Globally Universal Fractal Pattern of Human Settlements in River Networks

Yu Fang[1], Serena Ceola[2], Kyungrock Paik[3], Gavan McGrath[4], P. Suresh C. Rao[5], Alberto Montanari[2], James W Jawitz[1]

[1]Soil and Water Sciences Department, University of Florida, FL, USA.

[2]Department of Civil, Chemical, Environmental and Materials Engineering, University of Bologna, Bologna, Italy.

[3]School of Civil, Environmental, and Architectural Engineering, Korea University, Seoul, South Korea.

[4]Ishka Solutions, Australia.

[5]Lyles School of Civil Engineering, Purdue University, IN, USA.

Corresponding author: James W Jawitz (jawitz@ufl.edu)

Key Points:

- Preferential downstream clustering of human settlements is consistent across continents.
- The outlets of large river basins are preferred for human settlements compared to other coastal areas.
- Hortonian scaling and power-law scaling of power spectra indicate universal fractal structure of human settlements.
- Future implications include growing human pressures on downstream water quality and quality, coupled with exacerbated flood risks.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2017EF000746

Abstract

River networks play a key role in the spatial organization of human settlements. Both river networks and human settlements have been found to exhibit regular self-similar patterns, but little is known about the generalized spatial patterns of human settlements embedded within river networks. Here, based on night light data, we find a universal fractal structure at the global scale, with both robust Hortonian scaling relationships with the extent of human settlements and statistically significant power-law scaling of the power spectra of human area functions. Globally, we find consistent patterns of power-law preferential downstream clustering of human settlements across all six populated continents, typically up to 40% of the maximum flow length. This downstream clustering suggests an optimum distribution of humans in large river basins for trade, transport and natural resource utilization, but with attendant implications for human impacts on rivers. Recognition of such spatial patterns helps generalize assessments of human impacts on rivers, with direct implications for management of water quality and biological diversity in river networks.

1. Introduction

The spatial pattern of human settlements is influenced by landscape heterogeneity and natural resource availability, with river networks playing a central role in enabling access and mobility. Historically, humans have followed river networks during migrations in the early phases of settlement (Bertuzzo et al., 2007; Campos et al., 2006) and have preferentially built settlements close to rivers, for the purpose of navigation, water supply, and trade (Ceola et al., 2015; Kummu et al., 2011). Over time, these established settlements gradually developed into towns and cities that attracted more people, and even grew into megacities with unprecedented population size with accelerating urbanization (Grimm et al., 2008a). Quantitative regularities from empirical studies, including city size distribution (Berry, 1961; Decker et al., 2007; Krugman, 1996) and universal urban scaling theory (Bettencourt, 2013; Bettencourt et al., 2007; Bettencourt & West, 2010), have improved our understanding of urban growth dynamics and social organization, and also the environmental impact of urbanization (Fragkias et al., 2013). However, how such striking regularities are reflected in spatial patterns embedded within river networks remains elusive.

River networks are characterized by bifurcating and hierarchical geometries with universally consistent scale-free topological features, resulting from self-organization driven by similar generating processes (Dodds & Rothman, 2000). One descriptor of such organizational structure is stream order, which describes the relative size of a stream in a tree-like river network (Horton, 1945). Generalized Horton's laws refer to scaling relationships of topologic and geometric variables (e.g., stream number, basin area, and stream length) with stream order (Horton, 1945; Peckham & Gupta, 1999; Rodríguez-Iturbe & Rinaldo, 2001). The Hortonian ordering framework can provide a basis to link the geomorphological structure of river networks with ecological and anthropogenic processes. Hortonian ordering and scaling frameworks have been applied to explain water quality (Kang et al., 2008) and ecological diversity, including fish (Beecher et al., 1988; Platts, 1979), diatom communities (Stenger-Kovacs et al., 2014), benthic macro-invertebrate communities (Crunkilton & Duchrow, 1991), and riparian vegetation (Dunn et al., 2011). In addition to stream order, another descriptor of the structure of river networks within river basins is the geomorphological area function (A_G) defined as river basin drainage area as a function of distance along the flow path (Marani et al., 1994; Moussa, 2008). The fractal morphology of

river networks can be indicated by the observed scaling relationship between the Fourier transform-based power spectrum of A_G and spatial frequency f (Marani et al., 1994).

In this work, we evaluate the general questions of whether human settlement patterns within river basins are structured according to stream order and river basin drainage area. Hortonian scaling relationships have been reported for human populations (Miyamoto et al., 2011). However, a series of important questions are raised by a purported link between human population and stream order. 1) Is there universality of such scaling beyond specific landscape characteristics, socioeconomic development history, and sociopolitical governance frameworks of individual study regions? 2) Is the fractal structure of river networks found from geomorphological area functions transferable to indicate scaling relationship for humans? 3) Are the scaling relationships between humans and stream network topology sufficient to infer spatial patterns within river basins? 4) What is the underlying mechanism for the spatial patterns of human settlements in river networks?

This study assessed these questions at a global scale over a broad range of climatic zones with large variation of vegetation types, topography, and natural resource endowments. We introduced the human area function (A_H), analogous to A_G, to indicate the size of human settlements along the hydrological flow path to the basin outlet, and evaluated the power spectra relationship with frequency for A_H to assess the existence of scaling of human settlements. We compared the power spectra relationships for A_H and A_G to evaluate whether human settlement patterns within river basins are structured according to river basin drainage area. We employed several metrics including stream order, area functions, and the spectral power of area functions to assess human settlement spatial patterns within river networks across basins and continents. We further examined the trends of factors that influence human settlements to infer what the possible reasons of human settlement patterns might be and investigate whether clustering of human settlements in coastal areas (Small & Nicholls, 2003; Vitousek et al., 1997) is related to river basin outlets.

Our study tested for commonality in the linkages between river geomorphologic structure and human settlement patterns despite the broad diversity of catchment features. Human activities have caused substantial alterations to the Earth, affecting ecosystem patterns and processes, altering global hydrological and biogeochemical cycles, amplifying resource exploitation and environmental deterioration, and contributing to climate change (Vitousek et al., 1997). With increasing urbanization, the environmental impacts of these human activities are intensively concentrated within small areas, raising significant concerns about sustainability (Grimm et al., 2008b). River networks can connect, and facilitate the transport of pollutants between, human settlements and terrestrial and marine environments (Schmidt et al., 2017). Investigating the spatial distribution of human settlements within river basins improves our generalized understanding of human-nature co-evolution, but also has direct implications for understanding human impacts on rivers and downstream ecosystems.

2. Methods

2.1. Data sources

Nightlight data, collected by the US Air Force Weather Agency under the Defense Meteorological Satellite Program Operational Linescan System (National Oceanic and Atmospheric Administration, 2013) (https://ngdc.noaa.gov/), represent cloud-free nocturnal luminosity from sites with protracted lighting (i.e., cities, towns, gas flares). Nightlights are available as raster products at a resolution of 30 arc seconds, corresponding to nearly 1 km at the equator, and nightlight values are expressed as an adimensional digital number (DN) value, ranging from 0 to 63, corresponding to conditions characterized by absence of lights

through pronounced luminosity, which is interpreted as proportional to the presence of human settlements. Nightlights have thus been widely employed for demographic, economic, and environmental purposes (Ceola et al., 2014). Here, based on 2013 global nightlight data, we excluded sunlit and moonlit data, observations from ephemeral phenomena like fires, and all data associated to gas flares from the data set, to proxy human settlements, with the total DN values indicating the size of human settlements.

