%. Therefore, as long as the ratio is less than two, the effect of discreteness of box lengths can be neglected.

Relationship between the range of the boxes sizes and the box-counting dimension

Take the network with threshold area equal to 150 as an example. The range of box sizes is from 40 to 4000. Figure 4 shows that the slope changes with the variation of the range of boxes. For

Table 1 Relationship between the increment ratio of box sizes and the box-counting dimension.

Increment ratio of box sizes	1.10	1.21	1.46	2.74
Box-counting dimension	1.24	1.25	1.27	1.28
One standard deviation	0.01	0.01	0.02	0.04

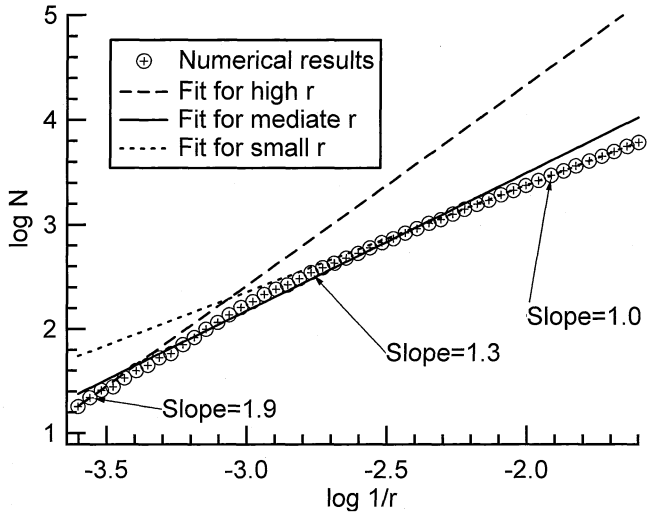

Fig. 4 Relationship between the range of box sizes and the box-counting dimension.

the smallest box size the slope of the curve approaches to 1, indicating D = 1, which is reasonable for the individual stream links consisting of straight lines. For the largest box sizes the slope will approach slope = 2 because all the points fall in only a few boxes. It can be concluded that the slope of the curve depends on the range of the boxes size, and it changes monotonically and slowly from 1 to 2, which indicates the box-counting dimension is monotonic to the range of calculated box size. This phenomenon also occurs when calculating the box-counting dimension of river networks with other threshold areas.

Theoretically, only when the size of boxes converges to 0, can one get the true value of the box-counting dimension. However, in numerical analysis, the box size can not really converge to 0 in mathematical terms. Our study suggests that in numerical analysis, smaller boxes can not necessarily bring more accurate results; in contrast, they sometimes make the results less accurate or even incorrect. Thus, the size of boxes does not necessarily need to converge to 0 in this numerical analysis.

It can be illustrated that the slope of the curve depends on the range of the box size, and it changes slowly from 1 to 2. The analysis shows that there is a theoretical causation of the unreliability of the box-counting method: the box length could only be limited discrete values and could not be less than the pixel length. This superficially dissatisfies the requirement of the definition of the box-counting dimension: the box-counting dimension is only reached when the box length is limited to zero. In contrast, at a scale smaller than the pixel, the generated river network is no longer fractal, which is natural for simulation at the pixel scale. Although the above preliminary conclusions are addressed, we cannot conclude from the analysis which range should be used to obtain the true fractal dimension unless further studies are carried out.

Relationship between the confluence area and the box-counting dimension

In this part, we calculate the box-counting dimension of Chabagou River networks, which is derived with threshold area from 22 to 7000. Results show they decrease as the threshold area increases.

From Fig. 5(a) and (b), it is found that the values of the box-counting dimensions vary according to the support area thresholds. Here the question arises of "identifying the exact value of the box-counting dimension of the real river channel". For the increasing application of fractal dimension as a parameter in hydrological responses and catchments geomorphology, etc. it is of consequence to answer this question.

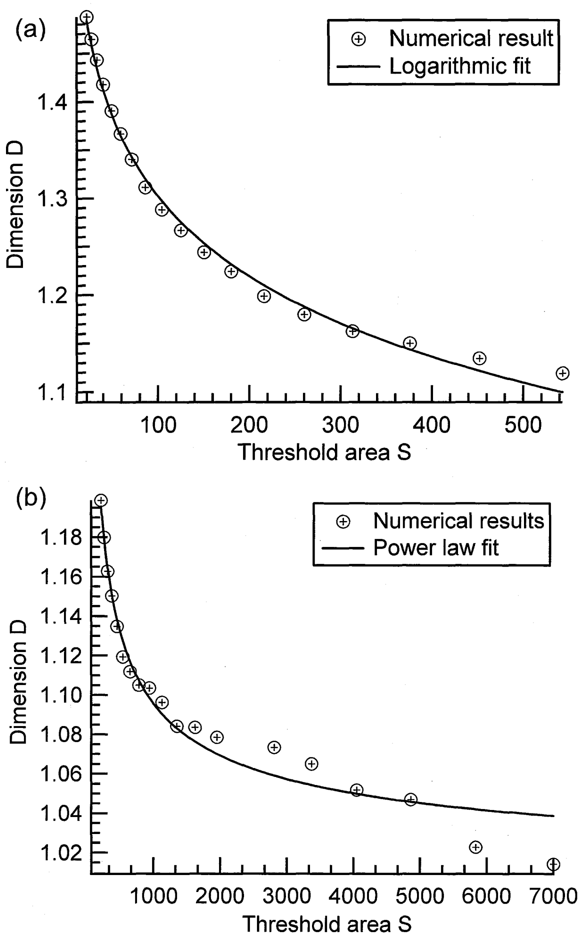

Fig. 5 Relationship between the threshold area and the box-counting dimension: (a) when threshold area is small, the fitting is $D = -0.12\ln S + 1.85$ with R^2=0.994; (b) when threshold area is large, the fitting is $D = 2.9S^{-0.50}$ with $R^2 = 0.87$.

