High Performance Computing to Support Multiscale Representation of Hydrography for the Conterminous United States

Lawrence V. Stanislawski[1], Yan Liu[2], Barbara P. Buttenfield[3], Kornelijus Survila[2], Jeffrey Wendel[1], Abdurraouf Okok[1]

[1] U.S. Geological Survey, Center of Excellence for Geospatial Information Science, 1400 Independence Road, Rolla MO 65401
Email: lstan@usgs.gov, jwendel@usgs.gov, aokok@usgs.gov

[2] University of Illinois at Urbana-Champaign
Email: yanliu@illinois.edu, survila2@illinois.edu

[3] University of Colorado-Boulder, Boulder, CO 80309-0260
Email: babs@colorado.edu

19th ICA Workshop, Automated Generalisation for On-Demand Mapping, Helsinki, Finland June 14, 2016

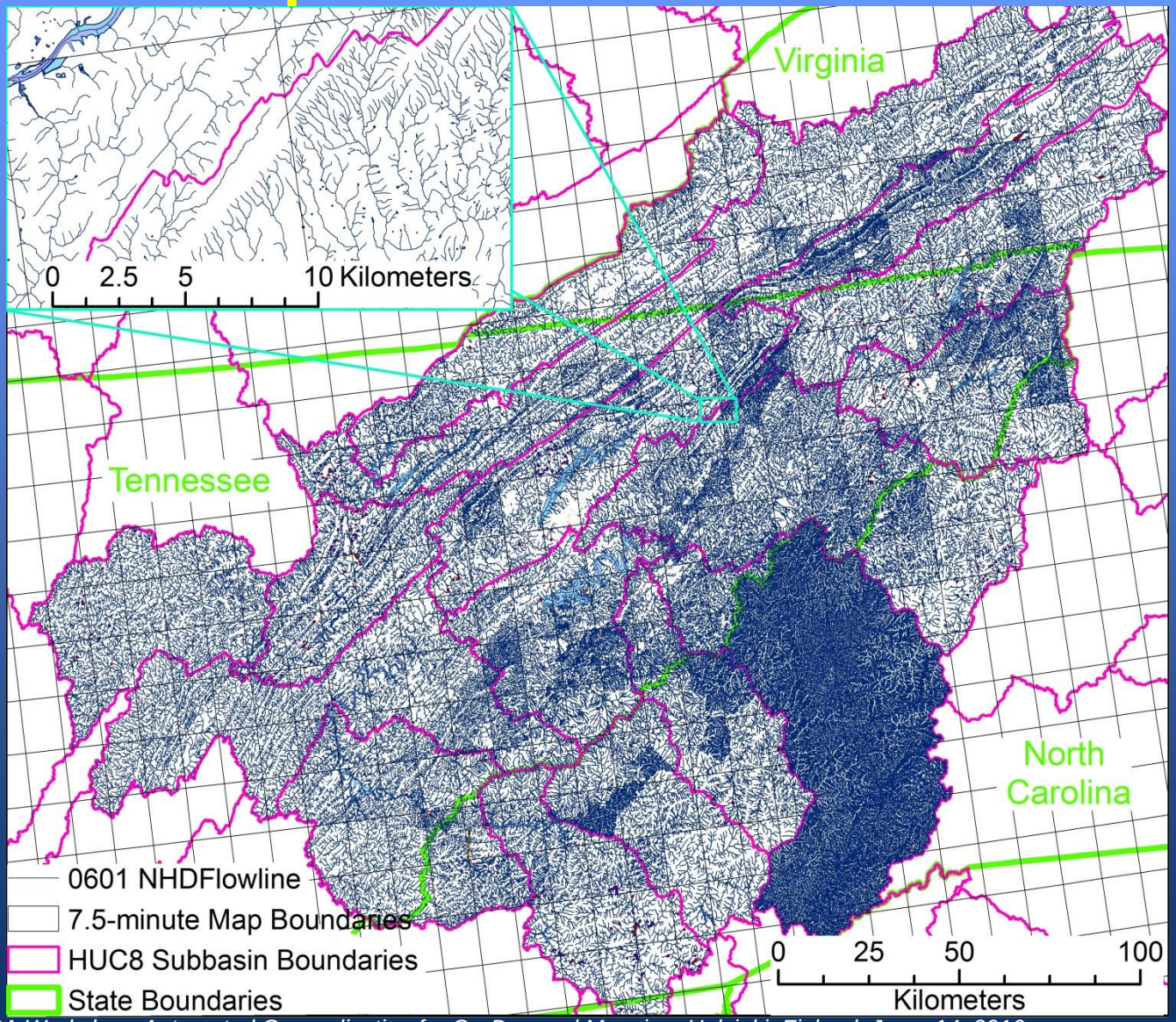
Outline

- Method to prune the high-resolution National Hydrography Dataset (NHD) in a hydrological consistent manner to 1:24,000 (24K) and smaller scales by estimating 24K target densities
  - Weighted Flow Accumulation (WFA) model
- Workflow to estimate target densities using open source and TauDEM on a Linux cluster high-performance computing environment
- Some results for NHD Regions 6, 7, and 10
- Summary and future work

USGS

Problem: Drainage density variations in high-resolution (HR) NHD are caused by natural conditions and compilation differences

16 subbasins in subregion 0601 (Region 6, subregion 1)



USGS

Objective: Estimate 1:24,000-scale (24K) density patterns from channels derived from a weighted flow accumulation (WFA) model

- Derive a natural density pattern of surface water channels (at 24K) that varies with terrain, runoff, soil permeability, soil depth, ground water, and vegetative cover
- Derived 24K density pattern is used to generalize high-resolution (HR) NHD to 24K and smaller scales (Radical Law)
  - Apply NHD Generalization Tools
- Advantages:
  - Can eliminate compilation inconsistencies in HR NHD while maintaining natural variation
  - Through automated comparisons of HR NHD and 24K derived channels, missing or improper content in HR NHD can be identified

USGS

Overview of workflow

Process	Commercial Software	Open Source/TauDEM
Estimate 24K parameters (total length, number of features, minimum length of first order tributaries, devoid polygons)	2 to 6 minutes per HUC8	
Extract 24K drainage channels from 1/3rd arc-second elevation model using weighted flow accumulation model	4 to 30 hours or more per HUC8, sequential processing	4 minutes to 6.4 hours per HUC8, simultaneous processing
Build line-density partitions for high-resolution NHD flowline	1 to 4 minutes per HUC8	
Compute 24K and smaller scale target densities for HR NHD partitions	5 to 10 seconds per HUC8	

