

## Article

# A Stream Order Family and Order-Based Parallel River Network Routing Method

Xi Yang <sup>1</sup>, Chong Wei <sup>1</sup>, Zhiping Li <sup>2</sup>, Heng Yang <sup>3</sup> and Hui Zheng <sup>4,\*</sup> 

<sup>1</sup> College of Surveying and Geo-Informatics, North China University of Water Resources and Electric Power, Zhengzhou 450046, China; yangxi940822@gmail.com (X.Y.); dixin313@163.com (C.W.)

<sup>2</sup> College of Geosciences and Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450045, China; lizhiping@ncwu.edu.cn (Z.L.)

<sup>3</sup> Science and Technology Research Institute, China Three Gorges Corporation, Beijing 100038, China; yang\_heng2@ctg.com.cn (H.Y.)

<sup>4</sup> Laboratory of Regional Climate-Environment Research for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China; zhenghui@tea.ac.cn (H.Z.)

\* Correspondence: zhenghui@tea.ac.cn

**Abstract:** River network routing is fundamental for flood forecasting but becomes computationally demanding with networks comprising thousands to millions of reaches. The sequential nature of upstream-to-downstream flow paths within river networks poses a significant challenge for parallel computation. This study introduces a family of stream orders coupled with an order-based parallel routing approach. We assign each reach an order that falls between one more than the maximum order of its upstream reaches and one less than the order of its downstream reach. This strategy enables parallel simulation of reaches with identical orders while sequentially processing those with different orders, thus maintaining the crucial upstream-to-downstream dynamic. To further enhance parallel scalability, we strategically relax the upstream-to-downstream relationship along the longest flow paths, dividing the network into independent subnetworks and introducing halo reaches to mitigate the impact of inexact inflows. We validate our approach using China's Yangtze River basin, the country's largest river network with 53,600 fully connected reaches. Employing a conceptual parallel execution machine, we demonstrate that our method achieves 80% parallel efficiency with up to 25 processors. By strategically introducing breakpoints, we further enhance scalability, enabling efficient simulations on 77 processors while maintaining 80% efficiency. These results highlight the scalability and efficiency of our methods for large-scale, high-resolution river network modeling within Earth system models. Our study also lays a theoretical groundwork for optimizing stream orders and halo reach placements, crucial for advancing river network modeling.

**Keywords:** stream order; river network routing; parallel computing; flood

**Citation:** Yang, X.; Wei, C.; Li, Z.; Yang, H.; Zheng, H. A stream order family and order based parallel river network routing method. *Water* **2024**, *1*, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

**Copyright:** © 2024 by the authors. Submitted to *Water* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

River networks play an essential role in the terrestrial water cycle [1]. They serve as the primary conveyance systems for water, linking terrestrial and aquatic ecosystems, and facilitating the transport of water, sediment, nutrients, and pollutants across diverse landscapes [2–8]. The representation of river network routing is a critical component in the development and refinement of land surface models and Earth system models [9]. The fidelity with which land surface models or Earth system models capture the spatial and temporal dynamics of river networks directly influences the models' ability to simulate key hydrological processes, such as evapotranspiration, infiltration, and runoff [10,11]. Moreover, the routing of water through river networks is a critical control on the timing and magnitude of floods [12,13]. As such, the inclusion of the detailed, fine-scale river network routing is not only vital for improving our understanding of the terrestrial water cycle but also for informing flood defense in regions where flood risk is a pressing concern [14].

Land surface models and Earth system models strive to capture the intricate details of river networks [15–17] over large domains, such as entire nations or the globe [13,18–20]. Fine-scale river network routing, which requires short river reaches and frequent time-stepping, is crucial for accurately simulating water flow through the complex channels of a river network. This level of detail is crucial for resolving small-scale features and processes that can significantly impact larger-scale river water flow patterns [21–24] and is essential for action-based flood forecasting [25,26]. The computational demand associated with the fine-scale river network routing over a large domain is substantial [23,27–30]. Moreover, the integration of fine-scale river routing with other model components, such as nutrient transport and human water use [25], further compounds the computational load. As a result, researchers and model developers must balance the desire for high-fidelity river network representation with the practical constraints of computational feasibility. The ongoing advancements in high-performance computing and the development of more sophisticated numerical methods are key to addressing these computational challenges [23, 27–30] and enabling the next generation of Earth system models and land surface models to incorporate large-domain fine-scale river network routing effectively.

The implementation of parallel processing in river network routing is essential for enhancing the performance of large-domain fine-scale simulations. However, this approach introduces considerable challenges. A primary concern stems from the inherently distributed structure of river networks, where numerous tributaries and river segments are intricately interconnected. This necessitates advanced data management protocols and parallel computation strategies to facilitate accurate information exchange regarding water flow dynamics between different segments of the river network [30]. It is imperative that this information exchange preserves the integrity of the upstream-to-downstream flow sequence, as the flow in any given segment is contingent upon the flow conditions in its upstream segments. Parallelization must respect this sequential dependency to ensure that the simulation outcomes are physically plausible and precise [23,28,30,31].

The inherent sequential nature of river flow significantly impedes the development of efficient and scalable parallel processing methods for river network routing. Furthermore, the spatial variability within river networks, marked by differences in channel widths, slopes, and flow velocities, adds to the complexity of parallelization. This variability necessitates dynamic load balancing techniques to prevent computational resources from being underutilized or overburdened. The development of robust parallel schemes is therefore critical to accommodate the diverse scales and dynamics inherent to river networks, without compromising the fidelity or reliability of the model's projections.

Various methods exist for parallelizing river network routing, such as domain decomposition [23,30] and halo reaches [31]. Domain decomposition involves partitioning river networks into independent tributaries and interconnected mainstreams. The tributaries are then allocated to various processors for concurrent simulation, while the mainstreams are processed in an upstream-to-downstream sequence. However, the parallelization scalability is limited. As the number of processors increases, the distribution of reaches within the tributaries diminishes, and the mainstream reaches become more dominant [23]. In an extreme scenario with an infinite number of processors, the majority of river reaches would be classified as mainstreams, effectively rendering the routing process predominantly sequential.

With the halo-reach method, the river network is divided into independent subnetworks, each of which is simulated in parallel without maintaining an upstream-to-downstream relationship at their boundaries [29]. To mitigate the inaccuracies arising from this division, halo reaches are prepended upstream of each subnetwork. These halo reaches serve to buffer the impact of boundary inexactitudes. This strategy offers scalability and efficiency, although careful configuration of the halo reaches is essential to ensure simulation accuracy [31].

