Mathematical characterization of fractal river networks

Author: Akira Terui[1,*]

Email: hanabi0111@gmail.com

[1]Department of Biology, University of North Carolina at Greensboro, Greensboro, NC 27410, USA

1 Maintext

Rivers form complex branching networks, and the ecological implication of river network complexity has gained a great interest over the past few decades[1–3]. To this end, there have been concerted efforts to construct virtual river networks, aiming to provide theoretical insights into how river network structure controls riverine ecological dynamics. Carraro and Altermatt[4] made a great contribution to this research area by comparing the scaling properties of virtual river networks produced by three different simulation methods - balanced binary trees (BBTs)[5], random branching networks (RBNs)[6,7], and optimal channel networks (OCNs)[8]. The first two methods have two parameters that control river network size (order or the number of nodes) and complexity (branching probability). In the meantime, OCNs are spanning trees that suffice a local minimum of a function describing total energy expenditure across the network. Carraro and Altermatt[4] performed extensive analysis and concluded that: (1) OCNs best represent the scaling properties of river networks, such as Horton's law and the exceedance probability of catchment area; (2) branching ratio is a scale-dependent quantity as the value changes across spatial resolutions at which river networks are extracted (expressed as the threshold catchment area A_T that initiates channels or pixel size l); (3) OCNs most accurately predicted the metapopulation stability and capacity. I agree that OCNs are capable of reproducing the scaling properties of real river networks. However, the supporting ground for the rest of the conclusions is, unfortunately, seriously flawed or rather inconclusive due to the improper mathematical definition or use of scale invariance, dimensions, and units.

First, the term "scale invariance" was falsely used in their article. To explain this, let y^* and y be the perimeter length of a perfect circle and the length of a coastline, respectively, measured as multiples of a ruler with a unit length x (the "observation scale") – a classical comparison between "scale-dependent" and "scale-invariant" objects used in Mandelbrot[9]. The approximated length measured by the ruler is $q(x)x$, where $q(x)$ is the number of dividing steps with a given ruler length x. In the case of perimeter length y^* (scale-dependent), one will obtain the exact value as the ruler length x approaches zero because a sufficiently short ruler can characterize the smooth curve of a perfect circle (the existence of characteristic scale). Clearly, $\lim_{x \to 0} q(x)x = 2\pi r$ (r is the radius) for the perimeter of a perfect circle. In contrast, the length of a complex coastline y (scale-invariant) reaches infinity as the scale x approaches zero because a shorter ruler can capture similar, but finer-scale complexity that is ignored by a longer ruler (the lack of characteristic scale; see Figure 1 in Mandelbrot[9] for the graphical comparison between the two types of objects). This property, i.e., the part is a reduced-scale image of the whole, is referred to as "statistical self-similarity" that underpins the mathematical characteristics of scale-invariance in the coastline length y. Such a scale-invariant structure can be eptiomized by the power-law function[10,11]:

$$y = f(x) = cx^z$$

where c is the scaling constant and z is the scaling exponent. The scale invariance can be easily proved by multiplying the scaling factor λ:

$$y' = f(\lambda x) = c\lambda^z x^z = \lambda^z f(x) = \lambda^z y$$

Thus, the observed object $y = f(x)$ is said to be scale-invariant because the multiplicative extension/shrink of observation scale x results in the same shape of the original object y but with a different scale[12].

Carraro and Altermatt[4] provided some evidence that branching ratio p_r follows a power law along the axes of observation scale A_T and pixel size l (i.e., length on a side) using OCNs (Equation 1 in the original article):

$$\mathrm{E}(p_r) \approx 1.531 A_T^{-0.523} A^{-0.032}$$

where A is the entire watershed area expressed in the number of pixels; thus, $A = A'l^{-2}$ (A' is the watershed area measured in the unit of l^2, such as km^2). Also, the authors empirically showed "visual" examples that the relationship between p_r and A_T follows a power-law using the data from 50 real river networks (Figure 3a in the original article). Provided that this relationship statistically holds true (see the next section for this issue), branching ratio has a property of scale-invariance. Despite this, the authors claimed that "*Here we show that an alleged property of such random networks (branching probability) is a scale-dependent quantity that does not reflect any recognized metric of rivers' fractal character…*" (Abstract) This interpretation is the opposite as it has been defined in the literature of scale-invariance[10–13], including the author's previous publication[14]. Therefore, the sentences that build upon this interpretation are inappropriate and misleading (**Table 1**). Importantly, many of these sentences are the concluding sentence of a paragraph. As such, the progression of supporting sentences in these paragraphs is no longer logical, unfortunately.

