Solving the mystery of vanishing rivers in China

Citation for published version:
Wang, Y, Ni, J, Yue, Y, Li, J, Borthwick, A, Cai, X, Xue, A, Li, L & Wang, G 2019, 'Solving the mystery of vanishing rivers in China', *National Science Review*.

Published In:
National Science Review

OXFORD
UNIVERSITY PRESS

National Science Review

Solving the mystery of vanishing rivers in China

Journal:	*National Science Review*
Manuscript ID	NSR_MS-2018-192
Manuscript Type:	Research Article
Date Submitted by the Author:	16-Oct-2018
Complete List of Authors:	Wang, Yichu; Peking University, The Key Laboratory of Water and Sediment Sciences, Ministry of Education; College of Environmental Sciences and Engineering; Peking University, Beijing Innovation Center for Engineering Science and Advanced Technology (BIC-ESAT) Ni, Jinren; Peking University, The Key Laboratory of Water and Sediment Sciences, Ministry of Education; College of Environmental Sciences and Engineering; Qinghai University, State Key Laboratory of Plateau Ecology and Agriculture Yue, Yao; Wuhan University, School of Water Resources and Hydropower Engineering Li, Jiaye; Tsinghua University, State Key Laboratory of Hydroscience and Engineering Borthwick, Alistair; The University of Edinburgh, School of Engineering Cai, Ximing ; University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering Xue, An; Peking University, The Key Laboratory of Water and Sediment Sciences, Ministry of Education; College of Environmental Sciences and Engineering Li, Li; Peking University, The Key Laboratory of Water and Sediment Sciences, Ministry of Education; College of Environmental Sciences and Engineering Wang, Guangqian; Qinghai University, State Key Laboratory of Plateau Ecology and Agriculture
Subject Categories:	GEOSCIENCES
Keywords:	vanishing rivers, river network, stream-segmentation, stream-order rules

Title: Solving the mystery of vanishing rivers in China

Authors: Yichu Wang[1, 2], Jinren Ni[1, 3, *], Yao Yue[4], Jiaye Li[5], Alistair G. L. Borthwick[6], Ximing Cai[7], An Xue[1], Li Li[1], Guangqian Wang[3]

Affiliations:

[1]The Key Laboratory of Water and Sediment Sciences, Ministry of Education; College of Environmental Sciences and Engineering, Peking University, Beijing, 100871, P. R. China

[2]Beijing Innovation Center for Engineering Science and Advanced Technology (BIC-ESAT), Peking University, Beijing, 100871, P. R. China

[3]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining, 810016, P. R. China

[4]School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan, 430072, P. R. China

[5]State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, 100084, P. R. China

[6]School of Engineering, The University of Edinburgh, The Kings Buildings, Edinburgh EH9 3JL, UK

[7]Ven Te Chow Hydrosystems Lab, Department of Civil and Environmental Engineering, University of Illinois, Urbana, IL 61801 USA

Author contributions:

J.R.N. designed the work and analyzed the cause of disappearing rivers; Y.C.W. completed drainage networks extraction and river statistical work with the help of Y.Y., J.R.N., J.Y.L., A.X. and L.L.; Y.C.W., J.R.N. and A.G.L.B. wrote the manuscript, G.Q.W. and X.M.C. contributed new ideas. All authors have read, and approved the final version of the manuscript.

[*]**Corresponding author:** Jinren NI

Address: Environmental Building 315, Peking University, Beijing, 100871, China

Tel.: +86-10-62751185

E-mail address: jinrenni@pku.edu.cn

Abstract

A major controversy was sparked worldwide by a recent national water census claiming that the number of Chinese rivers with watersheds ≥100 km^2 was less than half the previous estimate of 50,000 rivers, which also stimulates debates on the potential causes and consequences. Here we estimated number of rivers in terms of stream-segmentation characteristics described by Horton, Strahler, and Shreve stream-order rules, as well as their mixed mode for named rivers recorded in the Encyclopedia of Rivers and Lakes in China. As results, the number of "vanishing rivers" is found highly relevant to statistical specifications in addition to erroneous inclusion of pseudo-rivers primarily generated in arid or frost-thaw areas. The modified Horton stream-order scheme reasonably depicts the configuration of complete natural streams from headwater to destination, while the Strahler largely projects the fragmentation of the named river networks associated with human aggregation to the hierarchical river systems.

Key Words: vanishing rivers, river network, stream-segmentation, stream-order rules

Introduction

River networks are of primary importance to the continental land mass, water cycle, and transportation of materials from land to ocean. In 2013, the First National Census of Water [1] estimated that the number of rivers in China with the catchment area greater than 100 km^2 was 22,909, a reduction of about 27,000 from the previously accepted figure of over 50,000 rivers listed in a geography textbook published in the 1990s [2]. This incredible discrepancy also led to international debate [3] as to why rivers are vanishing in China.

Hitherto, tremendous efforts have been made to identify natural river networks from a physical perspective, as hierarchical branching systems. Drainage networks are usually extracted from Digital Elevation Models (DEMs) assuming that water always flows in the steepest downhill direction, and precipitation and geomorphic conditions are uniform [4]. Although these assumptions are generally valid, modification is needed for river networks in arid or semiarid areas [5]. In practice, the generated drainage networks could be further quantified in terms of different stream-order rules such as the conventional Horton [6], Strahler [7, 8], and Shreve [9] schemes, besides other options including extrapolation from exemplary basins in absence of accurate data [2] or statistics from river encyclopedia [10]. Meanwhile, flow intermittency and hydrologic dis-connectivity of rivers due to climate variability [11, 12] and human interventions (e.g. dam construction, urbanization, and irrigation) [13-15] over the past decades have considerably increased the complexity of understanding river networks.

Comparing with natural river fragmentations, there exists a gap concerning the identification of named river fragmentations. River naming is of fundamental importance to understanding river systems, and the complexity of its identification system derives from various humanity factors [10]. Traditional stream-order schemes, based on the physical structure of rivers, are poorly suited to describing river systems coded with human connotations. To obtain insights into the essential characteristics of river networks, a hybrid stream-order model specifically for estimating the number of named rivers is proposed. This mixed mode, together with other conventional stream-order rules, enables interpretation of the huge discrepancy between the most well-known estimates of river numbers in China.

Number of rivers estimated by conventional stream-order rules

From a hydrological perspective, China can be divided into ten large basins. The drainage networks of those basins are identified from a 30 m × 30 m resolution DEMs using a binary search method [16] after pretreatment to remove land surface sinks from the ends of inland rivers (see Supplementary Fig. 1). Supplementary Fig. 2 shows the resulting hierarchy of drainage networks, of order up to 10. High-order rivers form backbones of the drainage networks, low-order tributaries complete the detailed skeleton.

Figure 1a presents the numbers of rivers with different stream orders in China estimated by Horton, Strahler, and Shreve stream-order schemes, respectively. Here, the Shreve orders are reclassified to correspond to the range of the Horton and Strahler orders. For all schemes, the number of rivers diminishes as stream order increases. The results highlight the scheme-dependent statistical characteristics with respect to the number of rivers. For orders 1 to 10 the number reduces from 2,999,117 to 7 using Horton, 3,792,656 to 7 using Strahler, and 3,792,656 to 3,044 using Shreve. At the lowest order, the river numbers derived from the three schemes are very close; at the high orders the numbers diverge. Figure 1b presents the numbers of rivers aggregated by catchment areas. For catchment areas from < 10 km^2 to ≥ 100,000 km^2, the number of rivers significantly decreases.

[INSERT FIGURE 1 HERE]

We now focus on rivers in China with the catchment area greater than 100 km^2, of which about 80% are order 4 to order 10. Supplementary Table 1 lists the number of rivers with catchment area ≥ 100 km^2 reported in the National Census [1] (N_{Census}), and estimated by Horton (N_{Horton}), Strahler (N_{Strahler}), and Shreve (N_{Shreve}), respectively. Here N_{Horton} (27,187) provides the closest match to N_{Census} (22,909) with a relative difference of about 19%. The Strahler scheme gives an estimate that is 40% higher than that of the National Census. The Shreve scheme provides an estimate that is 29 times larger than N_{Census} because of the different stream-ordering schemes used. The similarity between the N_{Horton} and N_{Census} estimates is unsurprising given that the national Census was implemented using a Horton framework. Discrepancies are evident, resulting from the generation of pseudo-rivers due to lacking of consideration of factors such as climate, landform, and water utilization [5].

Number of rivers after exclusion of pseudo-rivers

Exclusion of pseudo-rivers is essential to approach the actual number of rivers approximately reflected by China's First National Census for Water. The pseudo-river could be regarded as a channel without runoff in the wet season. For precipitation-dependent rivers (Figure 2a), a basin is more capable of producing runoff when it has higher precipitation and lower evaporation [17], coincides with an aridity index (AI) [18] less than certain prescribed threshold value (Supplementary Fig. 3, more details see Supplementary data). Figure 2b maps the distribution of AI as it increases from southeast to northwest across China. For basins where rivers are mainly supplied by glacial meltwater and groundwater [19-22], modification is made based on the Water Occurrence (WO) [23] representing the percentage of available observations in presence of water during 1984-2015 (Figure 2c, more details see Supplementary data). As a result, the total number of rivers (N'_{Horton}) of catchment area ≥ 100 km^2 estimated after removal of pseudo-rivers (N_{P} = 4,577 $^{+60}_{-235}$) reduces to 22,610, which is only 1.3% less than the Census value (Supplementary Table 2, Supplementary Table 3). Figure 2d maps the density of pseudo-rivers with catchment area ≥ 100 km^2 for each province (D_{P}), which is equal to N_{P} divided by the area of the provinces. D_{P} increases from southeast to northwest across China. Pseudo-rivers are mainly located in

Northwestern China (NW), Inner Mongolia (IM), and part of Tibet Plateau (TP) (corresponding to Inland, Yellow, and international river basin). Inland, Yellow, and international river basin contribute 78%, 10%, and 8% of pseudo-rivers, respectively. In humid regions (i.e. Yangtze, Huai, Pear river basins, and rivers in Zhejiang and Fujian), no pseudo-rivers are identified. Remaining basins (i.e. Amur, Liao, Hai) contribute 4% of pseudo-rivers altogether. About 70% of pseudo-rivers with catchment area ≥ 100 km^2 are generated in arid or frost-thaw areas (i.e. inland river basin and international river basin) due to local violations of the assumptions underpinning runoff simulation (Supplementary Fig. 4). The remaining 30% pseudo-rivers distributed in the North China (NC), Northeast China (NE), and Southwest China (SW) are gullies or dried-out rivers resulted from an arid regional climate with scarce precipitation and strong evaporation, excessive exploitation of water resources, or special landform. In China's arid and semi-arid areas, the river discharges tended to decrease in recent decades [24-26]. Consequently, some perennial streams have converted to intermittent rivers, even to permanently dry gullies [27]. Long-term overconsumption of water resource also ultimately led to the drying out of rivers [27, 28]. In Southwest China, most precipitation rainfall infiltrates directly underground due to the widespread Karst landform, causing many channels to be unable to maintain runoff [29]**.**

[INSERT FIGURE 2 HERE]

After correction of the errors induced by pseudo-rivers, the Strahler scheme gives an estimate ($N'_{\mathrm{Strahler}} = 31{,}615$) that is 38% higher than that of the National Census; the modified Shreve scheme provides an estimate ($N'_{\mathrm{Shreve}} = 557{,}161$) that is 24 times larger than N_{Census} (Supplementary Table 1) due to different stream-segmentation characteristics of river networks described by Horton, Strahler, and Shreve stream-order rules. However, the influences of incomplete topographic data and landform changes should be fully considered in quantitative studies on river networks (Supplementary Table 4–8, Supplementary Fig. 5, Supplementary Fig. 6).

Number of named rivers based on a hybrid stream-order model

Another option for numbering rivers is from records in various geographical chorographies. Historically, river courses have been segmented by random naming which has evolved alongside human culture related to water [10]. However, traditional stream-ordering schemes are not suited to describing river systems fragmented with historically river naming. Here, a systematic comparison is undertaken for 107 river networks (Supplementary Fig. 7) listed in the *Encyclopedia of Rivers and Lakes in China* [10] with DEM-generated river networks analyzed in terms of different stream-order schemes. We find that arbitrary named river networks can be reasonably replicated using a hybrid of Horton, Strahler, and Shreve stream-order schemes. Each basin has a corresponding best-fit proportion in terms of the occurrence possibilities of the three stream-order systems ($P_f = P_{HO} : P_{ST} : P_{SH}$), and the most common P_f is 0.1 : 0.8 : 0.1, which occurs with 21 of the 107 representative basins (more detail see Supplementary data).

In the spirit of the attraction of inhabitants towards rivers [30, 31], we find that P_f for naming river networks is moderately relevant to population density in the year of 1776 during Early Qing Dynasty (D_{1776}) for 107 representative basins (R = 0.4, Supplementary Table 9). Thus, six types of basins are classified in terms of D_{1776} (Supplementary Table 10). For each class, the most frequently occurring value of P_f has been identified as ($P_{HO} : P_{ST} : P_{SH}$ = 0.4 : 0.6 : 0), (0.3 : 0.6 : 0.1), (0.2 : 0.7 : 0.1), (0.1 :0.8 : 0.1), (0 : 1 : 0), and (0 : 0.9 : 0.1), respectively in basin types of I, II, III, IV, V and VI. The population densities are selected from census information conducted in the year of 1776 during the Qing Dynasty, when complete provincial demographic statistics [32] and river naming systems [33] became commonly established. Since then, the naming system of river networks has remained almost unchanged [10, 33] even though China has experienced rapid urbanization [34] and population redistribution [35] in recent decades.

Having determined P_{f} in terms of human aggregation scale, multiple random tests are then implemented in conjunction with the Monte Carlo method [36], employing a pair-wise scheme for consistency conflict analysis based on a tree structure from graph theory (Figure 3, more detail see Supplementary data). Taking the above proportions of Horton: Strahler: Shreve as input (Supplementary Fig. 8), the total number of named rivers with catchment area ≥ 100 km^2 (N_{Named}) in China in the 1990s is estimated as $30{,}201^{+1,299}_{-1,187}$ (Supplementary Table 11), which is $7{,}591^{+1,299}_{-1,187}$ more than that generated solely by the Horton scheme.

[INSERT FIGURE 3 HERE]

We further investigated the distribution of both natural river density and named river density in China. Figure 4a maps the distribution of modified compound topographic index (*MCTI*) which combines surface flow accumulation, aridity, and topographic slope (see Supplementary data). *MCTI* shows an apparent decrease from east to west China, with areas of higher *MCTI* corresponding to greater humidity and flatter topography. Figure 4b maps the natural stream frequency distribution (D_{Horton}) throughout China. High values of D_{Horton} are distributed in Central China (CC), Southeast China (SE) and North China (NC), with the mean values of 3.26×10^{-3}, 3.14×10^{-3} and 4.11×10^{-3} km^{-2}. River networks are sparse in the Northwest China (NW), Southwest China (SW), and Tibet Plateau areas (TP), with mean values of D_{Horton} 2.62×10^{-3}, 2.93×10^{-3}, and 2.95×10^{-3} km^{-2}. It appears that river networks of higher density occur more frequently in humid and flat areas. Figure 4c maps the distribution of population density in early Qing Dynasty (D_{1776})

throughout China (36). The most populated regions are located in SE, NC, and CC, with mean values of D_{1776} 170.26, 98.57, and 80.03 person · km^{-2}, respectively. Less populated regions are located in TP (1.29 person · km^{-2}), NE (1.44 person · km^{-2}), and Inner Mongolia (IM, 2.61 person · km^{-2}). In Figure 4d, the distribution of river naming segment density (D_{Named}) defining the number of named rivers per unit area is demonstrated. Higher values of D_{Named} are found in NC (4.84×10^{-3} km^{-2}), SE (4.83×10^{-3} km^{-2}), and SC and CC (4.52×10^{-3} km^{-2}), whereas lower values occur in NW (3.28×10^{-3} km^{-2}), TP (3.60×10^{-3} km^{-2}) and NE (3.61×10^{-3} km^{-2}). It appears that the most heavily populated areas coincide with areas which have high densities of named rivers.

