Supplementary Information: Optimal Channel Networks accurately model ecologically-relevant geomorphological features of branching river networks

Luca Carraro[1,2,*], Florian Altermatt[1,2]

[1] *Department of Aquatic Ecology, Swiss Federal Institute of Aquatic Science and Technology, Eawag, 8600 Dübendorf, Switzerland*

[2] *Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zürich, Switzerland*

[*] *corresponding author: luca.carraro@eawag.ch*

Contents

Supplementary Note 1 **2**

Supplementary Methods **3**

- Branching probability and branching ratio of a BBT 3
- Details on the algorithm for generating RBNs 4
- Details on Fig. 5 5

Supplementary Tables **6**

Supplementary Figures **9**

Supplementary References **18**

Supplementary Note 1

This is a glossary of the main technical terms used in the manuscript. An illustration for most of these terms is provided in Supplementary Figure 11.

Node. The fundamental unit into which a river network is discretized. From a geomorphological perspective, it has the extent of a pixel of the digital elevation model. From an ecological perspective, it has the extent of a local, homogeneous community, possibly limited by physical constrains (e.g. riffle-pool sequences[1]); in metapopulation/metacommunity lexicon, it is often referred to as a ”patch”.

Source. A node with no upstream connections (i.e., whose indegree is 0). In hydrological terms, it is sometimes referred to as a ”headwater”.

Confluence. A node with two upstream connections (i.e., whose indegree is 2). It is sometimes referred to as a ”branching node”.

Root (or outlet). A node with no downstream connections (i.e., whose outdegree is 0).

Link. A sequence of flow-connected nodes that originates from a source (or confluence) and ends at the next confluence (or root).

Indegree. The number of edges directed towards a node in a directed graph.

Outdegree. The number of edges exiting from a node in a directed graph.

Spanning tree. A directed graph containing all nodes of a region (”spanning”) and in which there are no loops (”tree”).

Metapopulation. A set of spatially structured populations of the same species that are connected by dispersal.

Supplementary Methods

Branching probability and branching ratio of a BBT

Ref.[2] found (see Eq. (2) therein with $\kappa = 2$) that, for a BBT with branching probability p, the expected number of nodes is:

$$E[N] = \frac{(1+p)^{\Omega} - 1}{p}, \tag{S1}$$

where Ω is the maximum order (i.e. the distance from a source node to the outlet) of the network. Note that in our application the number of nodes N is fixed (equal to the N of a corresponding OCN), which results in BBT networks where the maximum order Ω is not "complete", i.e. some source nodes belong to the order $\Omega - 1$ (as the generating algorithm is stopped as soon as the number of network nodes is equal to N). Conversely, ref.[2] used networks where the maximum order Ω was "complete": N was not determined a priori, but resulted from a stochastic process where p and Ω were fixed. As a result, Eq. (S1) provides an overestimation of N with respect to the values used in this application (because of the "completeness" assumption under which Eq. (S1) was derived). We here aim for an approximated formulation, and hence neglect these differences and hypothesize that $N \approx E[N]$. The validity of our approximation is warranted by noting that the resulting values of branching ratio p_{r} for the generated BBTs are aligned with those observed for the corresponding OCNs (Supplementary Figure 8).

From Eq. (S1), it is

$$\Omega = \frac{\log(pN+1)}{\log(1+p)}. \tag{S2}$$

The expected number of sources $E[N_{\mathrm{S}}]$ is equal to the difference between N and the expected number of nodes of order lower than Ω:

$$E[N_{\mathrm{S}}] = N - \frac{(1+p)^{\Omega-1} - 1}{p}. \tag{S3}$$

By substituting Eq. (S2) in Eq. (S3), one gathers

$$E[N_{\mathrm{S}}] = \frac{pN-1}{1+p}. \tag{S4}$$

The expected number of links is equal to the expected number of source nodes plus the expected number of confluence nodes:

$$E[N_{\mathrm{L}}] = E[N_{\mathrm{S}}] + p(1 - E[N_{\mathrm{S}}]) = \frac{2pN+1-p}{1+p} \tag{S5}$$

Hence, the expected branching ratio reads:

$$E[p_{\mathrm{r}}] = \frac{E[N_{\mathrm{L}}]}{N} = \frac{2pN+1-p}{N(1+p)}. \tag{S6}$$

Solving for p yields

$$p = \frac{E[p_{\mathrm{r}}]N - 1}{(2 - E[p_{\mathrm{r}}])N - 1}, \tag{S7}$$

which is the branching probability value that must be used in the BBT generating algorithm to create a network with expected branching ratio $E[p_{\mathrm{r}}]$. When N is very large (as is the case in this application), Eq. (S8) can be simplified as

$$p = \frac{E[p_{\mathrm{r}}]}{2 - E[p_{\mathrm{r}}]}. \tag{S8}$$

We generated BBTs by using Eq. (S8), where $E[p_{\mathrm{r}}]$ was set equal to the branching ratio value of a corresponding OCN.

Details on the algorithm for generating RBNs

First of all, we need to generate a sequence of link length values from a geometrical distribution with mean $1/p_{\mathrm{r}}$ such that their sum be equal to N. To achieve this, we generate a sequence of random values $x_i,\ i = 1, \ldots, 2N$ (the length of the sequence is actually arbitrary, but it needs be considerably larger than $p_{\mathrm{r}}N$) from the aforementioned distribution. Such sequence represents the tentative distribution of link lengths in the RBN. We then consider the first part of such sequence, i.e. the first n_{L} values, where n_{L} is such that

$$\sum_{i=1}^{n_{\mathrm{L}}-1} x_i < N \leq \sum_{i=1}^{n_{\mathrm{L}}} x_i. \tag{S9}$$

We then distinguish three cases:

1. If $\sum_{i=1}^{n_{\mathrm{L}}} x_i = N$ and n_{L} is uneven, then the selected sequence $x_i,\ i = 1, \ldots, n_{\mathrm{L}}$ is the actual distribution of link lengths, and we impose $N_{\mathrm{L}} = n_{\mathrm{L}}$.

2. If $\sum_{i=1}^{n_{\mathrm{L}}} x_i > N$ and n_{L} is uneven, then the sequence of link lengths (excluding the last term of the sequence) is constituted by $x_i,\ i = 1, \ldots, n_{\mathrm{L}} - 1$, while we impose $x_{n_{\mathrm{L}}} = N - \sum_{i-1}^{n_{\mathrm{L}}-1} x_i$ and $N_{\mathrm{L}} = n_{\mathrm{L}}$.

