Development of a Routing Procedure to Assist an Earth Systems Model with Long Term Coastal Discharge Predictions

Josefine Wilms[1] and Marcus Thatcher[2]

[1] Council for Scientific and Industrial Research, Stellenbosch 7600, South Africa,
jwilms@csir.co.za,
http://www.csir.co.za
[2] Oceans and Atmosphere, Commonwealth Scientific and Industrial Research Orginisation, Aspendale, Australia, marcus.thatcher@csiro.au,
http://www.csiro.au

Abstract. Nearest neighbour searches, scaling, and a flow accumulation method were applied to improve predictions for freshwater deposits from land surfaces to the ocean for an earth systems model. Runoff, generated by the Conformal Cubic Atmospheric Model (CCAM), was read at a coarse resolution and downscaled, whereas digital elevation- and accumulation values were obtained from the HydroSHEDS database and upscaled. The accumulation, digital elevation, and runoff values were matched using a KDTree nearest neighbour algorithm. Starting from a zero-valued initial water body, CCAM runoff was routed to neighbouring cells. Flow direction was determined with a maximum flow accumulation method whereas the Manning equation was used to calculate the discharge rate. Inland reservoirs and coastal waters were removed and added to an outflow term. Mass conservation checks confirmed that the proposed procedure conserves mass and a 25-year simulation shows that the relative discharge rates, river routes, and outflow locations were sufficiently predicted.

Keywords: Long Term Forecasting, Runoff Routing, Earth Systems Model, K-d tree, Scaling, Manning Equation

1 Introduction

Runoff is the residual water from precipitation after evapotranspiration. This moisture, not absorbed by soil or plants, results in continental freshwater discharges into the ocean. Water evaporates from the ocean's surface, is transported back to the land as atmospheric moisture and reaches the land surface as precipitation. This process is known as the land-ocean water cycle [3].

Evapotranspiration and precipitation vary spatially but the return of runoff into the ocean is mostly concentrated at the world's largest river mouths. This significant freshwater discharge at mouth locations results in the salinity of ocean-water to be less within these regions. Salinity differentials, in turn, result in regional changes of the ocean's density [11]. Estimates of freshwater fluxes into the ocean is therefore needed to study oceanic freshwater budgets. Since stream discharge can be measured quantitatively, such estimates are also important to check that earth systems models (ESMs) are adhering to mass conservation and performing with reasonable accuracy.

The performance of mapping- and prediction methods for the routing of terrestrial runoff fields are constantly improved upon. Fekete et al. [6] used river discharge information from gauging stations from the World Meteorological Organization Global Runoff Data Centre (GRDC) to calculate annual inter-station runoff. In addition, they simulated river discharge with a water balance model (WBM), driven by long-term mean monthly climate data. These WBM simulation results were then weighted by multiplying the value of each simulation point with the ratio of its discharge to the observed runoff of the corresponding inter-station region from the GRDC data. Using this method, a set of spatially distributed runoff fields were created at a 0.5° resolution.

Recently, Mizukami et al. [10] developed the routing tool, mizuRoute. MizuRoute can use both small- and large scale runoff outputs from land-surface models as input and produces a spatially distributed streamflow at various spatial scales. It can use both grid- and vector based river networks and applies two different river routing schemes: kinematic wave tracking and impulse response function-unit-hydrograph routing [10].

Another, widely used, routing method is a river network model termed the Routing Application for Parallel Computation of Discharge (RAPID). It was developed for the National Hydrography Dataset Plus river network for which lateral inflow is obtained from a land surface model. A matrix-based version of the Muskingum method is applied to calculate flow and water volumes in all reaches [4].

However, for water budget analyses in ESMs, only the discharges at coastal river mouths are of interest. To estimate continental discharge with runoff fields, a river transport model that routes the terrestrial runoff into the correct river mouths is required. This study therefore follows a similar approach to that of Dai et al. [3] who used a river transport model (RTM), developed by Branstetter et al. [2], to route surface runoff to the ocean. The RTM they used, implemented a linear advection scheme at a resolution of 0.5°.

