Reducing Horton-Strahler Stream Order Can Enhance Flood Inundation Mapping Skill with Applications for the U.S. National Water Model

Fernando Aristizabal[1,2,3], Fernando Salas[3], Gregory Petrochenkov[4], Trevor Grout[1,3], Brian Avant[1,3], Bradford Bates[1,3], Ryan Spies[1,3], Nick Chadwick[3], Zachary Wills[3,5], Jasmeet Judge[2]

[1]Lynker, Leesburg, VA, USA

[2]Center for Remote Sensing, Agricultural and Biological Engineering Department, University of Florida, Gainesville, FL, USA

[3]National Water Center, Office of Water Prediction, National Weather Service, National Oceanic and Atmospheric Administration, Tuscaloosa, AL, USA

[4]New York Water Science Center, Hydrologic Applied Innovations Lab, United States Geological Survey, Troy, NY, USA

[5]Cooperative Institute for Satellite Earth System Studies, University of Maryland, College Park, MD, USA

Key Points:

- Flood inundation maps derived from Height Above Nearest Drainage (HAND) are subject to limitations at river junctions.
- A means of resolving this limitation is provided by reducing HAND processing units to level paths of unit stream order.
- Changing the scale of the stream network for HAND processing affects the stage-discharge relationship and leads to higher skill inundation.

Corresponding author: Fernando Aristizabal, `fernando.aristizabal@noaa.gov`

Abstract

The National Water Model (NWM) currently requires the post-processing of forecast discharges to produce forecast flood inundation maps (FIM) for protecting life and property. Height Above Nearest Drainage (HAND), a drainage normalizing terrain index, is worthy of producing high-resolution FIMs at large spatial scales and frequent time steps using reach-averaged synthetic rating curves. However, HAND based FIMs suffer from a known limitation caused by independent catchments that lack the ability to cross catchment boundaries and ridgelines. To counter this constraint, a version of HAND known as Generalized Mainstems (GMS) is proposed that reduces the Horton-Strahler stream order of the stream network. GMS contains all segments within the NWM stream network but instead of deriving HAND by accounting for all river segments at once, it is derived independently at the level path (LP) scale. LPs are unique identifiers propagated upstream from a sub-basin's outlet along the direction of maximum flow distance and repeated recursively until all segments are assigned LP identifiers. These FIMs are then mosaiced together, effectively turning the stream network into discrete groups of homogenous unit stream order by removing the influence of neighboring tributaries. Improvement in mapping skill is observed when compared to HEC-RAS 1D models by significantly reducing false negatives at river junctions. A more marginal reduction in the false alarm rate is also observed due to a bias introduced in the stage-discharge relationship by increasing the size of the catchments.

Plain Language Summary

Flooding is one of the most impactful natural disasters on life and property. The United States National Water Model (NWM) provides flood forecasts for the entire country so that adequate warnings can be raised to the public to enable safe evacuations and protective measures. In order to convert forecasted flow rates from the NWM to flood inundation maps (FIM), a model is used that converts elevation data from height above mean sea-level to height above the nearest river bottom. This model known as Height Above Nearest Drainage (HAND) suffers from issues in mapping performance where rivers converge. We developed a technique that mitigates these errors by removing consideration for neighboring tributaries in the relative elevation computation process. We compared these HAND derived FIMs to maps from more realistic models and found improvement in mapping performance.

1 Introduction

Flooding is one of the most significant natural disasters in the United States (US) affecting both the loss of life and property. In 2017 and 2019, river and flash flooding combined represented the leading cause of death and the second leading cause in 2018 among all natural disasters in the US (National Weather Service, 2020b, 2019, 2018). More than an average of 104 deaths per year are attributed to flood events from the 10 year period ending in 2019 (National Weather Service, 2020a). With respect to property damages, river and flash flooding have contributed to 60.7, 1.6, and 3.7 billion non-inflation adjusted US dollars in the annual periods of 2017 to 2019, respectively (National Weather Service, 2020b, 2019, 2018) with the large spike in 2017 attributed to the Hurricane Harvey event along the Gulf coast. Trends related to flood damages and fatalities have been steadily increasing over recent decades. (Mallakpour & Villarini, 2015; Downton et al., 2005; Kunkel et al., 1999; Pielke Jr & Downton, 2000; Corringham & Cayan, 2019). Some are expecting that the hydrologic cycle will intensify which will lead to more extreme precipitation in some areas along with a greater risk of flooding (Tabari, 2020; Milly et al., 2002; Wing et al., 2018). Increasing trends in frequency and risk are not uniform across spatial regions with work by Slater and Villarini (2016) indicating that trends are increasing across the US Midwest/Great Lakes region while decreasing in coastal Southeast, Southwest and California.

1.1 Operational Forecasting

Operational flood forecasting systems are primary tools in developing accurate forecasts for public awareness prior to life or property damaging events occur. One of these operational systems is the Advanced Hydrologic Prediction System (AHPS) maintained by the National Oceanic Atmospheric Administration (NOAA) National Weather Service (NWS) with thousands of forecasting points across the US at typically short forecast horizons of 24 or 72 hours (McEnery et al., 2005). AHPS provides forecasting services in the form of ensemble stream flows at more than 3,600 locations and flood inundation maps (FIM) at more than 150 of those points shown in Figure 1. Additionally, two forecasting networks relevant to the National Water Model (NWM) which will be introduced in Section 1.2 are rendered in Figure 1. AHPS implements a series of advances including model calibration techniques (Z. Zhang, 2003; Hogue et al., 2003; Duan, 2003; Gupta et al., 2003; Parada et al., 2003), distributed modeling approaches (Reed et al.,

2004; Koren et al., 2004; Duan & Schaake, 2002), ensemble forecasting (Day, 1985; Seo et al., 2000; Mullusky et al., 2002; Herr et al., 2002), enhanced data analysis procedures (McEnery et al., 2005), flood-forecasting inundation maps (Cajina et al., 2002), hydraulic routing models (Fread, 1973; Cajina et al., 2002), and multi-sensor precipitation techniques (Breidenbach et al., 1999; Kondragunta, 2001; Seo & Breidenbach, 2002; Bonnin, 1996). On an approximate basis, there is only one forecast point every 1,450 km of river and one forecast point with FIM every 29,000 km. Despite the AHPS advances in operational flood forecasting, it lacks sufficient spatial coverage and long-range forecast horizons to address the increasingly complex water challenges facing the US.

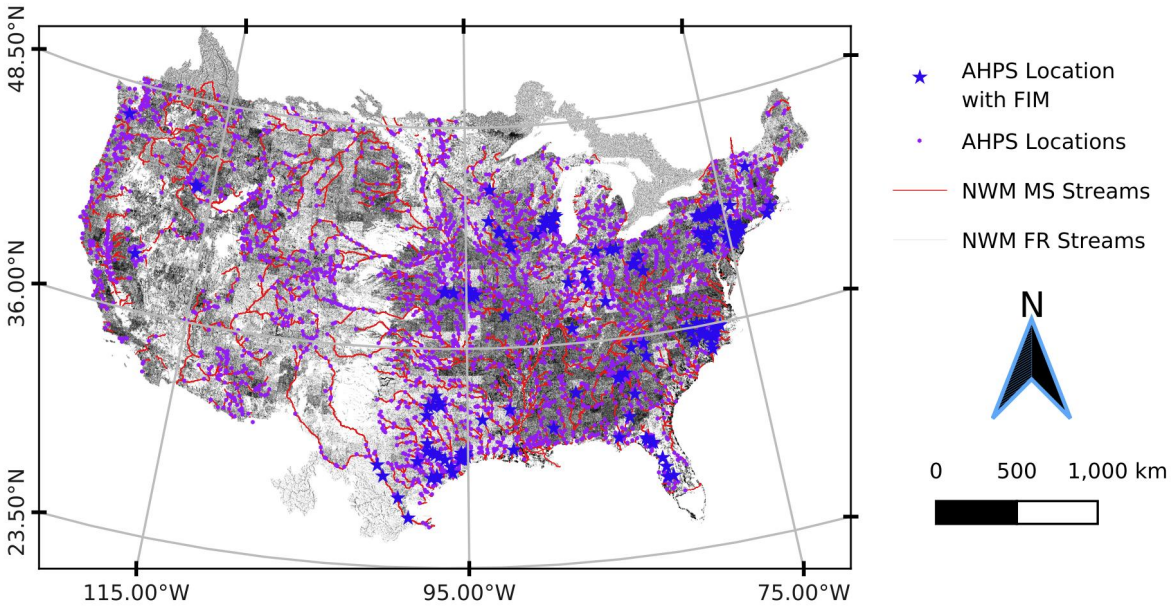


Figure 1. Forecast points with and without FIM in United States' Advanced Hydrologic Prediction System. Also show are the National Water Model stream networks at the full resolution (FR) and Mainstems (MS) resolution.

1.2 National Water Model

Additional work is required to fill-in the gaps that the AHPS leaves in terms of spatial and temporal coverage. In response to growing stakeholder demand for enhanced and

integrated water resource forecasts, the Office of Water Prediction (OWP) at the National Water Center (NWC) along with its partners has developed and implemented operationally the National Water Model (NWM) which is a version of the Weather Research and Forecast Hydrologic Model (WRF-Hydro) (Gochis et al., 2018; Cosgrove et al., 2019). The NWM forecasts river discharges at more than 2.7 million forecast points at a variety of time horizons including some medium (10 day) and long (30 day) range forecast horizons. The NWM enhances the spatial and temporal domain of the current AHPS capabilities operated at the 13 River Forecast Centers (RFC) in areas known as 'hydro-blind'. As a complement to the operational NWM, OWP has also developed a configuration of the NWM that extends RFC forecasts downstream by assimilating and routing forecast streamflows to the next downstream AHPS forecast point. This configuration of the NWM is used to enhance forecasting skill by leveraging best available regional-scale data. The river network upon which this special configuration operates is herein referred to as the Main-Stem (MS) modeling stream network. Figure 1.2 shows the Full-Resolution (FR) modeling stream network as well as the MS network. The MS network contains roughly 120 thousand forecasting points or roughly 4.4% of the reaches of the FR stream network.

The National Hydrography Dataset Plus (NHDPlus) V2.1 is the basis for the hydrofabric in the NWM due to its comprehensive use with the hydrologic communities' stakeholders (McKay et al., 2012). The term hydrofabric is used within the NWM jargon to describe the subset of hydrography composed of the geospatial datasets required for hydrologic modeling including but not limited to stream networks, catchments, channel properties, and elevation data. The Muskingam-Cunge routing method is used within the NWM to reduce computational requirements of a continental scale model (Bedient et al., 2008; Ponce & Changanti, 1994; Gochis et al., 2018). Muskingam-Cunge routing scheme has been demonstrated by Cunge (1969) to be equivalent to the convective-diffusive wave method without consideration to wave dampening. As a result of high computational costs and large spatial domains, the need for high-resolution FIM at 10 m or better requires additional post-processing from the principal output of the NWM which is forecast river discharges at the reach scale. The Height Above Nearest Drainage (HAND) terrain model is one such technique that can be used, along with synthetic rating curves (SRC), to convert riverine discharges to stages then finally to inundation extents.

1.3 Height Above Nearest Drainage

HAND normalizes topography along the nearest drainage path and its been demonstrated to be a good proxy and indicator of a series of important environmental conditions including soil environments, landscape classes, soil gravitational potentials, geomorphologies, soil moisture, and groundwater dynamics (Rennó et al., 2008; A. Nobre et al., 2011). A. D. Nobre et al. (2016) showed evidence for utilizing the drainage normalizing HAND dataset as a proxy for flood potential to make static flood inundation maps from known stages. A core assumption made for HAND based FIM is enforcing drainage across the entire area of interest which requires significant digital elevation maps (DEM) manipulations to make a reality. The terrain index has even gone on to provide additional utility in the observation of riverine flood inundation mapping from remote sensing especially in areas of high electromagnetic interference such as vegetated and anthropogenic areas (Aristizabal et al., 2020; Shastry et al., 2019; Huang et al., 2017; Twele et al., 2016). Zheng, Tarboton, et al. (2018) developed a methodology for determining stage-discharge relationships known as synthetic rating curves (SRC) by sampling reach-averaged parameters from HAND datasets and inputting into the Manning's equation (Gauckler, 1867; Manning et al., 1890). This collection of methods, coupling HAND with SRCs, have been experimented with and compared to other sources of FIM including engineering scale models, in-situ observation, and remote sensing based observation with solid results in large spatial scale applications (Godbout et al., 2019; Johnson et al., 2019; Garousi-Nejad et al., 2019; A. D. Nobre et al., 2016; Afshari et al., 2018; Zheng, Maidment, et al., 2018; Teng et al., 2015, 2017; J. Zhang et al., 2018).

1.4 HAND Implementations

Due to significant advances in high-performance computing (HPC) and large scale high-resolution DEMs such as the 3D Elevation Program (3DEP) seamless at the 1/3 arc-second (10 m) scale, HAND has been implemented into software for large-scale, continental computation. As part of the OWP's Innovators Program and NWC's Summer Institute, the National Flood Interoperability Experiment (NFIE) generated FIM hydrofabric (will be used interchangeably with the datasets produced by HAND) rapidly on a high-performance computer (HPC) (Maidment, 2017; Y. Y. Liu et al., 2016). NFIE used open-source dependencies including the Terrain Analysis Using Digital Elevation Models (TauDEM) (Tarboton, 2005) and the Geospatial Data Abstraction Library (GDAL)

(Warmerdam, 2008) to compute HAND for the Continental United States (CONUS) at 331 Hydrologic Unit Code (HUC) 6 processing units in 1.34 CPU years. By allocating 31 nodes at 20 cores per for a total of 620 available cores to the overall operation, it enabled the production to finish up in 36 hours consuming 3.2TB of peak memory and 5TB of total disk space. Originally, NFIE utilized the NHD Medium Resolution (MR) to etch or burn flowlines prior to further conditioning but more recent work has advanced this to the more current NHDPlus High Resolution (HR) (Y. Liu et al., 2020). The original NFIE dataset was employed by the NWC as an unofficial demonstrations to produce forecast FIM from the NWM for additional guidance in hydro-blind regions. Further work by Djokic (2019), implemented a series of improvements to HAND including equidistant reaches, updates to use with NHDPlusHR hydrography, and AGREE-DEM reconditioning (Hellweger & Maidment, 1997) into an ESRI Arc-Hydro workflow with use in ArcGIS. More notably the software added the ability to derive drainage potentials on both the NWM FR and MS stream networks which leverages the lower drainage density and Horton-Strahler stream order of the MS network.

Related to these efforts, the USGS has invested in relative elevation HAND-like methods via work in the GIS Flood Tool (GFT) that also uses synthetic rating curves with cross-sections for stage-discharge relationships (Verdin et al., 2016). Additional investment by Petrochenkov (2020) was able to successfully scale this approach by transitioning the method to open-source Python source code (PyGFT) and implementing novel interpolation methods to help address some of the catchment boundary discontinuities discussed more in this paper. In addition to the domestic work done in the US, some studies have expanded upon HAND to cover global domains at 30 m resolutions (Yamazaki et al., 2019; Donchyts et al., 2016).

1.5 OWP FIM

Many of those in the academic community assessing HAND's efficacy for producing FIM have noted opportunities for improvement. Godbout et al. (2019) found how reach length and slope are important parameters for maximizing mapping skill with the moderate values performing best. The co-linearity of reach length and slope led Godbout et al. (2019) to propose that reaches of extreme lengths performed worse because of the extreme slope values, a parameter directly represented in Manning's equation. Issues with the reach-average approaches have been documented in Tuozzolo et al. (2019) where large

within reach variance of the roughness Manning's n coefficient have been observed. Furthermore, Garousi-Nejad et al. (2019) noted improvements to mapping efficacy by conditioning monotonically decreasing thalweg elevations, adjusting the Manning's n roughness coefficient, and using higher resolution (3 m) Digital Elevation Model's (DEM) derived from light detection and ranging (Lidar). The use of higher resolution DEMs in that study was motivated by previous work with Lidar DEMs and least-cost thalweg derivations (Zheng, Maidment, et al., 2018). Further work by Johnson et al. (2019) noted the general under-prediction of HAND and suggested tuning the Manning's n parameter to better align SRC's with observations. Additionally, the sensitivity to low topographic relief and channel slope have been observed (Johnson et al., 2019; Godbout et al., 2019). Most notably, HAND has been shown to demonstrate sensitivity to drainage network density known colloquially as the catchment boundary problem (J. Zhang et al., 2018; McGehee et al., 2016; Li et al., 2020; A. D. Nobre et al., 2016). This sensitivity is noted at higher order streams with large flows and low Froude numbers and is manifested by a constriction in the inundation extents at the junction location. These findings suggest improvements to HAND are required that utilize advanced computational algorithms and software to compute a FIM hydrofabric required for producing continental-scale FIM.

With all of the latest developments in the realm of continental FIM (CFIM) noted in the previous paragraph, a fast, open-source framework for accelerating the research to operations pipeline is required. Here we introduce OWP FIM that utilizes a few of the latest techniques in HAND based FIM oriented for use with the NWM in continental scale operational forecasting settings. This framework is open-source and modular enabling and accelerating the research to operations development pipeline. Automated evaluation tools and processed test case data enable the rapid testing and evaluation of new techniques in consistent contexts. In addition to developed tooling, we introduce research demonstrating how FIM performance skill with HAND can be improved by reducing Horton-Strahler stream orders (Horton, 1945; Strahler, 1952, 1952) of the stream networks. Previous authors dating back to the first HAND for FIM work by A. D. Nobre et al. (2016) have noted a sensitivity of mapping skill to stream threshold which serves as a proxy for stream density and the maximum Horton-Strahler stream order (or simply stream order) of the processing unit employed (J. Zhang et al., 2018; McGehee et al., 2016; Li et al., 2020). Here we demonstrate how reducing a HAND processing unit's stream network to a singular stream order discretized by level paths, can enhance FIM

skill. This capability will be referred to as Generalized Mainstems (GMS). The following methods and results describe the work in more detail and demonstrate its efficacy in producing enhanced FIM for the NWM with applications.

2 Materials and Methods

OWP FIM is a fully operational pipeline of software tools to help acquire datasets, cache hydrofabrics, produce FIMs, and evaluate results.

2.1 Software Dependencies and Architecture

OWP FIM exclusively utilizes free and open source software dependencies including Python 3, GDAL, TauDEM, Geographic Resource Analysis Support System (GRASS), GNU Parallel, and MPICH (Python Core Team, 2019; GDAL/OGR contributors, 2020; Tarboton, 2005; GRASS Development Team, 2020; Tange, 2015; Amer et al., 2021). Within the Python 3 ecosystem, many common packages are employed including but not limited to RichDEM, GeoPandas, Rasterio, Rasterstats, and Numba (Barnes, 2018; Jordahl, 2014; Lam et al., 2015). To simplify setup and enhance portability across host operating systems OWP FIM packages all dependencies up in a Docker image (`https://docs.docker.com/engine/install/`). A user only needs to install Docker on their host machine and build the image from the provided recipe. Source code is made available for this project on GitHub (Aristizabal et al., 2022b).

The pipeline is discretized into key areas that a user can interact with to reproduce the results of this study. Preprocessing acquires and prepares datasets for production of the FIM hydrofabric. The FIM hydrofabric is defined as the datasets required to make an inundation map from discharges including the relative elevation model (REM) or HAND grid, the catchments in vector and raster form, and the hydro-table (contains synthetic rating curves and cross-walk information). Functionality is included to turn FIM hydrofabric and streamflows into inundation maps represented in both binary and depths. Lastly, a test suite includes means of calculating evaluation metrics compared to a variety of preprocessed test case data. A user should visit the Readme.md page on GitHub for more information on how to acquire the datasets and reproduce the pipeline.

2.2 Datasets

Data sources used within OWP FIM are publicly available from a variety of government sources including the USGS, NWC, Federal Emergency Management Agency (FEMA), and US Army Core of Engineers (USACE) to enhance reproducibility and collaboration among government, academia, and industry. Instructions for accessing data

are provided on the project's GitHub page via an Amazon Web Services (AWS) S3 bucket furnished by Earth Science Information Partners (ESIP) (Aristizabal et al., 2022a). The National Hydrography Dataset Plus High Resolution (NHDPlusHR) Beta Version is the latest hydrography dataset used for land surface hydrologic modeling in the US (Moore et al., 2019). It is used in conjunction with the hydrofabric of the NWM V2.1 to help define flowlines for OWP FIM while the NWM hydrofabric is also used to define reservoirs for exclusion and catchments to cross-walk to for forecasting purposes. For enforcing levee data into the DEMs, the USACE National Levee Database (NLD) is used to burn feature elevations (ENGINEERS, 2016). Since NHDPlusHR datasets extend beyond land borders into sea and Great Lake regions, we used the land-sea border from OpenStreetMap (OSM) and the land-lake border from Great Lakes Hydrography Dataset (GLHD) to exclude those areas from production of FIMs (OpenStreetMap contributors, 2017; Great Lakes Aquatic Habitat Framework contributors, 2020). Additionally, the Base Level Engineering (BLE) datasets within FEMA Region 6 spanning parts of 9 states including Colorado, New Mexico, Texas, Oklahoma, Kansas, Arkansas, Louisiana, Missouri and Mississippi at two recurrence intervals, 1% (100 year) and 0.2% (500 year), are used as validation in this study and furnished by the Interagency Flood Risk Management (InFRM) consortium (*Base Level Engineering (BLE) Tools and Resources*, 2021; *Estimated Base Flood Elevation (estBFE) Viewer*, 2021). These BLE datasets are provided at the watershed scale (HUC8) utilizing best available simulations and DEMs. The full input datasets presented by source are listed in Table 1. Areas with all the required data (from the NWM and the USGS) are labeled as the FIM domain which includes 2,188 HUC8s for the FR and GMS networks and 1,604 HUC8s for the MS method. These methods will be explained more later. An enhancement of OWP FIM over previous HAND based FIM versions is the support for Hawaii and Puerto Rico which the NWM V2.1 will cover.