Second, dimensions and units are improperly treated in their analysis. A dimension is the power of an axis along which a physical quantity is measured, and a unit is a way to assign a number to a particular dimension to make it relative. For example, length is a dimension and meter is a unit of length. Throughout the article, the authors used the number of pixels to measure the river length N, the total catchment area A, and the threshold catchment area A_T. There is no issue to use the number of pixels as a unit. However, a critical problem in their article is that they obscured the dimensions of pixels. For example, they made very unclear that the river length N has a dimension of pixel "length", while the total catchment area A and the threshold catchment area A_T have a dimension of pixel "area". In particular, the authors incorrectly defined p_r as a dimensionless quantity (Methods) despite its unit is [pixel length^{-1}] with a dimension of length.

The improper use of units and dimensions causes serious problems in the analysis of "real" river networks; the authors used different pixel sizes (length on a side) among watersheds (103 m to 1268 m; Supplementary Table 2 in the original article), meaning that the same number of pixels translates into very different lengths and areas. For example, in Figure 3 in the original article, the observation scale A_T ranges $20 - 500$ pixels. This pixel range translates into $0.2 – 5.3$ km^2 for the Toss watershed (with smallest pixel size) and $32.2 – 803.9$ km^2 for the Stikine watershed (with largest pixel size). Further, the branching ratio p_r is also affected by this variation in pixel size as its unit is [pixel length^{-1}]. When the pixel length was converted to a unit of km, the branching ratio represents the number of links per 0.1 km in Toss, while it represents the number of links per 1.3 km in Stikine. Hence, the authors compared values that are uncomparable.

To explore the consequences of the improper use of pixels, I re-analyzed their data with MERIT Hydro that has a constant pixel size of 3 arc-second (~90 m at the equator) across the globe[15]. I extracted river networks in R[16] using the R package `whitebox` with ten values of A_T [km^2] ($A_T = 1, ..., 1000$ with an equal interval at a $\log_{10}$ scale). Following the author's definition, I calculated branching ratio p_r [km^{-1}] as $\frac{N_L}{L}$ (N_L the number of links in a network [dimensionless]; L the total river length [km]) for each river network extracted at a given observation scale A_T. In the original article, the authors did not perform any statistical analysis despite the fact that the estimates of p_r are subject to statistical uncertainty, just like any other geospatial layers and ecological measures. Instead, they "picked" specific values of p_r to claim that the rank of p_r varies across scales (Table 1). To avoid such a subjective argument, and to statistically substantiate the power-law of p_r along the axis of A_T, I fitted the following log-linear models (i represents an individual data point of p_r estimated in a given watershed at a given scale A_T).

$$\log_{10} p_{r,i} = \log_{10} c_{w(i)} + z \log_{10} A_{T,i} + \varepsilon_i \qquad \text{(M0)}$$
$$\log_{10} p_{r,i} = \log_{10} c_{w(i)} + z_{w(i)} \log_{10} A_{T,i} + \varepsilon_i \quad \text{(M1)}$$

where ε_i is the normally-distributed error term. The first model (M0) assumes the "universal" scaling with the constant scaling exponent z across watersheds; i.e., the branching ratio at all the 50 watersheds follows the same power law with the watershed-specific scaling constant $\log_{10} c_{w(i)}$ ($w(i)$ is watershed w for a data point i). The second model (M1) assumes the "local" scaling with the watershed-specific scaling exponent $z_{w(i)}$. I estimated the evidence ratio of the two models using the approximated Bayes Factor (BF)[17], which is defined as $\mathrm{BF} = \exp\left(\frac{\mathrm{BIC(M1)} - \mathrm{BIC(M0)}}{2}\right)$. In this definition, a value of BF > 1 gives the support for the universal model M0 over M1; for example, if BF = 2, the model M0 is twice as likely as the model M1.

The analysis provided decisive support for M0 with BF = 7.55×10^{18}. In addition, the "local" scaling exponent in M1 added little explanatory capacity to M0, as indicated by the almost identical values of adjusted R^2 (0.996 for M0 vs. 0.997 for M1). Consequently, as evident from the model structure of M0, the rank of the expected branching ratio $\mathrm{E}(\log_{10} p_r)$ never changed across scales, maintaining the rank of the watershed-specific scaling constant $\log_{10} c_{w(i)}$ (Figure 1; see regression lines). This result is inconsistent with the author's statement "*by extracting different river networks at various scales (i.e., various A_T values) and assessing the rivers' rank in terms of p_r, one observes that rivers that look more "branching" (i.e., have higher p_r) than others for a given A_T value can become less "branching" for a different A_T value (Fig. 3).*" I also must note that I did not find any significant correlation between watershed area A [km^2] and branching ratio p_r [km^{-1}] when extracted with a constant value of $A_T = 1$ km^2 across watersheds (Spearman's rank correlation = -0.16; p-value = 0.28), as opposed to "*…if different river networks spanning different catchment areas (say, in km2) are compared, all of them extracted from the same DEM (same l and same AT in km2), then the larger river network will appear more branching (i.e., have larger p_r)*"[4]. The lack of correlation has also been reported in previous studies[6,7]. These author's statements merely reflect the lack of appropriate quantitative analysis and/or a statistical artifact of inconsistent units across watersheds.