[INSERT FIGURE 4 HERE]

Implications of different statistical specifications

The conventional Horton, Strahler and Shreve stream-order schemes imply different understandings of hierarchical river networks. Horton scheme [6], describing rivers extending from headwater to outlets, is suitable for understanding natural configuration of a complete river system. The Strahler scheme [7, 8] generates a nested hierarchy of drainage-basin forms, each of which could serve as an open physical system in terms of inputs of precipitation and outputs of runoff, and potentially describes the scale of human aggregation and activities along a large river defined by Horton. Shreve scheme [9] sums the number of sources in each catchment above a stream gauge or outflow, and is preferred in more accurate considerations in hydrodynamic and source control studies.

[INSERT FIGURE 5 HERE]

In summary, compared with the previous estimate of 50,000 rivers with watersheds larger than 100 km^2 in 1990s mostly extrapolated from exemplary river networks in southeast areas [2],

Figure 5 indicates a much smaller number of rivers in terms of Horton (22,610), which seems an exact reflection of natural composition of river networks from headwater to destination described by the latest national census (22,909). On the other hand, the greater number of rivers derived from the Strahler scheme (31,615) is very close to the number of historically named rivers (30,201), as if a good projection of human aggregation and water culture to the hierarchical river networks. Erroneous inclusion of pseudo-rivers mostly in arid or frost-thaw areas and topographic inputs may also contribute to the huge discrepancy in number of rivers. Overall, this study is not only of great significance to revealing the mystery of "vanishing rivers" in China in terms of statistical specifications, but also helpful to getting insights into the essentiality of representative river-segmentation modes and their relationship upon global river networks.

Methods

An efficient method proposed by Bai *et al*. (2015) [16] was adopted for extraction of drainage networks. For inland rivers, pretreatment is undertaken (Supplementary Fig. 1).

Three stream-order schemes including Horton [6], Strahler [7, 8], and Shreve [9] stream-ordering schemes are used to describe the hierarchical river networks in this paper.

Pseudo-rivers are identified based on two parameters, these are, Aridity index (*AI*) and Water Occurrence (*WO*).

The number of named rivers in China was estimated with a multiple random tests in conjunction with the Monte Carlo method [36], employing a pair-wise scheme for consistency conflict analysis based on a tree structure from graph theory.

Details of Methods and Data Sources are given in the Supplementary data.

Acknowledgments

Discussions on the river network by J. W. Pang (Ministry of Water Resources, P.R. China), J. Y. Zhang, J. Deng, and Z. D. Lin are particularly acknowledged.

Funding

This work was supported by National Natural Science Foundation of China (grant numbers 51721006 and 91647211).

References

1. Ministry of Water Resources, PRC and National Bureau of Statistics, PRC. *Bulletin of First National Census for Water*. Beijing: China Water &Power Press, 2013.
2. Huang, XQ. *China's Rivers (in Chinese)*. Beijing: The Commercial Press, 1995.
3. Hsu, A, Miao, W. 28,000 Rivers Disappeared in China: What Happened? *the Atlantic*.https://www.theatlantic.com/china/archive/2013/04/28-000-rivers-disappeared-in-china-what-happened/275365/ (29 April 2013).
4. O'Callaghan, JF and Mark, DM. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* 1984; **28**: 323–344.

5. Schneider A, Jost, A, Coulon, C *et al.* Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density. *Geophys. Res. Lett.* 2017; **44**: 2773–2781.
6. Horton, RE. Erosional development of streams and their drainage basins: hydro-physical approach to quantitative morphology. *Geol. Soc. Am. Bull.* 1945; **56**: 275–370.
7. Strahler, AN. Hypsometric (area-altitude) analysis of erosional topography. *Geol. Soc. Am. Bull.* 1952; **63**: 1117–1142.
8. Strahler, AN. Quantitative analysis of watershed geomorphology. *EOS Trans. AGU* 1957; **38**: 913–920.
9. Shreve, R. Statistical Law of Stream Numbers. *J. Geol.* 1966; **74**: 17–37.
10. Editorial Committee of Encyclopedia of Rivers and Lakes in China. *Encyclopedia of Rivers and Lakes in China (in Chinese).* Beijing: China Water Power Press, 2014.
11. Milly, PCD, Dunne, KA and Vecchia, AV. Global pattern of trends in streamflow and water availability in a changing climate. *Nature* 2005; **438**: 347–350.
12. Dai, A. Increasing drought under global warming in observations and models. *Nat. Clim. Chang.* 2013; **3**: 171–171.
13. Nilsson, C, Reidy, CA and Dynesius, M *et al.* Fragmentation and Flow Regulation of the World's Large River Systems. *Science* 2005; **308**: 405–408.
14. Haddeland, I, Heinke, J and Biemans, H *et al.* Global water resources affected by human interventions and climate change. *Proc. Natl. Acad. Sci. USA* 2014; **111**: 3251–3256.
15. Cheng, GD, Li, X and Zhao, WZ *et al.* Integrated study of the water–ecosystem–economy in the Heihe River Basin. *Natl. Sci. Rev.* 2014; **1**: 413–428.
16. Bai, R, Li, TJ and Huang, YF *et al.* An efficient and comprehensive method for drainage network extraction from DEM with billions of pixels using a size-balanced binary search tree.

Geomorphology 2015; **238**: 56–67.

17. Rui, XF. *Principles of Hydrology (in Chinese).* Beijing: China Water & Power Press, 2004.
18. Middleton, N, Thomas, DS and UNEP. *World Atlas of Desertification.* London: Edward Arnold, 1992.
19. Sorg, A, Bolch, T and Stoffel, M *et al.* Climate change impacts on glaciers and runoff in Tien Shan (Central Asia). *Nat. Clim. Chang.* 2012; **2**: 725–731.
20. Xiao, J, Jin, ZD and Wang, J *et al.* Hydrochemical characteristics, controlling factors and solute sources of groundwater within the Tarim River Basin in the extreme arid region, NW Tibetan Plateau. *Quat. Int.* 2015; **380**: 237–246.
21. Lutz, AF, Immerzeel, WW and Shrestha, AB *et al.* Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nat. Clim. Chang*. 2014; **4**: 587–592.
22. Li, Z, Lin, XQ and Coles, AE *et al.* Catchment-scale surface water-groundwater connectivity on china's loess plateau. *Catena* 2017; **152**: 268–276.
23. Pekel, JF, Cottam, A and Gorelick, N *et al.* High-resolution mapping of global surface water and its long-term changes. *Nature* 2016; **540**: 418–422.
24. Ma, ZM, Kang, SZ and Zhang, L *et al.* Analysis of impacts of climate variability and human activity on streamflow for a river basin in arid region of northwest China. *J. Hydrol.* 2008; **352**: 239–249.
25. Hang, YF, Guan, DX and Jin, CJ *et al*. Analysis of impacts of climate variability and human activity on streamflow for a river basin in northeast china. *J. Hydrol.* 2011; **410**: 239–247.
26. Zhao, GJ, Li, EH and Mu, XM *et al.* Changing trends and regime shift of streamflow in the Yellow River basin. *Stoch. Environ. Res. Risk Assess.* 2015; **29**: 1331–1343.
27. Acuña, V, Datry, T and Marshall, J *et al.* Why should we care about temporary waterways?

Science 2014; **343**: 1080–1081.

28. Guo, Y and Shen, YJ. Quantifying water and energy budgets and the impacts of climatic and human factors in the Haihe River Basin, China: 2. Trends and implications to water resources. *J. Hydrol.* 2015; **527**: 251–261.
29. Liu, MX, Xu, XL and Wang, DB *et al.* Karst catchments exhibited higher degradation stress from climate change than the non-karst catchments in southwest China: An ecohydrological perspective. *J. Hydrol.* 2016; **535**: 173–180.
30. Qi, W, Liu, SH and Zhao, MF *et al.* China's different spatial patterns of population growth based on the "Hu Line". *J. Geogr. Sci.* 2016; **26**: 1611–1625.
31. James, PE. *All Possible Worlds: A History of Geographical Ideas.* Indianapolis: The Bobbs-Merrill Company, 1972.
32. Ge, JX. *History of Population in China (Volume 5), Qing Dynasty (in Chinese)*. Shanghai: Fudan University Press, 2001.
33. Qi, ZN. *An Outline of Waterways (in Chinese).* Shanghai: Wenruilou Press, 1878.
34. Yang, L, Xu, YP and Han, LF *et al.* River networks system changes and its impact on storage and flood control capacity under rapid urbanization. *Hydrol. Process.* 2016; **30**: 2401–2412.
35. Chen, Q and Song, Z. Accounting for China's urbanization. *China Econ. Rev.* 2014; **30**: 485–494.
36. Roberts, GO and Rosenthal, JS. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* 2007; **44**: 458–475.
37. Zanardo, S, Zaliapin, I and Foufoula-Georgiou, E. Are American rivers Tokunaga self - similar? New results on fluvial network topology and its climatic dependence. *JGR* 2013; **118**: 166–183.

Abbreviation List

AI: Aridity Index.

D_{P}: Number of pseudo-rivers with catchment area ≥ 100 km^2 per unit area, $10^{-3}\cdot\mathrm{km}^{-2}$.

D_{Horton}: Number of rivers estimated by Horton stream-ordering scheme per unit area, 10^{-3} km^{-2}.

D_{1776}: Population density in Early Qing Dynasty (1776), person·km^{-2}.

$MCTI$: modified compound topographic index, a function of surface flow accumulation, aridity, and topographic slope.

N_{Census}: Number of rivers with catchment area ≥ 100 km^2 reported in National Census.

N_{Horton}: Number of rivers with catchment area ≥ 100 km^2 estimated by Horton stream-ordering scheme.

N'_{Horton}: Number of rivers with catchment area ≥ 100 km^2 estimated by Horton stream-ordering scheme after exclusion of pseudo-rivers.

N_{Named}: Number of named rivers with catchment area ≥ 100 km^2.

N_{P}: Number of pseudo-rivers with catchment area ≥ 100 km^2.

N_{Shreve}: Number of rivers with catchment area ≥ 100 km^2 estimated by Shreve stream-ordering scheme.

N'_{Shreve}: Number of rivers with catchment area ≥ 100 km^2 estimated by Shreve stream-ordering scheme after correction of the errors induced by pseudo-rivers.

: Number of rivers with catchment area ≥ 100 km^2 estimated by Strahler stream-ordering scheme.

N'_{Strahler}: Number of rivers with catchment area ≥ 100 km^2 estimated by Strahler stream-ordering scheme after correction of the errors induced by pseudo-rivers.

P_{f}: Proportions of three stream-ordering schemes (P_{HO} : P_{ST} : P_{SH}).

P_{HO}: Probability of occurrence of segments consistent with the Horton scheme.

P_{SH}: Probability of occurrence of segments consistent with the Shreve scheme.

P_{ST}: Probability of occurrence of segments consistent with the Strahler scheme.

WO: Water occurrence.

Figure Legends

Figure 1 Numbers of rivers in China with (a) different stream orders, and (b) different catchment areas estimated by Horton, Strahler, and Shreve stream-ordering scheme.

Figure 2 Pseudo-rivers in China. Distribution of: (a) Annual mean precipitation. (b) Aridity index in June, 2011. (c) Water occurrence in the period from 1984 to 2015. (d) Number of pseudo-rivers with catchment area ≥ 100 km^2 per unit area in each province. Here, China is divided into the following nine sub-regions: Northeast China (NE), North China (NC), Central China (CC), Southeast China (SE), South China (SC), Southwest China (SW), Northeast China (NE), Inner Mongolia (IM), and Tibet Plateau (TP).

Figure 3 Hybrid stream-order model for estimation of number of named rivers based on Monte Carlo method and graph theory.

Figure 4 Density of natural rivers and named rivers. Distribution of: (a) Modified compound topographic index (*MCTI*). (b) Natural river density (D_{Horton}). (c) Population density in early Qing Dynasty (D_{1776}). (d) Named river density (D_{Named}).

Figure 5 Interpretation of discrepancy in numbers of rivers ≥ 100 km^2 in terms of representative stream-segmentation modes (outer circles) and their relationship.

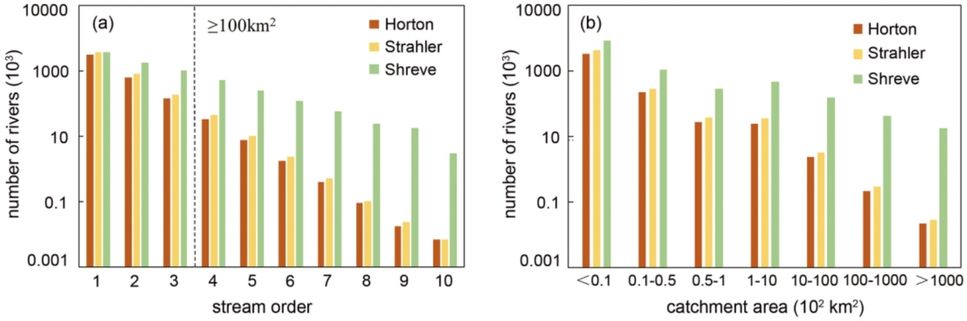


Figure 1 Numbers of rivers in China with (a) different stream orders, and (b) different catchment areas estimated by Horton, Strahler, and Shreve stream-ordering scheme.