2.3 Hydro-conditioning

The DEM is subject to a series of hydro-conditioning procedures to enhance its suitability for riverine flood inundation mapping. These techniques are specific for making OWP FIM and differ from the conditioning methods used by the NHDPlusHR Beta (Moore et al., 2019). HAND inherently requires all areas eligible for inundation to drain to the designated drainage network so DEMs must undergo significant manipulation to make this the case. In other words, all areas within a given processing unit for HAND must

Table 1. Data sources, names, and descriptions.

Source	Name	Description
USGS	NHDPlusHR BurnLineEvents	Stream lines used by NHDPlus HR for hydro-enforcement
USGS	NHDPlusHR Value-Added Attributes	Database of additional attributes associated with the BurnLineEvents that enhance navigation, analysis, and display
USGS	NHDPlusHR DEM	DEM used for NHDPlus HR at 1/3 arc-second (10 m) spatial resolution and vertical units in centimeters
NOAA-OWP	NWM Streams	Stream network center lines used by NWM for routing and forecasting.
NOAA-OWP	NWM Catchments	Surface drainage area corresponding to each reach in the NWM.
NOAA-OWP	NWM Waterbodies	Waterbodies considered by the NWM as reservoirs or lakes.
USACE	NLD	Levee database of locations and elevations
OSM	Land-Sea Border	Border of land and sea.
GLHD	Land-Great Lakes Border	Border of land and Great Lakes.
InFRM	Cross-Sections	HEC-RAS 1-D cross-sections used for modeling in BLE datasets. Includes discharges for 1% and 0.2% recurrence interval events.
InFRM	Flood Inundation Extents	Inundation depths produced by InFRM BLE HEC-RAS 1D for 1% and 0.2% recurrence interval events.

have monotonically decreasing elevations if we wish to have them be eligble for flooding. Hydro-conditioning is implemented to obtain many objectives including enforcing the location of hydrologically relevant features such as flowlines, lakes, or drainage divides whether natural or anthropogenic. It can also be used to simulate more accurate bathymetry which is not accounted for in the 10 m DEM (Gesch et al., 2002).

Specifically within the context of OWP FIM, the hydro-conditioning operations that take place in sequential order are presented. Prior to any hydro-conditioning, all input datasets must be subset from their original spatial domain scales into the processing units of size HUC8. The subsetting is done by spatial query for the cases of the levees, DEM, and NWM hydrofabric while the NHDPlusHR BurnLineEvents are subset via attribute query for the given reach code's membership in the processing unit. Hydro-conditioning raster operations take place on buffered boundary definitions to avoid edge contamination and effects (Lindsay & Seibert, 2013).

2.3.1 Stream Network Enforcement

The location of the stream network is enforced to ensure general agreement with established stream networks. The NHDPlusHR Beta Burnline Event layer is used to enforce stream locations in the NHDPlusHR workflow so it is also used here for hydro-enforcement (Moore et al., 2019). However, to better match the drainage density of the NWM V2.1 stream network which is based on the NHDPlus Medium Resolution, the Burnline Events are pruned utilizing a nearest neighbor search around the NWM flowlines. For every NWM headwater segment a headwater point is derived and linearly interpolated to the nearest Burnline Event segment. Burnline Event headwater segments are split at the adjusted headwater point to match NWM flowlines. The resulting pruned NHD stream network is what gets hydro-enforced in subsequent operations. This procedure is best illustrated in Figure 2 which shows that the pruned NHD network corresponds to the full density NHD network at NWM V2.1 headwater locations only. Additionally, the NHDPlusHR pruned headwaters are later used for seeding a new FIM drainage network that best agrees with the DEM after all hydro-conditioning takes place. This results in a stream network that has the same density as the NWM V2.1 flowline network but utilizes the locations of the NHDPlusHR Beta BurnLineEvents.

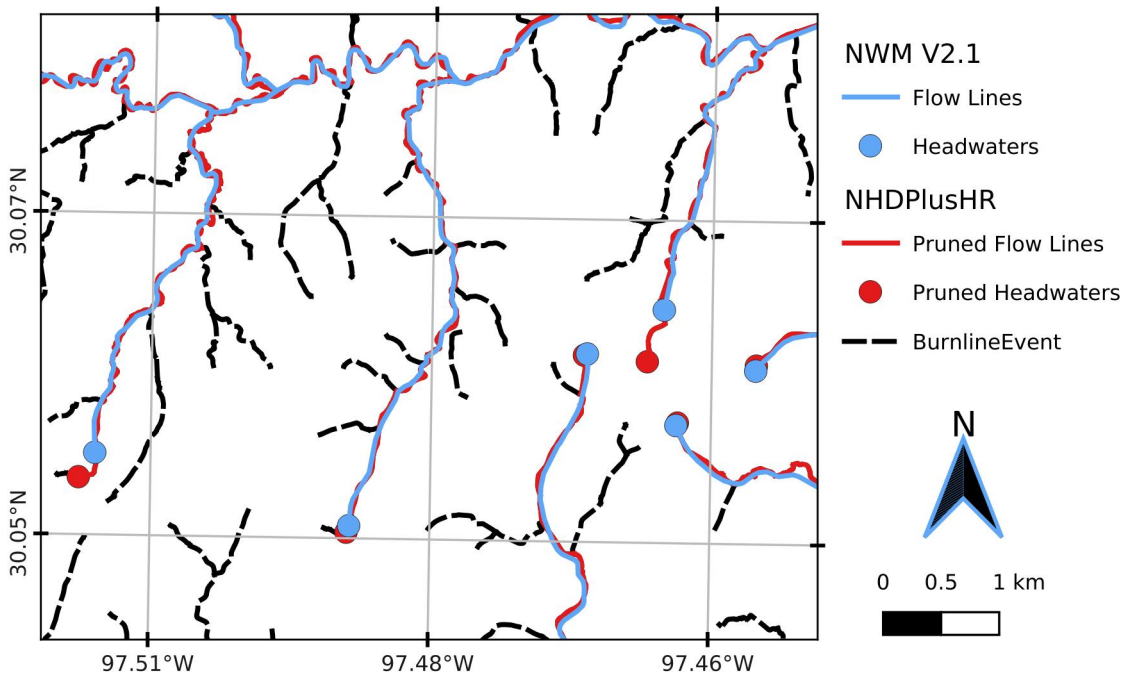


Figure 2. Pruning of NHDPlus HR Beta Burnline Events (dotted black) to NWM V2.1 stream density (blue) using nearest neighbor selection and linear interpolation. Resulting stream network (red) matches the drainage density of NWM V2.1 while corresponding spatially with the NHDPlusHR Burnline Events.

The pruned stream network is then utilized to hydro-enforce the DEM with a methodology developed by Hellweger and Maidment (1997) known as the AGREE DEM Surface Reconditioning System. The AGREE algorithm seeks to burn artificially deep thalweg elevations by a uniform value known as sharp drop. The modification continues by excavating an area of a given buffer distance from the thalweg by a depth proportional to the distance from the channel given by the smooth drop. The resulting enforcement of the thalweg and general bathymetric region results in a cross-section resembling a trapezoidal shape with a significantly lower elevation along the thalweg line only. In total, the AGREE algorithm requires three parameters including the buffer distance, smooth drop, and sharp drop. Using the AGREE method as opposed to simple thalweg burning techniques helps prevent distortions in the delineation of streams as well as the catchment boundaries (W. Saunders & Maidment, 1995; W. K. Saunders & Maidment, 1996; Mizgalewicz & Maidment, 1996; Hellweger & Maidment, 1997; Quenzer & Maidment, 1998; Baker et al., 2006). Baker et al. (2006) noted AGREE produced satisfactory results when

compared to other enforcement especially when computational costs are considered. Downsides to the technique include the possibility of exhibiting parallel streams where the burned stream and real stream are both represented (Hellweger & Maidment, 1997; W. Saunders, 1999) and some distortion of the catchment boundaries can also be observed (W. Saunders, 1999; W. K. Saunders & Maidment, 1996). Some of these drawbacks are later addressed by additional conditioning techniques later on.

2.3.2 Levee Enforcement

The DEM at 10 m resolution lacks sufficient representation of fine grain features such as embankments, floodwalls, and closure structures. In order to better represent the influences of these features upon hydraulics and inundation extents, the National Levee Database (NLD) published by USACE was used to enforce elevations within the 10 m DEM. The elevations found in the NLD are burned into the DEM if those elevations were found to exceed those already in place.

2.3.3 Depression Filling

Local depressions are naturally occurring features of a DEM but must be addressed if a connected drainage network with continuous catchments are to be derived for flood modeling purposes. The conditioned DEM was removed of depressions by filling areas with pits while preserving the stream and levee information previously enforced. Priority-Flood developed by Barnes et al. (2014b) is an algorithm for filling said depressions and shown to have improved performance over early works in the field by Jenson and Domingue (1988) implemented in Tarboton (2005) as well as Planchon and Darboux (2002). The depression filling algorithm used in our pipeline is a Priority-Flood variant developed by (Zhou et al., 2016) with enhanced single-thread performance and a time complexity of O(n log n) for floating point grids. This performance was enabled by limiting the processing queue with a region-growing method to exclude many of the slope cells (Zhou et al., 2016). The depression technique employed here does leave the existence of flat regions where pits existed a prior thus later requiring the need for resolving these flats. The enhanced variant of Priority-Flood is implemented and made available by Barnes (2018) and Zhou et al. (2015).

2.3.4 Stream Thalweg Elevation Conditioning

Thalweg elevations are critical components of relative elevation based inundation mapping thus much is performed to ensure the best available, monotonically decreasing elevations are derived prior to normalizing of elevations. In order to prevent situations where the burned thalweg and the thalweg endemic to the DEM run parallel to one another, the normalized excavation algorithm (W. Saunders, 1999) is used to seek a zonal (nearest neighbor) elevation minimum for each thalweg pixel. Each zone is defined as the thalweg's pixel nearest neighborhood within a maximum distance of 50 m. The zonal minimum is computed for each thalweg pixel zone and the minimum is used to replace the existing thalweg elevation value.

The next step involves conditioning these local minimums along the thalweg to enforce monotonically decreasing thalweg elevations for FIM. Garousi-Nejad et al. (2019) proposed an algorithm that traverses stream thalweg pixels in a depth first manner starting with adding all the headwater pixels to a queue. The connectivity of the thalweg pixels is defined by the D-8 flow directions further discussed in Section 2.4.1. At every thalweg pixel, the minimum elevation among itself and its upstream contributing thalweg pixels is taken as shown in Equation 1,

$$\mathbf{D}[x] = \min_{y \ drains \ to \ x} (\ \mathbf{D}[x] \ , \ \mathbf{D}[y] \ ) \tag{1}$$

, in which $\mathbf{D}$ represents the array of thalweg adjusted elevations indexed by x and y where by y is upstream of x. When a pixel's upstream neighbors are all evaluated, the downstream pixel is added into the queue thus the depth first traversal of the drainage network. This procedure enforces the location of streams and ensures that thalweg elevations are hydrologically correct which yielded a 7% improvement in Critical Success Index (CSI) per an evaluation for an event in 2017 on the Malad river (Garousi-Nejad et al., 2019).

2.4 Deriving FIM Hydrofabric

The FIM Hydrofabric is defined here as the collection of geospatial datasets that are used for converting NWM discharges into inundation extents.

2.4.1 Flow Directions and Flats Resolution

To facilitate the generation of a connected stream network and its associated catchments from the conditioned DEM, the depression-filled DEM is used to derive connectivity in the form of D-8 flow directions. D-8 seeks to allocate a drainage direction for every pixel based on the adjacent eight pixel neighborhood with the steepest slope (O'Callaghan & Mark, 1984). The horizontal component of slope is defined as one for the 4 neighboring pixels in the main cardinal directions while the intercardinal pixels are designated a horizontal component of $\sqrt{2}$. The flow direction is encoded as integers 1 through 8 corresponding with the cardinal direction East as 1 and continuing counter-clockwise to the Southeast direction as 8. Flow directions are derived for non-depression filled regions trivially with the above procedure but to define connectivity for every grid cell the remaining flats corresponding to depression filled cells must be resolved.

Flat resolution from flats endemic to the DEM or from depression filled regions is a costly, non-trivial procedure which was originally addressed by Garbrecht and Martz (1997). Software implementations have developed means to partition the problem and resolve flats iteratively with communication across processes (Tarboton et al., 2009; Tesfa et al., 2011; Wallis et al., 2009; Tarboton, 2005). The excessive iteration and communication leads to poor computational performance which motivated further work on how to optimize flat resolution (Survila et al., 2016; Barnes et al., 2014a). Specifically the work by Survila et al. (2016) enables the use of parallel processing and made smoother catchments from our informal experience than those from Barnes et al. (2014a). By processing flats local to each partition separately from flats shared with other partitions, the accelerated flat resolution algorithm demonstrated an average speed up of 468x when compared to prior implementations (Survila et al., 2016). OWP FIM utilized a CyberGIS implementation of the D-8 flow direction algorithm with the accelerated resolution of flats (Survila et al., 2016; Y. Liu et al., 2016).

2.4.2 Deriving FIM Stream Network

The derivations of relative elevations and catchments from the newly conditioned DEM involves re-deriving a new FIM stream network. The FIM stream network is of similar drainage density as the NWM V2.1 network but fully converges at all junctions leaving no divergences in the network. This is accomplished by using the seed points gen-

erated from the stream network enforcement process (Section 2.3.1). These seeds points are headwater locations of the NHDPlusHR Beta Burnline Events layer that spatially correspond to the headwater definitions in the stream network of the NWM V2.1. Feeding the seed points and previously computed flow directions into flow accumulation methods (Wallis et al., 2009; Tarboton, 1997, 2005) yields a stream link accumulation raster that can be converted to a vector file for further processing. Each stream link in this derived FIM stream network is split into equidistant reaches. Stream links are defined here as segments of rivers discretized by junctions with other NWM river segments. Stream links are then further segmented at NWM lakes and HUC8 boundaries. Discretizing at NWM lakes isolates reaches and catchments associated with lakes and reservoirs to avoid mapping them using the Manning's equation and could potentially enable volume based mapping in the future as a feature enhancement. Based on previous research, splitting each remaining stream link into equidistant reaches not to exceed a parameterized value of 1.5 km helps improve synthetic rating curve and mapping skill (Garousi-Nejad et al., 2019; Godbout et al., 2019; Zheng, Maidment, et al., 2018). Small reaches can lead to unrealistic variances in channel geometries while oversized reaches can lead to grouping too much slope variance into one discretation of the stream network. Short stream segments that are introduced as a result of forced network breaks due to reservoir, levee, or HUC boundaries inherent the synthetic rating curve properties of the upstream or downstream segment, depending on the topology. Section 2.4.5 details the derivation of the synthetic rating curves and the dependence on channel length. Additionally every reach (and later catchment) is assigned a globally unique identifier based on the HUC 8 membership. This stream network is important since it drives the HAND calculation and derivation of catchments.

2.4.3 Catchments

Catchments were derived using the D8 connectivity established by O'Callaghan and Mark (1984). Outlet points are set at the pixel center points of the delineated stream lines explained in Section 2.4.2. The outlets act as root nodes in a tree structure and the connectivity is traversed to derive the contributing area for each gage. Two sets of catchments are derived, one of which assigns the contributing area for each thalweg pixel which is used for relative elevation calculation. The other catchments are derived for the contributing area for each stream reach as defined in Section 2.4.2.

2.4.4 Height Above Nearest Drainage

Once the pixel level catchments are derived, the final relative elevations can be computed. Every non-thalweg elevation is subtracted from the thalweg elevation within the same pixel-level catchment described in Section 2.4.3. The DEM used for this operation is the DEM resulting from the thalweg conditioning procedures described in Section 2.3.4. Outside of the excavated channel from the AGREE DEM method, the native non-drainage enforced elevations are used to reduce sources of error in relative elevations due to pit filling.

2.4.5 Synthetic Rating Curves

A method for converting forecast river discharges from the NWM to stages or river depths is necessary for producing FIMs with HAND. For one-dimensional models such as the NWM, the typical procedure is to establish the stage-discharge relationship by sampling data from the DEM to derive a synthetic rating curve at discrete cross-sections (Quintero et al., 2021; Di Baldassarre & Claps, 2011). For this application, we utilized the reach averaged approach for developing synthetic rating curves (SRC) (Zheng, Tarboton, et al., 2018). The reach averaged approach seeks to sample the geometry parameters in the Manning's equation (Gauckler, 1867; Manning et al., 1890) on a reach scale then dividing those by length. The reach averaged Manning's formula is derived to be

$$\mathbf{Q} = \frac{1}{n}\frac{V^{5/3}S^{1/2}}{LB^{2/3}} \tag{2}$$

where Q is discharge, y indicates the stage, n is the Manning's n roughness coefficient, V is volume at stage y, S is channel slope, L is along flow length, and B is wetted bed area at stage y. Q, V, and B are taken a specific y values so are more formally written as $Q = Q(y)$, $V = V(y)$, and $B = B(y)$, respectively. All units are international given the 1 numerator above n. The reach averaged method has been compared to rating curves from Hydrologic Engineering Centerś River Analysis System (HEC-RAS) and USGS gages yielding comparable results for estimating the river bottom elevation profile, channel width at given stages, and stage-discharge relationships (Zheng, Tarboton, et al., 2018). The reach averaged geometry parameters including number of wet cells, bed area, and volume are sampled from the thalweg conditioned AGREE DEM using TauDEM's catchhydrogeo utility. Using the split reaches described in Section 2.4.2, the channel slope is

sampled from the thalweg conditioned DEM at the end points of the reaches while the same reaches are used to calculate the channel length.

Setting of the Manning's n roughness coefficient has precedent in previous CFIM studies (Maidment, 2017; Y. Y. Liu et al., 2016; Y. Liu et al., 2020; Djokic, 2019; Garousi-Nejad et al., 2019; Zheng, Maidment, et al., 2018) with two noted values of 0.05 and 0.06 for NFIE and Djokic (2019) respectively. These values are applied universally to the entire forecasting domain across space, time, and discharge profiles. We note significant opportunity to enhance CFIM skill by better parameterizing Manning's n according to available data including but limited to land cover, land use, stream order, stream geometry, drainage area, reach length, and discharge percentiles (Garousi-Nejad et al., 2019; Johnson et al., 2019). For now and for the purpose of this study, we examine the developed ecosystem of tools with Manning's n set to both 0.06 and 0.12 which we hope will shed some light on the sensitivity of this parameter to HAND based FIMs. After all the parameters to the Manning's equation have been determined with either hydrofabric sampling or user parameterization, we select 84 stage values from 0 to 25 meters in depth at one third of a meter increments to calculate the discharge values for each stage value.

2.4.6 Cross-walking with NWM Stream Network

The stream network derived in Section 2.4.2 must be associated with a NWM reach identifier so that a discharge can be converted to stage and later inundation extent. For the methods already discussed, we overlap the reach catchments derived in Section 2.4.3 with the NWM catchments matching the ID of the NWM catchment that most overlaps the derived catchment for HAND. For two subsequent methods discussed in Sections 2.5.1 and 2.5.2, we find the mid-point of the derived stream reach line described in Section 2.4.2 and find the NWM catchment that contains the mid-point. Additionally, only relevant catchments from the NWM for the given level path are selected for cross-walking for methods in Sections 2.5.1 and 2.5.2. While these conflation methods are approximate, they work for many instances just fine but do lead to areas with substantial error. More discussion on this will follow in Section 4.