The objective of this work is to update and improve the continental discharge estimates of CCAM [9] and, in so doing, improve its water budget approximations and forecasting capabilities. However, ESMs such as CCAM are run at coarse scales and currently operate at a resolution of 1°, or marginally higher, for global simulations [1]. It would be computationally too expensive to run the entire globe at a resolution high enough to incorporate smaller scale events.

Finding and incorporating accurate but computationally cheap methods for up- and downscaling as well as mapping non-matching grids therefore form an integral part of our study: Runoff, generated by CCAM, is downscaled and mapped to upscaled elevation- and flow accumulation data that are obtained from HydroSHEDS [8]. A matrix based RTM is then used to route the terrestrial runoff to the river mouths. In so doing, the freshwater fluxes and locations can be mapped back to the CCAM grid and used to drive its ocean model.

In the current work, the developed model is applied to the African continent only, but it should be noted that the procedures, discussed here, can be used for any landmass.

2 Methods

The modeling procedure, used in this study, is done in three stages and all code is written in Python 2.7: During the preprocessing stage, a digital elevation model (DEM)- and flow accumulation (FA) data are read and upscaled to a target resolution at which routing will be done. Runoff data, obtained from CCAM's NetCDF output at a coarse resolution are, in turn, downscaled to the target resolution and subsequently mapped to the new FA and DEM data locations. Flow direction is then determined from the upscaled accumulation set.

The second stage entails the routing of the water by utilising the flow direction information, obtained during the preprocessing procedure.

Postprocessing is done in the third stage: At the end of each simulation month, water budget results, obtained during the second stage, are written to a NetCDF file and visualisation of the results is done by opening this file in an open source, integrated data viewer called Panoply.

A detailed discussion on the procedures underlying each of the aforementioned modelling stages is given in Sections 2.1 to 2.3.

2.1 Preprocessing: Scaling and Mapping

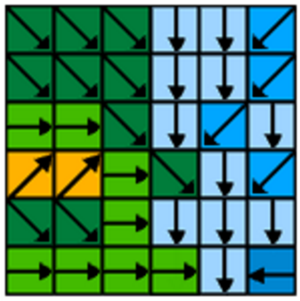
During preprocessing, FA- and DEM data are obtained from the HydroSHEDS [8] website. The DEM and FA data coincide spatially and is provided as binary interlaced (bil) format at a $30''$ $(1/120°)$ resolution.

FA and DEM data are read as two-dimensional arrays. The value of each location in the FA array denotes the number of locations that donate water to it. FA data are derived from DEMs by implementing a method developed by Jenson and Domingue [7] which calculates flow direction as the direction of steepest descent from DEM locations. For clarity, an example of such a process is shown in Figures 1a and 1b. For each (i, j) location in the FA array, the accumulation value is determined by summing the number of surrounding positions that deposit water into the particular location. An example of the relationship between flow direction and an FA is shown in Figures 1b and 1c.

FA and DEM arrays are upscaled to a $0.25°$ resolution. This is done by utilising a Python implementation of the Geospatial Data Abstraction Library (GDAL) that is specifically tailored to read and interpret binary geospatial data. Each bil file is accompanied by a header file that contains information about the number of rows, the number of columns, the coordinates of the upper left corner, and the step size between points. For the example used here, Africa, the upper left coordinates are given by $(-18.99583, 37.99583)$, the step size, for both

78	72	69	71	58	49
74	67	56	49	46	50
69	53	44	37	38	48
64	58	55	22	31	24
68	61	47	21	16	19
74	53	34	12	11	12

(a) Elevations.



(b) Flow directions.

0	0	0	0	0	0
0	1	1	2	2	0
0	3	7	5	4	0
0	0	0	20	0	1
0	0	0	1	24	0
0	2	4	7	35	1

(c) Flow accumulation.

