

## Probabilistic structure of the distance between tributaries of given size in river networks

Matteo Convertino,<sup>1</sup> Riccardo Rigon,<sup>2</sup> Amos Maritan,<sup>3</sup> Ignacio Rodriguez-Iturbe,<sup>4</sup> and Andrea Rinaldo<sup>1</sup>

Received 16 May 2007; revised 1 August 2007; accepted 9 August 2007; published 24 November 2007.

[1] We analyze the distribution of the distances between tributaries of a given size (or of sizes larger than a given area) draining along either an open boundary or the mainstream of a river network. By proposing a description of the distance separating prescribed merging contributing areas, we also address related variables, like mean (or bankfull) flow rates and channel and riparian area widths, which are derived under a set of reasonable hydrologic assumptions. The importance of such distributions lies in their ecological, hydrologic, and geomorphic implications on the spreading of species along the ecological corridor defined by the river network and on the propagation of infections due to water-borne diseases, particularly in view of exact theoretical predictions explicitly using the alongstream distribution of confluences carrying a given flow. Use is made here of real river networks, suitably extracted from digital elevation models, optimal channel networks, and exactly solved tree-like constructs like the Peano and the Scheidegger networks. The results obtained redefine theoretically in a coherent and general manner and verify observationally the distribution function of the above distances and thus provide the general probabilistic structure of tributaries in river networks. Specifically, we find that the probability of exceedence of the alongstream distance  $d$  of tributaries of size larger than  $a$  has the explicit form  $P(\geq d) = \exp(-Cd/a^{H/(1+H)})$ , where  $C$  is a constant that depends on the choice of boundary conditions and  $H \leq 1$  is the Hurst exponent.

**Citation:** Convertino, M., R. Rigon, A. Maritan, I. Rodriguez-Iturbe, and A. Rinaldo (2007), Probabilistic structure of the distance between tributaries of given size in river networks, *Water Resour. Res.*, 43, W11418, doi:10.1029/2007WR006176.

### 1. Introduction

[2] The role of the structure of river networks in modeling hydrochory (the spreading of species along ecological corridors), human-range expansions (i.e., predicting how populations migrate when settling into new territories), or the spreading of water-borne infections has been recently recognized through quantitative models of dispersion along fractal networks coupled with proper reactions at their nodes [Méndez *et al.*, 2004; Campos *et al.*, 2006; Muneeppeerakul *et al.*, 2007a, 2007b; Bertuzzo *et al.*, 2007]. An essential ingredient therein is the fact that settlers, infective agents or colonizing species do not occupy all the territory isotropically, but rather follow river patterns as general corridors and pathways. Among the key results, it was argued in a quan-

titative manner that landscape heterogeneities play an essential role in several physical and biological processes of great social, biological or physical importance that involve river basins as substrate [Ammermann and Cavalli-Sforza, 1984; Holmes, 1993; Fort and Méndez, 1999; Campos *et al.*, 2006; Muneeppeerakul *et al.*, 2007a, 2007b; Bertuzzo *et al.*, 2007].

[3] One interesting corollary of any analysis of migration/spreading fronts is the central role of the network structure acting as the substrate for wave propagation. Specific and reliable structural models thus need to be employed, as well as the probabilistic description of key random variables affecting the process. Empirical representations of natural forms in the river basin and their mathematical models, possibly addressing their fractal nature, have been thoroughly studied in the past [e.g., Rodriguez-Iturbe and Rinaldo, 1997]. We are specifically concerned here with features that might affect spreading processes on networks, like for instance the type and strictness of self-similar properties appearing at different scales of observation, or the looplessness of the structure [Rinaldo *et al.*, 1998, 2006]. The type of self-similarity observed for riverine trees is particularly important. In fact, whether or not one would observe the same distribution of confluences regardless of the stream chosen as receiving water body (see Figure 1), is central to the generality of the distributions studied herein, and deeply related to the similarity of the parts and the whole embedded in the fractal form [Rodriguez-Iturbe and Rinaldo, 1997]. Moreover, similarity is allowed only in a

<sup>1</sup>Dipartimento di Ingegneria Idraulica, Marittima Ambientale e Geotecnica, and International Center for Hydrology “Dino Tonini”, Università di Padova, Padova, Italy.

<sup>2</sup>Dipartimento di Ingegneria Civile e Ambientale and Centro Universitario per la Difesa Idrogeologica Dell’Ambiente Montano (CUDAM), Università di Trento, Trento, Italy.

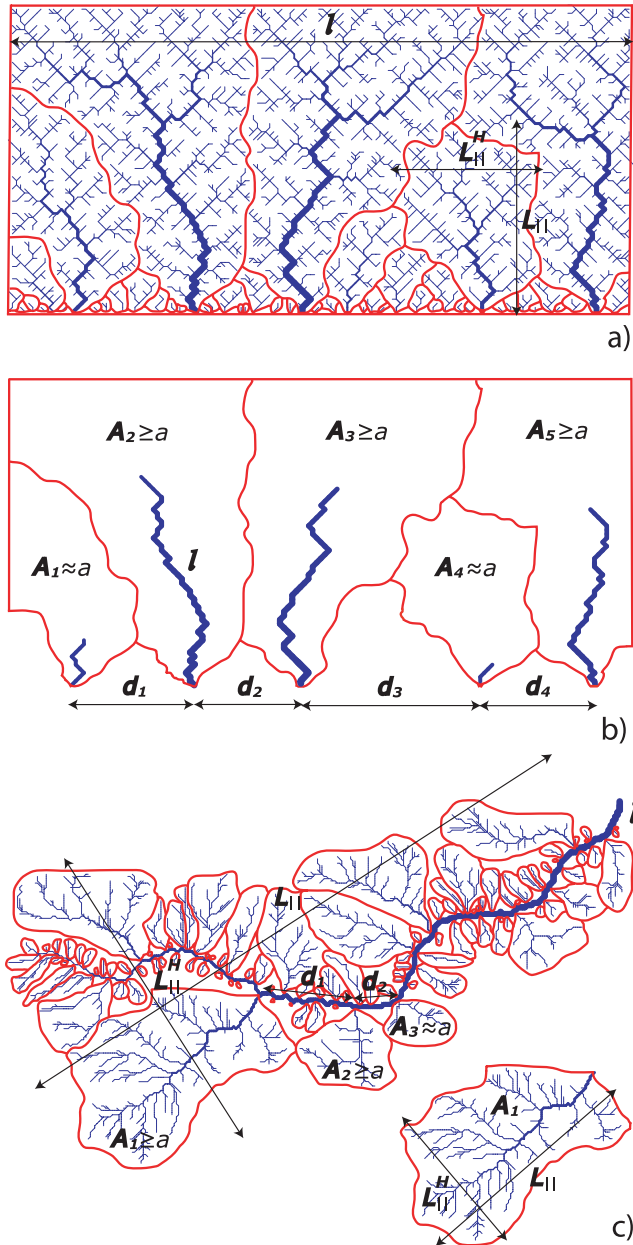
<sup>3</sup>Dipartimento di Fisica “Galileo Galilei” and Consorzio Nazionale Interuniversitario per la Scienze Fisiche della Materia (CNISM)–Istituto Nazionale per la Fisica della Materia (INFM), Università di Padova, Padova, Italy.

<sup>4</sup>Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA.

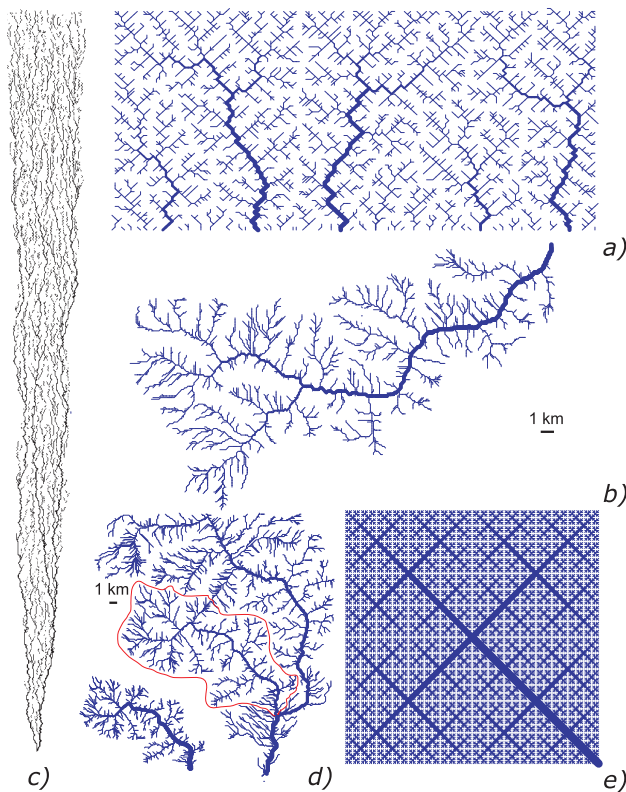
finite range of scales, as it needs proper finite-size corrections because upper and lower cutoffs in the aggregation structure, which reflect, respectively, the drainage density defining where channels begin and the loss of statistical significance of areas approaching the maximum basin size [e.g., Maritan et al., 1996; Rodriguez-Iturbe and Rinaldo, 1997]. Quite useful in this context is the use of Optimal Channel Network models [Rodriguez-Iturbe et al., 1992a, 1992b; Rinaldo et al., 1992, 1993], i.e., the robust determination of network structures that minimize total energy dissipation, that mimic quite efficiently the fluvial landforms over arbitrary ranges of scales and boundary conditions. Note that this property has also been shown to be an exact attribute of the steady state of the general (small-gradient) landscape evolution boundary value problem [Banavar et al., 2000, 2001].

[4] For general numerical calculations of the structure of distances of confluences of tributaries of a given area we shall adopt here (Figures 1 and 2): (1) the topology and

geometry of real rivers, chosen from our archives and suitably extracted from digital terrain maps of variable resolution [see, e.g., Rodriguez-Iturbe and Rinaldo, 1997] and (2) local minimum Optimal Channel networks (OCNs) [Rodriguez-Iturbe et al., 1992a, 1992b; Rinaldo et al., 1992, 1993; Rodriguez-Iturbe and Rinaldo, 1997]. They hold fractal characteristics that are obtained through a specific selection process from which one obtains a rich structure of optimal scaling forms that are known to closely conform to the scaling of real networks. To derive exact results, instead, we shall resort to Peano and Scheidegger networks. The former is a deterministic fractal [Mandelbrot, 1983] whose main topological and scaling features have been determined analytically [Marani et al., 1991; Colaiori et al., 1997], and the latter a random aggregation model of directed network [Scheidegger, 1967]. The former has been recently employed [Bertuzzo et al., 2007] to prove that geometrical constraints imposed by a fractal network imply strong corrections on the speed of migration fronts. It was noted therein that it is not surprising that Peano and OCNs lead to similar results, because the speed of the front depends chiefly on topological features that are indeed quite similar for real, OCN and Peano networks. In fact, the wave speed is affected mostly by the gross structure encountered by the front while propagating along the network, chiefly the bifurcations. Hence topology, rather than the fine structure of the subpaths, dominates the process. Because of this, we also address in detail the Scheidegger network (Figure 2), an exactly treatable tree-like pattern constructed by self-avoiding unbiased random aggregation pattern with injection [Scheidegger, 1967; Huber, 1991; Takayasu et al., 1991]. Indeed its topology and aggregation is known exactly to be rather different from real rivers and OCNs or Peano's construction. Optimal channel networks (OCNs) used in our computations are local minima (that is, dynamically accessible) of global energy dissipation, and are obtained by a



**Figure 1.** Samples of networks on which calculations are performed: (a) an OCN with cylindrical boundary conditions (i.e., periodic in the transverse direction) and a no-flux condition along the upper side. The overall dimensions are  $780 \times 78$  pixels in the case shown herein ( $m = \frac{L_{\perp}}{L_{\parallel}} = 10$ ). The lower open boundary border is sampled for tributary area in this case; here  $d_j$ ,  $H \sim 1$  (owing to the arrangement chosen, whereas single-outlet OCNs usually yield lower  $H$  and higher  $d_j$  values [see, e.g., Rigon et al., 1996] (see Table 1)). In such a case the length  $l$  of each tributary subbasin is roughly perpendicular to the draining side. The threshold area for channelization is chosen at 10 pixels. (b) The same OCN drawn with threshold area at 1000 pixels, which indeed shows tributary subbasins with area greater than  $a = 1000$ . The random distances among them are highlighted. (c) A real river network, the Tanaro basin (globally  $\sim 8000$  km<sup>2</sup>) closed at Garessio (Piemonte, Northern Italy), suitably extracted from digital terrain map (DTM), of dimensions  $178 \times 296$  pixels where we have chosen for computational reasons relatively large pixels (area  $100 \times 100$  m<sup>2</sup>). Note the particular arrangement of tributaries draining from both sides. The inset illustrates the notation where  $L = L_{\parallel}$  and  $L_{\perp} \sim L^H$  are the longitudinal and transverse length scales along the principal axes of inertia, respectively,  $H$  is the Hurst parameter, and  $l$  is the mainstream length.