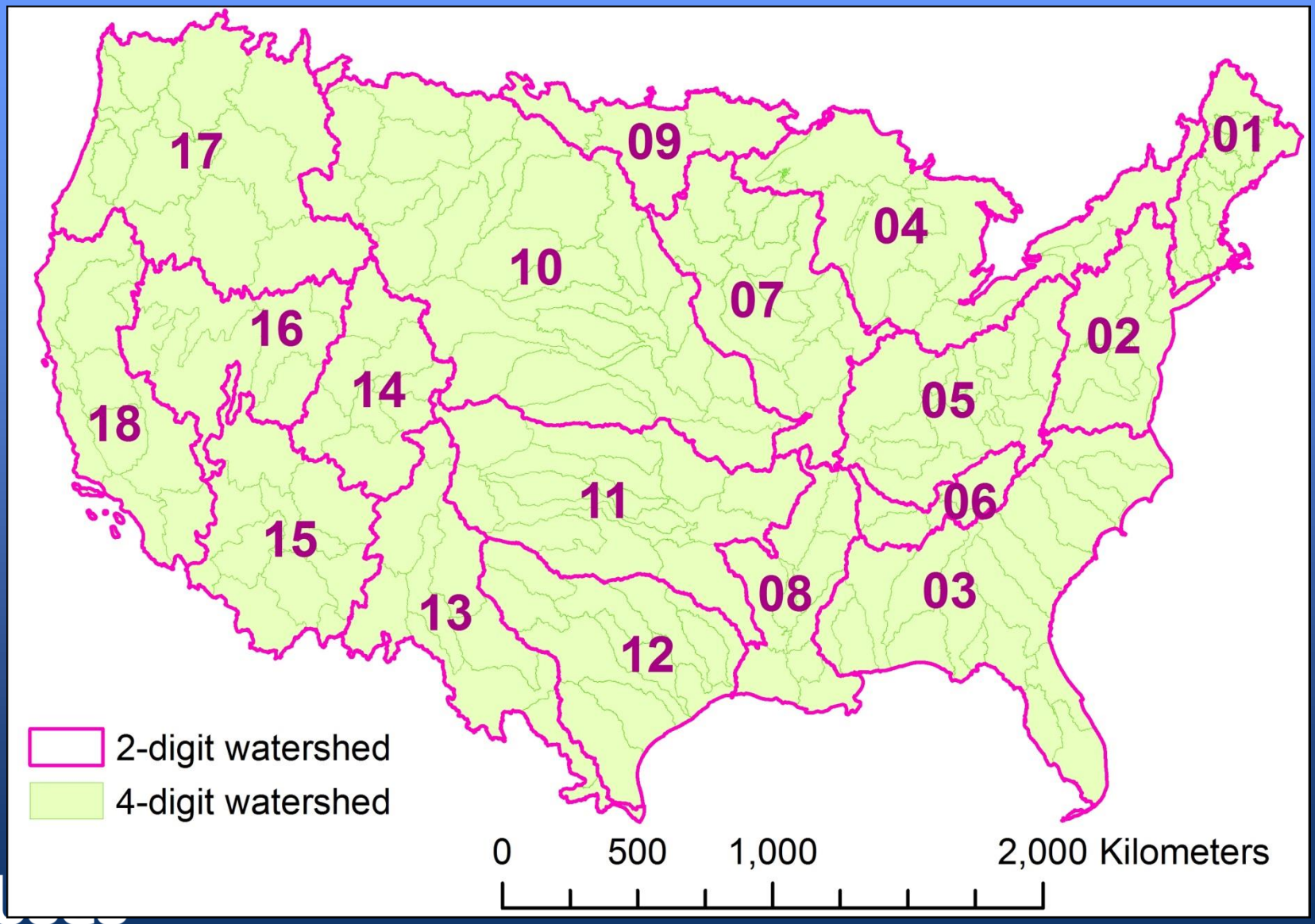
447 HUC8 subbasins in about 16 hours

High-Performance Processing on Linux Cluster with Open Source tools

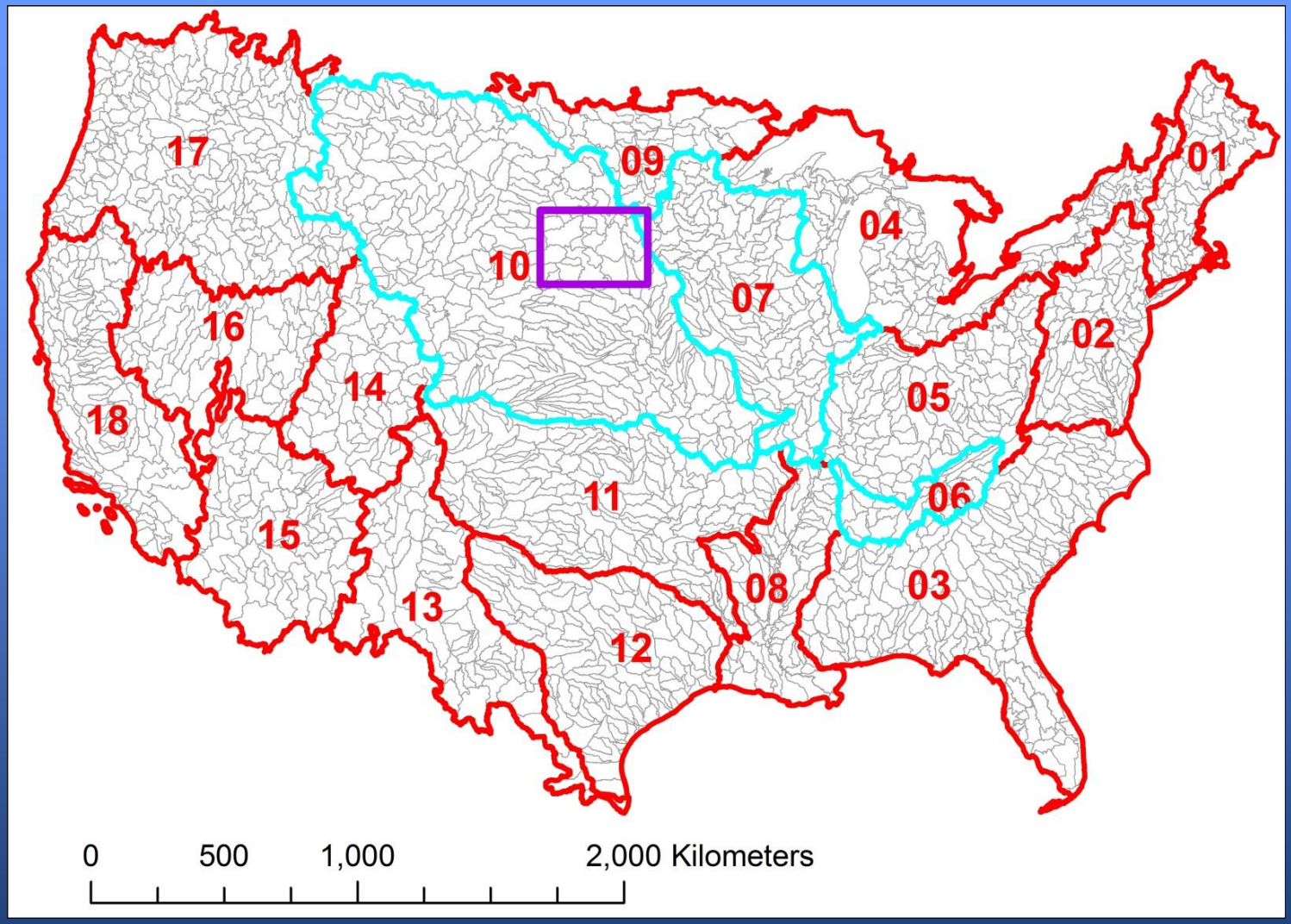
- Vector and raster operations – Geospatial Data Abstraction Library (GDAL) deployed through C, C++, and Python libraries
- Raster channel extraction – Terrain Analysis Using Digital Elevation Models (TauDEM) employs parallel programming to enable data decomposition and parallel input and output of large DEMs
- Workflow implemented through Python and Simple Linux Utility for Resource Manager (SLURM) programs.
- Hardware: 5-node Linux cluster with 20 processing cores and 64 GB of RAM per node, with parallel Lustre file system and high-speed Infiniband interconnect.
- One similar hardware node connected to the cluster is running Windows for completing the geoprocessing steps that use commercial software (Esri ArcGIS® Desktop tools implemented through Python).

USGS

National Hydrography Dataset (NHD)
Region (HUC2) and Subregion (HUC4) Watersheds



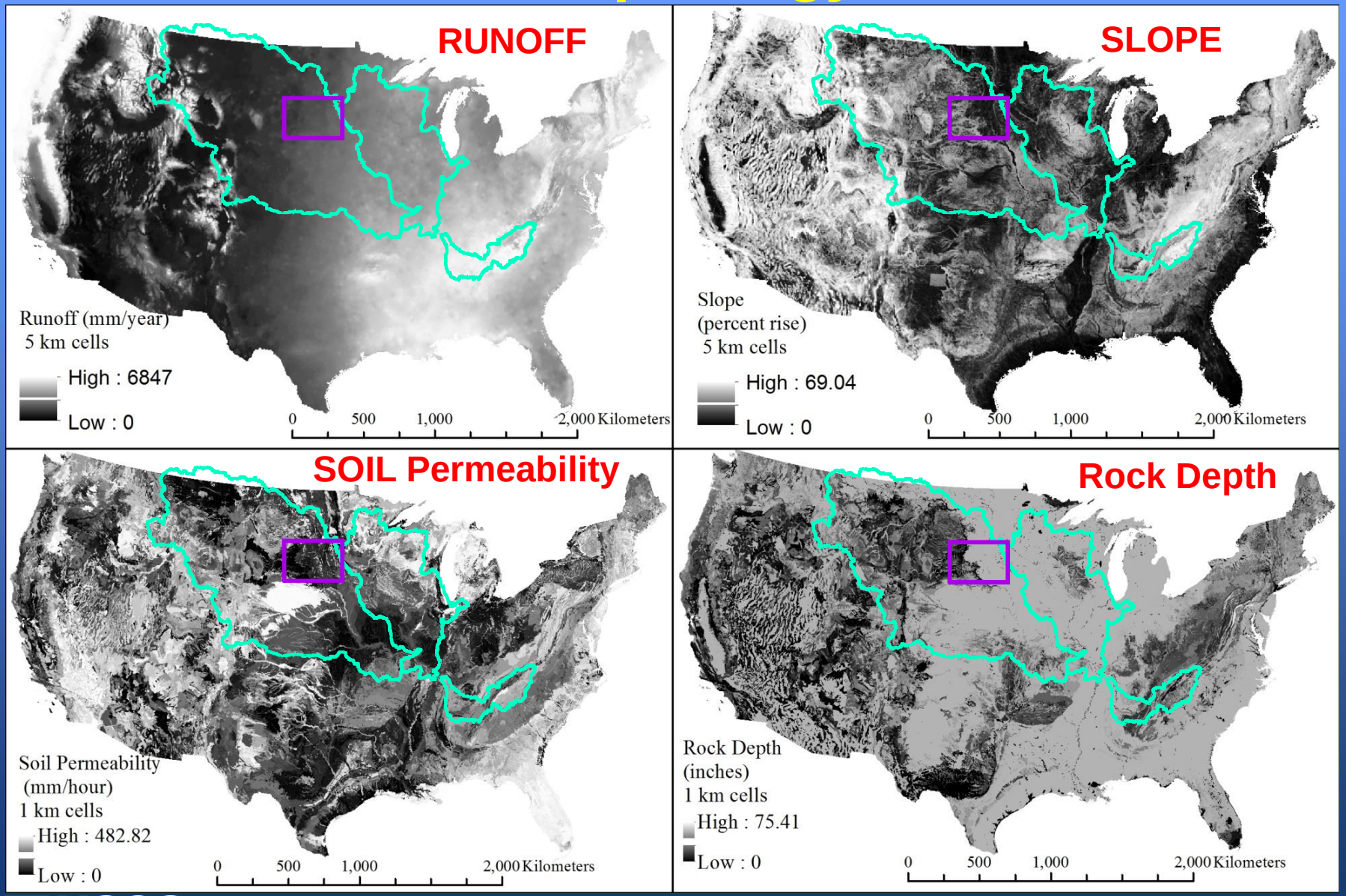
Test Area: Regions 6, 7 and 10 HUC8 subbasins



Hydrologic Unit 2 (HUC2) Region watersheds: 6, 7, and 10, excluding a two subregions that straddle the Canadian border. This includes 32, 131, and 284 subbasin watersheds (total 447).

USGS

Stream Geomorphology Conditions

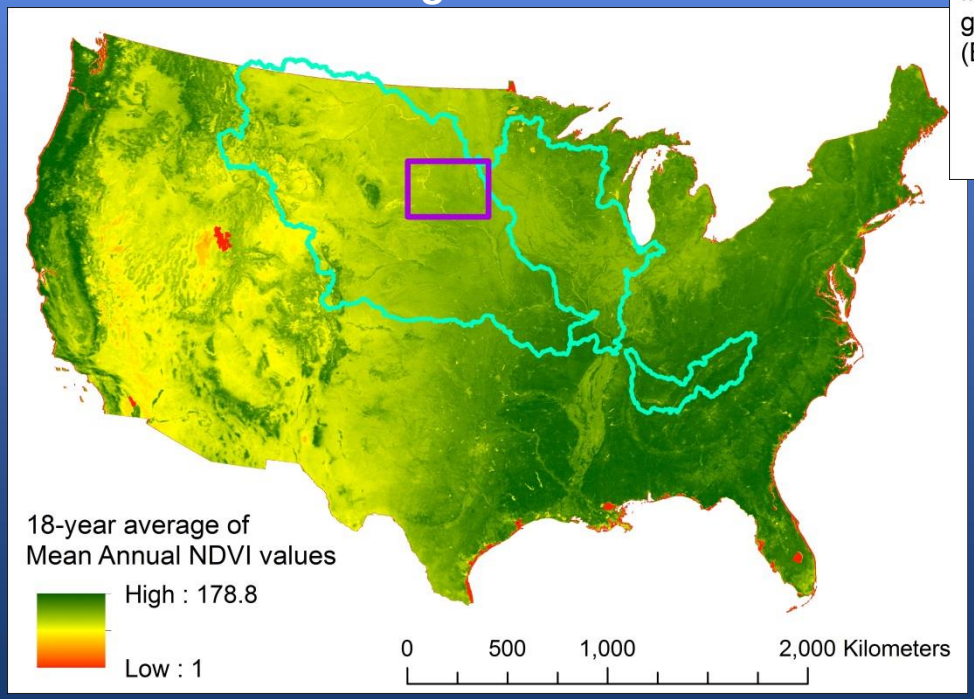


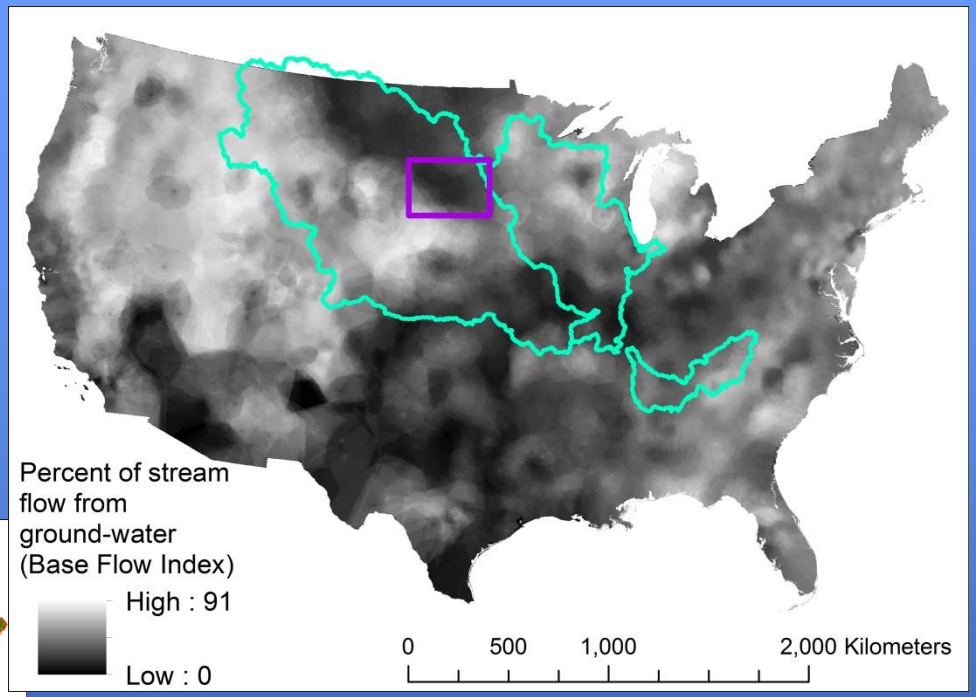
USGS

Stream Geomorphology Conditions

Purple box is sample area in the mid-west plains spanning to non-glacial to glacial area (from west to east)

1-km resolution 18-year NDVI average. Tested in the weighted flow accumulation model to adjust runoff for the effects of vegetation.