139x49mm (300 x 300 DPI)

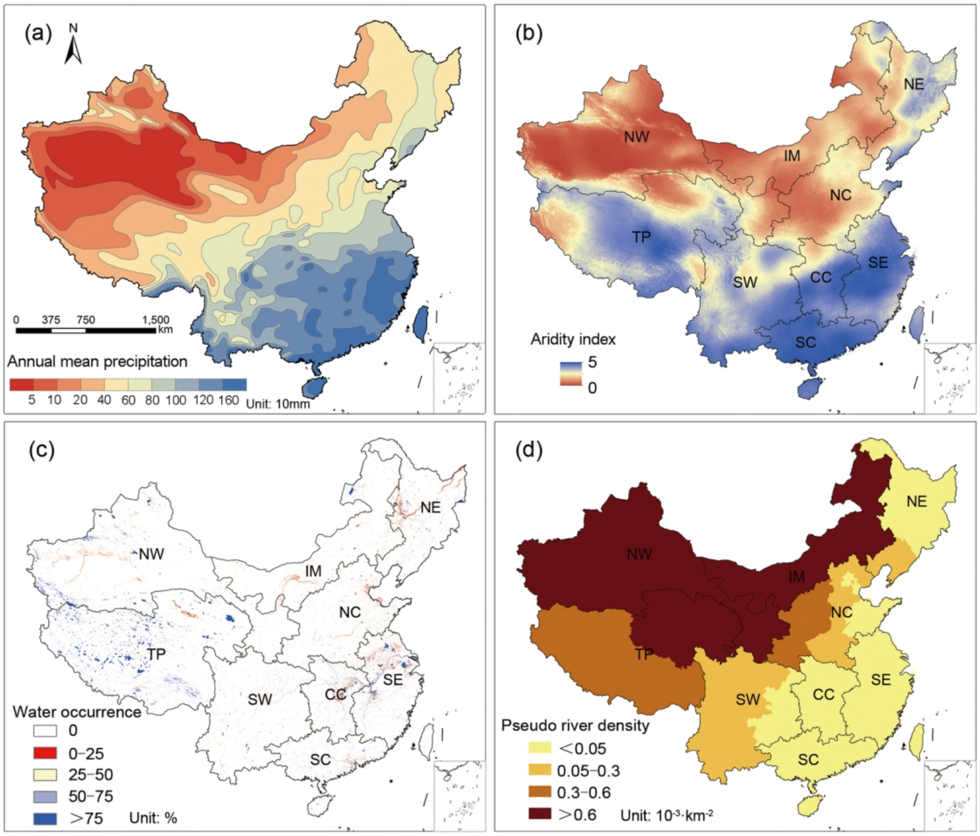


Figure 2 Pseudo-rivers in China. Distribution of: (a) Annual mean precipitation. (b) Aridity index in June, 2011. (c) Water occurrence in the period from 1984 to 2015. (d) Number of pseudo-rivers with catchment area ≥ 100 km^2 per unit area in each province. Here, China is divided into the following nine sub-regions: Northeast China (NE), North China (NC), Central China (CC), Southeast China (SE), South China (SC), Southwest China (SW), Northeast China (NE), Inner Mongolia (IM), and Tibet Plateau (TP).

140x119mm (300 x 300 DPI)

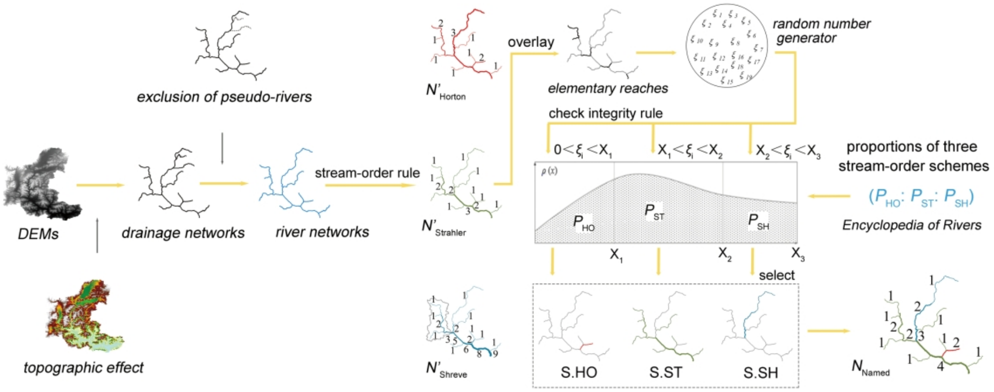


Figure 3 Hybrid stream-order model for estimation of number of named rivers based on Monte Carlo method and graph theory.

128x53mm (600 x 600 DPI)

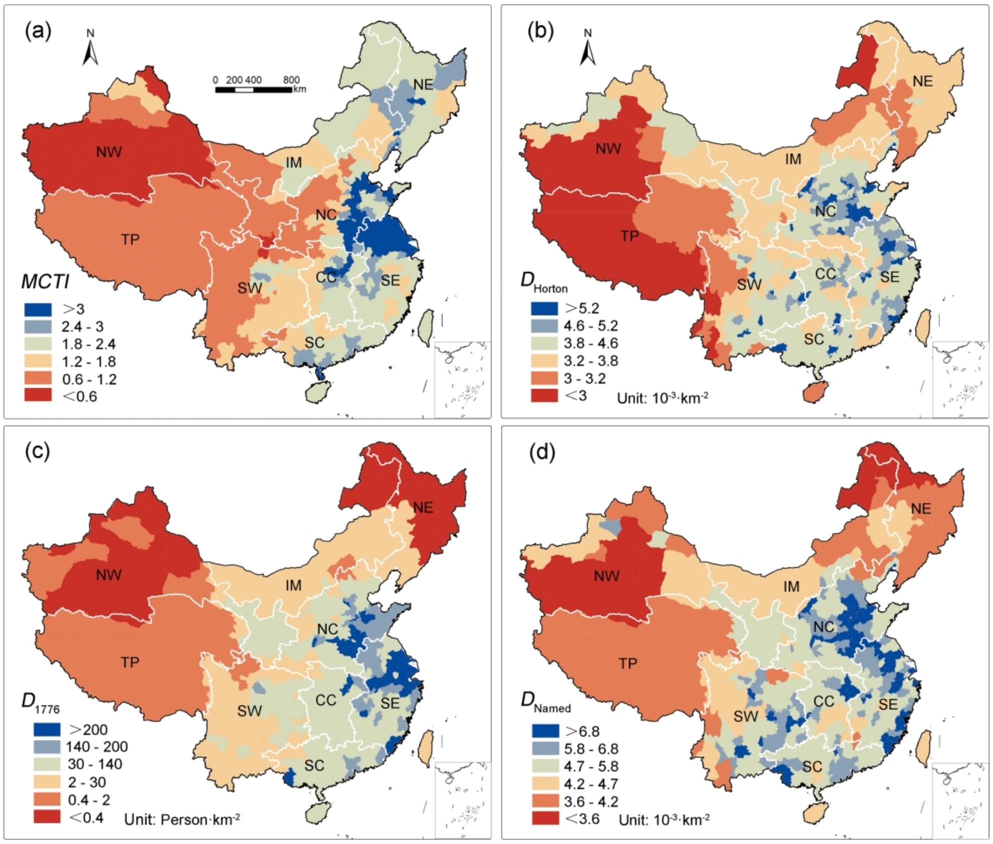


Figure 4 Density of natural rivers and named rivers. Distribution of: (a) Modified compound topographic index (*MCTI*). (b) Natural river density (D_{Horton}). (c) Population density in early Qing Dynasty (D_{1776}). (d) Named river density (D_{Named}).

140x119mm (600 x 600 DPI)

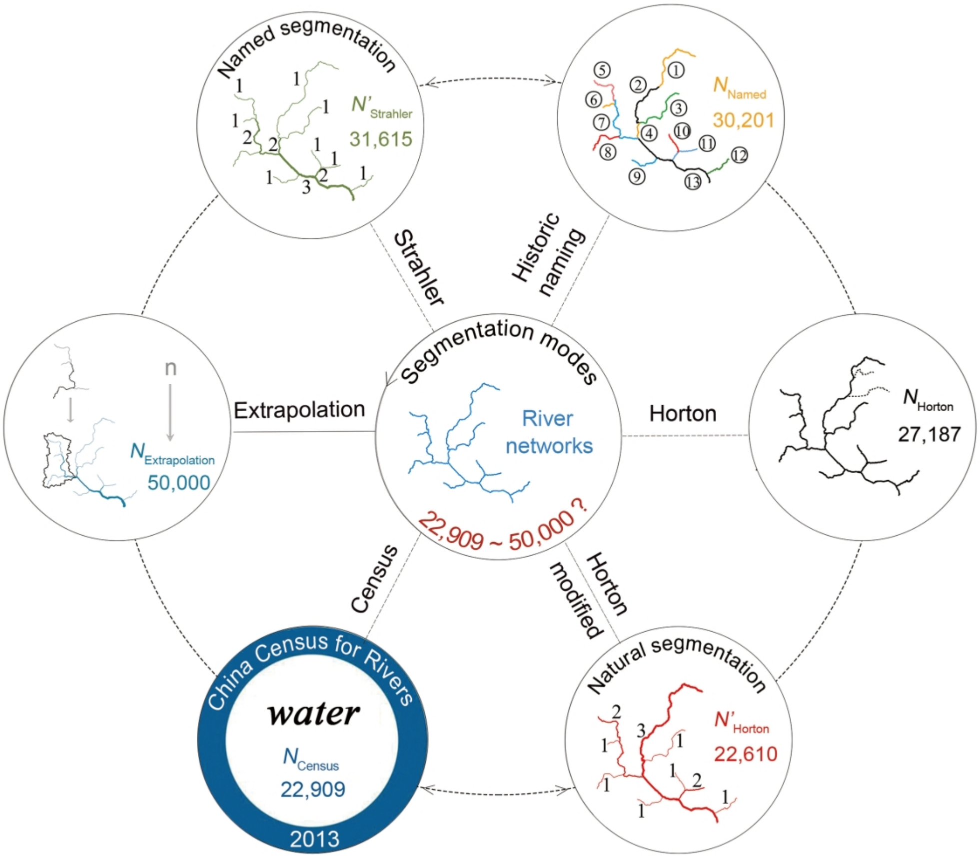


Figure 5 Interpretation of discrepancy in numbers of rivers ≥ 100 km^2 in terms of representative stream-segmentation modes (outer circles) and their relationship.

97x85mm (600 x 600 DPI)

Supplementary Data for

Solving the mystery of vanishing rivers in China

Yichu Wang, Jinren Ni, Yao Yue, Jiaye Li, Alistair G. L. Borthwick, Ximing Cai, An Xue,
Li Li, Guangqian Wang

Corresponding author: Jinren Ni

Email: jinrenni@pku.edu.cn

This PDF file includes:

Supplementary text
Supplementary Figs. 1 to 8
Supplementary Tables 1 to 11
References for SD reference citations
Complete Abbreviation List

1. Extraction of Drainage networks

We adopt Bai *et al*.'s (2015) [1] compound method to extract drainage networks. This computationally efficient method comprises five steps: I. pretreatment to fill local depressions; II. determination of flow directions; III. accumulation of upslope areas; IV. extraction of drainage channel; and V. vectorization of geographic features (Supplementary Fig. 1).

For interior rivers, pretreatment is undertaken. Lakes and deserts act as sinks for inland rivers. Lake sinks are identified by overlaying information from the global lake and wetland database (GLWD) [2] on the DEMs. The GLWD data relate to lakes of area larger than 0.1 km^2, and so only sinks of area ≥ 0.1 km^2 are identified. The remaining rivers are considered to end in the desert. For these basins we set a critical elevation value based on the Hydrosheds dataset [3], below which a region is a sink. After identifying the sinks, we use Arc info to erase the sink pixels. Thus, the sinks of inland rivers form the boundaries of the DEMs in a similar way to the sinks of exorheic rivers.

The extracted drainage networks were compared with the Hydrosheds Data [3]. And it appears that location of major rivers (stream order ≥ 6) is consistent with that derived from the Hydrosheds Data.

The hierarchical drainage networks which are extracted from 30 m resolution Aster GDEMs based on the above method for ten basins of China are illustrated in Supplementary Fig. 2.

2. River system hierarchy

Three stream-order schemes including Horton, Strahler, and Shreve stream-ordering schemes are used to describe the hierarchical river networks in this paper.

2.1 Horton stream-ordering scheme

In applying the Horton scheme [4], we first specify the maximum stream order, $n_{\max}$. The river corresponding to stream order $n = n_{\max}$ is the longest drainage path in the basin, and hence forms the backbone of the river network. The remainder of the network is then ordered into river bifurcations, at progressively lower orders until $n = 1$ is reached and the finest-scale rivers with shortest drainage paths leading from higher order river segments have been identified. Rivers of n-th stream order are identified by searching for the longest drainage paths that connect to the n+1-th order rivers. This procedure is repeated for n = n-1, n-2, ... 1. This method gives exactly the same results for river segmentation as the original Horton scheme even though its implementation is slightly different.

It should be noted that based on each stream ordering scheme, rivers can also be classified by catchment area. For example, in China's First National Census of Water [5], the number of rivers with catchment area ≥ 10,000 km^2, 1,000–10,000 km^2, 100–1,000 km^2, and 50–100 km^2 are estimated separately, based on the Horton scheme. And Supplementary Table 2 displays the number of rivers (and pseudo rivers) aggregated by catchment areas based on the Horton scheme identified from 30 m × 30 m DEMs. The number of rivers decreases as catchment area increases, for catchment area 10–50 km^2, to ≥ 100,000 km^2 the number reduces from 215,385 to 23.

2.2 Strahler stream-ordering scheme

In the Strahler scheme [6, 7], first-order rivers have headwaters without bifurcation. And when there is a confluence, the stream-order of downstream rivers obeys the following rules:

$$n_d = \begin{cases} n_1 + 1, & \text{for} \quad n_1 = n_2 \\ \max(n_1, \ n_2), & \text{for} \quad n_1 \neq n_2 \end{cases} \tag{1}$$

where n_d is the stream order of the downstream river, and n_1 and n_2 are the respective stream orders of the upstream rivers at the bifurcation.

2.3 Shreve stream-ordering scheme

In the Shreve scheme [8], the first-order rivers are headwaters without bifurcation, and form the branches at the finest level of the network. At a confluence, the stream order of the downstream river is the sum of the orders of the two upstream rivers.

3. Identification of pseudo-rivers

3.1 Pseudo-river identification hypothesis

We cannot ensure that channels identified via DEMs are in fact rivers because, on the one hand, the drainage networks extraction method is not totally suitable for the rivers in arid areas [9], and on the other hand, the topographic data do not contain information on local climate conditions, soil characteristics, vegetation cover, etc [10, 11]. To make direct comparison with number of rivers reported in China’s First National Census for Water [5], modification is needed by excluding pseudo-rivers from the extracted drainage networks. The most obvious difference between a real river and a pseudo-river is determined by the runoff condition in the channel. From the climate perspective, a basin is more capable of producing runoff when it has higher precipitation and lower evaporation [12]. Thus, the following pseudo-river identification hypothesis is proposed, based on common sense. A pseudo-river is defined as a channel that has zero runoff in the wet season and is located in a river basin whose Aridity index (AI) [13] is less than a prescribed threshold value.

Based on the standard dry-wet climate classification [14], the threshold value is taken to be 0.2, at the boundary between arid and semi-arid zones. When $AI = 0.2$, the ratio of actual evaporation to precipitation is equal to unity according to the Budyko Curve (Supplementary Fig. 3) [15-18]. This means all precipitation is converted into evaporation, and no runoff is generated, and is in accordance with findings from previous studies that a dry climate in semi-arid and arid zones promotes the development of pseudo-rivers [19, 20]. The proposed hypothesis is mainly suitable for rivers that are supplied by rainfall, and so for inland basins, international river basins, and the Yellow River basin where rivers are mainly supplied by glacial meltwater and groundwater [21-24], modification is carried out, based on the Water Occurrence (WO) [25] parameter which represents water dynamics from 1984 to 2015 expressed as a percentage of available observations when water is present. Specifically, we identify a channel as a pseudo-river when any element in the channel's series of water occurrence values between 1984 and 2015 is zero. Moreover, noting that large rivers with catchment area ≥ 1,000 km^2 always have stable runoff [26], we focus solely on rivers with catchment area < 1,000 km^2 during the identification process. Based on AI and WO, 4,577 pseudo-rivers with catchment areas ≥ 100 km^2 (N_{P}) within the main river basins are identified (Supplementary Table 3).