Fig. 1. An example of a flow accumulation derivation from DEM values.

directions, is given as $0.008\overline{3}\,^{\circ}$ and the number of rows and columns are denoted by 8760 and 8880, respectively.

The information within the header files is used to construct an array of latitudes and longitudes. These are subsequently upscaled from 1/120° to 0.25° by extracting every 30^{th} value. The resulting array lengths are then used to split the original accumulation matrix into equally sized sub-matrices, of which each contains 30×30 entries that signify the values within a 0.25° spacial range.

Upscaling of the accumulation matrix is done by following [5] and applying the Network scaling algorithm (NSA) by using a maximum value operator to aggregate each of the 30×30 grid values, i.e. the maximum accumulation value within each of the blocks is kept. An example case of the NSA method is illustrated in Figure 2 for upscaling a 6×6 to a 3×3 grid.

The DEM is upscaled in a similar manner, but, instead of keeping the maximum value, the average of each of the 30×30 elevation values is calculated.

Downscaling of routing values begins by reading runoff values in NetCDF format. Again, a Python library, NetCDF4, is utilised to obtain the data in the correct format. Each file contains the runoff values for one month at six-hourly intervals. The average runoff value for each day of the month is computed in order to have a single daily runoff matrix.

0	0	0	0	0	0
0	1	1	2	2	0
0	3	7	5	4	0
0	0	0	20	0	1
0	0	0	1	24	0
0	2	4	7	35	1

(a) Original flow accumulation.

1	2	2
3	20	4
2	7	35

(b) Upscaled flow accumulation.

Fig. 2. An example of the NSA to scale a 6×6 to a 3×3 grid.

Python's PySAL library [12] is used to construct a k-d tree from the upscaled target latitudes and longitudes. The k-d tree is used to map each runoff value, in equal parts, to its nine nearest neighbours in the FA- and DEM arrays. The original runoff values are of such a nature that it coincides to one of the accumulation value coordinates. For clarity, the refinement procedure is shown in Figure 3.

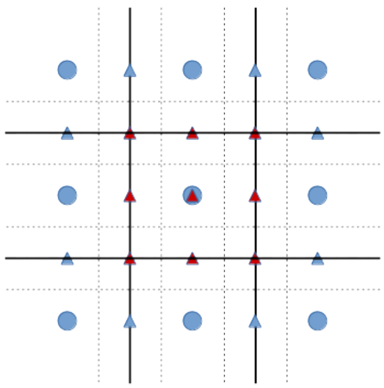


Fig. 3. An example of grid refinement from 0.5° to 0.25° resolution.

The preprocessing stage concludes with runoff-, FA-, and DEM arrays that coincide spatially at a resolution of 0.25°.

2.2 Routing

The routing procedure is initiated by determining out- and inflow locations for each coordinate value on the 0.25° resolution grid. The indices of each point's 8 neighbours are determined and the outflow location is set to the indices of the neighbour with the highest accumulation value. Each grid location can therefore deposit its contents to a single neighbouring location only.

Subsequently, it is determined whether a location will receive water. A two dimensional inflow array is initialised with a size equal to that of the outflow-location matrix and all values are set to False. If a neighbour's outflow-location is equal to the indices of this cell, then the inflow value is changed to True. Once the Truth condition of each point has been determined, the True values are isolated by adding their indices to a list of inflow-locations.

At this stage, it is therefore known to where a point deposits its contents and if it will receive content from neighbouring positions.

The routing procedure starts with initialising a two-dimensional zero-valued water body array. At the start of each day, runoff-volumes are added to the water body array. It is assumed that water is equally distributed within the $0.25° \times 0.25°$ area and the water level within each cell is therefore determined by dividing the runoff volume V by the area of its host cell.

For each cell the water body slope to its outflow location is calculated as

$$S = \frac{\left(z_{DEM}^{host} + z_{WL}^{host}\right) - \left(z_{DEM}^{out} + z_{WL}^{out}\right)}{\Delta x} \ , \tag{1}$$

where S denotes the slope, z_{DEM}^{host} and z_{DEM}^{out}, are the elevations of the host and its outflow location, respectively, whereas, z_{WL}^{host} and z_{WL}^{out} denote the water levels within these locations. The Haversine distance between the cell centres is given by Δx.

The discharge from each cell is given by the Manning equation,

$$Q_{out} = \left(\frac{1}{n}\right) A R^{2/3} \sqrt{S} \ , \tag{2}$$

where Q_{out} is the discharge rate in $m^3 s^{-1}$ and n is Manning's roughness coefficient, which, following [13], is set to 0.025. The width of the domain through which water travels is approximated as the square root of its area, $D = \sqrt{A}$, which allows the hydraulic radius to be expressed as $R = \left(D z_{WL}^{host}\right) / \left(D + 2 z_{WL}^{host}\right)$.

The water volume that is discharged during a time step Δt is calculated as $V_{out} = Q_{out} \Delta t$. The amount of water deposited from any location on the grid within Δt is therefore known at this stage.

To calculate the volume of water V_{in} that a position will receive, the neighbours of each location, recorded in the inflow-locations list, are examined: For each neighbour it is determined whether its outflow corresponds to the inflow-location. If it does, the discharge from this neighbour is added to the inflow volume of the point that is being analysed.

Once the inflow to- and outflow from each point have been calculated, the water body values can be updated as

$$V(t + \Delta t) = V(t) - V(t)_{out} + V(t)_{in} \ . \tag{3}$$

Finally, the updated water body is examined and water is removed from locations that either receive water but does not have an outflow location or are located on the coast.

At the end of each month discharge, water level, water volume, and the total amount of water that is deposited as outflow are written to binary files.

The routing algorithm is shown in Figure 4.

2.3 Post-Processing

For visualisation purposes, the binary output files are converted to a single Network Common Data Form (NetCDF) file. and an open source software, Panoply, is used to view results.

The discharge values can then be mapped back to the CCAM grid and used as input to CCAM's ocean procedure.

3 Results

A simulation was done for the African continent for a 24-year period spanning from 1981/01/01 to 2005/03/01. Courant numbers, discharge rates, and water volumes were recorded for each grid location on a daily basis. To orientate the reader a map, illustrating the actual locations of the largest African rivers, is shown in Figure 5. Courant numbers Co for 2005/03/01 are shown in Figure 3 and were calculated as

$$Co = \frac{|u|\Delta t}{\Delta x} \ , \tag{4}$$

where $|u| = Q_{out}/A$ is the velocity at which water is discharged at time t. Figure 3 shows the discharge rates for 2005/03/01 that were calculated using Eq. 2.

Adherence to mass conservation is checked daily by calculating the total mass of water on land using two methods: Firstly, a landmass water volume is determined by the summation of all values within the water body array, located on land. Secondly, a land mass water volume is calculated by computing the amount of runoff that has entered the system up to this point in time and subtracting the amount of water that has left the system as output. The aforementioned output includes locations that have no outflow but contains water (inland reservoirs) as well as water within cells that are located on the coast.

Values obtained from these two computations are then subtracted and should be close to zero if mass conservation is adhered to. Results for mass conservation are shown in Figure 7.

The solution method used in this model is explicit and therefore the Courant numbers should be significantly smaller than one. Figure 3 shows that the Courant number satisfies the aforementioned CFL condition in that the maximum value is $7.7e-4$. Visual inspection of the discharge rates show that the Congo yields the highest discharge rate followed by the Nile, Niger, and the Zambezi. In reality, the Niger should dominate the Nile and the Zambezi. When comparing the simulation results with the actual positions of the major African rivers, shown in Figure 5, it is concluded that the locations of the largest rivers and their mouths are located at approximately the correct locations.

Output from the original CCAM routing algorithm is shown in Figure 8.

Visual inspection confirms that the updated routing procedure is an improvement on the original since the rivers are now more clearly defined and the severe growth of inland water bodies has been subdued.

The simulation is computationally cheap: A decade was simulated in less than two hours on a single processor of a Lenovo laptop with 7.7 GiB Memory and an Intel Core i7-6500U CPU @ 2.50GHz.

4 Conclusions

A model has been presented for determining the locations and discharges of rivers into the ocean on a $0.25°$ grid and was applied to the African continent. The accuracy of predictions for river- as well as river-mouth locations have been improved upon when the discharge results are compared to the original discharge output from CCAM, shown in Figure 8. It should be noted that there is a slight increase in volume of the total water body over a period of 24 simulation years. Since it is assumed that the density of water is constant, this amounts to a slight increase in mass. However, the total water volume on land for which a maximal increase of 0.017 m^3 is recorded, is of the order $10e+11$. The increase is therefore comparatively small.

The discharge of the Niger relative to that of the Nile and Zambezi has been under-predicted. This could be due to the fact that the simulation is started with a zero water level in that it is forced solely by runoff and the rivers may therefore not have stabilised after 24 simulation years. The discharge discrepancies could also be the result of using an incorrect roughness factor in the Manning equation. Currently, the roughness coefficient is kept constant for all rivers. This may be an unrealistic assumption and the method used for discharge prediction warrants further investigation.

Although the algorithm does not take long to run, parallelisation of the procedure would allow the authors to test its performance for longer simulation periods at finer spatial and temporal scales.

Acknowledgments

The authors wish to thank the CSIR for funding and CSIRO for providing the ESM.

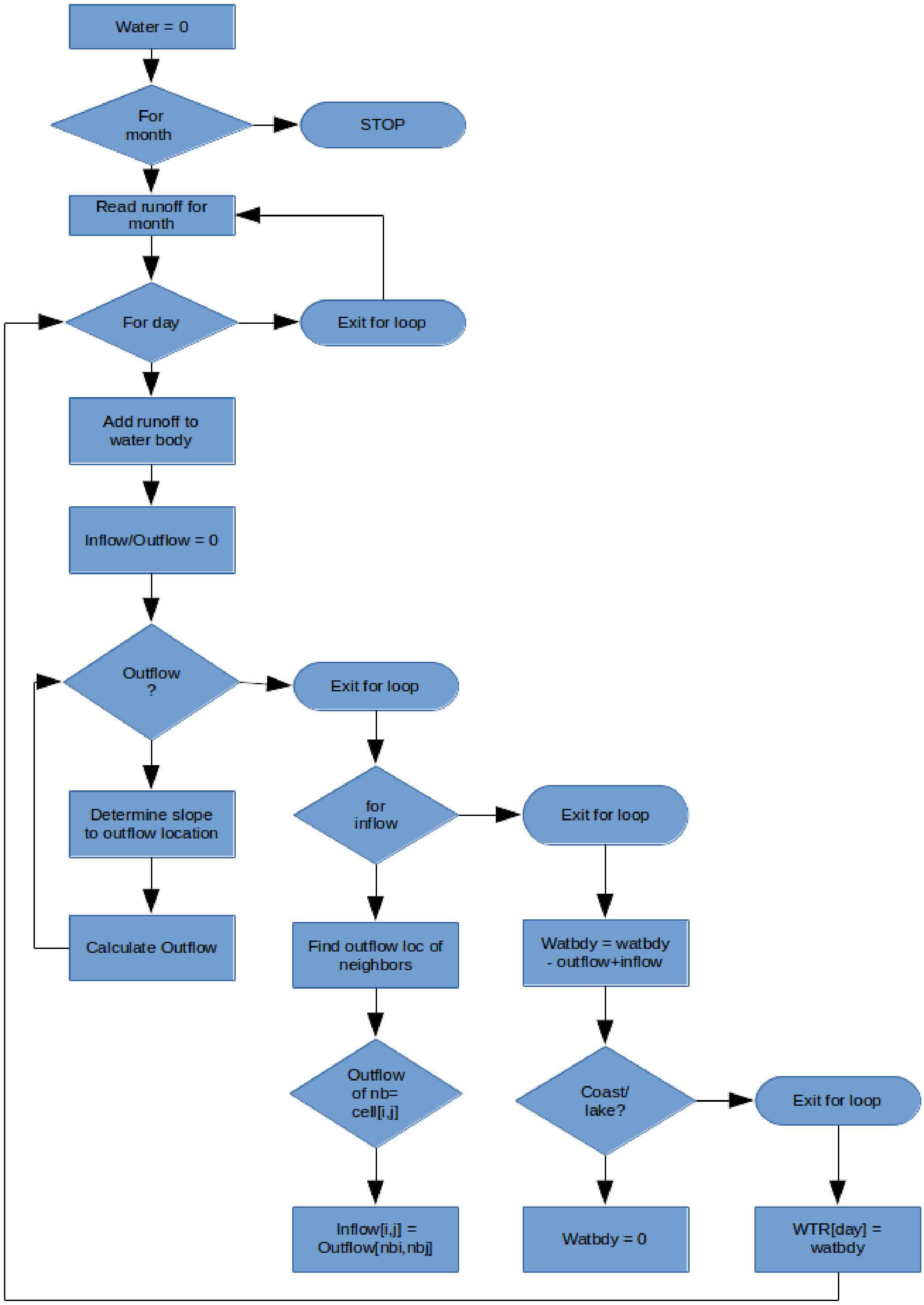
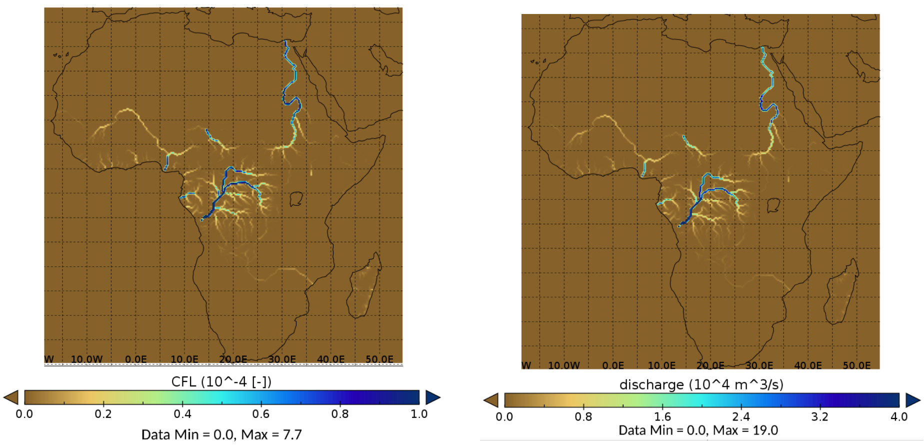


Fig. 4. Flowchart for routing.

Fig. 5. African rivers.



(a) Courant numbers. (b) Discharge rates.

Fig. 6. Courant numbers and discharge rates for 2005/03/01.

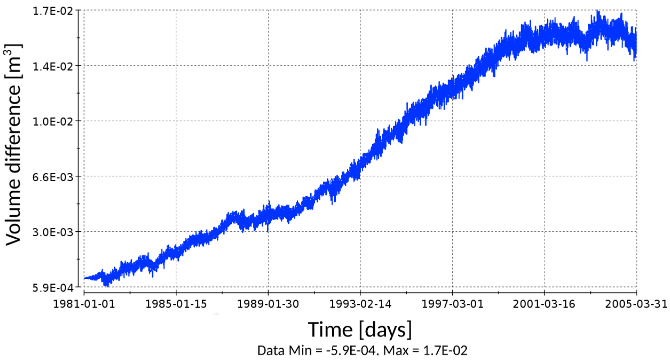


Fig. 7. Total Volume (Mass) conservation check results for 1981/01/01-2005/03/01.