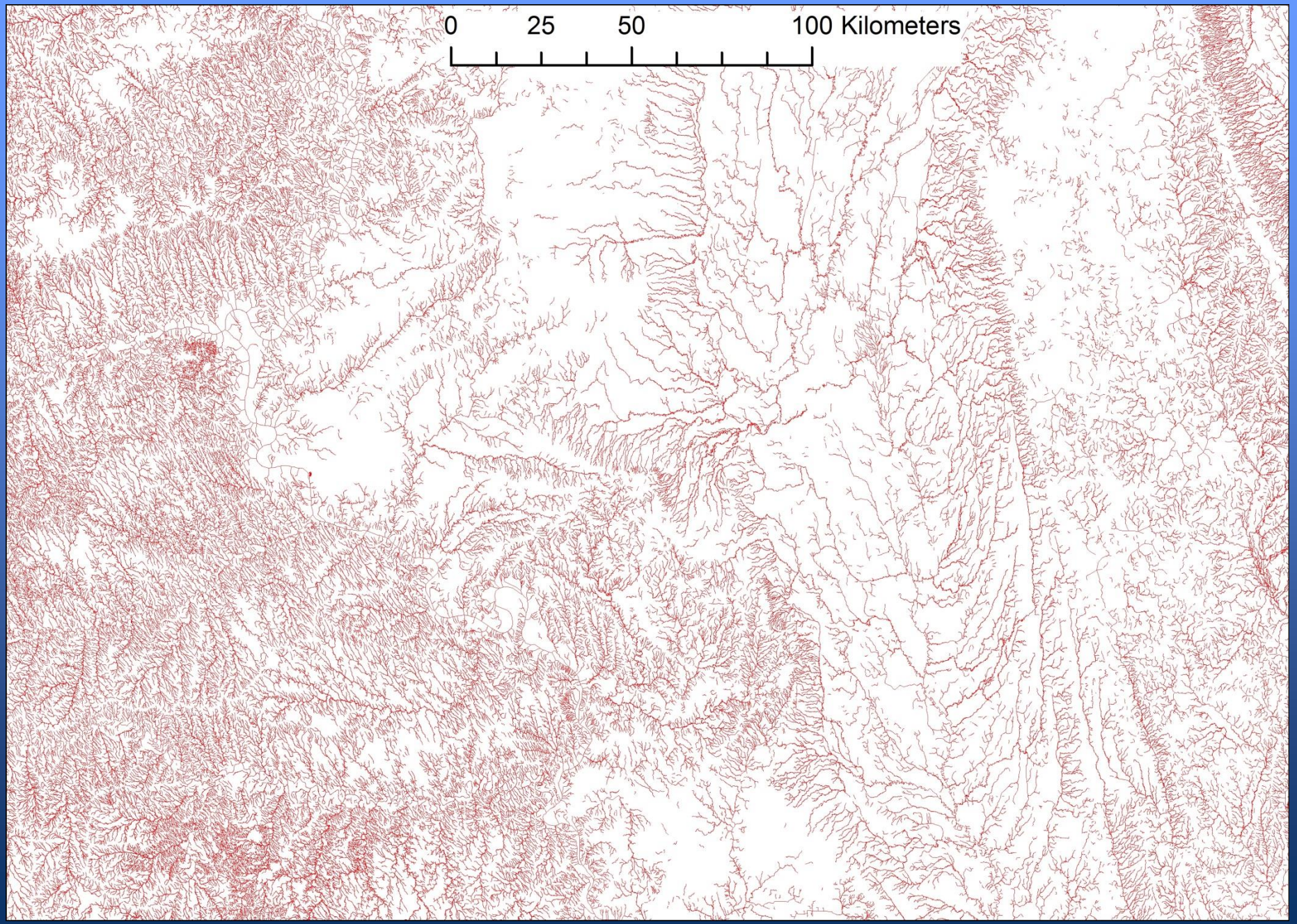




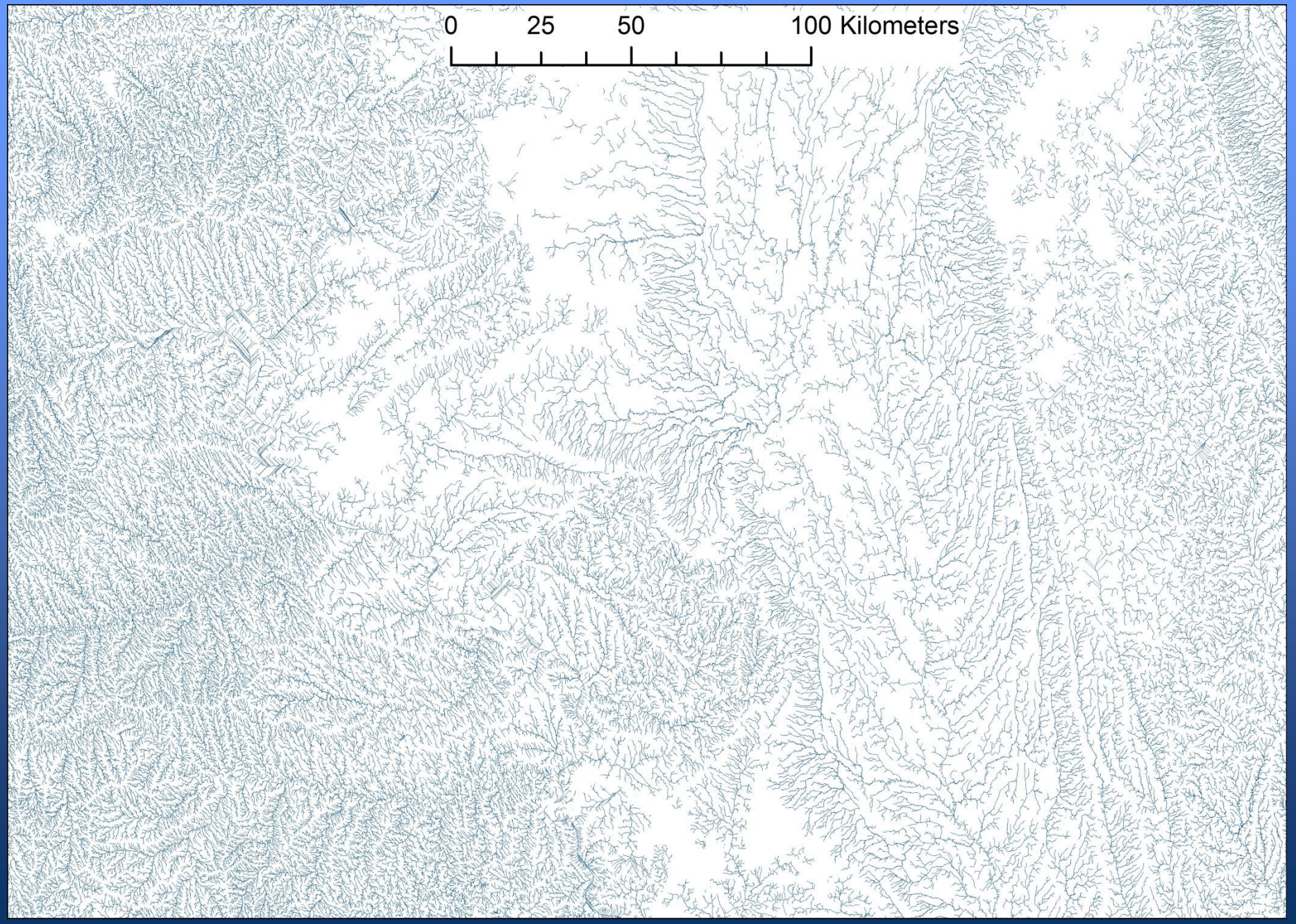
1-km resolution Base Flow Index, or percent of stream flow contributed by ground water.

USGS

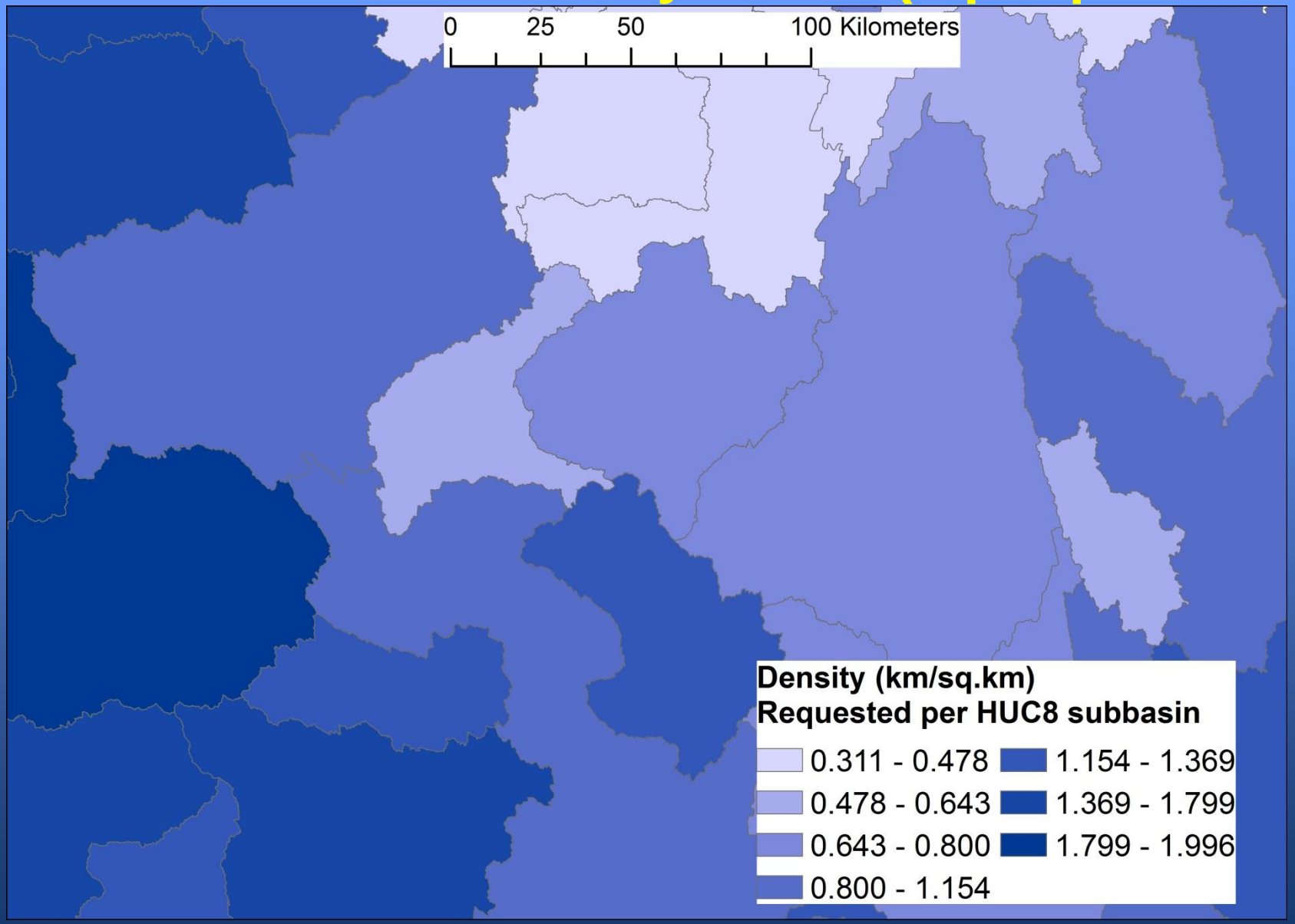
Results: 24K NHD Flowline Features



Results: 24K Extracted Channels (TauDEM)