**Figure 2.** Sample networks used in the comparative analysis: (a) multiple-outlet OCN already shown on Figure 1, (b) the Tanaro river basin (Po' basin, Piemonte, Northern Italy) closed at Garessio with extension of about 530 km<sup>2</sup> and mean elevation of 1382 m, (c) Scheidegger network realized as the original construct and with the parallel side strongly longer than the shorter one in grant to obtain  $H = 1/2$ , it is a subbasin extracted from the basin lattice Figure A1a, (d) Mis, closed at Mis (detected by the red drainage divide line), and Cordevole, closed at Sospirolo, are nested subbasins (Piave basin, Veneto, Northern Italy), suitably extracted from the relevant DTM ( $255 \times 271$ ), the former with extension of about 218 km<sup>2</sup> ( $126 \times 173$  pixels) and the latter of about 691 km<sup>2</sup> with mean elevation of 1721 m, and (e) Peano network at the seventh stage of the multiplicative growth process ( $258 \times 258$  pixels). For a detailed description of the specific construction rules and their references, see *Rinaldo et al.* [2006]. OCN, real basins, and Scheidegger network have been extracted with a threshold on area equal to  $A_{th} = 10$ . Each pixel of DTMs utilized have length equal to 100 m.

simulated annealing optimization procedure. With relaxed constraints like periodic boundary conditions, both for single and multiple outlets, carefully annealed OCNs exhibit a value for the exponent of the aggregated area close to the mean field one, which is known exactly to characterize the ground state and Peano [Rodríguez-Iturbe and Rinaldo, 1997].

[5] Ecological implications of the distribution of tributaries relate to the availability of flow, slope and riparian area available in defining an ecological corridor [e.g., Campos et al., 2006; Bertuzzo et al., 2007; Banavar et al., 2007; Muneepeerakul et al., 2007b]. Riparian zones, in fact, play many important roles in regulating ecosystem function

within streams, surrounding environments, and upland areas. They are usually related to the stream width or to topographically concave areas, features that inevitably relate to landscape-forming discharges and hence to the distribution of tributary areas. Spreading of species or infections on river networks also requires flow specifications defining, for instance, the bias affecting dispersion of water-borne agents or propagules. Moreover, although models of biodiversity in two-dimensional landscapes had been extensively studied [e.g., Tilman and Downing, 1994; Hubbel, 2001], only recently Muneepeerakul et al. [2007a] have shown that the combined effects of directionality of dispersal produced by the network landscape significantly alter various biodiversity patterns of vegetation communities obeying the neutral model. Be it metapopulation models coupled with the strictly hierarchical competition-colonization trade-off model to study the resulting biodiversity patterns in the river basin [Muneepeerakul et al., 2007b], or patterns of infection or migration spreading [Bertuzzo et al., 2007], it is clear the fundamental role played by the network substrate and of its properties on several ecological processes. About the purported connection with riverine ecology, suffice here to say that the distribution we seek defines analytically the distribution of the distance of mean flow rates, channel width or riparian areas alongstream. All are, in fact, proxies of cumulative basin area, and hence of tributaries' contributing area, which conditions such distribution.

[6] Three sample basins in Northern Italy of different size are investigated in detail here as source of real-world data. It is easily overlooked, in fact, that the very definition of the confluence of streams with a tributary depends on specifications like the threshold area for channelization, or the size of the confluent basin that defines a meaningful tributary. On this basis alone we argue that a tenet of geomorphology, i.e., that the link length distribution is the same random variable regardless of the link magnitude [e.g., Leopold et al., 1964], not uniform at all, needs be revisited. A link, in fact, is only defined as the alongstream distance separating two consecutive tributaries of any size. As we shall see, this concept is neither theoretically or empirically wrong, but it is strongly correlated to an arbitrary threshold in tributary area being assigned. Even operationally in the extraction of a river network, the concept is thus of fundamental significance.

[7] This paper is organized as follows. Section 2 proposes our new theoretical framework for the quantities of interest. Section 3 discusses the theoretical and empirical resources mobilized for the validation of the theory, specifically treating tributaries draining on mainstreams (in the case of real rivers and Peano networks), or tributaries draining onto a draining multiple-outlet line (in the case of OCNs and Scheidegger networks). Section 4 proposes results and discussion of the comparative analyses of theory and data from various sources. A set of conclusions, Section 5, then closes the paper.

## 2. Theoretical Framework

[8] Figure 1 illustrates the definition of the random variable that we address throughout this paper. The distance separating confluences of tributary areas larger than a threshold  $a$  is shown along the baseline of multiple outlet OCNs (Figures 1a and 1b) and the mainstream of a real river network (Figure 1c). Note that a river basin, and any



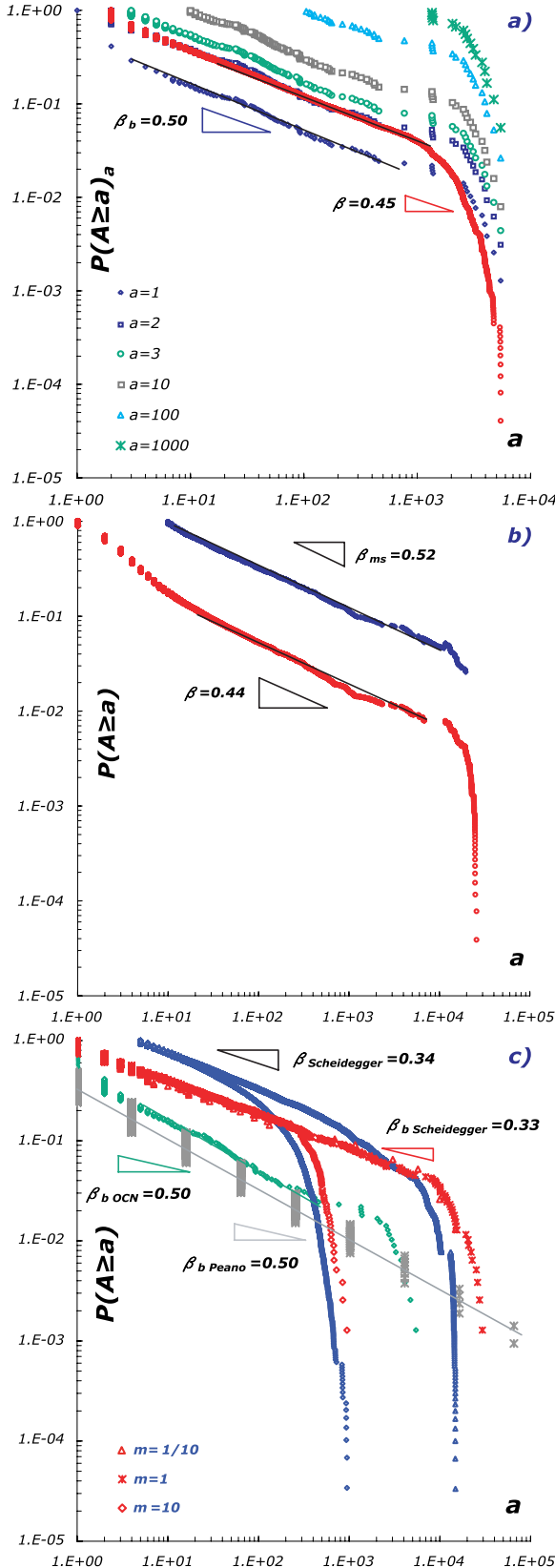
subbasin nested into it, is generally an anisotropic system defined by a longitudinal typical length  $L = L_{\parallel}$  (which we will identify with the linear size of the system), and a typical transverse length  $L_{\perp} \leq L_{\parallel}$ , both measured along the principal inertia axes (Figure 1). We shall assume that the

width scales as  $L_{\perp} \sim L_{\parallel}^H$  (with  $0 \leq H \leq 1$ ) and call the basin self-affine if  $H < 1$  and self-similar if  $H = 1$ .  $H$  is called the Hurst coefficient because of the analogies with fractional Brownian motion contexts and is typically in the range  $H = 0.75 \div 1.00$  [Rodriguez-Iturbe and Rinaldo, 1997, pp. 174–175]. As a consequence the maximum characteristic area  $A_{\max}$  is postulated to scale as  $A_{\max} = L_{\parallel} L_{\perp} \sim L^{1+H}$  [Maritan et al., 1996; Rinaldo et al., 1998].

[9] Figure 2 shows the networks on which we base the validation of the ensuing theoretical predictions. They are (a) multiple-outlet optimal channel networks (OCNs), where distances are usually measured along the lower draining boundary, although in some instances we may resort to calculation of single-outlet OCNs to test mainstream distributions, (b) and (d) three real river networks, suitably extracted from their landscape topography, where distances among tributaries' areas will be measured along the mainstream, (c) subbasins of a Scheidegger's network draining at a multiple-outlet boundary (see also Appendix A), and (e) Peano's basin, where tributaries are computed along the mainstream. For an accurate description of the statistical properties of the networks employed, and for an updated list of pertinent references the reader is referred to Rinaldo et al. [2006].

[10] In all cases we can define, along either mainstreams or multiple-outlet boundaries or else anywhere within the basin, the random contributing area at a site, say  $A$ , whose probability distribution is here computed through  $P(A \geq a)$ , that is, the relative proportion of areas, sampled anywhere, larger than the given value  $a$  (Figure 3) [Rodriguez-Iturbe et al., 1992a, 1992b; Maritan et al., 1996; Rodriguez-Iturbe and Rinaldo, 1997] whose scaling form is:

$$P(A \geq a) \sim a^{-(\tau-1)} F\left(\frac{a}{A_{\max}}\right) \sim a^{-(\tau-1)} F\left(\frac{a}{L^{1+H}}\right), \quad (1)$$



**Figure 3.** Cumulative probability distributions of areas  $P(A \geq a|L)$  sampled on the whole basin and  $P_{ms}(A \geq a|L)$  sampled along a mainstream. (a) Statistics on OCNs for different threshold values  $a$  (see legend). Here we have computed either  $\tau - 1 = \beta$  or  $\tau_b - 1 = \beta_b$  from the estimate of the slope of the double logarithmic plot. Note that the less constrained arrangement of the periodic boundary conditions calls for values of  $\tau$  closer to the exact value  $3/2$  of the ground state. (b) Sample statistics of the Tanaro river basin. Note that  $\tau_{ms} - 1 = \beta_{ms} \sim 0.52$  for areas along the mainstream and  $\tau \sim 0.44$  for the entire basin, as we expected. Note also the noteworthy lower cutoff imposed by the extent of unchanneled sites owing to the relatively large extraction threshold chosen, a step quite suitable for  $8000 \text{ km}^2$  networks. (c) Sample statistics of the Scheidegger network with  $L_{\perp} \sim L^{1/2}$  and  $L_{\parallel} \sim L$ . The Scheidegger model requires  $\tau_b = \tau$  and might be treated as a single- or multiple-outlet construction, with directed, subvertical drainage directions and periodic boundary conditions (see also Appendix A). Here we show the results for different ratios of sides (see legend) starting with the lattice characterized by  $m = \frac{L_{\perp}}{L_{\parallel}} = 1/10$ , which allows subbasins properly characterized by  $H = 1/2$ . Statistics for the single-outlet Peano basin are also shown for which, as for the Scheidegger model, the equivalence  $\beta_b = \beta$  is true.

