Through the Looking Glass: Designing a Multi-Scale National Hydrographic Database

Barbara P Buttenfield, University of Colorado
Lawrence V. Stanislawski, CEGIS-USGS

babs@colorado.edu
lstan@usgs.gov

Supported by USGS-CEGIS grant # 04121HS029,
"Generalization and Data Modeling for New Generation Topographic Mapping"

GIS Day University of Kansas
20 November 2013

Overview

- National Hydrography Dataset (NHD) represents surface water of coterminous United States
- Compilation scales
  - High resolution 1:24,000 (24K)
  - Medium resolution 1:100,000 (100K)
  - Local resolution by densification of 24K to 1:5,000 (5K) or larger scale
- Users work with data at many mapping scales (topographic base maps) and analytical resolutions (hydrologic models, climate change) - commonly from 5K to 1M scale
- USGS cannot produce data for every scale and resolution; so build software tools to generalize data appropriately to desired target scales.

USGS/CEGIS Generalization Initiative

Generalization tasks:

1) **Categorize landscape types** to tailor generalization strategy
2) Partition data to **maintain local density variations**, geographically valid shapes, textures
3) **Conditional feature pruning** that maintains data model integrity
4) Simplification for data production at multiple levels of detail
5) **Establish feature prominence** that may vary with scale
6) **Metric validation (assessment)** of results with an existing data benchmark, or with an expected outcome.

Generalization constraints = developed methods must maintain:

1) Connectivity (topology) of hydrographic network
2) Local density variations that typify physiographic or climate variations but remove density variations due to inconsistent compilation
3) Full attribution and all database relations

Enrich

Add attributes describing geometry / hydrography to support later tasks

Prune

Eliminate entire tributaries / polygon appendages to reduce inappropriate clutter