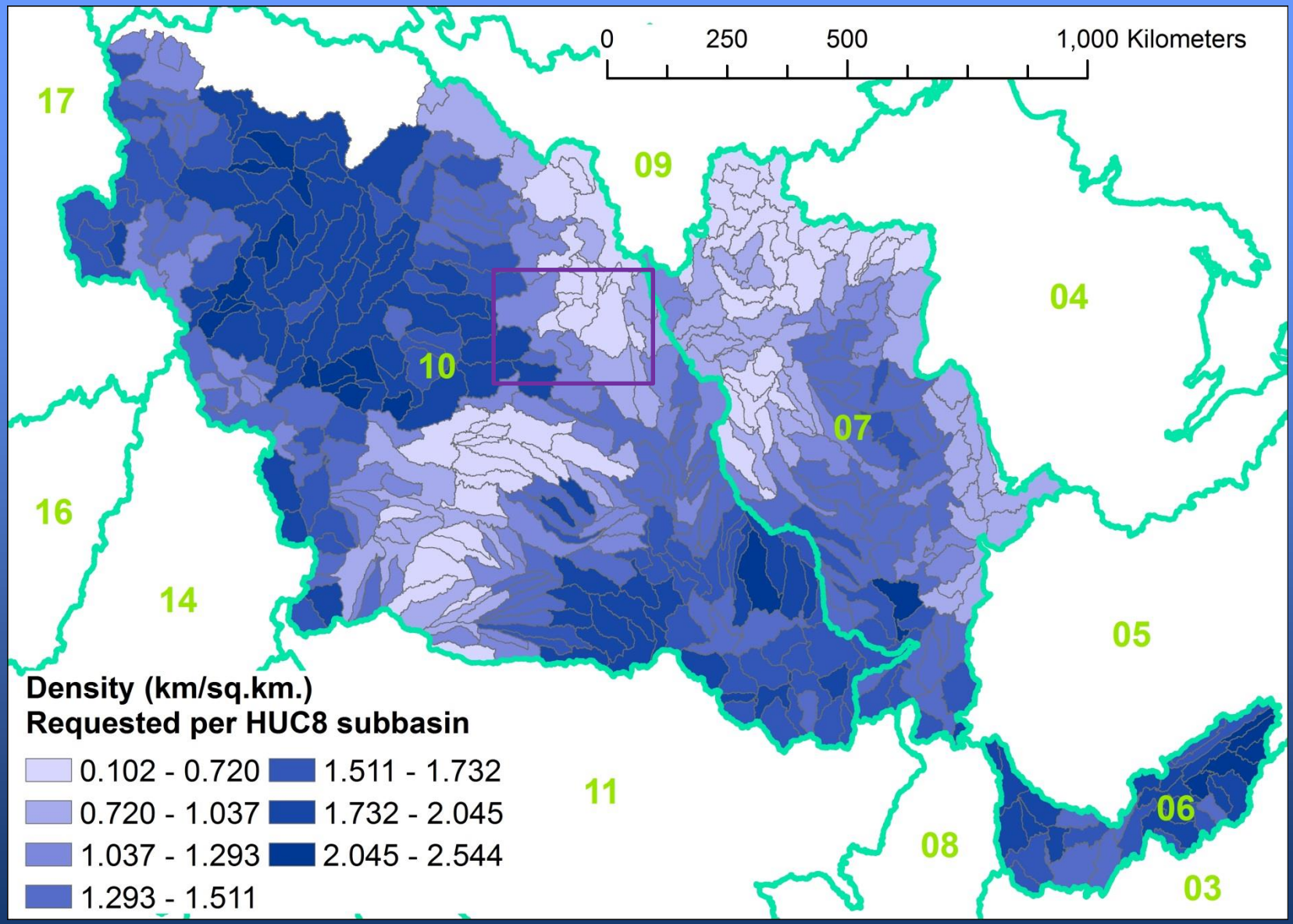


Results: Subbasin Density at 24K (input parameter)



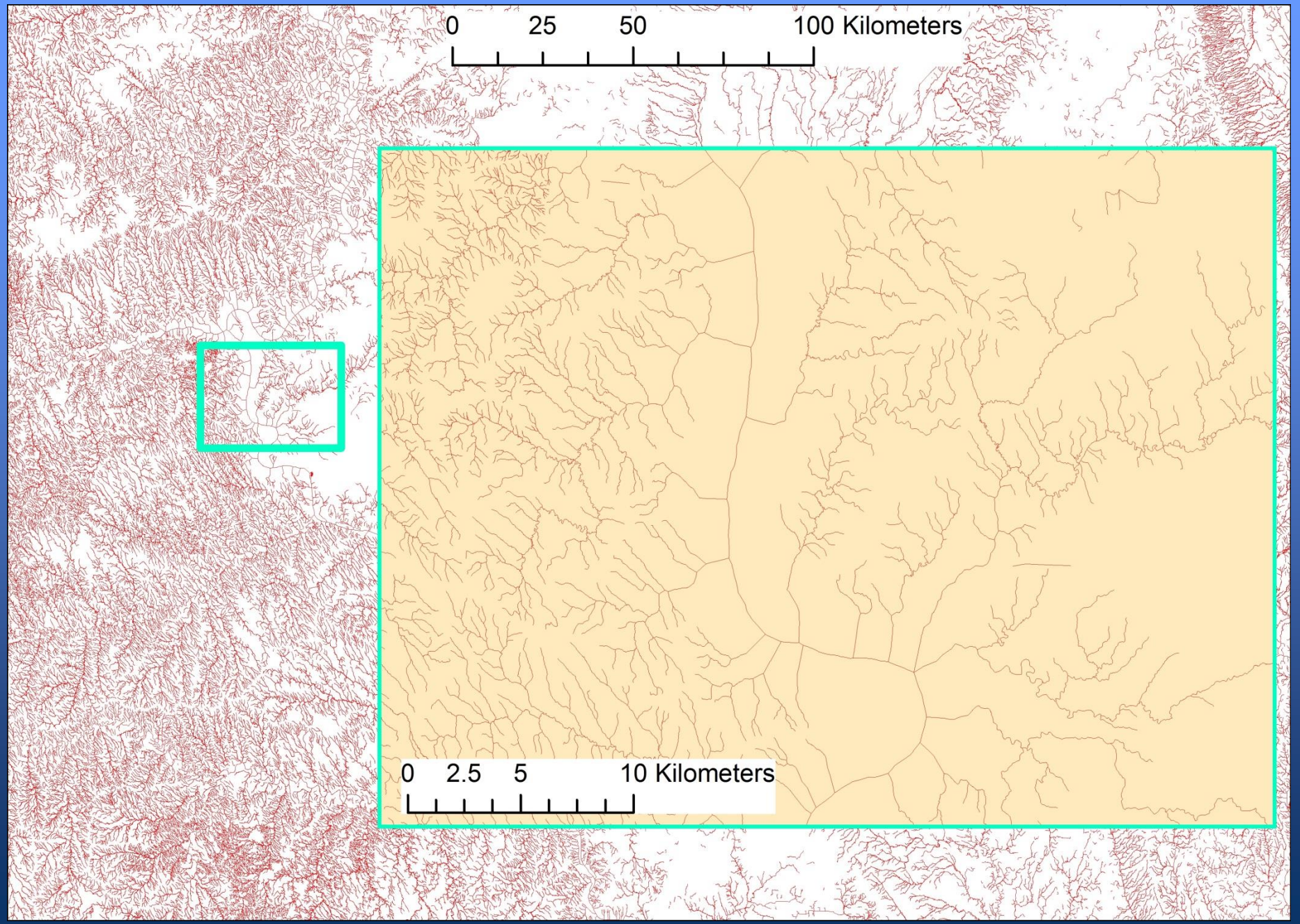
USGS

Results: Subbasin Density at 24K (input parameter)