For different threshold areas, the behaviour of the change is different, and therefore should be discussed separately. When the threshold area is small (e.g. from 22 to 546), there is a logarithmic fit (cf. Fig. 5(a)) between the threshold area and box-counting dimension. The relationship can be expressed as $D = -0.12\ln S + 1.85$, with the coefficient of determination equal to 0.99. However, from the figure it is noticed that when S becomes large, the fitting line falls below the true data value. It is probably because the logarithmic function diverges when S becomes large, while the box-counting dimension converges to 1. Hence, when the threshold area is large (e.g. from 546 to 7000), another behaviour is presented, which can be fit into a power law (cf. Fig. 5(b)) relation $D = 2.9S^{-0.50}$. The results are consistent with the observation which indicates that when the threshold area increases, the box-counting dimension decreases to be closer to 1.

If the fittings are appropriate, we can obtain the box-counting dimension D of a certain network by S (the threshold area which defined the network). Also, based on our fittings, we carry on theoretical analysis in the future to provide the possibility of investigating the real fractal river network according to the relationship between S and D, which could eliminate the effect of threshold area. Figure 5(a) shows the logarithmic relationship between the values of box-counting dimension and the support area thresholds, which may be demonstrated by the multifractal theory.

Our present work is only an exploratory approach, and further investigation remains to be done. Additional data from other natural rivers are needed in future investigations.

CONCLUSION

In the present work we have applied the digital elevation model (DEM) to analyse the box-counting dimension of the Chabagou River channel. The results obtained in this work support the observation of the complicated nature of fractal river networks and the sensitivity of the character of river to the threshold of confluence area. (1) The box-counting dimension is monotonic to the range of box sizes while insensitive to the increment ratio of box sizes. (2) The box-counting dimension has a logarithmic dependence on S; $D = -0.12\ln S + 1.85$ for smaller threshold areas; it has a power law asymptotic behaviour $D = 2.9S^{-0.50}$ for larger threshold areas.

In this paper an approach is proposed to evaluate the sensitivity of threshold area in box-counting dimension calculations, and it is expected that real box-counting dimensions can be identified from the calculation. However, dimensions derived from this method are not accurate enough because of its dependence on the discreteness of box sizes. Meanwhile, considerable uncertainty coming from the fitting and calculation are added together and make the final result more inaccurate than we can accept. Moreover, these uncertainties are intrinsic and cannot be eliminated by technical advances such as use of a finer grid.

On the other hand, our research is based on the assumption that the river is mono-fractal. Nevertheless, there is all likelihood that the nature river networks are multifractal. No matter how mono-fractal or multifractal the river network is, the problem of scaling always exists.

The results presented in this paper only touch upon the geometrical properties of the river network, which may serve as a foundation for the analysis of scaling properties of hydrophysical and hydrological processes.

Acknowledgements The authors would like to thank Professor Dawen Yang for his valuable comments. This research was supported by the National Natural Science Foundation of China (project no. 50779022).

REFERENCES

Helmlinger, K. R., Kumar, P. & Foufoula-Geogriou, E. (1993) On the use of digital elevation model data for Hortonian and fractal analyses of channel networks. *Water Resour. Res.* **29**(8), 2599–2613.

Horton, R. E. (1945) Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology. *Geol. Soc. Am. Bull.* **56**, 275–370.

Huang, Y. (1996) How to get a reliable fractal dimension by box-counting. *Communications in Nonlinear Science* **2**, 22–27.

Ichoku, C., Karnieli, A. & Verchovsky, I. (1996) Application of fractal techniques to the comparative evaluation of two methods of extracting channel networks form digital elevation models. *Water Resour. Res.* **32**(2), 389–399.

Kusumayudha, S. B., Zen, M. T., Notosiswoyo, S. & Gautama, R. S. (2000) Fractal analysis of the Oyo River, cave systems, and topography of the Gunungsewu karst area, central Java, Indonesia. *Hydrogeology J.* **8**, 271–278.

La Barbera, P. & Rosso, R. (1989) On the fractal dimension of stream networks. *Water Resour. Res.* **25**(4), 735–741.

Lovejoy, S., Schertzer, D. & Tsonis, A. A. (1987) Functional box counting and multiple elliptical dimensions in rain. *Science* **235**, 1036–1038.

Mandelbrot, B. B. (1983) *Fractal Geometry of Nature*. W. H. Freeman, New York, USA.

Nikora, V. I. & Sapozhnikov, V. B. (1993) River network fractal geometry and its computer simulation. *Water Resour. Res.* **29**(10), 3569–3575.

Radziejewski, M. & Kundzewicz, Z. W. (1997) Fractal analysis of flow of the River Warta. *J. Hydrol.* **200**, 280–294.

Rodríguez-Iturbe, I. & Rinaldo, A. (1997) *Fractal River Basins: Chance and Self-Organization.* Cambridge University Press, New York, USA.

Veltri, M., Veltri, P. & Maiolo, M. (1996) On the fractal description of natural channel networks. *J. Hydrol.* **187**, 137–144.

Yang, D., Herath, S. & Musiake, K. (1997). Analysis of stability of river network geomorphologic parameters using DEM data of Japanese catchments. In: *Proc. 27th Congress of the Intl. Association for Hydraulic Research*, vol. 2, 973–978.

Yang, D., Herath, S. & Musiake, K. (2001). Spatial resolution sensitivity of catchment geomorphologic properties and the effect on hydrological simulation. *Hydro. Processes* **15**, 2085–2099.