HydroSHEDS and Hydro1K data are two major global river network databases, with the former (Lehner et al., 2008) (http://hydrosheds.org/) primarily derived from elevation data of NASA's Shuttle Radar Topography Mission and the latter (US Geological Survey, 2000) (https://lta.cr.usgs.gov/HYDRO1K) from the USGS 30 arc-second digital elevation model of the world (GTOPO30). The threshold areas adopted to delineate river networks are 20 km^2 and 1000 km^2 for HydroSHEDS and Hydro1K, respectively (Stein et al., 2014). To extract stream-order-based basin and sub-basin boundaries for Hortonian analysis, we applied HydroSHEDS river networks vector and drainage direction raster together to utilize its higher accuracy. For area function analyses, we used Hydro1K flow direction raster to take advantage of its feasibility to derive hydrological distance. We also employed the GEODATA 9 arc-second DEM (DEM-9s) Version 3 (Hutchinson et al., 2008) from Geoscience Australia (http://www.ga.gov.au/) as a supplement to Hydro1K data since the coverage of Australia is lacking from Hydro1K. We derived a total of 2988 river basins with stream order ≥ 1 from Hydro1K data, and added 148 Australian basins delineated using GEODATA DEM-9s data based on the same contributing area threshold as Hydro1K, 1000 km^2. Thus, in total we conducted area function analyses on 3136 river basins. The numbers of river basins delineated from different databases are shown in Table S1. Differences in basin boundaries when extracted from the two primary databases are illustrated for the St. Lawrence and Saskatchewan-Nelson basins (Figure S1). All data we used were at (for HydroSHEDS and Hydro1K data) or resized to (for GEODATA DEM-9s) 30 arc seconds, the same resolution as the nightlight data.

2.2. Analyses

Horton's laws characterize the morphology of river networks, and the law of stream contributing areas is expressed (Rodríguez-Iturbe & Rinaldo, 2001),

$$\bar{A}_w = \bar{A}_1 R_A^{w-1} \quad (1)$$

where w indicates stream order, $\bar{A}_w$ is the mean area contributing to streams of order w, and the dimensionless area ratio is $R_A = \bar{A}_w/\bar{A}_{w-1}$. We hypothesize that human settlement size (H), using nightlights as a proxy, also show similar relationship with w as follows:

$$\bar{H}_w = \bar{H}_1 R_H^{w-1} \quad (2)$$

where $\bar{H}_w$ is the mean of the sum of nightlight DN values for each order w in each basin, and R_H is defined here as human settlement ratio ($R_H = \bar{H}_w/\bar{H}_{w-1}$). A minimum of stream order 3 is required to conduct Hortonian analysis, and here we analyzed 2705 river basins globally with $w \geq 4$, derived from HydroSHEDS data. We computed total areas and total nightlight DN values for all the embedded sub-basins. Since the logarithms of Equation (1) and (2) lead to equations linear in stream order, w, we linearly regressed the log mean areas and log mean nightlights (mean human presence) against w ($p < 0.05$) and calculated R_A and R_H as the antilog value of the slope for each river basin. We tested the significant difference between R_A and R_H by comparing the 95% significance intervals of the two slopes using R and applied the ratio $\theta = R_H/R_A$ to directly evaluate the spatial structure of human settlements within river basins.

The geomorphological area function, A_G, relates the contributing area to the distance to the basin outlet following the hydrological path as (Rodríguez-Iturbe & Rinaldo, 2001):

$$\int_0^{D_{max}} A_G(x)\, dx = A \quad (3)$$

where A is the total area of the river basin, x is the hydrological distance, and D_{max} is the longest distance to the outlet ($0 \le x \le D_{max}$). Similarly, we defined A_H as the total nightlight DN values at a distance x from the basin outlet to represent the structural characteristics of human settlements in river basins:

$$\int_0^{D_{max}} A_H(x)\, dx = H \quad (4)$$

where H is the total nightlight DN values in the river basin. We computed A_G and A_H for 3136 global river basins, using hydrological distance calculated based on the GEODATA (for Australia) and Hydro1K (for the rest of world) databases. Hydrological distance was computed for each cell in each basin as the distance from the cell to the closest river segment plus the distance from the river segment to the final river basin outlet (46 km $\le D_{max} \le$ 6488 km). Note that hydrological distance was not supported by the HydroSHEDS database used above for the Hortonian analyses.

To better compare geomorphological and human area functions among river basins, we normalized all variables in equations (3) and (4) by considering normalized distance, $\hat{x} = x/D_{max}$, normalized geomorphological area, $\hat{A}_G = A_G/A$, and normalized human area, $\hat{A}_H = A_H/H$. The ratio, $\rho_H = \hat{A}_H/\hat{A}_G$, represents the normalized human settlement density. The normalized values, $\hat{x}$, $\hat{A}_G$, and $\hat{A}_H$, range between 0 and 1, and the integrals (normalized forms of equations (3) and (4)) equal unity:

$$\int_0^1 \hat{A}_G(\hat{x}) d\hat{x} = 1 \quad (5)$$

$$\int_0^1 \hat{A}_H(\hat{x}) d\hat{x} = 1 \quad (6)$$

Spectral analysis can measure the strength of periodic components of a signal at different frequencies and was conducted here using the "spei" function in R through Fourier transform of the two area functions, A_G and A_H, from spatial dimensions into power spectrum as functions of frequency. The spectra, S_G and S_H, are frequency-domain representation of A_G and A_H and describe their variance structure, and the frequency, f (km^{-1}), indicates how often the signal occurs per unit of distance. The power spectrum, S_G, is expected to follow a power-law scaling relationship with the spatial frequency, f(Marani et al., 1994):

$$S_G = \alpha f^{-\beta_G} \quad (7)$$

Here, we also tested whether a similar scaling relationship exists between S_H and f. The values of the associated spectral slopes, β_G and β_H, indicate how fast the variance of A_G and A_H change across hydrological distance, and we thus compared the range of the two spectral slopes (β_G and β_H) to explore the spatial pattern of human settlements. Again, the difference between β_G and β_H was tested through comparison of the 95% significance intervals of the two slopes using R. In order to provide sufficient flow path distance for spectral analysis, our analysis was restricted to $w \ge 3$, which resulted in a global set of 563 river basins. Regressions were restricted to frequencies > 10^{-3} km^{-1} to minimize artifacts associated with approaching the finite basin boundaries, and only river basins with $R^2 > 0.5$ were included for analyses.

Continental-average patterns in average human settlement density, $\overline{\boldsymbol{\rho}}_{\boldsymbol{H}}$, were computed across all river basins on each continent at 40 equal-length intervals along $\hat{\boldsymbol{x}}$. The

number of river basins in each continent is shown in Table S1. Hydrological distance was not available for smaller coastal basins (SCBs) with area < 1000 km^2. Therefore to compare human settlement density with distance in coastal areas, we used Euclidean distance to the coast, *d*, for both large river basins (LRBs) and SCBs for d < 50 km. This threshold was selected because the area of SCBs diminishes to near zero, while the area of LRBs increases to maximum when *d* approaches 50 km from the coast for all six continents (Figure S2). Additionally, we examined the average trends of slope and upstream contributing areas for each continent in order to provide insight to the drivers of the human settlement patterns.