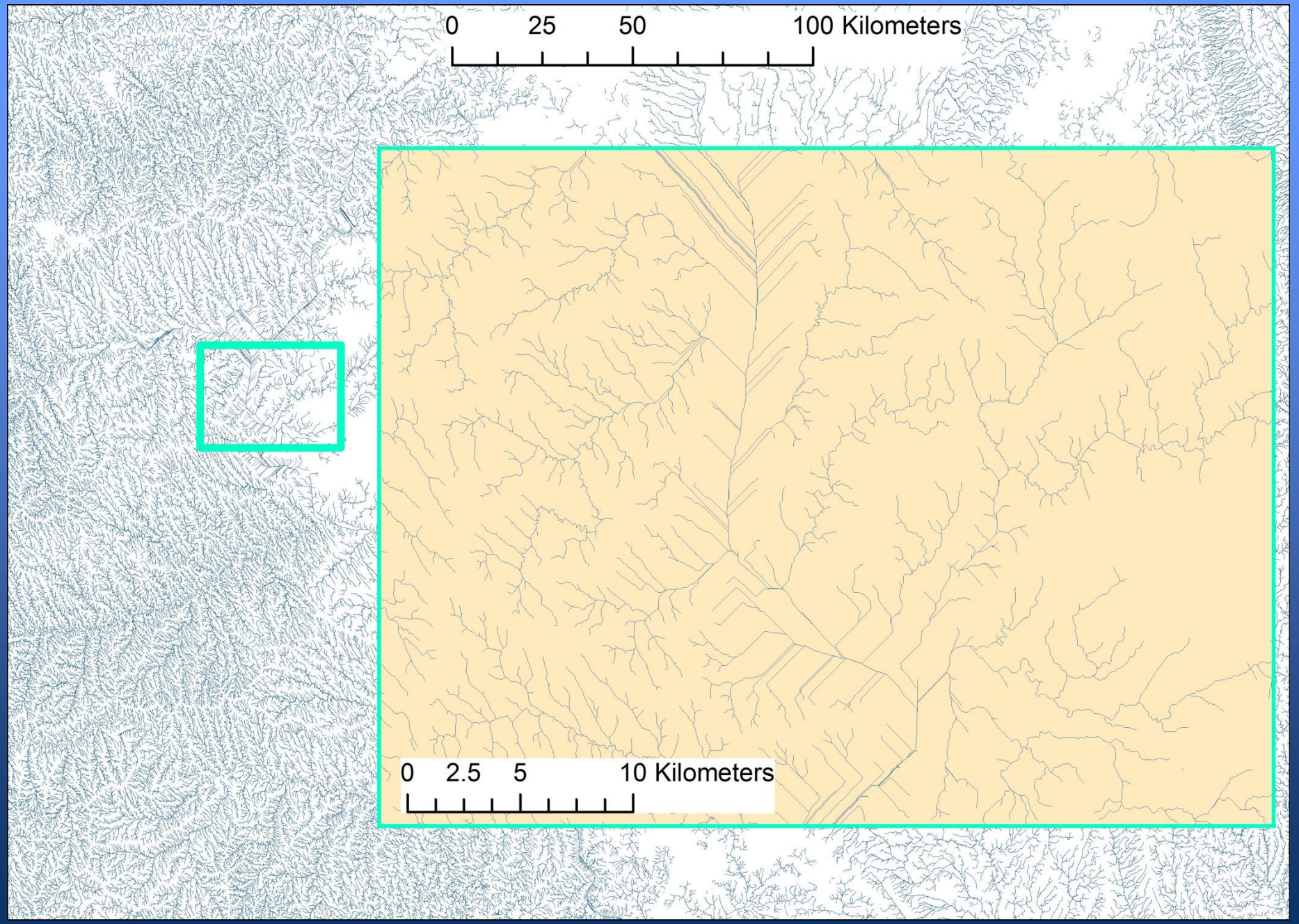


USGS

Results: 24K NHD Flowline Features

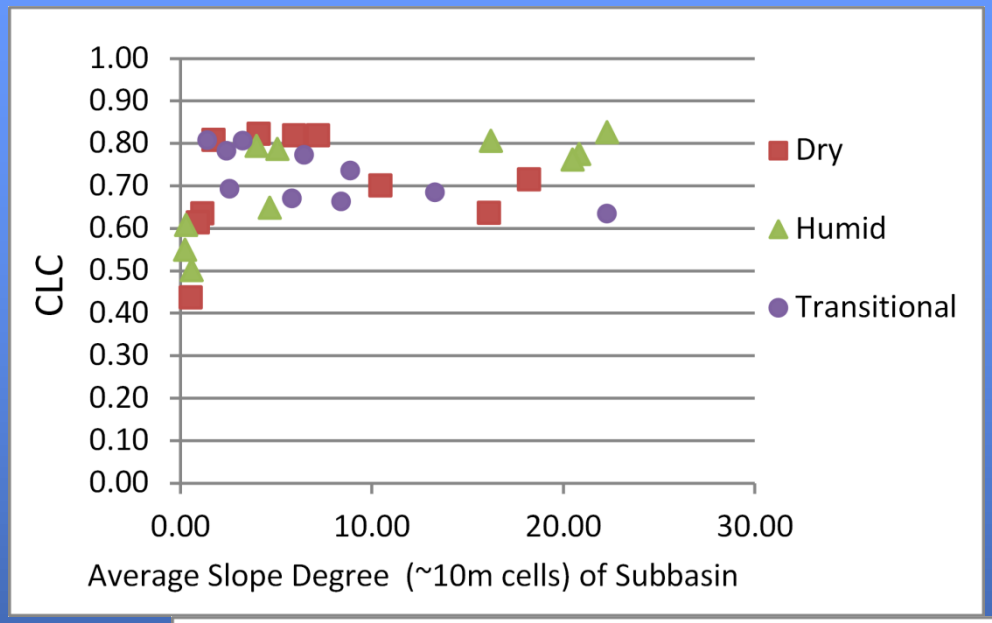


Results: 24K Extracted Channels (TauDEM)

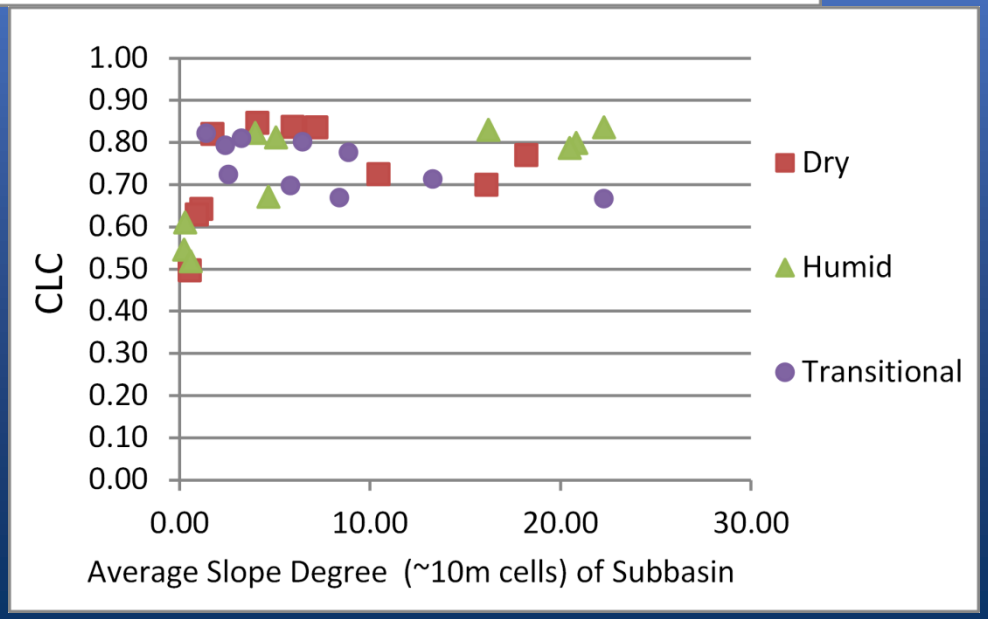


Coefficient of Line Correspondence (CLC): 24K NHD

Flowlines compared to extracted channels, 30 HUC8 subbasin study



Commercial Software extraction
Average: 0.71
Minimum: 0.44
Maximum: 0.83

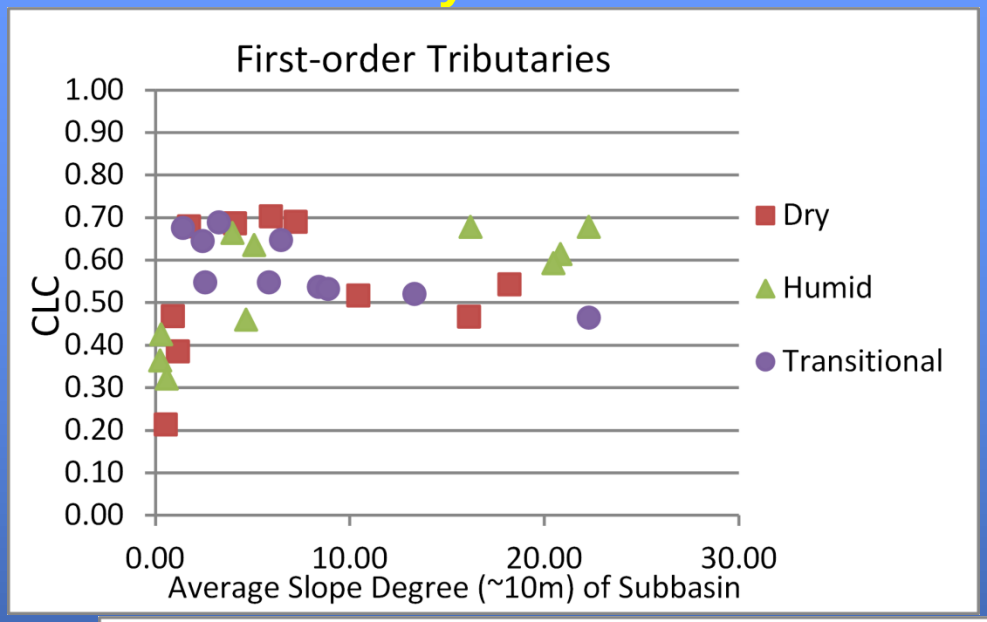


Extraction with open source and TauDEM
Average: 0.73
Minimum: 0.50
Maximum: 0.85

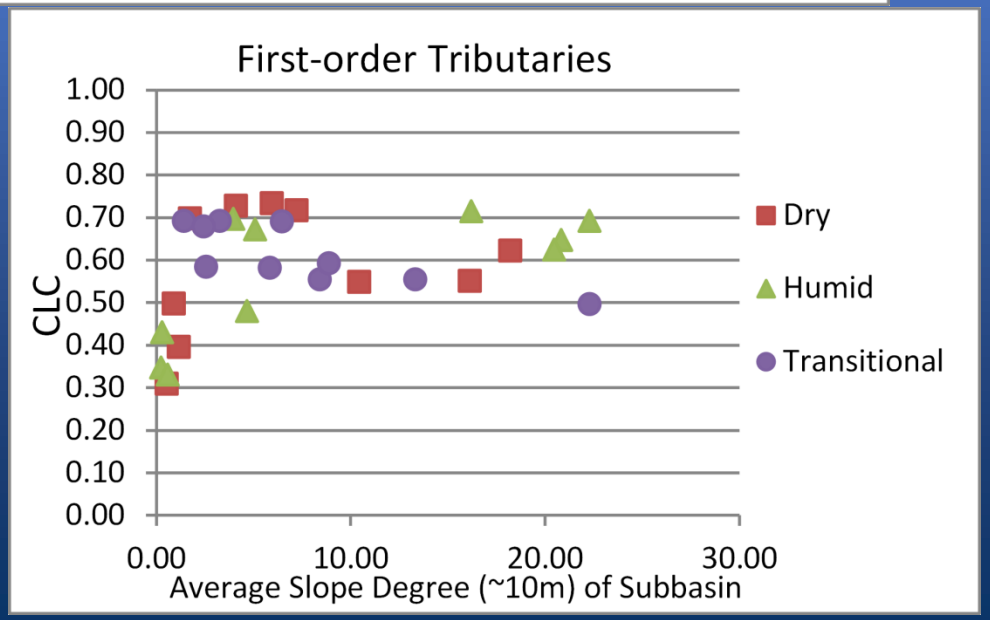
USGS

Coefficient of Line Correspondence (CLC):

Comparison of first-order 24K NHD Flowline to extracted channels 30 HUC8 subbasin study