To eliminate the influence of pseudo-rivers for N_{Strahler} and N_{Shreve}, we use the ratio of pseudo-rivers (R_{P}) to revise the results:

$$R_{\mathrm{P}} = N_{\mathrm{P}}/N_{\mathrm{Horton}} \tag{2}$$

$$N'_{\mathrm{Strahler}} = N_{\mathrm{Strahler}} \times (1 - R_{\mathrm{P}}) \tag{3}$$

$$N'_{\mathrm{Shreve}} = N_{\mathrm{Shreve}} \times (1 - R_{\mathrm{P}}) \tag{4}$$

The revised number of rivers ≥ 100 km^2 estimated by Horton, Strahler, and Shreve stream-order scheme (N'_{Strahler}, N'_{Strahler}, N'_{Shreve}) for ten basins of China is displayed in Supplementary Table 1.

3.2 Constituent analysis of pseudo-rivers

Pseudo-rivers generally occur where the runoff simulation assumption is invalid, gullies, and dried-out rivers. Now we examine the contributions of these two types of pseudo-rivers in turn.

The runoff simulation assumption is often invalid in non-water erosion areas (Supplementary Fig. 4) where both precipitation and the underlying geomorphic condition are non-uniform [19]. We find the number of pseudo-rivers ≥ 100 km^2 distributed in non-water erosion areas (http://cese.pku.edu.cn/chinaerosion/) of China is 3,150, which is 69% of the total number of pseudo-rivers.

By subtracting the number of pseudo-rivers which generally occur where the runoff simulation assumption is invalid, and dried-out rivers from the total number of pseudo-rivers, we obtain the number of gullies and dried-out rivers ≥ 100 km^2 of China to be 1427, which is 31% of the total number of pseudo-rivers. Gullies and dried-out rivers are mainly located in Loess Plateau, Karst landform areas, in addition to Liao, Amur, and Hai River basin.

3.3 Error analysis and sensitivity analysis

By comparing the values of N_{Horton}' and N_{Census} for ten basins of China (Supplementary Table 1), we find the mean error range for N_{P} is ± 5.5%. This error partly arises from the -1%–5% error in the water occurrence dataset, allowing for 1% false water detections and 5% missing data [25]. So the total number of pseudo-rivers with catchment area ≥ 100 km^2 for the Yellow, inland, and international river basins ranges from 4,162 to 4,425. Meanwhile, there is an 8% error in the aridity index, so the total number of pseudo-rivers with catchment area ≥ 100 km^2 for the remaining seven basins ranges from 180 to 212. Summing these two parts together, the number of pseudo-rivers ≥ 100 km^2 is estimated to be $4{,}577\,^{+60}_{-235}$.

A sensitivity analysis is conducted to test whether 0.2 is reasonable for the threshold value of aridity index. In addition to the Census information and the original study, five further scenarios are considered, with threshold values of 0.1, 0.15, 0.2, 0.25, 0.3, respectively. It should be noted that Scenario 3 corresponds to the case in Section 3.1. The exclusion rates of pseudo rivers with catchment area ≥ 100 km^2 for Yangtze, Amur, Liao, Hai, Huai, and Pearl river basins, and rivers in Zhejiang and Fujian for Scenarios 1, 2, 3, 4, and 5 are 0.2%, 0.4%, 1.5%, 3.1%, and 5.9%, respectively. Thus, the pseudo-river exclusion rates for these basins are not sensitive to the threshold value of *AI.*

3.4 Extrapolation of results

Supplementary Table 3 also displays the number of pseudo-rivers aggregated according to catchment area. The number of pseudo-rivers decreases with basin area. Small rivers are in a more dynamic hydrologic condition because they are more sensitive to precipitation [26].

4. Incomplete topographic data

4.1 Discrepancy in numbers of rivers obtained from different topographic data

Moreover, the influences of incomplete topographic data and changes to land surface topography should also be fully considered in quantitative studies on river networks. Another two DEM datasets were used to investigate the effect of DEM resolution on numbers of rivers.

Supplementary Figure 5 displays the numbers of rivers of China aggregated by catchment areas extracted from 900 m × 900 m (N_{900}), 90 m × 90 m (N_{90}), and 30 m × 30 m DEMs (N_{30}). We find for large rivers (catchment area ≥ 1000 km^2), N_{900}, N_{90}, and N_{30} are similar. The relative difference between N_{90} and N_{30} maintain at the values of about 5% until the catchment areas reduces to 50 km^2. For rivers with catchment area < 50 km^2, the relative difference between N_{90} and N_{30} reaches 10% (Supplementary Table 4). However, the low-resolution topographic data in the 1990s can only identify large rivers, and fails fully to identify small rivers (catchment area < 1000 km^2). For small rives with catchment area < 1000 km^2, the relative difference between N_{900} and N_{30} is as much as 30% (Supplementary Table 5). Overall, it appears that the 90 m × 90 m DEMs in the 2000s can identify rivers with catchment area ≥ 50 km^2, while the low-resolution topographic data of the 1990s (900 m × 900 m DEMs) are only suitable for identification of large rivers with catchment area ≥ 1000 km^2.

One approach to calculating the number of rivers ≥ 100 km^2 in the 1990s would be to extrapolate from the number of large rivers. Here, we calculate the number of rivers of area < 1,000 km^2 in the 1990s by extrapolation from the number of rivers ≥ 1,000 km^2. For a Horton-Strahler self-similar drainage network, we calculate the number of rivers at each order via R_B [27]. Here we assume that when a Horton scheme transforms into a catchment area scheme, the drainage network is also self-similar, and there is a pseudo-bifurcation ratio (called R_B') between rivers with different catchment area grades. Here it is not possible to obtain the pseudo-bifurcation ratio using the least-squares method [4], due to incomplete identification of small rivers. Here we adopt another method to approximately calculate the pseudo-bifurcation ratio (R_B') [28] for the ten basins of China, defined as:

$$R_B' = N_{900(1000\text{-}10000)} / N_{900(10000\text{-}100000)} \tag{5}$$

$$R_B'' = N_{30(50\text{-}100)} / N_{30(100\text{-}1000)} \tag{6}$$

Thus,

$$N_{900(100\text{-}1000)} = N_{900(1000\text{-}10000)} \times R_B' \tag{7}$$

$$N_{900(10\text{-}100)} = N_{900(1000\text{-}10000)} \times (R_B')^2 \tag{8}$$

$$N_{900(50\text{-}100)} = N_{900(1000\text{-}10000)} \times (R_B') \times (R_B'') \tag{9}$$

$$N_{900(10\text{-}50)} = N_{900(1000\text{-}10000)} \times (R_B')^2 - N_{900(1000\text{-}10000)} \times (R_B') \times (R_B'') \tag{10}$$

$$N_T = N_{900(100\text{-}1000)} + N_{900(1000\text{-}10000)} + N_{900(10000\text{-}100000)} + N_{900(\geq 100000)} \tag{11}$$

$N_{900(\geq 100000)}$, $N_{900(10000\text{-}100000)}$, $N_{900(1000\text{-}10000)}$, $N_{900(100\text{-}1000)}$, $N_{900(50\text{-}100)}$, $N_{900(10\text{-}50)}$, $N_{900(10\text{-}100)}$ represent the numbers of rivers with catchment areas ≥ 100,000 km^2, 10,000–100,000 km^2, 1,000–10,000 km^2, 100–1000 km^2, 50–100 km^2, 10–50 km^2, and 10–100 km^2 extracted from 900 m × 900 m DEMs, respectively. $N_{30(50\text{-}100)}$, $N_{30(100\text{-}1000)}$ are numbers of rivers with catchment areas 50–100 km^2, 100–1000 km^2. N_T is the number of rivers with catchment area ≥ 100 km^2 extrapolated from 900 m × 900 m DEMs.

Based on equation (5), (6), the value for R_B', and R_B'' of China is 10.9, and 1.1, respectively. By inserting $N_{900(\geq 100{,}000)}$, $N_{900(10{,}000\text{ - }100{,}000)}$, $N_{900(1{,}000\text{ - }10{,}000)}$ for ten basins of China, R_B', and R_B'' into equations (7) – (11), we obtain values for estimated N_{900} with different catchment areas, and N_T (Supplementary Table 6).

The resulting number of rivers ≥ 100 km^2 extrapolated from low-resolution topographic data (N_T) is $28{,}250^{+7297}_{-13418}$, about 1,063 more than the result based on 30 m × 30 m DEM (N_Horton). Previous research has suggested that use of a low-resolution DEM causes overestimates of river catchment areas [29], providing further evidence for the overestimate of the number of small rivers after extrapolation.

4.2 Error Analysis

Noting that R_B' has a range of 5.2 ~ 14.0, then N_T also has a corresponding range. Inserting both $R_\mathrm{B}' = 4.2$, and $R_\mathrm{B}' = 14.0$ into equation (7) – (10), we obtain the minimum value, and maximum value of N_T of China to be 14832, and 35547, respectively (Supplementary Table 6).

5. Changes to land surface topography

5.1 Discrepancy in the number of rivers due to change in topography

During the recent two decades, channel geometries and river network topologies have altered due to intensive urbanization and rapid land-use changes in China, which may also cause the extraction of river networks to vary [30, 31]. Here we used GTOPO30DEMs

(900 m × 900 m, launched in 1996) and AsterGDEMs (30 m × 30 m, launched in 2011) to represent China's land surface topography in the 1990s and 2010s, respectively. During this period, China experienced intensive urbanization and land-use changes [32]. Drainage networks extracted from each DEMs are compared to explore the effect of network alteration caused by topographic changes on the number of rivers. To minimize the effect of DEMs resolution, we re-sampled 30 m AsterGDEMs at 900 m. The re-sampled DEMs is called 900 m DEMs' herein.

Let N_{900} and N_{900}' be the numbers of rivers with catchment area ≥ 100 km^2 identified from the 900 m DEM and 900 m DEMs'. Supplementary Table 7 lists values of N_{900}' with different catchment areas. Comparing N_{900} with N_{900}', we find relative difference between N_{900}' and N_{900} increase from -0.2 % to -14% when the catchment area of rivers decreases from 1,000-100,000 km^2 to 10-50 km^2. It means the topographic change mainly cause variation of number of small rivers instead of large rivers.

Here we assume that both the rivers identified by 900 m DEMs' and those not identified by 900 m DEMs' contain the same percentage of migrated rivers (P_M). Thus, we approximately estimated the difference in numbers of rivers with catchment area ≥ 100 km^2 due to topographic change (N_M):

$$N_\mathrm{M} = N_{900}' - N_{900} + (N_{30} - N_{900}') \times P_\mathrm{M} \tag{12}$$

$$P_\mathrm{M} = (N_{900}' - N_{900}) / N_{900} \tag{13}$$

P_M is the percentage of migrated rivers (P_M).

Based on equation (12), (13), we also obtain the difference in numbers of rivers with catchment areas ≥ 100 km^2 for ten river basins of China (Supplementary Table 8).

It appears that changes in land topography from 1996 to 2011 caused the number of rivers with catchment area ≥ 100 km² (N_{M}) to decrease in by $1{,}953^{+623}_{-958}$. The difference in numbers of rivers due to topographic change reduces across eastern and central China except for a slight increase in the Hai River and Huai River Basin (Supplementary Table 1) owing to the construction of artificial channels [33]. The topographic changes in elevation during 1996–2010 are concentrated in the Hai, Huai, Pearl River basin, rivers in Zhejiang and Fujian, and estuary regions of the Yangtze River and Yellow River basins (Supplementary Fig. 6) which comprise relatively flat (Supplementary Table 8 lists mean slopes), urbanized areas.

5.2 Error Analysis

Since P_{M} has a range of -4.3% – 13.7% (Supplementary Table 8), then N_{M} also has a corresponding range. Inserting both $P_{\mathrm{M}} = -4.3\%$ and $P_{\mathrm{M}} = 13.7\%$ into equation (12), gives $N_{\mathrm{M}} = 1,953^{+623}_{-958}$.

6. Random river naming

6.1 Hybrid Stream-Order Scheme based on Monte Carlo method

Drainage networks are split into segments following prescribed rules when using a river-ordering scheme. Based on numerous comparisons of the named river systems (*Encyclopedia of Rivers and Lakes in China* [34]) and river networks generated using the three different stream-order rules for various-scale basins in China, we found the historical river naming practice was best replicated by a mix of Horton, Strahler, and Shreve ordering schemes. This implies the number of rivers in a named river system can be simulated using different stream-order schemes in different regions. However, no information can be gleaned from historical records about the spatial distribution of river reaches determined by each of the three schemes. Consequently, the number of historically named rivers in China can only be assessed by combining information about the probabilities of occurrence of usage of the three stream-order schemes, which is approximated by recorded frequencies of occurrence in the sample watersheds in the *Encyclopedia of Rivers and Lakes in China* [34]. By intersecting different layers of river networks according to the stream-order schemes of Horton, Strahler, and Shreve, a set of elemental river reaches is generated. Then, assessments by the three stream-order schemes are distributed according to their occurrence probabilities to each of the river reaches, and the total number of river reaches estimated using the hybrid stream-order model.

To avoid the subjective error that may be incurred by manually assigning a stream-order scheme to a given element reach, the Monte Carlo method [35] is implemented to simulate the random choice of stream-order scheme. This method also enables scenario analysis by assigning different combinations of occurrence probabilities to the three stream-order schemes. The core idea behind the Monte Carlo method is to simulate a random process according to an appropriate statistical model, and then to approximate the true solution by a suitable estimate. Fig. 3 displays the framework of this naming system. The procedure underpinning the naming system is as follows:

(I) Estimate probabilities of occurrence of Shreve, Strahler, and Horton river reaches (i.e. P_{HO}, P_{ST}, P_{SH}) by analyzing occurrence frequencies of the three stream-order schemes in the sample watersheds considered in the *Encyclopedia of Rivers and Lakes in China* [34].

(II) Select an appropriate probability density function, $f(x)$.

(III) Set up correspondence between the three stream-order schemes and different intervals of *f*(*x*), which meets the following requirements:

$$\begin{cases} P_{\mathrm{H0}} = \int_{0}^{\mathrm{X}_1} f(x)\mathrm{d}x \\ P_{\mathrm{ST}} = \int_{\mathrm{X}_1}^{\mathrm{X}_2} f(x)\mathrm{d}x \\ \mathrm{P}_{\mathrm{SH}} = \int_{\mathrm{X}_2}^{\mathrm{X}_3} f(x)\mathrm{d}x \end{cases} \tag{14}$$

Suppose that *I(x)* is the integral function of *f(*x*)*:

$$I(x) = \int_{0}^{\mathrm{x}} f(x)\mathrm{dx} \tag{15}$$

Then, X_1, X_2, and X_3, which define the intervals of *f*(*x*), can be determined as:

$$\begin{cases} \mathrm{X}_1 = I(\mathrm{X}_1) \\ \mathrm{X}_2 = I(\mathrm{X}_2) - \mathrm{X}_1 \\ \mathrm{X}_3 = I(\mathrm{X}_2) - \mathrm{X}_2 \end{cases} \tag{16}$$

(IV) Set up a pseudo-random number generator to simulate a uniformly distributed random variable, ξ, within the range from 0 to X_3. The naming scheme is as follows:

When $0 < \xi \leq \mathrm{X}_1$, the element reach of interest is selected as a Horton river;

When $\mathrm{X}_1 < \xi \leq \mathrm{X}_2$, the reach is classified as a Strahler river; and

When $\mathrm{X}_2 < \xi \leq \mathrm{X}_3$, the reach is chosen as a Shreve river.