**Table 1.** Summary of Computed and Predicted Scaling Exponents for Peano's Network, Scheidegger's Constructions, Multiple-Outlet OCNs Where Areas Are Sampled at the Lower Boundary, Three Real River Basins, and a Single-Outlet OCN Where Areas Are Sampled Along the Mainstream<sup>a</sup>

Scaling Exponent	Peano	Scheidegger	Multiple-Outlet OCN	Tanaro	Mis	Cordevole	Single-Outlet OCN
$H$	1.00	1/2	0.97	0.75	0.78	0.82	0.85
$d_f$	1.00	1.00	1.01	1.03	1.02	1.02	1.01
$\tau$	3/2	4/3	1.45	1.44	1.43	1.43	1.45
$\phi + 1$	3/2	$1.33 \pm 0.02$	$1.46 \pm 0.02$	$1.42 \pm 0.03$	$1.45 \pm 0.03$	$1.44 \pm 0.03$	$1.45 \pm 0.02$
$2 - \frac{d_f}{1+H}$	3/2	4/3	1.48	1.41	1.43	1.44	1.45
$\psi$	1/2	$0.33 \pm 0.02$	$0.50 \pm 0.01$	$0.43 \pm 0.02$	$0.45 \pm 0.02$	$0.46 \pm 0.02$	$0.47 \pm 0.02$
$\frac{Hd_f}{1+H}$	1/2	1/3	1/2	0.44	0.45	0.46	0.46
$\tau_{ms}$	3/2	$1.66 \pm 0.02$	—	$1.52 \pm 0.03$	$1.55 \pm 0.03$	$1.54 \pm 0.03$	$1.53 \pm 0.02$
$1 + \frac{d_f}{1+H}$	3/2	5/3	—	1.58	1.57	1.56	1.54
$\tau_b$	—	$1.33 \pm 0.02$	$1.50 \pm 0.02$	—	—	—	—
$2 - \frac{1}{1+H}$	—	4/3	1.49	—	—	—	—

<sup>a</sup>For Peano's Network, see Figure 2e. For Scheidegger's construction, see Figures 2c and A1. For multiple-outlet OCNs, see Figures 1a, 1b, and 2a. For the Tanaro river basin, see Figures 1a and 1b, and for the Mis and Cordevole river basins, see Figure 2d. The computed scaling coefficients are  $H$ ,  $d_f$ ,  $\tau$ ,  $\tau_{ms}$ , and  $\tau_b$  estimated numerically by fitting scaling laws (for  $\tau$ ,  $\tau_b$ , and  $\tau_{ms}$ , see Figure 3) and checked against exact solutions when available (see equations (1)–(9)); the scaling exponent  $\phi + 1 = \tau$  is estimated from the relationship  $\langle a \rangle_a \propto a^\phi$  (equation (11) and Figure 4); the third independent estimate of  $\tau = 2 - \frac{d_f}{1+H}$  (equation (3)) is then directly computed from the computational estimates of  $H$  and  $d_f$ ; the scaling exponent  $\psi$  is computed from the relationship  $\langle d \rangle_a \propto a^\psi$  (Figure 8); the scaling relationship  $\psi = \frac{Hd_f}{1+H}$ , derived theoretically (equation (19)), is compared with the independent estimates of  $H$  and  $d_f$ ; and the scaling exponent  $\tau_{ms}$  of the distribution of areas sampled along the mainstream is computed, when appropriate, from our data. Note that we have not computed it only for multiple outlet OCNs; the theoretical prediction  $\tau_{ms} = 1 + d_f/(1 + H)$  (equation (6)) is analyzed from our data; the scaling exponent  $\tau_b$  is computed via the proper area distributions sampled along a line boundary where multiple outlets occur. This calculation is possible only for multiple-outlet OCNs (Figures 1a, 1b, and 2a) and for the particular Scheidegger-like constructions shown in Figure A1; the theoretical prediction  $\tau_b = 2 - 1/(1 + H)$  (equation (9)) is compared against our data. Variabilities of measured exponents are standard errors found by Reduced Major Axis Regression (RMA or SMA) bootstrapping over cases and deriving slopes by the linear and the Jackknife models [Warton et al., 2006].

where:  $A_{\max} = L^{1+H}$  [Maritan et al., 1996] is the largest area that defines the upper cutoff for the distribution (the length scale  $L$  is defined later), and  $\tau = 1.43 \pm 0.02$  with remarkably small scatter from a surprisingly large archive of real river basins regardless of climate, vegetation, exposed lithology or mainstream sinuosity of the different river basins (Figure 3) [Rodriguez-Iturbe et al., 1992a; Maritan et al., 1996; Rigon et al., 1996; Rinaldo et al., 1998]. Obviously the probability density function (pdf)  $p(a) = -dP(A \geq a)/da$  follows as:

$$p(a) = p(a|L) = a^{-\tau} F\left(\frac{a}{L^{1+H}}\right), \quad (2)$$

where  $f$  follows from direct integration of equation (1) and the notation  $p(a|L)$  stresses the conditional character of the pdf on the cutoff size  $L^{1+H}$  [Maritan et al., 1996]. Further relationships involving scaling exponents [Rinaldo et al., 1998] are Hack's law  $l \sim A_{\max}^h$  (where  $h \sim 0.57$  is the Hack's exponent with implications on the type of self-similarity operating within the river basin [Rigon et al., 1996, 1998]) and  $l \sim L^{d_f}$  (with  $d_f \sim 1.0$ , see Table 1), where  $l$  is the length of the longest stream in the drained area itself function of the contributing area value  $a$  (see Figure 1), from which one derives linkages that constitute stringent comparative tools, for example,

$$\tau = 2 - \frac{d_f}{1+H} = 2 - h, \quad (3)$$

which we shall usually simplify by assuming  $d_f \sim 1$ , hence leading to  $2 - \tau \sim 1/(1 + H)$ . Moreover, the relation  $\tau + h \sim 2$  naturally conforms to the common values  $\tau \sim 1.43$ ,  $h \sim 0.57$  [Maritan et al., 1996]. As shown in Figure 3 the above relationships are verified computationally or exactly derived for all the constructs shown in Figure 2 (Figure 1), including the Scheidegger construction which exhibits the exact value  $\tau = 4/3$  for the distribution of areas

along the baseline [Huber, 1991; Takayasu et al., 1991] and a slow convergence to the exact asymptotic values for area at any point (Figure 3) [e.g., Maritan et al., 1996; Rigon et al., 1996; Rodriguez-Iturbe and Rinaldo, 1997]. A summary of the relevant scaling exponents derived in this study is reported in Table 1 jointly with comparative results to be discussed later.

[11] The constraint of conservation of total area suggests [Maritan et al., 1996; Rodriguez-Iturbe and Rinaldo, 1997] that the distribution of areas sampled along a given line boundary where multiple outlets occur, say  $p_b(a|L)$ , or along a mainstream draining from both sides, say  $p_{ms}(a|L)$ , differs from equations (1) and (2) because tributary area sampled anywhere within the tributary is a different random variable, sampled from a different population from those constrained in relation to their possible spatial locations. Indeed if at  $i$  sites one collects the areas  $A_i$  and must enforce the constraint  $\sum_i A_i = A_{\max}$  (where  $A_{\max}$  is the total area), the resulting population is different from that leading to equation (1) because the analog areas  $A_i$  sampled anywhere do not add to total area [Rodriguez-Iturbe and Rinaldo, 1997]. If we assume that the distribution at the mainstream sites is given by [Maritan et al., 1996; Rodriguez-Iturbe and Rinaldo, 1997]:

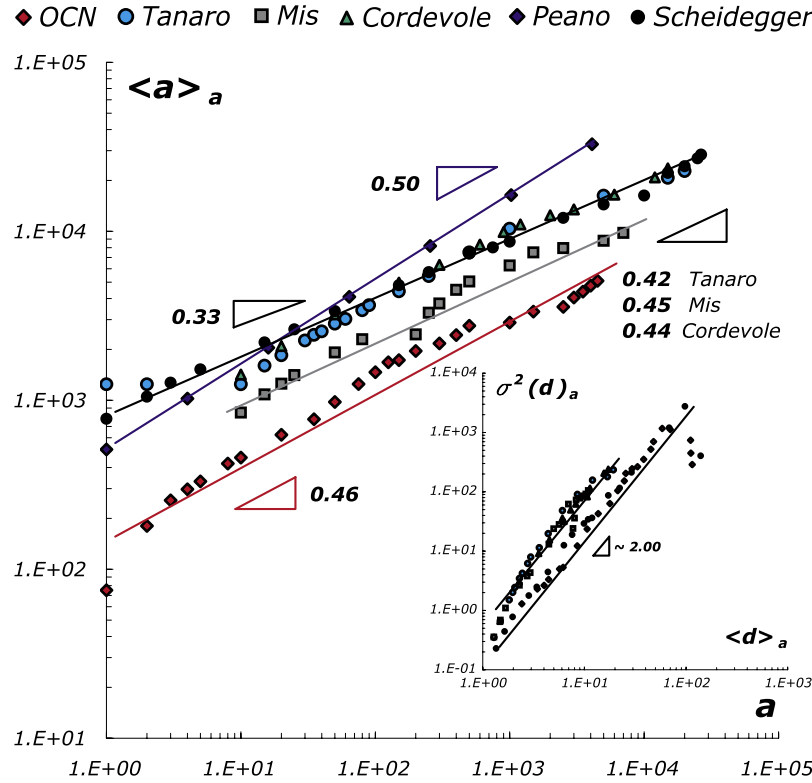
$$p_{ms}(a|L) = a^{-\tau_{ms}} f_{ms}\left(\frac{a}{L^{1+H}}\right), \quad (4)$$

the area conservation constraint, i.e.,

$$l \langle a \rangle_{ms} \sim L^{1+H}, \quad (5)$$

(where  $l$  is the length of the mainstream) leads, by explicitly calculating the mean and power counting, to the relationship

$$\tau_{ms} = 1 + \frac{d_f}{1+H}, \quad (6)$$



**Figure 4.** When we average the size of basins larger than  $a$  encountered along a mainstream or a multiple-outlet boundary, we observe a scaling relationship between the average size  $\langle a \rangle_a$  and the threshold area  $a$  that defines the tributary, i.e.,  $\langle a \rangle_a \sim a^{\tau-1}$ , where  $\tau$  is the scaling exponent of the distribution of areas anywhere in the basin. Here we document an excellent agreement of theory and data, in all cases, specifically, Peano's construction (diamonds), Scheidegger's network (solid dots), and real river basins (circles, squares, and triangles). The estimated scaling exponents are also shown. Inset: the logarithmic corrections to the relation between  $\sigma^2(\langle d \rangle_a)$  versus  $\langle d \rangle_a$  shown here as a function of  $a$  (see Figures 8 and 9). An average exponent for all basins is roughly equal to 2, confirming the variance mean theoretical expectation  $\sigma^2(\langle d \rangle_a) \sim \langle d \rangle_a^{2\alpha}$  with  $\alpha \simeq 1$ .

and ultimately to the inequality:

$$\tau \leq 1.5 \leq \tau_{ms}, \quad (7)$$

indicating there are significant differences between  $\tau$  and  $\tau_{ms}$ . For mainstream boundaries, draining areas of overall size  $L^{1+H}$  we have normally  $\tau \sim 1.43$  and  $\tau_{ms} \sim 1.5$ , although in particular cases like Scheidegger's (where  $H=1/2$ ,  $d_f=1$  and  $\tau=4/3$  exactly [Huber, 1991]) one has  $\tau_{ms}=5/3$  (see Table 1). Figure 3 shows the relevant numerical computations, where finite-size effects are particularly evident.

[12] Finally note that yet another random variable could be derived sampling area along a straight boundary perpendicular to the main axis  $L_{\parallel}$  of the basin. In such a case, independently on the size of the boundary as long as it exceeds  $L^H$ , one has

$$p_b(a|L) = a^{-\tau_b} f_b\left(\frac{a}{L^{1+H}}\right), \quad (8)$$

where

$$\tau_b = 2 - \frac{1}{1+H}, \quad (9)$$

with the usual meaning of the symbols. For straight multiple-outlet boundaries in Scheidegger constructs (of the type discussed in Appendix A) we find the exact result

$\tau_b = \tau = 4/3$ . Notice that if  $d_f = 1$ ,  $\tau = \tau_b$  and if  $H = 1$  all area scaling exponents equal  $3/2$ .