Commercial Software extraction
Average: 0.55
Minimum: 0.21
Maximum: 0.70



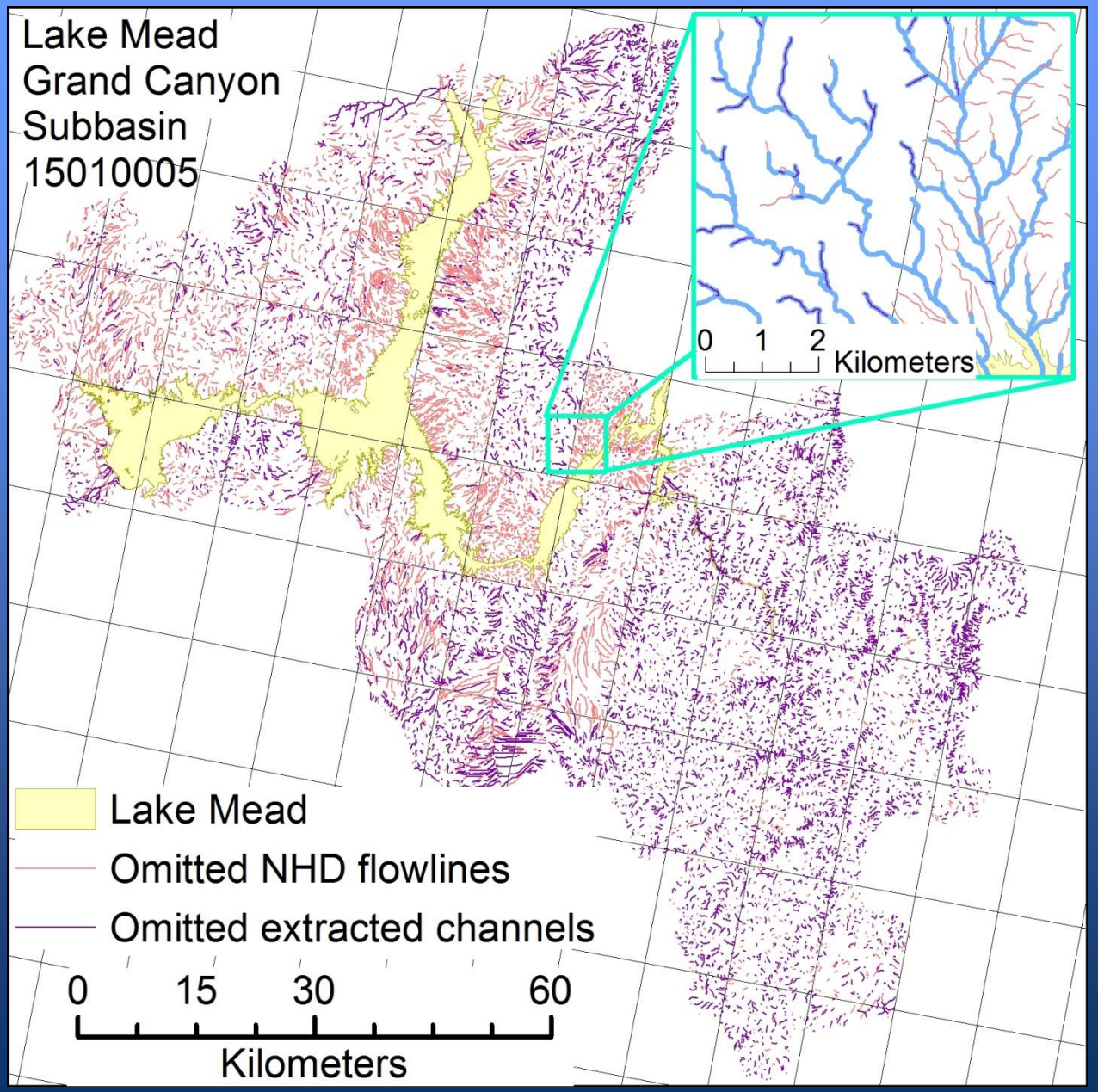
Extraction with open source and TauDEM
Average: 0.59
Minimum: 0.31
Maximum: 0.73

USGS

Omitted features

CLC process identifies line features in one dataset that are omitted from the other dataset.

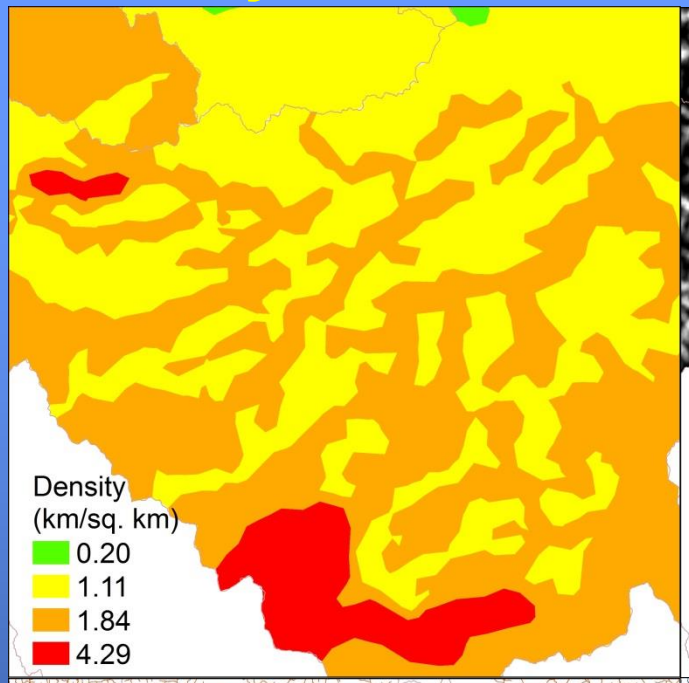
Extracted channels that are omitted from the NHD can help guide updates to the NHD where collection is too sparse.

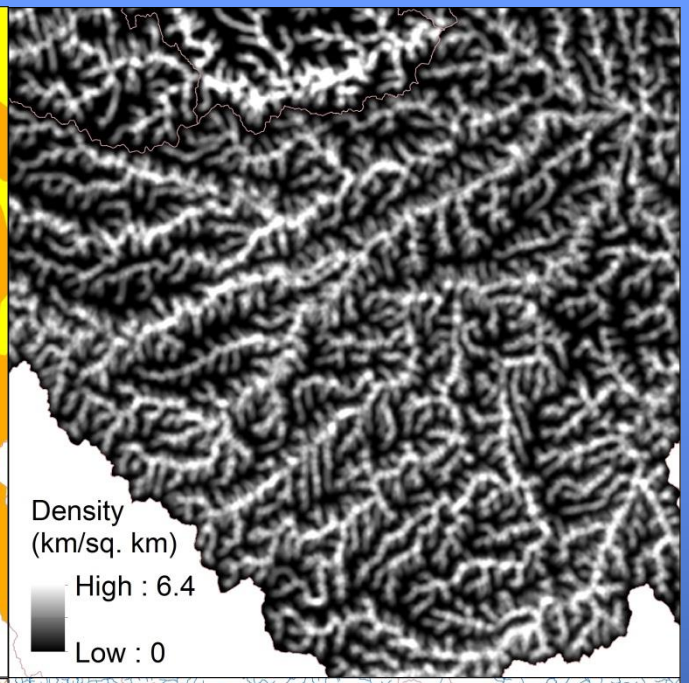


USGS

Using 24K Density Pattern to Prune HR NHD Flowlines to 24K

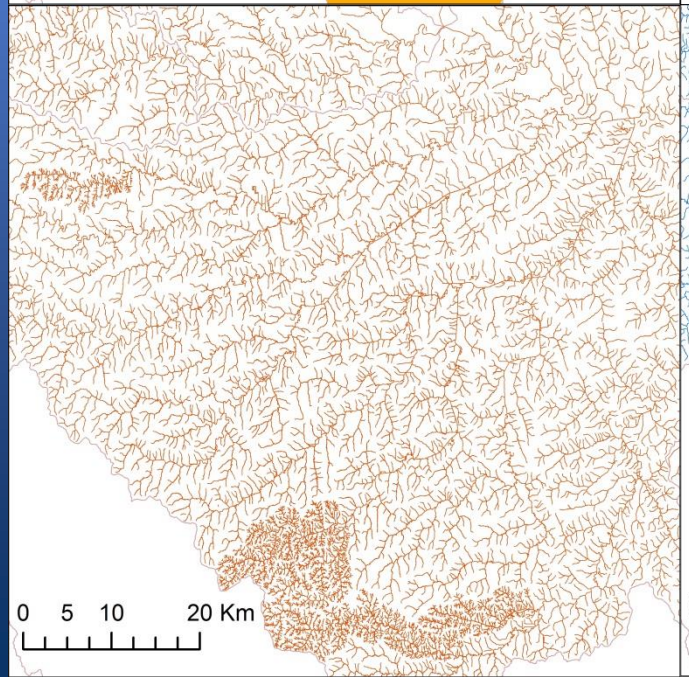
HR NHD Flowline Density Partitions (step 5)

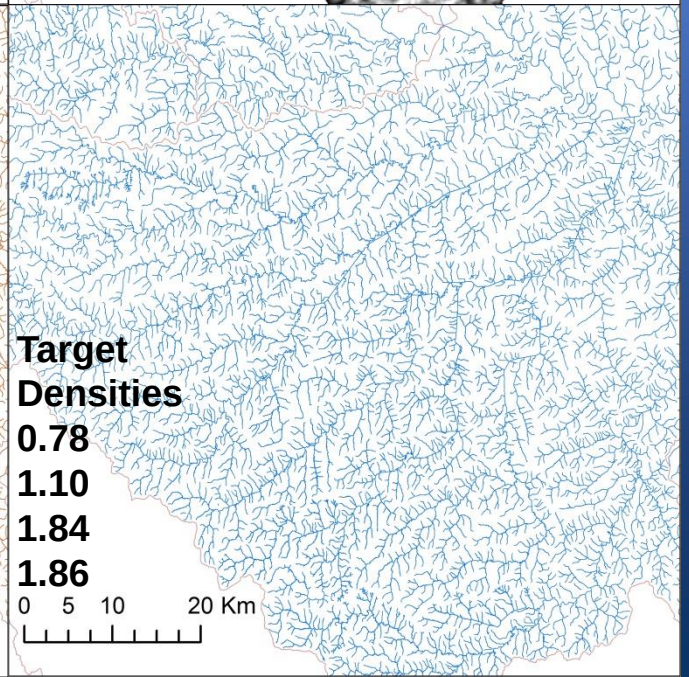




Derived 24k Density Pattern (steps 1-4 and 6)

HR NHD Flowlines



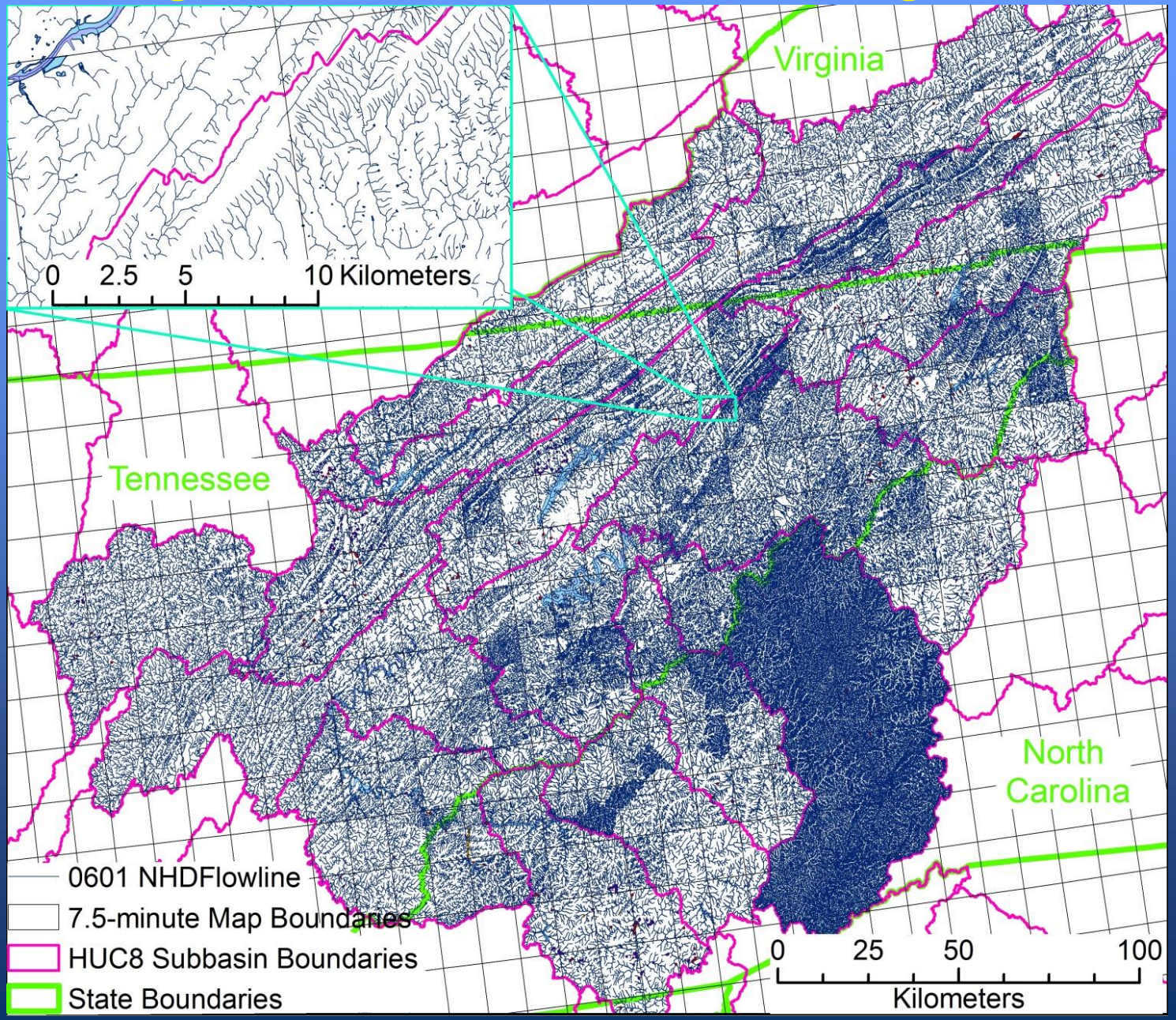


HR Flowlines Thinned to 24k Target Densities (step 7 and thinning)