2.5 Stream Order Reduction

FIM skill has been shown to be sensitive to the drainage density of the stream network employed as the datum for HAND (J. Zhang et al., 2018; McGehee et al., 2016; Li et al., 2020; A. D. Nobre et al., 2016). In our evaluations, we note negative effects at the confluence of lower flow tributaries with higher flow rivers partly due to the independent nature of the catchments within HAND methods. Figure 3 illustrates this exact situation where two tributaries converge with a higher order stream segment. An actual map with OWP FIM is generated using the NWM full-resolution stream network and compared with a FEMA 100 year extent (see Section 2.7 for more details) showing significant under-prediction in inundation extent. The higher discharge along the MS of 1,900 cubic meters per second (CMS) does not translate to the lower flow rates along the tributaries of 84 and 195 CMS. This is due to a lack of representation of backwater conditions in the hydraulic routing techniques used. As a parallel problem, there is excess water accumulated along the MS that cannot extend in either a fluvial or pluvial manner beyond the boundaries of the MS catchments. We seek to resolve this problem by deriving HAND for processing units with stream networks of reduced stream order. We present two successive methods implemented that reduce drainage densities by reducing Horton-Strahler stream orders of the networks employed and test our presented hypothesis that unary stream order networks enhance FIM performance skill with HAND. The resulting FIMs from the overlapping HAND processing units are mosaiced together taking any inundated area to be inundated but more will be explained in Section 2.6.

2.5.1 NWS Main-stems

The Mainstems (MS) network is a subset of the NWM full-resolution (FR) network at and downstream of AHPS forecast points as seen in Figure 1. The MS network comprises about 200 thousand km of stream length which is less than 4% of the FR total stream length of 5.5 million km. It also spans 121,724 reaches across 1,608 HUC8s. HAND was originally derived for this stream network to enhance mapping skill along these critical MS segments (Djokic, 2019). The inundation derived from this stream network is mosaiced with the inundation from the FR network to form the MS FIMs. Within each HUC, you'll typically only find a MS stream network of stream order 1 (i.e. headwater) but this can vary if more than one AHPS forecasting point is found within or upstream of the HUC in question.

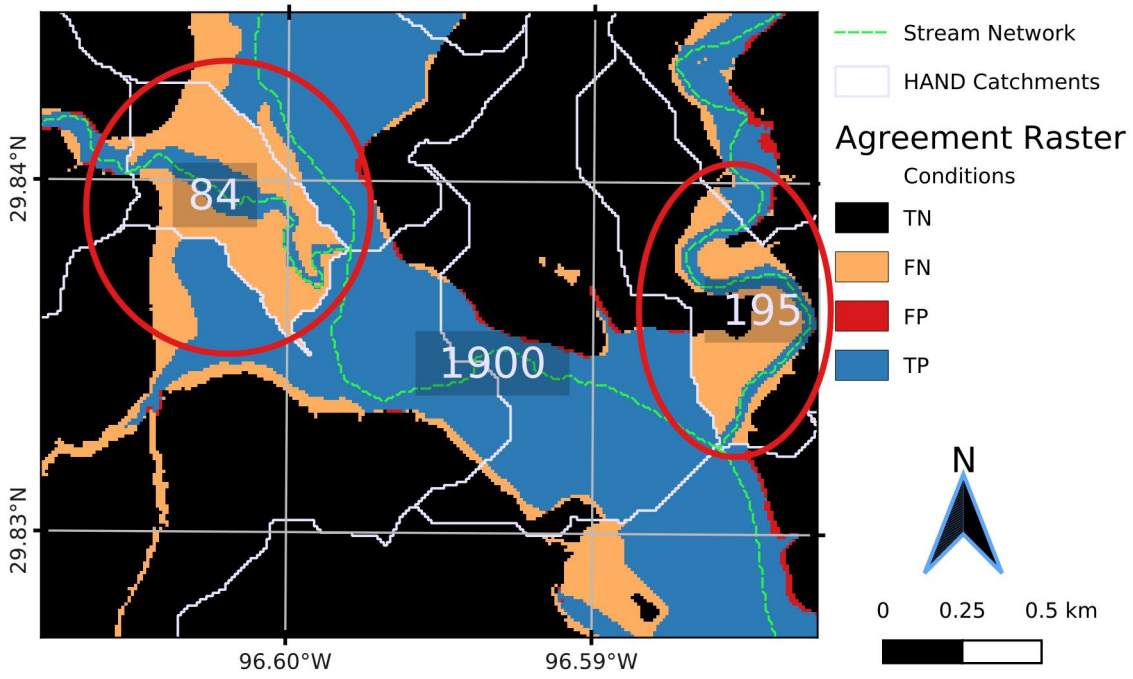


Figure 3. False negatives associated with confluence of tributaries with MS. Integers represent flow values from BLE 100 year event for the associated areas. No backwater consideration is implemented and the independent nature of the HAND catchments prohibits pluvial inundation from taking place.

2.5.2 Generalized Mainstems

To further the efforts implemented by MS, we sought to derive HAND at a level path scale which we call GMS. Since the MS network only covers a small percentage of the NWM forecasting domain, we sought to expand the benefits of stream order reduction within HAND processing units to the entire FR domain. Level paths group flowlines by maximizing the length of each flow path and minimizing the number of level path identifiers within a given domain (Moore et al., 2019; McKay et al., 2012). Starting at the outlet, a unique level path is propagated upstream. At every confluence, the direction of maximum flow path length is sought to propagate the current level path identifier. For the remaining parent reaches of the given junction, a new level path identifier is assigned and the process recursively continues with them. Figure 4 illustrate how level paths (symbolized by unique colors) are propagated upstream by the value of arbolate sum. Each HUC8 is discretized into level paths independently and relevant in-

puts are assigned to each level path processing unit given a buffer of 7 km. At the level path scale, the methods in Sections 2.3 and 2.4 are executed leaving out any tributaries of the level path in question at the time. The only exception to this is the use of the NWM stream network directly for use with hydro-enforcement which was motivated by the difficulty in deriving level paths in the NWM stream network with high agreement with the NHDPlusHR stream lines.

To illustrate the GMS procedure, we reference Figure 5 to show how deriving HAND and FIMs from GMS works. In Figure 5a, we uniquely color code the level paths derived for the NWM stream network. For each one of these lines, we derive HAND and its associated datasets including catchments, crosswalks, and rating curves. Each level path is buffered to a polygon with a user-available distance parameter of 7 km and this polygon is used to subset the original DEM for two selected level paths in Figure 5b. We illustrate two HAND grids for two of the level paths in this HUC8 in Figure 5c. Once the FIM hydrofabrics for each level path are generated, we can inundate them individually also shown in Figure 5d. Lastly these individual FIMs are mosaiced together as explained in Section 2.6 and shown in Figure 5c.

2.6 Inundation Mapping

The FIM hydrofabric consisting of the relative elevations grid, catchments grid, catchment polygons, rating curve, and cross-walking data are all used to convert forecasts from the NWM into forecasts extents. For operational situations, one would cache the FIM hydrofabric then either produce libraries of FIM for a sample of discharges or stages or also produce the FIM in near real-time (NRT). From the cached FIM hydrofabric and design or forecast discharges including those extracted from the NWM, inundation maps can be generated at HUC 8 spatial processing units in a rapid, parallel operation. The discharges are associated with NWM reach identifiers and cross-walked over to reach identifiers in the FIM hydrofabric.

Utilizing the stage-discharge relationships in the synthetic rating curves, each forecast for each catchment identifier is assigned a stage value. The catchments grid encoded with the reach identifiers are used to map the stages by thresholding to the forecast stage. We use the basic logic already established in previous works to conduct this (A. D. Nobre et al., 2016; Y. Y. Liu et al., 2016; Maidment, 2017). Mathematically, the HAND

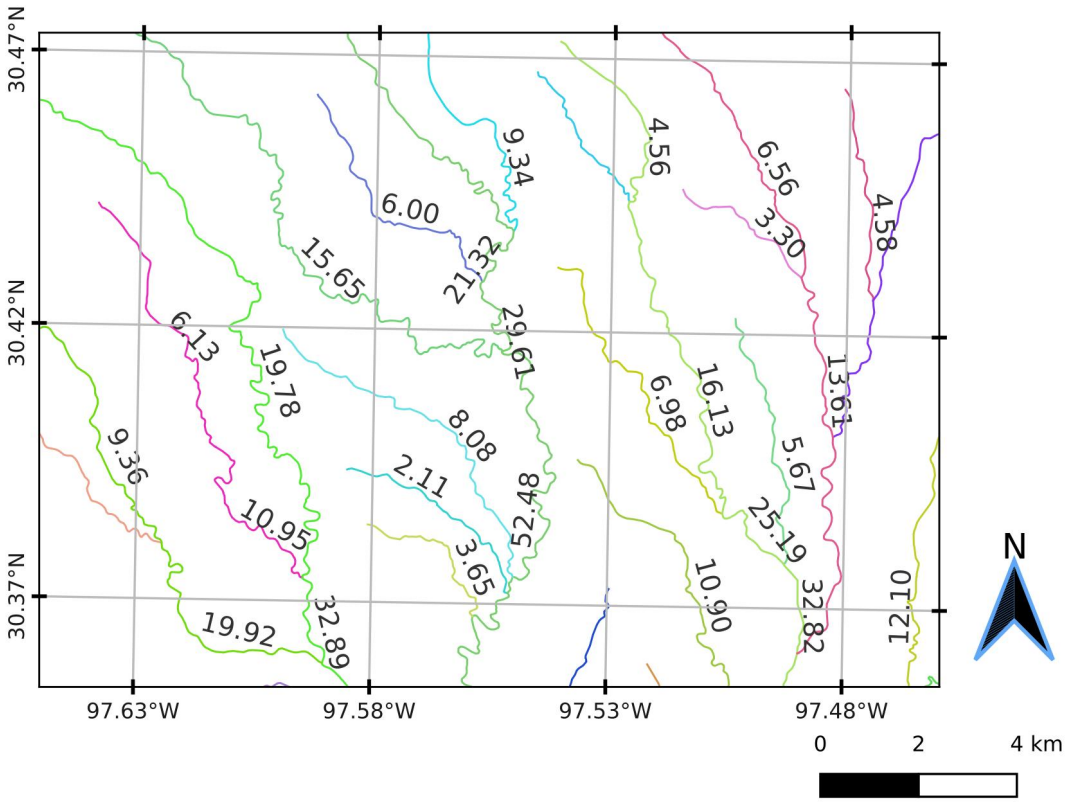


Figure 4. Illustrates how level paths for the NWM are derived. Level paths symbolized by lines of unique colors are propagated upstream following the direction of maximum arbolate sum at each junction.

values, H_{ij}, can be indexed by the reach identifiers, i, and pixel indices, j. For each forecast stage, S_i, one can express the formula for D_{ij}, a continuous variable denoting water depth at a given pixel with reach and pixel identifiers i and j respectively in Equation 3. For each forecast stage, S_i, one can express the formula for F_{ij}, a binary variable denoting inundation condition in Equation 4 in terms of D_{ij} by simply thresholding at zero depths.

$$D_{ij} = S_i - H_{ij} \tag{3}$$

$$F_{ij} = D_{ij} > 0 \tag{4}$$

For the cases of MS and GMS, the inundation maps produced for the respective processing units at lower maximum stream orders must be mosaiced together to form a seamless forecast in the form of a single raster file. For mosaicing the depths, we select the maximum inundation depth from the all the contributing areas K index by its lower case

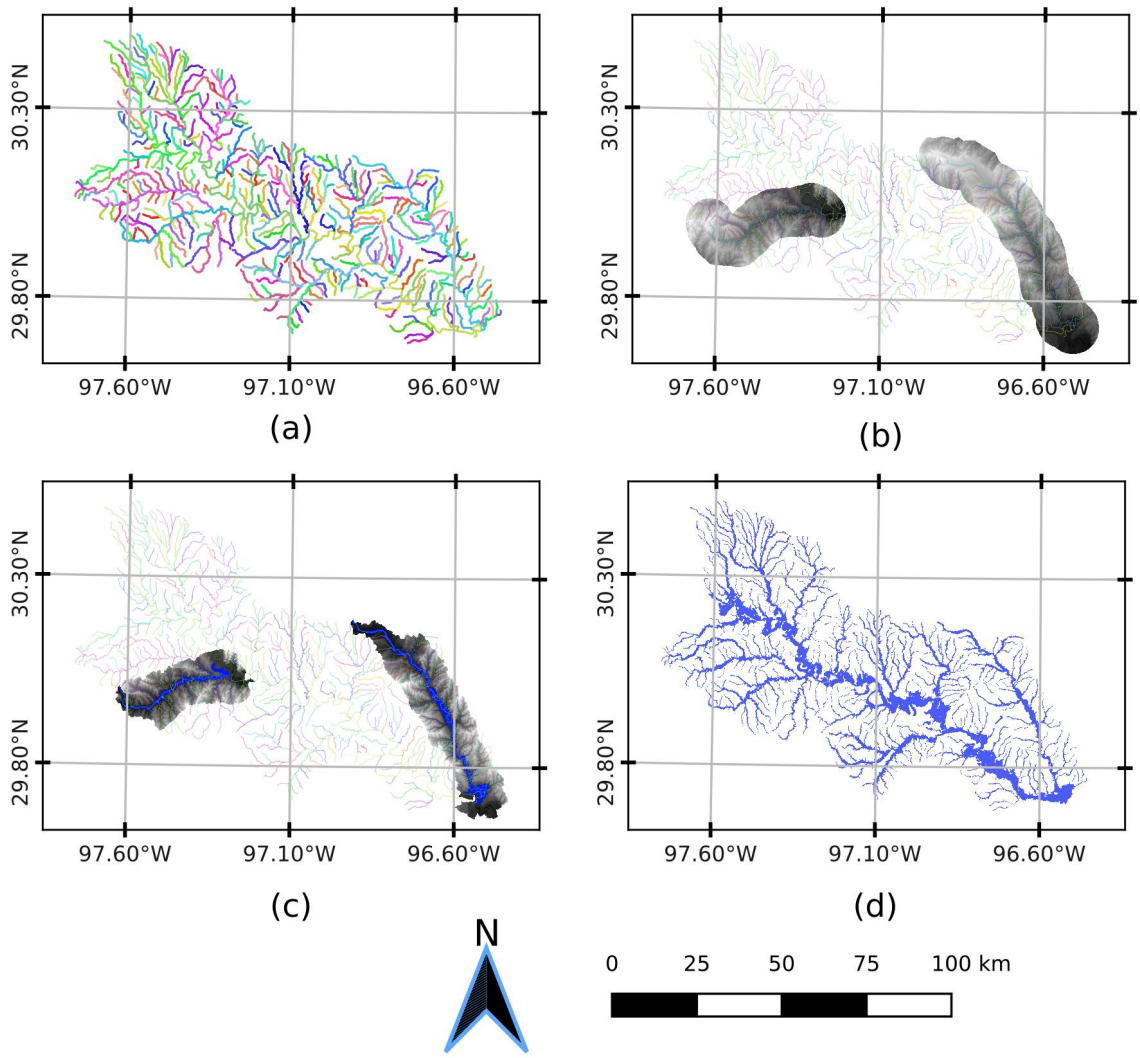


Figure 5. Overall procedure for GMS HAND. In (a), we illustrate all NWM stream lines symbolized by their level path. Meanwhile (b), demonstrates the DEM clipped to a 7 km buffer around two selected level paths. In (c), we show how HAND can be computed just for each one of these two level paths independently. We also show inundation maps created for these two level paths in (c). In (d), we show all the inundation maps for all the level paths mosaiced together.

character, k. Equation 5 illustrates how the maximum depth from all the contributing areas, k, to each pixel j in catchment i. Equation 6 illustrates the same process but for mosaicing the binary inundation maps.

$$D_{ij} = \max_{k=[1,\ldots,K]} D_{ijk} \tag{5}$$

For the MS and GMS methods, the contributing areas are defined differently. For MS, the FIM from MS HAND and FR HAND are mosaiced together to form a singular inundation map thus K is set to 2 for that case. For GMS, all FIMs from all the level paths in a given area are mosaiced together then K is set to this number of level paths. Figures 5a and 5b, illustrate how inundation maps are created for lower stream order processing units then mosaiced together.

$$F_{ij} = \max_{k=[1,\ldots,K]} F_{ijk} \tag{6}$$

2.7 Evaluation

Evaluation of our relative elevation CFIM method is conducted by comparison to the HEC-RAS 1D derived models produced within FEMA region 6 (*Base Level Engineering (BLE) Tools and Resources*, 2021; *Estimated Base Flood Elevation (estBFE) Viewer*, 2021). 49 HUC 8's spanning about 185 thousand square km were available at the time (now more) across nine states and shown in Figure 6. The maps to the 1% recurrence flow (1 in 100 year) and the 0.2% recurrence flow (1 in 500 year) are furnished by InFRM so we used those corresponding discharges and mapping extents for evaluation. We did exclude NWM V2.1 Reservoirs from evaluation because these are not properly accounted for in the inundation. By using the same HEC-RAS derived discharges and FIM extents, we are able to separate out errors introduced by hydrology, atmospheric forcings, hydraulic routing, etc that we would have potentially seen if we used NWM forecasted discharges. Figure 7 illustrates both NWM V2.1 and BLE stream lines as well as the BLE cross-sections that have recurrence discharges associated with them. We elected to spatially intersect the HEC-RAS cross sections with the NWM stream network assigning the 1% and 0.2% flow rates to each NWM reach. To handle multiple intersections, we opted to use a filter to select the median discharge value attributed to each NWM reach. This partially handles the influence of neighboring cross sections that could cause flow discontinuities and mass conservation issues. Additionally, the stream network of the InFRM furnished models are of higher stream densities and bifurcation ratios, as evident in Figure 7, leading to a significant amount of false negatives (FN) (under-prediction) along headwater streams with Horton-Strahler orders of one due to the lack of representation of these additional headwater streams in the NWM network. While the limitations are noted, this method does best to detangle the influence of exogenous variables that we do not wish to study in this comparison. The metrics employed in this study

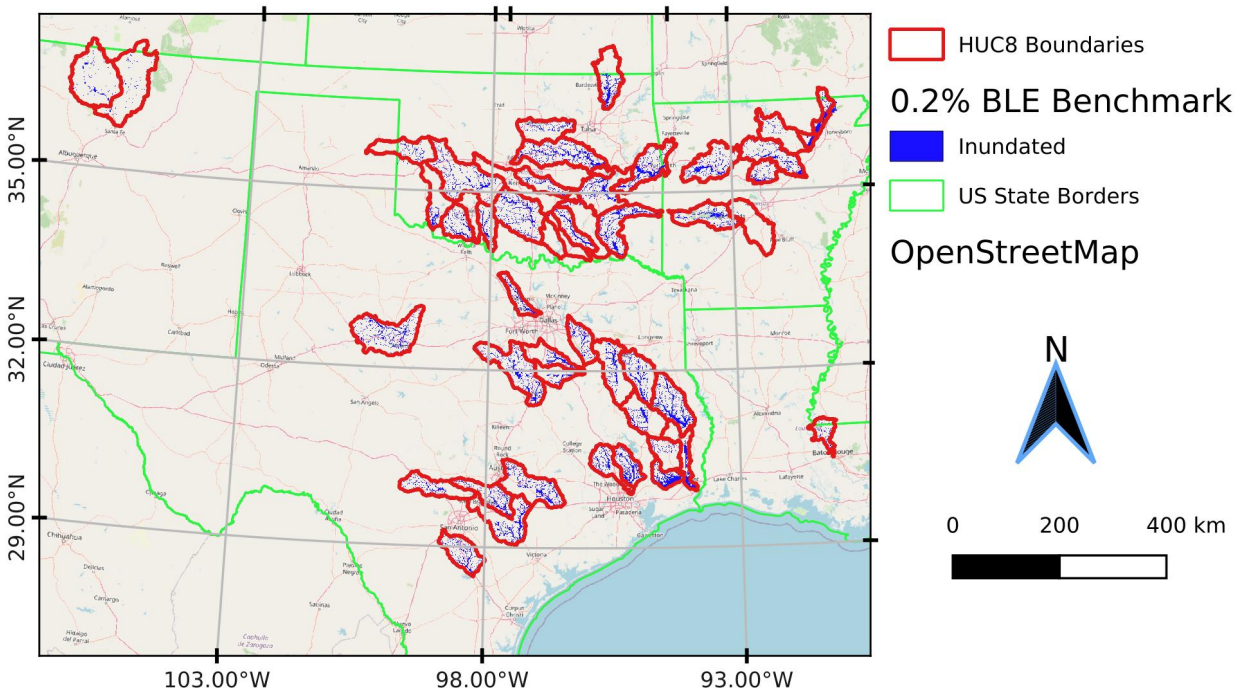


Figure 6. Shows 185 thousand square km of modeled areas for BLE domain of 49 HUC8s across 9 states. This dataset for 1% and 0.2% recurrence flows were used as benchmarks.

to evaluate inundation extents include Critical Success Index (CSI), Probability of Detection (POD), and False Alarm Ratio (FAR) and presented in Equations 7, 8, 9, respectively. To calculate these secondary metrics, one must define three primary metrics including true positives (TP) which is predicted wet and wet in benchmark dataset. The two types of errors consist of false positives (FP), or type I errors, which is dry in benchmark but predicted wet and false negatives (FN), or type II errors, which is wet in benchmark by predicted dry. Lastly, the reader may come across true negatives (TN) which is defined as dry in both the benchmark and predicted datasets. Maximizing POD indicates a model's ability to detect the given threat of interest, inundation, while minimizing FAR is sought to indicate a models ability in reducing FN errors. Some work by Gerapetritis and Pelissier (2004) while at the NWM denotes CSI a good proxy for measuring a forecasting system's utility in protecting life and property and has been shown to be optimized mathematically when $POD = 1 - FAR$. While these metrics are commonly employed in the evaluation of FIM and binary weather prediction communities in general, they do come with some notable limitations including frequency dependence