[13] An interesting feature that can be derived from equation (2) and its variations (4) and (8) is  $\langle a \rangle_a$ , i.e., the mean area of all tributaries whose size is larger than a given threshold area  $a$ . This is tantamount to averaging the size of all incoming tributaries larger than  $a$ , resulting in the obvious fact  $\langle a \rangle_a \geq a$ . Indeed this quantity matters here, because we might want to define sites with flow, width or riparian area larger than a threshold value to define viable minimum values, for example, for biological conservation. One thus has:

$$\langle a \rangle_a = \frac{\int_a^\infty dx x p(x|L)}{\int_a^\infty dx p(x|L)} = \frac{\int_a^\infty dx x^{-z} f(x/L^\phi)}{\int_a^\infty dx x^{-\tau} f(x/L^\phi)}, \quad (10)$$

(where  $z = \tau - 1$  and  $\phi = 1 + H$ , defined for reasons to appear in the latter) which poses a nontrivial scaling result that we shall use in the following. The basic result is that when  $L \rightarrow \infty$  and  $z > 1$  one has the asymptotic behavior  $\langle a \rangle \propto a^{1-z}$ , whereas if  $z < 1$  one has  $\langle a \rangle \propto L^{\phi(1-z)}$  (see equation (18)). In the case at hand, regardless of boundary conditions in order to average sizes of tributaries larger than  $a$  we sample anywhere in the basin, thus  $p(a|L)$  is the

appropriate distribution and  $\tau$  the relevant scaling exponent, one has:

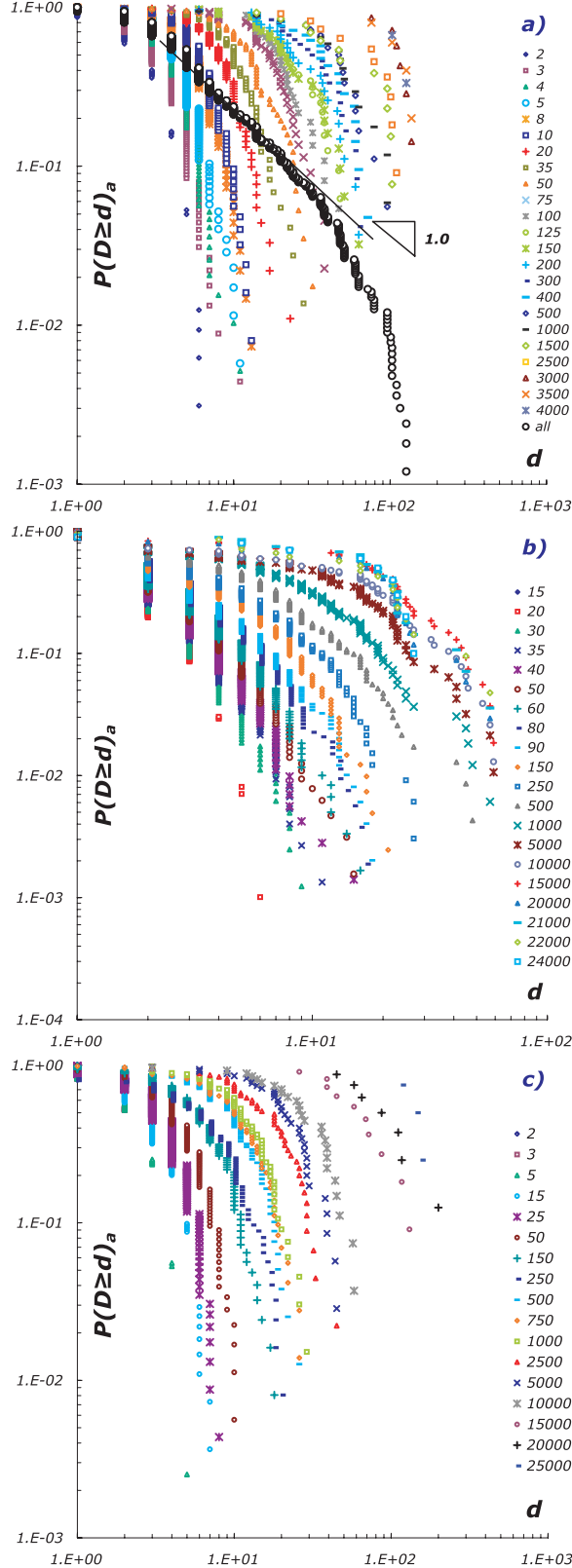
$$\langle a \rangle_a = \frac{L^{(1+H)(2-\tau)}}{a^{1-\tau}} \propto a^{\tau-1} = a^{1-\frac{d_f}{1+H}}, \quad (11)$$

a result that we shall compare with data in the following. Note that for scaling with  $\tau_b$  as in equation (9) one would

have had  $\langle a \rangle_a \sim a^{H/(1+H)}$  and for  $\tau_{ms}$  as in equation (6) one would have obtained  $\langle a \rangle_a \sim a^{d_f/(1+H)}$ .

[14] We shall present now the derivation of the distribution of distances of tributaries characterized by their area larger than a threshold  $a$  (and thus by related characteristic flow rates, riparian proper, cross-section width, depth and velocity among other features). Notationwise, we shall define the random variable drawn from the random population of possible distances as  $D$ , and as  $P(D \geq d)_a$  the probability of exceeding a given distance  $d$  conditional on the confluence of tributaries of area (at least)  $a$ .

[15] We approach this derivation initially for the multiple outlet boundary leading to a distribution of areas with scaling exponent  $\tau_b$ . We assume that the distribution of distances (conditional on a given tributary area  $a$ ) is defined by the probability of finding  $d-1$  consecutive tributaries with area smaller than  $a$ , and finding two tributaries of area equal or greater than  $a$  at both ends. Distances are walks in terms of steps between tributaries, so  $d$  is in pixel units. If statistical independence holds for tributaries area, for  $a \gg 1$  so  $d \gg 1$ , such probability would be given by  $P(A < a)^{d-1} = (1 - P(A \geq a))^{d-1} \propto \exp[-(a^{1-\tau_b}) d]$ . One doubts, however, that the statistical independence might indeed hold for small distances. At small areas, in fact, the determination of the probability is most uncertain owing to the issue of the proper drainage density for network extraction, hence probabilities become reliable only for sizable areas. For large areas, however, the occurrence of consecutive tributaries cannot be independent at small distances, as one cannot possibly have two large consecu-



**Figure 5.** Exceedence probability distributions  $P(D \geq d)_a$  of the random distance  $D$  between tributaries versus the current value  $d$  of distance, clearly a function of the threshold size  $a$  that defines a meaningful confluent tributary. (a) Distribution of distances for the multiple-outlet OCN draining along a straight line of Figure 2a (see also Figures 1a and 1b), for different values of  $a$  namely reported in the legend on right of the graphic; darker line (empty circles) is obtained by ensemble averaging of the exceedence probabilities putting together all distances for each value of  $a$  (note that such distances span the entire range of possible values of  $D$ ). (b) Statistics of distances made along the mainstream of Tanaro basin (Figure 1c and Figure 2b). The chosen values of  $a$  are on the legend attached on the right. (c) Statistics of the Scheidegger network for  $m = L_{\perp}/L_{\parallel} = 1$  draining onto a straight boundary. Values of the parameter  $a$  related to each distribution are on the legend on right of the plot; observe how for higher values of  $a$ , distributions are quite detached from the other ones, confirming the fast deviation of scaling exponents noted computationally for the model (Figure A2). Note also that the larger the size that defines tributary confluent basins, the longer it takes to find them (on the average) walking along the chosen draining boundary. See for whichever type of networks that distributions of distances for each value of the parameter  $a$  lead to a characteristic length  $d(a)$  as a function of the parameter  $a$ , so they can be considered as exponential distribution, like  $P[D \geq d]_a = \text{const} \exp(-\beta d)$  with  $\beta \rightarrow 0$ .



tive basins for distances smaller than  $L_{\perp} \sim L^H$ . This assumption will also be tested in the following.

[16] For two tributaries of size  $\approx a$ , draining in a quasi-parallel arrangement along a boundary, one may speculate that the minimum distance is roughly proportional to their width  $L_{\perp}$ . This implies that the mainstream is somewhat straight and that for smaller distances the receiving boundary can only collect smaller basins. The above relationships imply:

$$d(a) \sim a^{\frac{H}{1+H}}. \quad (12)$$

Note that in case the boundary is draining tributaries from both sides, as in the case of a mainstream, we expect the distribution to be a fraction of that given by the ansatz of equation (12). Note also that the unconditional probability  $p(d)$  would be given by the relationship  $p(d)dd = p(a)da$  and  $da/dd \propto d^{1/H}$ , from which one obtains, after some calculations, using equation (6) with  $d_f = 1$ :

$$p(d) \sim d^{-\frac{1+H}{H}} f_d\left(\frac{d}{L^H}\right), \quad (13)$$

more generally the power law is  $d^{-\frac{\tau_{ms}(1+H)}{H}}$ , from which the unconditional probability of exceedence is given by  $P(D \geq d) \propto d^{-(\tau_{ms}-1)\frac{1+H}{H}}$ , with the scaling argument  $d^{-1/H}$ . From the computational viewpoint, one finds easier to compute the probability  $P(D \geq d)_a$  of finding a distance conditional on a given size  $a$  of the tributary. The basic relationship would be given by

$$p(d) = \int da p(d|a) p(a), \quad (14)$$

from which, after some calculations and power counting by assuming  $p(d|a) \sim d^{-\psi} \hat{f}_d(d/a^{H/(1+H)})$ , one obtains  $\psi + 1/H = (1+H)/H$  and hence  $\psi = 1$  independently of  $\tau$ , or:

$$p(d|a) = d^{-1} \hat{f}_d\left(\frac{d}{a^{H/(1+H)}}\right), \quad (15)$$

which, interestingly, leads to the testable scaling form

$$P(\geq d) = P(D \geq d)_a = \hat{F}_d\left(\frac{d}{a^{H/(1+H)}}\right), \quad (16)$$

(where both notations are freely interchanged) that follows from equation (15) with:

$$\hat{F}_d(x) = \int_x^\infty dy y^{-1} \hat{f}_d(y), \quad (17)$$

whose collapse for different values of  $a$  can be directly compared with observations from real or artificial networks. Notationwise (in analogy with previous choices), we shall use the notation  $F_{ms}(x)$  and  $F_b(x)$  for mainstream or multiple-outlet boundary estimates of the scaling function in equation (16), respectively. The theoretical prediction, in fact, suggests that the larger the threshold area  $a$  for tributaries of equal size, the larger, in the average, the random distance  $D$  among them, hence different values of  $a$  should produce broader distributions with larger mean and variance. However, when probabilities  $P(D \geq d)_a$  are plotted against  $d/a^{H/(1+H)}$  a collapse onto a single curve is expected. Notice that the scaling function  $F_d(x) \rightarrow 1$  when  $x \rightarrow 0$  and  $F_d(x) \rightarrow 0$  when  $x \rightarrow \infty$ , and it depends on the boundary conditions adopted (e.g., if drainage occurs along a mainstream, areas flow in from both sides, and distances are shorter).

[17] The expected distance between two rivers of area  $\geq a$  would therefore be given in terms of the probability of finding such area  $a$ , i.e., (for a mainstream)

$$\langle d \rangle_a \sim \langle a^{\frac{Hd_f}{1+H}} \rangle \propto \frac{\int_a^\infty p_{ms}(a|L) a^{Hd_f/(1+H)} da}{\int_a^\infty p_{ms}(a|L) da}. \quad (18)$$

Standard finite-size arguments require:

$$\int_a^\infty a^{-z} F\left(\frac{a}{L^{1+H}}\right) da \sim \begin{cases} L^{(1+H)(1-z)} & \text{for } z < 1 \\ a^{1-z} & \text{for } z > 1 \end{cases}, \quad (19)$$

where  $z = \tau_{ms} - \frac{Hd_f}{1+H}$ , and the integral is dominated by the upper/lower cutoff when  $z < 1$  and  $z > 1$  respectively. The coefficient  $z = \tau_{ms} - \frac{Hd_f}{1+H} \geq 1$  because  $H \leq 1 \leq d_f$ . Since  $\tau > 1$ , one finds the mean distance between two subbasins of area  $\geq a$  as:

$$\langle d \rangle_a \sim \frac{a^{1-(\tau_{ms}-\frac{Hd_f}{1+H})}}{a^{1-\tau_{ms}}} \sim a^{\frac{Hd_f}{1+H}} \sim \langle a \rangle_a^H, \quad (20)$$

which can be compared with field observations directly. On average for real basins  $d_f \simeq 1$  so simplifying the scaling relation is  $\langle d \rangle_a \sim a^{\frac{H}{1+H}}$ . In equation (19) we have used the result  $\langle a \rangle_a \sim a^{\tau_{ms}-1}$  which can be derived with the same above procedure. Analogously one may compute the second moment  $\langle d^2 \rangle_a = \langle a^{2H/(1+H)} \rangle$  which poses a few additional finite-size problems. In fact, through scaling arguments similar to the ones seen before, with  $d_f \simeq 1$ , we obtain

$$\langle d^2 \rangle_a \sim a^{\tau_{ms}-1} L^{2H-1} = a^{\frac{1}{1+H}} L^{2H-1}, \quad (21)$$

**Figure 6.** Collapse tests for  $P(D \geq d)_a$ . (left) The collapse of  $P(D \geq d)_a$  is tested as a function of  $d/a^{\frac{Hd_f}{1+H}}$ , as predicted by theory for mainstreams (equation (16)). (right) The collapse of  $P(D \geq d)_a$  is tested as a function of  $d/\langle a \rangle_a^H$  or  $d/\langle a \rangle_a^{d_f}$  depending on boundary conditions (equations (20) and (23) into equation (16)). Here we show (a) and (b) for OCNs, (c) and (d) for the Tanaro river network, and (e) and (f) for the Scheidegger model with side ratio  $m = 1$ . Sensitivity tests are shown on the estimate of the best exponents (insets in Figures 6a–6d), where  $R^2 = \sum_{i=1}^N \sigma_i^2$ , i.e., the sum of variations with respect to the mean for each  $i$ th logarithmic bin considered. Larger red dots represent mean values for each bin, and error logarithmic bars are standard deviations. The fitting shown with the red continuous line is carried out by an exponential function. For Scheidegger networks draining on a line, the collapse for higher tributaries' areas has to be done with the deviation exponents from  $H/1 + H$  and  $d_f$  (insets in Figures 6e–6f), as observed before and as supported by our calculations (Figure A2).