USGS

Stratified Pruning Results: HR NHD, Subregion 0601

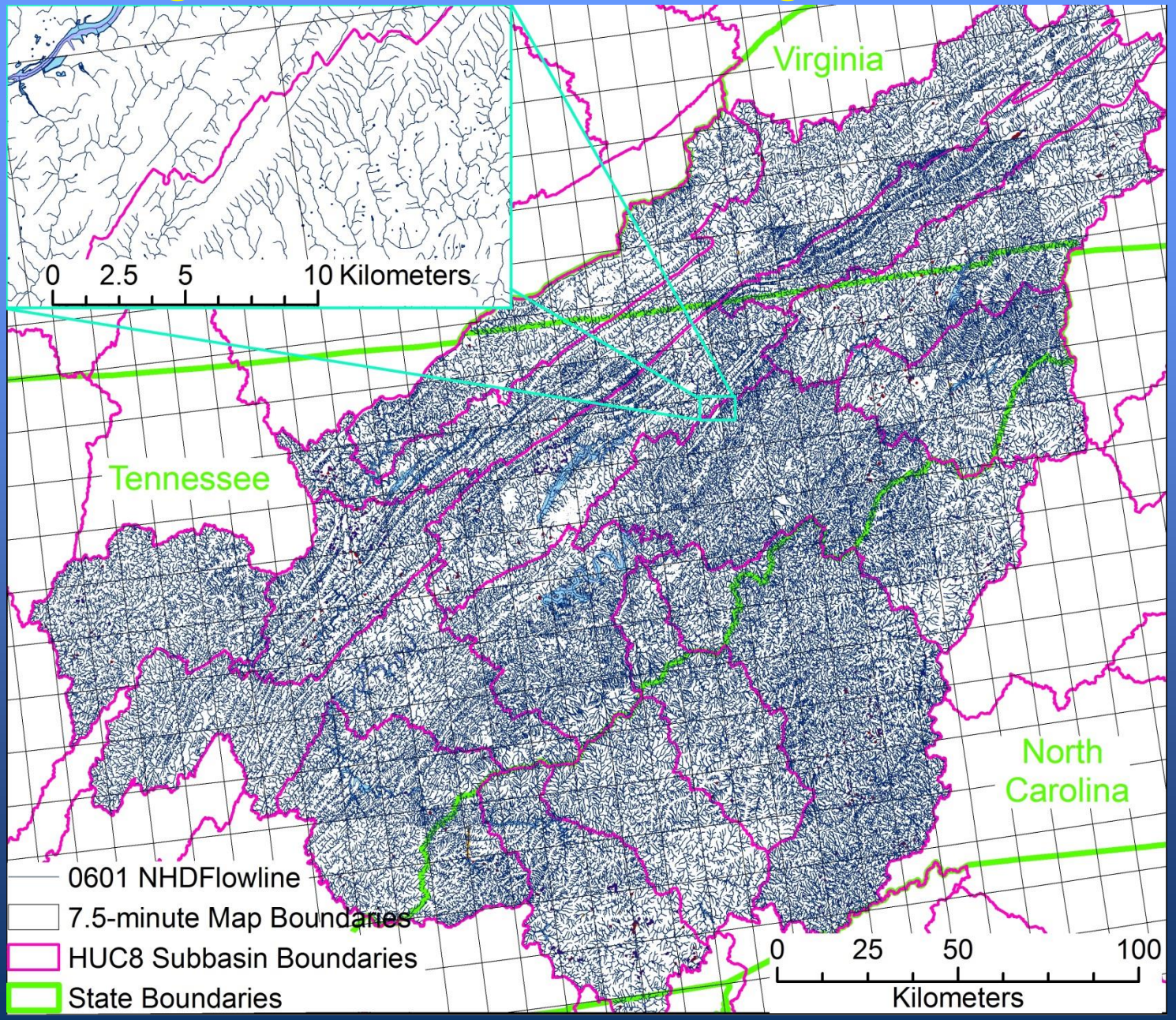
16 subbasins in subregion 0601 (Region 6, subregion 1)



USGS

Stratified Pruning Results: 24K, Subregion 0601

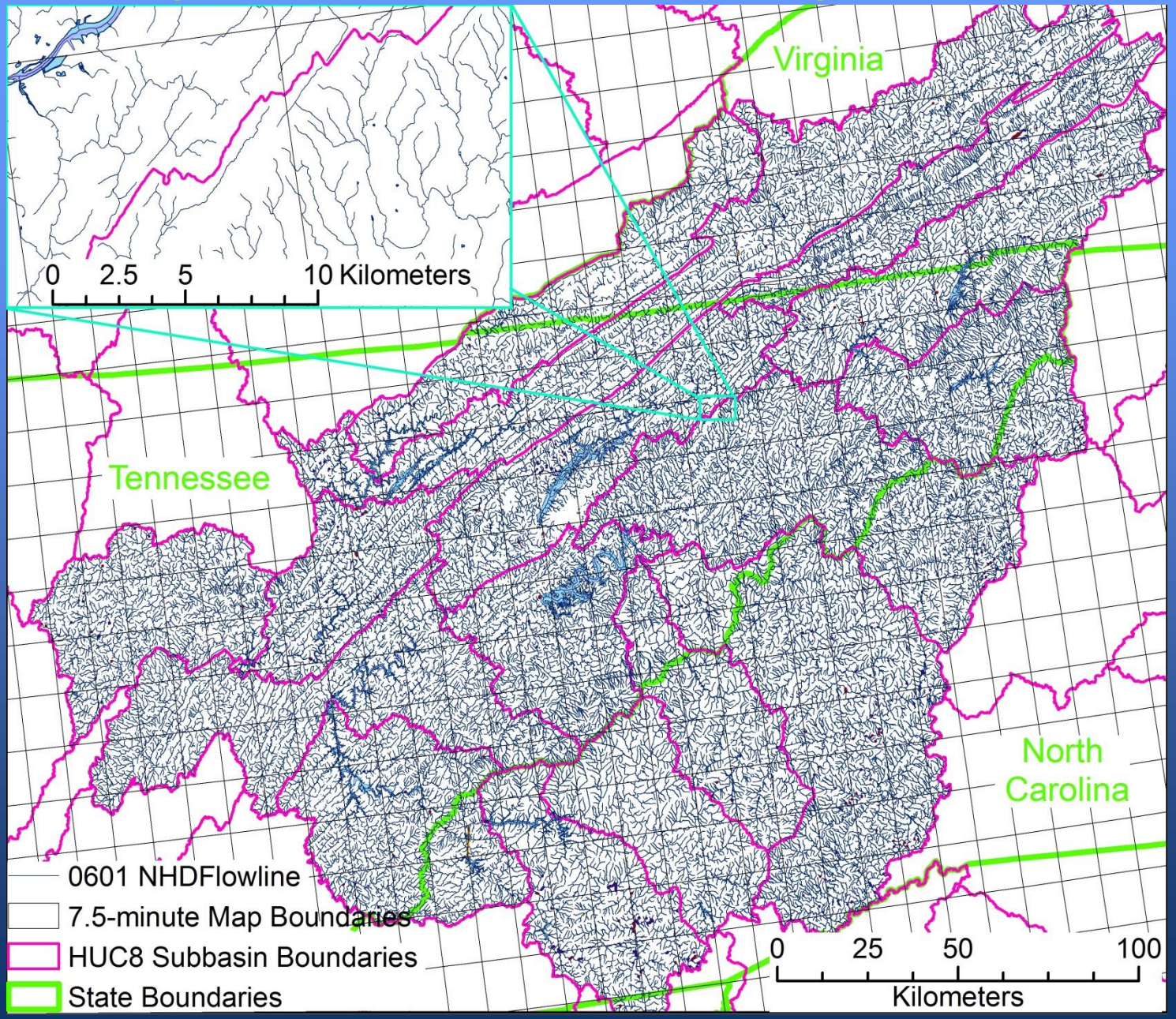
1:24,000-scale
16 subbasins in subregion 0601 (Region 6, subregion 1)



USGS

Stratified Pruning Results: 100K, Subregion 0601

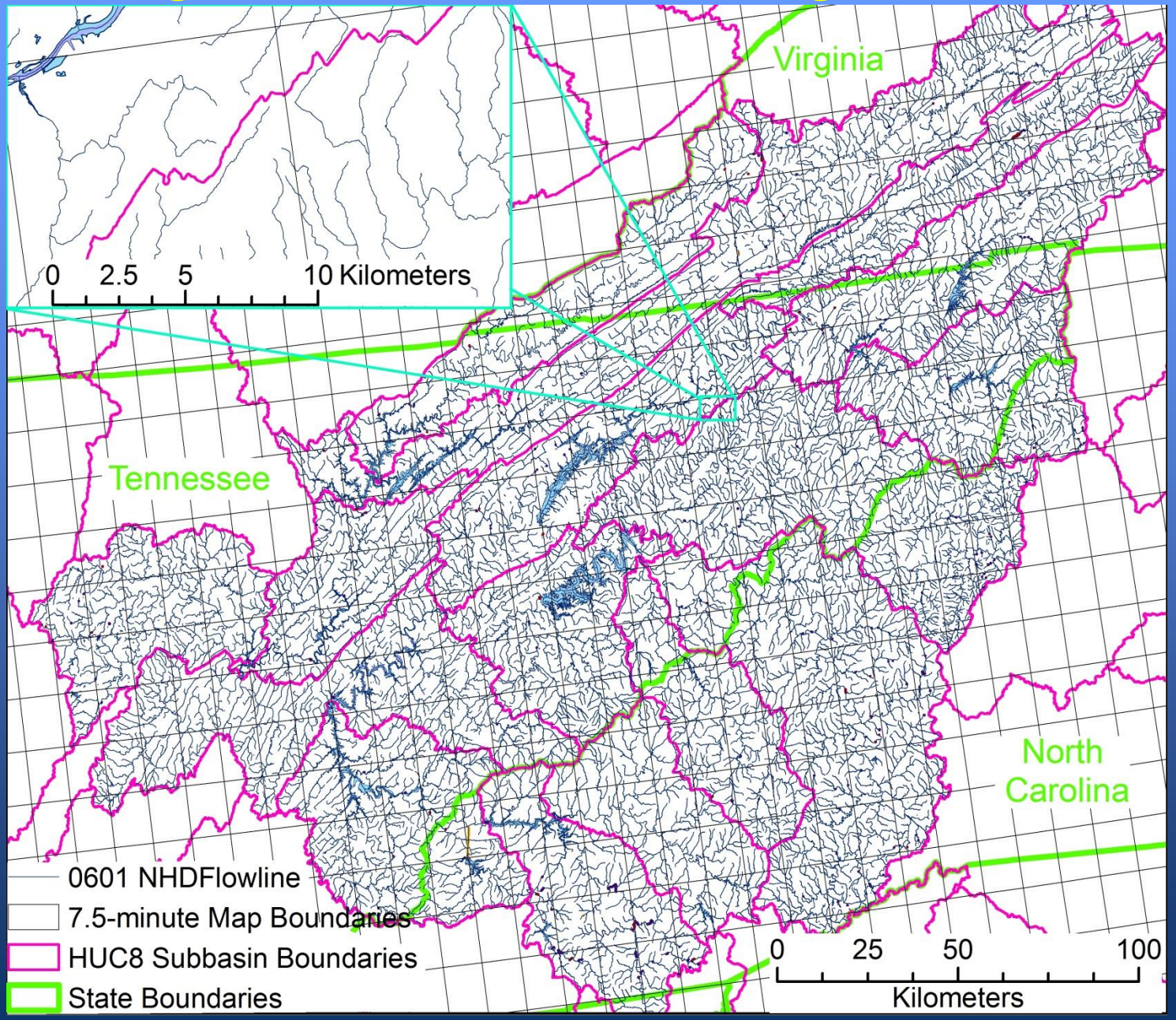
1:100,000-scale
16 subbasins in subregion 0601 (Region 6, subregion 1)