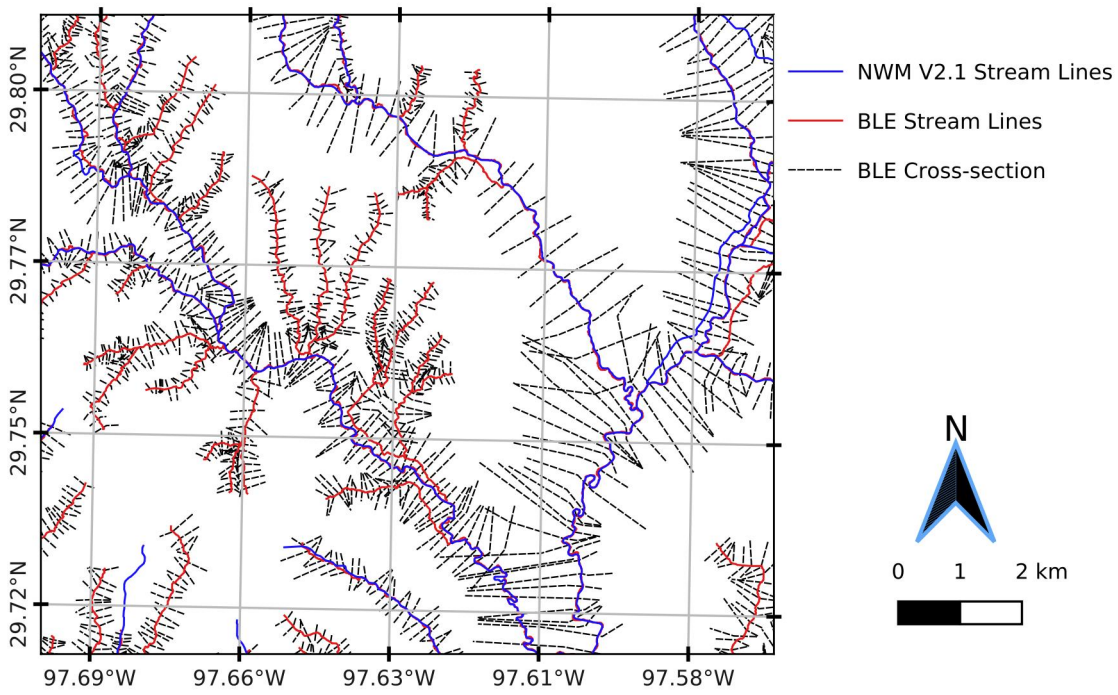


Figure 7. Illustrates Base Level Engineering (BLE) cross sections and stream lines. The BLE stream network, which is denser than the NWM V2.1 stream lines, is also shown. BLE cross sections are intersected with NWM reaches and the median recurrence discharge for 1% and 0.2% levels are selected per NWM reach.

in the case of CSI and FAR (Gerapetritis & Pelissier, 2004; Stephens et al., 2014; Schaefer, 1990; Jolliffe & Stephenson, 2012). Thus, frequency dependent statistics should be used with caution when comparing across sites with varying frequencies. Lastly, approximately 6 HUC8s do not have NWM MS reaches thus we imputed the metrics for FR for these sites as the best available forecasting capability.

$$CSI = \frac{TP}{TP + FN + FP} \tag{7}$$

$$POD = \frac{TP}{TP + FN} \tag{8}$$

$$FAR = \frac{FP}{TP + FP} \tag{9}$$

3 Results

3.1 Mapping Performance

We produced FIMs for the entire BLE domain within the 49 HUC8 areas across several states in the south central US. The forecasted FIMs using the discharges for the 1% (100 year) and 0.2% (500 year) recurrence flows directly from HEC-RAS were used to avoid noise and errors from hydrological processes. We computed the statistics (CSI, POD, and FAR) for both 100 and 500 year events for Mannings N set to 0.06 and 0.12. The distribution of these statistics can be examined in Figure 8 as violin plots. Each half of a violin plot represents the kernel density estimation (KDE) for a given model (FR, MS, GMS), given Manning's n value (0.06, 0.12), and given recurrence interval (1%, 0.2%), and performance metric (CSI, POD, FAR). We also denote trend lines for each metric and Manning's n setting as well as their respective slope estimate and one-tailed p-value denoting the level of significance of the trend.

Aggregating the metrics in the method above treats each HUC8 as it's own unit and does little to consider the size differences of the HUCs. In an opposing aggregation method, we illustrate in Table 2 the CSI, POD, and FAR recomputed for the entire domain using the sum of all the TPs, FPs, and FNs. From Figure 8 and Table 2, we denote several meaningful trends. Using CSI as an overall proxy for skill of the FIM, we

Table 2. Recomputed CSI, POD, and FAR using the primary metrics, TPs, FPs, and FNs, aggregated for BLE domain. The best value across models is highlighted in bold.

Metric	Manning's n	FR		MS		GMS	
		100yr	500yr	100yr	500yr	100yr	500yr
CSI	0.06	0.5576	0.5839	0.5717	0.5990	**0.5796**	**0.6075**
	0.12	0.5915	0.6149	0.6054	0.6288	**0.6182**	**0.6435**
POD	0.06	0.6354	0.6575	0.6524	0.6755	**0.6633**	**0.6863**
	0.12	0.7255	0.7446	0.7460	0.7648	**0.7606**	**0.7810**
FAR	0.06	0.1800	0.1609	0.1787	**0.1589**	**0.1778**	**0.1589**
	0.12	0.2379	0.2208	0.2374	0.2204	**0.2324**	**0.2148**

note that generally speaking the skill is correlated with a reduction of the stream orders of the processing units used for HAND. In other words, the more we derive HAND on networks of unit drainage density and mosaic the resulting FIMs, the better those FIMs perform. While this trend is evident for both sets of Manning's n values, the trend is slightly more significant for the higher value of 0.12. Other trends related to this Figure include the general performance premium for 0.2% events as opposed to lower 1% events. We also note how the higher Manning's n value enhances performance for both of these recurrence intervals across all models.

Dissecting the improvements and trends presented in the previous paragraph comes down mostly to improvement in POD or a reduction in absolute amount of FNs. POD being the primary driver in skill enhancement is evident across models by comparing the slope of the POD lines with the slope of the FAR lines. Even though aggregating metrics by HUC8 yields a statistically zero trend, one does see a slight reduction in FAR across models that reduce HAND's maximum stream order. Additionally, we note that POD is a primary driver in enhancing performance across Manning's n values as well. This significant improvement comes at a cost of false alarms as the FAR increases significantly across Manning's n values.

3.2 Computational Performance

The NFIE experiments were able to produce HAND for 331 HUC6's in 1.34 CPU years (Y. Y. Liu et al., 2016) and estimates using work from Djokic (2019) put producing HAND at the FR NWM at 0.55 CPU years. For our work, we were able to produce HAND at the full NWM resolution in 0.13 CPU years which represents a substantial speed-up compared to previous works. For the MS resolution an additional, 0.05 CPU years is required on top of this bringing the total to about 0.18 CPU years to produce 2,188 HUC8s that span additional areas not covered in previous HAND versions including Hawaii and Puerto Rico. GMS which generalizes HAND production to level path scale adds a significant amount of CPU time to the process bringing the estimate total to about 1.17 CPU years.

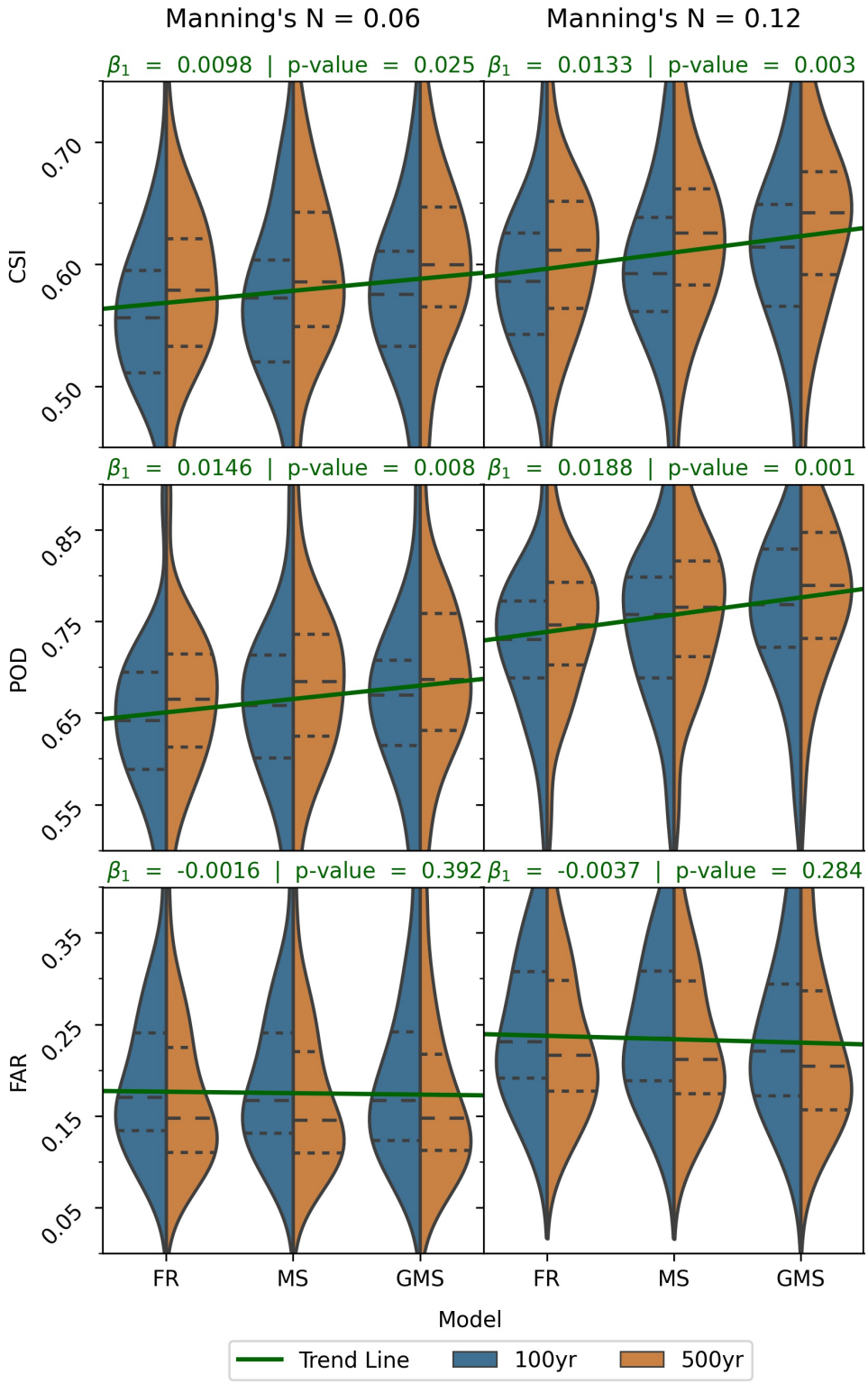


Figure 8. Shows kernel density estimation of the distributions (sample size = 49) for 1% (100 year) and 0.2% (500 year) along with horizontal, dashed lines for the 25th, 50th, and 75th percentiles (in order from bottom to top). The sub-figures separate the combination of three metrics (CSI, POD, and FAR) for two settings of Manning's n (0.06 and 0.12). Trend lines for each metric - Mannings combination are shown (sample size = 294) along with associated slope and p-value of slope testing one-tailed significance.

4 Discussion

Overall, we note a positive relationship between FIM skill and a reduction of the stream order of the stream network we use to derive the HAND datasets. Most of this change is accounted for by increasing POD thus reducing FNs especially along higher order rivers with higher flow magnitudes. We note that reducing stream order does in turn suffer from diminishing returns in which the increase in mapping skill for applying stream order reduction to roughly 4-5% of the stream network is about the same as the increase for applying stream order reduction to the remaining 95-96% of the stream network. This motivates further work in identifying what the optimal coverage of stream order reduction could be and how to parameterize that coverage. One option could be removing stream orders ones and possibly twos and threes from stream order reduction and simply using the inundation from FR from these areas.

In analyzing the data, we found a slight reduction in FAR was detected and more digging pointed to a bias in rating curves introduced by stream order reduction. Figure 9 illustrates the general effect that stream order reduction has on synthetic rating curves. Sub-figure 9a shows how the average rating curves for all reaches for stage values 0 to 25 meters at one-third meter intervals tend to bias down (and to the right) with ever increasing stream order reduction (FR to MS to GMS). This bias is more pronounced for GMS since that implements stream order reduction down to the unit level for the entire FR network while MS only does so for 4-5% of the network. Attempting to diagnose this bias in the SRC leads one to Equation 2 which shows the reach averaged synthetic rating curve relationship between stage and discharge. Across the three methods explored, FR, MS, and GMS, one identifies differences in the inputs and outputs and notes no difference in the stages and Manning's n values. While the channel slope and reach lengths are not exactly the same across methods, their averaged differences are very negligible which only leaves room for deviations in volume and bed area. Again, volume (V(y) or simply V) is synonymous to reach-averaged cross-sectional area and bed area (B(y) or B) is analogous to reach-averaged hydraulic radius. Discharge, Q, is directly related to volume and inversely related to bed area and each parameter is weighed according to the magnitude of its exponent which are $\frac{5}{3}$ and $\frac{2}{3}$ respectively (see Equation 2). Figures 9 b and c show how volume and bed area compare across the three methods with GMS having significantly greater values than MS which has greater values than FR. Again the relative discrepancy between FR vs MS and MS vs GMS is explained by the extent of

their spatial coverages. Both V and B values increase but are weighed differently by their respective exponents and pull Q in different directions. We show in Figure 9d the relationship of $\frac{V^{5/3}}{B^{2/3}}$ and plot this ratio against stage, y, to show how these two parameters collectively pull the rating curve Q up and biases the rating curve down. In other words, the magnitude and weight of the volume at each stage level exceeds the influence of the magnitude and weight of the bed area. Both parameters are set to increase mainly due to much larger catchments leading to more pixels at each stage level as shown in Figure 9e. Much of the increase in inundated pixels, volume, and bed area can be explained by much larger catchments that encompass neighboring tributaries. These tributaries have a significant amount of bathymetry that is low-lying thus easily including the SRC derivation. They also contribute volume and bed area that is technically not perpendicular to the flux of streamflow being accounted for in the stream in question. Careful examination of Figure 10b shows how much larger catchments include neighboring tributaries and the geometry associated with those tributaries. This geometry is not perpendicular to the flow that is associated with the main reach thus leading to biases in the SRC. We consider this fact to have a nuanced effect on skill, while reducing the rate of FPs it also can lead to FNs due to biases in the SRC.

Additional careful analysis of Figure 10a, leads one to note many catchments that don't have inundation or significant inundation. While the cause of these errors can be varied, we assert here that conflating 4 networks for use in evaluations leads to significant error. As one may remember, Section 2.4.6 details how reach identifiers are conflated for the FIM network back to that of the NWM. One of the issues is when a reach of given stream order accidentally conflates to that of a neighboring tributary that is of lower order which leads to areas of FNs. The utilization of MS and GMS only conflates to NWM catchments directly associated with the level path in question which is inherently easy to do with those methods. Thus part of the improvement in MS and GMS methods is due to a slight improvement in cross-walking methodology. The NWM stream network was derived using the NHD medium resolution dataset which was derived from coarser DEMs than those used here. Additional conflation is identified in cross-walking the stream network used by the BLE maps and those of HAND. Until a singular stream network is used for the NWM, BLE benchmark, and for HAND based FIM, conflation will continue being a source of error.

Our qualitative analysis suggests that the synthetic rating curves offer a significant opportunity for improvement in HAND based FIM for future development. The bathymetry of the 10 m DEM from 3DEP is known to be lacking proper representation thus leading to inadequate representation of volume and bed area with all three methods employed. Manning's n which typically accounts for roughness could be tuned to account for these DEM limitations or could be held fixed to some local value associated with a given flood magnitude. Some adjusting parameter must be introduced to enhance the estimation of the bathymetric representation. Lidar DEMs from the USGS at 3 m and 1 m scale could be utilized to derive HAND as well which we conject should show better agreement with higher fidelity FIMs also derived from the same Lidar based DEMs.

Lastly, after errors introduced by conflation, poor roughness estimation, bathymetric/elevation adjustment are accounted for, HAND still has another fundamental limitation that is inherently baked into how it works. For HAND to be derived and thus create a FIM for a given area, that area must entirely drain to the stream network and the stream network must also drain itself. In other words, an entire area eligible for flooding must monotonically decrease in elevation. DEM's naturally don't do this and the dynamics of true flood events don't follow drainage patterns. Enforcing this assumption for HAND leads to significant amount of DEM manipulations that introduce basic errors. These errors are deep into the assumptions of HAND and thus more difficult to disentangle. Ultimately, the use of more advanced 2-D hydrodynamic models should be considered for dealing with this limitation of HAND but would come at significant expense at the given high resolution across very large spatial scales and frequent forecast resolutions.

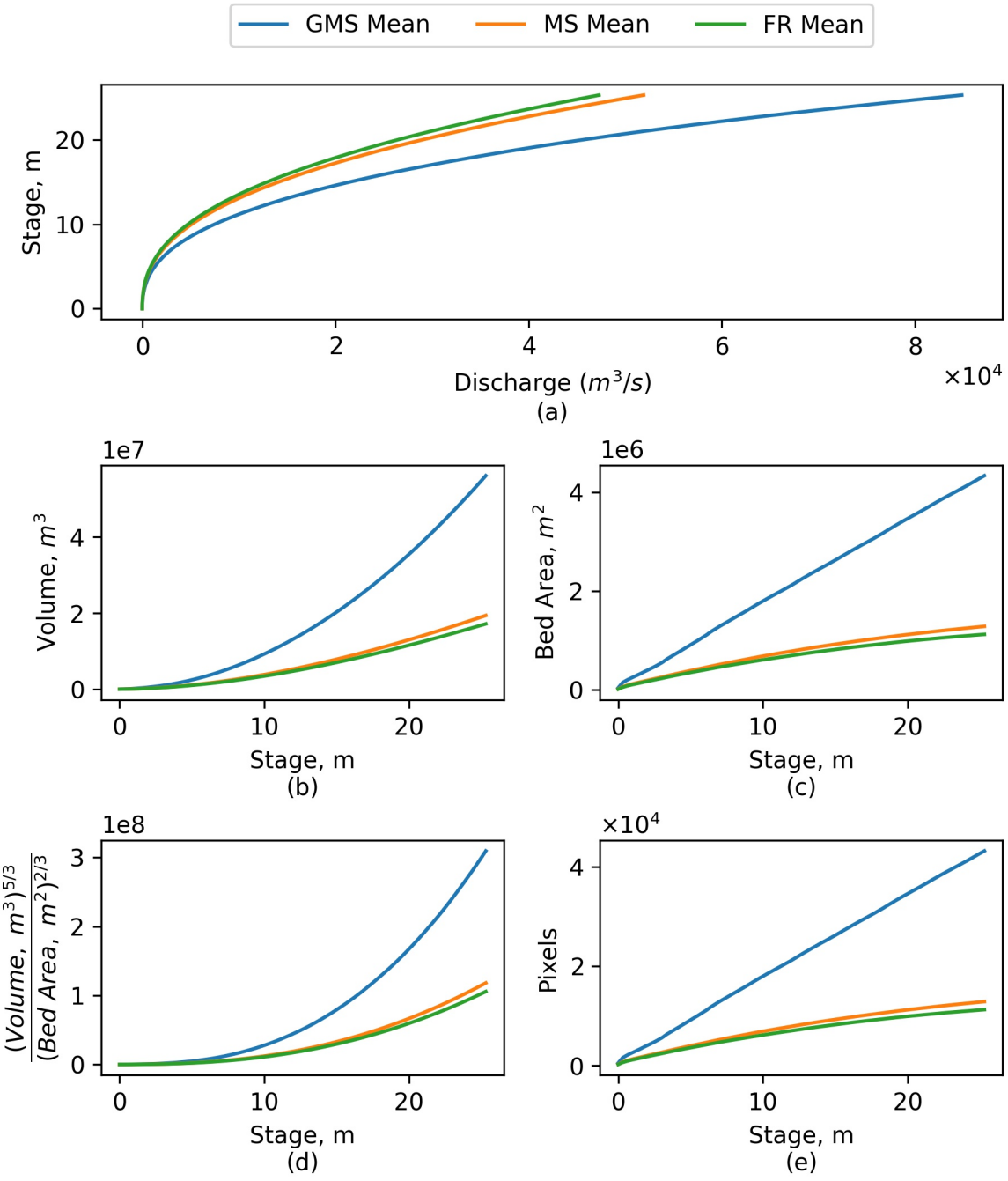


Figure 9. Illustrates average quantities for the three methods, FR, MS, and GMS, for each stage value (m). The values are (a) Discharge $m^3 s^{-1}$, (b) Volume m^3, (c) Bed Area m^2, (d) a function of Volume and Bed Area, and (e) number of pixels.

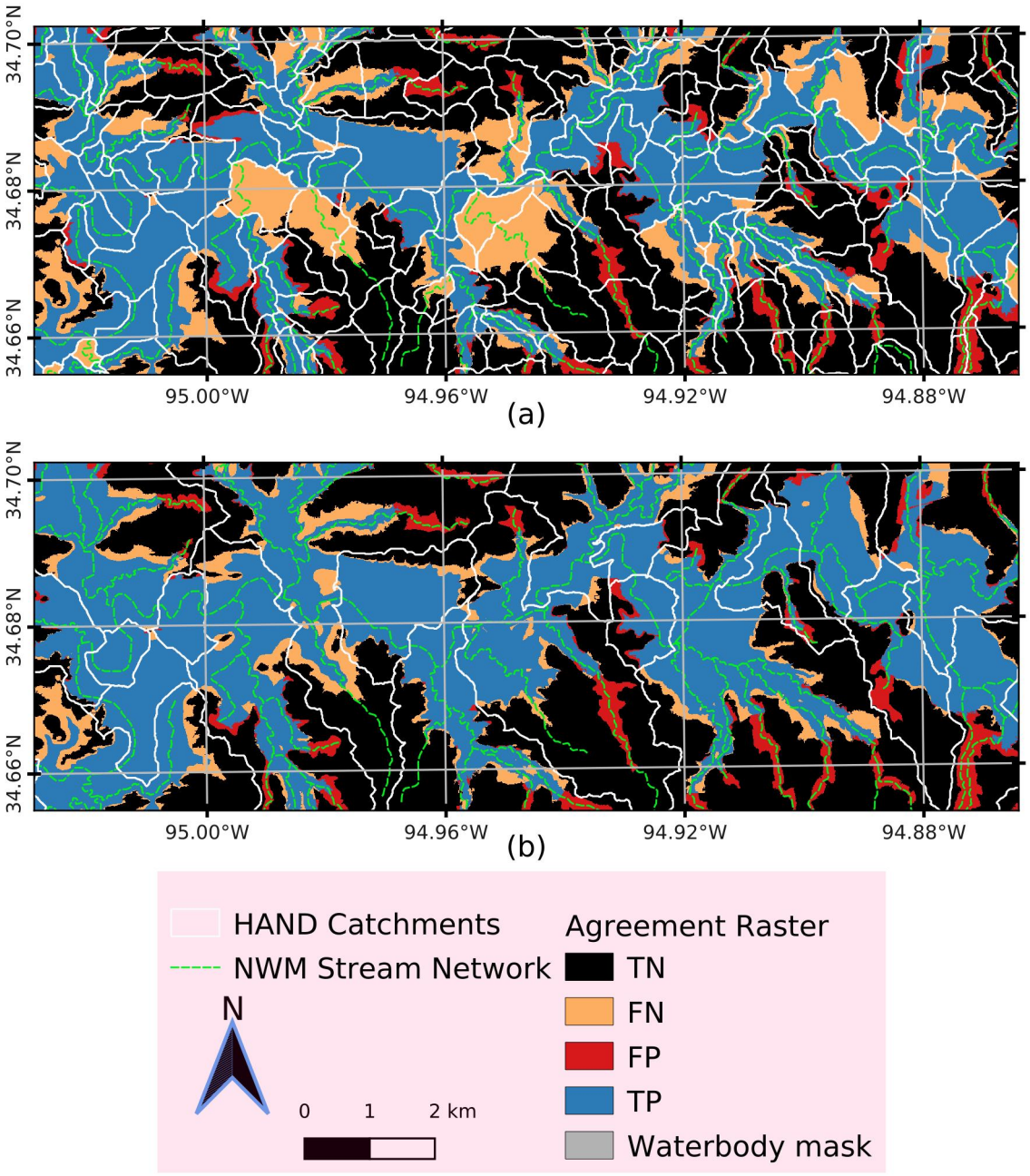


Figure 10. OWP FIM inundation agreement, TP, FP, FN, and TN, with BLE HEC-RAS maps in HUC 11140105. Catchment boundaries and stream lines are shown in white and dotted green, respectively. Sub-figure (a) shows agreement of FR HAND denoting significant areas of under-prediction due to junctions and catchment boundaries. Meanwhile, (b) shows the agreement for GMS and much larger catchments leading to much better inundation agreement for this given reach. Overall, this illustrates the benefits of stream order reduction for deriving HAND datasets.

5 Conclusions

Floods present a significant, under-served, and expanding risk to life, property, and resources. Previous flood forecasting systems lacked the coverage to adequately inform society of these risks. The National Water Model (NWM) developed by the National Oceanic and Atmospheric Administration's Office of Water Prediction, along with partners, provides increased spatial coverage and resolution as well as additional forecast time horizons on mostly hourly intervals. Additional processing is required to convert streamflows from the NWM to river stages and finally to flood inundation maps (FIM). Height Above Nearest Drainage (HAND) is a means of detrending digital elevations maps (DEM) by normalizing elevation to the nearest relevant drainage point. HAND coupled with the use of reach averaged synthetic rating curves (SRC) provide such a means of creating continental scale FIM capabilities at high resolutions (1/3 arc-second, 10 m) and high temporal frequencies (up to 1 hr). Scalable, open-source software was developed to produce HAND and associated datasets (catchments, SRCs, and cross-walking data) for the NWM forecasting area including Hawaii and Puerto Rico (Aristizabal et al., 2022b). HAND is produced using the latest hydro-conditioning techniques to enforce monotonically decreasing elevations including stream burning, levee enforcement, pit-filling, stream channel excavation, thalweg breaching, headwater seeding, stream reach resampling, and more. Finally, we use this implementation to investigate the skill of the FIMs by varying the scale of the processing units used to derive HAND. We illustrate that reducing the Horton-Strahler stream order of a HAND processing unit down to one enhances skill by significantly reducing false negatives at junctions of major streams. This also affects the parameters used to compute stage-discharge relationships biasing discharge higher at given stages which reduced the rate of false positives. FIM skill was evaluated over large spatial scales by comparsion to HEC-RAS 1D models. Further investment in the SRC's is warranted by accounting for bathymetric errors inherited by the DEM and better accounting for localized friction values at varying flow magnitudes. Utilizing the highest resolution Lidar and bathymetric data should also improve the vertical accuracy of HAND and better account for fine grain features that greatly affect inundation extents. Due to inherent limitations with HAND, scalable, physics-based methods are necessary to consider to provide a better representation of flood extent dynamics in steady and unsteady conditions.

Acknowledgments

This work was funded by the OWP which is part of NOAA's National Weather Service. Lynker, under contract with OWP, facilitated this work and computational resources used in research and development. We would like to thank some notable contributors of this work including Chief Scientist at OWP, Fred Ogden for his technical expertise. Additionally, David Blodgett from the USGS Water Mission Area was instrumental in helping define level paths and other hydrographic work. More information on code availability, usage, and data retrieval for OWP FIM is available on GitHub (Aristizabal et al., 2022b). Thanks to the Earth and Space Science Informatics Partnership (ESIP) for storing data from this study for public use and dissemination helping to provide transparent datasets for further collaboration with the research community (Aristizabal et al., 2022a).

References

Afshari, S., Tavakoly, A. A., Rajib, M. A., Zheng, X., Follum, M. L., Omranian, E., & Fekete, B. M. (2018). Comparison of new generation low-complexity flood inundation mapping tools with a hydrodynamic model. *Journal of Hydrology*, *556*, 539–556.

Amer, A., Balaji, P., Bland, W., Gropp, W., Guo, Y., Latham, R., ... others (2021). *Mpich user's guide.* Mathematics and Computer Science Division at Argonne National Laboratory. (Version 3.4.1)

Aristizabal, F., Bates, B., Avant, B., Chadwick, N., Grout, T., Spies, R., ... Cocks, G. (2022a). *noaa-nws-owp-fim.* `s3://noaa-nws-owp-fim/`. Earth Science Information Partners (ESIP).

Aristizabal, F., Bates, B., Avant, B., Chadwick, N., Grout, T., Spies, R., ... Cocks, G. (2022b). *Noaa-owp/inundation-mapping.* `https://github.com/NOAA-OWP/inundation-mapping`. GitHub.

Aristizabal, F., Judge, J., & Monsivais-Huertero, A. (2020). High-resolution inundation mapping for heterogeneous land covers with synthetic aperture radar and terrain data. *Remote Sensing*, *12*(6), 900.

Baker, M. E., Weller, D. E., & Jordan, T. E. (2006). Comparison of automated watershed delineations. *Photogrammetric Engineering & Remote Sensing*, *72*(2), 159–168.

Barnes, R. (2018). *Richdem: High-performance terrain analysis* (Tech. Rep.). PeerJ Preprints.

Barnes, R., Lehman, C., & Mulla, D. (2014a). An efficient assignment of drainage direction over flat surfaces in raster digital elevation models. *Computers & Geosciences*, *62*, 128–135.

Barnes, R., Lehman, C., & Mulla, D. (2014b). Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Computers & Geosciences*, *62*, 117–127.

Base level engineering (ble) tools and resources. (2021, Apr). Interagency Flood Risk Management Consortium, US Federal Emergency Management Agency. Retrieved from `https://www.fema.gov/media-collection/base-level-engineering-ble-tools-and-resources`

Bedient, P. B., Huber, W. C., Vieux, B. E., et al. (2008). *Hydrology and floodplain analysis*. Prentice Hall Upper Saddle River, NJ.

Bonnin, G. (1996). The noaa hydrologic data system. In *Preprints, 12th int. conf. on interactive information and processing system (iips) for meteorology, oceanography, and hydrology, atlanta, ga, amer. meteor. soc* (pp. 410–413).

Breidenbach, J., Seo, D., Tilles, P., & Roy, K. (1999). Accounting for radar beam blockage patterns in radar-derived precipitation mosaics for river forecast centers. In *Preprints, 15th int. conf. on interactive information and processing systems (iips) for meteorology, oceanography, and hydrology, dallas, tx, amer. meteor. soc* (Vol. 5).

Cajina, N., Sylvestre, J., Henderson, E., Logan, M., & Richardson, M. (2002). Fldview: The nws flood forecast mapping application. In *Proc. of the interactive symp. on the advanced weather interactive processing system (awips)* (pp. 170–172).

Corringham, T. W., & Cayan, D. R. (2019). The effect of el niño on flood damages in the western united states. *Weather, Climate, and Society*, *11*(3), 489–504.

Cosgrove, B., Gochis, D., Graziano, T. M., Clark, E. P., & Flowers, T. (2019). The evolution of noaa's national water model: An overview of version 2.1 and future operational plans. *AGUFM*, *2019*, H51D–01.

Cunge, J. (1969). On the subject of a flood propagation computation method (musklngum method). *Journal of Hydraulic Research*, *7*(2), 205–230.

Day, G. N. (1985). Extended streamflow forecasting using nwsrfs. *Journal of Water Resources Planning and Management*, *111*(2), 157–170.

Di Baldassarre, G., & Claps, P. (2011). A hydraulic study on the applicability of flood rating curves. *Hydrology Research*, *42*(1), 10–19.

Djokic, D. (2019). *Arc hydro: Developing hand from nhdplushr and nwm data – detailed workflow* (1st ed.). Environmental System Research Institute.

Donchyts, G., Winsemius, H., Schellekens, J., Erickson, T., Gao, H., Savenije, H., & van de Giesen, N. (2016). Global 30m height above the nearest drainage. *HAND*, *1000*(0).

Downton, M. W., Miller, J. Z. B., & Pielke Jr, R. A. (2005). Reanalysis of us national weather service flood loss database. *Natural Hazards Review*, *6*(1), 13–22.

Duan, Q. (2003). Global optimization for watershed model calibration. *Calibration of watershed models*, *6*, 89–104.

Duan, Q., & Schaake, J. (2002). Results from the second international workshop on model parameter estimation experiment (mopex). In *Paper presentation, second federal interagency hydrologic modeling conf.*

ENGINEERS, U. A. C. O. (2016). National levee database. *Accessed March*.

Estimated base flood elevation (estbfe) viewer. (2021, Apr). Interagency Flood Risk Management Consortium, US Federal Emergency Management Agency. Retrieved from `https://webapps.usgs.gov/infrm/estBFE/`

Fread, D. L. (1973). Technique for implicit dynamic routing in rivers with tributaries. *Water Resources Research*, *9*(4), 918–926.

Garbrecht, J., & Martz, L. W. (1997). The assignment of drainage direction over flat surfaces in raster digital elevation models. *Journal of hydrology*, *193*(1-4), 204–213.

Garousi-Nejad, I., Tarboton, D. G., Aboutalebi, M., & Torres-Rua, A. F. (2019). Terrain analysis enhancements to the height above nearest drainage flood inundation mapping method. *Water Resources Research*, *55*(10), 7983–8009.

Gauckler, P. (1867). *Etudes théoriques et pratiques sur l'ecoulement et le mouvement des eaux*. Gauthier-Villars.

GDAL/OGR contributors. (2020). GDAL/OGR geospatial data abstraction software library [Computer software manual]. Retrieved from `https://gdal.org` (Version 3.1.2)

Gerapetritis, H., & Pelissier, J. M. (2004). On the behavior of the critical success index.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., & Tyler, D. (2002). The national elevation dataset. *Photogrammetric engineering and remote sensing*, *68*(1), 5–32.

Gochis, D., Barlage, M., Dugger, A., FitzGerald, K., Karsten, L., McAllister, M., ... others (2018). The wrf-hydro modeling system technical description,(version 5.0). *NCAR Technical Note*, *107*.

Godbout, L., Zheng, J. Y., Dey, S., Eyelade, D., Maidment, D., & Passalacqua, P. (2019). Error assessment for height above the nearest drainage inundation mapping. *JAWRA Journal of the American Water Resources Association*,

941 *55*(4), 952–963.

942 GRASS Development Team. (2020). Geographic resources analysis support sys-
943 tem (grass gis) software [Computer software manual]. USA. Retrieved from
944 `https://grass.osgeo.org`

945 Great Lakes Aquatic Habitat Framework contributors. (2020). *Great Lakes Hydrog-*
946 *raphy Dataset (GLHD).* `https://www.glahf.org/download/159/`.

947 Gupta, H. V., Sorooshian, S., Hogue, T. S., & Boyle, D. P. (2003). Advances in au-
948 tomatic calibration of watershed models. *Calibration of watershed models*, *6*,
949 9–28.

950 Hellweger, F., & Maidment, D. (1997). Agree-dem surface reconditioning system.
951 *University of Texas, Austin*.

952 Herr, H., Welles, E., Mullusky, M., Wu, L., & Schaake, J. (2002). Simplified short
953 term precipitation ensemble forecasts: Theory. In *Preprints, 16th conf. on hy-*
954 *drology, orlando, fl, amer. meteor. soc., j1–j16.*

955 Hogue, T. S., Gupta, H. V., Sorooshian, S., & Tomkins, C. D. (2003). A multi-step
956 automatic calibration scheme for watershed models. *Calibration of Watershed*
957 *Models*, *6*, 165–174.

958 Horton, R. E. (1945). Erosional development of streams and their drainage basins;
959 hydrophysical approach to quantitative morphology. *Geological society of*
960 *America bulletin*, *56*(3), 275–370.

961 Huang, C., Nguyen, B. D., Zhang, S., Cao, S., & Wagner, W. (2017). A comparison
962 of terrain indices toward their ability in assisting surface water mapping from
963 sentinel-1 data. *ISPRS International Journal of Geo-Information*, *6*(5), 140.

964 Jenson, S. K., & Domingue, J. O. (1988). Extracting topographic structure from
965 digital elevation data for geographic information system analysis. *Photogram-*
966 *metric engineering and remote sensing*, *54*(11), 1593–1600.

967 Johnson, J. M., Munasinghe, D., Eyelade, D., & Cohen, S. (2019). An integrated
968 evaluation of the national water model (nwm)–height above nearest drainage
969 (hand) flood mapping methodology. *Natural Hazards and Earth System Sci-*
970 *ences*, *19*(11), 2405–2420.

971 Jolliffe, I. T., & Stephenson, D. B. (2012). *Forecast verification: a practitioner's*
972 *guide in atmospheric science*. John Wiley & Sons.

Jordahl, K. (2014). Geopandas: Python tools for geographic data. *URL: https://github. com/geopandas/geopandas*.

Kondragunta, C. (2001). An outlier detection technique to quality control rain gauge measurements. *AGUSM*, *2001*, H22A–07.

Koren, V., Reed, S., Smith, M., Zhang, Z., & Seo, D.-J. (2004). Hydrology laboratory research modeling system (hl-rms) of the us national weather service. *Journal of Hydrology*, *291*(3-4), 297–318.

Kunkel, K. E., Pielke Jr, R. A., & Changnon, S. A. (1999). Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: A review. *Bulletin of the American Meteorological Society*, *80*(6), 1077–1098.

Lam, S. K., Pitrou, A., & Seibert, S. (2015). Numba: A llvm-based python jit compiler. In *Proceedings of the second workshop on the llvm compiler infrastructure in hpc* (pp. 1–6).

Li, Z., Mount, J., & Demir, I. (2020). Evaluation of model parameters of hand model for real-time flood inundation mapping: Iowa case study. *EarthArXiv. July*, *1*.

Lindsay, J. B., & Seibert, J. (2013). Measuring the significance of a divide to local drainage patterns. *International Journal of Geographical Information Science*, *27*(7), 1453–1468.

Liu, Y., Tarboton, D. G., & Maidment, D. R. (2020). *Height above nearest drainage (hand) and hydraulic property table for conus* (Tech. Rep.). Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States). Oak Ridge ....

Liu, Y., Yin, D., Gao, Y., & Muite, B. (2016). *Cybergis toolkit.* `https://github.com/cybergis/cybergis-toolkit`. GitHub.

Liu, Y. Y., Maidment, D. R., Tarboton, D. G., Zheng, X., Yildirim, A., Sazib, N. S., & Wang, S. (2016). A cybergis approach to generating high-resolution height above nearest drainage (hand) raster for national flood mapping.

Maidment, D. R. (2017). Conceptual framework for the national flood interoperability experiment. *JAWRA Journal of the American Water Resources Association*, *53*(2), 245-257. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/1752-1688.12474` doi: 10.1111/1752-1688.12474

Mallakpour, I., & Villarini, G. (2015). The changing nature of flooding across the central united states. *Nature Climate Change*, *5*(3), 250–254.

Manning, R., Griffith, J. P., Pigot, T., & Vernon-Harcourt, L. F. (1890). *On the flow of water in open channels and pipes.*

McEnery, J., Ingram, J., Duan, Q., Adams, T., & Anderson, L. (2005). Noaa's advanced hydrologic prediction service: building pathways for better science in water forecasting. *Bulletin of the American Meteorological Society*, *86*(3), 375–386.

McGehee, R., Li, L., & Poston, E. (2016). The modified hand method. In D. R. Maidment, A. Rajib, P. Lin, & E. P. Clark (Eds.), *National water center innovators program summer institute report 2016* (Vol. 4).

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., & Rea, A. (2012). *Nhdplus version 2: User guide; national operational hydrologic remote sensing center: Washington, dc, 2012.*

Milly, P. C. D., Wetherald, R. T., Dunne, K., & Delworth, T. L. (2002). Increasing risk of great floods in a changing climate. *Nature*, *415*(6871), 514–517.

Mizgalewicz, P. J., & Maidment, D. R. (1996). *Modeling agrichemical transport in midwest rivers using geographic information systems* (Unpublished doctoral dissertation). Center for Research in Water Resources, University of Texas at Austin.

Moore, R. B., McKay, L. D., Rea, A. H., Bondelid, T. R., Price, C. V., Dewald, T. G., & Johnston, C. M. (2019). *User's guide for the national hydrography dataset plus (nhdplus) high resolution* (Tech. Rep.). US Geological Survey.

Mullusky, M., Wu, L., Herr, H., Welles, E., Schaake, J., Ostrowski, J., & Pryor, N. (2002). Simplified short term precipitation ensemble forecasts: Application. In *Preprints, 16th conf. on hydrology, orlando, fl, amer. meteor. soc., jp1* (Vol. 19).

National Weather Service. (2018, Apr). *Summary of natural hazard statistics for 2017 in the united states.* NOAA. Retrieved from `https://www.weather.gov/media/hazstat/sum17.pdf`

National Weather Service. (2019, Apr). *Summary of natural hazard statistics for 2018 in the united states.* NOAA. Retrieved from `https://www.weather.gov/media/hazstat/sum19.pdf`

National Weather Service. (2020a, Nov). *Nws preliminary us flood fatality statistics.* NOAA's National Weather Service. Retrieved from `https://`

www.weather.gov/arx/usflood

National Weather Service. (2020b, Jun). *Summary of natural hazard statistics for 2019 in the united states.* NOAA. Retrieved from https://www.weather.gov/media/hazstat/sum19.pdf

Nobre, A., Cuartas, L., Hodnett, M., Rennó, C., Rodrigues, G., Silveira, A., ... Saleska, S. (2011). Height above the nearest drainage – a hydrologically relevant new terrain model. *Journal of Hydrology*, *404*(1), 13 - 29. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022169411002599 doi: https://doi.org/10.1016/j.jhydrol.2011.03.051

Nobre, A. D., Cuartas, L. A., Momo, M. R., Severo, D. L., Pinheiro, A., & Nobre, C. A. (2016). Hand contour: a new proxy predictor of inundation extent. *Hydrological Processes*, *30*(2), 320–333.