3. Hypothesized human settlement archetypes

Considering the regularities found both in the structure of river networks (Horton's laws and geomorphological area function) and city organization (city size distribution and scaling theory), we postulate the existence of a direct relationship between the spatial patterns of human population distributions within river basins and the geomorphological structure of river networks. We hypothesize a structured downstream clustering of human settlements along river networks, and employ several metrics including stream order, area functions, and the spectral power of area functions to assess spatial patterns. We illustrate in Figure 1 archetypal patterns of unstructured and structured clustering along the river network, and how these would manifest for the applied metrics. While it is well known that human populations are not uniformly distributed (Small & Cohen, 2004; Fang & Jawitz, 2018), we also illustrate the homogeneous case as a reference.

We hypothesize log-linear Hortonian relationships between average human presence and stream order (Figure 1a, red lines), similar to the underlying basin area distribution (blue lines) (Rodríguez-Iturbe & Rinaldo, 2001). The Horton ratios, R_H and R_A, are calculated from the slopes of these log-linear plots. The ratio $\theta = R_H/R_A$ indicates the attractiveness for human settlement by basin order: $\theta \sim 1$ indicates neutral attractiveness (i and ii), and $\theta > 1$ indicates a larger attractiveness for human settlement in larger order basins (iii). The hypothesized log-linear Hortonian relationship between average human presence and stream order described in Figure 1a should also manifest in power-law scaling of the power spectrum of A_H (Figure 1b, red lines), as is observed for A_G (blue lines) (Marani et al., 1994). For homogeneous organization, we expect the spectral slopes to be similar ($\beta_H/\beta_G \sim 1$) (i), while greater spatial autocorrelation for clustered human settlements (ii and iii) should produce larger spectral slopes for A_H ($\beta_H/\beta_G > 1$). We also hypothesize A_H to be diagnostic of the spatial organization of human settlements, with different patterns illustrated for the three archetypes (Figure 1c, red lines): no distinct peaks are observed for the low-variability homogeneous case (i), while in cases with a wider distribution of settlement sizes, the largest settlements (cities) appear as distinct peaks – either independent of location along the stream main channel for the unstructured case (ii), or with a pronounced right skew for the downstream clustered case (iii). Finally, note that the peaks in A_H observed in Figure 1c may be driven by the variability in catchment area as a function of hydrological distance (Figure 1c blue lines). Therefore, we introduce a human settlement density function ρ_H describing the preference of human settlements in relation to the available area of catchment at a given hydrological distance from the basin outlet, which is expected to clearly outline differences among the proposed archetypes (Figure 1d).

4. Results

4.1. Horton relationship for human settlements

The 2705 global river basins with $w \geq 4$ accounted for 77% of total human settlements. Horton's law of stream areas was supported, at a significance level $p < 0.05$, for 2702 river basins, with $R^2 = 0.99 \pm 0.01$ (mean ± standard deviation), and the remaining three basins also conformed to Horton's law of areas at 0.1 level of significance. Significant relationships ($p < 0.05$) were found between log-scaled average human presence and stream order for 2473 river basins, 91% of the river basins studied, with $R^2 = 0.98 \pm 0.03$ (Figures S3a-S3c). Example Horton relationships for area and human presence are shown for the St. Lawrence and Saskatchewan-Nelson basins in North America (Figures 2a and 2b). These results demonstrate a globally robust fractal structure of human settlements in river basins, despite diverse heterogeneous landscapes and varied human-environment interactions. However, such scaling is itself insufficient to describe the spatial pattern of human settlements within river basins and in the worst case could still occur had human settlements been homogeneously distributed, as illustrated in Figure 1a. A Hortonian law for human settlements is thus expected for each of the proposed archetypes since the distribution of human settlements is related with the available basin area, which increases regularly with stream order (Carrara et al., 2012).

As shown in Figure 1a, a difference between the area and human settlement Horton ratios is required in order to discern a stream order-based pattern of human settlement. Our global analysis found R_A values between 2.2 and 8.9, with mean 4.2, similar to the relatively narrow range of 3 to 6 suggested based on statistical considerations (Kirchner, 1993). While R_H had a similar mean value of 4.5, the range (2.1 to 15.8) was much wider than for R_A, illustrating a larger variety of human settlement patterns (Figure S3d). Despite this, we found significantly different values for R_A and R_H for only 68 of the 2473 river basins, although in 81% of these (55 of 68), the human settlement Horton ratio was greater than that for basin area ($\theta > 1$) (Figure S3e), indicating a human preference for larger order basins, corresponding to archetype (iii) in Figure 1. Nevertheless, based on the overall similarity between the two Horton ratios in the global set of river basins, we applied area functions to further investigate preferential locations of human settlements in river basins.

4.2. Power-law scaling in power spectra for human settlements

We found significant power-law scaling in the power spectra of A_G for all of the 563 river basins with $w \geq 3$ for which area functions could be computed, and significant power-law scaling in the power spectra of A_H for 561 basins (Figures S4a and S4b). Power spectra for A_H and spectral slopes, β_H and β_G are illustrated here for the 10 largest river basins (Figure 3). The log-log spectral slopes are linear over approximately 3 orders of magnitude, from ~ 1km to up to 1000 km. The Mississippi, Nile, and Saskatchewan-Nelson basins presented smaller values of both β_G and β_H than the other basins, indicating slower changes of the variance of A_G and A_H across hydrological distance. A narrow range for β_G has been reported (1.7-1.9 for 11 major river basins) (Marani et al., 1994), while our global analysis found a wider range for β_G (1.23-2.06) with mean 1.53, and an even greater range for β_H (0.85-2.87) with mean 1.65. However, significant differences were found between β_G and β_H

for only 84 river basins, reflecting broadly common regulation mechanisms on the organization of river networks and human settlements in river basins, and again demonstrating the effect of the underlying basin area constraining human settlements. Among these 84 basins, $\beta_H > \beta_G$ for 83% (Figure S4c), including seven of the largest basins: Amazon, Congo, Niger, Amur, Parana, Yangtze, and Mississippi (Figure 3). Larger β_H indicates higher spatial auto-correlation of A_H than A_G, and thus clustering of human settlements into towns and cities along hydrological paths. The reference archetype of homogeneously distributed settlements was thus eliminated as a potential model for those basins with significantly larger β_H values than β_G. However, the power spectra do not reveal the locations of clustering, and thus archetypes ii and iii (Figure 1) still remained as possible patterns. To explore this issue, we therefore evaluated the normalized area functions and the human settlement density function.

4.3. Human settlement patterns from transferrable area functions

Area functions are key indicators of the organization of human settlements linked to river basin structure, as shown in Figures 2c and 2d for the St. Lawrence and Saskatchewan-Nelson basins. The ratio of A_H and A_G, defined here as the normalized human settlement density, ρ_H, plotted as a function of normalized flow length, $\hat{x}$, shows human settlements concentrated downstream for the St. Lawrence and upstream in the Saskatchewan-Nelson (Figures 2e and 2f). We classified the 3136 global basins as upstream or downstream clustered based on the ratio of the normalized flow lengths required to include half of the human settlements ($\hat{x}_{H,50}$) and half of the total basin area ($\hat{x}_{G,50}$). For example, for the St. Lawrence and Saskatchewan-Nelson river basins $\hat{x}_{H,50}/\hat{x}_{G,50}$= 0.7 and 1.2, indicating downstream and upstream clustering, respectively (Figure S5).