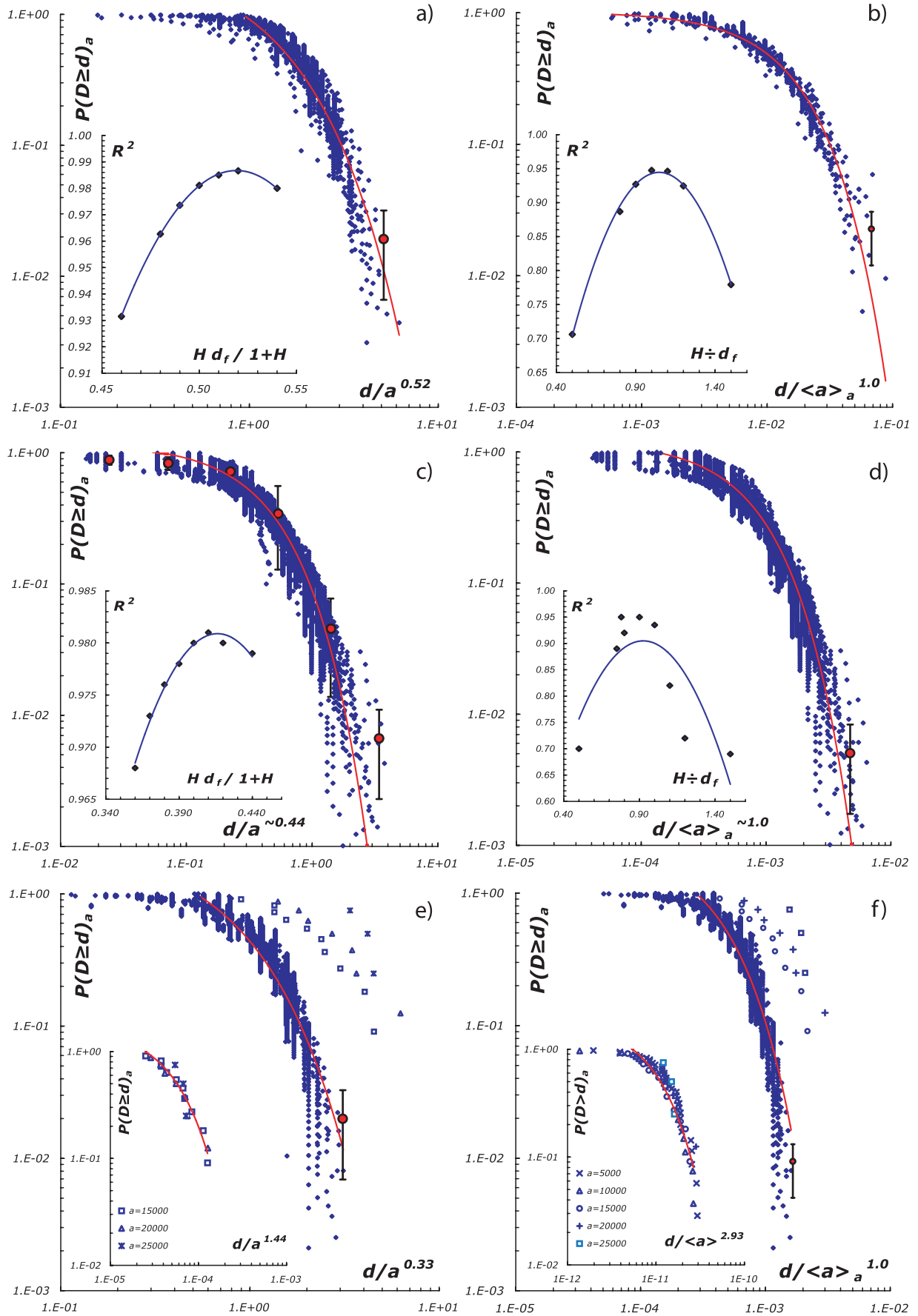
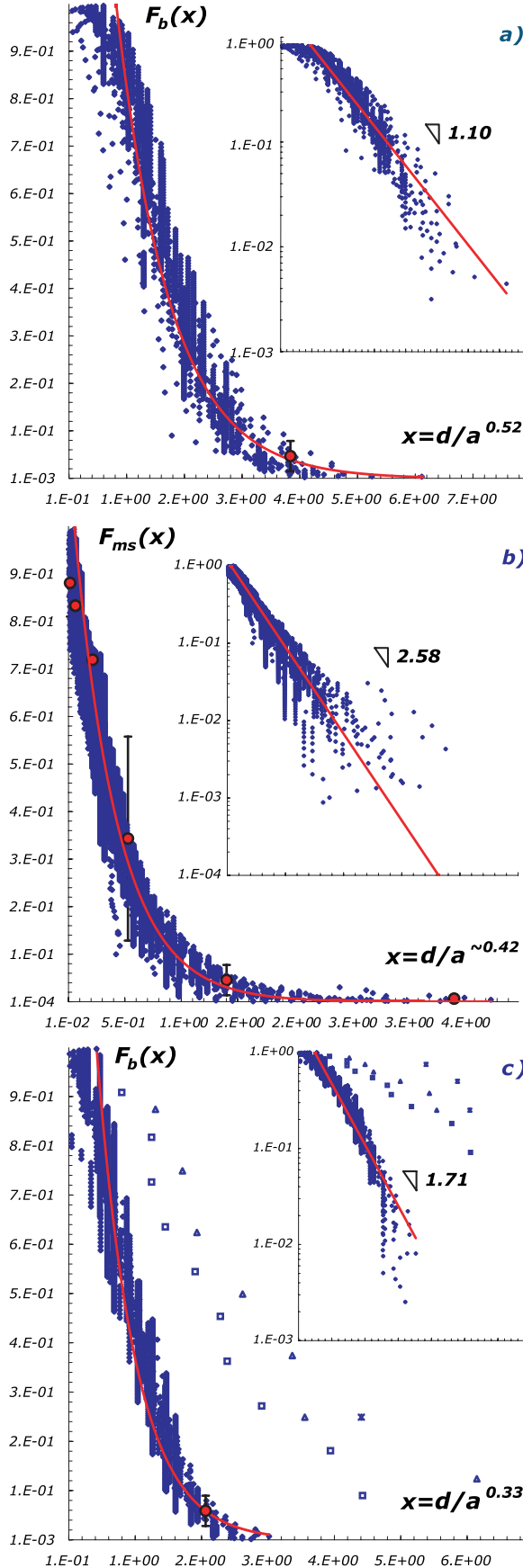


Figure 6



and the variance  $\sigma^2(d)_a = \langle d^2 \rangle - \langle d \rangle^2 = \langle (d - \langle d \rangle)^2 \rangle$ , using equation (20) and the square of equation (18), is given by

$$\sigma^2(d)_a \sim a^{\frac{2H}{1+H}} \left( -1 + \text{const} \frac{L^{1+H}}{a} \right)^{\frac{2H-1}{1+H}}, \quad (22)$$

Both statistics foresee notable corrections dependent on size  $L$ , which will also be tested against data (see inset in Figure 4).

[18] We finally note that in the case of straight, long enough (i.e. longer than  $L^H$ ) multiple-outlet boundary one has to carry out the same operations via the distribution  $p_b(a|L)$  with  $\tau_b = 2 - 1/(1 + H)$  (see equation (9)), thus obtaining:

$$\langle d \rangle_a \sim a^{\frac{Hd_f}{1+H}} \sim \langle a \rangle_a^{d_f/L}, \quad (23)$$

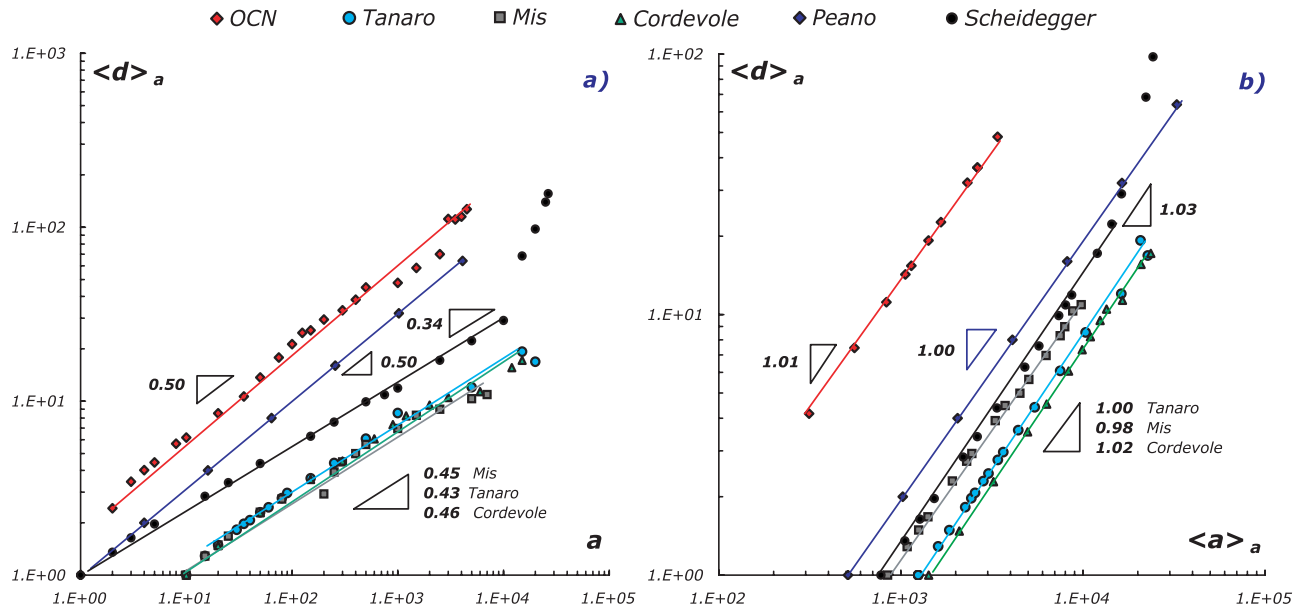
which indicates that Scheidegger's network draining along a straight boundary should scale approximately like real rivers along mainstems (for both  $d_f = 1$ , exactly in the case of Scheidegger and approximately for real rivers).

[19] All the above relationships will be compared against data in the next Section. Note, finally, that for Peano, OCNs and real basins one has  $H \simeq 1 \simeq d_f$ , whereas for Scheidegger river networks  $d_f = 1$  while  $H = 1/2$  derive from the exact properties of the random walk [Hughes, 1995] that defines the outer boundaries. Thus we value comparative studies among them, as notable differences in scaling structures should arise in the observational properties.

### 3. Comparative Analysis: Tributaries Draining Onto Multiple-Outlet Boundaries or Mainstems

[20] The data set specifically used for the computation of the distributions of distances for real rivers pertains three sample river basin in Northern Italy (see caption of Figure 2), the Tanaro river basin (Piemonte, Italy), the Cordevole and Mis river basins (Veneto, Italy), part of our relatively large geomorphologic database where we have chosen the actual basins for their differing size, relief and geologic and vegetational context. Table 1 illustrates the rather narrow range of variation for the possible values of  $H$ ,  $d_f$ ,  $\tau$  of real networks,

**Figure 7.** The detailed structure of the collapse function (i.e., homogeneity or scale function)  $F_b(x)$  and  $F_{ms}(x)$  plotted versus  $x$ , where  $x$  is the argument  $d/a^{H/(1+H)}$ . Here we choose an exponential that would be the case should the probability obey  $P(D \geq d) \propto P(A < a)^{d-1}$ . We note that as expected, the mainstream collapse function is generally shorter than in the boundary case, with higher exponents in semilog plot, because tributaries drain independently on both sides, whereas for line boundaries, all other conditions being equal, the distances are larger in the average. (c) For Scheidegger's construct the exponent  $\beta$  on semilog plot is higher than (a) the OCN one and lower than (b) the Tanaro one. The analytic form of the exponential function is of the type  $F(x) = \exp(-Cx)$  where  $C$  is a suitable constant (reported in each inset) where  $x_c = 1/C$  is a cutoff value of the distribution of  $d/a^{H/(1+H)}$ . Note the finite-size effects for Scheidegger's networks (see also Figure A1).



**Figure 8.** Scaling of the mean distance  $\langle d \rangle_a$  with  $a$  and  $\langle a \rangle_a$ : (a) the scaling exponent  $\frac{Hd_f}{1+H}$  is derived from the relation between  $\langle d \rangle_a$  versus  $a$ , and (b) we recover the exponents:  $d_f$ , in the case of straight draining boundary, and  $H$ , in the case of a fractal mainstream.

which follows usual patterns [see, e.g., Maritan *et al.*, 1996; Rodriguez-Iturbe and Rinaldo, 1997; Rinaldo *et al.*, 1998]. Conversely, conceptual constructs like Peano or Scheidegger have rather different features [Rinaldo *et al.*, 1998], some exactly determined [Huber, 1991; Marani *et al.*, 1991; Takayasu *et al.*, 1991], thus allowing us to check theoretical results against the scaling exhibited by computational entities, such as real river networks, and OCNs (single and multiple outlet). All geomorphological quantities, including distance  $d$  and area  $a$ , are expressed in pixel units.