O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer vision, graphics, and image processing*, *28*(3), 323–344.

OpenStreetMap contributors. (2017). *Planet dump retrieved from https://planet.osm.org .* https://www.openstreetmap.org.

Parada, L. M., Fram, J. P., & Liang, X. (2003). Multi-resolution calibration methodology for hydrologic models: application to a sub-humid catchment. *Calibration of Watershed Models*, *6*, 197–212.

Petrochenkov, G. (2020). pygft: Rapid flood inundation modeling tool [Computer software manual]. Reston, VA: U.S. Geological Survey. Retrieved from https://code.usgs.gov/gft/python-gis-flood-tool

Pielke Jr, R. A., & Downton, M. W. (2000). Precipitation and damaging floods: Trends in the united states, 1932–97. *Journal of climate*, *13*(20), 3625–3637.

Planchon, O., & Darboux, F. (2002). A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena*, *46*(2-3), 159–176.

Ponce, V. M., & Changanti, P. (1994). Variable-parameter muskingum-cunge method revisited. *Journal of Hydrology*, *162*(3-4), 433–439.

Python Core Team. (2019). Python: A dynamic, open source programming language [Computer software manual]. Retrieved from https://www.python.org/ (Python version 3.8.2)

Quenzer, A. M., & Maidment, D. R. (1998). *A gis assessment of the total loads and water quality in the corpus christi bay system* (Unpublished doctoral dissertation). Center for Research in Water Resources, University of Texas at Austin.

Quintero, F., Rojas, M., Muste, M., Krajewski, W. F., Perez, G., Johnson, S., ... Zogg, J. (2021). Development of synthetic rating curves: Case study in iowa. *Journal of Hydrologic Engineering*, *26*(1), 05020046.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., & Participants, D. (2004). Overall distributed model intercomparison project results. *Journal of Hydrology*, *298*(1-4), 27–60.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., & Tomasella, J. (2008). Hand, a new terrain descriptor using srtm-dem: Mapping terra-firme rainforest environments in amazonia. *Remote Sensing of Environment*, *112*(9), 3469–3481.

Saunders, W. (1999). Preparation of dems for use in environmental modeling analysis. In *Esri user conference* (pp. 24–30).

Saunders, W., & Maidment, D. (1995). Grid-based watershed and stream network delineation for the san antonio-nueces coastal basin. *Proceedings of Texas Water*, *95*, 16–17.

Saunders, W. K., & Maidment, D. R. (1996). *A gis assessment of nonpoint source pollution in the san antonio-nueces coastal basin* (Tech. Rep.). Center for Research in Water Resources, University of Texas at Austin.

Schaefer, J. T. (1990). The critical success index as an indicator of warning skill. *Weather and forecasting*, *5*(4), 570–575.

Seo, D.-J., & Breidenbach, J. (2002). Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *Journal of Hydrometeorology*, *3*(2), 93–111.

Seo, D.-J., Perica, S., Welles, E., & Schaake, J. (2000). Simulation of precipitation fields from probabilistic quantitative precipitation forecast. *Journal of Hydrology*, *239*(1-4), 203–229.

Shastry, A., Egbert, R., Aristizabal, F., Luo, C., Yu, C.-W., & Praskievicz, S. (2019). Using steady-state backwater analysis to predict inundated area from national water model streamflow simulations. *JAWRA Journal of the*

American Water Resources Association, *55*(4), 940–951.

Slater, L. J., & Villarini, G. (2016). Recent trends in us flood risk. *Geophysical Research Letters*, *43*(24), 12–428.

Stephens, E., Schumann, G., & Bates, P. (2014). Problems with binary pattern measures for flood model evaluation. *Hydrological Processes*, *28*(18), 4928–4937.

Strahler, A. N. (1952). Hypsometric (area-altitude) analysis of erosional topography. *Geological Society of America Bulletin*, *63*(11), 1117–1142.

Survila, K., Li, T., Liu, Y. Y., Tarboton, D. G., & Wang, S. (2016). A scalable high-performance topographic flow direction algorithm for hydrological information analysis. In *Proceedings of the xsede16 conference on diversity, big data, and science at scale* (pp. 1–7).

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, *10*(1), 13768. Retrieved from `https://doi.org/10.1038/s41598-020-70816-2` doi: 10.1038/s41598-020-70816-2

Tange, O. (2015). Gnu parallel-the command-line power tool.; login: The usenix magazine, 36 (1): 42–47, feb 2011. *URL http://www. gnu. org/s/parallel*, *101*.

Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water resources research*, *33*(2), 309–319.

Tarboton, D. G. (2005). Terrain analysis using digital elevation models (taudem). *Utah State University, Logan*.

Tarboton, D. G., Schreuders, K., Watson, D., & Baker, M. (2009). Generalized terrain-based flow analysis of digital elevation models. In *Proceedings of the 18th world imacs congress and modsim09 international congress on modelling and simulation, cairns, australia* (Vol. 20002006).

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software*, *90*, 201–216.

Teng, J., Vaze, J., Dutta, D., & Marvanek, S. (2015). Rapid inundation modelling in large floodplains using lidar dem. *Water Resources Management*, *29*(8), 2619–2636.

Tesfa, T. K., Tarboton, D. G., Watson, D. W., Schreuders, K. A., Baker, M. E., & Wallace, R. M. (2011). Extraction of hydrological proximity measures from dems using parallel processing. *Environmental Modelling & Software*, *26*(12), 1696–1709.

Tuozzolo, S., Langhorst, T., de Moraes Frasson, R. P., Pavelsky, T., Durand, M., & Schobelock, J. J. (2019). The impact of reach averaging manning's equation for an in-situ dataset of water surface elevation, width, and slope. *Journal of Hydrology*, *578*, 123866.

Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing*, *37*(13), 2990–3004.

Verdin, J., Verdin, K., Mathis, M. L., Magadzire, T., Kabuchanga, E., Woodbury, M., & Gadain, H. (2016). *A software tool for rapid flood inundation mapping* (Tech. Rep.). US Geological Survey.

Wallis, C., Watson, D., Tarboton, D., & Wallace, R. (2009). Parallel flow-direction and contributing area calculation for hydrology analysis in digital elevation models. *power*, *11*(8), 7.

Warmerdam, F. (2008). The geospatial data abstraction library. In *Open source approaches in spatial data handling* (pp. 87–104). Springer.

Wing, O. E., Bates, P. D., Smith, A. M., Sampson, C. C., Johnson, K. A., Fargione, J., & Morefield, P. (2018). Estimates of present and future flood risk in the conterminous united states. *Environmental Research Letters*, *13*(3), 034023.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). Merit hydro: a high-resolution global hydrography map based on latest topography dataset. *Water Resources Research*, *55*(6), 5053–5073.

Zhang, J., Huang, Y.-F., Munasinghe, D., Fang, Z., Tsang, Y.-P., & Cohen, S. (2018). Comparative analysis of inundation mapping approaches for the 2016 flood in the brazos river, texas. *JAWRA Journal of the American Water Resources Association*, *54*(4), 820–833.

Zhang, Z. (2003). Hydrologic model calibration in the national weather service. *Calibration of watershed models*, 133.

Zheng, X., Maidment, D. R., Tarboton, D. G., Liu, Y. Y., & Passalacqua, P. (2018). Geoflood: Large-scale flood inundation mapping based on high-resolution

terrain analysis. *Water Resources Research*, *54*(12), 10–013.

Zheng, X., Tarboton, D. G., Maidment, D. R., Liu, Y. Y., & Passalacqua, P. (2018). River channel geometry and rating curve estimation using height above the nearest drainage. *JAWRA Journal of the American Water Resources Association*, *54*(4), 785–806.

Zhou, G., Sun, Z., & Fu, S. (2015). *Filldem.* `https://github.com/zhouguiyun-uestc/FillDEM`. GitHub.

Zhou, G., Sun, Z., & Fu, S. (2016). An efficient variant of the priority-flood algorithm for filling depressions in raster digital elevation models. *Computers & Geosciences*, *90*, 87–96.

Figure 1.

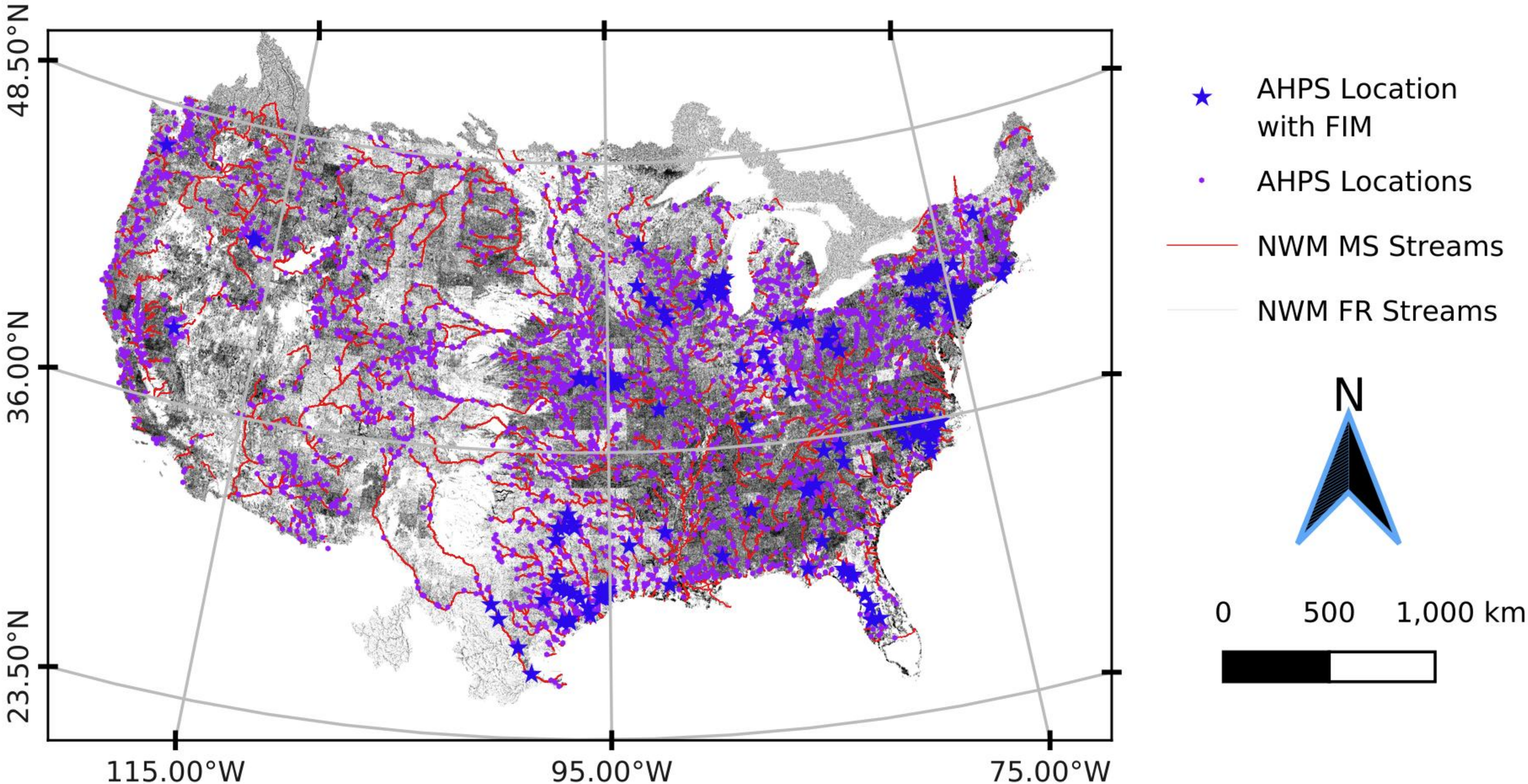

48.50°N
36.00°N
23.50°N
115.00°W
95.00°W
75.00°W
AHPS Location with FIM
AHPS Locations
NWM MS Streams
NWM FR Streams
N
0
500
1,000 km

Figure 2.

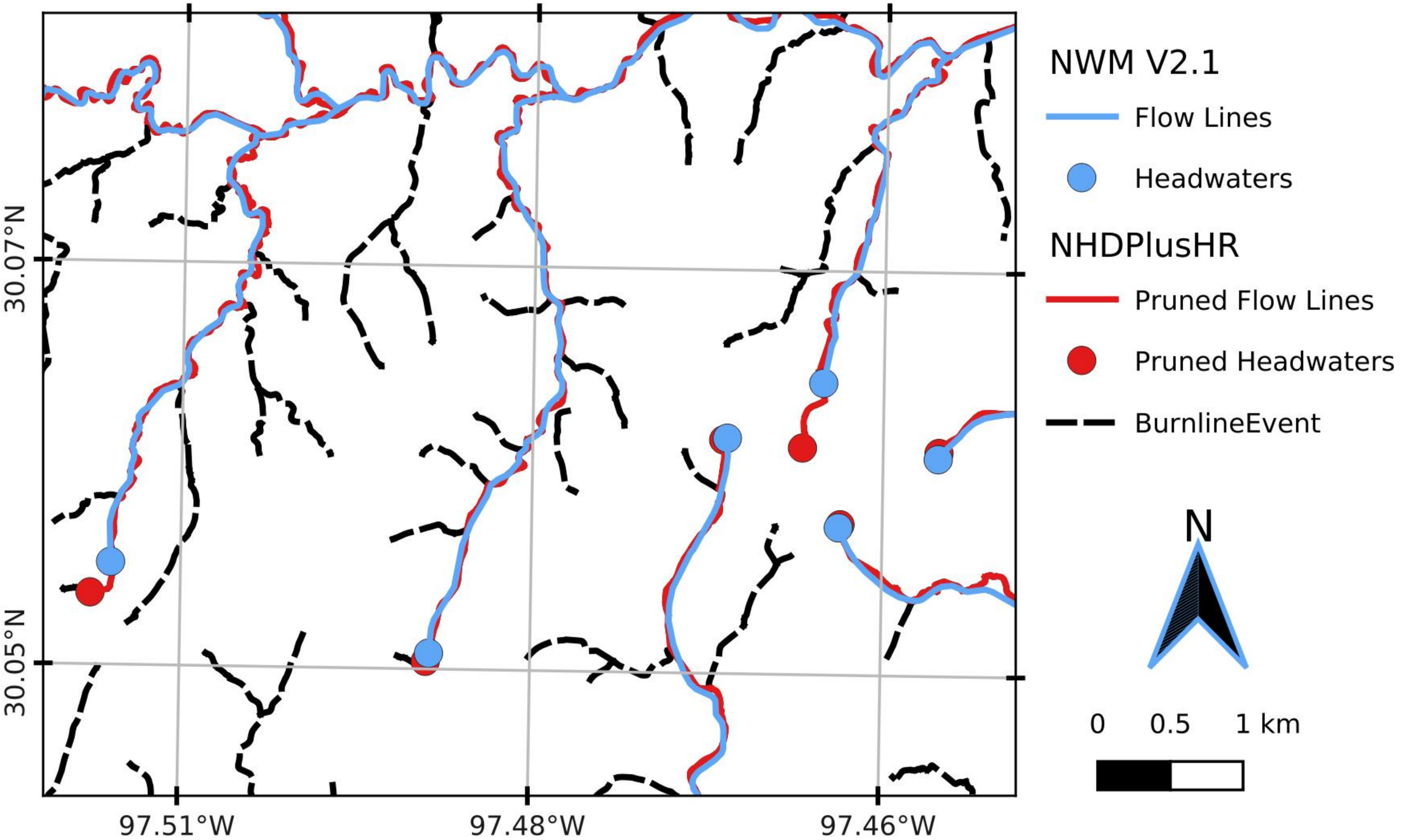

30.07°N
30.05°N
97.51°W
97.48°W
97.46°W
NWM V2.1
Flow Lines
Headwaters
NHDPlusHR
Pruned Flow Lines
Pruned Headwaters
BurnlineEvent
N
0
0.5
1 km

Figure 3.

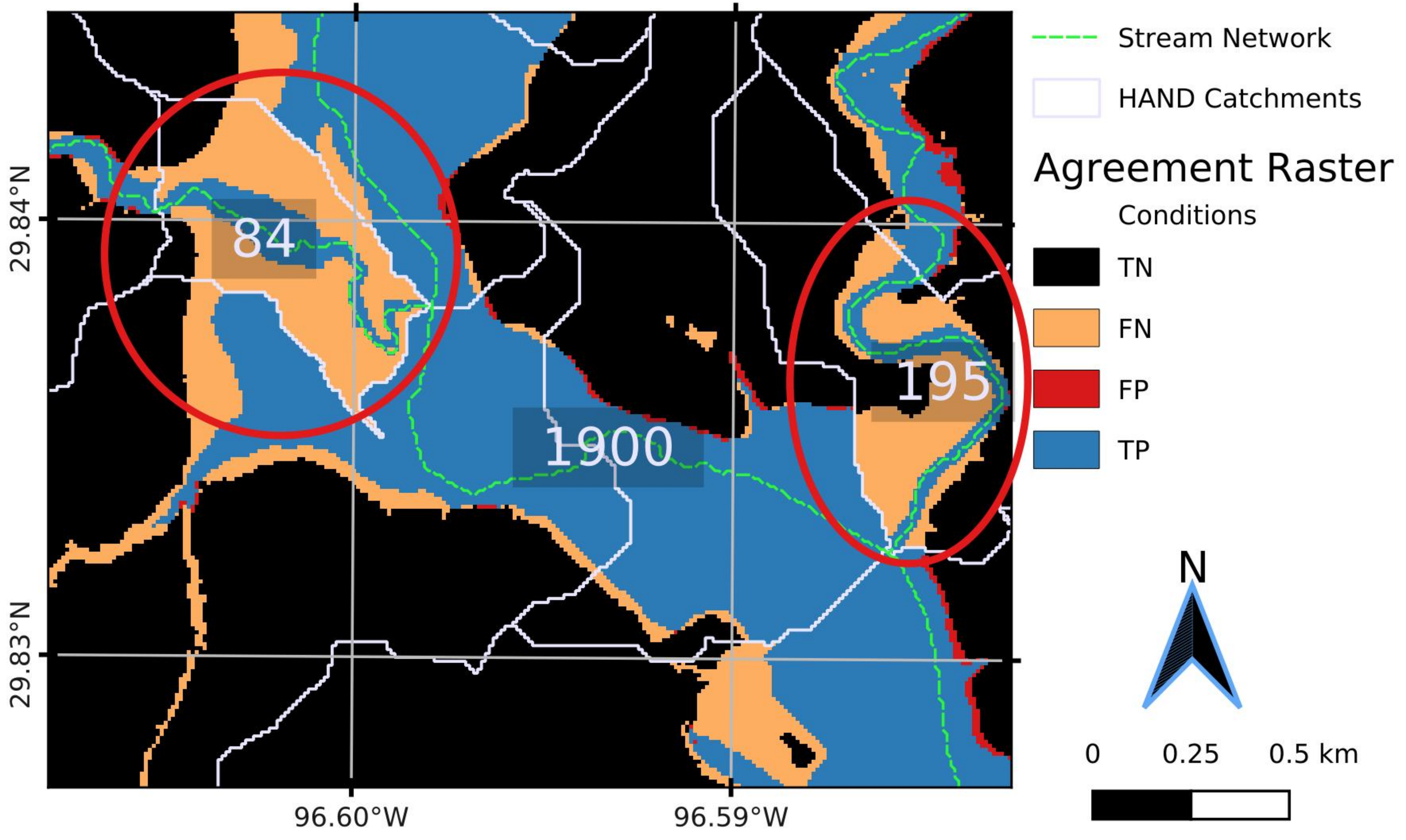

84
1900
195
29.84°N
29.83°N
96.60°W
96.59°W
Stream Network
HAND Catchments
Agreement Raster
Conditions
TN
FN
FP
TP
N
0
0.25
0.5 km

Figure 4.