We introduce a family of stream orders and utilize a parallel strategy termed stream order-based parallelism in this study. This approach involves categorizing river reaches

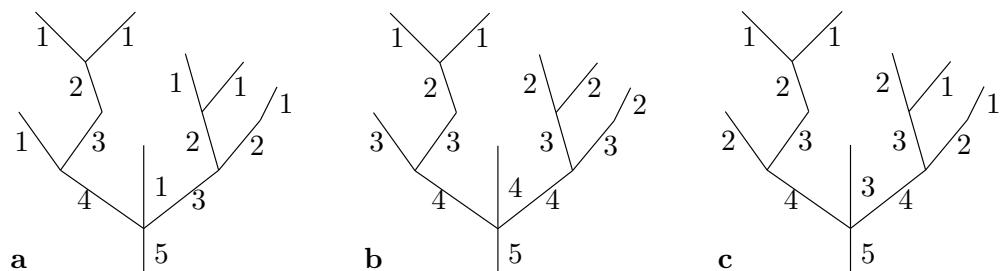
into groups based on similar stream orders. Reaches within the same order group are simulated in parallel, while those of different orders are processed sequentially. This method is designed to achieve both parallel scalability and simulation precision, strictly maintaining the upstream-to-downstream flow hierarchy. Compared to domain decomposition, this method demonstrates superior parallel scalability. In the theoretical case of infinite processors, computation time would approximate that of sequential routing along the longest flow path. Furthermore, this method can be effectively combined with the halo-reach strategy, with a few strategically placed halo reaches on the main flow path to enhance parallel scalability. The efficacy of these methods has been validated through testing in the Yangtze River basin and mainland China, confirming their scalability and efficiency.

The structure of this paper is as follows: Section 2 elaborates on the proposed stream ordering methods and the order-based parallel routing technique. Section 3 outlines the study area and the river network geometry dataset utilized. Section 4 presents the outcomes of applying the proposed methods. Finally, Section 5 summarizes the findings and discusses their implications.

## 2. Stream Orders and Parallel River Network Routing

### 2.1. A Family of Stream Orders Designed for Parallel River Routing

Our research is inspired by several foundational studies. The mizuRoute river network routing model, for instance, employs the Strahler stream order to segment mainstreams into branches [23]. Within each branches, reaches are sequentially simulated from upstream to downstream, while branches of the same Strahler order are processed in parallel. This approach underscores the potential of order-based parallelism. However, the Strahler order categorizes branches rather than individual reaches, which results in a granularity too coarse for efficient parallel routing. The disparity in the number of branches across Strahler orders and in the number of reaches in a branch introduces significant variability in the computational load, posing a considerable challenge for load balancing and scalable parallel processing.



**Figure 1.** Illustration of the proposed stream orders. **a** The lower limits of the stream orders. **b** The upper limits of the stream orders. **c** A feasible set of stream order values conforming the two ordering rules.

In response to these challenges, our proposed method assigns stream orders directly to river reaches, enabling more granular parallelization. The Spatially Explicit Integrated Modeling System (SEIMS) model introduces a novel layer-based parallelism [28,32,33], where reaches are stratified by their topological distance from the outlet. Reaches within the same layer are processed in parallel, with layers sequenced. Yet, the allocation of a reach to a specific layer is predetermined by the river network's topography. If the number of reaches in a layer does not align with the number of processors, processor idleness ensues [28,33]. The SEIMS model lacks the flexibility required to optimize reach distribution across layers.

Our proposed stream ordering method offers a more adaptable alternative to the SEIMS model. The ordering is defined by two rules: (1) a reach's order falls within an integer range that is one more than the highest order of its upstream reaches and one less than the order of its downstream reach, and (2) the order value is an integer range from one up to the number of the reaches along the longest flow path within all the river networks

under examination. This ordering ensures that reaches with identical orders are free from upstream-downstream dependencies, allowing for parallel processing.

The ordering rules generate a spectrum of feasible stream order values. The lower and upper bounds of the feasible values for each reach can be estimated as follows, as shown in Figure 1. The lower bounds are calculated by repeating the following two steps: (1) assigning the upmost reaches an order of one, and (2) setting a reach's order to one plus the highest upstream order. This calculation parallels the upstream-downstream strategy utilized in SEIMS [33, Fig. 6b]. Conversely, the upper bounds are determined by: (1) designating the outlet reach an order equivalent to the number of reaches along the longest flow path, and (2) subtracting one from the downstream reach's order. This calculation aligns with the downstream-upstream strategy in SEIMS [33, Fig. 6c]. The flexibility in stream order assignment provides an opportunity to optimize reach distribution, potentially minimizing processor idleness and enhancing parallel efficiency, as evidenced by our results.

As the number of processors increases significantly, computation time will predominantly be dictated by the reaches along the longest flow path. To further enhance parallel scalability, our proposed ordering methods can be complemented with halo reaches. By strategically inserting halo reaches upstream of select reaches along the longest flow path, the river network is effectively partitioned into subnetworks with shorter longest flow paths. The orders for these subnetworks and halo reaches are assigned according to the above proposed methods. The use of halo reaches should be judicious, reserved for scenarios where the longest flow path emerges as the primary determinant of computation time.

## 2.2. A Conceptual Shared-Memory Parallel Execution Machine for River Network Routing

In this study, we utilize a conceptual shared-memory parallel execution model to assess the parallel performance and scalability of our proposed stream ordering method. This model offers a theoretical framework for evaluation, factoring out the impact of inefficient code implementation and hardware architectural differences. The parallel performance and scalability metrics obtained are instrumental in guiding the optimization of the proposed stream orders.

The conceptual machine is equipped with  $m$  processors capable of simultaneously processing up to  $m$  reaches. The river networks under consideration comprise a total of  $n$  reaches. The machine selects groups of reaches from the river networks, beginning with the lowest order. Each group consists of reaches of the same order, which are selected randomly. The size of each group not exceeding  $m$ . This process is repeated until all reaches have been simulated, as outlined in Algorithm 1.

We disregard the time spent on fetching and potential data saving operations. The total computation time is quantified by the total number of groups required to process the entire river network. Parallel performance is gauged using the metrics of parallel speedup and efficiency. Parallel speedup is calculated as the ratio of the computation time for sequential simulation (i.e., the total number of river reaches) to that for parallel simulation. Perfect scaling is achieved if the speedup ratio increases linearly with the number of processors. Parallel efficiency is determined by the ratio of actual parallel speedup to the ideal speedup under perfect scaling conditions. This efficiency metric is crucial for evaluating the parallel performance and scalability of the proposed stream orders.

During computation, there may be instances of processor idleness. This idleness is attributed to the mismatch between the number of reaches in a group and the number of available processors. Such idleness can diminish parallel efficiency and scalability. To identify potential sources of this idleness, we will calculate the computation deficiency for each individual reach. The computation deficiency of a reach is defined as the ratio of the number of idle processors during the routing of the reach to the number of total processors. If no processors are idle during the routing of a reach, the computation deficiency is zero. If idle processors are present, the computation deficiency is positive; as the number of idle

processors increases, the deficiency approaches one. This metric is valuable for identifying reaches that contribute to processor idleness and for guiding the optimization of stream orders.

---

**Algorithm 1** A shared-memory parallel execution machine of river network routing.
 

---

**Require:** There are  $n$  streams.  $O_i \in \mathbb{Z}^+$  is the order of the  $i$ -th stream,  $i = 1 \dots n$ . The order is sorted:  $O_{i+1} - O_i \geq 0$ .

