# Large-scale watershed delineation strategy

**Luc Rébillout**, PhD, NCCHE, Oxford, MS, luc@ncche.olemiss.edu
**Nuttita Pophet**, PhD, NCCHE, Oxford, MS, nuttita@ncche.olemiss.edu
**Yavuz Ozeren**, PhD, NCCHE, Oxford, MS, yozeren@ncche.olemiss.edu
**Ahmed Sahin**, Grad. Student, NCCHE, Oxford, MS, nazim@ncche.olemiss.edu
**Mohammad Al-Hamdan**, PhD, NCCHE, Oxford, MS, mzalhamd@olemiss.edu
**Paul Smith**, IT Specialist, NCCHE, Oxford, MS, cvpsmith@ncche.olemiss.edu
**Ron Bingner**, PhD, USDA-ARS-NSL, Oxford, MS, ron.bingner@usda.gov
**Glenn Herring**, USDA-ARS-NSL, Oxford, MS, glenn.herring@usda.gov

## Introduction

This study is part of the joint NCCHE-USDA-ARS-NSL partnership under the AIMS (Agricultural Integrated Management Systems) project. The goal of AIMS is to provide a web-based decision support tool for watershed development providing access to high quality watershed numerical simulation management decision tools for experts, stakeholders and non-engineers alike. AIMS hosts the USDA-ARS software suite TopAGNPS and Annualized Agricultural Non-Point Source (AnnAGNPS) for watershed simulation and eventual integration with CCHE1D, NCCHE's channel network simulation solution for water routing. The goal of this study is to prepare continuous and complete input datasets for AIMS which will allow automated input data preparation for AnnAGNPS for any watershed in the United States. AnnAGNPS is a continuous-simulation, watershed- scale, mixed land use, surface-runoff modeling and simulation tool (Momm et al. 2014). For AnnAGNPS simulations, the watershed is subdivided into concentrated flows (reaches) and sub-catchments (cells) using TopAGNPS, a topographic parameterization program utilizing digital elevation models (DEMs) to identify and measure topographic features, define surface drainage spatial extent, and channel network pattern using standard geographic information systems (GIS) assisted flow-routing algorithms. As part of the input preparation efforts, procedures were developed for generating channel and cell data for the entire United States using TopAGNPS.

## Methods

### Source datasets

For this project the two main datasets used are digital elevation maps from the 3D Elevation Program (3DEP,) and local watershed boundaries from the Watershed Boundary Dataset (WBD) both maintained by USGS.

**3D Elevation Program (3DEP):** Accessible at: https://www.usgs.gov/3d-elevation-program. 3DEP is a coordinated effort by USGS National Geospatial Program aiming to provide high quality topographic data all over the country. At the end of fiscal year 2021, "84% of the Nation has available or in progress elevation  data that meet 3DEP specifications for high accuracy and resolution."

**Watershed Boundary Dataset (WBD):** Accessible at: https://www.usgs.gov/national-hydrography/watershed-boundary-dataset. This dataset is a seamless set of hydrologic units. The largest units are the 2-Digit Hydrologic units, containing 4-Digit units, all the way down to 12-Digit hydrologic units hereafter referred to as HUC12s. Each unit may represent all or part of the total drainage area to an outlet point.

## Watershed agglomeration algorithm

The objective of the strategy consists of an optimized grouping of HUC12s into a single hydrologic unit called TopAGNPS Hydrologic Unit Codes (T-HUC). The overall idea is to group HUC12s such that they satisfy two conditions:

1. They have a single outlet
2. Their grouped drainage area is inferior to an arbitrary threshold *T*.

The relevant attributes for this algorithm of the WBD are the area, the type of HUC12s and the connectivity between HUC12s.