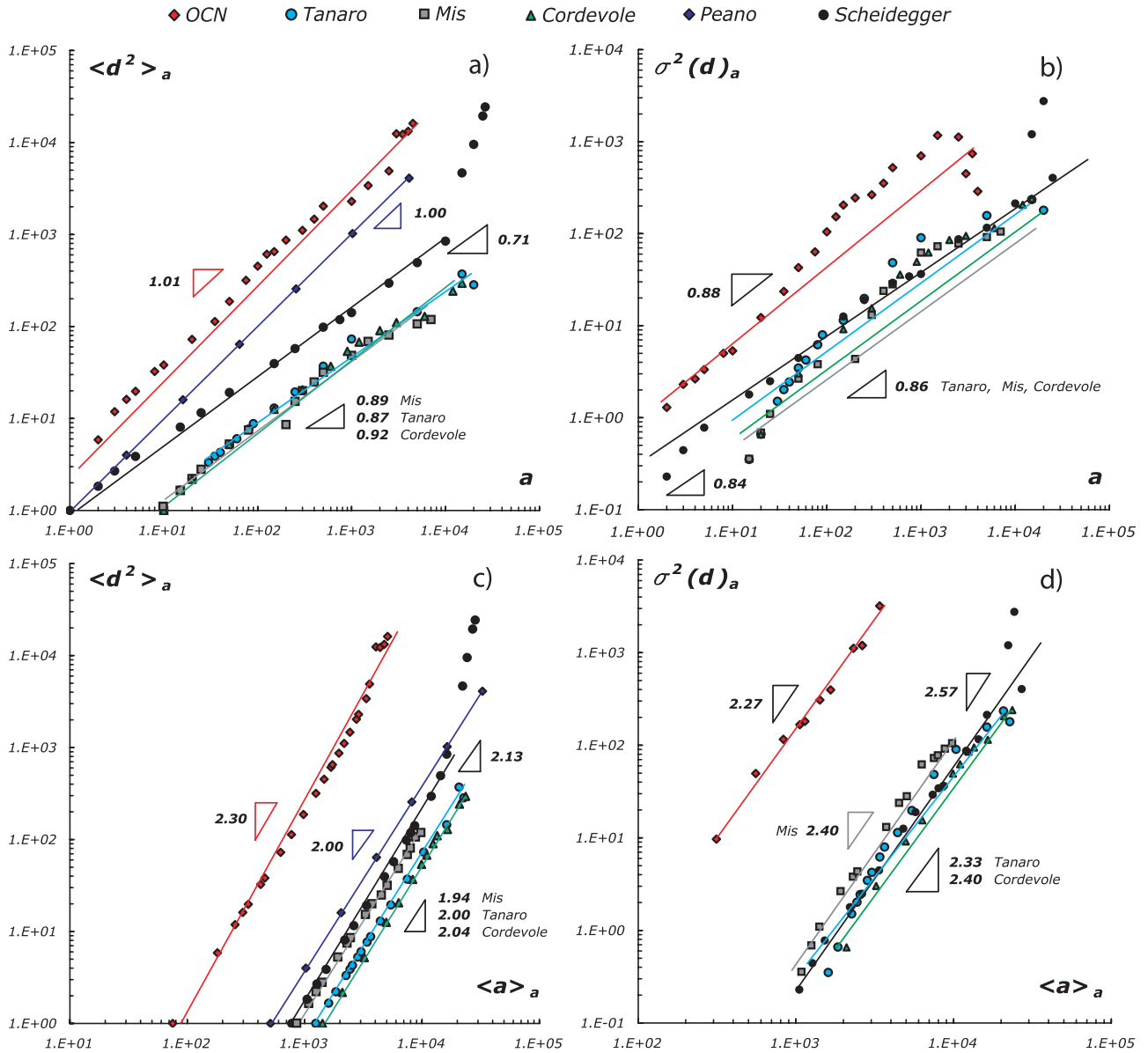
[21] Figure 4 illustrates a preliminary test of the theoretical scaling results that are derived solely from the probability of areas  $p(a|L)$ . Specifically, we discuss the scaling relationship of  $\langle a \rangle_a$  (the mean area of all basins larger than a threshold area  $a$  encountered along a boundary or a mainstream) with the minimum tributary area  $a$ . From theory we predict (equation (11)) that such scaling is of the type  $\langle a \rangle_a \propto a^{\tau-1}$  where  $\tau$  is the scaling exponent of the distribution of area anywhere within the network, and thus independent of boundary conditions [Rodriguez-Iturbe *et al.*, 1992a; Maritan *et al.*, 1996]. The computed values match quite accurately Peano's exact value  $\tau - 1 = 1/2$  [Marani *et al.*, 1991; Colaiori *et al.*, 1997; Rodriguez-Iturbe and Rinaldo, 1997], the Scheidegger's exact value  $\tau - 1 = 1/3$  [Huber, 1991; Takayasu *et al.*, 1991] and the expected value from real networks ( $\tau - 1 = 0.42 - 0.46$ ). The implied validation of the scaling argument leading to equation (11) is deemed particularly significant.

[22] Figure 5 shows the empirical distributions of distance conditional on area  $P(D \geq d|a)$ , computed for the sample networks shown in Figure 2. Notice that at larger values of  $a$ , the broader become the distributions and the larger the random distance  $D$  in the average, as expected. The double logarithmic plot chosen emphasizes the lack of a power law character (predicted by equation (16)) strictly conditional either on the threshold area  $a$  or on the drainage sites chosen as boundary conditions (a line or a main-

stream). In fact, Figure 5a shows multiple-outlet OCNs draining on a line; Figure 5b shows the results for mainstreams of Tanaro basin; and Figure 5c illustrates Scheidegger's distribution of distances draining along a line orthogonal to the mainstream. Figure 5a shows also an approximation to the unconditional probability  $P(D \geq d)$  obtained by ensemble averaging. This is operationally mimicked by the average of the exceedence probabilities calculated by putting together all distances computed for each value of the parameter  $a$  (note that such distances span the entire range of possible values of  $D$ ). Although the sampling of  $a$  was not designed to produce a proper ensemble average, we find that, as theory predicts and regardless of details on how this choice approximates the unconditional probability, the computational slope is rather close to the theoretical value of  $-1$  for the exceedence probability (which of course implies a scaling of  $-2$  for the unconditional pdf). The effects of the different aggregation structure are clearly reflected by the observational conditional exceedence probabilities of  $D$ . In all cases, in fact,  $P(D \geq d|a)$  proves to be a function of the threshold value  $a$  for the confluent tributary area, as larger thresholds naturally imply larger distances among equally sized tributaries.

[23] Figure 6 shows the collapse plot attempted because of the theoretical prediction (equation (16)), with estimates of the best collapse that indeed conforms well to the theoretical predictions (see caption for details). Two theoretical predictions are validated, namely the scaling form where  $d$  is rescaled with respect to the mean area  $\langle d \rangle_a \sim \langle a \rangle_a^{Hd_f}$  (with the exponent  $H$  for real basins,  $d_f$  for straight mainstream) of all tributaries with area  $\geq a$ , and the scaling with  $a^{H/(1+H)}$ . Values of  $H$  for the best collapse match the viable values from direct observation. It is safe to assume, in this context, that the estimate of  $d_f \approx 1$ . In all cases the theoretical prediction is verified quite well, and differences arising in the scalings are explained by the different boundary conditions or aggregation structures, like, for example, in the case of Scheidegger's





**Figure 9.** Scaling of the second moment,  $\langle d^2 \rangle$ , and variances,  $\sigma^2(d)_a$ , with (a) and (b)  $a$  and (c) and (d)  $\langle a \rangle_a$ . Note that the scaling exponents are affected by the cutoff dependence on  $L$  foreseen by equations (20) and (21).

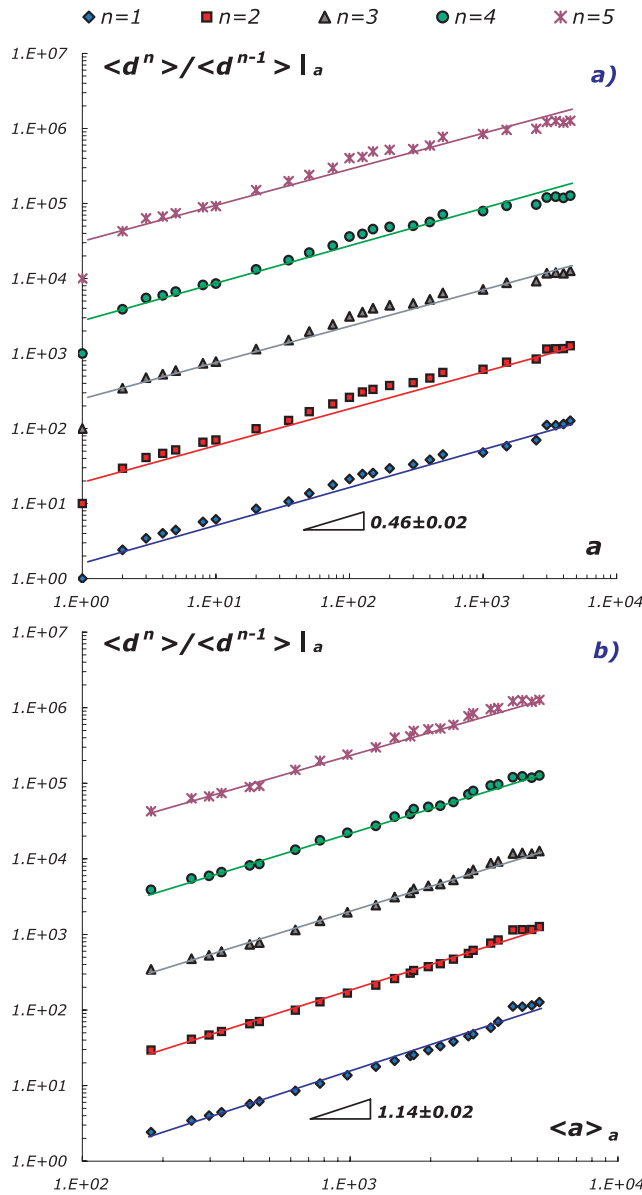
construction which exhibits by construction  $H = 1/2$ ,  $d_f = 1$  and  $\tau = 4/3$  [Huber, 1991]. Overall, the theoretical framework seems quite robustly verified.

[24] Note also that it is of great practical interest the fitting of the collapse functions  $F_{ms}(x)$  and  $F_b(x)$ , in this case simply addressed via exponential functions (Figure 7). Although such choice clearly oversimplifies the fitting problem, the single parameter serves well our purposes and clearly represents the inverse of the characteristic value (of  $d/a^{Hd_f/(1+H)}$ , or  $d/\langle a \rangle_a$  respectively) for cutoff to occur. Here we obtain  $F_{ms}(x) \sim \exp(-2.58x)$  and  $F_b(x) \sim \exp(-1.10x)$ , respectively. These may indeed be useful for practical applications.

[25] Figures 8 and 9 show the results of an articulate set of calculations required to describe the scaling of mean and variance of  $d$  given  $a$ ,  $\langle a \rangle_a$  for all networks shown in Figure 2. Here we illustrate both the scaling of the second moment  $\langle d^2 \rangle$

and of the variance  $\sigma^2(d)_a = \langle d^2 \rangle_a - \langle d \rangle_a^2$  as a function of  $a$ . Note that the theoretical predictions (equations (19)–(22)) are reasonably satisfied regardless of the fractal properties of the network, noting that in the case of Scheidegger networks this represents quite a demanding test owing to its exact scaling coefficients ( $H = 1/2$ ,  $d_f = 1$ ,  $\tau = 4/3$ ,  $\tau_{ms} = 5/3$ ).