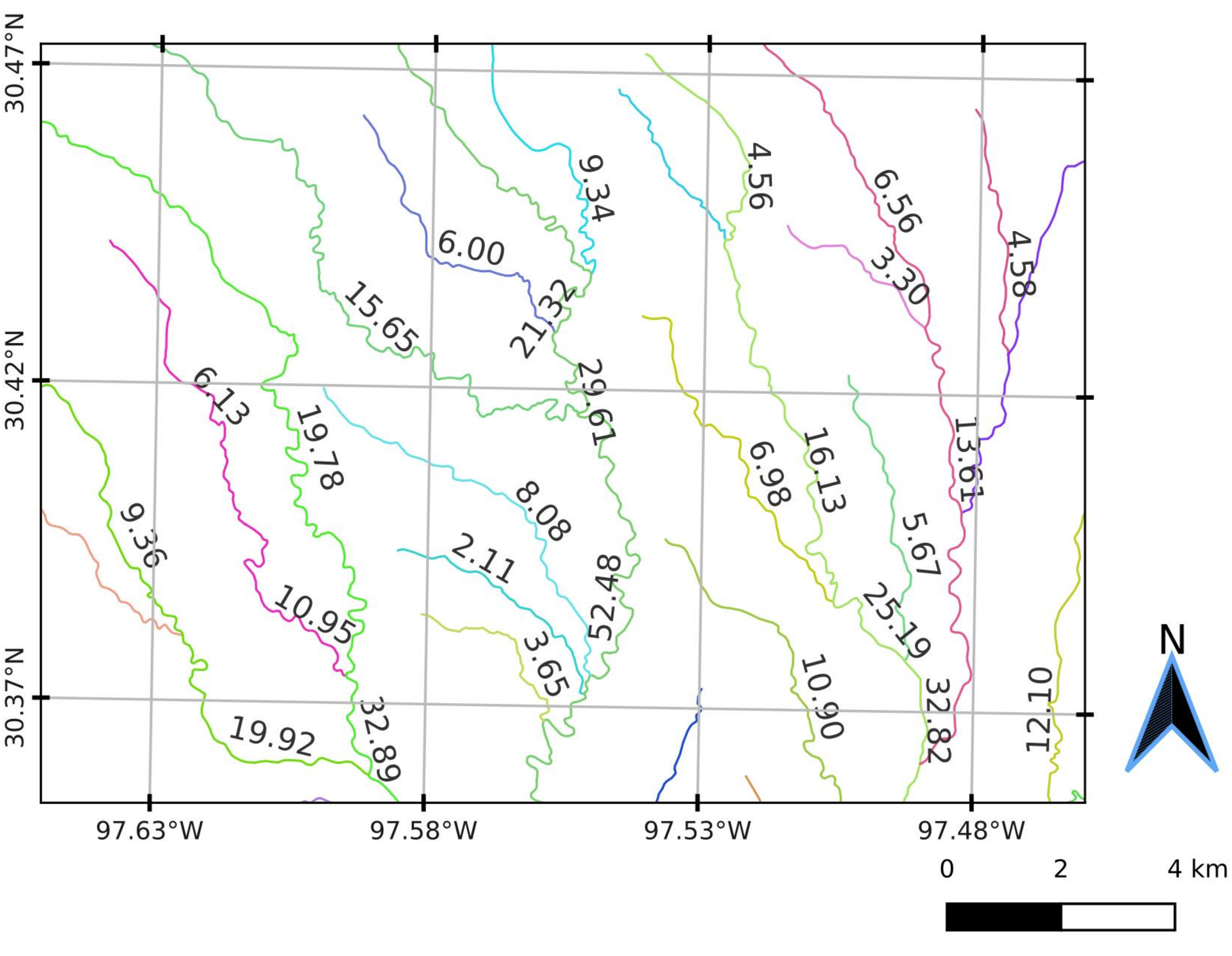
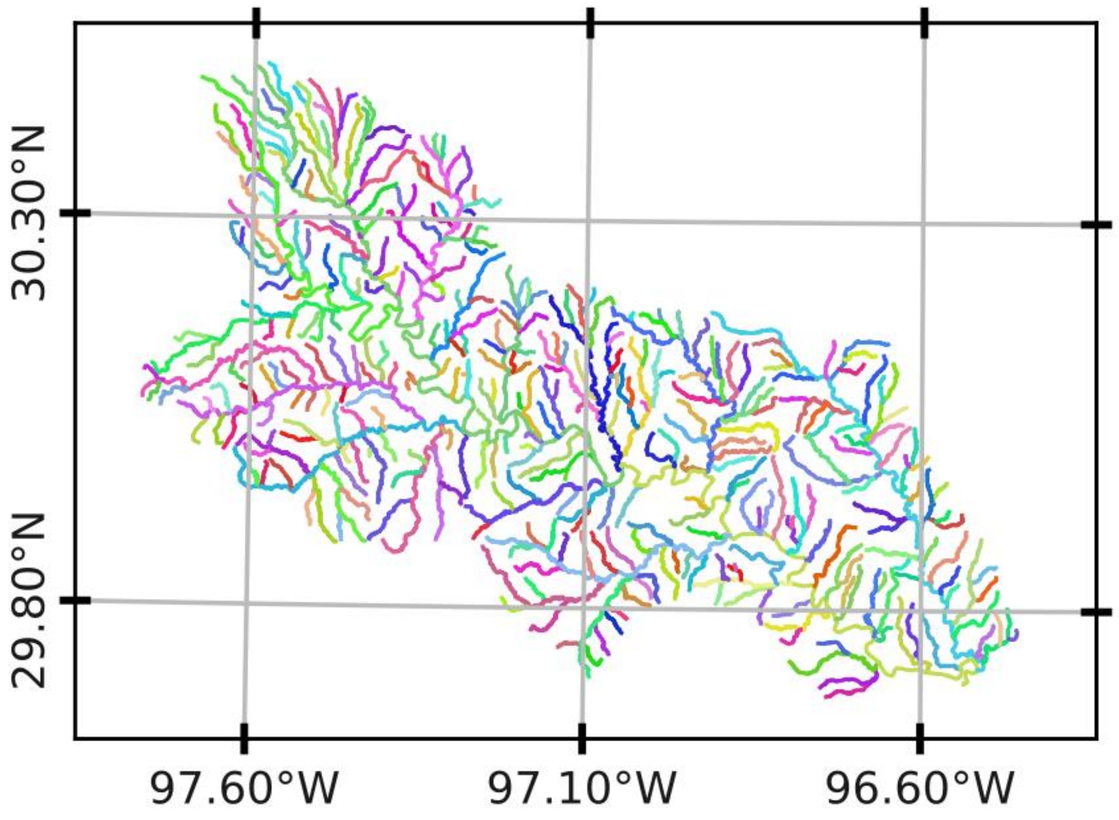
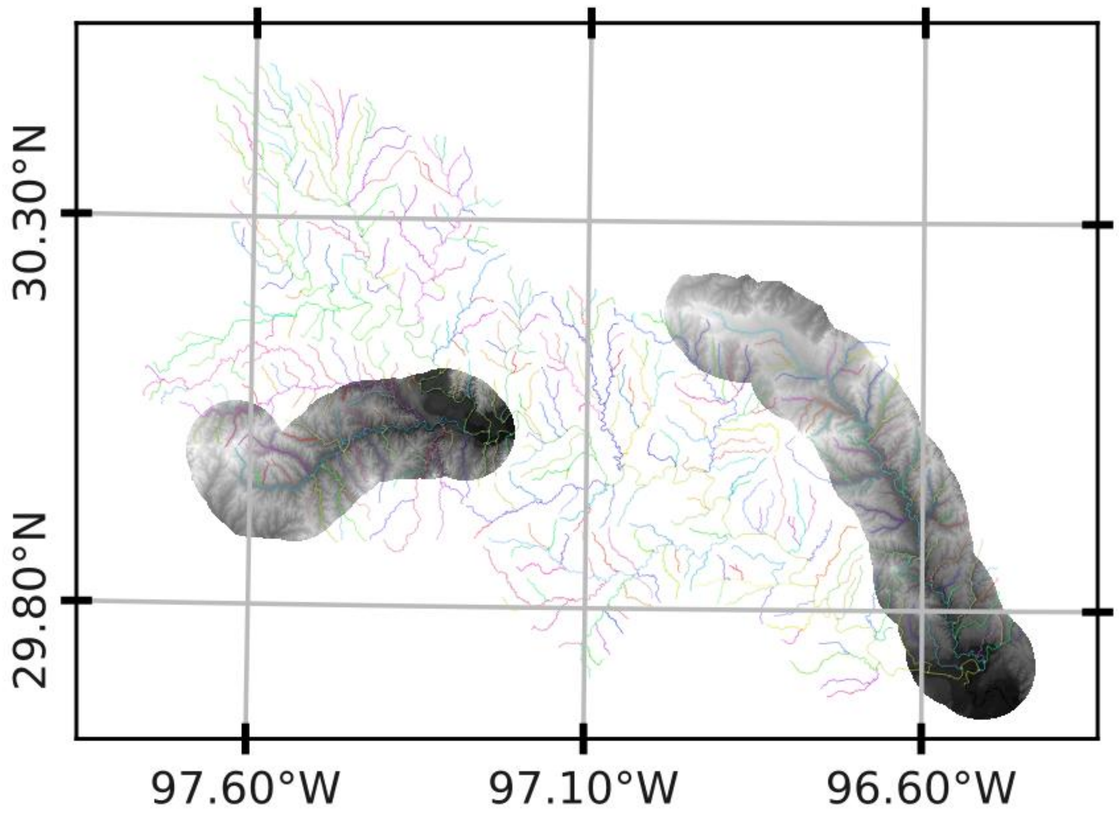

30.47°N
30.42°N
30.37°N
97.63°W
97.58°W
97.53°W
97.48°W
9.34
6.00
15.65
21.32
29.61
4.56
6.56
3.30
4.58
6.13
19.78
8.08
2.11
52.48
3.65
9.36
10.95
19.92
32.89
6.98
16.13
5.67
25.19
32.82
10.90
13.61
12.10
N
0
2
4 km

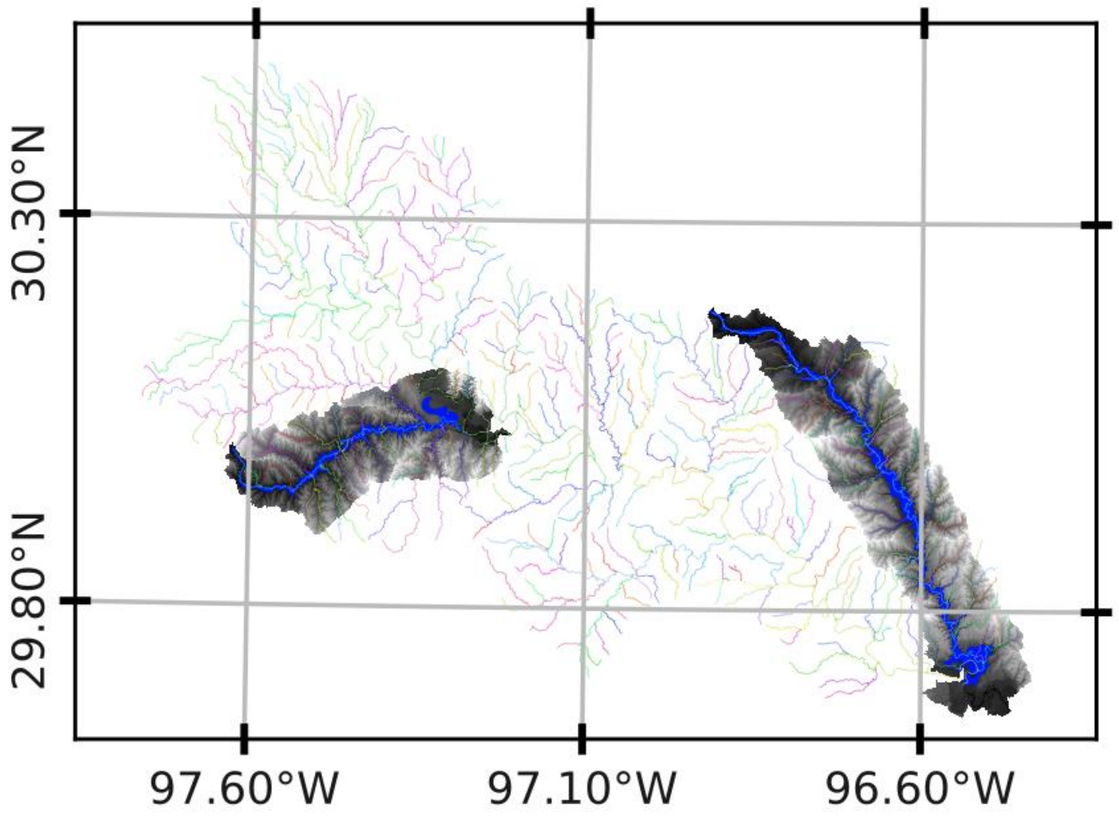
Figure 5.


30.30°N
29.80°N
97.60°W
97.10°W
96.60°W

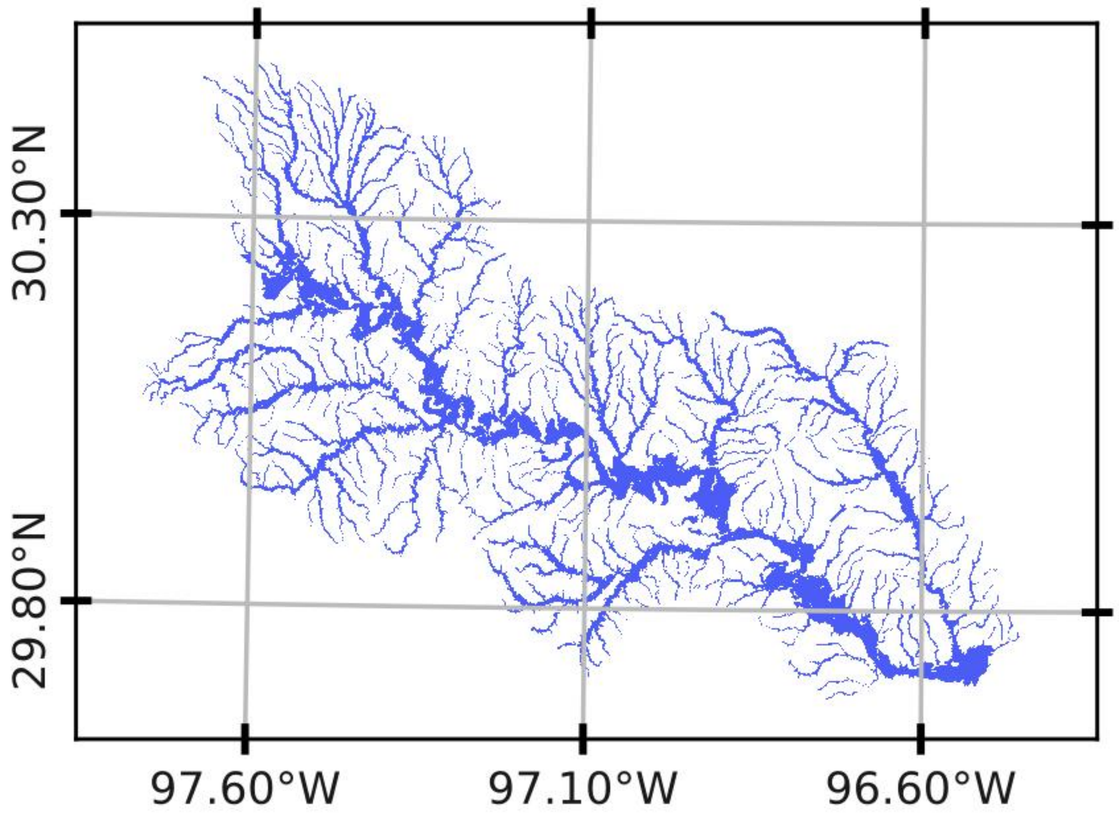
(a)


30.30°N
29.80°N
97.60°W
97.10°W
96.60°W

(b)


30.30°N
29.80°N
97.60°W
97.10°W
96.60°W

(c)


30.30°N
29.80°N
97.60°W
97.10°W
96.60°W

(d)


N

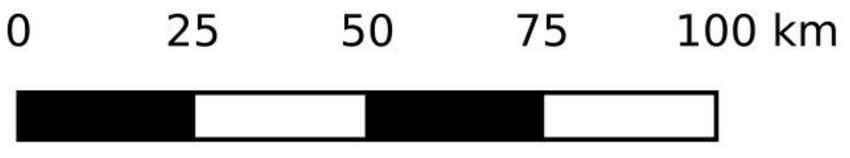

0
25
50
75
100 km

Figure 6.

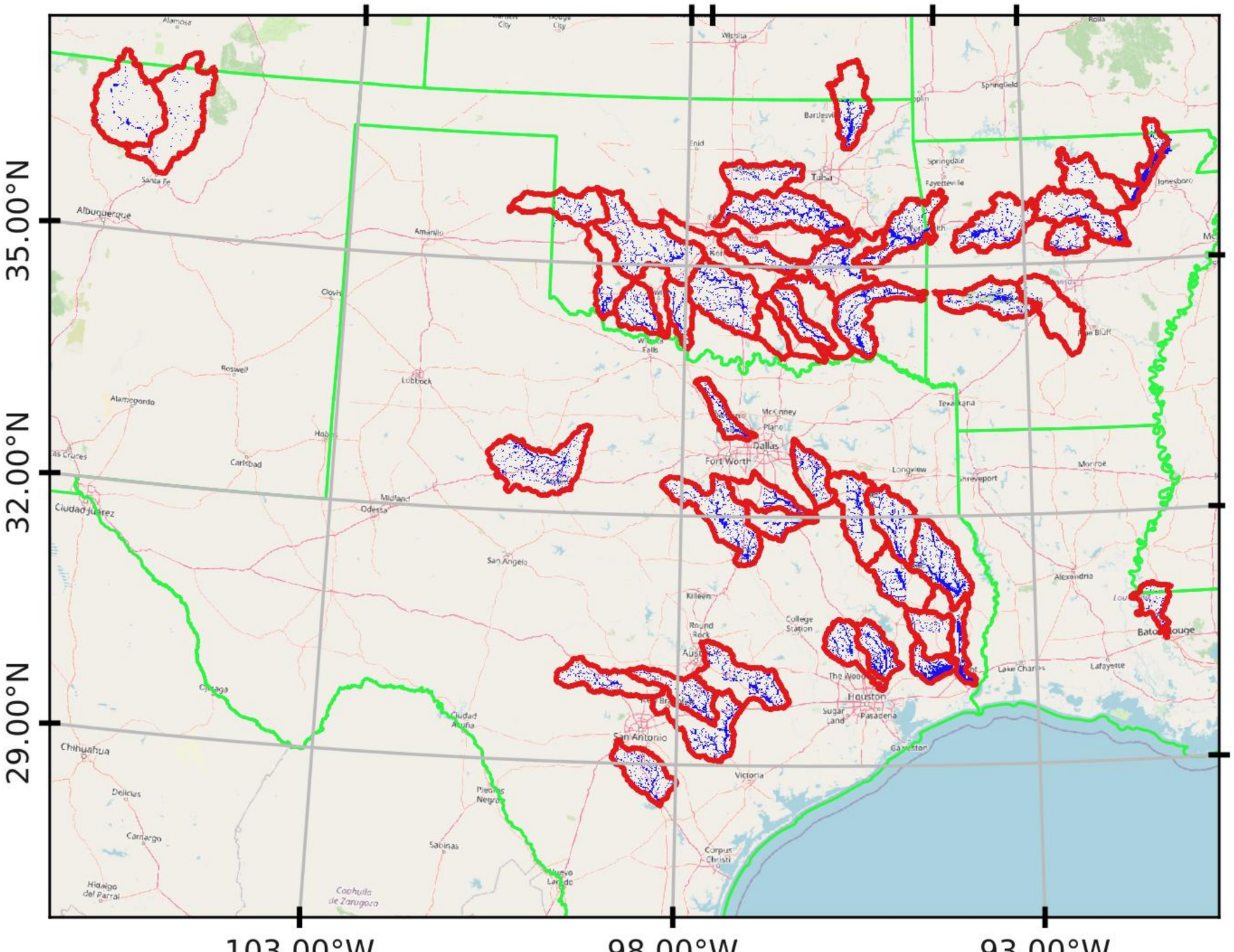

35.00°N
32.00°N
29.00°N
103.00°W
98.00°W
93.00°W
Albuquerque
Santa Fe
Amarillo
Lubbock
Roswell
Carlsbad
Midland
Odessa
San Angelo
Abilene
Wichita Falls
Dallas
Fort Worth
McKinney
Plano
Tulsa
Enid
Springfield
Springdale
Fayetteville
Jonesboro
Texarkana
Longview
Shreveport
Monroe
Alexandria
Lafayette
Lake Charles
Houston
Pasadena
Sugar Land
The Woodlands
College Station
Killeen
Round Rock
Austin
San Antonio
Victoria
Corpus Christi
Galveston
Laredo
Chihuahua
Ciudad Juárez
Coahuila de Zaragoza
HUC8 Boundaries
0.2% BLE Benchmark
Inundated
US State Borders
OpenStreetMap
N
0
200
400 km

Figure 7.