USGS

Stratified Pruning Results: 250K, Subregion 0601

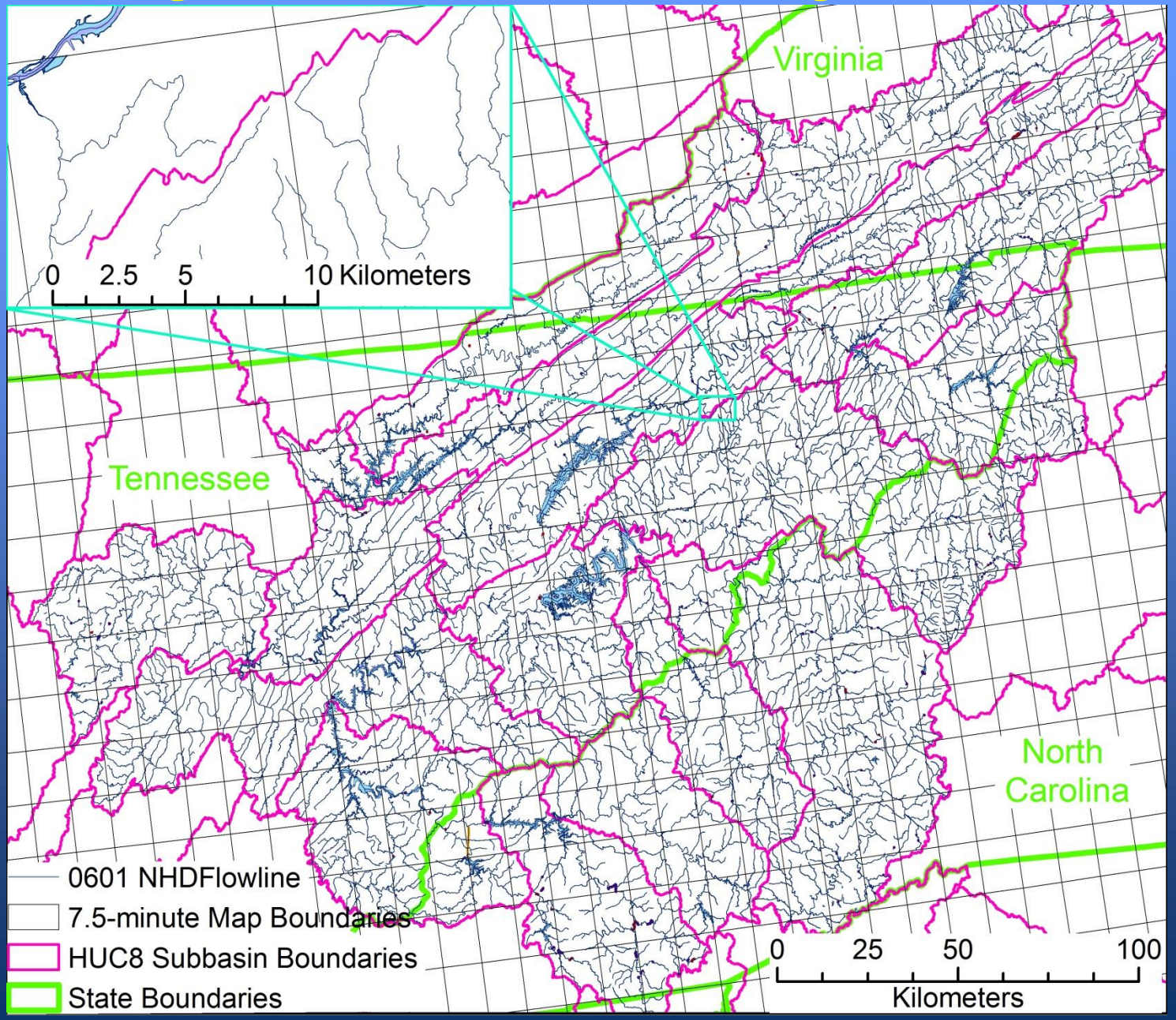
1:250,000-scale
16 subbasins in subregion 0601 (Region 6, subregion 1)



USGS

Stratified Pruning Results: 500K, Subregion 0601

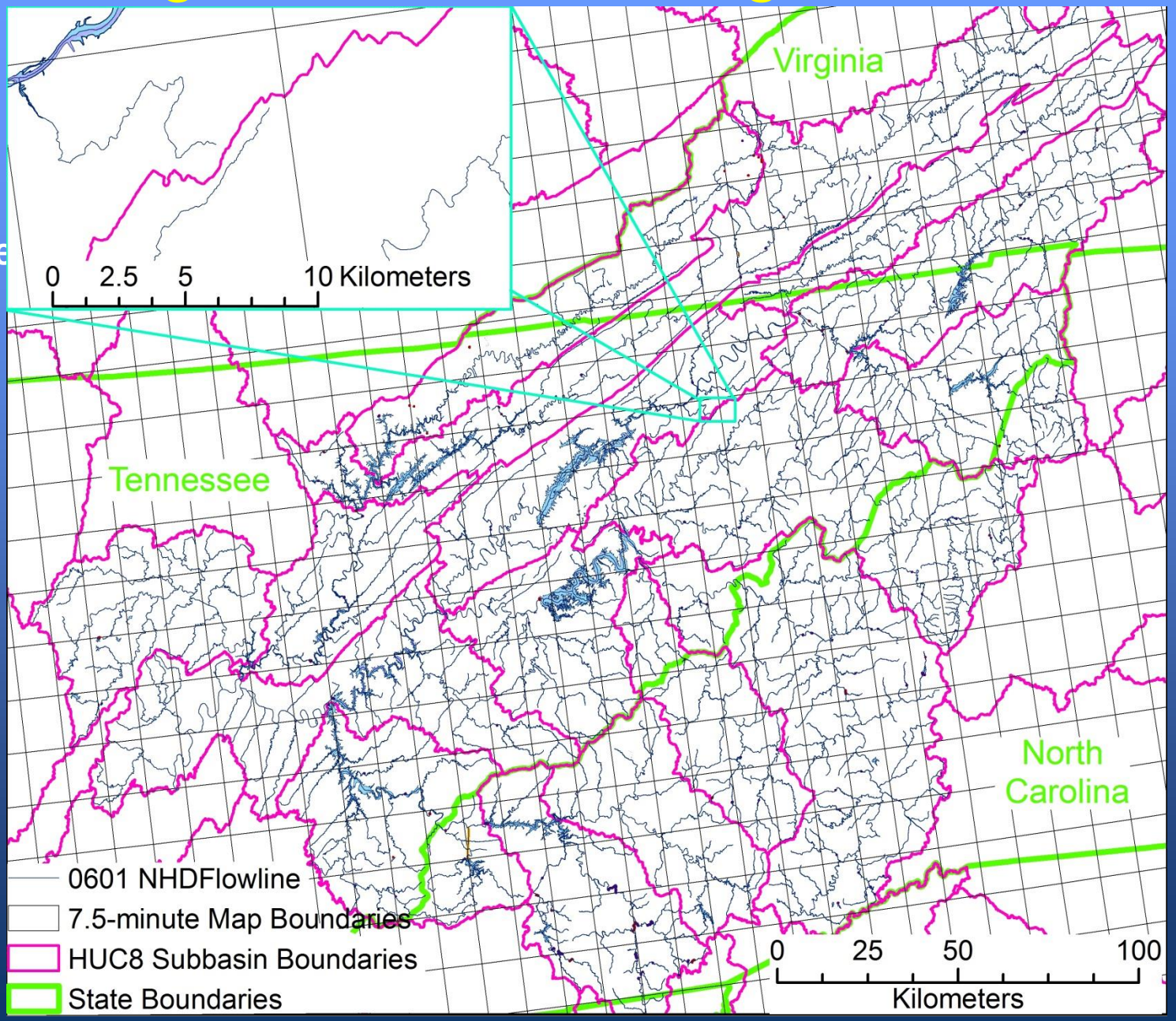
1:500,000-scale
16 subbasins in subregion 0601 (Region 6, subregion 1)



USGS

Stratified Pruning Results: 1M, Subregion 0601

1:1,000,000-scale
16 subbasins in subregion 0601 (Region 6, subregion 1)



USGS

Summary of Methods

- **Use of open source tools and TauDEM in the Linux parallel processing environment provides a big performance boost over the commercial software methods on a Windows machine, along with slightly better channel extraction results. (But there is a learning curve to use open source tools.)**
- **Model Results**
  - **Thus far extracted 24K natural drainage density patterns are effective for estimating target densities to prune the HR NHD to multiple scales.**
  - **Methods primarily extracts natural drainage. Does not perform as well where drainage is controlled by man-made features.**
  - **Initial approximation for requested length controls between subbasin density variations, and weights control within subbasin variations.**
  - **Braided channels are not extracted, which can affect density variations.**
  - **Parallel channel extraction in flat areas and waterbodies can affect connectivity of the network and may cause over extraction in flat areas.**
  - **CLC comparisons indicate 44 to 83 percent of extracted channels match the 24K NHD Flowlines**
    - **Lower matching occurs for subbasins with average slope less than 1.5 degrees.**
    - **Most mismatching occurs in first-order headwater features.**
    - **Extracted features omitted from HR NHD can guide updates where NHD is too sparse.**

USGS

Future Work

- Testing of TauDEM tools with improved parallel processing capabilities that can handle HUC6 subbasins (about 6 or 7 HUC8).
- Further evaluation of WFA model to adjust weights and test methods to limit over extraction of channels.
- Conversion of other steps in workflow to open source methods.
- Test methods on lidar-derived DEMs (~2 or 3 meter resolution).
- Develop automated methods to capture hydrographic features in low relief, coastal and swampy areas.

USGS

Thank you!

USGS