Integrate *f*(x) as:

$$P_{\mathrm{i}} = \int_{\mathrm{Xj-1}}^{\mathrm{Xj}} f(x)\mathrm{d}x, \ (i = \mathrm{Horton}, \mathrm{Strahler}, \mathrm{Shreve}; j = 1, 2, 3; \mathrm{X}_0 = 0) \tag{17}$$

Define another random variable, $\eta = f(\xi)$. Then,

$$P_{\mathrm{in_i}} = \frac{\Sigma_1^{n_i}\eta_{\mathrm{j}}}{n_{\mathrm{i}}} = \frac{\Sigma_1^{n_{\mathrm{i}}}f(\xi_{\mathrm{j}})}{n_{\mathrm{i}}} \tag{18}$$

Hence,

$$P_{in_i} = E(\eta) = \lim_{n_{\mathrm{i}}\to\infty} P_{in_i} \tag{19}$$

Which implies that $P_{\mathrm{in_i}}$ is the unbiased estimator of P_{i} when n_{i} is sufficiently large.

(V) Assign IDs to each of the integrated river units in terms of the three stream-order schemes, which implies each elemental river reach has three IDs (namely, ID_{Horton}, ID_{Strahler}, and ID_{Shreve}) derived from overlay of the three river networks.

(VI) Generate integrity information and marked information for each of the elemental river reaches.

(VII) Randomly assign a stream-order scheme according to IV, and check if the assignment is valid according to the flags in the marked system. For valid assignment, simply change the marked information of other elements in the same basic unit of rivers to that of the assigned scheme.

(VIII) By defining the function, $ID_{(\mathrm{ni})}$, to return the total number of distinct IDs (i = Horton, Strahler, and Shreve, as in Step (IV)), the Monte Carlo estimate of the total number of river reaches, N_{MC}, is as follows:

$$N_{\mathrm{MC}} = \Sigma_{\mathtt{Horton,\ Strahler,\ Shreve}} ID(n_{\mathrm{i}}) \tag{20}$$

(IX) Repeat Steps (IV) and (VIII) M times, and we can obtain the series of N_{MC}s: $N_{\mathrm{MC\text{-}1}}$, $N_{\mathrm{MC\text{-}2}}$, ……, $N_{\mathrm{MC\text{-}M}}$. The mean N_{MC} is:

$$\overline{N_{\mathrm{MC}}} = \frac{\Sigma_1^{\mathrm{M}} N_{\mathrm{MC-j}}}{\mathrm{M}} \tag{21}$$

According to the law of large numbers, the actual number of river reaches, n_{M}, is:

$$n_{\mathrm{M}} = \lim_{\mathrm{M}\to\infty} \overline{N_{\mathrm{MC}}} \tag{22}$$

Which means that the mean Monte Carlo estimator approaches n_{M} for sufficiently large numbers of simulations.

6.2 Proportions of the three ordering schemes in the named river systems

When simulating the named river system using the Monte Carlo method, the first step is to obtain the proportions of the three stream-ordering schemes (P_{f}). To quantify suitable values for P_{f} in the named river system, we collect data on 107 representative basins whose catchment areas range from 2,690 km^2 to 137,633 km^2 (Supplementary Fig. 7) from *Encyclopedia of Rivers and Lakes in China* [34]. For each basin, best-fit segment proportions are determined for the three stream-ordering schemes. The procedure for determining P_{f} for each basin is as follows:

(1) Vectorize the basin's drainage network map, obtained from the *Encyclopedia of Rivers and Lakes in China*.

(2) Create a series of elemental reaches for the river network by splitting the rivers at their vertices; the total number of elemental reaches in the series is n_{Total}.

(3) Name each elemental reach according to information extracted from the *Encyclopedia of Rivers and Lakes in China*.

(4) Compare the resulting named river system with the river network defined by the Strahler stream-ordering scheme. Assign a value of 2 to named elementary reaches that are consistent with the Strahler scheme.

(5) Consider those named reaches that are not consistent with Strahler. Then compare these reaches with results from the Shreve and Horton stream-ordering schemes, and assign a value of 1 to those consistent with Horton, and 3 to those consistent with Shreve. It should be noted that if a reach is assigned the value 1, its downstream reaches with the same Horton order should also be assigned the value 1 (according to the Horton ordering rule, and overwriting any previously assigned values).

(6) Total numbers of reaches assigned values 1, 2, and 3, are denoted $n_{\mathrm{HO},}$ n_{ST}, and n_{SH}.

(7) Segment proportions for the Horton (P_{HO}), Strahler (P_{ST}), and Shreve (P_{SH}) schemes are evaluated from $n_{\mathrm{HO}} / n_{\mathrm{Total}}$, $n_{\mathrm{ST}} / n_{\mathrm{Total}}$, and $n_{\mathrm{SH}} / n_{\mathrm{Total}}$.

The reason why we first compare the named river system with rivers defined by Strahler ordering scheme in Step (4) is that the Strahler scheme gives results closest to those from traditional named river segmentation.

The use of 107 sub-basins to represent the drainage behavior of their parent basins is acceptable given that: (i) they cover 26% of China's total land area; (ii) they cover most geomorphic types in China including loess, fluvial landform, karst, and desert; (iii) their surface morphology ranges from flat plain to very steep mountains; and (iv) their aridity indices range from semi-arid to hyper-humid.

6.3 Method validation

After determining P_f for 107 representative basins, we compute the number of named rivers using random simulation based on the Monte Carlo method for each basin (n_{Mixed}) in order to validate the model. We denote the relative difference between n_{Mixed} and n_{Named} for the 107 representative basins as *Dif*, There are 31, 42, and 34 basins with values of *Dif* equal to zero, > zero, and < zero, respectively, which implies the error is random. Meanwhile, 93 basins have 87% of their *Dif* values lying between -10% and 10%, which confirms the model has satisfactory accuracy.

6.4 Correlation between river naming and population density

The naming of rivers is essentially a human activity, and so we assume the river naming process is also associated with population density. To confirm the validity of this assumption, a correlation analysis would normally be required. However, the proportions of the three schemes (P_f) which represent the naming orientation of rivers cannot be directly fitted to a correlation analysis. In such cases, we replace P_f with $N\mathrm{P}_f$.

We consider 66 scenarios with different proportions of the Horton, Strahler, and Shreve river-ordering schemes, and compute the total number of rivers N_{Mixed} of China for each scenario, using random simulation based on the Monte Carlo method (Fig. 3). Supplementary Table 9 is a ranked list in order of the least value to the highest value of N_{Mixed}, in which each scenario is signed a number. Here, N_{Mixed} is quite insensitive to the relative proportions of Horton and Strahler, when the Shreve proportion is small (NP_f 1 to 20). N_{Mixed} is highly sensitive to the proportion of the Shreve stream-ordering scheme, and increases substantially for Scenarios 56–66. It is interesting to note that when the drainage network is composed of 10 % rivers defined by Horton, 20% rivers defined by Strahler, and 70 % rivers defined by Shreve, the total value of N_{Mixed} for China reaches ~ 50,000 (the textbook value at the center of the recent controversy). Altering the mix further, it is possible to obtain 669,946 rivers when Shreve is 1, and Horton and Strahler are both zero.

The value of $N\mathrm{P}_{\mathrm{f}}$ increases with N_{Mixed}. Since N_{Mixed} is the number of named rivers, a high value of NP_{f} corresponds to a high frequency of naming of rivers. A correlation has been determined between the Scenario $N\mathrm{P}_{\mathrm{f}}$ and the population density in Qing Dynasty (1776) [36] for 107 representative basins. The analysis shows that Scenario $N\mathrm{P}_{\mathrm{f}}$ and the population density in Qing Dynasty (1776) are positively correlated ($R = 0.4$, $P = 0.01$), confirming that the assumption is reasonable.

6.5 The number of named rivers in China

We separately compute the local N_{Named} of each of the four population modes using the corresponding proportions of the three ordering schemes (Supplementary Table 10), then sum the values together. The local N_{Named} values for basins of type I, II, III, IV, V, and VI are 6453, 8697, 9155, 8594, 1811, and 1605, respectively (Supplementary Fig. 8). In this way, we obtain N_{Named} for China to be $36{,}315^{+1562}_{-1428}$. After eliminating the influence of pseudo-rivers for N_{Named} by the ratio of pseudo-rivers (R_{P}), the N_{Named} for China ultimately turns out to be $30{,}201^{+1,299}_{-1,187}$.

Even though the main focus here is on rivers in China with catchment area ≥ 100 km^2, the hybrid stream-order model is capable of estimating numbers of named rivers for varying catchment areas. Supplementary Table 11 lists the extrapolated results for the number of named rivers aggregated by catchment areas obtained from the six typical P_{f}. In China, the number of named rivers decreases with basin area, whereas the error range related to the number of named rivers increases with catchment area. Low-order rivers which have higher numbers are more likely to lead to an accurate result during random simulation.

Data sources

Bulletin of First National Census for Water [37]. The bulletin separately lists the numbers of rivers with catchment areas ≥ 10,000 km^2, ≥ 1,000 km^2, ≥ 100 km^2, and ≥ 50 km^2; the values for areas ≥ 100 km^2 are taken as reference.

DEMs. Three DEMs data sets are utilized in this study. First, the drainage networks of China are extracted from the 30 m resolution Aster GDEM (available at http://reverb.echo.nasa.gov/reverb/), which provides the highest resolution data freely available at global scale. These data were launched in the same year as China's First National Census for Water, and so any disparity caused by a time gap is minimized. Secondly, 90m resolution SRTM DEMs (available at http://www.gscloud.cn/), lunched in 2003, is used to represent the terrain in 2000s. Thirdly, the 900 m resolution GTOPO 30 DEMs (available at http://earthexplorer.usgs.gov/), launched in 1996, is used to represent the terrain in 1990s.

Vector data of streams and lakes. Three sets of vector stream and lake data are utilized. National Geomatical Data (available at http://www.ngcc.cn/) at 1:4 million resolution provide burned streams in the DEM preprocessing procedure [38]. Hydrosheds Data [3] (available at http://hydrosheds.cr.usgs.gov) are used to validate the extracted drainage networks. The Global Lakes and Wetlands Database [2] (GLWD) (available at http://www.worldwildlife.org) is used to identify boundaries of sinks when extracting interior rivers.

Annual mean precipitation of China. Annual mean precipitation map of China is obtained from the *Atlas of Physical geography of China* [39].

Aridity Index. The aridity index [13] is a function of precipitation and potential evapotranspiration, and is defined as:

$$AI= P/PET^0 \tag{23}$$

where P is mean annual precipitation; and PET^0 is the mean annual potential evapotranspiration. The aridity index is widely used as an indicator of dry-wet climate. We adopt the climate classification scheme for aridity index values proposed by the Meteorological Standards of China (see Supplementary Information). Two sets of aridity index data are utilized: the annual average aridity index for 1950-2000 (available at https://perswww.kuleuven.be/~u0055544/aridity/Global_Aridity_PET_Methodolgy.htm); and the aridity index for June, 2011 (available at http://www.dsac.cn/DataProduct/Detail/201004). The resolution of both data sets is 1 km. The first data set is used in estimating the number of dried-out rivers, and the second is used in identifying which pseudo rivers to exclude.

Water Occurrence. The Water Occurrence dataset [25] is accessible via http://global-surface-water.appspot.com/, and presents information on the spatial distribution of surface water at 30 m resolution over the Earth's surface in the period from 1984 to 2015. The information is given as percentage occurrence frequency, previously treated to ensure temporal consistency.

Encyclopedia of Rivers and Lakes in China [34]. Information of representative named river networks of China is obtained from *Encyclopedia of Rivers and Lakes in China*. This encyclopedia lists rivers of area ≥ 1,000 km^2 and provides examples of the traditional approach to naming rivers in China.

Modified Compound Topographic Index (MCTI). The *MCTI* is modified from *CTI*, which is a function of slope and upstream flow accumulation, and is defined as:

$$CTI = \ln\left(\alpha/\tan\left(\pi S/180\right)\right) \tag{24}$$

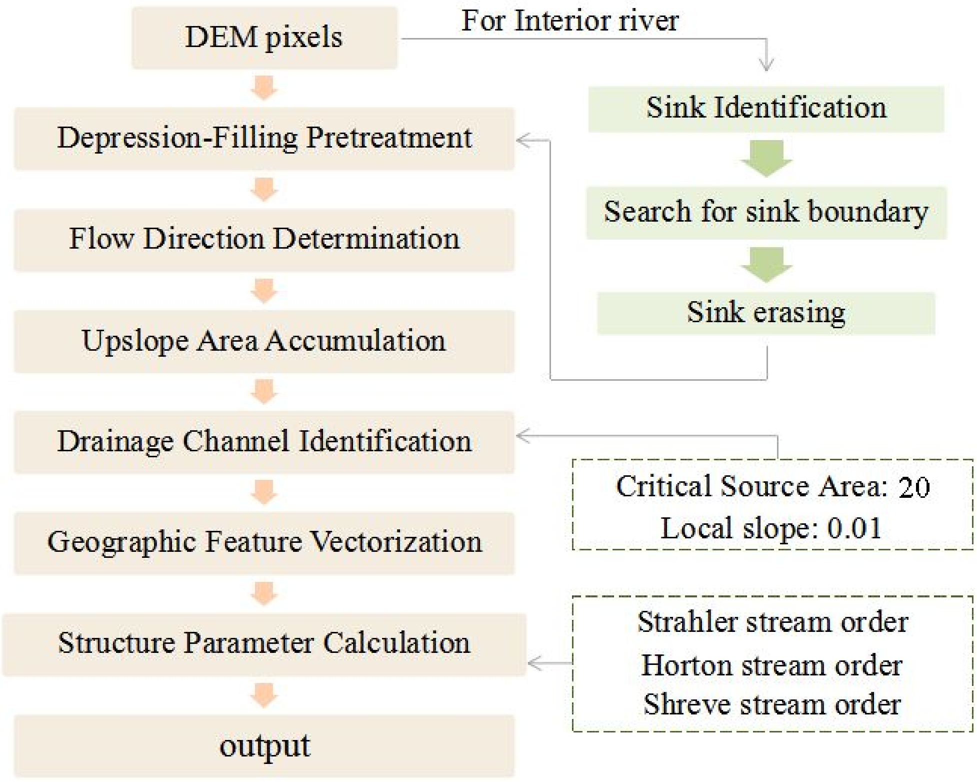
where α is the accumulated flow; and S is the slope (°). *CTI* is obtained from DEMs, and partly reflects the hydrologic characteristics of a basin [40]. However, because both α and S are computed from DEMs, *CTI* does not account for climate conditions, and so is deficient (considering that river networks are products of both climate and geology). Here, we modify traditional *CTI* by replacing α with the annual average aridity index for 1950-2000 which describes the climatic character of a basin. Thus, *MCTI* is defined as:

$$MCTI = \ln\left(AI/\tan\left(\pi S/180\right)\right) \tag{25}$$

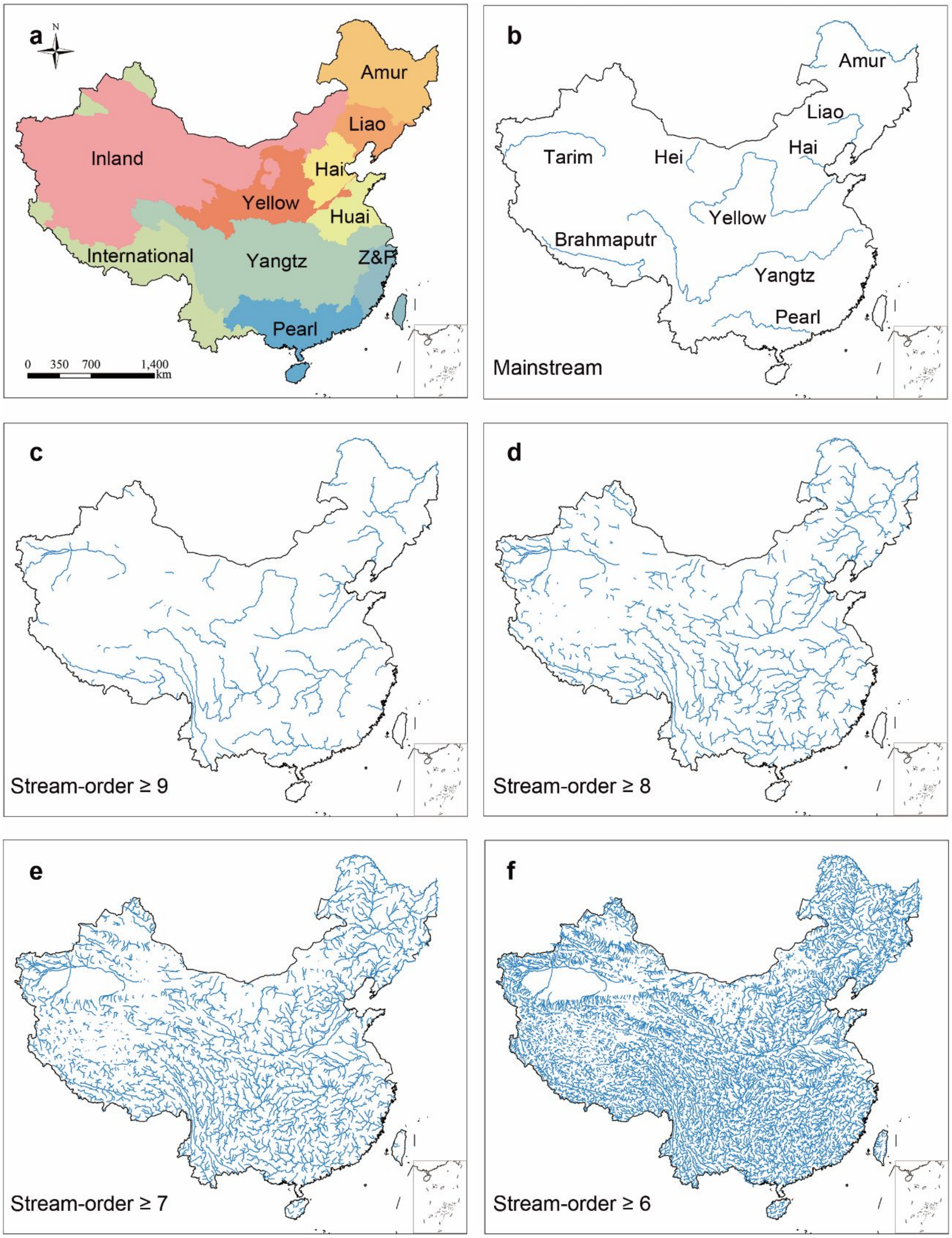
Population density data. Two sets of population data are used to calculate population density. Population data for the Qing Dynasty are obtained from the *History of Population in China (Volume 5), Qing Dynasty* [36]. Population data in 2014 are taken from the *China Statistical Yearbook of 2014* [41].

Data availability

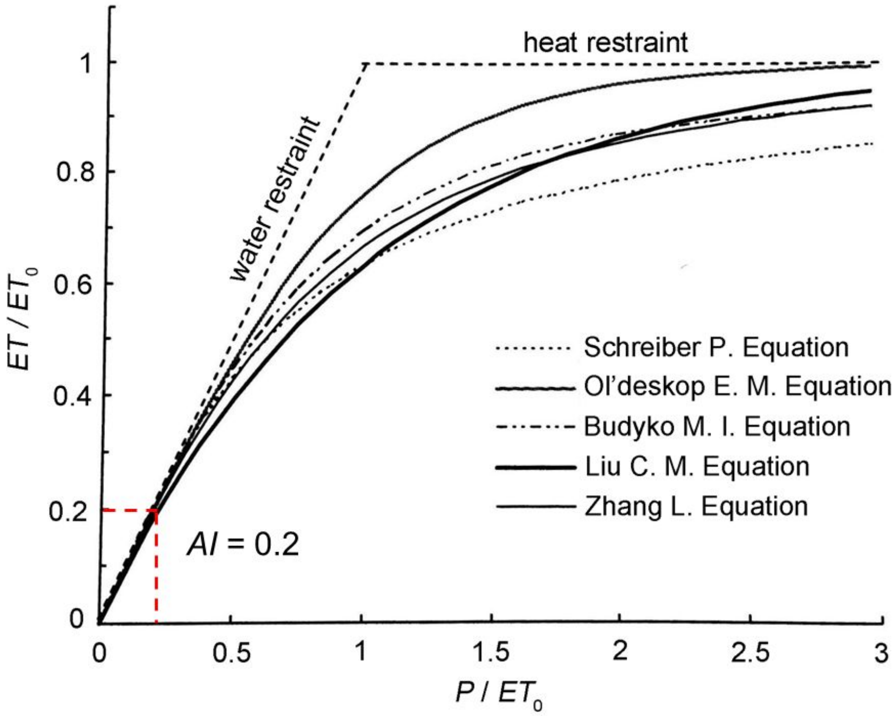
The data that support the findings of this study are available from the corresponding author upon reasonable request.



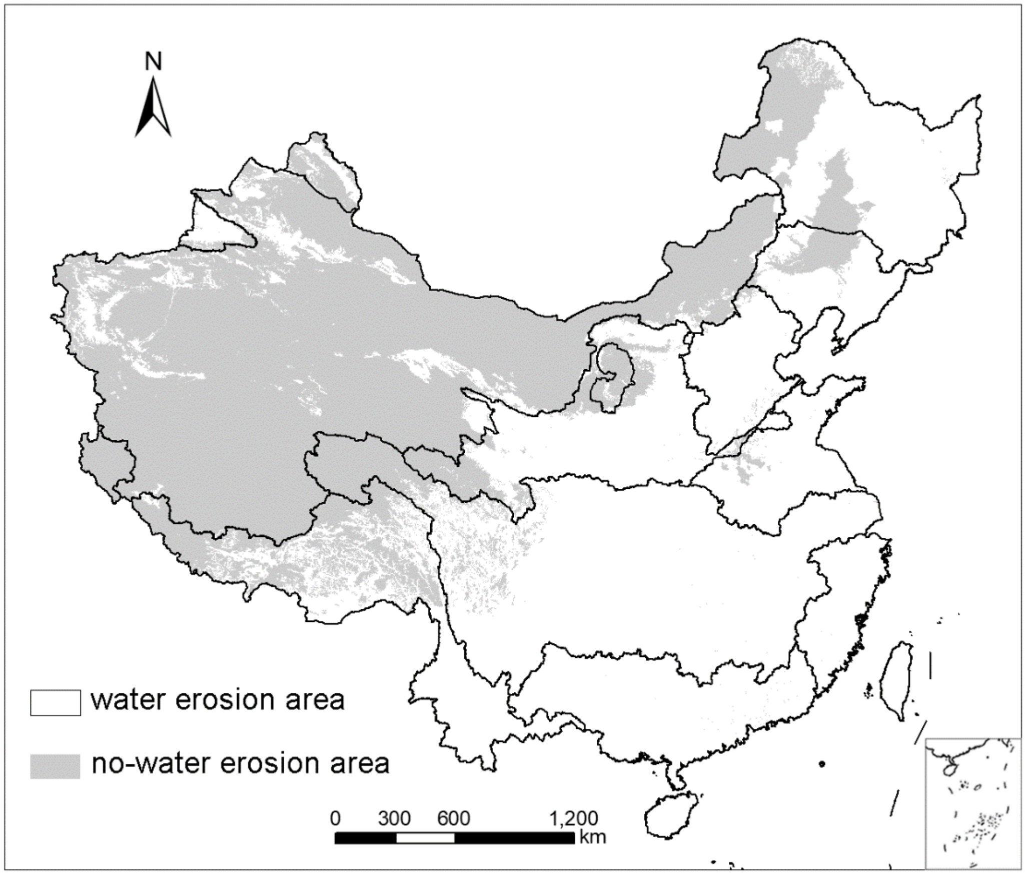
Supplementary Fig. 1. Flow chart showing procedure for drainage network extraction.



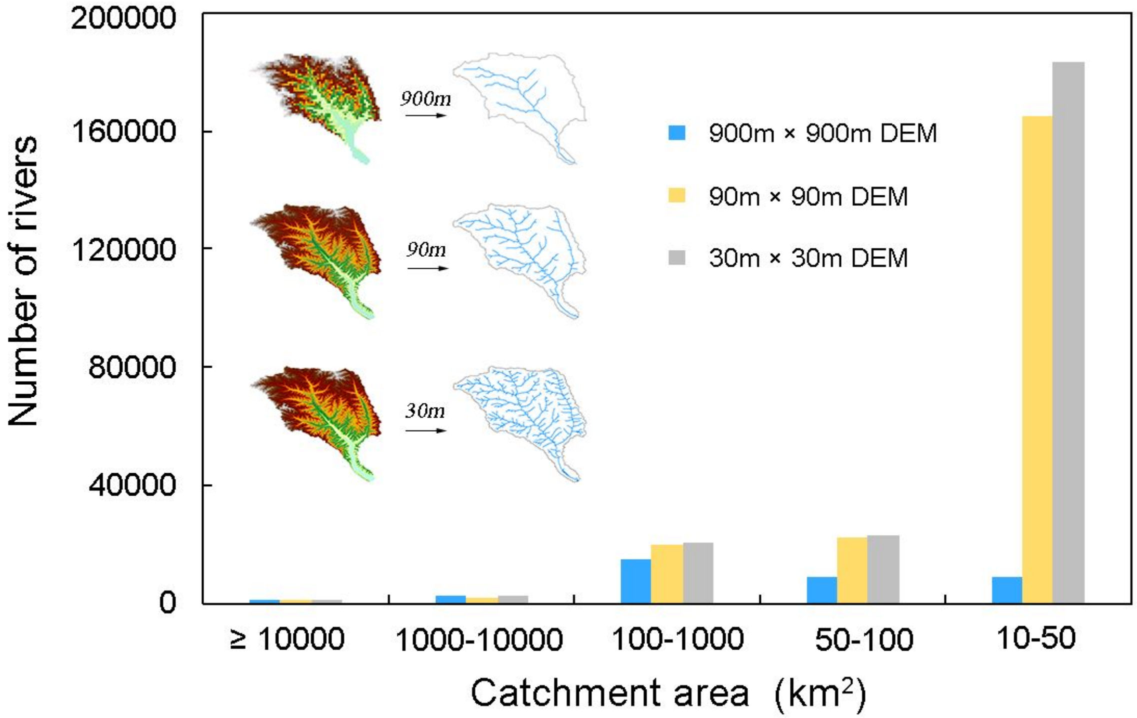
Supplementary Fig. 2. Drainage networks of China. (a) Division into ten basins. (b) Mainstreams of largest rivers. (c) Rivers of stream-order ≥ 9. (d) Rivers of stream-order ≥ 8. (e) Rivers of stream-order ≥ 7. (f) Rivers of stream-order ≥ 6.



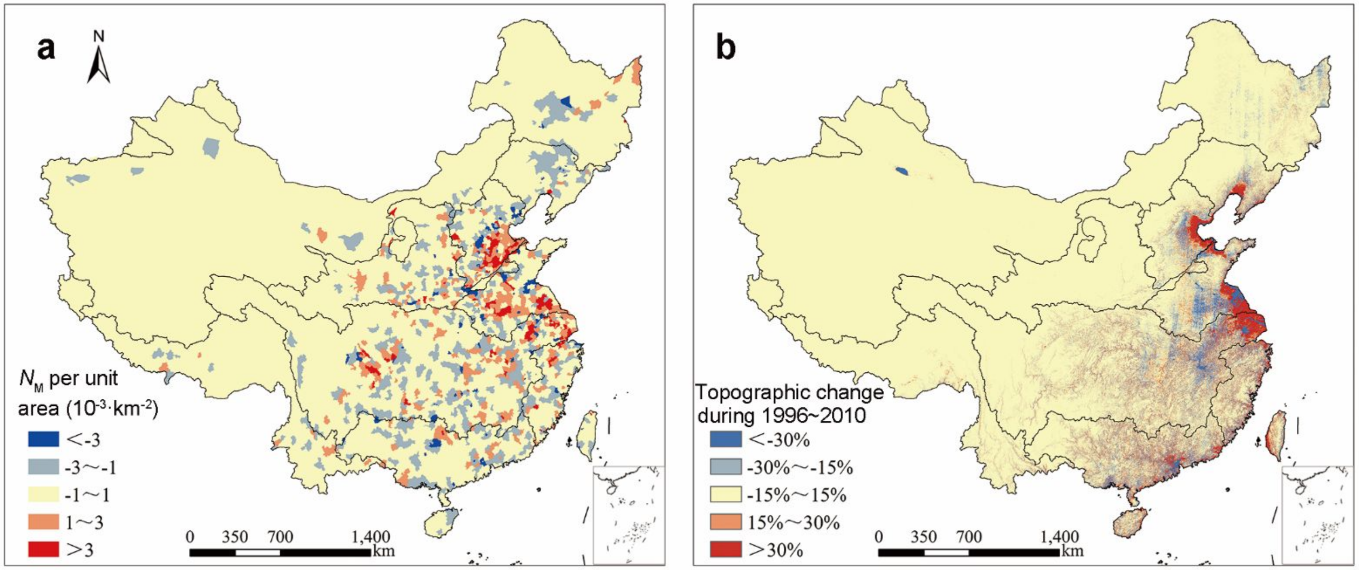
Supplementary Fig. 3. Comparison of different actual evaporation formulas (revised from Liu, 2014 [42]).



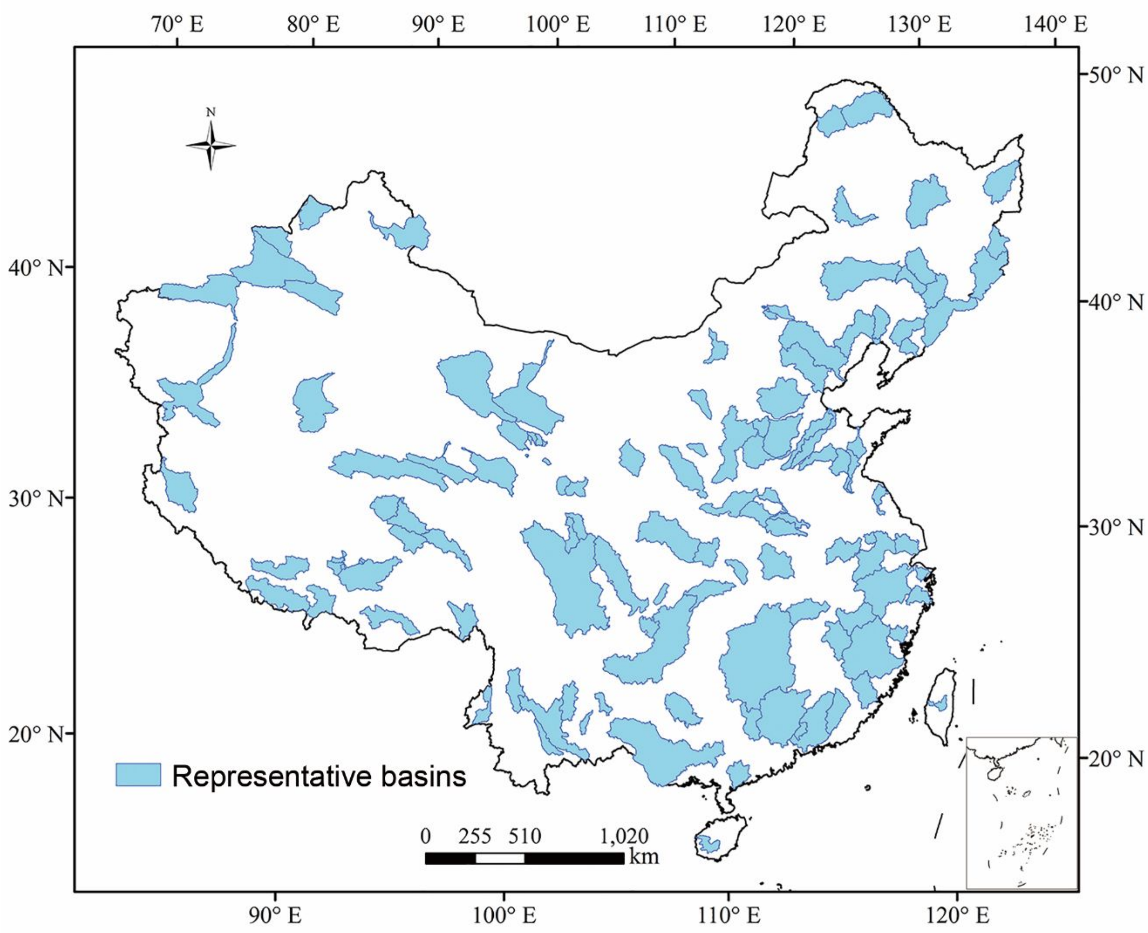
Supplementary Fig. 4. Distribution of water-erosion area and no-water erosion area in China.



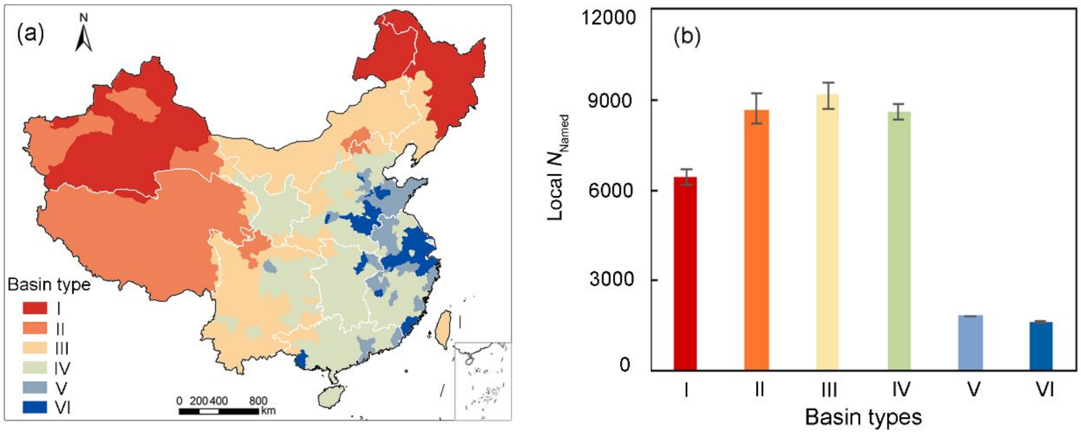
Supplementary Fig. 5. Numbers of rivers with different catchment areas extracted from different DEMs.



Supplementary Fig. 6. Distributions across China of: (a) N_M per unit area. (b) Relative difference of topographic change in terms of elevation values during 1996–2010.



Supplementary Fig. 7. Distribution of 107 representative basins throughout China, taken from *Encyclopedia of Rivers and Lakes in China*.



Supplementary Fig. 8. Estimated N_{Named} in China: (a) Distribution of six types of basins in China with different population densities. (b) Histogram displaying local N_{Named} using corresponding proportions of the three stream-ordering schemes.

Supplementary Table 1.

Numbers of rivers ≥ 100 km^2 in China obtained from National Census and different stream-ordering methods, numbers of rivers after excluding pseudo-rivers, rivers extrapolated from low-resolution topographic data, difference in number of rivers due to topographic change, and numbers of named rivers.

Basin	N_{Census}	N_{Horton}	N_{Stahler}	N_{Shreve}	N'_{Horton}	N'_{Strahler}	N'_{Shreve}	N_{T}	N_{M}	N_{Named}
Yangtze	5,276	5,194	7,246	125,518	5,194	6,026	104,387	4,714	419	30,201
Yellow	2,061	2,253	3,096	72,532	1,809	2,575	60,321	1,963	130	
Amur	2,428	2,737	3,839	60,266	2,676	3,193	50,120	4,577	259	
Liao	791	903	1,298	19,184	809	1,079	15,954	1,439	123	
Hai	892	949	1,340	28,060	908	1,114	23,336	1,700	-13	
Huai	1,266	993	1,405	20,011	993	1,168	16,642	1,308	-42	
Fujian & Zhejiang	694	701	993	22,378	701	826	18,611	784	74	
Pearl	1,685	1,723	2,413	47,092	1,723	2,007	39,164	1,831	259	
International	2,467	2,834	3,961	123,858	2,488	3,294	103,007	3,009	148	
Interior	5,349	8,900	12,424	151,047	5,309	10,332	125,618	6,930	596	
$\sum$	22,909	27,187	38,015	669,946	22,610	31,615	557,161	28,254	1,953	

The minus sign means the number of rivers increases due to topographic change.

$N_{\text{Name d}}$: the number of named rivers in China estimated calculated using different P_f in different regions.

Supplementary Table 2.

Numbers of rivers (and pseudo-rivers) ≥ 100 km^2 in China based on Horton scheme within different catchment areas for the ten basins of China using 30 m × 30 m DEMs.

Basin	Catchment Area (km^2)					
	≥100,000	10,000–100,000	1,000–10,000	100–1,000	50–100	10–50
Yangtze	8	41	450	4,695	5,193	39,962
Yellow	2	16	190	2,045	2,280	16,204
Amur	2	33	235	2,467	2,849	20,724
Liao	1	11	85	806	812	5,496
Hai	2	12	80	855	940	7,369
Huai	2	7	100	884	1,003	7,593
Fujian & Zhejiang	0	7	57	637	622	4,970
Pearl	1	13	162	1,547	1,706	13,132
International	3	23	246	2,562	2,915	22,065
Interior	2	55	789	8,054	8,806	77,870
∑	23	218	2,394	24,552	27,126	215,385

Supplementary Table 3.

Numbers of pseudo-rivers within a given catchment area for ten basins of China.

Basin	Catchment Area (km^2)		
	100–1000	50–100	10–50
Yangtze	0	0	0
Yellow	444	952	10,530
Amur	61	70	528
Liao	94	88	568
Hai	41	63	355
Huai	0	1	0
Fujian & Zhejiang	0	0	0
Pearl	0	0	0
International	346	857	12,118
Interior	3,591	5,110	45,110
∑	4,577	7,141	69,209

Supplementary Table 4.

Numbers of rivers in China with different catchment area extracted from 90m ×90m DEMs.

Basin	Catchment Area (km^2)					
	≥100,000	10,000–100,000	1,000–10,000	100–1,000	50–100	10–50
Yangtze	6	39	459	4624	5405	39008
Yellow	2	18	195	2137	2313	17640
Amur	2	35	260	2742	3076	21537
Liao	1	10	74	744	777	6101
Hai	2	10	82	822	909	7336
Huai	2	6	100	774	906	7291
Fujian & Zhejiang	0	5	44	469	456	3698
Pearl	1	12	150	1435	1595	12548
International	3	20	230	2334	2711	20383
Interior	1	50	692	7084	7865	58823
∑	20	205	2,286	23,165	26,013	194,365

Supplementary Table 5.

Numbers of rivers in China with different catchment area extracted from 900m ×900 m DEMs.

Basin	Catchment Area (km^2)					
	≥100,000	10,000–100,000	1,000–10,000	100–1,000	50–100	10–50
Yangtze	8	36	469	3,020	1,777	1,371
Yellow	2	15	190	987	214	32
Amur	2	35	245	2,215	1,580	1,652
Liao	1	11	80	687	495	483
Hai	1	13	67	551	391	409
Huai	1	10	81	483	342	281
Fujian & Zhejiang	0	6	58	430	222	162
Pearl	1	14	166	992	545	375
International	3	23	256	1,532	941	819
Interior	2	53	742	6,327	4,084	4,876
∑	21	216	2,354	17,224	10,591	10,460

Supplementary Table 6.

Numbers of rivers with different catchment areas extrapolated from 900 m × 900 m DEMs

Basin	$N_{900(100\text{-}1000)}$	$N_{900(50\text{-}100)}$	$N_{900(10\text{-}50)}$	N_{T}	N_{Tmin}	N_{Tmax}
Yangtze	5112	5623	50099	5625	2952	7079
Yellow	2071	2278	20296	2278	1195	2867
Amur	2671	2938	26171	2953	1556	3712
Liao	872	959	8546	964	508	1212
Hai	730	803	7157	811	429	1019
Huai	883	971	8652	975	513	1226
Fujian & Zhejiang	632	695	6196	696	366	876
Pearl	1809	1990	17732	1990	1044	2505
International	2790	3069	27346	3072	1613	3866
Interior	8088	8897	79260	8885	4655	11185
∑	25659	28224	251454	28250	14832	35547

Supplementary Table 7.

Numbers of rivers in China with different catchment area extracted from resampled 900m ×900 m DEMs.

Basin	Catchment Area (km^2)					
	≥100,000	10,000–100,000	1,000–10,000	100–1,000	50–100	10–50
Yangtze	8	42	464	2748	1523	1112
Yellow	2	15	177	933	208	21
Amur	2	33	249	1994	1382	1183
Liao	1	10	79	593	382	352
Hai	1	10	77	553	371	349
Huai	1	6	107	486	300	178
Fujian & Zhejiang	0	6	56	383	185	91
Pearl	1	15	163	833	446	313
International	3	22	254	1443	883	778
Interior	1	58	724	5887	4095	4578
∑	20	217	2350	15853	9775	8955

Supplementary Table 8.

Numbers of rivers ≥ 100 km^2 in ten basins in China related to changes in topography.

Basin	N_{30}	N_{900}	N_{900}'	N_{900}- N_{900}'	N_{30} - N_{900}'	P_M	N_M	N_{Mmin}	N_{Mmax}	Slope(°)
Yangtze	5,194	3,533	3,262	271	1,932	7.67%	419	187	536	13.4
Yellow	2,253	1,194	1,127	67	1,126	5.61%	130	18	222	9.3
Amur	2,737	2,497	2,278	219	459	8.77%	259	199	282	4.6
Liao	903	779	683	96	220	12.32%	123	86	126	5.4
Hai	949	632	641	-9	308	-1.42%	-13	-22	33	6.32
Huai	993	575	600	-25	393	-4.35%	-42	-42	29	2.1
Fujian & Zhejiang	701	494	445	49	256	9.92%	74	38	84	14.2
Pearl	1,723	1,173	1,012	161	711	13.73%	259	130	259	11.1
International	2,834	1,814	1,722	92	1,112	5.07%	148	44	245	18.1
Interior	8,900	7,124	6,670	454	2,230	6.37%	596	357	760	5.8
∑	27,187	19,815	18,440	1,375	8,747	6.94%	1,953	995	2,576	-

P_M is the percentage of migrated rivers defined as the ratio of (N_{900}- $N_{900'}$) and N_{900}.

Supplementary Table 9.

Scenarios, listing proportions of the three stream-ordering schemes.

NP_f	P_{HO}	P_{ST}	P_{SH}	N_{Mixed}
1	1	0	0	27,143
2	0.9	0	0.1	27,209
3	0.9	0.1	0	27,225
4	0.8	0	0.2	27,240
5	0.8	0.1	0.1	27,284
6	0.7	0	0.3	27,284
7	0.8	0.2	0	27,331
8	0.6	0	0.4	27,368
9	0.7	0.1	0.2	27,379
10	0.7	0.2	0.1	27,476
11	0.5	0	0.5	27,522
12	0.6	0.1	0.3	27,553
13	0.7	0.3	0	27,600
14	0.6	0.2	0.2	27,720
15	0.5	0.1	0.4	27,857
16	0.4	0	0.6	27,876
17	0.6	0.3	0.1	27,919
18	0.6	0.4	0	28,134
19	0.5	0.2	0.3	28,193
20	0.5	0.3	0.2	28,510
21	0.4	0.1	0.5	28,565
22	0.5	0.4	0.1	28,749
23	0.3	0	0.7	28,880
24	0.5	0.5	0	29,047
25	0.4	0.2	0.4	29,107
26	0.4	0.3	0.3	29,532
27	0.4	0.4	0.2	29,923
28	0.4	0.5	0.1	30,171
29	0.3	0.1	0.6	30,200
30	0.4	0.6	0	30,434
31	0.3	0.2	0.5	31,102
32	0.3	0.3	0.4	31,643
33	0.3	0.4	0.3	32,047
34	0.3	0.5	0.2	32,232
35	0.3	0.6	0.1	32,327
36	0.3	0.7	0	32,378
37	0.2	0	0.8	32,496
38	0.2	0.8	0	34,777
39	0.2	0.1	0.7	34,950
40	0.2	0.7	0.1	35,146

Supplementary Table 9.

Scenarios, listing proportions of the three stream-ordering schemes. *(Continued)*

NP_f	P_{HO}	P_{ST}	P_{SH}	N_{Mixed}
41	0.2	0.6	0.2	35,513
42	0.2	0.5	0.3	35,830
43	0.2	0.2	0.6	35,968
44	0.2	0.4	0.4	36,119
45	0.2	0.3	0.5	36,307
46	0.1	0.9	0	37,102
47	0	1	0	38,015
48	0.1	0.8	0.1	38,057
49	0.1	0.7	0.2	39,350
50	0	0.9	0.1	39,911
51	0.1	0.6	0.3	40,790
52	0	0.8	0.2	42,316
53	0.1	0.5	0.4	42,625
54	0.1	0.4	0.5	44,718
55	0	0.7	0.3	45,364
56	0.1	0.3	0.6	47,363
57	0	0.6	0.4	49,446
58	0.1	0.2	0.7	50,427
59	0.1	0.1	0.8	53,341
60	0.1	0	0.9	53,520
61	0	0.5	0.5	55,132
62	0	0.4	0.6	63,499
63	0	0.3	0.7	77,460
64	0	0.2	0.8	103,854
65	0	0.1	0.9	171,600
66	0	0	1	669,946

Supplementary Table 10.