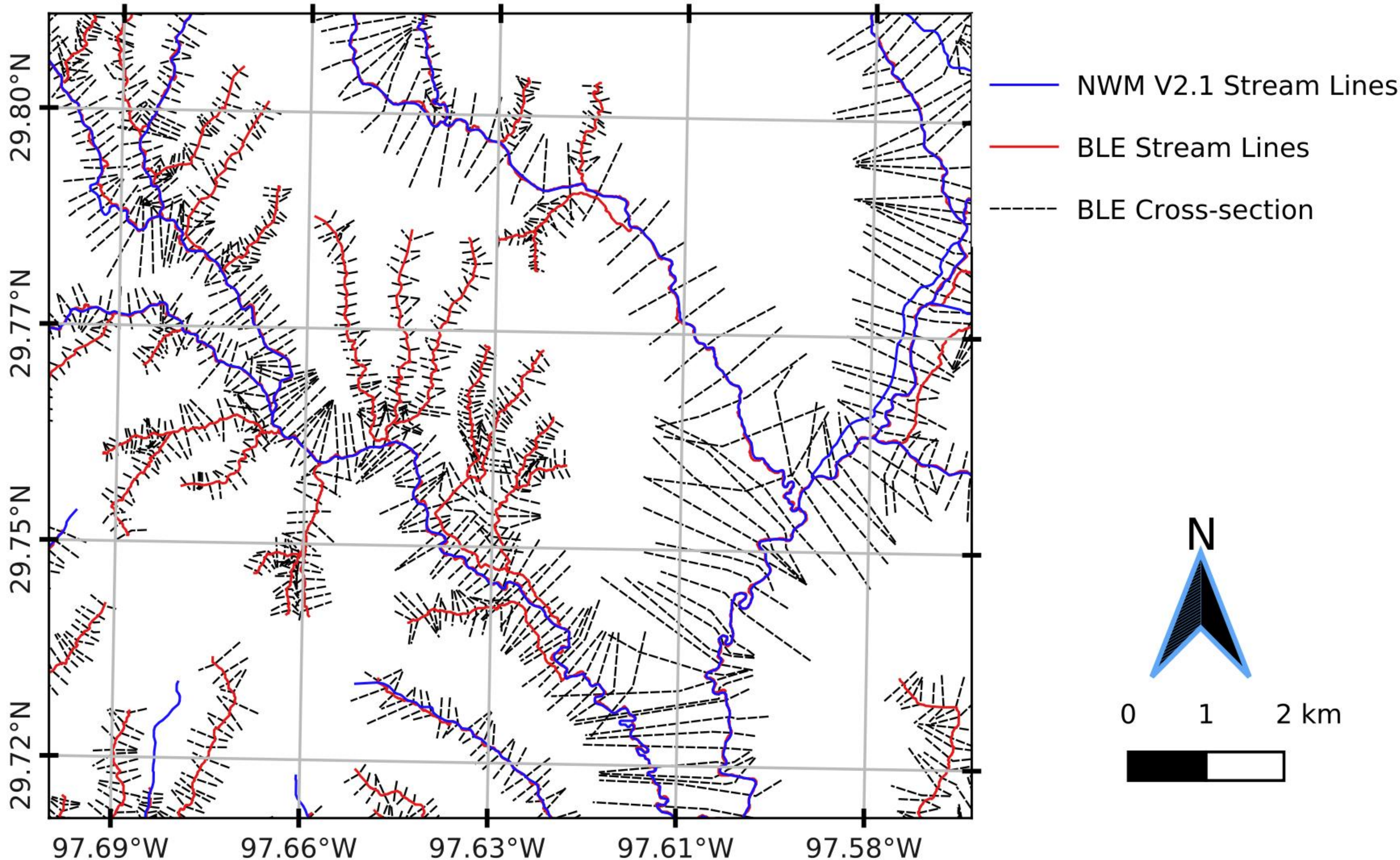

29.80°N
29.77°N
29.75°N
29.72°N
97.69°W
97.66°W
97.63°W
97.61°W
97.58°W
NWM V2.1 Stream Lines
BLE Stream Lines
BLE Cross-section
N
0
1
2 km

Figure 8.

Manning's N = 0.06

Manning's N = 0.12

β_1 = 0.0098 | p-value = 0.025

β_1 = 0.0133 | p-value = 0.003

β_1 = 0.0146 | p-value = 0.008

β_1 = 0.0188 | p-value = 0.001

β_1 = -0.0016 | p-value = 0.392

β_1 = -0.0037 | p-value = 0.284

CSI

POD

FAR

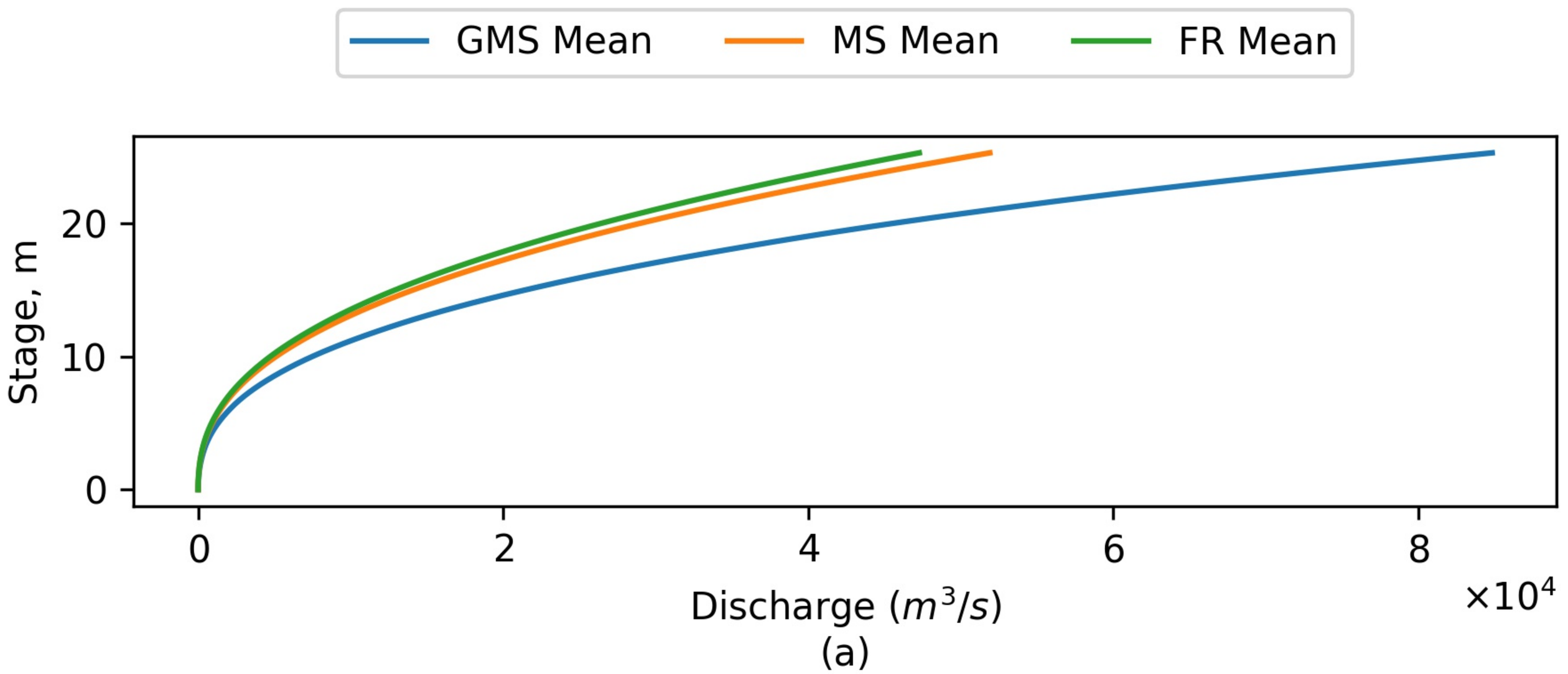
0.70
0.60
0.50
0.85
0.75
0.65
0.55
0.35
0.25
0.15
0.05

FR MS GMS FR MS GMS

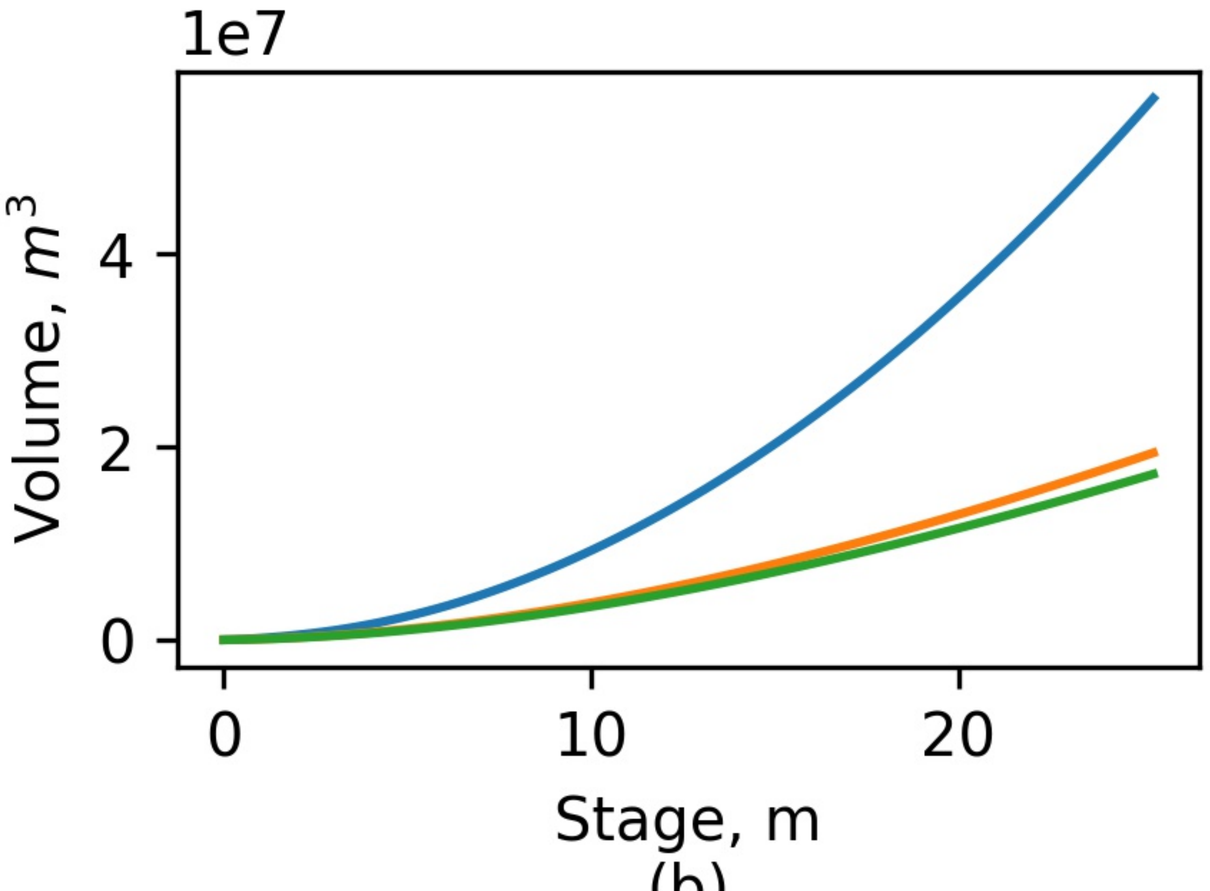
Model

Trend Line 100yr 500yr

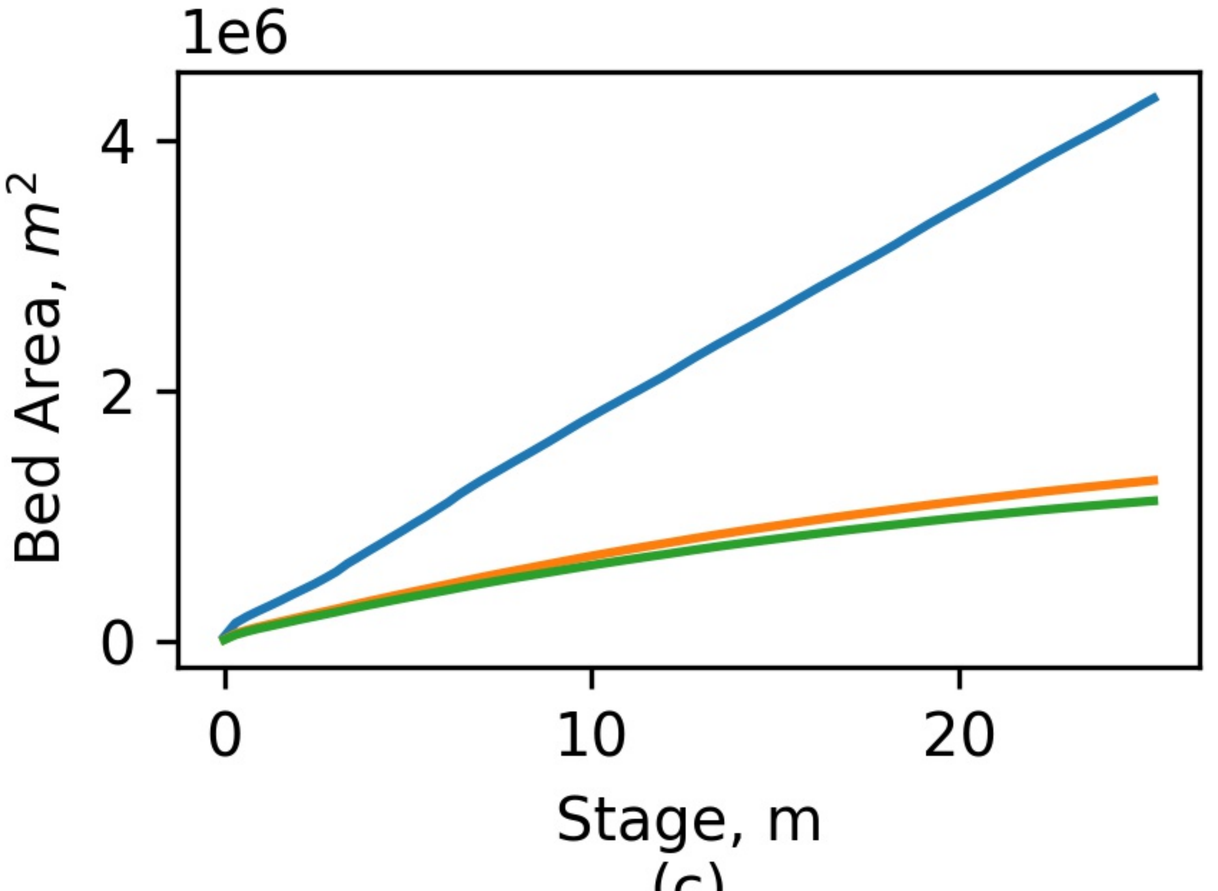
Figure 9.


GMS Mean
MS Mean
FR Mean
Stage, m
20
10
0
0
2
4
6
8
×10^4
Discharge (m^3/s)

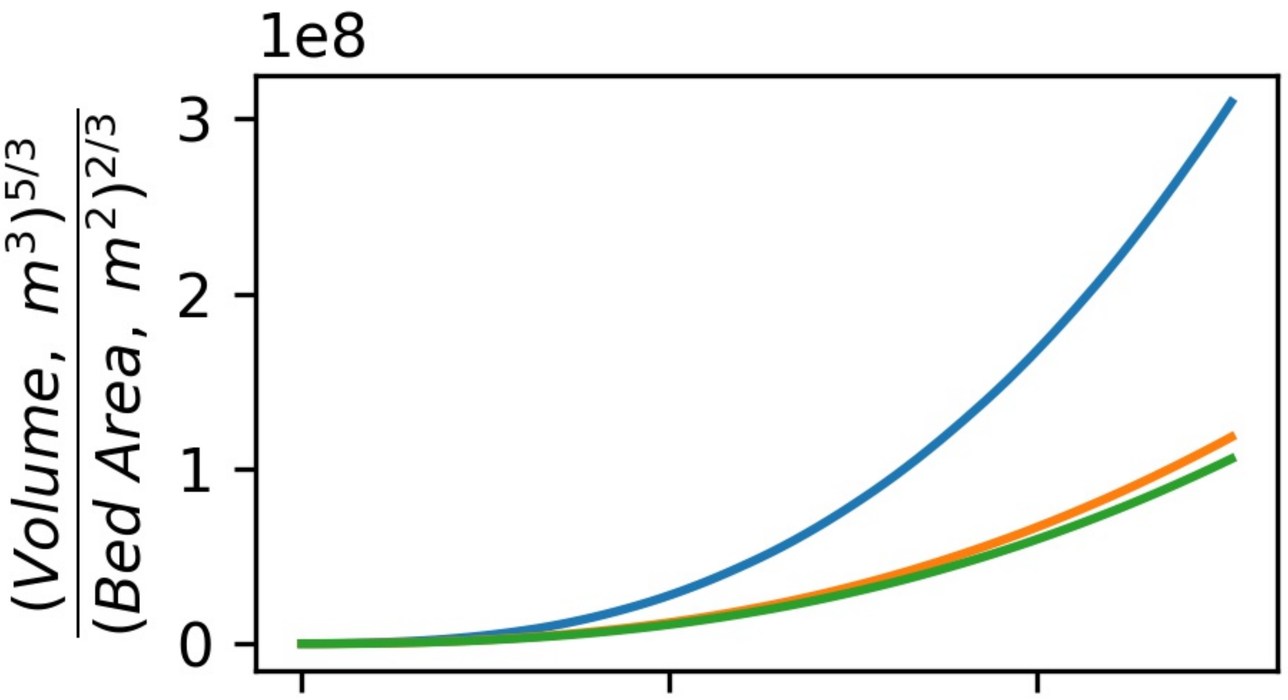
(a)


1e7
Volume, m^3
4
2
0
0
10
20
Stage, m

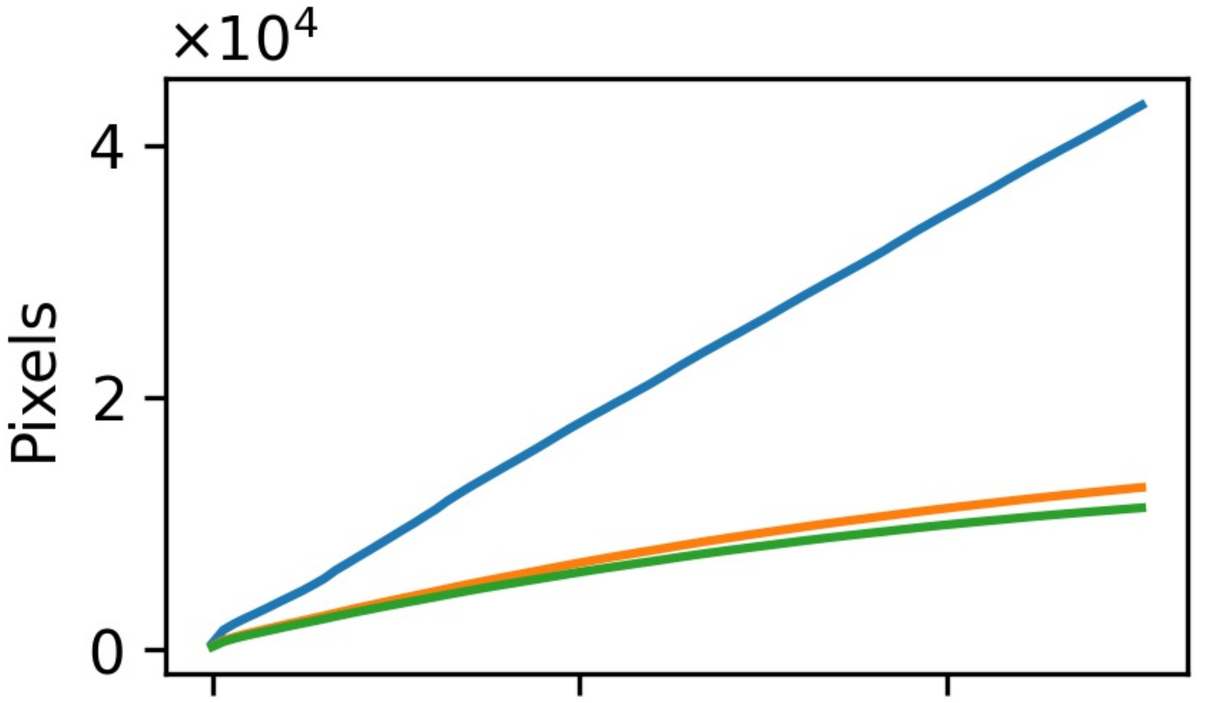
(b)


1e6
Bed Area, m^2
4
2
0
0
10
20
Stage, m

(c)


1e8
$\frac{(Volume,\ m^3)^{5/3}}{(Bed\ Area,\ m^2)^{2/3}}$
3
2
1
0
0
10
20
Stage, m

(d)


×10^4
Pixels
4
2
0
0
10
20
Stage, m

(e)

Figure 10.

34.70°N
34.68°N
34.66°N
95.00°W
94.96°W
94.92°W
94.88°W

(a)

34.70°N
34.68°N
34.66°N
95.00°W
94.96°W
94.92°W
94.88°W

(b)

HAND Catchments
NWM Stream Network
N
0 1 2 km

Agreement Raster
TN
FN
FP
TP
Waterbody mask