**Require:** The machine has  $m$  processors.  $m \in \mathbb{Z}^+$ .

**Ensure:** The simulation of a stream must be after its upstream.

```

 $i \leftarrow 1$  {the first stream of a group}
 $j \leftarrow 1$  {the last stream of a group}
while  $i \leq n$  do
  for  $k = 1$  to  $m$  do
    if  $i + k - 1 \leq n$  and  $O_{i+k-1} = O_i$  then
       $j \leftarrow i + k - 1$ 
    end if
  end for
  Parallel Simulation Simulate the streams from  $i$  to  $j$  in parallel.
   $i \leftarrow j + 1$ 
end while
```

---

### 3. Study Domain and Data

We applied the proposed stream ordering and order-based parallel routing methods to the Yangtze River, the largest river network in China. The river network is characterized by full connectivity among its reaches. This case provides a rigorous evaluation of our proposed methods at a substantial scale and under challenging conditions for parallel routing.

For the delineation of the Yangtze River network, we employed the Multi-Error-Removed Improved-Terrain Hydrography (MERIT-Hydro) dataset [16,34], a global hydrography dataset with a 3 arc-second spatial resolution. This dataset includes a hydrologically adjusted digital elevation model, an eight-direction flow model, flow accumulation area, river channel width, and height above the nearest drainage. Our delineation process utilized the flow accumulation area and the eight-direction flow model. The threshold for the flow accumulation area was set at  $20 \text{ km}^2$  for catchment delineation and  $10 \text{ km}^2$  for river centerline extraction. The extraction yielded a single river network comprising 53,600 reaches.

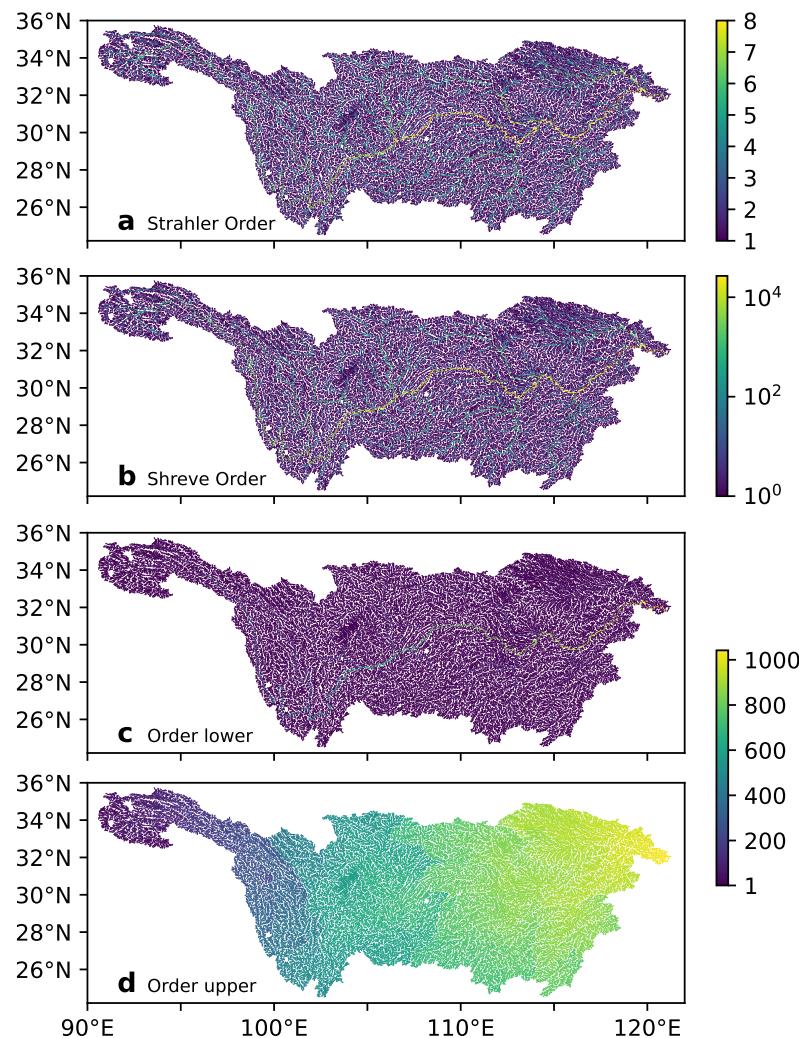
### 4. Results

#### 4.1. Stream Orders for the Yangtze River Network

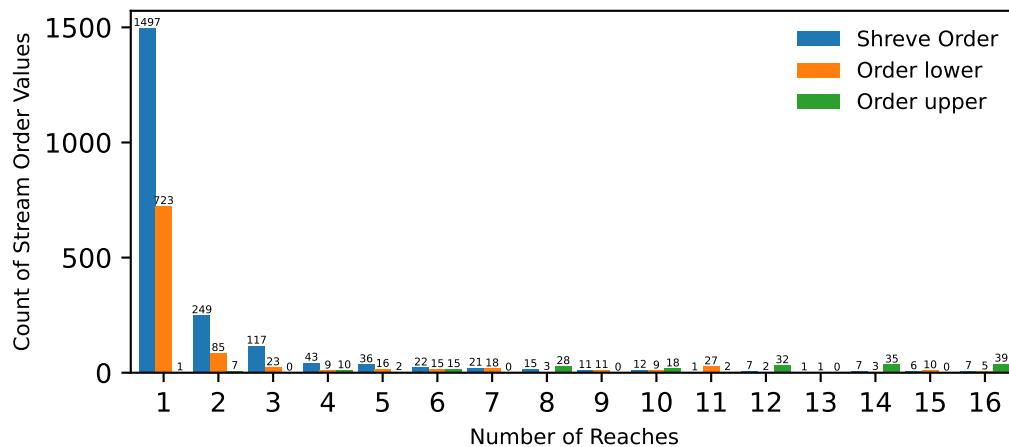
Figure 2 illustrates the spatial distribution of the proposed stream orders in the Yangtze River basin, juxtaposed with those of the Strahler and Shreve stream orders. The Strahler system exhibits a narrow range of order values, with a maximum value of eight observed in this case. This limited range restricts its application to parallelizing branches, akin to the mizuRoute model [23], rather than individual reaches. The coarse granularity of the Strahler order poses challenges for achieving effective load balancing.

In contrast, both the Shreve order and our proposed stream order offer adequate granularity for order-based parallel routing. It is important to note that the Shreve order is applicable for parallel routing only when no upstream-downstream pairs share the same order value, a condition that is coincidentally met in this study. However, the Shreve order spans a range from 1 to 26,828, which is excessively broad. As depicted in Figure 3, 2,378 unique Shreve order values are assigned to only one reach. This extensive range, coupled with the scarcity of reaches sharing the same order value, can lead to inefficiencies in parallelization.

Figure 2c and Figure 2d display the spatial distribution of the lower and upper bounds of our proposed order family, respectively. These bounds range from 1 to 1,042, which corresponds to the number of reaches along the longest flow path. The upper bound demonstrates an even distribution of order values throughout the river network, while the lower bound shows a greater concentration in the lower order values. For the upper bound, a limited number of order values are assigned to a small set of reaches, as shown in Figure 3. The balanced distribution of order values, along with the adequate granularity and the presence of sufficient reaches sharing the same order value, render the upper bounds of our proposed stream order family particularly well-suited for parallel river network routing.



**Figure 2.** Spatial distribution of stream orders over the Yangtze River basin. **a** The Strahler order. **b** The Shreve order. **c** The lower bounds of the proposed order family. **d** The upper bounds of the proposed order family.



**Figure 3.** The count of stream order values with a limited number of reaches. For the sake of clarity, only those stream orders with 16 or fewer reaches are illustrated.

#### 4.2. Parallelization Performance, Scalability, and Hotspot of Computation Deficiencies

Figure 4 displays the parallel speedup, efficiency, and scalability of the proposed stream orders compared to the Shreve order. Initially, as the number of processors increases, the ideal speedup is linearly proportional to the number of processors allocated. When the allocated processors become abundant enough, the parallelization process encounters a bottleneck due to the sequentiality along the longest flow path. The black dashed lines in the figure indicate this theoretical limit. In this instance, the number of processors at which the ideal speedup reaches the theoretical limit is 51, which corresponds to the ratio of the total number of reaches in the river network to the number of reaches along the longest flow path.

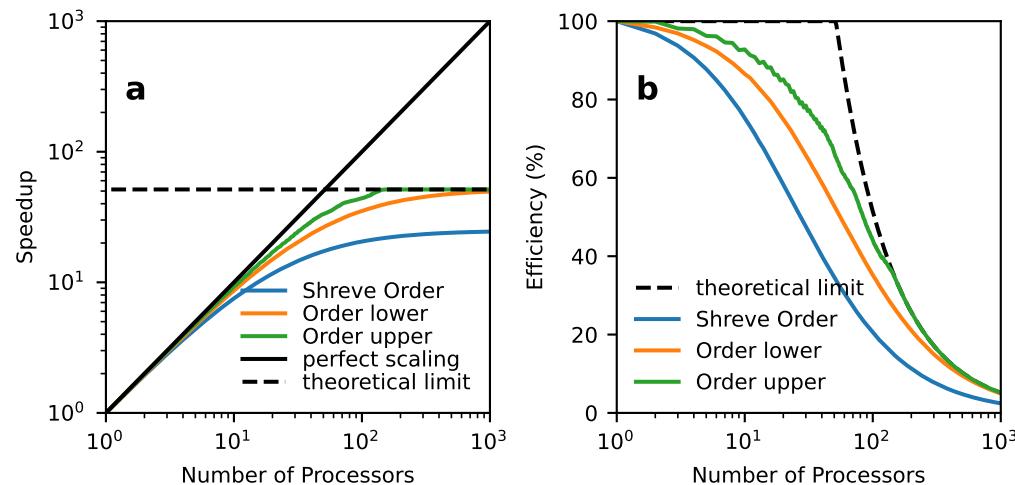
The Shreve order shows the lowest parallelization performance and scalability. Its efficiency falls below 80% as soon as the processor count exceeds 6. When the number of processors is approximately 150 or above, the speedup is nearly saturated. The saturated speedup is significantly lower than the theoretical limit, with a value of approximately 24.5. This value is close to the ratio of the total number of reaches to the number of unique Shreve order values (i.e., 24.87). This indicates that the low saturated speedup is likely due to the broad range of Shreve order values and the scarcity of reaches sharing the same order value, as shown in Figure 2 and Figure 3.

The lower bounds of our proposed stream order family offer superior speedup and efficiency compared to the Shreve order. As the processor count increases, the parallel speedup increases more rapidly. Once the number of processors exceeds approximately 280, further increases have minimal impact on speedup. The saturation speedup closely matches the theoretical limit, indicating the granularity of our proposed stream order family is well-suited for parallel processing. The lower bounds also scale better. The efficiency remains above 80% when the number of processors is less than 14.

The lower bounds of the proposed stream order family demonstrate better speedup and efficiency than the Shreve order. As the allocated processor count escalates, the parallel speedup increases at a more rapid rate than that of the Shreve order. However, once the number of processors surpasses about 280, the enhancement in parallel speedup becomes negligible. The saturation speedup is close to the theoretical limit, showing that the number of unique order values is appropriate. This suggests that our proposed stream order family is ideally granular for parallel river network routing. Additionally, the lower bounds showcase enhanced scalability, maintaining efficiency above 80% when the processor count does not exceed 14.

The upper bounds demonstrated the best performance and scalability among the three methods. Efficiency is maintained above 80% when the number of processors is 25 or fewer. As the number of processors increases, the speedup rapidly saturates, closely approaching the theoretical limits at a processor count of 141. These results indicate that the upper

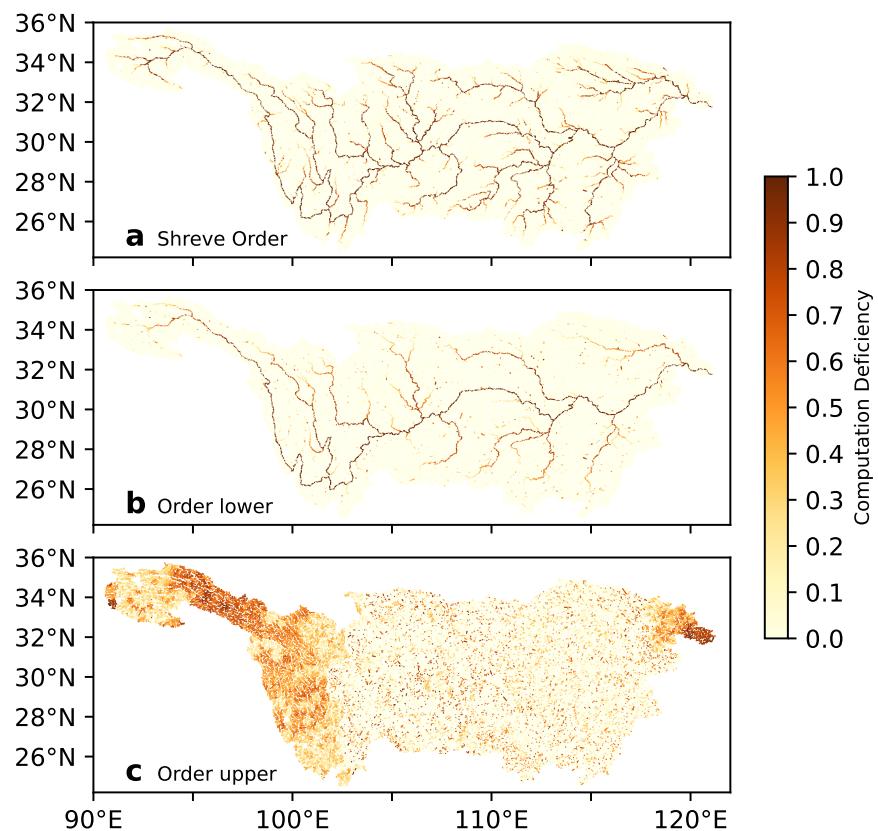
bounds of the proposed order family are particularly well-suited for parallel river network routing. The scalability of the upper bounds is either comparable to or surpasses those found in the existing literature [23,30]. It is important to note that scalability in this study is measured using a conceptual parallel execution model rather than with a real-world implementation.



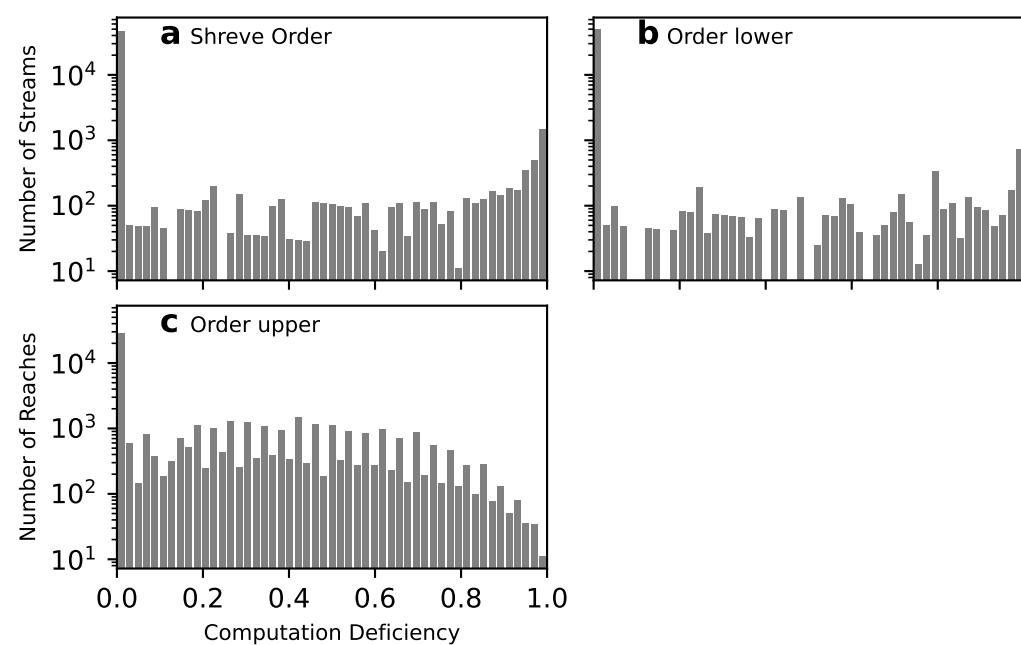
**Figure 4.** Parallelization speedup (a) and efficiency (b) for the Shreve orders and the proposed stream orders, respectively. The solid black lines represent the ideal speedup or efficiency under conditions of perfect scaling. The dashed black lines indicate the limits imposed by the upstream-to-downstream sequentiality along the longest flow path.

Figure 5 and Figure 6 respectively illustrate the spatial distribution and the histogram of computation deficiencies within the Yangtze River network. These deficiencies were measured using 51 processors, the point at which the ideal speedup aligns with the theoretical limit. For the Shreve order and the lower bounds of the proposed stream order family, reaches with significant computation deficiencies are primarily located in the mainstreams. This inefficiency arises from the limited number of mainstream reaches that share the same order values. In contrast, due to the coarse granularity of the upper bounds, the likelihood of computation deficiencies occurring is lower, and they are more concentrated.

The upper bounds exhibit a markedly different pattern, with reaches showing high computation deficiencies predominantly found in the upstream reaches, especially in the headwater regions. The greater number of upstream reaches, compared to mainstream reaches, makes it more probable that they will share order values. This sharing reduces the probability of processor idleness and computation deficiencies. The contrasting patterns observed between the lower and upper bounds suggest that there is considerable potential for optimizing the order values to improve parallel efficiency.



**Figure 5.** Computation deficiency of routing each individual reach observed at a processor count of 51 for the Yangtze River network.



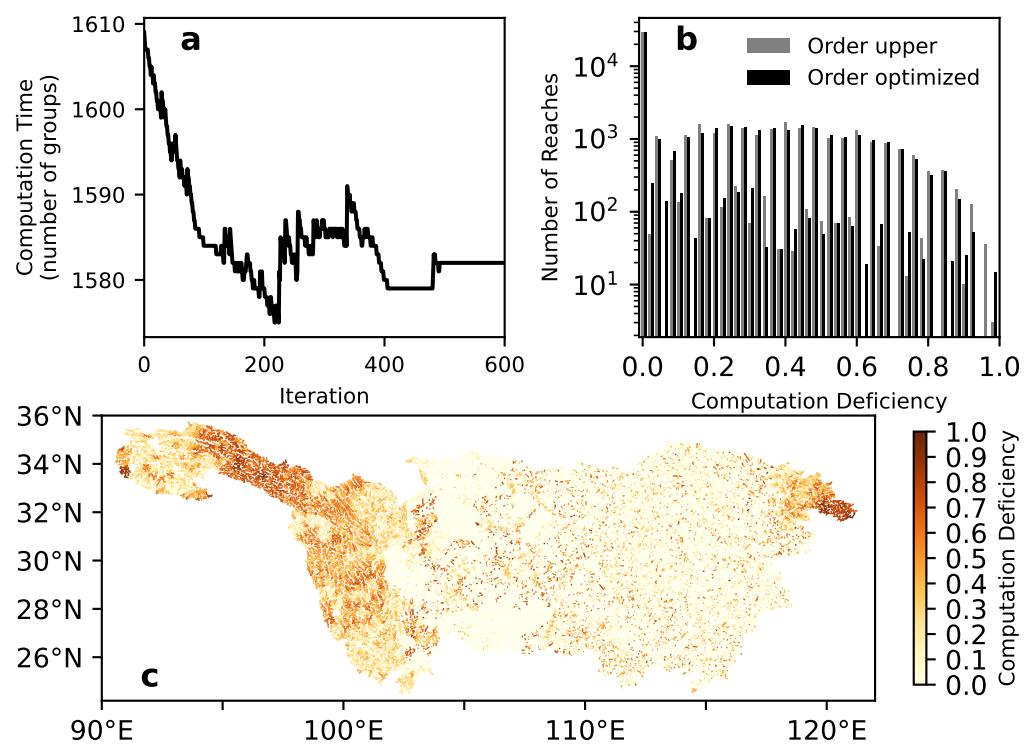
**Figure 6.** Histogram of the computation deficiency for the Yangtze River network observed at a processor count of 51.

#### 4.3. Optimization of the Stream Order in between the Two Bounds

We show that the stream order can be optimized using a simple strategy that iteratively minimizes total computation time. The iterative optimization operates only on the reaches

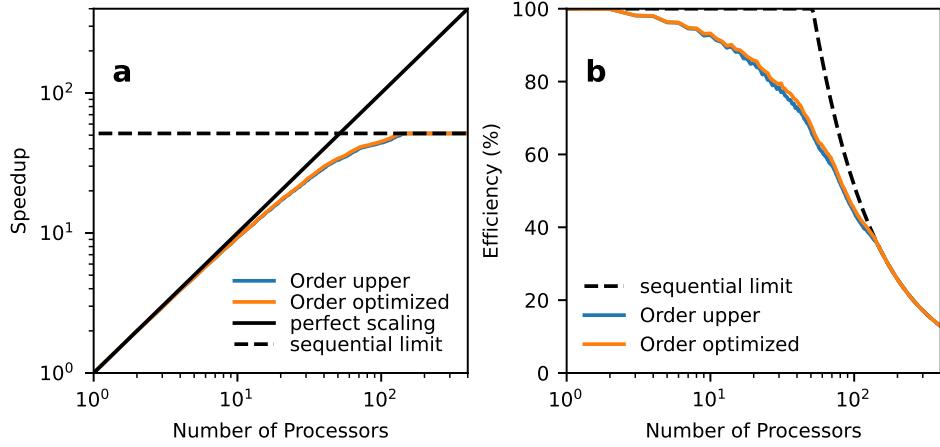
where the upper and lower bounds differ and starts with the upper bounds of the stream order family, which show the best parallel performance yet. In each iteration, the reach with the most significant computation deficiency is targeted, with its order being reduced by one. Concurrently, appropriate adjustments are made to the orders of upstream reaches to ensure adherence to the two rules outlined in Section 2.1.

Figure 7a illustrates the reduction in total computation time at each optimization iteration. The total computation time experiences a rapid decline during the initial iterations, followed by a gradual stabilization. The stream orders that yield the minimum total computation time are then selected for further analysis. A comparison in the histograms of computation deficiencies between the optimized orders and the upper bounds of the stream order family is presented in Figure 7b. The optimization results in a slight reduction in the number of reaches with high computation deficiencies. Figure 7c displays the spatial distribution of computation deficiencies, showing that the optimization strategy effectively reduces these deficiencies, particularly in the middle reaches. However, the headwater and downstream regions exhibit negligible changes, indicating that a more advanced optimization strategy may be necessary to achieve enhanced parallel performance uniformly across the river network.



**Figure 7.** Optimization of the stream order. **a** Decrease in the total computation time with the optimization iteration. **b** Comparison in the histogram of the computation deficiency for the optimized stream orders and the upper bounds of the stream order family. **c** Spatial distribution of the computation deficiency.

Figure 8 compares the upper bounds and the optimized stream orders for different number of processors. The optimized stream orders are obtained from iterations five times of the number of processors. The optimized stream orders exhibit a slight but consistent improvement in parallel speedup and efficiency, when the number of processors is limited. When the parallel performance of the upper bounds have already hit theoretical limits imposed by the sequentiality along the longest flow path, the optimized stream orders show no further improvement. These results suggest that relaxation of the upstream-to-downstream relationship is necessary for better parallel performance when the processor count is significant enough.



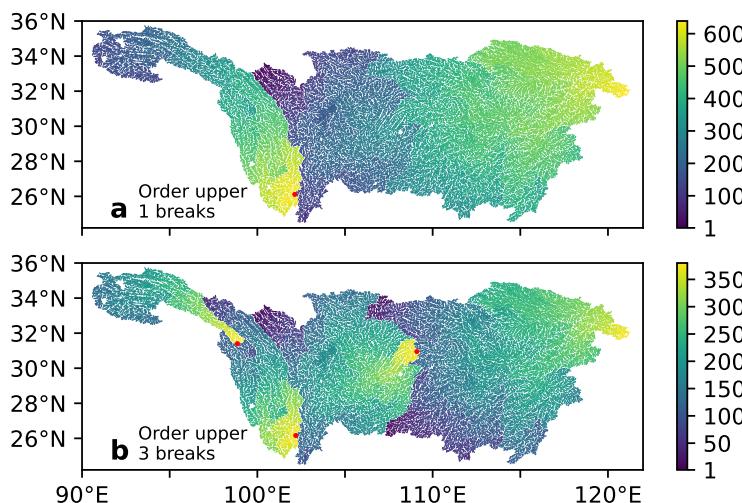
**Figure 8.** Same as Figure 4 but for the optimized stream orders and the upper bounds of the proposed order family.

#### 4.4. A few relaxations of the upstream-to-downstream relationship along the longest flow path

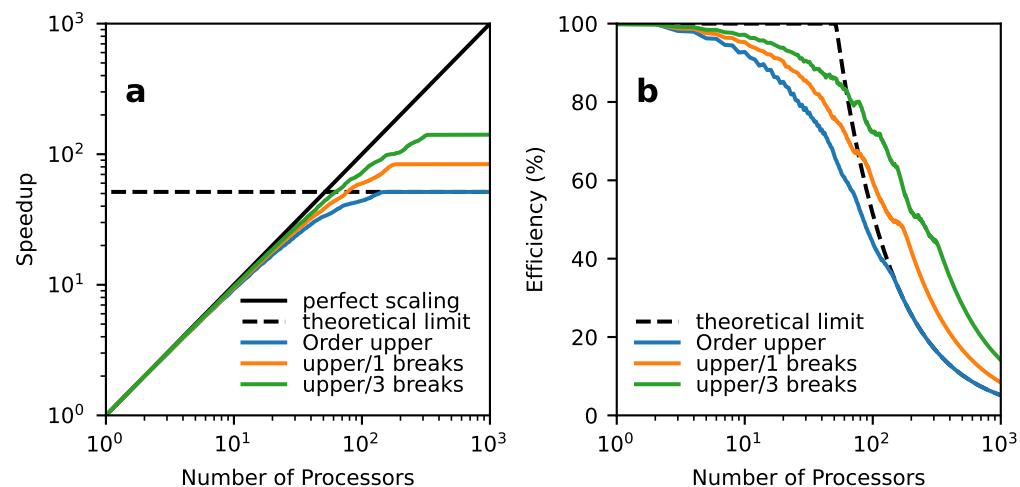
In this section, we introduce breakpoints along the longest flow path. The fully connected Yangtze river network is partitioned into independent subnetworks with these breakpoints. Halo reaches, which are duplicates of the upstream reaches at these breakpoints, are prepended upstream to the subnetworks. The inflow to the uppermost halo reaches is established and fixed during each routing time step and updated using the outflow from the upstream subnetworks before subsequent steps. We neglect the computation and communication time associated with these updates from our analysis, as our proposed parallel methods necessitate only a few breakpoints. In this study, we set the buffer length—the minimum number of reaches along any flow path within the halo-reach region—to six. The buffer length, approximately 80 km, is deemed sufficient to mitigate the impact of inexact inflows resulting from the relaxation [31]. Apart from these adjustments, the segmented river networks can be processed using the ordering methods detailed in Section 2.1 and routed using the conceptual machine outlined in Section 2.2.

Figure 9 illustrates the spatial distribution of the upper bounds of the stream orders for the Yangtze River network with one and three breakpoints. The breakpoints are placed in the equal intervals along the longest flow path. As analyzed above, the saturating parallel speedup equals to the ratio of the total number of reaches to the number of reaches along the longest flow path within the river networks. Such equally placed breakpoints can achieve the near-optimal saturating parallel speedup. However, the segmentation of the river network and appending of the halo reaches lead to the changes in part of the longest flow path. The placement of the breakpoints should be optimized by considering the interplay among the breakpoint insertion, halo reach appending, and the changes in the longest flow path. An optimization algorithm that can minimize the number of reaches along the longest flow path should be developed to achieve the best saturating parallel performance. However, such an optimization algorithm is beyond the scope of this study.

Figure 10 illustrates the parallel speedup and efficiency of the stream orders for the Yangtze River network with one and three breakpoints. As discussed, introducing breakpoints effectively raises the saturating parallel performance by shortening the longest flow path. The saturating parallel speedup can now exceed the theoretical limits imposed by the upstream-to-downstream sequentiality along the longest flow path. Moreover, the improvement in parallel speedup and efficiency is consistent across all different allocations of processor numbers. An 80% parallel efficiency can be maintained for 40 processors with one breakpoint and for 77 processors with three breakpoints in the Yangtze River network. Performance saturation occurs when the number of processors exceeds approximately 150 and 280 for the network with one and three breakpoints, respectively. The enhancements in parallelization scalability are significant.



**Figure 9.** Same as Figure 2d, but for the stream orders with one (a) and three (b) breaks along the longest flow path. Red dots denote the breakpoints. The halo reaches are not shown for clarity.



**Figure 10.** Same as Figure 4, but for the upper bounds of the order family with two (the orange lines) or four (the blue lines) breaks along the longest flow path.

## 5. Discussion and Conclusions

In conclusion, this study introduces a novel family of stream orders and an associated order-based parallel routing method, designed to preserve the upstream-to-downstream flow sequence while facilitating parallel processing. The application of these methods within the Yangtze River basin has yielded superior performance and scalability, outperforming current strategies.

The potential for future research is abundant, with the optimization of stream orders and strategic placement of breakpoints offering promising avenues for further enhancing parallel performance. The findings of this study lay a solid theoretical groundwork for subsequent investigations in this domain.

Furthermore, the convergence of high-performance computing advancements and the sophisticated numerical methods introduced herein is poised to significantly advance land surface and Earth system modeling. By enabling the practical integration of large-domain fine-scale river network routing, forthcoming models will be better positioned to simulate intricate hydrological processes, thereby enabling more precise and actionable forecasts. This research represents a pivotal step towards the development of more accurate and efficient modeling frameworks.

347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363

**Author Contributions:** Conceptualization: H.Z.; methodology: H.Z.; software: H.Z.; validation: C.W. and Z.L.; formal analysis: X.Y.; investigation: X.Y.; resources: H.Y.; data curation: H.Z.; writing—original draft preparation: X.Y.; writing—review and editing: C.W. and Z.L.; visualization: X.Y. and H.Y.; supervision: H.Z.; project administration: H.Y.; funding acquisition: H.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (grant number 42275178 and 42075165).

**Data Availability Statement:** The MERIT hydrography dataset is available at [http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT\\_Hydro](http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro) (accessed on 2024-05-01). The delineated river network, along with the scripts for river network analysis and figure generation used throughout this paper, can be found at the GitHub repository <https://github.com/hzheng88/paper-2024-stream-order-parallel-yangtze>.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Trenberth, K.E.; Smith, L.; Qian, T.; Dai, A.; Fasullo, J. Estimates of the global water budget and its annual cycle using observational and model data. *J. Hydrometeorol.* **2007**, *8*, 758–769. <https://doi.org/10.1175/JHM600.1>.  
364  
365  
366  
367  
368
2. Dai, A.; Trenberth, K.E. Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *J. Hydrometeorol.* **2002**, *3*, 660–687. [https://doi.org/10.1175/1525-7541\(2002\)003<0660:EOFDFC>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0660:EOFDFC>2.0.CO;2).  
369  
370  
371  
372  
373  
374  
375
3. Dai, A.; Qian, T.; Trenberth, K.E.; Milliman, J.D. Changes in continental freshwater discharge from 1948 to 2004. *J. Clim.* **2009**, *22*, 2773–2792. <https://doi.org/10.1175/2008JCLI2592.1>.  
376  
377  
378  
379
4. Regnier, P.; Resplandy, L.; Najjar, R.G.; Ciais, P. The land-to-ocean loops of the global carbon cycle. *Nature* **2022**, *603*, 401–410. <https://doi.org/10.1038/s41586-021-04339-9>.  
380  
381  
382  
383  
384  
385
5. Sohail, T.; Zika, J.D.; Irving, D.B.; Church, J.A. Observed poleward freshwater transport since 1970. *Nature* **2022**, *602*, 617–622. <https://doi.org/10.1038/s41586-021-04370-w>.  
386  
387  
388  
389
6. Liu, C.; Liang, X.; Ponte, R.M.; Vinogradova, N.; Wang, O. Vertical redistribution of salt and layered changes in global ocean salinity. *Nat. Commun.* **2019**, *10*, 3445. <https://doi.org/10.1038/s41467-019-11436-x>.  
390  
391  
392  
393  
394
7. Li, M.; Peng, C.; Zhou, X.; Yang, Y.; Guo, Y.; Shi, G.; Zhu, Q. Modeling global riverine DOC flux dynamics from 1951 to 2015. *J. Adv. Model. Earth Syst.* **2019**, *11*, 514–530. <https://doi.org/10.1029/2018MS001363>.  
395  
396  
397  
398  
399
8. Aufdenkampe, A.K.; Mayorga, E.; Raymond, P.A.; Melack, J.M.; Doney, S.C.; Alin, S.R.; Aalto, R.E.; Yoo, K. Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. *Front. Ecol. Environ.* **2011**, *9*, 53–60. <https://doi.org/10.1890/100014>.  
390  
391  
392  
393  
394
9. David, C.H.; Famiglietti, J.S.; Yang, Z.L.; Habets, F.; Maidment, D.R. A decade of RAPID—Reflections on the development of an open source geoscience code. *Earth Space Sci.* **2016**, *3*, 226–244. <https://doi.org/10.1002/2015ea000142>.  
395  
396
10. Senatore, A.; Mendicino, G.; Gochis, D.J.; Yu, W.; Yates, D.N.; Kunstmann, H. Fully coupled atmosphere-hydrology simulations for the central Mediterranean: Impact of enhanced hydrological parameterization for short and long time scales. *J. Adv. Model. Earth Syst.* **2015**, *7*, 1693–1715. <https://doi.org/10.1002/2015MS000510>.  
397  
398  
399
11. Wagner, S.; Fersch, B.; Yuan, F.; Yu, Z.; Kunstmann, H. Fully coupled atmospheric-hydrological modeling at regional and long-term scales: Development, application, and analysis of WRF-HMS. *Water Resour. Res.* **2016**, *52*, 3187–3211. <https://doi.org/10.1002/2015WR018185>.  
400  
401  
402
12. Lin, P.; Hopper, L.J.; Yang, Z.L.; Lenz, M.; Zeitler, J.W. Insights into hydrometeorological factors constraining flood prediction skill during the May and October 2015 Texas Hill Country flood events. *J. Hydrometeorol.* **2018**, *19*, 1339–1361. <https://doi.org/10.1175/JHM-D-18-0038.1>.  
403  
404  
405
13. Alfieri, L.; Burek, P.; Dutra, E.; Krzeminski, B.; Muraro, D.; Thielen, J.; Pappenberger, F. GloFAS - global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1161–1175. <https://doi.org/10.5194/hess-17-1161-2013>.  
406  
407
14. Willner, S.N.; Otto, C.; Levermann, A. Global economic response to river floods. *Nat. Clim. Chang.* **2018**, *8*, 594–598. <https://doi.org/10.1038/s41558-018-0173-2>.  
408  
409
15. Lehner, B.; Grill, G. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol. Process.* **2013**, *27*, 2171–2186. <https://doi.org/10.1002/hyp.9740>.  
410  
411
16. Yamazaki, D.; Ikeshima, D.; Sosa, J.; Bates, P.D.; Allen, G.H.; Pavelsky, T.M. MERIT Hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resour. Res.* **2019**, *55*, 5053–5073. <https://doi.org/10.1029/2019wr024873>.  
412  
413
17. Chaney, N.W.; Torres-Rojas, L.; Vergopolan, N.; Fisher, C.K. HydroBlocks v0.2: enabling a field-scale two-way coupling between the land surface and river networks in Earth system models. *Geosci. Model Dev.* **2021**, *14*, 6813–6832. <https://doi.org/10.5194/gmd-14-6813-2021>.  
414  
415  
416
18. Yang, Y.; Pan, M.; Lin, P.; Beck, H.E.; Zeng, Z.; Yamazaki, D.; David, C.H.; Lu, H.; Yang, K.; Hong, Y.; et al. Global reach-level 3-hourly river flood reanalysis (1980–2019). *Bulletin of the American Meteorological Society* **2021**, *102*, E2086–E2105. <https://doi.org/10.1175/BAMS-D-20-0057.1>.  
417  
418  
419

19. Lin, P.; Pan, M.; Beck, H.E.; Yang, Y.; Yamazaki, D.; Frasson, R.; David, C.H.; Durand, M.; Pavelsky, T.M.; Allen, G.H.; et al. Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resour. Res.* **2019**, *55*, 6499–6516. <https://doi.org/10.1029/2019WR025287>. 420  
421  
422
20. Wu, H.; Kimball, J.S.; Li, H.; Huang, M.; Leung, L.R.; Adler, R.F. A new global river network database for macroscale hydrologic modeling. *Water Resour. Res.* **2012**, *48*, 2012WR012313. <https://doi.org/10.1029/2012WR012313>. 423  
424
21. Yamazaki, D.; Oki, T.; Kanae, S. Deriving a global river network map and its sub-grid topographic characteristics from a fine-resolution flow direction map. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 2241–2251. <https://doi.org/10.5194/hess-13-2241-2009>. 425  
426
22. Thober, S.; Cuntz, M.; Kelbling, M.; Kumar, R.; Mai, J.; Samaniego, L. The multiscale routing model mRM v1.0: simple river routing at resolutions from 1 to 50 km. *Geosci. Model Dev.* **2019**, *12*, 2501–2521. <https://doi.org/10.5194/gmd-12-2501-2019>. 427  
428
23. Mizukami, N.; Clark, M.P.; Gharari, S.; Kluzek, E.; Pan, M.; Lin, P.; Beck, H.E.; Yamazaki, D. A vector-based river routing model for Earth system models: Parallelization and global applications. *J. Adv. Model. Earth Syst.* **2021**, *13*, e2020MS002434. <https://doi.org/10.1029/2020ms002434>. 429  
430  
431
24. Nguyen-Quang, T.; Polcher, J.; Ducharne, A.; Arsouze, T.; Zhou, X.; Schneider, A.; Fita, L. ORCHIDEE-ROUTING: revising the river routing scheme using a high-resolution hydrological database. *Geosci. Model Dev.* **2018**, *11*, 4965–4985. <https://doi.org/10.5194/gmd-11-4965-2018>. 432  
433  
434
25. Wada, Y.; De Graaf, I.E.M.; Van Beek, L.P.H. High-resolution modeling of human and climate impacts on global water resources. *J. Adv. Model. Earth Syst.* **2016**, *8*, 735–763. <https://doi.org/10.1002/2015MS000618>. 435  
436
26. Coughlan de Perez, E.; van den Hurk, B.; van Aalst, M.K.; Amuron, I.; Bamanya, D.; Hauser, T.; Jongma, B.; Lopez, A.; Mason, S.; Mendler de Suarez, J.; et al. Action-based flood forecasting for triggering humanitarian action. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 3549–3560. <https://doi.org/10.5194/hess-20-3549-2016>. 437  
438  
439
27. Yamazaki, D.; de Almeida, G.A.M.; Bates, P.D. Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resour. Res.* **2013**, *49*, 7221–7235. <https://doi.org/10.1002/wrcr.20552>. 440  
441  
442
28. Liu, J.; Zhu, A.X.; Liu, Y.; Zhu, T.; Qin, C.Z. A layered approach to parallel computing for spatially distributed hydrological modeling. *Environ. Model. Softw.* **2014**, *51*, 221–227. <https://doi.org/10.1016/j.envsoft.2013.10.005>. 443  
444
29. David, C.H.; Famiglietti, J.S.; Yang, Z.L.; Eijkhout, V. Enhanced fixed-size parallel speedup with the Muskingum method using a trans-boundary approach and a large subbasins approximation. *Water Resour. Res.* **2015**, *51*, 7547–7571. <https://doi.org/10.1002/2014WR016650>. 445  
446  
447
30. Liu, Y.H.; Yang, Z.L.; Lin, P.R. Parallel river channel routing computation based on a straightforward domain decomposition of river networks. *J. Hydrol.* **2023**, *625*, 129988. <https://doi.org/10.1016/j.jhydrol.2023.129988>. 448  
449
31. David, C.H.; Yang, Z.L.; Famiglietti, J.S. Quantification of the upstream-to-downstream influence in the Muskingum method and implications for speedup in parallel computations of river flow. *Water Resour. Res.* **2013**, *49*, 2783–2800. <https://doi.org/10.1002/wrcr.20250>. 450  
451  
452
32. Liu, J.; Zhu, A.X.; Qin, C.Z.; Wu, H.; Jiang, J. A two-level parallelization method for distributed hydrological models. *Environ. Model. Softw.* **2016**, *80*, 175–184. <https://doi.org/10.1016/j.envsoft.2016.02.032>. 453  
454
33. Zhu, L.J.; Liu, J.; Qin, C.Z.; Zhu, A.X. A modular and parallelized watershed modeling framework. *Environ. Model. Softw.* **2019**, *122*, 104526. <https://doi.org/10.1016/j.envsoft.2019.104526>. 455  
456
34. Yamazaki, D.; Ikeshima, D.; Tawatari, R.; Yamaguchi, T.; O'Loughlin, F.; Neal, J.C.; Sampson, C.C.; Kanae, S.; Bates, P.D. A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.* **2017**, *44*, 5844–5853. <https://doi.org/10.1002/2017GL072874>. 457  
458

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content. 459  
460  
461