We found downstream clustering in the majority of river basins, 70% of the global 3136 basins, with upstream clustering in the remainder (Figure 4). The percentage of downstream clustered basins by continent (Figure S6) is 73% for Europe, 72% for Asia and Australia, 69% for North America and South America, and 63% for Africa. However, when taking account of human settlement size, the largest percentage of human settlements located in downstream clustering river basins was found for Asia (80%), while only 36% of human settlements in South America are located in downstream clustering river basins, due to the dominant effect of the two largest river basins, Amazon and Parana. In the Amazon, the largest basin in the world, settlements are preferentially upstream, although sparsely populated as a whole. Sao Paulo, the largest city in South America, is located upstream on the Parana.

4.4. Drivers of human settlement patterns in river networks

The natural advantages of downstream reaches, including low-slope flatlands, fertile soils, and deeper rivers with larger discharge facilitating navigation (Leopold & Maddock, 1953; Rodríguez-Iturbe & Rinaldo, 2001), give rise to a convergence of human settlements downstream. However, multiple factors could lead to upstream clustering of human settlements, including climatic gradients along the hydrological flow path (e.g., Calgary and Winnipeg in the more-temperate southerly upstream reaches of the northward-flowing Saskatchewan-Nelson basin), geographic constraints (e.g., Raleigh, Fayetteville, and Columbia along the Piedmont fall line in the eastern US), inland transport accessibility (e.g.,

Atlanta and Charlotte emerged as modern airport hubs from railroad origins), and political history (e.g., the establishment of inland capital cities, such as Delhi, Madrid, and Moscow in the Ganges, Tagus, and Volga basins).

A general spatial pattern of human settlements across continents emerges from the area function-based continental average normalized human settlement density, $\bar{\rho}_H$, which was much larger than one in areas close to the basin outlet ($\hat{x} \to 0$), indicating a strong human preference for these areas (Figure 5). Human settlement density decreased following power-law scaling (Figure 5, inset, $R^2 = 0.90$, $p < 0.01$) as the distance from the outlet increased, until approximately $\hat{x} = 0.4$. Beyond this distance human settlements were found to be approximately uniform with mean $\bar{\rho}_H \sim 1$. This downstream clustered pattern was consistent in each of the six continents studied and corresponded to archetype iii (Figure 1).

The observed commonality in human settlement patterns emerged from diversity and complexity, and thus may have some underlying optimality principles. The overall similarity found between the two Horton ratios and between the two spectral slopes for global river basins suggests the role of habitat availability in determining human settlement patterns. However, we also explored the relative effects of landscape attributes and proximity to basin outlets. First, the observed power law scaling of $\bar{\rho}_H$ may be related to landscape attributes. For example, across all six continents, with distance from basin outlets we find power law scaling of upstream contributing area and increasing trend of average slope (Figure S7). The normalized distance cutoff of $\hat{x} \sim 0.4$ thus can be considered as an indicator of the distance at which the advantages of the river architecture diminish for preferential human settlements.

Second, we evaluated whether the observed downstream clustering is driven entirely by the previously observed preferential location of human settlements in coastal areas (Small & Nicholls, 2003; Vitousek et al., 1997). Our area function analyses included all larger river basins, LRBs, with upstream contributing area > 1000 km^2, however, many smaller basins drain directly to the coast (Figure 6a). These smaller coastal basins, SCBs, accounted for 49% of the total global human settlements, with greater area than LRBs at the coast. Globally, we found the average human settlement density within 50 km of the coast to be higher in LRBs than in SCBs (Figure 6b). This indicates that the downstream clustering observed in Figure 5 is not just a coastal effect, but a human settlement location preference near the outlets of large river basins, perhaps related to inland waterway access to continental interiors. Human settlement densities in LRBs and SCBs in coastal areas are compared by continent in Figure S8. A preference for LRBs over SCBs in coastal areas was observed for North America, South America, and Asia. In Europe, the same preference for LRBs was observed for $d > 10$ km, while SCBs were preferred for $d < 10$ km, which may indicate prioritization of flood risk avoidance. In Africa and Australia, no preference was found between LRBs and SCBs, perhaps related to higher aridity in portions of the continental interiors, reducing the importance of inland accessibility for coastal settlement locations.

5. Conclusions

In this study, we identified a globally consistent pattern of the spatial structure of human settlements along river networks. The discovery of general patterns of human settlements within river networks at a global scale has only recently become feasible due to the convergence of advanced computation capacity and improved data availability, including global topographic and hydrographic data, and nightlight images utilized in this study. Our analyses resulted in the following important findings. First, human settlements appear to

follow a fractal structure identified by both Hortonian scaling with stream order, similar to Horton's law of areas, and power-law scaling for the power spectra of A_H, analogous to A_G. Although found robustly across the globe, the topological similarity based on Horton's ratios (91%) and power spectra (99%) mostly reflected the underlying relationship between human settlement extents and habitat availability. However, for basins with significant differences in scaling relationships for area and human settlements, the majority revealed human settlement preferences for higher-order basins (81%) and clustered structure along hydrological paths (83%).

Second, we found that multiple socio-hydromorphic factors (A_G, A_H, and $\bar{\rho}_H$) suggested a downstream clustered structure of human settlements in river basins (Figure 1, archetype iii). These hydrological distance-based analyses enabled more specific identification of human settlement location preferences than Hortonian analyses, which may not differentiate between equivalent-order basins that could be located either upstream or downstream. We found that in 70% of global river basins, the largest cities tend to be more downstream than upstream (Figure 4), with this effect diminishing at normalized flow distance beyond approximately $\hat{x} > 0.4$ (Figure 5). Further, we found that downstream clustering of human settlements is preferential near the outlets of large river basins compared to other coastal areas (Figure 6). The global clustering of human settlements within 40% of river flow length is an indicator of the preferred locations of intensified human activities.

Third, we found an emergent general pattern of downstream clustering with power-law scaling of mean normalized human settlement density along normalized hydrological length across continents (Figure 5), despite diverse characteristics in climate, geology, topography, and vegetation, and the complex interactions between human and natural systems. In the absence of global historical human settlement data, our analysis is a modern snapshot. Nevertheless, regardless of initial conditions, continental-average human settlements were found to have all converged. Seeking explanations for these emergent human settlement patterns can contribute to making future predictions; however, such drivers are usually complex and not directly observable. The fractal structure of river networks has been suggested as a natural product of least energy dissipation (Rodríguez-Iturbe et al., 1992), and the adaptation of the natural distribution of vegetation supported as a process of maximizing water use (Gao et al., 2014). Similarly, the emergent human settlement patterns may reflect an optimum self-organization for humans in river basins to better utilize natural resources, ecological assets, and geographic advantages.

Continued preferential downstream clustering of human settlements at the outlets of large river basins has several important implications for the future. Downstream human settlements will face increasing threats, considering accelerating sea-level rise (Nicholls & Cazenave, 2010) and increasing flood risks (Hirabayashi et al., 2013) due to climate change. These pressures will necessitate broader establishment and continued maintenance of robust mitigation measures. This will exert especially severe financial pressure on developing countries where increasing and rapid urbanization are expected (Cohen, 2006). However, humans may make irrational decisions away from optimality (Sivapalan, 2018), and the socio-economic optimality may change, given dramatic shifts in the external and internal drivers. For example, people might shift further inland or otherwise redistribute away from the current spatial patterns for the sake of avoiding upcoming flood risks. Holistic understanding of the underlying processes that drive the emergent spatial patterns is needed for future prediction (Sivapalan, 2018).

Last, the observed human settlement patterns within river networks can be used to generalize human impacts on rivers, improve our understanding of hydrologic and

biogeochemical responses, and advance integrated watershed management. Hydrologic science and water management are increasingly driven by future challenges including global climate change and rapid urbanization, and the further necessity to actively incorporate the dynamics of human society is demonstrated by the introduction of socio-hydrology science (Sivapalan et al., 2012; Sivapalan, 2018). Human settlement spatial patterns are key to the expansion to the interaction with human-social systems, as the preferential clustering of human settlements exerts extra pressure on local resources (Ellis & Ramankutty, 2007). Basin-wide water quality modeling mostly considers population as the major stressor and aggregates the total basin population loadings of pollutants without accounting for their spatial patterns (Schmidt et al., 2017; Van Drecht et al., 2009). However, the locations of the pollutant loadings, if varied within river networks, can lead to different effects on water quality at the river outlet, due to heterogeneous dilution and attenuation rates within river networks. The general downstream clustering of human settlement patterns across continents can imply similar effects from stressors. For example, water quality in downstream areas is often poor due to the accumulated impact of densely distributed population and intensified economic activities (UNEP, 2016; Vorosmarty et al., 2010). Meanwhile, the observed general patterns of human settlement in river basins can also imply similar basin-scale mitigation schemes. Our study could thus contribute to improved water quality modelling by incorporating spatial perspectives and facilitate targeted treatment of pollutants.

Acknowledgements

The authors thank the organizers and attendees of the following workshops for fruitful discussions and exchange of ideas: Synthesis Workshop on “Dynamics of Structure and Functions of Complex Networks” held 29 June - 16 July 2015 at Korea University, South Korea, Synthesis 1.9 Workshop on “Complexity & Feedbacks” held 21-25 March 2016 at Purdue University, USA, and Synthesis 2 Workshop on “From Dynamics of Structure to Functions of Complex Networks” held 8-19 August 2016 at Technische Universität Dresden, Germany. We also thank JongChun Kim for providing initial assistance with sub-basin boundary extraction. SC and AM acknowledge the support SWITCH-ON project, funded from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 603587. KP was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (grant number 2015R1A2A2A05001592). YF and JWJ were supported in part by USDA National Institute of Food and Agriculture Hatch project FLA-SWS-005461.

References

Beecher, H., Dott, E., & Fernau, R. (1988). Fish species richness and stream order in Washington State streams. *Environmental Biology of Fishes, 22*(3), 193-209. https://doi.org/10.1007/BF00005381

Berry, B. J. L. (1961). City size distributions and economic development. *Economic Development and Cultural Change, 9*(4), 573-588. https://doi.org/10.2307/1151867

Bertuzzo, E., Maritan, A., Gatto, M., Rodriguez-Iturbe, I., & Rinaldo, A. (2007). River networks and ecological corridors: Reactive transport on fractals, migration fronts, hydrochory. *Water Resources Research, 43*(4), W04419. https://doi.org/10.1029/2006WR005533

Bettencourt, L. (2013). The origins of scaling in cities. *Science, 340*(6139), 1438-1441. https://doi.org/10.1126/science.1235823

Bettencourt, L., Lobo, J., Helbing, D., Kuehnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America, 104*(17), 7301-7306. https://doi.org/10.1073/pnas.0610172104

Bettencourt, L., & West, G. (2010). A unified theory of urban living. *Nature, 467*(7318), 912-913. https://doi.org/10.1038/467912a

Campos, D., Fort, J., & Méndez, V. (2006). Transport on fractal river networks: Application to migration fronts. *Theoretical Population Biology, 69*(1), 88-93. https://doi.org/10.1016/j.tpb.2005.09.001

Carrara, F., Altermatt, F., Rodriguez-Iturbe, I., & Rinaldo, A. (2012). Dendritic connectivity controls biodiversity patterns in experimental metacommunities. *Proceedings of the National Academy of Sciences of the United States of America, 109*(15), 5761-5766. https://doi.org/10.1073/pnas.1119651109

Ceola, S., Laio, F., & Montanari, A. (2014). Satellite nighttime lights reveal increasing human exposure to floods worldwide. *Geophysical Research Letters, 41*(20), 7184-7190. https://doi.org/10.1002/2014GL061859

Ceola, S., Laio, F., & Montanari, A. (2015). Human-impacted waters: New perspectives from global high-resolution monitoring. *Water Resources Research, 51*(9), 7064-7079. https://doi.org/10.1002/2015WR017482

Cohen, B. (2006). Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technology in society*, 28(1-2), 63-80.

Crunkilton, R., & Duchrow, R. (1991). Use of stream order and biological indices to assess water quality in the Osage and Black river basins of Missouri. *Hydrobiologia*, *224*(3), 155-166. https://doi.org/10.1007/BF00008465

Decker, E. H., Kerkhoff, A. J., & Moses, M. E. (2007). Global patterns of city size distributions and their fundamental drivers. *Plos One, 2*(9), e934. https://doi.org/10.1371/journal.pone.0000934

Dodds, P. S., & Rothman, D. H. (2000). Scaling, universality, and geomorphology. *Annual Review of Earth and Planetary Sciences, 28*, 571-610. https://doi.org/10.1146/annurev.earth.28.1.571

Dunn, W., Milne, B., Mantilla, R., & Gupta, V. (2011). Scaling relations between riparian vegetation and stream order in the Whitewater River network, Kansas, USA. *Landscape Ecology, 26*(7), 983-997. https://doi.org/10.1007/s10980-011-9622-2

Ellis, E. C., & Ramankutty, N. (2007). Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment, 6*(8), 439-447. https://doi.org/10.1890/070062

Fang, Y., & Jawitz, J. W. (2018). High-resolution reconstruction of the United States human population distribution, 1790 to 2010. *Scientific data*, 5, 180067.

Fragkias, M., Lobo, J., Strumsky, D., & Seto, K. C. (2013). Does size matter? Scaling of CO2 emissions and US urban areas. *Plos One, 8*(6), e64727. https://doi.org/10.1371/journal.pone.0064727

Gao, H., Hrachowitz, M., Schymanski, S. J., Fenicia, F., Sriwongsitanon, N., & Savenije, H. H. G. (2014). Climate controls how ecosystems size the root zone storage capacity at catchment scale. *Geophysical Research Letters*, 41(22), 7916-7923.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., & Briggs, J. M. (2008a). Global change and the ecology of cities. *Science*, 319(5864), 756-760.

Grimm, N. B., Foster, D., Groffman, P., Grove, J. M., Hopkinson, C. S., Nadelhoffer, K. J., . . . Peters, D. P. C. (2008b). The changing landscape: ecosystem responses to urbanization and pollution across climatic and societal gradients. *Frontiers in Ecology and the Environment, 6*(5), 264-272. https://doi.org/10.1890/070147

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., ... & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9), 816.

Horton, R. E. (1945). Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin, 56*(3), 275-370.

Hutchinson, M. F., Stein, J. L., Stein, J. A., Andersion, H., & Tickle, P.K. (2008). *GEODATA 9 Second DEM Version 3 and Flow Direction Grid*. Geoscience Australia, http://www.ga.gov.au/metadata-gateway/metadata/record/gcat_66006.

Kang, S., Lin, H., Gburek, W. J., Folmar, G. J., & Lowery, B. (2008). Baseflow nitrate in relation to stream order and agricultural land use. *Journal of Environmental Quality, 37*(3), 808-816. https://doi.org/10.2134/jeq2007.0011

Kirchner, J. W. (1993). Statistical inevitability of Horton's laws and the apparent randomness of stream channel networks. *Geology, 21*(7), 591-594. https://doi.org/10.1130/0091-7613(1993)021<0591:SIOHSL>2.3.CO;2

Krugman, P. (1996). Confronting the mystery of urban hierarchy. *Journal of the Japanese and International Economies, 10*(4), 399-418. https://doi.org/10.1006/jjie.1996.0023

Kummu, M., de Moel, H., Ward, P. J., & Varis, O. (2011). How close do we live to water? A global analysis of population distance to freshwater bodies. *Plos One, 6*(6), e20578. https://doi.org/10.1371/journal.pone.0020578

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *EOS, Transactions, American Geophysical Union, 89*(10), 93-94. https://doi.org/10.1029/eost2008EO10

Leopold, L. B., & Maddock, T. (1953). *The hydraulic geometry of stream channels and some physiographic implications* (Vol. 252). US Government Printing Office.

Marani, M., Rinaldo, A., Rigon, R., & Rodriguez-Iturbe, I. (1994). Geomorphological width functions and the random cascade. *Geophysical Research Letters, 21*(19), 2123-2126. https://doi.org/10.1029/94GL01933

Miyamoto, H., Hashimoto, T., & Michioku, K. (2011). Basin-wide distribution of land use and human population: stream order modeling and river basin classification in Japan. *Environmental Management, 47*(5), 885-898. https://doi.org/10.1007/s00267-011-9653-0

Moussa, R. (2008). What controls the width function shape, and can it be used for channel network comparison and regionalization? *Water Resources Research, 44*(8), W08456. https://doi.org/10.1029/2007WR006118

National Oceanic and Atmospheric Administration. (2013). *Version 4 DMSP-OLS nighttime lights time series*. National Geophysical Data Center, https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

Nicholls, R. J., & Cazenave, A. (2010). Sea-level rise and its impact on coastal zones. *Science*, 328(5985), 1517-1520.

Peckham, S. D., & Gupta, V. K. (1999). A reformulation of Horton's Laws for large river networks in terms of statistical self-similarity. *Water Resources Research*, *35*(9), 2763-2777. https://doi.org/10.1029/1999WR900154

Platts, W. S. (1979). Relationships among stream order, fish populations, and aquatic geomorphology in an Idaho river drainage. *Fisheries*, *4*(2), 5-9. https://doi.org/10.1577/1548-8446(1979)004<0005:RASOFP>2.0.CO;2

Rodríguez-Iturbe, I., Rinaldo, A., Rigon, R., Bras, R. L., Ijjasz-Vasquez, E., & Marani, A. (1992). Fractal structures as least energy patterns: The case of river networks. *Geophysical Research Letters*, 19(9), 889-892.

Rodríguez-Iturbe, I., & Rinaldo, A. (2001). *Fractal river basins: chance and self-organization.* New York: Cambridge University Press.

Schmidt, C., Krauth, T., & Wagner, S. (2017). Export of Plastic Debris by Rivers into the Sea. *Environmental science & technology*, 51(21), 12246-12253.

Sivapalan, M., Savenije, H. H., & Blöschl, G. (2012). Socio-hydrology: A new science of people and water. *Hydrological Processes*, 26(8), 1270-1276.

Sivapalan, M. (2018). From engineering hydrology to Earth system science: milestones in the transformation of hydrologic science. *Hydrology and Earth System Sciences*, 22(3), 1665.

Small, C., & Cohen, J. E. (2004). Continental physiography, climate, and the global distribution of human population. *Current Anthropology*, *45*(2), 269-277. https://doi.org/10.1086/382255

Small, C., & Nicholls, R. J. (2003). A global analysis of human settlement in coastal zones. *Journal of Coastal Research*, *19*(3), 584-599.

Stein, J. L., Hutchinson, M. F., & Stein, J. A. (2014). A new stream and nested catchment framework for Australia. *Hydrology and Earth System Sciences*, *18*(5), 1917-1933. https://doi.org/10.5194/hess-18-1917-2014

Stenger-Kovacs, C., Toth, L., Toth, F., Hajnal, E., & Padisak, J. (2014). Stream order-dependent diversity metrics of epilithic diatom assemblages. *Hydrobiologia*, *721*(1), 67-75. https://doi.org/10.1007/s10750-013-1649-8

UNEP. (2016). *A Snapshot of the World's Water Quality: Towards a global assessment*. United Natons Environment Programme, Nairobi, Kenya.

Van Drecht, G., Bouwman, A. F., Harrison, J., & Knoop, J. M. (2009). Global nitrogen and phosphate in urban wastewater for the period 1970 to 2050. *Global Biogeochemical Cycles*, 23(4).

US Geological Survey. (2010). *HYDRO1k elevation derivative database*. National Center for Earth Resource Observations and Science, https://lta.cr.usgs.gov/HYDRO1K.

Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*, *277*(5325), 494-499. https://doi.org/10.1126/science.277.5325.494

Vorosmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., . . . Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*(7315), 555-561. https://doi.org/10.1038/Nature09440

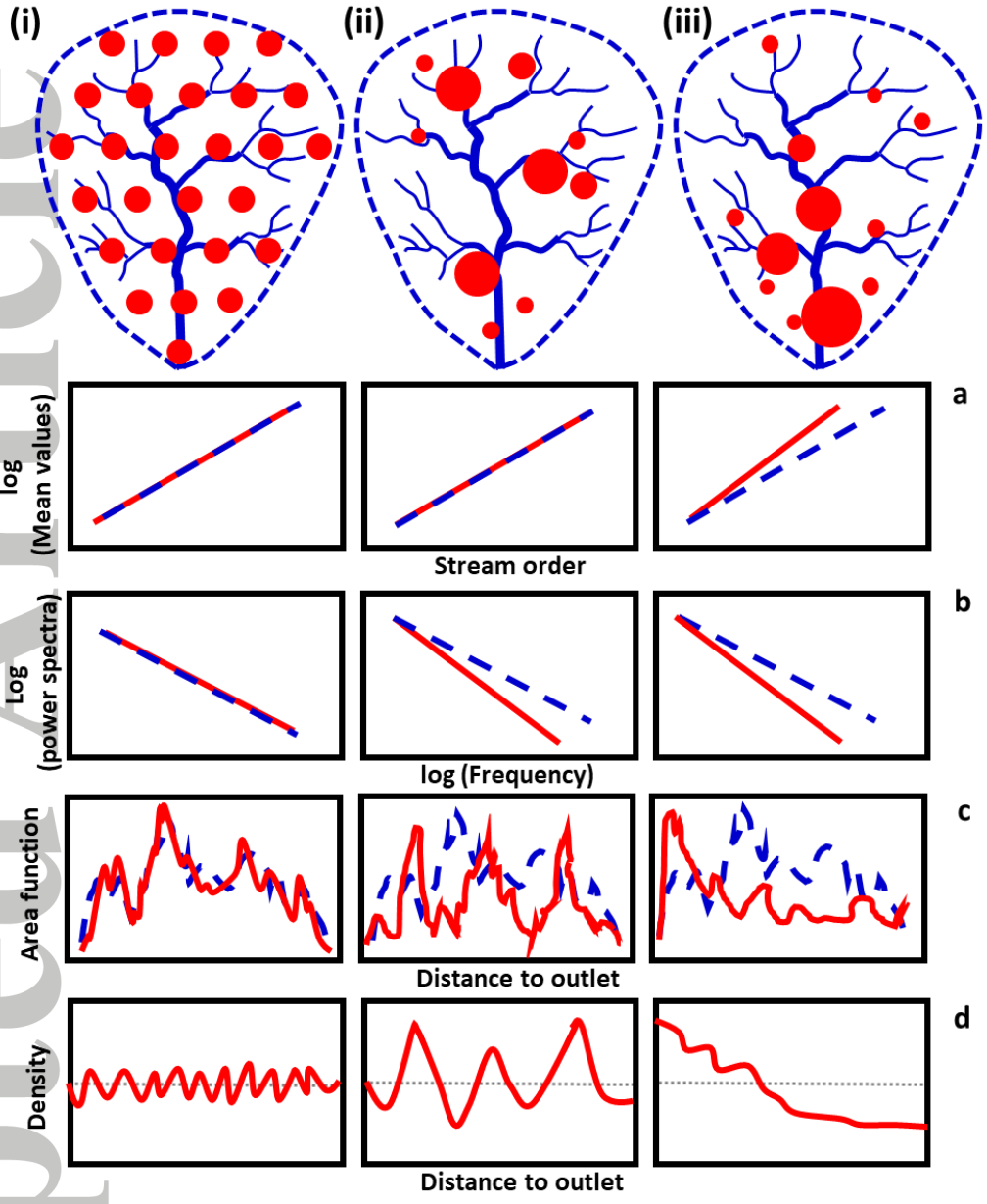

Figure 1: Archetypes of hypothesized human settlement organization in catchments, and associated metrics of pattern organization: (i) homogeneity, (ii) unstructured clustering along river networks, and (iii) downstream clustering. Solid and dashed blue lines indicate simplified river networks and basin boundary; red circles indicate human settlements in different patterns of spatial organization. The metrics applied to basin area (dashed blue lines) and human settlements (solid red lines) are: (a) Horton's laws, (b) power spectra of geomorphologic area function (A_G) and human area function (A_H), (c) A_G and A_H, and (d) human settlement density function (ρ_H).

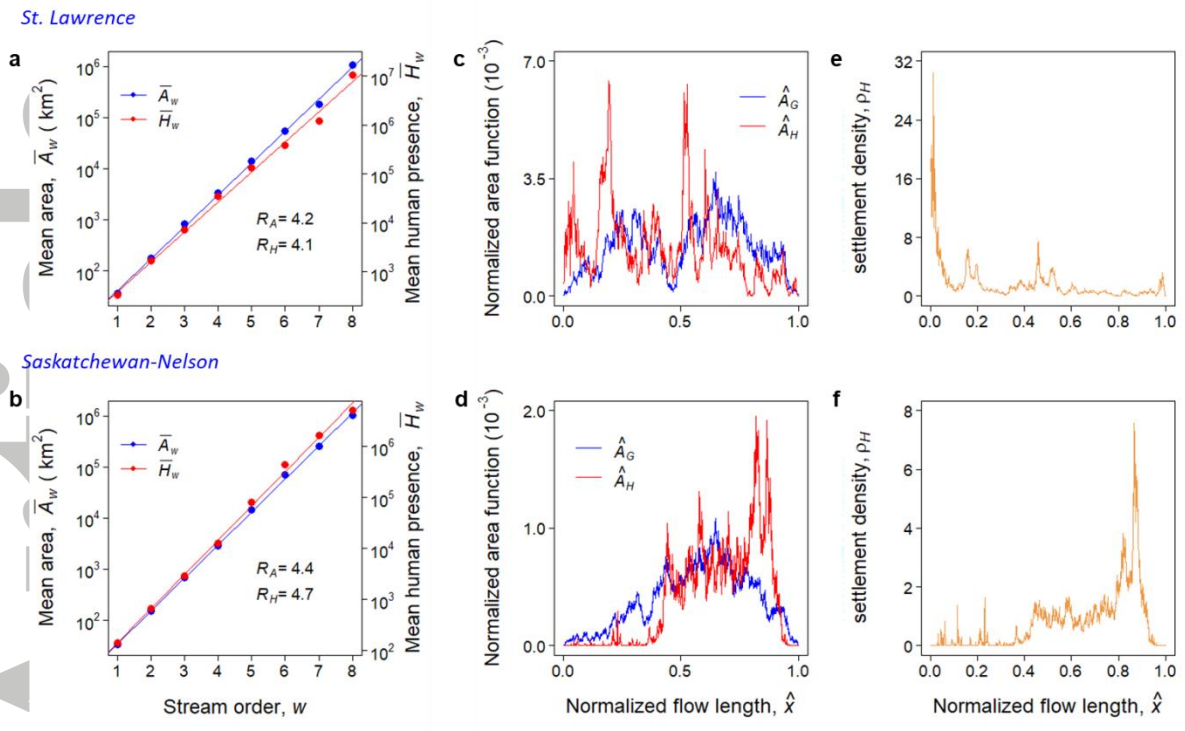

Figure 2: Hortonian analyses and area function-based analyses of human settlement patterns for the St. Lawrence and Saskatchewan-Nelson basins. (a-b) Hortonian analysis of mean area and mean human presence vs stream order. Values of the area and human settlement ratios, R_A and R_H, are inset. (c-d) Normalized geomorphological area function ($\hat{A}_G$) and normalized human area function ($\hat{A}_H$) vs normalized flow length, $\hat{x}$. (e-f) normalized human settlement density (ρ_H) vs $\hat{x}$.

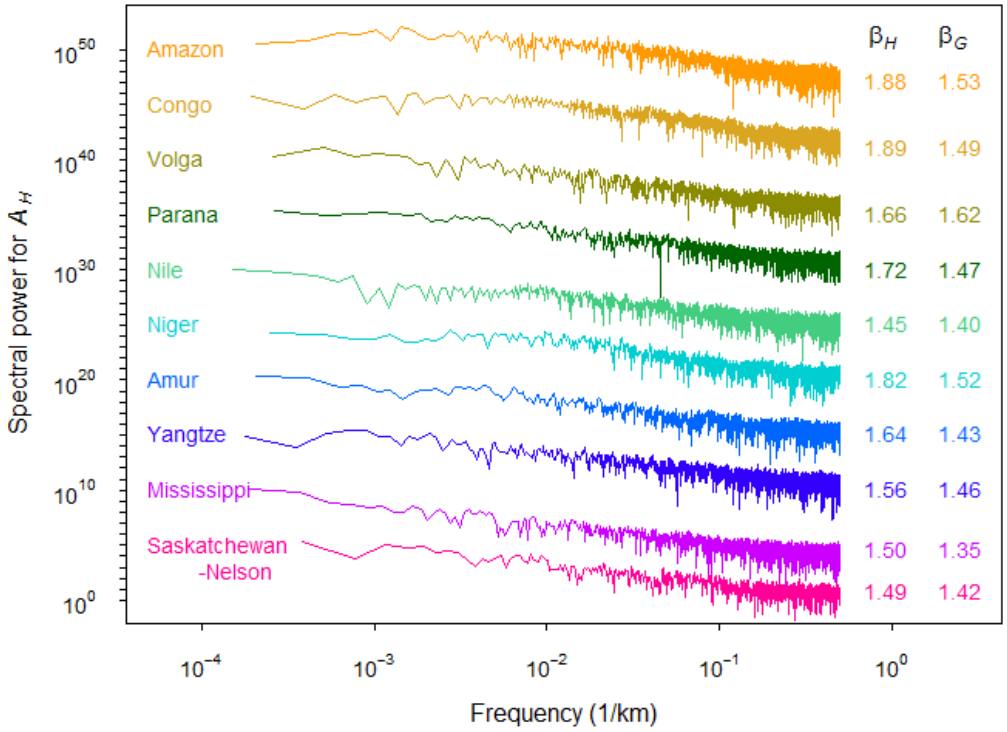

Figure 3. Power spectra for human area function (A_H) of the 10 largest river basins, and their associated spectral slopes β_H. Also shown are geomorphological area function power spectra slopes, β_G. Spectra for individual basins have been arbitrarily shifted on the y-axis for visualization.

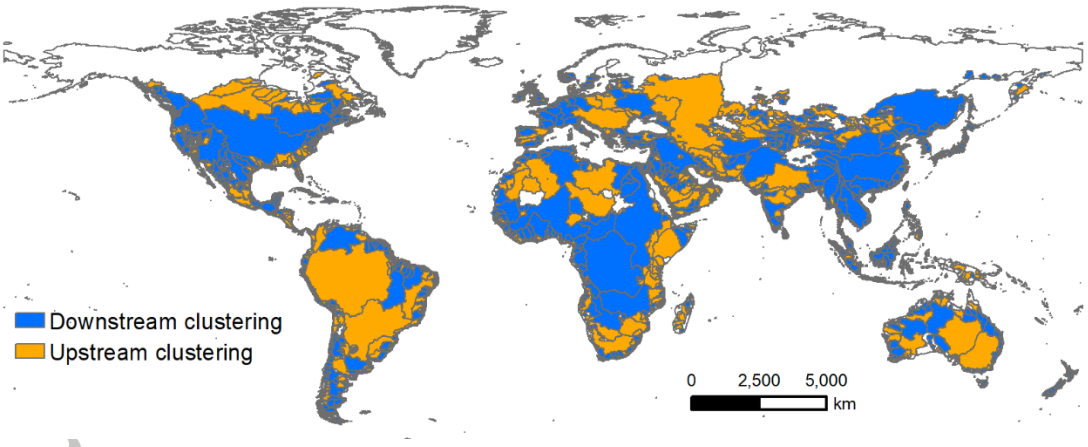

Figure 4. Downstream (blue) or upstream (orange) clustering of human settlements in 3136 global river basins. Downstream clustering is observed in 70% of river basins.

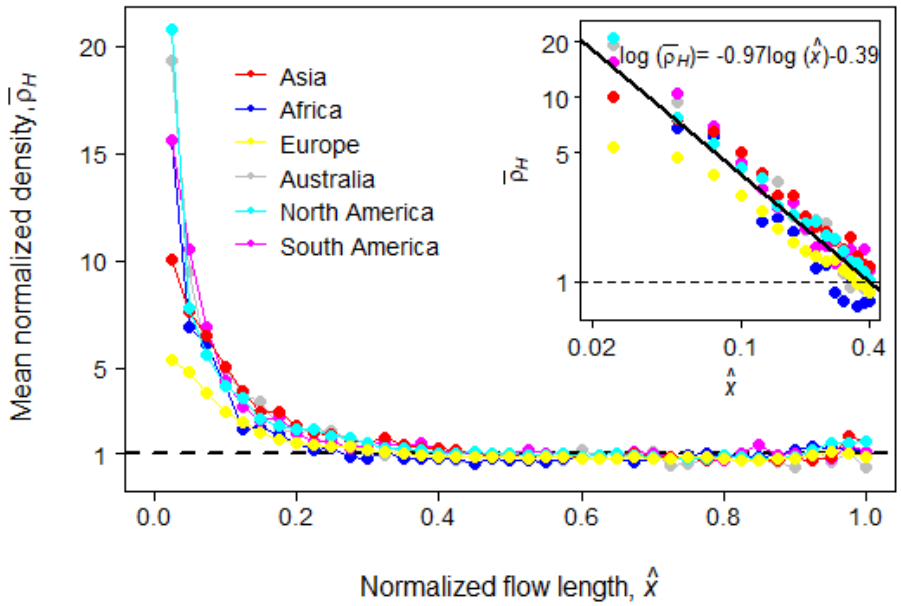

Figure 5. Continent-based analysis of mean normalized human settlement density ($\bar{\rho}_H$) along hydrological flow paths in river basins. The dashed line ($\bar{\rho}_H = 1$) indicates uniformly distributed human settlement density. Inset shows power law scaling between $\bar{\rho}_H$ and $\hat{x}$ for $\hat{x} < 0.4$.

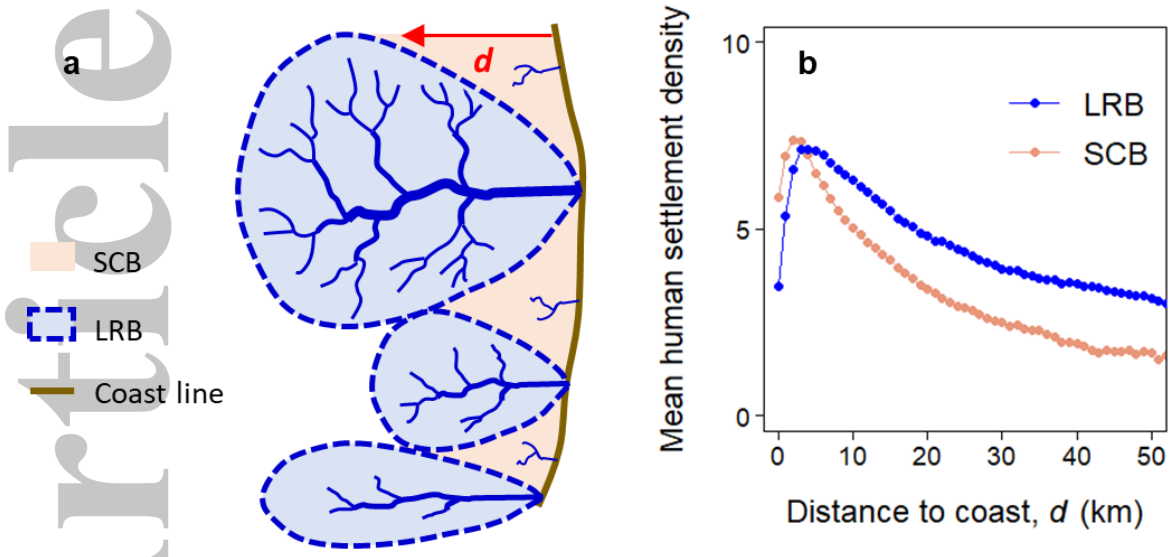

Figure 6. Human settlements in coastal areas. (a) Large river basins (LRBs) and smaller coastal basins (SCBs) in coastal areas. The red arrow indicates Euclidean distance from the coast, *d*. (b) Global mean human settlement density as a function of *d*. Overall, densities are greater in LRBs than in SCBs. Note LRBs are river basins delineated based on Hydro1K dataset, while SCBs refer to basins along the coast with area < 1000 km^2.