It is worthwhile to note that the above argument remains valid even when the rank of "dimensional" branching ratio changes across scales due to the existence of local scaling laws (i.e., line crossing can occur in Figure 1A when the scaling exponent z [i.e., slope in a log-log space] varies by watershed). If scaling exponent z varies by watershed, the dimensional p_r is no longer comparable across watersheds because the dimension of p_r ($= 2z$) varies by watershed (just like one cannot compare length and area). It is crucial for this argument to recognize that, by writing $p_r = c_w A_T^z$, we equate the dimension in both sides of the equation. The dimension of A_T is two [km^2]. Hence, the dimension of p_r is $2z$ since the scaling exponent z applies to the unit of A_T as $p_r = c_w A_T^z$ [km^{2z}]. Typically, branching ratio p_r, or the inverse of the average link length by definition, has a dimension of greater than one because of the fractal nature of river networks, as the author's analysis of OCNs indicated ($2 \times 0.52 = 1.04$). As such, to compare branching ratio across watersheds, one must non-dimensionalize p_r as $\bar{p}_r = \frac{p_r}{A^z} = c_w$ (see Figure 1B). As evident from the equation, the non-dimensionalized branching ratio $\bar{p}_r$ is constant across scales (c_w) with no possibility of changes in rank of the expected value. This technique of dimensional analysis has been widely used when comparing the self-similar structure of scale-invariant objects (see Figure 1 in Rinaldo et al.[13] and equation (2.2) in Rodriguez-Iturbe and Rinaldo[10] for examples). In fact, some past studies used $A_T = 1$ km^2 such that the estimated dimensional p_r approximates the non-dimensional $\bar{p}_r$ ($p_r \approx c_w$ when $A_T = 1$ in a given unit regardless of the value of z)[6,7]. Therefore, when z varies by watershed, the comparison of dimesional branching ratio p_r is meaningless (unless properly rescaled as in Terui et al.[6,7]). Perhaps, a valid implication from this re-analysis is that one should use small values of A_T (e.g., 1 km^2) as p_r fluctuate more unpredictably at coarser observation scales (larger A_T; Figure 1). This is simply because of a small sample size (i.e., the number of links) at coarser resolutions, which inflates the statistical uncertainty. Yet, this is not a problem unique to branching ratio. Any geospatial layers (e.g., land use) contain statistical uncertainty, which should increase with observation scales.

The issue of dimension/unit is pertinent to their metapopulation simulations as they used pixel size l as a unit of "population scale" and "dispersal distance". This makes the results of their metapopulation simulations hardly interpretable. For example, they assumed a population scale of 103 m in the Toss watershed to 1268 in the Stikine watershed. A reasonable interpretation of this simulation setup is that the authors simulated metapopulation dynamics of "different" species among watersheds. This is not trivial as the population scale defines the number of local populations within a metapopulation – one of the most influential parameter dictating metapopulation CV (see Equations (2) and (3) in the original article). Their estimates
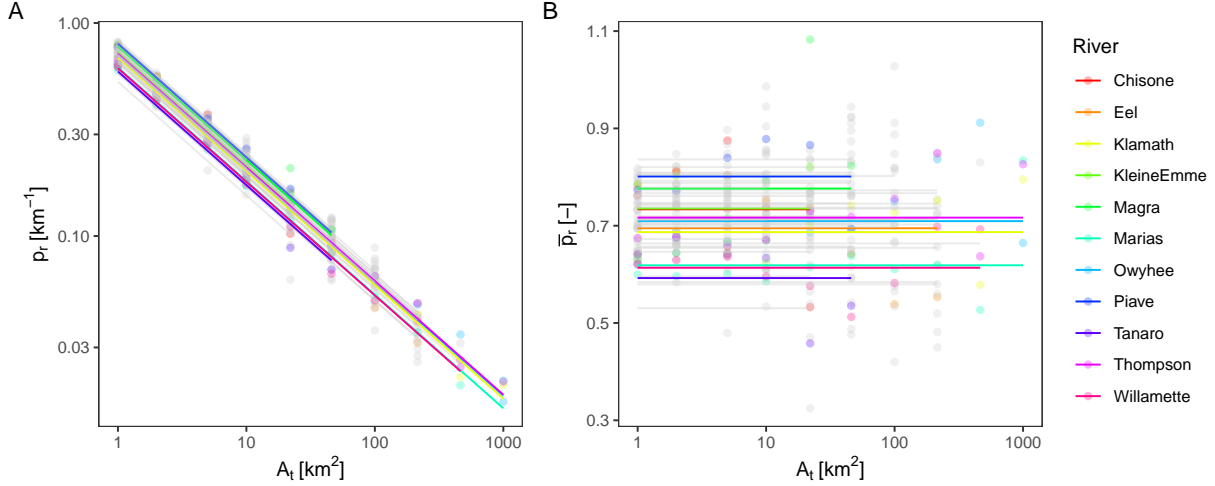
Figure 1: (A) Power law scaling of brnaching ratio. Dots are individual data points. Lines are predicted values from the model M0 (i.e., the expected value of branching ratio). (B) Non-dimesionalized branching ratio. Dots are individual data points of branching ratio after proper rescaling. Lines are predicted values from the model M0 with proper rescaling.