3. If n_{L} is even, we discard $x_{n_{\mathrm{L}}-1}$ and $x_{n_{\mathrm{L}}}$, and replace them with the first two terms of the unused portion of the randomly generated sequence. We then re-evaluate n_{L} via Eq. (S9) until case 1 or 2 are satisfied.

Note indeed that N_{L} must be uneven by construction (see also Fig. 1a,b).

Once the sequence of link lengths is established, the network is assembled as follows: first, one link is randomly picked as the most downstream one; at each following step, a pair of unattributed links is randomly picked and connected to a randomly chosen network link having no upstream connection yet. The procedure ends when all links have been attributed to the network.

Details on Fig. 5

In Fig. 5a, concerning the evaluation of bifurcation ratios for the different network types, each boxplot contains up to 50 points, depending on whether all 50 networks of a given type had at least one link of the relative stream order. Boxplot elements are as follows: center line, median; box limits, upper and lower quartiles; whiskers, extending up to the most extreme data points that are within $\pm 1.5 \cdot \mathrm{IQR}$, where IQR is the interquartile range; circles, outliers. The displayed trend lines are fitted on the ensemble of 50 networks. Horton's bifurcation ratios R_{B} were calculated for each network as $\exp(-\gamma)$, where γ is the slope of the aforementioned trend line. Displayed R_{B} values are mean $\pm$ standard deviation across the 50 networks.

In Fig. 5b, concerning the evaluation of length ratios, plot construction is as in Fig. 5a, except that here trend lines are fitted by discarding values corresponding to the largest stream order $\overline{\omega}$ for each network type (as the network might not be representative of a fully developed $\overline{\omega}$-th order basin[3]). Similarly, Horton's length ratios R_{L} were calculated for each network as $\exp(\gamma)$, where γ is the slope of the linear regression on a semilog L_{ω} - ω plot, where values corresponding to the largest stream order were excluded. Displayed R_{L} values are mean $\pm$ standard deviation across the 50 networks.

In Fig. 5c, rivers and OCNs were obtained with $A_{\mathrm{T}} = 1$ pixel; RBNs and BBTs were derived accordingly (i.e., by calculating 50 (N, p_{r}) pairs for the OCNs extracted with $A_{\mathrm{T}} = 1$, and then generating a BBT and a RBN for each so-obtained (N, p_{r}) pair.). For RBNs and BBTs, drainage area values are evaluated as number of nodes upstream of a node. In Supplementary Table 1 and Supplementary Figure 3 values of β fitted separately for each real river network are reported. For each network type, the vectors of drainage area values across the 50 replicates are pooled; dots represent the exceedance probability for any natural value of a from 1 to the maximum drainage area value contained in the pooled vector. Linear fits are then performed on the subsets of points with a not exceeding 2,000 pixels.

Power-law scaling of drainage areas is recognized as the signature of fractal behaviour in river networks[4]. Power-law distributions are the only scale-free distributions[5]: that is, the ratio between the probability to pick a pixel with area smaller than 20 pixels and the probability to pick one pixel with area smaller than 10 pixels is equal to the ratio between the probabilities to pick pixels with area smaller than 200 and 100 pixels, respectively.

Supplementary Tables

Supplementary Table 1: Values of the bifurcation ratio R_B, the length ratio R_L and the exponent β of the scaling of drainage areas fitted separately for each of the 50 rivers used in the analysis.

ID	River	R_B	R_L	β	ID	River	R_B	R_L	β
1	Wigger	4.362	2.056	0.492	26	French Broad	4.464	2.136	0.489
2	Sense	3.539	2.283	0.458	27	Santa Ynez	4.355	1.846	0.496
3	Toss	4.669	2.303	0.430	28	Yadkin	3.457	1.983	0.561
4	Emme	3.503	2.116	0.479	29	Danube	4.472	2.410	0.514
5	Fella	4.857	2.428	0.438	30	Lillooet	4.540	2.047	0.486
6	Kleine Emme	3.667	2.262	0.405	31	Salzach	4.612	2.568	0.477
7	Chisone	3.703	2.249	0.491	32	Sava	3.498	1.712	0.485
8	Ardeche	3.480	1.909	0.366	33	Rogue	4.355	1.993	0.495
9	Pigg	3.519	2.555	0.362	34	Inn	4.740	2.940	0.439
10	Goose Creek	4.647	1.966	0.510	35	Durance	4.631	2.272	0.500
11	Ararat	3.483	2.127	0.433	36	Adige	4.590	2.330	0.494
12	Carmel	3.532	1.934	0.509	37	Boise	3.518	2.154	0.429
13	Tanaro	4.381	2.579	0.413	38	Trinity	3.519	1.994	0.447
14	Magra	3.470	1.628	0.572	39	New	4.670	2.645	0.469
15	Hinterrhein	4.603	2.581	0.460	40	Eel	4.790	2.883	0.390
16	Thur	3.481	1.972	0.403	41	James	3.463	2.140	0.454
17	Var	4.809	2.838	0.459	42	Salinas	4.933	2.445	0.444
18	Rhone	3.605	2.065	0.493	43	Frances	4.433	2.288	0.451
19	Arc	4.836	2.382	0.465	44	Thompson	3.417	1.695	0.418
20	Ticino	4.820	2.665	0.403	45	Willamette	4.313	2.015	0.518
21	Adda	4.853	2.329	0.467	46	Marias	3.514	2.245	0.499
22	Piave	3.606	2.300	0.446	47	Flathead	4.504	1.776	0.425
23	Dora Baltea	3.556	2.050	0.415	48	Klamath	3.456	1.823	0.459
24	Applegate	4.717	2.016	0.489	49	Owyhee	3.604	2.144	0.547
25	Rivanna	3.515	1.889	0.547	50	Stikine	4.491	1.984	0.474

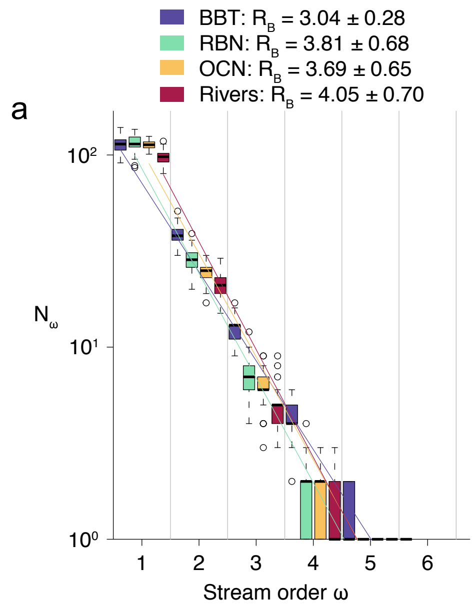
Supplementary Table 2: Information on river basins used in the analysis. ID identifies watersheds in Fig. 2; "no. px": number of pixels; "Zoom" is the value of the *zoom* option in function *get_elev_raster* of *elevatr*; "Country" refers to the country where the river outlet is located; "CH": Switzerland; "Latitude", "Longitude" are the approximate WGS84 coordinates of the river outlet (the actual outlet is placed on the extracted river network, by following the steepest descent path from the approximate outlet).

ID	River	A [km^2]	A [no. px]	l [m]	Zoom	Country	Latitude [°]	Longitude [°]
1	Wigger	367.15	34113	103.74	9	CH	47.28654638	7.93404089
2	Sense	419.75	38311	104.67	9	CH	46.90412746	7.236513547
3	Toss	426.91	40379	102.82	9	CH	47.52846262	8.583028031
4	Emme	445.01	41347	103.74	9	CH	46.95469014	7.753921279
5	Fella	451.66	41162	104.75	9	Italy	46.3966107	13.22661856
6	Kleine Emme	461.69	42897	103.74	9	Italy	47.05332118	8.253575743
7	Chisone	476.05	41253	107.42	9	Italy	44.95289911	7.193817035
8	Ardeche	478.06	41456	107.39	9	France	44.62699325	4.394486774
9	Pigg	534.29	36355	121.23	9	USA	36.92667671	-79.61239861
10	Goose Creek	552.39	37578	121.24	9	USA	37.15747676	-79.47602493
11	Ararat	563.89	37811	122.12	9	USA	36.42398192	-80.56982276
12	Carmel	593.28	39774	122.13	9	USA	36.52305093	-121.8171978
13	Tanaro	1601.82	33817	217.64	8	Italy	44.5780968	7.909598156
14	Magra	1682.68	35524	217.64	8	Italy	44.06387912	9.973121665
15	Hinterrhein	1690.06	38232	210.25	8	CH	46.80828259	9.412934092
16	Thur	1733.39	40624	206.56	8	CH	47.59709053	8.68189986
17	Var	1841.35	40274	213.82	8	France	43.90061037	7.188240385
18	Rhone	1850.30	41846	210.28	8	CH	46.308703	7.827632055
19	Arc	1874.63	42479	210.07	8	France	45.48830382	6.295638562
20	Ticino	1876.39	42436	210.28	8	CH	46.16383105	8.89518776
21	Adda	1925.02	43572	210.19	8	Italy	46.16219969	9.84047307
22	Piave	1929.40	43616	210.32	8	Italy	46.17265146	12.28352335
23	Dora Baltea	1961.58	42867	213.91	8	Italy	45.73786207	7.453772748
24	Applegate	1985.82	39200	225.07	8	USA	42.39723797	-123.4570604
25	Rivanna	1988.46	34578	239.81	8	USA	37.76389972	-78.18496046
26	French Broad	2073.53	34014	246.90	8	USA	35.5654141	-82.5833704
27	Santa Ynez	2141.22	34093	250.61	8	USA	34.66826814	-120.4578087
28	Yadkin	2155.53	35359	246.90	8	USA	36.23928814	-80.84975726
29	Danube	5245.03	33707	394.47	7	Germany	48.36125922	9.962855488

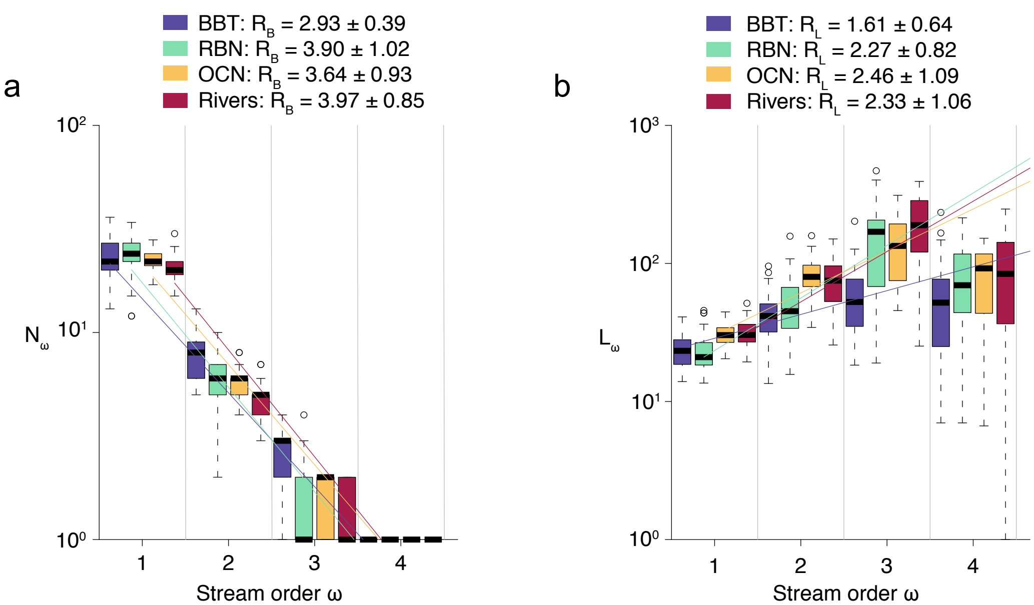
Supplementary Table 2: Information on river basins used in the analysis. ID identifies watersheds in Fig. 2; ”no. px”: number of pixels; ”Zoom” is the value of the *zoom* option in function *get_elev_raster* of *elevatr*; ”Country” refers to the country where the river outlet is located; ”CH”: Switzerland; ”Latitude”, ”Longitude” are the approximate WGS84 coordinates of the river outlet (the actual outlet is placed on the extracted river network, by following the steepest descent path from the approximate outlet).

ID	River	A [km^2]	A [no. px]	l [m]	Zoom	Country	Latitude [º]	Longitude [º]
30	Lillooet	5694.16	39469	379.83	7	Canada	49.8055843	-122.2287205
31	Salzach	6238.97	37253	409.24	7	Austria	47.93972189	12.93863557
32	Sava	6266.67	34840	424.11	7	Slovenia	45.90547135	15.57538896
33	Rogue	6472.04	33580	439.02	7	USA	42.43377682	-123.3525715
34	Inn	6567.87	39214	409.25	7	Austria	47.26853064	11.44602469
35	Durance	6734.73	37483	423.88	7	France	44.10882949	6.007341155
36	Adige	6801.34	40608	409.25	7	Italy	46.47656603	11.31435568
37	Boise	7232.91	37530	439.00	7	USA	43.661357	-116.2791958
38	Trinity	7663.40	37201	453.87	7	USA	41.17459387	-123.7050914
39	New	7787.78	33363	483.14	7	USA	37.31400334	-80.64365526
40	Eel	7887.53	38289	453.87	7	USA	40.45314096	-124.0187222
41	James	8406.80	38282	468.62	7	USA	37.4999327	-79.29489547
42	Salinas	8505.79	36441	483.13	7	USA	36.12801116	-121.0318833
43	Frances	12329.49	39623	557.83	6	Canada	60.85367208	-129.0854007
44	Thompson	19327.20	37627	716.70	6	Canada	50.75150894	-120.3451663
45	Willamette	22386.29	32216	833.60	6	USA	45.21648703	-123.0191039
46	Marias	23165.53	38637	774.32	6	USA	47.93204278	-110.5080309
47	Flathead	23476.71	39156	774.32	6	USA	47.31333546	-114.7048914
48	Klamath	30551.72	38293	893.22	6	USA	41.34860382	-123.5047449
49	Owyhee	30757.21	38560	893.11	6	USA	42.90111597	-117.6988698
50	Stikine	57949.16	36041	1268.02	5	Canada	56.50995498	-132.412977

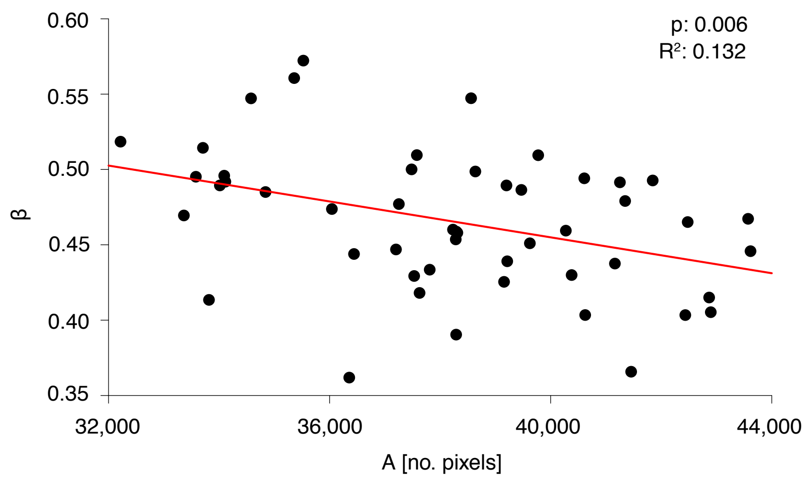
Supplementary Figures



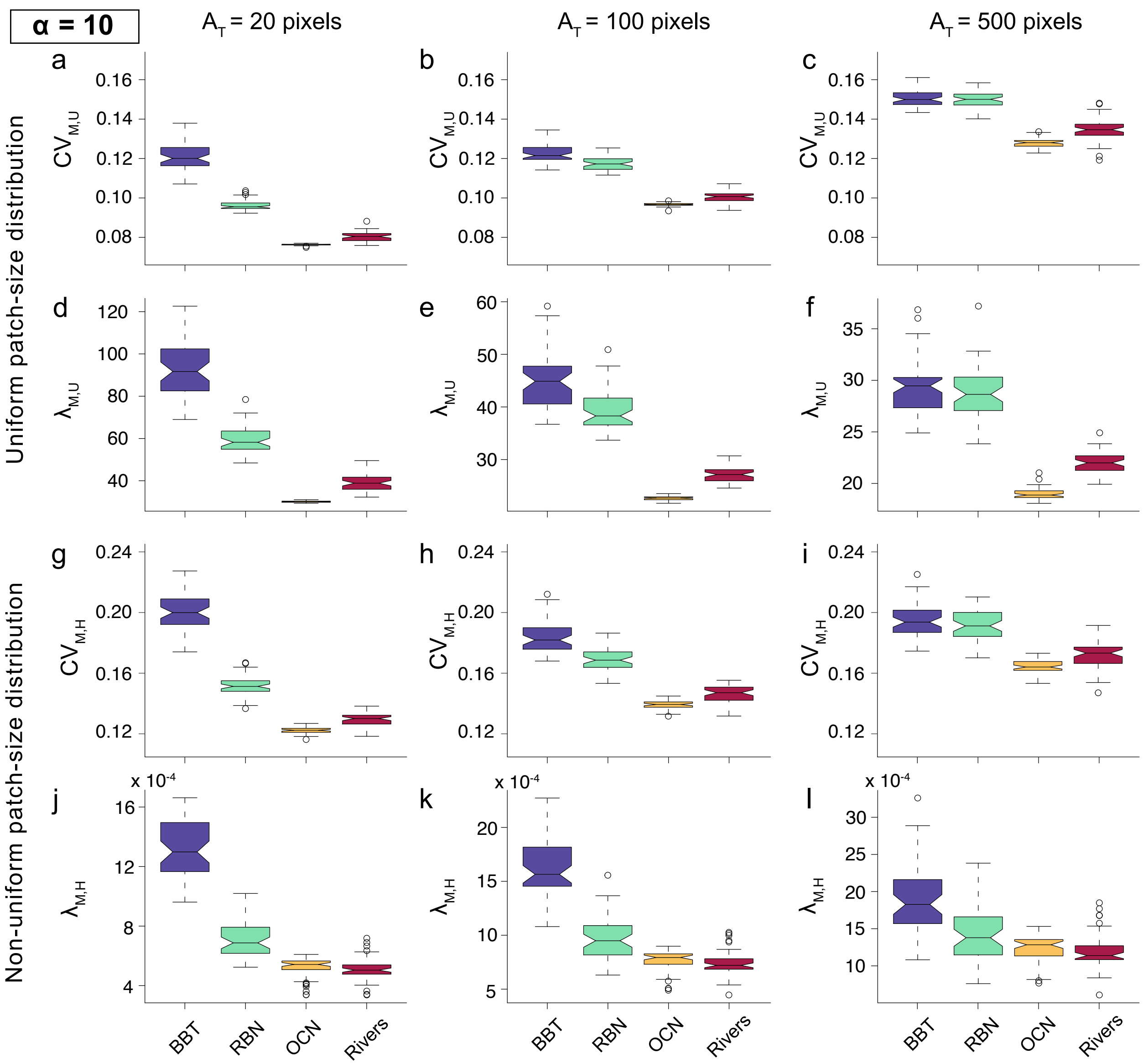
Supplementary Figure 1: Horton's laws for bifurcation (a) and length (b) ratios for the networks derived with $A_T = 100$. Plot construction as in Fig. 5a,b.



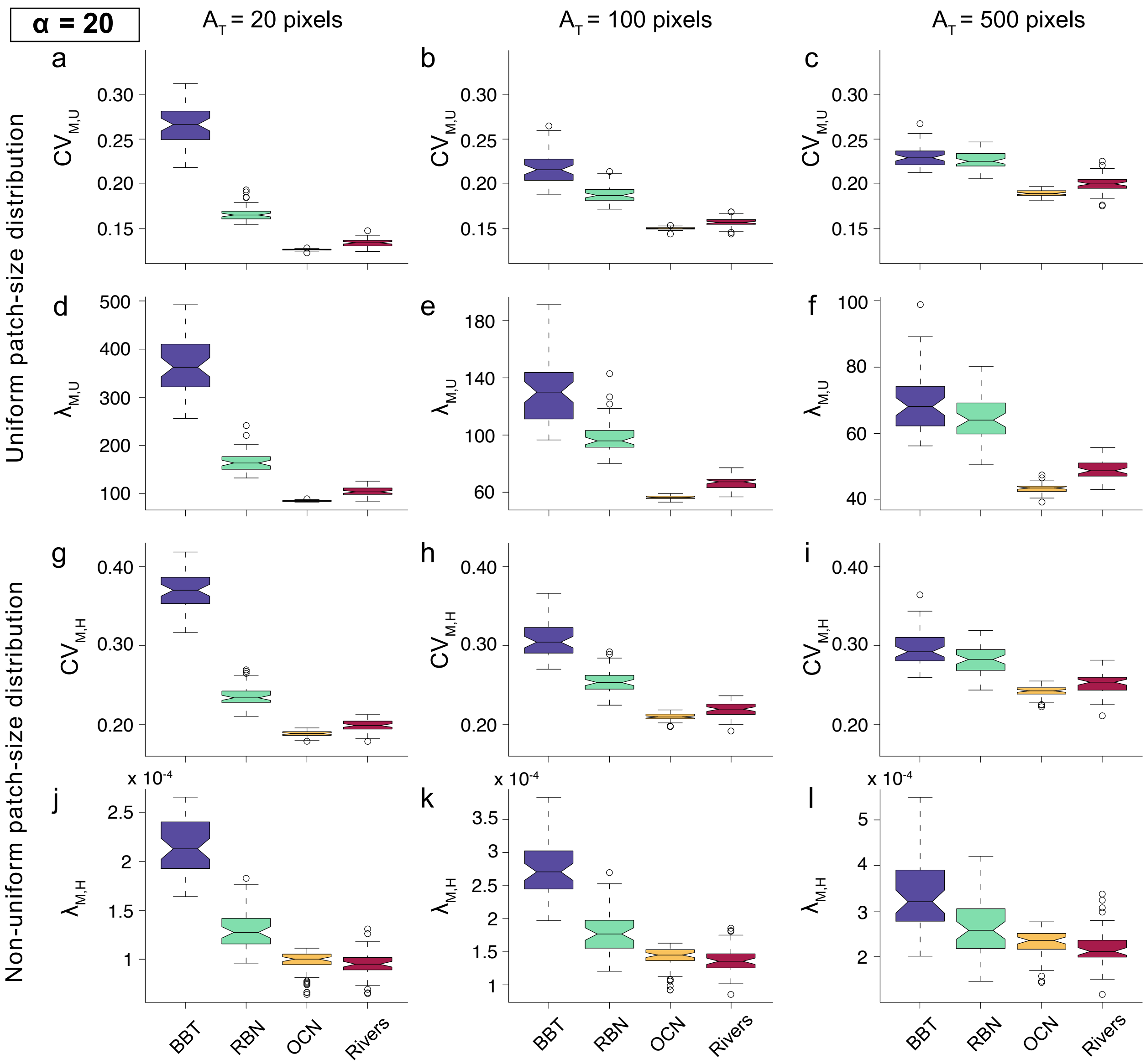
Supplementary Figure 2: Horton's laws for bifurcation (a) and length (b) ratios for the networks derived with $A_T = 500$. Plot construction as in Fig. 5a,b.



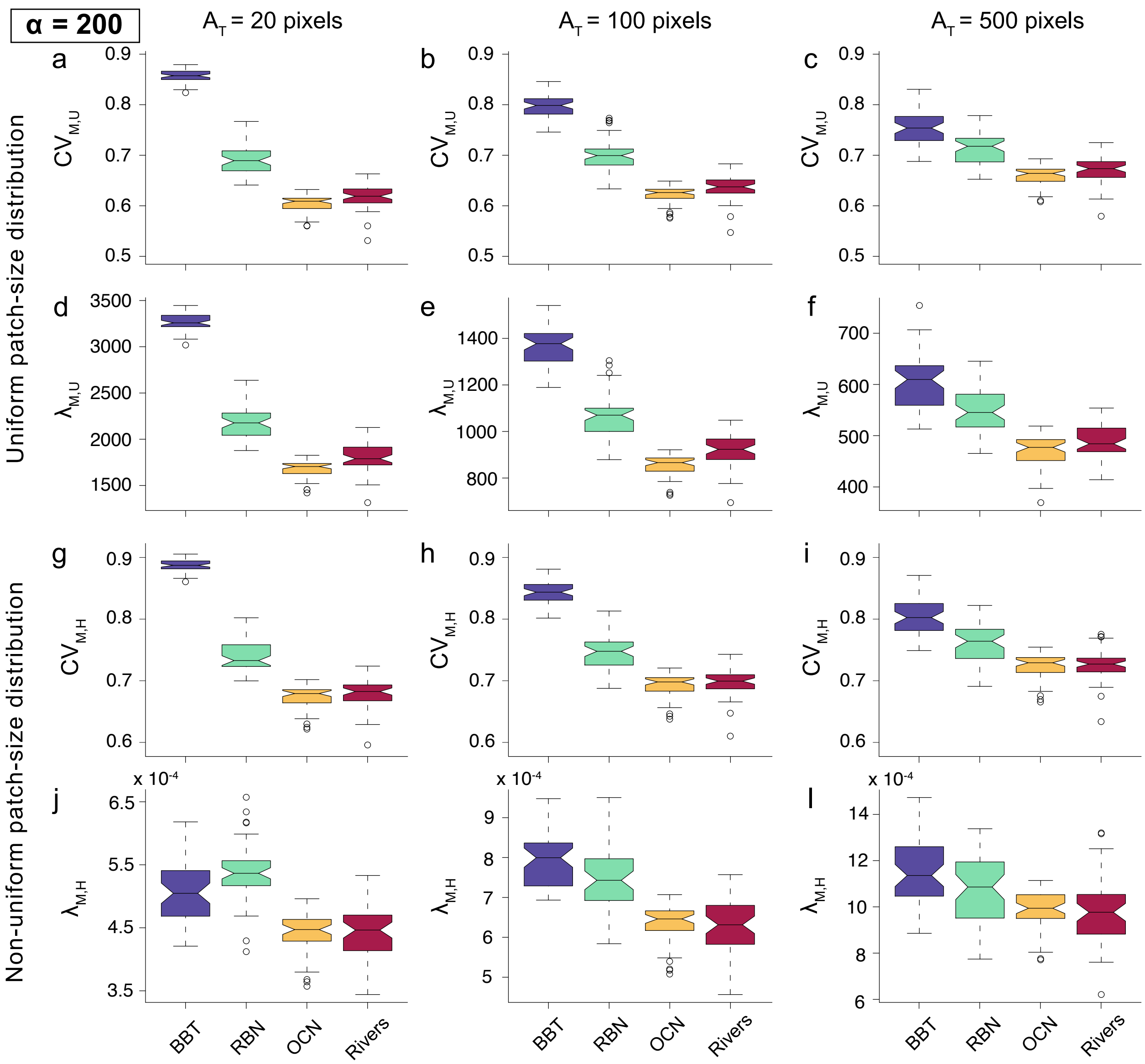
Supplementary Figure 3: Exponents β of the scaling relationship of drainage area fitted separately the 50 rivers used in the analysis (as in Supplementary Table 1) as a function of total drainage area expressed in number of pixels. Black circles identify single river basins; red line shows linear fit. p-value and R^2 of the linear fit are reported. Note that larger basins are associated with a lower value of β, closer to the predicted range[6] $\beta = 0.43 \pm 0.02$.



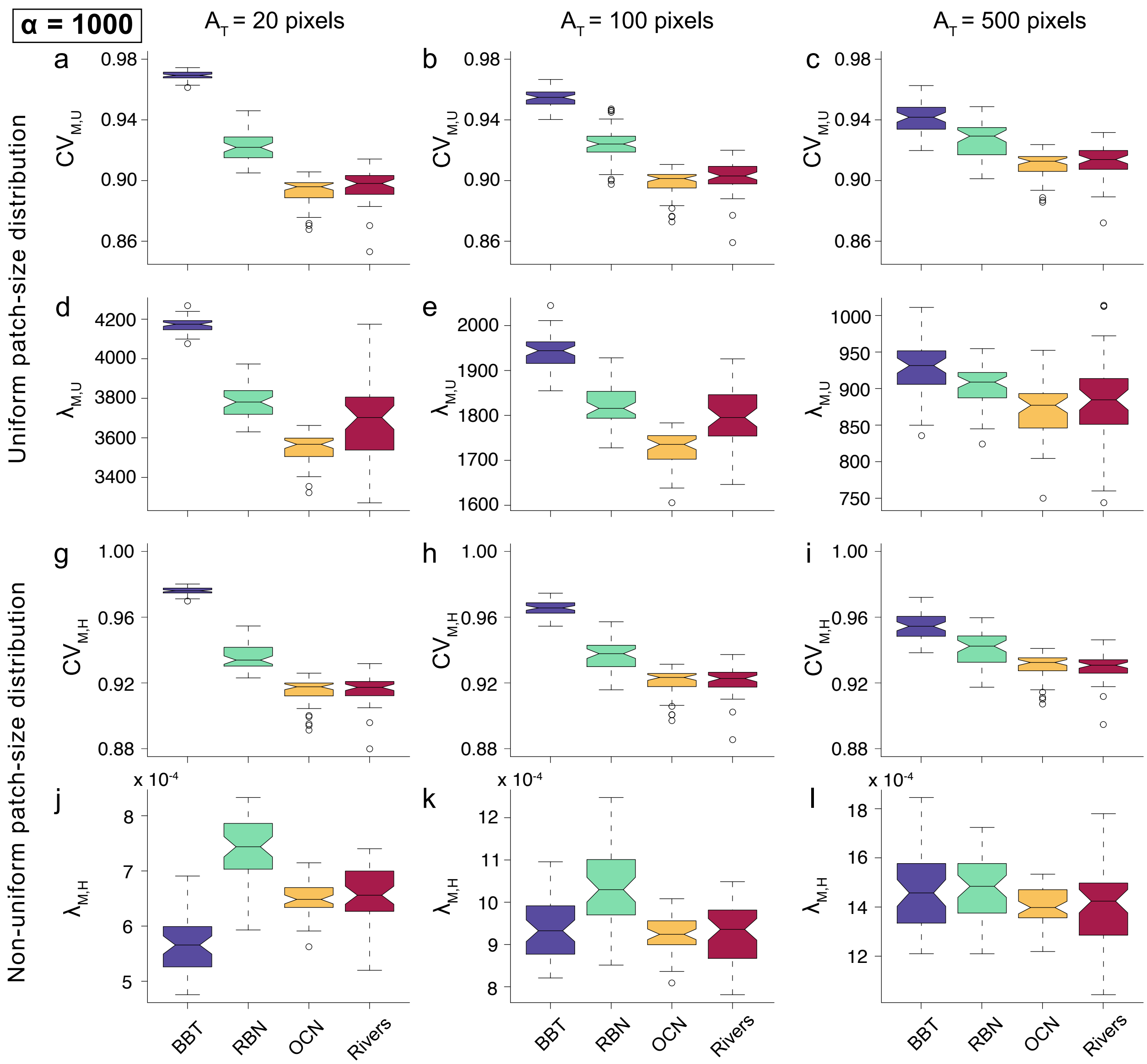
Supplementary Figure 4: Values of metapopulation metrics $CV_{M,U}$, $\lambda_{M,U}$, $CV_{M,H}$ and $\lambda_{M,H}$ obtained by setting $\alpha = 10\,l$. Plot construction as in Fig. 4.



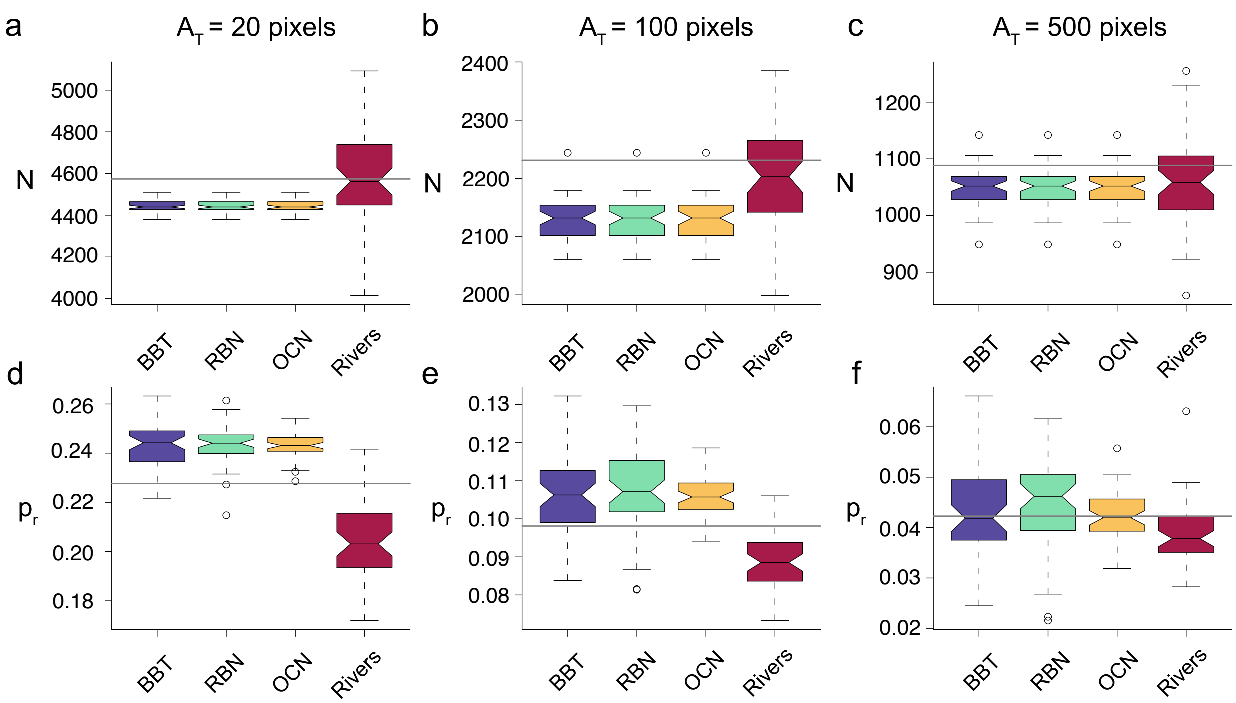
Supplementary Figure 5: Values of metapopulation metrics $CV_{M,U}$, $\lambda_{M,U}$, $CV_{M,H}$ and $\lambda_{M,H}$ obtained by setting $\alpha = 20l$. Plot construction as in Fig. 4.



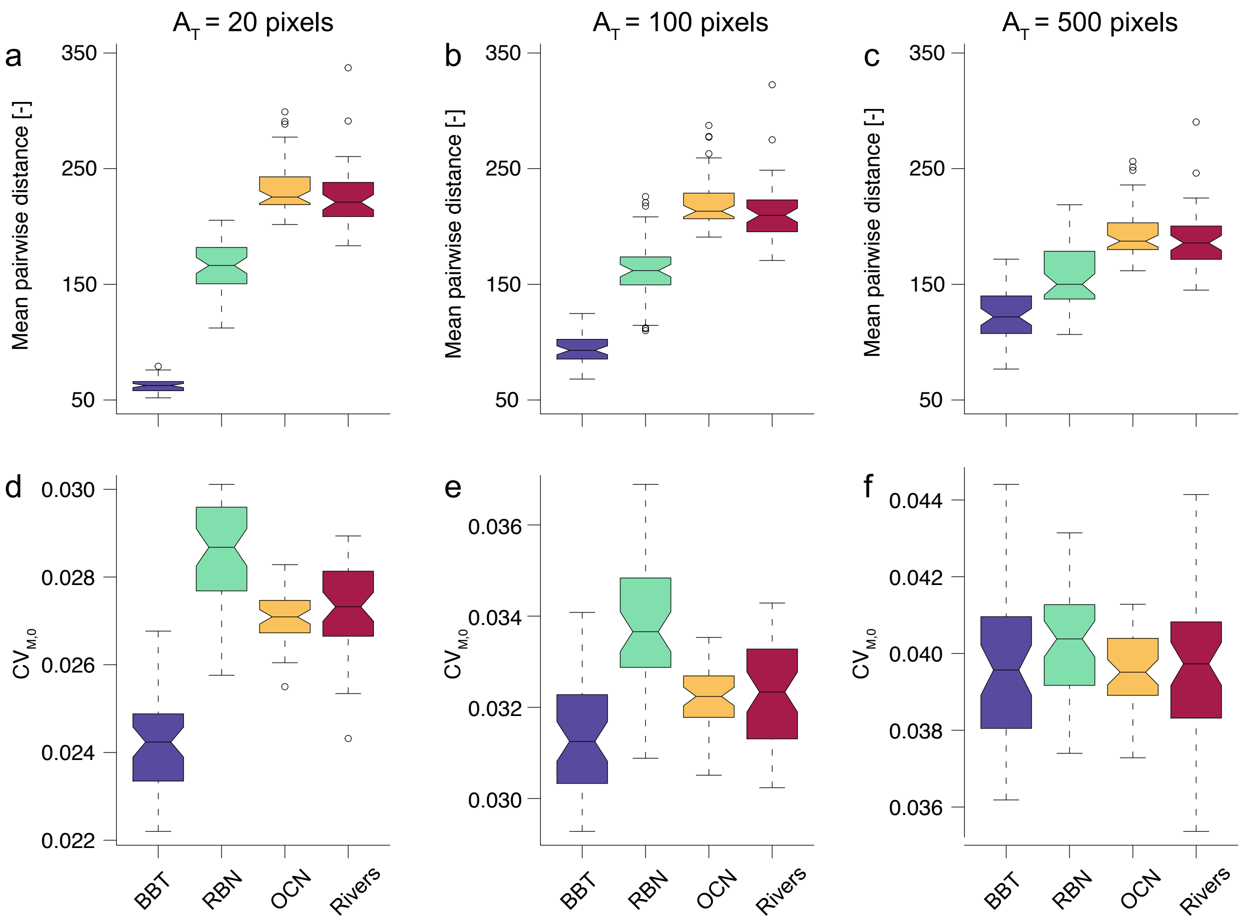
Supplementary Figure 6: Values of metapopulation metrics $CV_{M,U}$, $\lambda_{M,U}$, $CV_{M,H}$ and $\lambda_{M,H}$ obtained by setting $\alpha = 200\,l$. Plot construction as in Fig. 4.



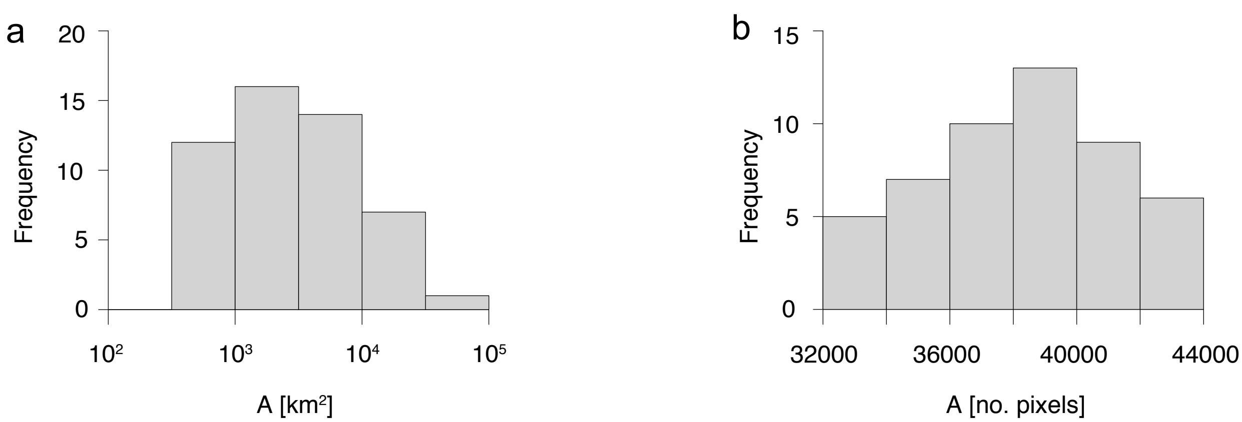
Supplementary Figure 7: Values of metapopulation metrics $CV_{M,U}$, $\lambda_{M,U}$, $CV_{M,H}$ and $\lambda_{M,H}$ obtained by setting $\alpha = 1000\,l$. Plot construction as in Fig. 4.



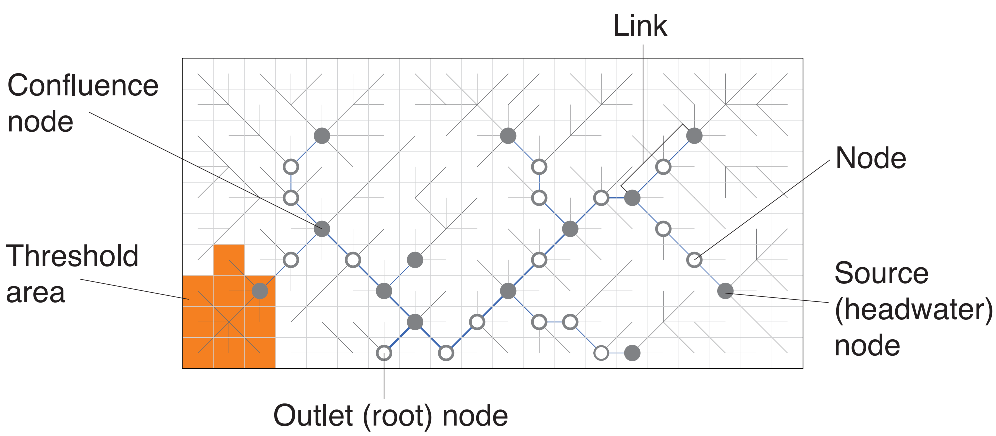
Supplementary Figure 8: Distributions of numbers of nodes N and branching ratios p_r for the analyzed networks (Fig. 4). Horizontal grey lines show values of N and p_r predicted by Eq. (1). Note that the distributions of N for OCN, RBN and BBT at a given A_T value are equal by construction.



Supplementary Figure 9: Values of mean pairwise distance (expressed in units of l) and of $CV_{M,0}$ for the networks used in the analysis. Plot construction as in Fig. 4.



Supplementary Figure 10: Distribution of drainage area values for the 50 rivers used in the analysis. a) expressed in km^2; b) expressed in number of pixels.



Supplementary Figure 11: A small OCN (same as in Fig. 1c) with illustration of the main terms used in the manuscript.

Supplementary References

[1] T. Erős and E. H. Campbell Grant. Unifying research on the fragmentation of terrestrial and aquatic habitats: patches, connectivity and the matrix in riverscapes. *Freshwater Biology*, 60(8):1487–1501, 2015.

[2] J. D. Yeakel, J. W. Moore, P. R. Guimarães, and M. A. M. de Aguiar. Synchronisation and stability in river metapopulation networks. *Ecology Letters*, 17(3):273–283, 2014. doi: 10.1111/ele.12228.

[3] S. D. Peckham and V. K. Gupta. A reformulation of Horton's laws for large river networks in terms of statistical self-similarity. *Water Resources Research*, 35(9):2763–2777, 1999.

[4] A. Maritan, A. Rinaldo, R. Rigon, A. Giacometti, and I. Rodríguez-Iturbe. Scaling laws for river networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 53(2): 1510–1515, 1996. doi: 10.1103/PhysRevE.53.1510.

[5] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[6] I. Rodríguez-Iturbe and A. Rinaldo. *Fractal River Basins. Chance and self-organization*. Cambridge University Press, New York, US, 2001.