[26] The validity of the finite-size scaling argument is strengthened by the analysis of the ratio of consecutive moments  $\langle d^n \rangle_a / \langle d^{n-1} \rangle_a \propto a^{H/(1+H)}$  for different (integer) values of  $n$ , which probes the structure of the cutoff function  $F_{ms}(x)$ . Once such ratio is plotted against  $a$  in a double logarithmic graph (as in Figure 10; see Rigon *et al.* [1996] for computational details on the binning requirements for sampling the moments for several samples of area  $a$ ), one might estimate the slope, which indeed should be equal to  $H/(1+H)$  from equation (16). The consistency of the ratio of several consecutive moments (here arbitrarily translated



**Figure 10.** Scaling of the moments of the tributaries' distances for the OCN of Figures 1a, 1b, and 2a. The plot shows the ratio  $\langle d^n \rangle / \langle d^{n-1} \rangle_a$  where  $n$  is the moment order. The lowest curve is the moment of order one; the others show higher-order moments (up to fifth order), which exhibit lower exponents than the meaningful but the stability results are excellent. Note that the curves are conveniently offset vertically by powers of 10. (a) and (b) The relation  $\langle d^n \rangle / \langle d^{n-1} \rangle_a \propto a^{H_d/H_f}$  with mean exponent of 0.46 and standard deviation of 0.02. (c) and (d) The relation  $\langle d^n \rangle / \langle d^{n-1} \rangle_a \propto \langle a \rangle_a^{d_f}$  with mean exponent equal to  $1.14 \pm 0.02$ .

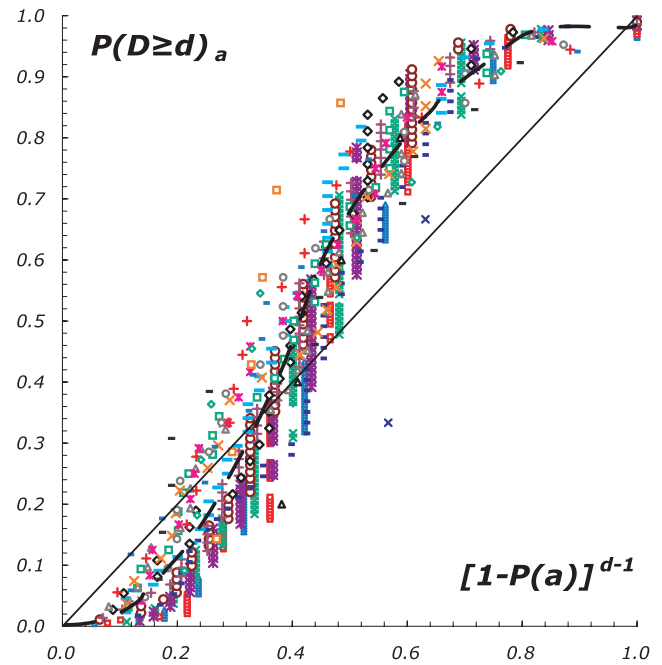
vertically for convenience) is a solid test for the scaling argument adopted.

[27] Figure 11 illustrates an a posteriori test of the validity of the statistical independence assumption, i.e.,  $P(D \geq d)_a \propto P(A < a)^{d-1}$ , and is carried out by plotting directly the probability of exceeding a distance  $d$  given the threshold area  $a$  with the theoretical value  $(1 - P(A \geq a))^{d-1}$  valid under strict statistical independence. We note, that, as we

expected, deviations appear both at large and small areas, thus denying generality to the position. Notice, however, that although the independence argument would have called for a scaling structure for  $P(D \geq d)_a$  which is actually unsupported by our data, differences do not seem overwhelming. Note that all relationships between tributaries' areas and distances are then rightly derived through the finite-size scaling assumption rather than by the expression postulating statistical independence.

#### 4. Discussion

[28] The collapse of all conditional distributions of distance  $P(D \geq d)_a$  when plotted against  $d/a^{H/(1+H)}$  or  $d/\langle a \rangle_a^{d_f}$  (depending on whether drainage is allowed along a mainstream or a multiple-outlet line boundary) defines a general and interesting ecological measure. As predicted by equation (16), the distribution is shown to exhibit a finite-size scaling form that derives from the probability distribution of contributing area at any point along a given boundary or mainstream, and its related scaling exponents  $\tau_{ms}$ ,  $\tau_b$ . The major effect of the constraints that can be imposed by the draining boundary is not surprising because areas tributary to mainstreams or boundaries define a different population of random variables.



**Figure 11.** Independence test between probabilities of areas and distances. We have assumed the possibility to simplify the relation  $P(D \geq d)_a (1 - P(A \geq a))^{(d-d_c)} \theta(d - d_c(a))$  assuming  $d_c = 1$  and neglecting the nonlinear term  $\theta(d, a)$  with  $d_c$  that can be approximated by the distance  $d \propto a^{H_d/H_f}$ . On the basis of the obtained curves we can infer the independence between areas along the mainstream at distances longer than  $a^{H_d/H_f}$ , so we have great areas with great distances only. In this case,  $P(A \geq a)$  is the exact value found on the complete exceedance probability distribution of areas (for  $a = 1$ ), relative to each exceedance probability distribution of distances  $P(D \geq d)_a$ .

[29] Most important is deemed the “universal” behavior of  $P(D \geq d)_a$  against  $d/a^\psi$  or, equivalently,  $d/\langle a \rangle^\phi$  (with appropriate  $\psi$ ,  $\phi$  depending on B.C., see Table 1 for a complete comparison and a detailed summary of scaling relationships), where  $a$ ,  $\langle a \rangle$  indicate, respectively, the threshold area for a tributary and the mean area of all tributaries whose area is larger than  $a$  (see Figure 1). The collapse of all distributions that are clearly distinct for different values of  $a$ , onto a single universal curve suggests that the relationship  $P(D \geq d)_a \sim F(d/a^\psi)$  is valid for all areas  $a$  and thus seemingly of general nature (Figures 5 and 6). This is a distinctive statistical tool that has notable implications. In fact, the traditional geomorphological tenet that assumes link lengths as uniformly distributed [e.g., see *Leopold et al.*, 1964] is incompatible with our findings that pose a few conceptual and practical problems. Uniformly distributed link lengths do not seem to hold, in fact, on one hand because given a threshold area for the definition of a meaningful tributary, no observational data set indeed conforms to some kind of uniform distribution: the cutoff is inevitably far from a step function. On the other hand, the distribution cannot escape the problem of the choice of  $a$ , which makes the very definition of link quite arbitrary otherwise. On this basis alone we suggest that the concept needs be readdressed in the present framework. Obviously this has implications for biodiversity and ecosystem services provided by the river basin, as the confluence of tributary areas is crucial to flow, width and riparian area at any basin site.

[30] The analytical specification of the cutoff function  $F(d/\langle a \rangle^\phi)$  differs from multiple-outlet boundaries of main-streams for the fact that allowing drainage from either sides naturally shortens the distance between comparable (or larger) confluences. We notice, however, that the scaling ansatz  $d \propto a^{H/(1+H)}$  seems to work properly also in this case, thereby suggesting that the distribution of distances along a support draining from either side is a obtained by rescaling the analog for an open boundary by some constant (Figure 10). Our guess of analytic form of the scaling function is  $F(x) = C \exp(-\beta x)$  where  $\beta = 1/x_c$  is the exponent found via linear estimates on semilog plot between  $F(x)$  and  $x$ . In such case  $x_c$  acts as a cutoff value of the distribution of  $d/\langle a \rangle^\phi$ . The numerical estimates produced for  $C$ ,  $x_c$  are described in Section 3 and are possibly general. Further confirmations from data will be sought elsewhere.

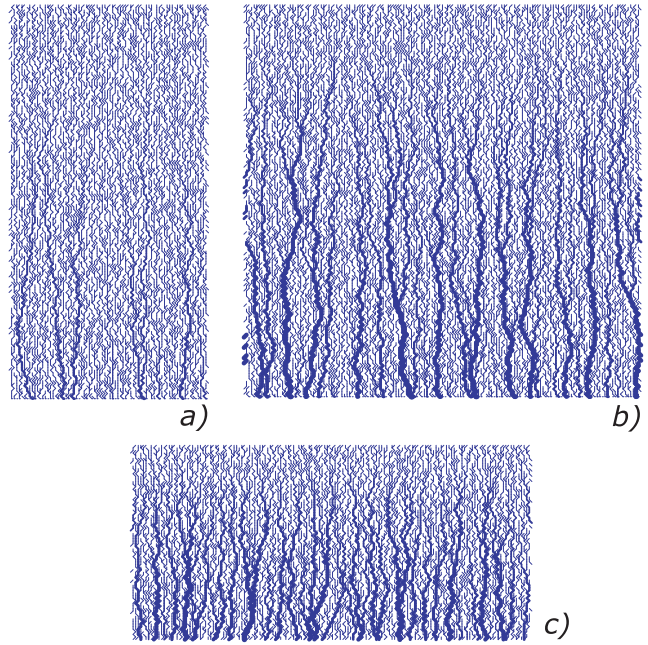
[31] The computation of mean and variance of the empirical distances matches nicely our theoretical predictions. The mean distance between tributaries of area larger than  $a$  is definitely proportional to  $\langle a \rangle^H$  (hence often  $\langle d \rangle_a \approx \langle a \rangle$ ), whereas the mean distance plotted against  $a$  obeys the predicted scaling, resulting in  $\langle d \rangle_a \sim a^{H/(1+H)} \approx a^{0.52 \pm 0.04}$  except for singular cases like Scheidegger's (where the exponent is close to the predicted value of 1/3 (see Appendix A)). The theoretical predictions, based on two different arguments, capture well the structure of the distribution.

[32] Overall the probabilistic structure of tributaries thus seems fully captured.

## 5. Conclusions

[33] The following conclusions are worth mentioning.

[34] 1. The probabilistic structure of the alongstream distance between tributaries of (any) comparable size  $a$  drawn from a river network has been defined theoretically



**Figure A1.** Scheidegger networks arranged, clockwise, with increasing aspect ratios, i.e., of perpendicular versus lateral dimension of the tributaries; specifically, it is in growing order  $m = L_\perp/L_\parallel =$  (a) 1/10, (b) 1, and (c) 10. Note that this geometric and topological arrangement is seen as a demanding test for the theoretical results owing to the pronounced anisotropy of the network forms (see Figure A2).

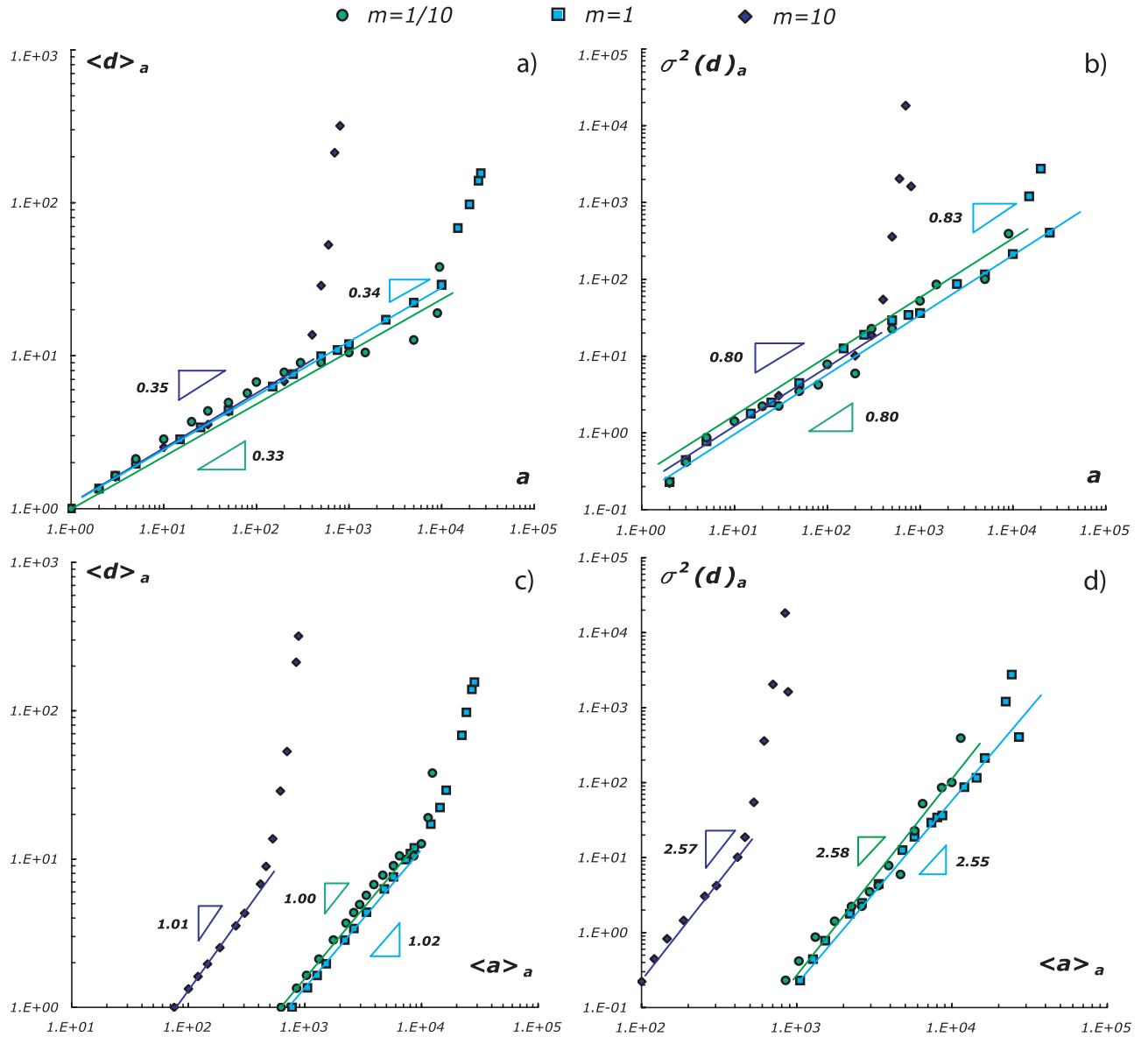
and verified observationally. Its importance is linked to the ecological implications of the availability of flow, width and riparian area available at any stream site, all bearing implications for the type of biodiversity and ecosystem services produced by the river basin.