of metapopulation CV is therefore confounded by the difference of simulated species' trait (i.e., population scale). Similarly, the average dispersal distance α was measured in the unit of [pixel length], meaning that it varies from 1.03×10^4 m and 1.268×10^5 m when converted to the unit of meter. It is difficult to envision that such a huge variation exists within the same species. Again, this parameter has a strong influence on metapopulation CV and capacity (see Methods in the original article). Thus, Figure 6 and Supplementary Figures $4-9$ of the original article are ecologically uninterpretable. Although I do not know how this alters one of the main conclusions "*OCNs most accurately predicted the metapopulation stability and capacity*", it is clear that a complete full re-analysis is desired to yield "interpretable" results.

Carraro and Altermatt[4] offered an important perspective to the use of the different classes of virtual river networks. Therefore, a full re-interpretation/re-analysis is warranted to provide valid insights into how we assess river network structure.

1.1 Data availability

There is no new data associated with this manuscript. Codes are available at https://github.com/aterui/public-proj_carraro-cee

Reference

1. Grant, E. H. C., Lowe, W. H. & Fagan, W. F. Living in the branches: Population dynamics and ecological processes in dendritic networks. *Ecology Letters* **10**, 165–175 (2007).

2. Carrara, F., Altermatt, F., Rodriguez-Iturbe, I. & Rinaldo, A. Dendritic connectivity controls biodiversity patterns in experimental metacommunities. *Proceedings of the National Academy of Sciences* **109**, 5761–5766 (2012).

3. Mari, L., Casagrandi, R., Bertuzzo, E., Rinaldo, A. & Gatto, M. Metapopulation persistence and species spread in river networks. *Ecology Letters* **17**, 426–34 (2014).

4. Carraro, L. & Altermatt, F. Optimal Channel Networks accurately model ecologically-relevant geomorphological features of branching river networks. *Communications Earth & Environment* **3**, 1–10 (2022).

5. Yeakel, J. D., Moore, J. W., Guimarães, P. R. & de Aguiar, M. A. M. Synchronisation and stability in river metapopulation networks. *Ecology Letters* **17**, 273–283 (2014).

6. Terui, A. *et al.* Metapopulation stability in branching river networks. *Proceedings of the National Academy of Sciences* **115**, E5963–E5969 (2018).

7. Terui, A., Kim, S., Dolph, C. L., Kadoya, T. & Miyazaki, Y. Emergent dual scaling of riverine biodiversity. *Proceedings of the National Academy of Sciences* **118**, e2105574118 (2021).

8. Carraro, L. *et al.* Generation and application of river network analogues for use in ecology and evolution. *Ecology and Evolution* **10**, 7537–7550 (2020).

9. Mandelbrot, B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* **156**, 636–638 (1967).

10. Rodríguez-Iturbe, I. & Rinaldo, A. *Fractal River Basins: Chance and Self-Organization.* (Cambridge University Press, 2001).

11. Brown, J. H. *et al.* The fractal nature of nature: Power laws, ecological complexity and biodiversity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357**, 619–626 (2002).

12. Proekt, A., Banavar, J. R., Maritan, A. & Pfaff, D. W. Scale invariance in the dynamics of spontaneous behavior. *Proceedings of the National Academy of Sciences* **109**, 10564–10569 (2012).

13. Rinaldo, A., Rigon, R., Banavar, J. R., Maritan, A. & Rodriguez-Iturbe, I. Evolution and selection of river networks: Statics, dynamics, and complexity. *Proceedings of the National Academy of Sciences* **111**, 2417–2424 (2014).

14. Giometto, A., Altermatt, F., Carrara, F., Maritan, A. & Rinaldo, A. Scaling body size fluctuations. *Proceedings of the National Academy of Sciences* **110**, 4646–4650 (2013).

15. Yamazaki, D. *et al.* MERIT hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research* **55**, 5053–5073 (2019).

16. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2021).

17. Wagenmakers, E.-J. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review* **14**, 779–804 (2007).