Information on four human aggregation modes of basins in China in terms of population density.

Basin types	Population density in Qing Dynasty (person · km^{-2})	Range of NP_f	Maximum occurrence NP_f	Proportions of the three schemes	Frequency (A/B)*
I	< 0.4	24–52	㉚	0.4:0.6:0	4/13
II	0.4–2	24–48	㉟	0.3:0.6:0.1	4/24
III	2–30	28–61	㊵	0.2:0.7:0.1	4/22
IV	30–140	27–48	㊽	0.1:0.8:0.1	11/30
V	140–200	36–52	㊼	0:1:0	3/12
VI	> 200	48–52	㊿	0:0.9:0.1	2/6

A is the number of basins with most frequent occurrences of NP_f; B is the number of basins located in each basin type.

Supplementary Table 11.

Number of named rivers in China aggregated by catchment area.

Basin Area (km^2)	N_{Named}	$N_{\text{Named - min}}$	$N_{\text{Named - max}}$
≥ 10000	457	256	735
1000–10000	2,917	2,225	3,719
100–1000	26,945	24,264	29,755
50–100	29,501	26,682	32,424
10–50	220,578	212,168	228,977

References

1. Bai, R, Li, TJ and Huang, YF *et al.* An efficient and comprehensive method for drainage network extraction from DEM with billions of pixels using a size-balanced binary search tree. *Geomorphology* 2015; **238**: 56–67.
2. Lehner, B and Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* 2004; **296**: 1–22.
3. Lehner, B, Verdin, K and Jarvis, A. New global hydrography derived from spaceborne elevation data. *Eos. Transactions.* 2008; **89**: 93–94.
4. Horton, RE. Erosional development of streams and their drainage basins: hydro-physical approach to quantitative morphology. *Geol. Soc. Am. Bull.* 1945; **56**: 275–370.
5. Leading group office of the first national census for water. *General Plan for the First National Census for Water (in Chinese)*. Beijing: China Water &Power Press, 2010.
6. Strahler, AN. Hypsometric (area-altitude) analysis of erosional topography. *Geol. Soc. Am. Bull.* 1952; **63**: 1117–1142.
7. Strahler, AN. Quantitative analysis of watershed geomorphology. *EOS Trans. AGU* 1957; **38**: 913–920.
8. Shreve, R. Statistical Law of Stream Numbers. *J. Geol.* 1966; **74**: 17–37.
9. O'Callaghan, JF and Mark, DM. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* 1984; **28**: 323–344.
10. Schneider, A, Jost, A and Coulon, C *et al.* Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density. *Geophys. Res. Lett.* 2017; **44**: 2773–2781.
11. Sangireddy, H, Carothers, RA and Stark, CP *et al.* Controls of climate, topography,

vegetation, and lithology on drainage density extracted from high resolution topography data. *J. Hydrol.* 2016; **537**: 271–282.

12. Rui, XF. *Principles of Hydrology (in Chinese).* Beijing: China Water & Power Press, 2004.
13. Middleton, N and Thomas, DS and UNEP. *World Atlas of Desertification.* London: Edward Arnold, 1992.
14. Mao, F, Sun, H and Yang, HL. Research Progress in Dry/wet Climate Zoning (in Chinese). *Progress in Geography* 2011; **30**: 17–26.
15. Budyko, MI. *Evaporation under Natural Conditions.* Jerusalem: Isr. Program for Sci. Transl, 1948.
16. Department of Hydrology and Water Resources, University of Arizona. *The Quantitative Features of China's Water Resources: An Overview.* Tucson, AZ: Department of Hydrology and Water Resources, University of Arizona, 1983. (Report on Natural Resource Systems No. 38).
17. Ol'deskop, EM. *On Evaporation from the Surface of River Basins (in Russia).* Tartu: Tartu Univ, 1911.
18. Zhang, L, Dawes, W and Walker, GR. Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water Resour. Res.* 2001; **37**: 701–708.
19. Tang, KL. *Soil and Water Conservation in China (in Chinese)*. Beijing: Science Press, 2004.
20. Zhu, TX. Gully and tunnel erosion in the hilly Loess Plateau region, China. *Geomorphology* 2012; **153**: 144–155.

21. Sorg, A, Bolch, T and Stoffel, M. Climate change impacts on glaciers and runoff in Tien Shan (Central Asia). *Nat. Clim. Chang.* 2012; **2**: 725–731.
22. Xiao, J, Jin, ZD and Wang, J *et al.* Hydrochemical characteristics, controlling factors and solute sources of groundwater within the Tarim River Basin in the extreme arid region, NW Tibetan Plateau. *Quat. Int.* 2015; **380**: 237–246.
23. Lutz, AF, Immerzeel, WW and Shrestha, AB *et al.* Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nat. Clim. Chang*. 2014; **4**: 587–592.
24. Li, Z, Lin, XQ and Coles, AE *et al.* Catchment-scale surface water-groundwater connectivity on china's loess plateau. *Catena* 2017; **152**: 268–276.
25. Pekel, JF, Cottam, A and Gorelick, N *et al.* High-resolution mapping of global surface water and its long-term changes. *Nature* 2016; **540**: 418–422.
26. Widder, S, Besemer, K and Singer, GA *et al.* Fluvial network organization imprints on microbial co-occurrence networks. *Proc. Natl. Acad. Sci. U. S. A.* 2014; **111**: 12799–12804.
27. Ni, JR and Ma, AN. *River dynamic geomorphology (in Chines).* Beijing: Peking University Press, 1998.
28. Zanardo, S, Zaliapin, I and Foufoula-Georgiou, E. Are American rivers Tokunaga self‐similar? New results on fluvial network topology and its climatic dependence. *JGR* 2013; **118**: 166–183.
29. Yang, DW, Herath, S and Musiake, K. Spatial resolution sensitivity of catchment geomorphologic properties and the effect on hydrological simulation. *Hydrol. Process.* 2001; **15**: 2085–2099.

30. Yang, L, Xu, YP and Han, LF et al. River networks system changes and its impact on storage and flood control capacity under rapid urbanization. Hydrol. Process. 2016; 30: 2401–2412.
31. Zheng, HB. Birth of the Yangtze River: age and tectonic-geomorphic implications. Natl. Sci. Rev. 2015; 2: 438–453.
32. Deng, XZ, Huang, JK and Rozelle, S. *et al.* Impact of urbanization on cultivated land changes in China. *Land Use Pol.* 2015; **45**: 1–7.
33. Meierdiercks, KL, Smith, JA and Baeck, ML *et al.* Analyses of urban drainage network structure and its impact on hydrologic response. *J. Am. Water Resour. Assoc.* 2010; **46**: 932–943.
34. Editorial Committee of Encyclopedia of Rivers and Lakes in China. *Encyclopedia of Rivers and Lakes in China (in Chinese)*. Beijing: China Water Power Press, 2014.
35. Roberts, GO and Rosenthal, JS. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* 2007; **44**: 458–475.
36. Ge, JX. *History of Population in China (Volume 5), Qing Dynasty (in Chinese).* Shanghai: Fudan University Press, 2001.
37. Ministry of Water Resources, PRC and National Bureau of Statistics, PRC. *Bulletin of First National Census for Water.* Beijing: China Water &Power Press, 2013.
38. Callow, JN, Van Niel, KP and Boggs, GS. How does modifying a DEM to reflect known hydrology affect subsequent terrain analysis? *J. Hydrol.* 2007; **332**: 30–39.
39. Department of Geography, Northwest Normal Universitys. *Atlas of Physical geography of China.* Beijing: SinoMaps Press, 1984.

40. Moore, ID, Grayson, RB and Ladson, AR. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* 1991; **5:** 3–30.

41. National Bureau of Statistics of China. *China Statistical Yearbook 2014 (in Chinese).* Beijing: China Statistics Press, 2014.

42. Liu, CM. *Hydrogeography of China (in Chinese).* Beijing: Science Press, 2014.

Complete Abbreviation List

AI: Aridity Index.

D_{P}: Number of pseudo-rivers with catchment area ≥ 100 km^2 per unit area, $10^{-3}\cdot\mathrm{km}^{-2}$.

D_{Horton}: Number of rivers estimated by Horton stream-ordering scheme per unit area, 10^{-3} km^{-2}.

Dif: Relative difference between n_{Mixed} and $n_{\mathrm{Named}}$for representative basins.

D_{M}: Change in number of rivers ≥ 100 km^2 due to topographic change per unit area, $10^{-3}\cdot\mathrm{km}^{-2}$.

D_{Named}: Number of named rivers estimated by hybrid stream-order rule per unit area, $10^{-3}\cdot\mathrm{km}^{-2}$.

D_{1776}: Population density in Early Qing Dynasty (1776), person·km^{-2}.

$f(x)$: Probability density function.

$I(x)$: Integral function of $f(x)$.

ID_{Shreve}, ID_{Strahler}, and ID_{Horton}: Value of ID for each elementary reach derived from overlay of the three river networks.

$MCTI$: modified compound topographic index, a function of surface flow accumulation, aridity, and topographic slope.

n: Stream order.

n_{d}: Stream order of downstream river reach.

n_{HO}: Total number of reaches consistent with Horton scheme in a representative basin.

n_{Horton}: Number of rivers estimated by Horton stream-ordering scheme in a representative basin.

n_M: Actual number of river reaches, $n_\mathrm{M} = \lim\limits_{\mathrm{M}\to\infty} \overline{N_\mathrm{MC}}$.

n_max: Maximum stream order.

n_SH: Total number of reaches consistent with Shreve scheme in a representative basin.

n_Shreve: Number of rivers estimated by Shreve stream-ordering scheme in a representative basin.

n_Mixed: Number of rivers obtained from random naming simulation based on Monte Carlo method for representative basins.

n_ST: Total number of reaches consistent with Strahler scheme in a representative basin.

n_Strahler: Number of rivers estimated by Strahler stream-ordering scheme in a representative basin.

n_Total: Total number of elemental reaches in a representative basin.

n_Named: Number of historic named rivers for representative basins.

$n_{1,}$ n_2: Respective stream orders of upstream rivers at the bifurcation.

N_Census: Number of rivers with catchment area $\geq$ 100 km^2 reported in National Census.

N_Horton: Number of rivers with catchment area $\geq$ 100 km^2 estimated by Horton stream-ordering scheme.

N_Horton': Number of rivers with catchment area $\geq$ 100 km^2 estimated by Horton stream-ordering scheme after exclusion of pseudo-rivers.

N_M: Discrepancy in number of rivers with catchment area $\geq$ 100 km^2 caused by changes in land topography from 1996 and 2011.

N_Mixed: Number of rivers (catchment area $\geq$ 100 km^2) estimated by hybrid stream-order rule.

N_Mmin, N_Mmax: Minimum and maximum values of N_M.

N_{MC}: Total number of river reaches obtained from Monte Carlo simulation.

$\overline{N_{\mathrm{MC}}}$: mean value of N_{MC}

N_{Named}: Number of named rivers in 1990s with catchment area ≥ 100 km^2.

N_{P}: Number of pseudo-rivers with catchment area ≥ 100 km^2.

NP_{f} : Scenario number related to P_{f}.

N_{Shreve}: Number of rivers with catchment area ≥ 100 km^2 estimated by Shreve stream-ordering scheme.

N_{Shreve}': Number of rivers with catchment area ≥ 100 km^2 estimated by Shreve stream-ordering scheme after correction of the errors induced by pseudo-rivers.

N_{Strahler}: Number of rivers with catchment area ≥ 100 km^2 estimated by Strahler stream-ordering scheme.

N_{Strahler}': Number of rivers with catchment area ≥ 100 km^2 estimated by Strahler stream-ordering scheme after correction of the errors induced by pseudo-rivers.

N_{30}: Number of rivers extracted from the 30 m × 30 m DEMs.

$N_{30(50\text{-}100)}$: Numbers of rivers with catchment areas 50–100 km^2.

$N_{30(100\text{-}1000)}$ are numbers of rivers with catchment areas 100–1000 km^2.

N_{90}: Number of rivers extracted from the 90 m × 90 m DEMs.

N_{900}: Number of rivers extracted from the 900 m × 900 m DEMs.

$N_{900(\geq 100000)}$: Numbers of rivers with catchment areas ≥ 100,000 km^2 extracted from the 900 m × 900 m DEMs.

$N_{900(10000\text{-}100000)}$: Numbers of rivers with catchment areas 10,000-100,000 km^2 extracted from the 900 m × 900 m DEMs.

$N_{900(1000\text{-}10000)}$: Numbers of rivers with catchment areas 1,000-10,000 km^2 extracted from the 900 m × 900 m DEMs.

$N_{900(100\text{-}1000)}$: Numbers of rivers with catchment areas 100-1000 km^2 extracted from the 900 m × 900 m DEMs.

$N_{900(50\text{-}100)}$: Numbers of rivers with catchment areas 50-100 km^2 extracted from the 900 m × 900 m DEMs.

$N_{900(10\text{-}50)}$: Numbers of rivers with catchment areas 10-50 km^2 extracted from the 900 m × 900 m DEMs.

$N_{900(10\text{-}100)}$: Numbers of rivers with catchment areas 10-100 km^2 extracted from the 900 m × 900 m DEMs.

N_{T}: Number of rivers with catchment area ≥ 100 km^2 extrapolated from the number of rivers ≥ 10,000 km^2.

N_{Tmin}, N_{Tmax}: Minimum and maximum values of N_{T}.

N_{900}': Number of rivers with catchment area ≥ 100 km^2 extracted from the 900 m × 900 m DEMs'.

P: Mean annual precipitation, mm.

PET^0: Mean annual potential evapo-transpiration, mm.

P_{f}: Proportions of three stream-ordering schemes ($P_{\mathrm{HO}} : P_{\mathrm{ST}} : P_{\mathrm{SH}}$).

P_{HO}: Probability of occurrence of segments consistent with the Horton scheme.

P_{i}: Probability density function, $P_{\mathrm{i}} = \int_{\mathrm{Xj-1}}^{\mathrm{Xj}} f(x)\mathrm{d}x$.

$P_{\mathrm{in_i}}$: Unbiased estimator of P_{i} when n_{i} is sufficiently large.

P_{M}: Percentage of migrated rivers.

P_{SH}: Probability of occurrence of segments consistent with the Shreve scheme.

P_{ST}: Probability of occurrence of segments consistent with the Strahler scheme.

R_B: Number of rivers at a given order divided by the number at next higher order.

R_B', R_B'': pseudo-bifurcation ratio. The ratio of numbers of rivers with different catchment area grades.

R_P: Ratio of pseudo-rivers.

S: Slope (°).

WO: Water occurrence.

X_1, X_2, and X_3: Intervals of $f(x)$.

ξ: Uniformly distributed random variable.

η: Random variable, $\eta = f(\xi)$.