[35] 2. The pdf of distance depends on the threshold size needed to define a tributary; obviously if a unit area is employed, the distance turns out to be unit (i.e., the link length is also unit) and grows, in a statistically well-defined sense, as basins grow larger. Interestingly, however, when the distribution is plotted against  $d/a$  or  $d/\langle a \rangle$  ( $\langle a \rangle$  being the mean area of all basins larger than  $a$ ) collapses onto a single curve. Notice that the dependence of link length on  $a$  proves that the traditional tenet that assumes uniform link length distributions is a spinoff of the rather arbitrary threshold for channelization used to extract fluvial patterns, and cannot hold conceptually as it depends on the choice of cumulated area that defines a tributary.

[36] 3. The distribution of distances that separate a given average injection of flow, width or riparian area (i.e., confluent area) seems to represent more general a concept than that of link length. Our empirical findings support well the theoretical arguments. In particular, exactly solved network constructions like Peano or Scheidegger networks (whose merits and limits in describing the geometry of river networks are described elsewhere but are deemed noteworthy) have been extensively used to validate our findings.

[37] 4. Overall we have proposed an explicit form of the probabilistic structure of tributaries in the form  $P(\geq d) = P(D \geq d)_a = \exp(-Cd/a^{H/(1+H)})$  where the constant  $C$  is dependent on the choice of boundary conditions (i.e.,





**Figure A2.** Scaling relations between  $\langle d \rangle_a$  and  $a$  and  $\langle a \rangle_a$  for the Scheidegger networks shown in Figure A1. Here various aspect ratios are used, specifically in growing order with  $m = L_{\perp}/L_{\parallel} = 1/10, 1$ , and  $10$ . The first case is perhaps the most appropriate to describe the exact results of the exponent related to the tributaries distribution, i.e.,  $\frac{Hd_f}{1+H} = 1/3$ , because it eases the formation of subbasins with the exact value  $H = 1/2$ . Other lattice ratios lead to a deviation from the exact values of exponents and deviations appear as shown because of the slow logarithmic growth of the construct. The largest deviations relate to the highest value of  $m$ . The computational relations (a)  $\langle d \rangle_a$  versus  $a$  and (c)  $\langle d \rangle_a$  versus  $\langle a \rangle_a$ . (b) and (d) The scaling of their related variance.

whether mainstream areas of multiple-outlet line sources are considered).

### Appendix A: Scheidegger's Construction and the Distance Between Tributaries

[38] Testing the theory against Scheidegger's construct proves particularly demanding and interesting, owing to the strong anisotropy that such structure produces, and to the slow convergence to exact power law statistics [Huber, 1991; Takayasu et al., 1991; Rodriguez-Iturbe and Rinaldo, 1997]. It also relates to subtrees rooted along the primary path of a binary tree [Troutman and Karlinger, 1993]. In

this context, we have performed our analyses for different networks varying the ratio, say  $m$ , between the perpendicular and the parallel main axes. Specifically, we have carried out the same analysis on  $\langle d \rangle_a$  as shown in Figures 8a and 8b and on  $\sigma(d)$  in Figures 9b and 9d. Figure A1 shows, clockwise, Scheidegger's networks with  $m = 1/10, 1, 10$ .

[39] It is clear that statistical measures of distances for the Scheidegger model are strongly dependent on the size of the lattice employed. It is relevant to note the deviation from the predicted value of the exponent  $\frac{H}{1+H} = 1/3$  (equation (22) and Figure A2a) for  $\langle d \rangle$  in function of the parameter  $a$ , and from the exponent  $d_f = 1.00$  (equation (22) and Figure A2c) for  $\langle d \rangle$  in function of the mean area  $\langle a \rangle_a$ .

Same deviations are found in the variance  $\sigma(d)^2$  in function of  $a$  and  $\langle a \rangle$  (Figures A2b and A2d). The slow logarithmic growth of Scheidegger's construct [Huber, 1991] is, in fact, equivalent to the use of a model with a high ratio  $m$ , which leads to incomplete aggregation. Several analyses have been performed on Scheidegger basins with very low value of  $m$  (for example using models with global area  $1000 \times 100$  and  $5000 \times 100$ ), and the statistics of areas at the boundary match the theoretical results. Otherwise exponents gradually deviate from the exact values (Figure A2). The explanation for these deviation lies on the low value of the Hurst exponent characterizing Scheidegger's construct,  $H = 1/2$ . Only considering low values of  $m$ , in fact, or equivalently considering a very long network, it is possible to tend to reproduce the correct statistics between tributaries distances and areas, a phenomenon fully understood theoretically by Huber [1991]. Therefore we note, also from Figure 3c, that boundary areas (whose distribution is a power law with exponent  $\tau_b$ ) and global areas (whose scaling exponent is  $\tau$ ) exhibit both the typical behavior as power laws, and are not susceptible to the level of growth of the stochastic construct. In this way only finite-size effects alter the scaling, and this happens is particularly significant for high values of  $m$ .

[40] **Acknowledgments.** Funding from AQUATERRA EU Project GOCE-2020 and PRIN 2006 Fenomeni di Trasporto nel Ciclo Idrologico and support from the University of Padua are gratefully acknowledged. The writers wish to thank Sergio Fagherazzi and two anonymous reviewers for their valuable comments on the first version of the manuscript.

## References

- Ammermann, A. J., and L. L. Cavalli-Sforza (1984), *The Neolithic Transition and the Genetics of Population in Europe*, Princeton Univ. Press, Princeton, N. J.
- Banavar, J. R., J. Damuth, A. Maritan, and A. Rinaldo (2000), Topology of the fittest transportation network, *Phys. Rev. Lett.*, **84**, 4745–4748.
- Banavar, J. R., F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo (2001), Scaling, optimality and landscape evolution, *J. Stat. Phys.*, **104**, 1–33.
- Banavar, J. R., J. Damuth, A. Maritan, and A. Rinaldo (2007), Scaling in ecosystems and the linkage of macroecological laws, *Phys. Rev. Lett.*, **98**, 068104, doi:10.1103/PhysRevLett.98.068104.
- Bertuzzo, E., A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo (2007), River networks and ecological corridors: Reactive transport on fractals, migration fronts, hydrochory, *Water Resour. Res.*, **43**, W04419, doi:10.1029/2006WR005533.
- Campos, D., J. Fort, and V. Méndez (2006), Transport on fractal river network: Application to migration fronts, *Theor. Popul. Biol.*, **69**, 88–93.
- Colaiori, F., A. Flammini, J. R. Banavar, and A. Maritan (1997), Analytical and numerical study of optimal channel networks, *Phys. Rev. E*, **55**, 1298–1310.
- Fort, J., and V. Méndez (1999), Time-delayed theory of the neolithic transition in Europe, *Phys. Rev. Lett.*, **82**, 867–870.
- Holmes, E. E. (1993), Are diffusion models too simple? A comparison with telegraph models of invasion, *Am. Nat.*, **142**, 779–795.
- Hubbell, S. P. (2001), *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton Univ. Press, Princeton, N. J.
- Huber, G. (1991), Scheidegger's rivers, Takayasu's aggregates and continued fractions, *Physica A*, **170**, 463–469.
- Hughes, B. D. (1995), *Random Walks and Random Environments*, Oxford Univ. Press, Oxford, U. K.
- Leopold, L. B., M. G. Wolman, and J. P. Miller (1964), *Fluvial Processes in Geomorphology*, W. H. Freeman, San Francisco, Calif.
- Mandelbrot, B. B. (1983), *The Fractal Geometry of Nature*, W. H. Freeman, San Francisco, Calif.
- Marani, A., R. Rigon, and A. Rinaldo (1991), A note on fractal channel networks, *Water Resour. Res.*, **27**, 3041–3049.
- Maritan, A., A. Rinaldo, I. Rodriguez-Iturbe, R. Rigon, and A. Giacometti (1996), Scaling in river networks, *Phys. Rev. E*, **53**, 1501–1513.
- Méndez, V., D. Campos, and S. Fedotov (2004), Front propagation in reaction-dispersal models with finite jump speed, *Phys. Rev. E*, **70**, 036121, doi:10.1103/PhysRevE.70.036121.
- Muneepeerakul, R., J. Weitz, S. A. Levin, A. Rinaldo, and I. Rodriguez-Iturbe (2007a), A neutral metapopulation model of biodiversity in river networks, *J. Theor. Biol.*, **245**, 2, 351–363.
- Muneepeerakul, R., S. A. Levin, A. Rinaldo, and I. Rodriguez-Iturbe (2007b), On biodiversity in river networks: A trade-off metapopulation model and comparative analysis, *Water Resour. Res.*, **43**, W07426, doi:10.1029/2006WR005857.
- Rigon, R., I. Rodriguez-Iturbe, A. Maritan, A. Giacometti, D. G. Tarboton, and A. Rinaldo (1996), On Hack's law, *Water Resour. Res.*, **32**, 3367–3374.
- Rigon, R., I. Rodriguez-Iturbe, and A. Rinaldo (1998), Feasible optimality implies Hack's law, *Water Resour. Res.*, **34**, 3181–3189.
- Rinaldo, A., I. Rodriguez-Iturbe, R. Rigon, R. L. Bras, E. Ijjasz-Vasquez, and A. Marani (1992), Minimum energy and fractal structures of drainage networks, *Water Resour. Res.*, **28**, 2183–2191.
- Rinaldo, A., I. Rodriguez-Iturbe, R. Rigon, E. Ijjasz-Vasquez, and R. L. Bras (1993), Self-organized fractal river networks, *Phys. Rev. Lett.*, **70**, 1222–1226.
- Rinaldo, A., I. Rodriguez-Iturbe, and R. Rigon (1998), Channel networks, *Annu. Rev. Earth Planet. Sci.*, **26**, 289–306.
- Rinaldo, A., J. R. Banavar, and A. Maritan (2006), Trees, networks and hydrology, *Water Resour. Res.*, **42**, W06D07, doi:10.1029/2005WR004108.
- Rodriguez-Iturbe, I., and A. Rinaldo (1997), *Fractal River Basins: Chance and Self-Organization*, Cambridge Univ. Press, New York.
- Rodriguez-Iturbe, I., E. Ijjasz-Vasquez, R. L. Bras, and D. G. Tarboton (1992a), Power-law distribution of mass and energy in river basins, *Water Resour. Res.*, **28**, 988–993.
- Rodriguez-Iturbe, I., A. Rinaldo, R. Rigon, R. L. Bras, and E. Ijjasz-Vasquez (1992b), Energy dissipation, runoff production and the three dimensional structure of channel networks, *Water Resour. Res.*, **28**, 1095–1103.
- Scheidegger, A. E. (1967), A stochastic model for drainage patterns into an intramontane trench, *Bull. Assoc. Hydrol.*, **12**, 15–20.
- Takayasu, H., M. Takayasu, A. Provata, and G. Huber (1991), Statistical models of river networks, *J. Stat. Phys.*, **65**, 725–740.
- Tilman, D., and J. A. Downing (1994), Biodiversity and stability in grasslands, *Nature*, **367**, 363–365.
- Troutman, B. M., and M. R. Karlinger (1993), A note on subtrees rooted along the primary path of a binary tree, *Discuss. Appl. Math.*, **42**, 87–93.
- Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby (2006), Bivariate line-fitting methods for allometry, *Biol. Rev.*, **81**, 259–291.

M. Convertino and A. Rinaldo, Dipartimento IMAGE, Università di Padova, via Loredan 20, I-35131 Padova, Italy. (rinaldo@idra.unipd.it)

A. Maritan, Dipartimento di Fisica “Galileo Galilei”, Università di Padova, I-35131 Padova, Italy.

R. Rigon, Dipartimento di Ingegneria Civile e Ambientale and Centro Universitario per la Difesa Idrogeologica Dell'Ambiente Montano (CUDAM), Università di Trento, I-35080 Trento, Italy.

I. Rodriguez-Iturbe, Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA.