Supporting Information

Rinaldo et al. 10.1073/pnas.1322700111

SI Text

Statics: Methods

The Basic Operational Problem. The optimal channel network (OCN) model was originally based on the ansatz that configurations occurring in nature are those that minimize a functional describing the dissipated energy and on the derivation of an explicit form for such a functional (1–4). Spanning, loopless network configurations characterized by minimum energy dissipation are obtained by selecting the configuration s that minimizes the functional:

$$H_\gamma(s) = \sum_{i=1}^{N} A_i^\gamma(s), \qquad \textbf{[S1]}$$

where i spans the lattice of N sites (A_i and γ are defined above). Given that $A_i = \sum_j W_{j,i} A_j + 1$, where $W_{j,i} = W_{j,i}(s)$ is the element of the adjacency matrix spanning the connectivity of every node j to i, the configuration s determines uniquely, on a spanning tree, the values of A_i. It is crucial, as we shall see later, that one has $\gamma < 1$ directly from the physics of the problem subsumed by the slope–area relation.

The global minimum (i.e., the ground state) of the functional in Eq. **S1** is exactly characterized by known mean field exponents (5), and one might expect to approach the mean field behavior by reaching a stable local minima upon careful annealing of the system. This is in fact the case. The proof of the above is not trivial: any stationary solution of the landscape evolution equation must locally satisfy the relationship $|\vec{\nabla} z_i| \propto A_i^{-1/2}$ between flux/area and the topographic gradient at any point i and gradients are approximated by Δz_i, the largest drop in elevation at i. One can thus uniquely associate any landscapes with an oriented (spanning and loopless) graph on a lattice. One thus can reconstruct the field of cumulative areas $\{A_i\}$ corresponding to a given oriented spanning graph. Note also that we want to emphasize the dependence on the configuration $s = \{A_1, A_2, \ldots, A_N\}$ that the system assumes on the features of the oriented spanning graph associated with the landscape topography z through its gradients ∇z that uniquely defines total contributing areas $A_i, i = 1, N$ in a N-site lattice.

Optimal arrangements of network structures and branching patterns result from the direct minimization of the functional in Eq. **S1**. The basic operational problem to obtain OCNs for a given domain is to find the connected path draining it that minimizes $H_\gamma(s)$ without postulating predefined features, e.g., the number of sources or the link lengths. One key problem is the assessment of the robustness of OCN configurations selected by any minimum procedure. This has been studied (4) with respect to the following: the strategy for minimum search; the role of initial conditions; the robustness of the functional dependence on γ; the role of lattice anisotropies; the effects of "quenched" randomness (3). The basic optimization strategies are similar to algorithms developed in the context of nonnumerable (NP-complete) problems where the exponential growth of possible configurations prevents complete enumeration. Iterated random searches work best in that context (4).

The basic algorithm proceeds as follows. An initial network configuration, s, is chosen as a spanning tree on the grid to drain an overall area made up by N sites. This defines an orientation and a connection for each pixel stating to which of the eight neighboring pixels its area is draining, neighbors being assumed at unit distance from the centroid. This in turn needed both preliminary and a posteriori speculations on whether a triangular lattice—with six neighboring nodes—or an anisotropic scheme in which diagonal connections were weighted by a $\sqrt{2}$ factor would be a better model of local interactions (4). A scalar state variable, $A_i(t)$, denotes the total area at a point i at stage t of the optimization process as follows:

$$A_i(t) = \sum_j W_{ij}(t) A_j(t) + 1,$$

where W_{ij} is the (now dynamic) functional operator that has the connectivity matrix as its static counterpart by the following:

$$W_{ij}(t) = \begin{cases} 1, & \text{if } i,j \text{ are connected} \\ 0, & \text{otherwise} \end{cases}$$

(that is $W_{ij}(t)$ implies that $j \to i$ is a drainage direction). Note that j spans the eight neighboring pixels of the arbitrary ith site. The unit mass added refers to the area representative of the actual site as a proxy of the distributed injection term. From the initial configuration (stage $t = 0$), the basic strategy consists of drawing a site at random and perturbing the system by assigning a change δW_{ij}, i.e., by modifying at random its connection to the former receiving pixel. Hence,

$$W_{ij}(t+1) = W_{ij}(t) + \delta W_{ij}.$$

This corresponds to perturbing the configuration s $(s \to s')$. Adjusting to such a local modification, all aggregated areas A_i are modified in the downstream region until the original and the modified path reconvene. The change is accepted if the modified value of $H_\gamma(s')$ is lowered by the random change $[H_\gamma(s') < H_\gamma(s)]$ and no loops are formed. Loops are excluded on a rigorous basis, as it was shown exactly that they lead to energetically unfavorable configurations [for the functional in Eq. **1** with $\gamma < 1$ every tree is a local minimum of total energy expenditure (6); *Thermodynamics of River Networks*]. As the new configuration is adopted as a base configuration, the process is iterated. Otherwise, the change is discarded [if $H_\gamma(s') \geq H_\gamma(s)$], and the t-stage configuration s is perturbed again.

The procedure leads to a configuration for which no improvement on total energy expenditure appears after a fixed (and large) number of iterations, i.e., an OCN. The whole process may or may not be then reset and restarted from the same initial configuration. This is done several times at times to allow the random process a fair chance to capture nonlocal minima—should they be of interest (see main text). The configuration attaining the lowest energy dissipation among the trials described before is chosen as the OCN.

A visual scheme of the progress of the basic algorithm is illustrated in refs. 4 (chapter 4) and 7.

This basic procedure, at times termed the Lin or the greedy approach because of the similarities with the N-city traveling salesman algorithm (4), respects the rules of a fair search for approximate solutions but is apt to yield trapping in local minimum energy. Variants of the basic algorithm, implemented to test the importance of choice of strategy for minimum search, include Lin–Kernighan's multiple simultaneous perturbations and simulated annealing schemes to avoid trapping of the configuration into unsuitable local minima. This is done by accepting perturbations of the current configuration $(s \to s')$ even if they yield $H_\gamma(s') \geq H_\gamma(s)$ with a probability depending on a state parameter T. In practice, the probability of acceptance of

the perturbation is given by the Metropolis rule, i.e., it is 1 if the resulting change corresponds to $H_\gamma(s') < H_\gamma(s)$) or, if

$$\Delta H = H_\gamma(s') - H_\gamma(s) \geq 0$$

and $e^{-\Delta H/T} > R$, and 0 otherwise ($R \in (0,1)$ is a random number $R \in (0,1)$). To carry out proper annealing, one makes changes in the parameter T from relatively high values at the start to low values toward the end of the analysis. Clearly, for high values of T, the likelihood of accepting unfavorable changes is high, whereas for $T \to 0$ the rule is equal to that of the basic algorithm. A "cooling" schedule for decreasing values of T as the procedure evolves is thus required (4).

Exact Results. We elaborate further here on the analytical results that complete our static view of dynamically accessible optimal states (4, 5, 8–11). Exact properties for the global minimum of the functional $H_\gamma(s)$ in Eq. **S1** are addressed first.

Let us consider two limit cases ($\gamma = 0$ and $\gamma = 1$). If we denote with x_i the along-stream length of the pathway connecting the ith site to the outlet, it is straightforward to show that $\sum_i A_i = \sum_i x_i$. Thus, the minimization of energy dissipation for $\gamma = 1$ corresponds to the minimization of the weighted path connecting every site to the outlet, i.e., the mean distance from the outlet. The case $\gamma = 1$ thus admits as global minimum the most direct network. The $\gamma = 0$ case, instead, implies the minimization of the total weighted length of the spanning tree where every configuration has the same energy because every spanning tree has the same number of links ($L^2 - 1$ for a $L \times L$ square lattice). For the values of $\gamma \in (0,1)$ the search for the features of the global minima proves interesting.

For $\gamma = 1$, the configurations yielding a minimum of the energy is realized on a large subclass of the set of spanning trees. These are all of the directed ones, where every link has positive projection along the diagonal, oriented toward the outlet. The $\gamma = 1$ case gives a minimum energy scaling $E \sim L^3$ (where L is, as usual, the characteristic linear size of the lattice) for each directed network. This follows from the observation that any directed network corresponds to the Scheidegger model of river networks (12), where all directed trees are equally probable by construction. Such model can be mapped into a model of mass aggregation with injection exactly solved (13–15), later shown (16) to be map exactly the time activity of the celebrated sandpile model of self-organized criticality (17). The corresponding scaling exponents, as stated in the main text, are

$$\beta = 1/3, \psi = 3/2, h = 2/3.$$

Because all directed trees are equally probable, having the same mean distance to the outlet, each stream behaves like a single random walk in the direction perpendicular to the diagonal through the outlet.

The $\gamma = 0$ case gives the same energy $E \sim L^2$ for every tree, in analogy to the problem of random 2D spanning trees, whose geometrical properties have been computed in the case of a square lattice to give the following (8):

$$\beta = 3/8, \psi = 3/5, \phi = 2, h = 5/8.$$

In the thermodynamic limit ($L \to \infty$), the global minimum in the space, $\mathcal{S}$, of all spanning trees of the functional

$$E(\gamma, \mathcal{S}') = \sum_i A_i(\mathcal{S}')^\gamma$$

scales as follows:

$$\min_{\mathcal{S}' \in \mathcal{S}} E(\gamma, \mathcal{S}') \sim \max(L^2, L^{1+2\gamma})$$

(where $\mathcal{S}'$ denotes the subset of optimal trees) for all $\gamma \in [0,1]$. Because $E(\gamma, \mathcal{S}')$ is an increasing function of γ and it is equal to L^2 for $\gamma = 0$, then for $\gamma \geq 0$ it is obvious that $E(\gamma, \mathcal{S}') \geq L^2$. The sum over all of the sites can be performed in two steps:

$$E(\gamma, \mathcal{S}') = \sum_{n=1}^{2L-1} \sum_{i \in \mathcal{D}_n} A_i(\mathcal{S}')^\gamma,$$

where $\mathcal{D}_n$ are the diagonals orthogonal to the one passing through the outlet. For directed spanning trees (DSTs), one notes that the sum of the areas in a given diagonal $\mathcal{D}_n$ is independent of the particular tree. It is thus obtained:

$$S_d(k) = \sum_{i \in \mathcal{D}_k} A_i = \begin{cases} k(k+1)/2 & k \leq L \\ L^2 - S_d(2L-1-k) & L+1 \leq k \leq 2L-1, \end{cases} \quad \textbf{[S2]}$$

where we defined $S_d(0) = 0$. Such quantities can be only increased considering generic undirected trees; thus, we can write for every spanning tree the following:

$$S(k, \mathcal{T}) = \sum_{i \in \mathcal{D}_k} A_i(\mathcal{T}) \geq S_d(k). \quad \textbf{[S3]}$$

Let us observe that for $k = 0, \ldots, (L-1)$:

$$S(k, \mathcal{T}) + S(2L-1-k, \mathcal{T}) \geq S_d(k) + S_d(2L-1-k) = L^2, \quad \textbf{[S4]}$$

making convenient to perform the summation over couples of diagonals. To get a lower bound for E, we need to assess that, for every set Γ:

$$\sum_{i \in \Gamma} A_i^\gamma \geq \left(\sum_{i \in \Gamma} A_i \right)^\gamma,$$

that follows from Schwartz inequality, being $A_i \geq 1$ and $0 \leq \gamma \leq 1$. The result is Now, using Eq. **S4**, and we can write the following:

$$\begin{aligned} E(\gamma, \mathcal{T}) &= \sum_{n=1}^{2L-1} \sum_{i \in \mathcal{D}_n} A_i(\mathcal{T})^\gamma = \sum_{n=0}^{L-1} \sum_{i \in \left(\mathcal{D}_n \cup \tilde{\mathcal{D}}_n\right)} A_i^\gamma \\ &\geq \sum_{n=0}^{L-1} \left(\sum_{i \in \left(\mathcal{D}_n \cup \tilde{\mathcal{D}}_n\right)} A_i(\mathcal{T}) \right)^\gamma \\ &= \sum_{n=0}^{L-1} [S(n, \mathcal{T}) + S(2L-1-n, \mathcal{T})]^\gamma \\ &\geq \sum_{n=0}^{L-1} L^{2\gamma} = L^{1+2\gamma}, \end{aligned}$$

where $\tilde{\mathcal{D}}_n = \mathcal{D}_{(2L-1-n)}$. Equality in the last inequality holds for directed networks. We can thus write the following:

$$E(\gamma, \mathcal{S}') \geq L^{1+2\gamma}. \quad \textbf{[S5]}$$

Eq. **S5** yields the lower bound,

$$E(\gamma, \mathcal{S}') \geq \max(L^2, L^{1+2\gamma}), \quad \textbf{[S6]}$$

holding for every tree $\mathcal{S}' \in \mathcal{S}$, and thus also for the minimum over $\mathcal{S}'$ (8).

From the above results, the relevant scaling exponents can be computed. For undirected networks, one has the following:

$$\langle A\rangle \sim L^{\varphi}, \qquad \textbf{[S7]}$$

where $\varphi \geq 1$ implies a value slightly bigger than 1 if one assumes a quasidirect behavior. Note that for a directed path one obtains instead the following:

$$\langle A\rangle \sim L.$$

Recalling that β is the scaling exponent for total contributing area, and the above result on the scaling of energy [$E = L^2\langle a^{\gamma}\rangle \sim L^{1+2\gamma}$], one gets $2\gamma - 1 = 2(\gamma - \beta)$ holding for $\gamma > \beta$ (5). Moreover, one may assume (5) that anisotropic scaling occurs in the basic planar scales, such that $A \propto L^{1+H}$, where $H \leq 1$ is a scaling coefficient that has had some use in geomorphology (see, e.g., ref. 4). The final result (8) follows from the scaling relations derived above solving for β, giving, for $\gamma \geq 1/2$,

$$\begin{cases} \beta = \dfrac{3(1-\gamma) + (\varphi - 1)(1+\gamma)}{2(1-\gamma) + \varphi - 1} - 1 \\ H = \dfrac{\varphi - \gamma}{1-\gamma} \end{cases} \qquad \textbf{[S8]}$$

(the constraint $\gamma > \beta$ become $\gamma > 1/2$, independently of the value of φ). Thus, if $H \leq 1$, for all $\gamma < 1$ must be $\varphi = 1$, yielding the following:

$$\beta = 1/2, H = 1, \psi = 2, \varphi = 1, h = 1/2,$$

valid for $\gamma \in (1/2, 1)$ (8).

The exact exponents corresponding to the ground state are the same as in mean-field theory and for the Peano basin (9, 18) referred to in the main text, and indeed significantly different from the range of scaling exponents observed in nature (10). The most striking result is that the exponent of the distribution of aggregated areas is $\beta = 1/2$, quite different from the consistent range 0.43 ± 0.02 for observational values regardless of climate, vegetation, exposed lithology, and geologic constraints (4). One could even make the prediction that river networks undergoing frequent, intense perturbations (like, e.g., in strongly seismic contexts) should perhaps approach mean-field conditions (and thus aggregation should have $\beta \to 0.50$).

Minimum Energy and Loopless Structures

We shall now define precisely the selective advantages of trees in the fluvial physics (19). Consider a square lattice. Fix an orientation for all lattice bonds. On each bond b, a flux J_b is defined (notice that, in the general context of networks, it is illegal to identify the flux from node, say, i with J_i because it is generally not unique, differently from the case of trees). We shall assume that $J_b > 0$ if it is flowing along an assigned orientation. Uniform (unit) injection is equivalent to the set of constraints $(\partial J)_x = 1$, where ∂ is a discrete version of the divergence, and is a measure of the net outflow from a site:

$$(\partial J)_x = \sum_{b \in x} J_b\, \theta(b,x) = 1, \qquad \textbf{[S9]}$$

where the unit value is the model injection, constant for every node in the simplest case; b spans all bonds (links) concurring on node x, and $\theta(b,x) = 1$ (-1) if b is oriented outward (inward) node x. Any local minimum of the following function:

$$E = \sum_b |J_b|^{\gamma} \qquad \textbf{[S10]}$$

when $0 < \gamma < 1$, corresponds to $J_b \neq 0$ only on the bonds of a spanning tree. The main point (19) is in the proof that the networks that correspond to local minima of the dissipated energy are loopless and tree-like. The tree must be spanning due to the constraints [**S9**]: one cannot have $J_b = 0$ for all b's connected to a site so that there must be at least one outlet from each site x. Some site (or sites) must also be declared to be the global outlet. We shall show that loopless structures emerge as optimal solutions of Eq. **S10** with the constraint [**S9**], which is precisely the case for river networks.

Fig. 7 of the main text illustrates an extremely simple example with just four sites: the left panel shows the setup for the elementary four-bond network. The dot is the outlet. Here, the current a is taken as the parameter regulating the entire distribution of fluxes owing to continuity. The center panel illustrates the only loopless configurations of the system generated by integer values of a. The right panel shows the plot of the function E vs. a from the following Eq. **S11** with $\gamma = 0.5$:

$$E = |a|^{\gamma} + |a+1|^{\gamma} + |1-a|^{\gamma} + |2-a|^{\gamma}, \qquad \textbf{[S11]}$$

which is derived from Eq. **S10** after implementation of [**S9**]. In particular, the center panel shows the plot $E(a)$, where one notices that there are local minima in correspondence with one of the four currents being zero (a = 2, 1, 0, −1), corresponding to the four trees shown in Fig. 7, *Right*, of the main text. The explanation is simple. Suppose that $a \sim 0$ (the other cases are equivalent). All of the terms in [**S11**] but $|a|^{\gamma}$ can be expanded in Taylor series around $a = 0$. Thus, locally one has the following:

$$E = 2 + 2^{\gamma} + |a|^{\gamma} + \mathcal{O}(a), \qquad \textbf{[S12]}$$

which has a cusp-like behavior because $0 < \gamma < 1$. Notice that $\partial E/\partial a|_{a=0^{\pm}} = \pm\infty$ and thus one cannot find the minima simply by imposing the condition $\partial E/\partial a = 0$. If $a \neq 0, \ \pm 1, 2$, $\partial^2 E/\partial a^2 < 0$ and there are no other minima of E (only maxima). The proof for the general case dealing with arbitrary lattices is elsewhere (19).

Fig. 7, *Right*, of the main text also shows the function E vs. a plotted for various values of γ (specifically, for $\gamma = 0.25, 0.5, 1$, and 2). It is deemed quite instructive, also in view of the unexpected twists in research mentioned in the introduction to the main text, as the original assumption of accepting only tree-like configurations to search for minima of Eq. **S1** turned out to be energetically favorable only when $\gamma < 1$—a condition clearly requested by the physics of the problem (Eq. **2**) and confirmed by real-life data. In that sense, we may call it a fluke. Also, note that for $\gamma = 1$ all directed configurations, loopless or not (with the currents going in the positive directions), have the same energy.

The case $\gamma = 2$ corresponding to a linear Ohm-like current-energy drop is akin to resistor networks for which there is just one minimum for a network-like (not a tree) configurations obtained at $a = 1/2$ [see for an account of the relevant literature (19)]. Because there is one unknown current for each bond, and one continuity equation for each site the number of independent variables is given by the number of bonds minus the number of sites (excluding the outlet), which for the simple topologies considered is equal to the number of elementary loops (this is a particular case of the Euler theorem) (see also ref. 4, chapter 4).

Because we have shown that local minima occur in singular configurations where some currents are zero, we cannot introduce the standard technique of Lagrange multipliers to find the minima of E with the constraint [**S9**]. To be able to do that, E must be regularized as

$$E = \sum_b \left(J_b^2 + \varepsilon^2\right)^{\gamma/2}$$

in the limit $\varepsilon \to 0$. The general proof is beyond the scope of this paper and given elsewhere for an arbitrary graph where the number l of independent loops is given by the number of bonds minus the number of sites plus the number of connected components (19). Note that, for the particular case where the graph must be a spanning structure, the number of connected components is unit (for example, in the case of an $n \times m$ rectangular lattice, one has $l = nm - n - m + 1$).

Obviously, for the set of dynamical rules postulated above, the energy landscape is riddled with a large number of local minima characterized by a range of similar values of E. In single realizations, boundary and initial conditions affect the feasible (i.e., dynamically accessible) optimal state to different degrees depending on their constraining power. This fact matches the observation that scaling exponents are coherently linked in a range of values (5, 10), narrow enough but significantly different from the ground state (8).

The truly important implications are twofold: on one side, in fact, all local optima are trees; on the other, imperfect optimal search procedures are capable of obtaining suboptimal networks, which nevertheless prove statistically indistinguishable from the forms observed in nature and quite different from the absolute minima (4, 10). Indeed, we believe that the worse energetic performance and yet the better representation of the patterns of nature mimic general evolutionary and selection processes—indeed, landscapes do not have a blueprint for evolution, but fall on a selective advantage out of stability, however myopic and aimless the evolutionary tinkering .

Thermodynamics of River Networks

Central to models postulating chance-dominated network selection is the assumption of equal likelihood of any possible tree-like configuration. However, the foundations of OCNs discussed above postulate that some spanning, loopless network configurations are more likely than others. Indeed, their overall likelihood is controlled by the minima of the functional Eq. **S1** defining total energy expenditure of the network structure both as a whole and in its parts. The set of possible configurations for the system is constituted by the ensemble of all rooted trees spanning a given lattice of sites defined by a complete set of oriented links among connected neighbor sites. In this section, we investigate, following ref. 20, the thermodynamic rationale behind the scaling properties of the energy and entropy of OCNs.

We start by assigning a probability $P(s)$ to each particular spanning tree configuration s. As usual, we take $P(s)$ to be a Boltzmann distribution as follows:

$$P(s) \propto e^{-H(s)/\mathcal{T}}, \qquad \textbf{[S13]}$$

where $\mathcal{T}$ is a suitable parameter resembling Gibbs' temperature of ordinary thermodynamic systems and the functional $H(s)$ is the Hamiltonian of the system, i.e., a global property related to energetic characters.

Network models where all spanning trees are equally likely is the limit case of the model described by Eq. **S13** for $\mathcal{T} \to \infty$. OCNs belong to the class of configurations described by Eq. **S13** where the Hamiltonian $H(s)$ reduces to $H(s) = H_\gamma(s) = \sum_i A_i^\gamma$ (where i spans the L^2 sites occupied by a $L \times L$ square lattice and with usual notation), the total energy dissipation of the spanning tree configuration. In fact, OCNs are obtained by selecting the spanning network configurations, s, which maximize the probability in Eq. **S13** by minimizing the Hamiltonian. OCNs thus represent the other extreme of random networks, because they constitute the maximum probability case for $\mathcal{T} \to 0$. Ordinary thermodynamic settings characterized by a finite value of $\mathcal{T}$ are thus of interest. We will show, following ref. 20, that OCN concepts work at any finite temperature where energy minimization always maximize the probability of a configuration provided that the network is large enough (the so-called thermodynamic limit).

We can distinguish different behaviors depending on the value of γ.

For $\gamma < 0$, one obtains Hamiltonian paths, i.e., spiral-like patterns with tendency to penalize aggregation. For $\gamma = 1$, one minimizes $H_1(s)$, a measure of the directedness of the channels constituting the channel tree—proportional to the mean distance to the outlet, $<\ell>$, of all sites of the lattice, such distance being measured along the drainage directions. For a spanning tree s, one may write $<\ell> \propto \sum_{i \in s} \sum_{j \in x(i)} \Delta x_{ij}$, where $x(i)$ is the path along the network from i to the outlet and Δx_{ij} is the spatial step or the lattice size, taken as unit in unbiased lattices. Rearranging the sum (Δx_{ij} is added A_i times for each site i, i.e., as many times as the contributing sites upstream of i) yields $<\ell> \propto \sum_{j=1}^{N} A_j$ for the N sites and $H_1(s) = \sum_{j \in s} A_j$ for the configuration s, where A_j is the total contributing area at the arbitrary site j within the tree s. For $\gamma > 1$, the patterns maximizing Eq. **S13** are such that the average length of the path from each site to the outlet is the shortest. The range $0 < \gamma < 1$ is more interesting and in this region the system exhibits rich scaling structures and aggregation patterns. From the physics of river erosion, to first order in the small gradient approximation (see *Dynamics* in the main text), one must minimize $H_{1/2}(s)$. Therein, for a given drainage basin overlain with a lattice of L^2 sites, let $\mathcal{S}$ be the set of spanning loopless trees rooted in a given point, say 0. For any configurations $s \in \mathcal{S}$, we define the probability of the tree s as in Eq. **S13**, i.e.,

$$P(s) \propto e^{-H_\gamma(s)/\mathcal{T}},$$

where, again, $\mathcal{T}^{-1}$ is Gibbs' parameter mimicking the inverse of temperature of classic thermodynamic systems. For a fixed γ, let $H_\gamma(\mathcal{S})$ denote the finite set of all possible values that may be taken on by $H_\gamma(s)$ for trees $s \in \mathcal{S}$. Given an energy level, E, let $N(E)$ be the degeneracy, or the number of different spanning trees s for which $H_\gamma(s) = E$. One has the following:

$$P\left(H_\gamma(\mathrm{s}) = E\right) = \sum_{s: H_\gamma(s) = E} P(s) \propto N(E) e^{-E/\mathcal{T}}.$$

Defining formally the thermodynamic entropy as $\sigma(E) = \log N(E)$, one obtains the following:

$$P\left(H_\gamma(s) = E\right) \propto e^{-F(E)/\mathcal{T}}, \qquad \textbf{[S14]}$$

where a free energy $F(E) = E - \mathcal{T}\sigma(E)$ has been introduced. Indeed, the most probable states correspond to an energy level E that minimizes $F(E)$.

Our definition of Hamiltonian is appropriate because the governing functional has the meaning of total energy dissipation. The central result (20) is that entropy, σ, scales subdominantly with the system size, L, compared with the energy term such that even for a nonzero value of Gibbs' parameter. Thus, the most probable spanning tree configurations determined by minimizing the free energy can be equally well obtained by minimizing total energy dissipation provided that $L \to \infty$—in practice, the lattice should simply be large enough (20). As a result, in the thermodynamic limit the system described by the probability [**S13**] always tends to operate like at zero temperature, i.e., when total energy dissipation is minimized. The proof is sketched in what follows. One exact result is that for the set $s \in \mathcal{S}_E$ of OCNs, one has $E = \min_s H_\gamma(s) \propto L^{2+\delta}$ with $\delta > 0$ for $\gamma \geq 1/2$. The easiest case to prove corresponds to $\gamma = 1$. For a given spanning tree s, we have seen (*Exact and Computational OCN Analyses* in main text)

that $H_1(s) = L^2 < \ell >$ where $< \ell >$ is the average distance of the L^2 sites from the outlet measured along the links belonging to s. Thus, the minimum of $H_1(s)$ is attained by the set $\mathcal{S}_D$ of all DSTs, i.e., the trees whose links have positive projection on the diagonal oriented in the outlet direction. DSTs have $< \ell > \propto L$, i.e., $H_1(\mathcal{S}_D) = \mathrm{Min}_s\ H_1(s) \propto L^3$. Using this result and Schwarz' inequality, it has been shown (20) that for $\gamma \in (1/2, 1]$ and any $s \in \mathcal{S}$, one has the following:

$$H_\gamma(s) \geq \text{const}\ L^{1+2\gamma},$$

where the constant depends on lattice properties. Thus, the above inequality implies that

$$\min H_\gamma(s) > \text{const}\ L^{2+\delta(\gamma)}$$

with $\delta > 0$ for $\gamma > 1/2$. For $\gamma = 1/2$, the global minimum has a different lower bound, i.e.,

$$E \propto L^2 \log L,$$

which is characterized by a logarithmic correction (20). This can be proven by explicitly constructing classes of structures that scale with the value of δ obtained by coarse graining the elevation field associated with the OCN structure (see main text).

Numerical and theoretical results for the $\gamma = 1/2$ case suggest that

$$\mathrm{Min}\, H_{1/2}(\mathcal{S}_E) \propto L^{2+\delta(1/2)},$$

where $\mathcal{S}_E$ denotes the set obtained by averaging over accessible local minima, with $\delta(1/2) = 0.1 - 0.2$. This result is consistent with the findings shown in Fig. 6 of the main text. Recall that the probability of exceedence $P[A \geq a]$ has a probability distribution $p(a, L) = a^{1+\beta} F(a/L^\phi)$, where $\beta = 0.43 \pm 0.02$ and the finite-size effect is defined by a coefficient $\phi = 1 + H \sim 2$. Then one has the following:

$$\min_s\ H_{1/2} \propto L^{2+\delta(1/2)},$$

with $\delta(1/2) = 1 - 2\beta \approx 0.13 > 0$ with $\beta = 0.43$ [recall that $L^2 < A^{1/2} > = \int_1^\infty dx\, x^{1/2-\beta} F(x/L^2)$ where F is the finite-size scaling function as in Fig. 1 of the main text]. Note that a scaling of energy with size in OCNs as $E \propto L^{2.2}$ has also been experimentally observed in multiple outlet OCNs (4) and is confirmed by the results of Fig. 6. The result that $\delta > 0$ can be proved in general (4, 20).

For spanning loopless trees, the number $N(E)$ of configurations s with given energy E scales as follows (20):

$$N(E) \propto \mu^{L^2}$$

such that

$$\sigma(E) \propto L^2,$$

where μ is a real number depending on lattice properties. For example, in a four-neighbor lattice the number N_D of DST is $N_D = 2^{(L-1)^2}$. The total number N of spanning trees is greater than N_D but less then the number of possible ways of choosing $L^2 - 1$ links (number of links of a spanning tree) among all of the $2L(L-1)$ possible links. Thus,

$$2^{(L-1)^2} = N_D < N < \binom{2L(L-1)}{L^2 - 1} \sim 2^{2L^2}.$$

Because the number $N(E)$ of configurations with a given energy is smaller than N, then

$$N(E) < 2^{2L^2},$$

and, in general, the above equation is satisfied with $2 < \mu < 4$ in the four-neighbor case. The general case follows directly from the same type of reasoning (20).

We thus conclude that, for OCNs generated minimizing Eq. **S1** with $\gamma \geq 1/2$, entropy scales subdominantly to the energy with system size. Thus, in the limit $L \to \infty$,

$$\min F(E) \propto L^{2+\delta}$$

because $\delta > 0$ [in fact, one has rigorously

$$F(E) \propto L^{2+\delta} - \text{const}\ TL^2$$

and, for $L \to \infty$, $F(E) \propto L^{2+\delta}$ when $\delta > 0$]. Hence, the configuration s that minimizes H_γ also minimizes $F(E)$ whatever the value Gibbs' parameter $\mathcal{T}$, provided that the system is large enough. Hence OCNs, which correspond to the $\mathcal{T} = 0$ limit case [i.e., the configuration yielding min $F(E)$ is that endowed with min$E$ only for $\mathcal{T} \to 0$], reproduce natural conditions for any "temperature." Because fluvial networks usually develop migration of divides and competition for drainage in the absence of geologic controls over domains large with respect to the scale of channel initiation, it is likely that natural networks evolve under conditions that well approximate the thermodynamic limit $L \to \infty$. We suggest that this is the reason for the outstanding ability of OCNs to reproduce observational evidence.

What are the implications of the above results? Every tree is a local minimum of total energy dissipation in the fluvial landscape, and imperfect optimal search procedures yield suboptimal networks, which nevertheless prove statistically indistinguishable from the forms observed in nature (and quite different from their ground states). Thus, OCNs have by construction a much constrained structure (loopless and spanning) and entail the minimization of total energy dissipation. These structures develop under generic conditions and do not exhibit a set of dynamically recursive states but freeze into static scale-free structure, i.e., they behave like a $\mathcal{T} = 0$, "frozen" system by the subdominant scaling of entropy with system size compared with the energy of the system. Indeed, we believe that the worse energetic performance and yet the better representation of fluvial patterns in nature is a consequence of imperfect selection mimicking the myopic tinkering of evolutionary processes.

1. Rodriguez-Iturbe I, Rinaldo A, Rigon R, Bras RL, Ijjasz-Vasquez E (1992) Energy dissipation, runoff production and the three dimensional structure of channel networks. *Water Resour Res* 28(4):1095–1103.
2. Rodriguez-Iturbe I, Rinaldo A, Rigon R, Bras RL, Ijjasz-Vasquez E (1992) Fractal structures as least energy patterns: The case of river networks. *Geophys Res Lett* 19(9): 889–892.
3. Rinaldo A, et al. (1992) Minimum energy and fractal structures of drainage networks. *Water Resour Res* 28(9):2183–2195.
4. Rodriguez-Iturbe I, Rinaldo A (1997) *Fractal River Basins: Chance and Self-Organization* (Cambridge Univ Press, New York).
5. Maritan A, Rinaldo A, Rigon R, Giacometti A, Rodríguez-Iturbe I (1996) Scaling laws for river networks. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 53(2):1510–1515.
6. Rinaldo A, Banavar JR, Maritan A (2006) Trees, networks and hydrology. *Water Resour Res* 42(6):W06D07.
7. Briggs LA, Krishnamoorthy M (2013) Exploring network scaling through variations on optimal channel networks. *Proc Natl Acad Sci USA* 110(48):19295–19300.
8. Maritan A, Colaiori F, Flammini A, Cieplak M, Banavar JR (1996) Disorder, river patterns and universality. *Science* 272:984–988.
9. Colaiori F, Flammini A, Maritan A, Banavar JR (1997) Analytical solutions for optimal channel networks. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 55:1298–1302.

10. Rinaldo A, Rigon R, Rodriguez-Iturbe I (1999) Channel networks. *Annu Rev Earth Planet Sci* 26:289–306.
11. Banavar JR, Colaiori F, Flammini A, Maritan A, Rinaldo A (2001) Scaling, optimality and landscape evolution. *J Stat Phys* 104:1–33.
12. Scheidegger A (1967) A stochastic model for drainage patterns into an intramontane trench. *Bull Ass Sci Hydrol* 12:15–20.
13. Takayasu H, Nishikawa I, Tasaki H (1988) Power-law mass distribution of aggregation with injection. *Phys Rev A* 37(8):3110–3117.
14. Takayasu H, Takayasu M, Provata A, Huber G (1991) Statistical properties of aggregation with injection. *J Stat Phys* 65(3-4):725–745.
15. Huber A (1991) Scheidegger's rivers Takayasu's aggregates and continued fractions. *Physica A* 170:463–469.
16. Dhar D (1997) The Abelian sandpile and related models. *Physica A* 263:4–28.
17. Bak P, Tang C, Wiesenfeld K (1988) Self-organized criticality. *Phys Rev A* 38(1):364–374.
18. Marani A, Rigon R, Rinaldo A (1991) A note on fractal channel networks. *Water Resour Res* 27:3041–3049.
19. Banavar JR, Colaiori F, Flammini A, Maritan A, Rinaldo A (2000) Topology of the fittest transportation network. *Phys Rev Lett* 84(20):4745–4748.
20. Rinaldo A, et al. (1996) Thermodynamics of fractal networks. *Phys Rev Lett* 76(18): 3364–3367.