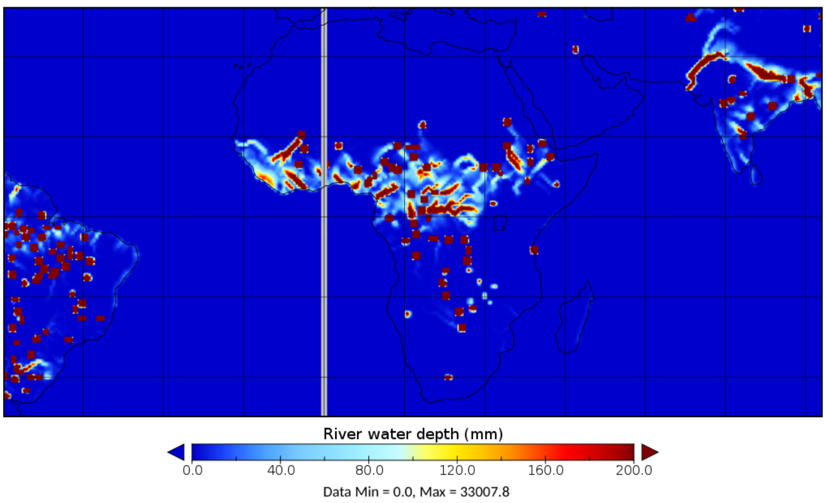


Fig. 8. Original routing output from CCAM simulation.

References

1. Encyclopedia of Sustainability Science and Technology, chap. Coupled Climate and Earth System Models. Springer-Verlag New York (2012)
2. Branstetter, M.: Development of a parallel river transport algorithm and applications to climate studies. Ph.D. thesis, University of Texas, Austin (2001)
3. Dai, A., Trenberth, K.: Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. Journal of Hydrometeorology 3, 660–687 (Jul 2002)
4. David, C., Maidment, D., Niu, G.Y., Yang, Z.L., Habets, F., Eijkhout, V.: River network routing on the nhdplus dataset. Journal of Hydrometeorology 12, 913–934 (Mar 2011)
5. Fekete, B., Vorosmarty, C., Lammers, R.: Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error. Water Resources Research 37, 1955–1967 (Jul 2001)
6. Fekete., B., Vrsmarty, C., Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances. Global Biogeochemical Cycles 16(3), 15–1–15–10 (2002), http://dx.doi.org/10.1029/1999GB001254
7. Jenson, S., Domingue, J.: Extracting topographic structure from digital elevation data for geographic information system analysis. Photogrammetric Engineering and Remote Sensing 54, 1593–1600 (1988)
8. Lehner, B., Verdin, K., Jarvis, A.: New global hydrography derived from spaceborne elevation data. Transactions, AGU 89, 93–94 (2008)
9. McGregor, John L.and Dix, M.R.: An Updated Description of the Conformal-Cubic Atmospheric Model, pp. 51–75. Springer New York, New York, NY (2008), http://dx.doi.org/10.1007/978-0-387-49791-4_4
10. Mizukami, N., Clark, M., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., Viger, R., Markstrom, S., Hay, L., Woods, R., Arnold, J., Brekke, L.: mizuroute version 1: a river network routing tool for a continental domain water resources applications. Geosci. Model Dev 9, 2223–2016 (Jun 2016)
11. Nakamura, M.: Effects of ice albedo and runoff feedbacks on the thermohaline circulation. Journal of Climate 9, 1783–1794 (Aug 1996)
12. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830 (2011)
13. Zhao, Y.: Modeling of River-flow Routing Using a Muskingum-and-Manning Method and Application in Basin of Seine. Ph.D. thesis, Pierre and Marie Curie University, Paris (2006)