HUC12s belong to one of six types: single outlet (most common), closed basin (sink), multiple outlet (e.g. swamp areas), frontal, island, indeterminate. In this study, only the single and multiple outlet units are kept. In practice, multiple outlet units end up being “absorbed” within larger units which typically remove ambiguities. The steps of the algorithm go as follows from that subset of the WBD (which represents the vast majority of the dataset):

1. Generate the directed graph of the selected HUC12s (referred as node within this context) using the “tohuc” attribute. Several disconnected networks are thus generated and each represents an independent watershed.
2. Each node is given an additional attribute of its own area summed with the cumulated area of its ancestors (i.e. all the nodes upstream).
3. A sub-network is identified of all the nodes with a cumulative area > *T*.
4. The sub-network is traversed in a depth-first fashion and the nodes are listed in a stack.
5. For each node in the stack, its predecessors (i.e. the immediate upstream contributors) are identified and all their respective ancestors are grouped in a T-HUC which is then removed from the main connected network. The cumulative area of each THUC is subtracted to all the downstream nodes.
6. Steps 3–5 are repeated until the entirety of the network is processed. If individual nodes still have an area > *T* they are considered their own T-HUC.

Each T-HUC is thus defined by the geometry of the union of all the contained HUC12s, its area, the HUC12 that is the outlet of the T-HUC, the HUC12 into which it drains (if it exists), the list of HUC12s “foreign inflows” (if they exist) that are HUC12 draining into the T-HUC but do not belong to it, and the area of its bounding box.

A final step 7 consists of merging the T-HUCs if one flows into another *and* they lie in the same bounding box: indeed, the condition of single outlet for the entire T-HUC is preserved and it does not present additional computational effort from TopAGNPS as the software processes the DEM from the bounding box of the T-HUC thus removing the overall number of T-HUCs to process. This algorithm was written in Python relying mainly on the networkx and geopandas libraries.

## Automated watershed delineation with TopAGNPS

The T-HUC boundaries were used as a guide to collect DEMs from the 3DEP service. The chosen resolution was 10-m cells with a buffer distance of 500 meters around the bounding box of each T-HUC. The python toolbox py3dep from the HyRiver software stack developed by Chegini et al. (2021) was used for retrieving DEMs and processing them in an appropriate format for TopAGNPS.

To carry out a complete watershed delineation, the location of the outlet is required. This information is often not available in advance. To address this issue, a pre-processing step was carried out using TopAGNPS to compute an accumulated flow raster map. By masking out all raster values outside of the original T-HUC boundary, the outlet was identified as the raster with the highest accumulated flow value within the boundary. Once the location of the raster was known, the full processing of the watershed could continue. A critical source area (CSA) value of 10 ha and a minimum source channel length (MSCL) value of 250 meters were used for this process to ensure cell areas are sufficiently small enough to capture the spatial variability of climate, topography, soils and land management.

When a delineation is completed, quality control is performed first by computing the fraction of the original T-HUC boundary that is indeed delineated and secondly how much area beyond the boundary is also covered and whether or not the delineated cells touch the edges of the DEM which might suggest that the T-HUC is not a watershed but has external inflows to be mindful of for further analysis.

# Results

## Watershed aggregation

The outlet of 4,800 T-HUCs obtained using the watershed aggregation algorithm described above and a drainage area threshold $T$ = 40,000 km$^2$ is shown in Figure 1.

The T-HUCs tend to have a larger area away from coastal and lake areas; this is explained by the single outlet condition of the algorithm and the fact that along coasts and lakes, multiple rivers pour into the body of water and each one must belong to a different T-HUC. A density of smaller T-HUCs is also observed coinciding with The “Great Basin” in the western portion of the country which is the largest endorheic watershed (closed basin) in North America spanning portions of Nevada, Utah, California, Baja California, Idaho, Oregon, and Wyoming. Many single outlet HUC12s eventually pour into a closed basin HUC12 which cuts them off from the rest of the network.

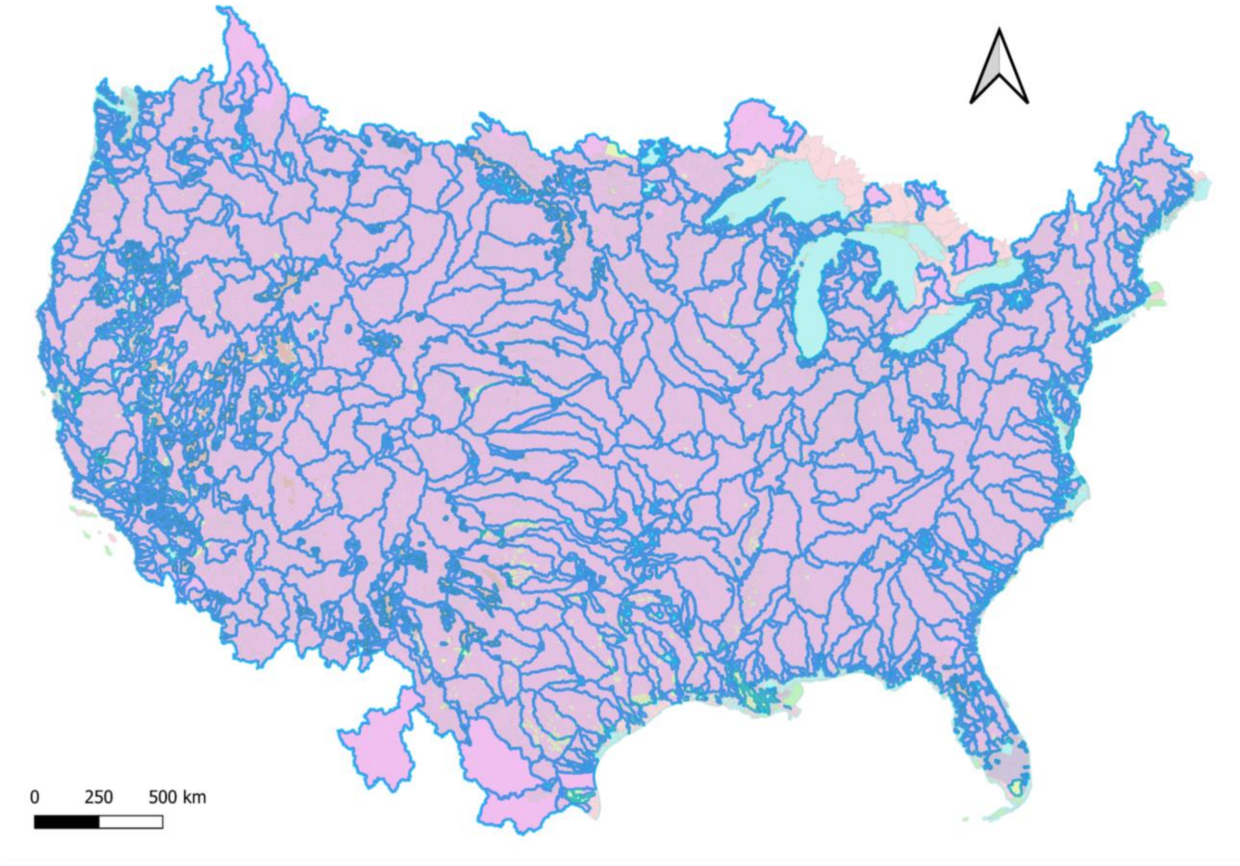

**Figure 1.** Collection of 4800 T-HUCs (bold blue lines, CONUS shown only) overlayed over WBD. Color codes for HUC12s are pink: single outlet; light green: multiple outlet; dark green: island; salmon: frontal; brown: closed basin; purple: indeterminate.

## Challenges and Issues

While this process works generally well, some pathological cases can arise particularly when the DEM features large flat areas where a flow path can not effectively be computed by the software. Another example of a pathological case would be a DEM traversed with a single river that is locally divided in two branches: in this case, the outlet will belong to one of those branches and only one side of the river will be delineated and the other side ignored (Figure 2). Finally, this analysis revealed that the connectivity between HUC12s as described by the WBD is not necessarily reflected in the final network as delineated by TopAGNPS due to the fact that the DEMs obtained from the 3DEP service are not hydraulically enforced.

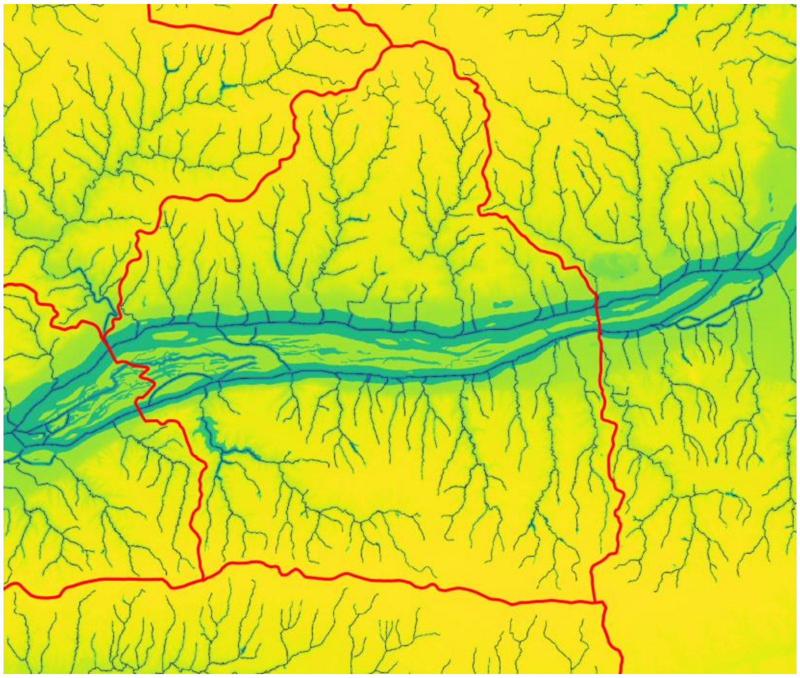

**Figure 2.** Example of a pathological case with a main river subdivided in two branches

## Conclusion

In this study a large-scale nation-wide watershed delineation strategy of the conterminous United States using TopAGNPS (USDA-ARS) is presented. An algorithm to efficiently group HUC12 hydrological units from the Watershed Boundary Dataset (WBD) was developed to optimize the use of computer resources and maximize the area covered for every delineation. A collection of T-HUCs (TopAGNPS Hydrologic Unit Codes) which consist of groups of HUC12s was then used as the basis to collect digital elevation maps (DEM) from USGS' 3D Elevation Program (3DEP). A method to automatically identify a T-HUC outlet was defined which allowed for a full automation of the watershed delineation process. While some challenges remain this process will have applications in the development of a web-based decision support system called AIMS.

## Acknowledgments

The research was supported by the USDA-ARS National Sedimentation Laboratory (NSL) and The University of Mississippi (UM) grant 58-6060-8-008 which is gratefully acknowledged.

## References

Chegini, T., Li, H.-Y., and Leung, L. 2021 "HyRiver: Hydroclimate Data Retriever", Journal of Open Source Software, 6(66):1–3, doi: 10.21105/joss.03175

Momm, H. G., Bingner, R., Yuan, Y., Locke, M., and Wells, R. 2014. "Spatial characterization of riparian buffer effects on sediment loads from watershed systems". Journal of Environmental Quality, 43(5): 1736-1753.

Momm, H., Bingner R., Wells, R., Moore, K., and Herring, G. 2021 “Integrated Technology for Evaluation and Assessment of Multi-Scale Hydrological Systems in Managing Nonpoint Source Pollution.” Water, 13(6): 842.