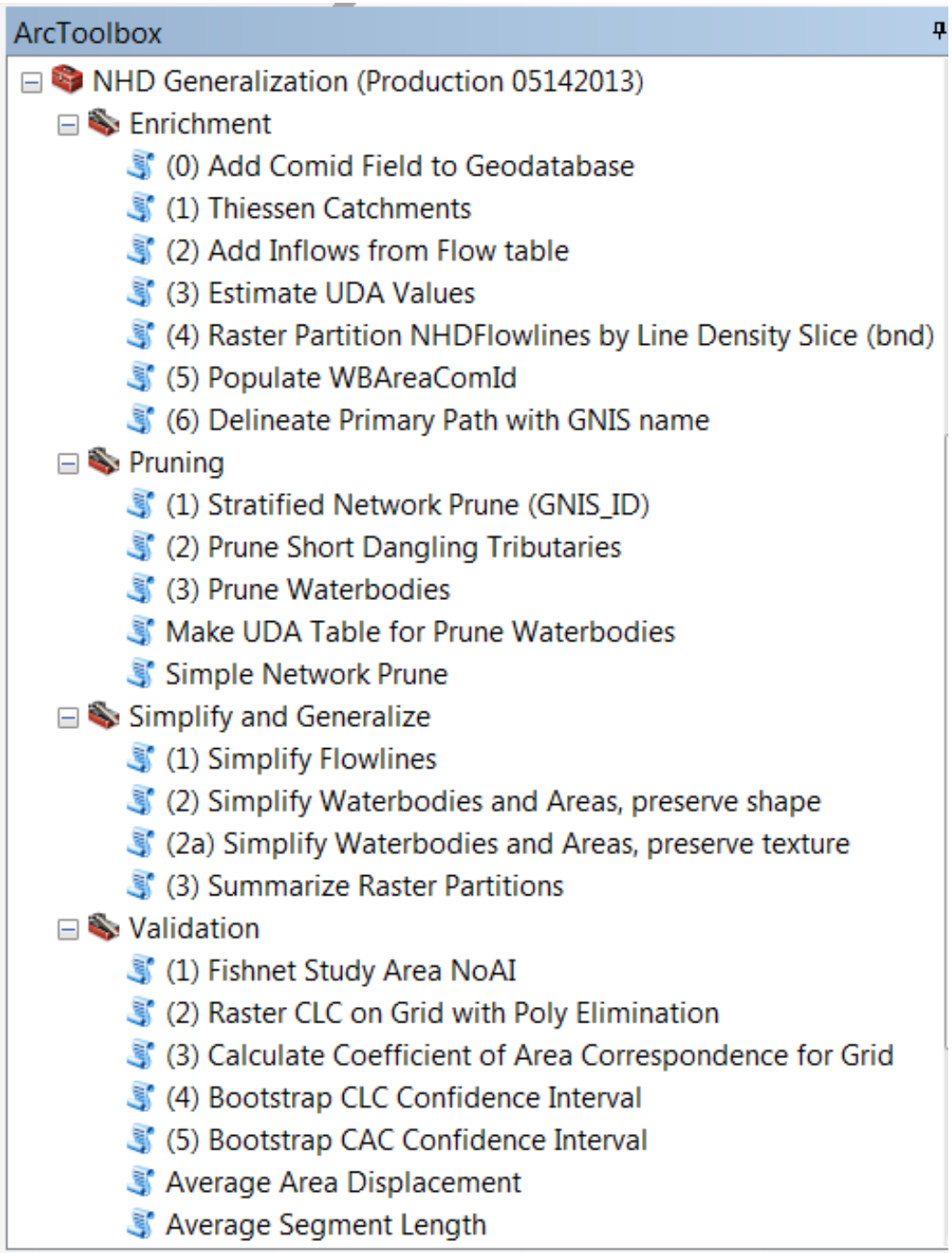
Simplify

Eliminate coordinates in lines and polygons; correct vertical integration

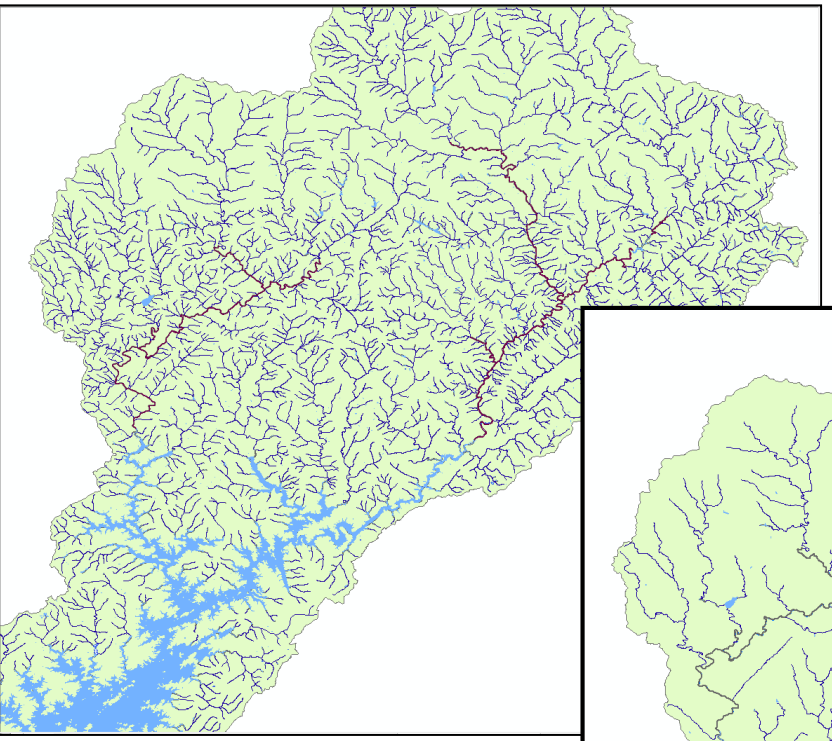
Validate

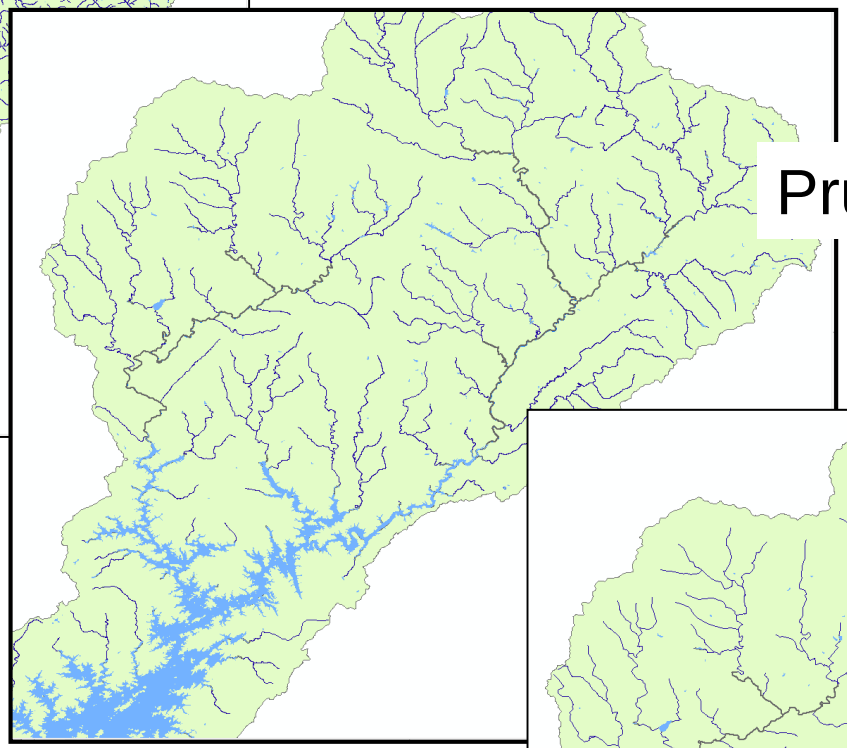
Assess quality of results

Move tools into parallel processing environment

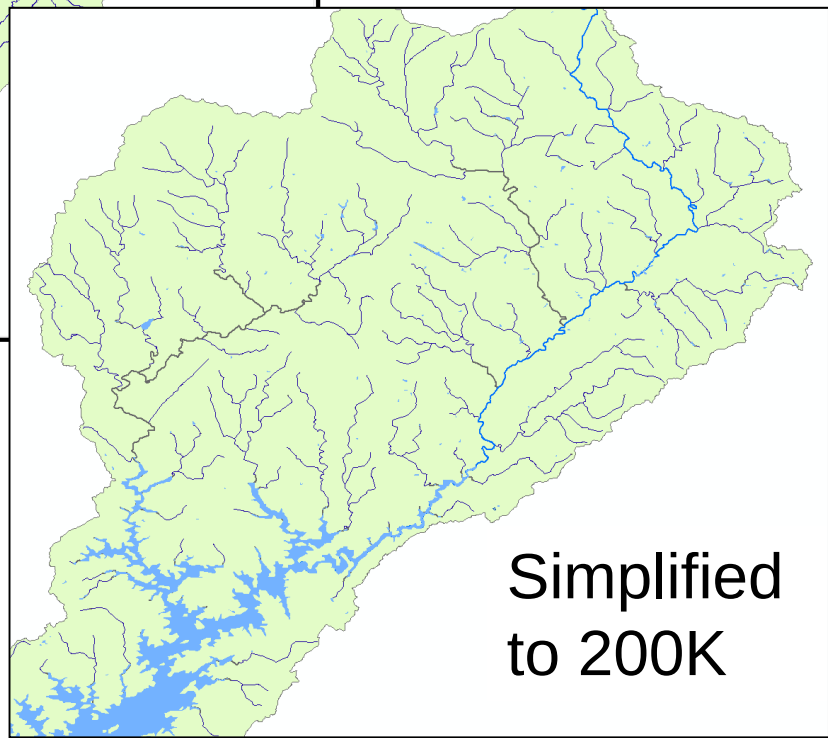


Distinguish Pruning from Simplification





Pruned to 200K



Simplified to 200K

24k Source

0 5 10 20

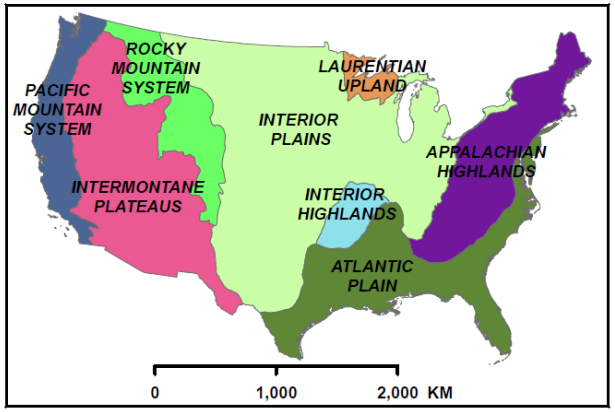
Kilometers

Atlanta (HUC 03010001)

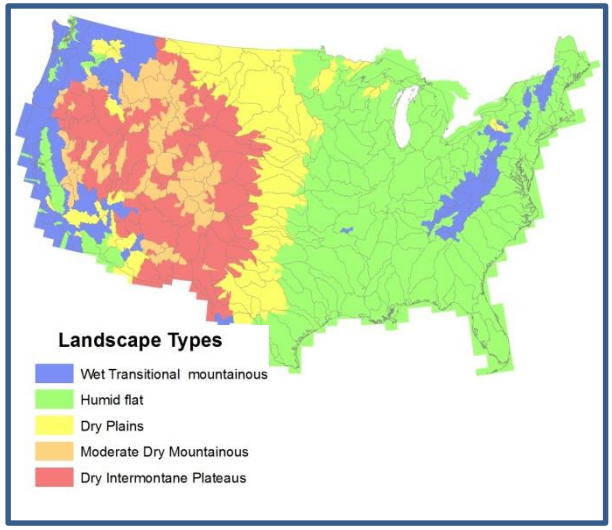
(1) Classify Landscape Types

Classify coterminous US on 7 variables for 5km grid

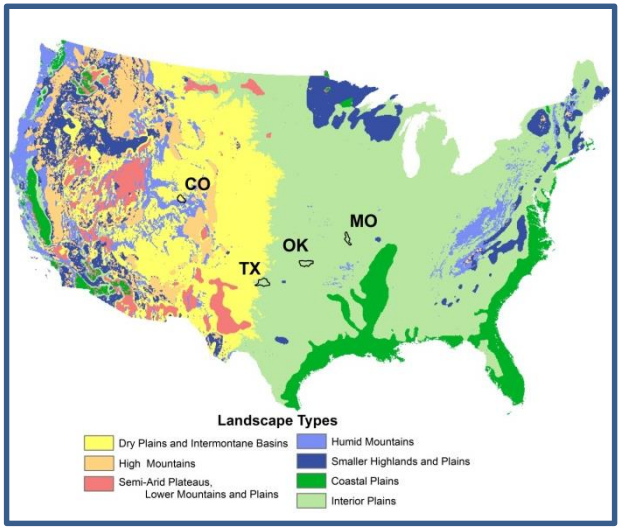
- Average Elevation
- Average Slope
- Standard Deviation of Elevation
- Drainage Density Estimates
- Runoff (mm/year)
- Bedrock Density
- Area of Inland Surface Water



Fenneman and Johnson, 1945



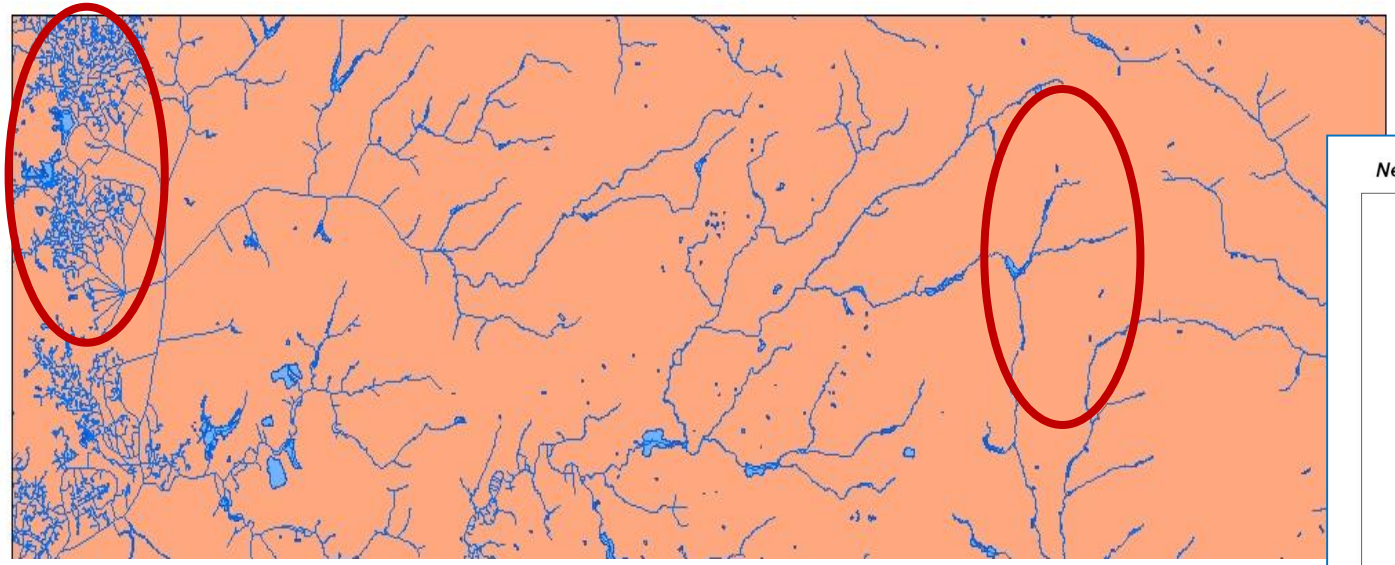
Stanislawski, 2009

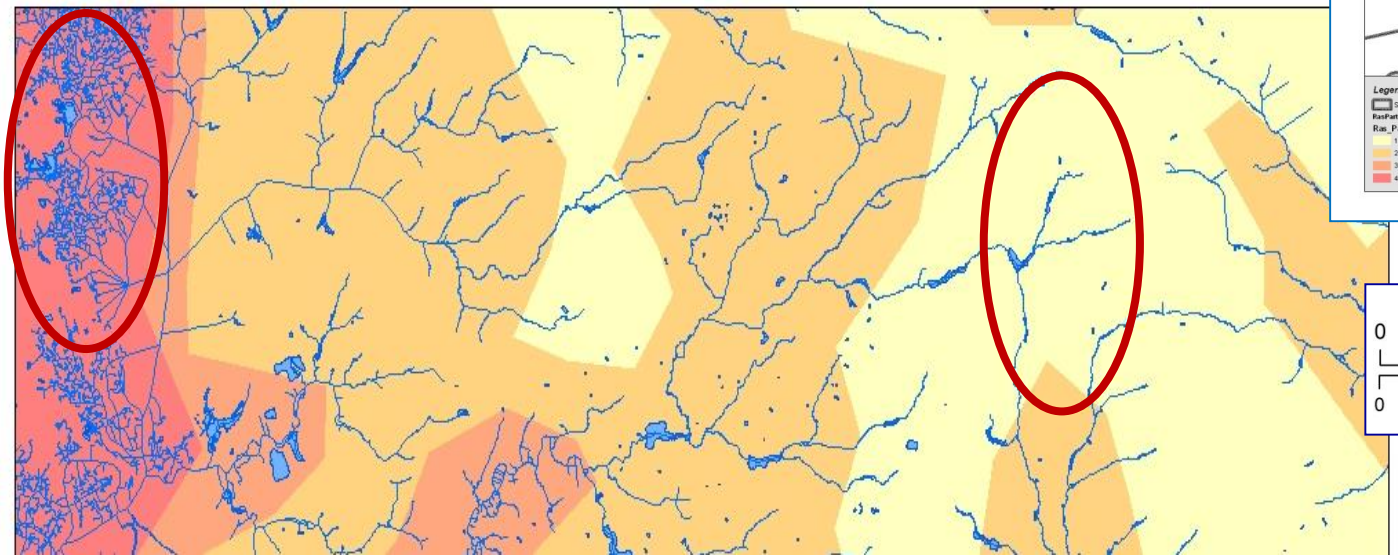


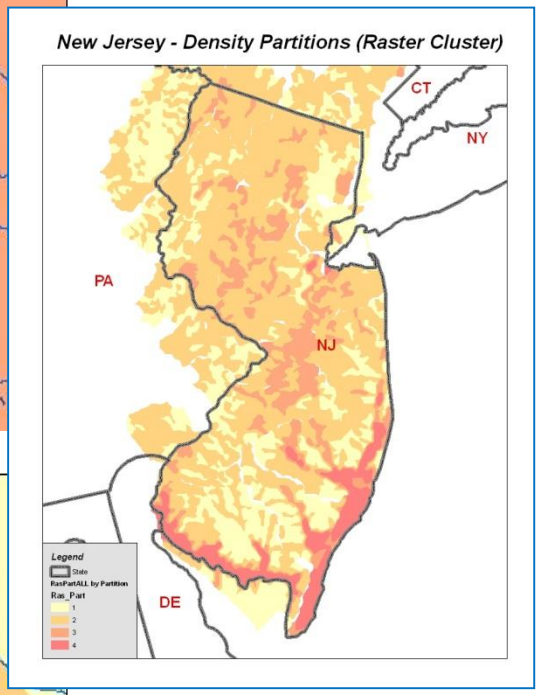
Stanislawski et al 2010

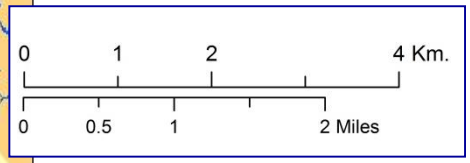
Maximum likelihood classification

(2) Why Generalize by Density Partition?









(2) Data Partitioning to Protect Local Density

Method

Generate smooth density surface for coterminous US and establish natural break density classes; apply these to local regions (HUC 8's, HUC 4's) that can be subdivided into 2-6 density partitions, with any patch within a partition spanning a minimum of 15 km^2.

Three tasks to partition locally:

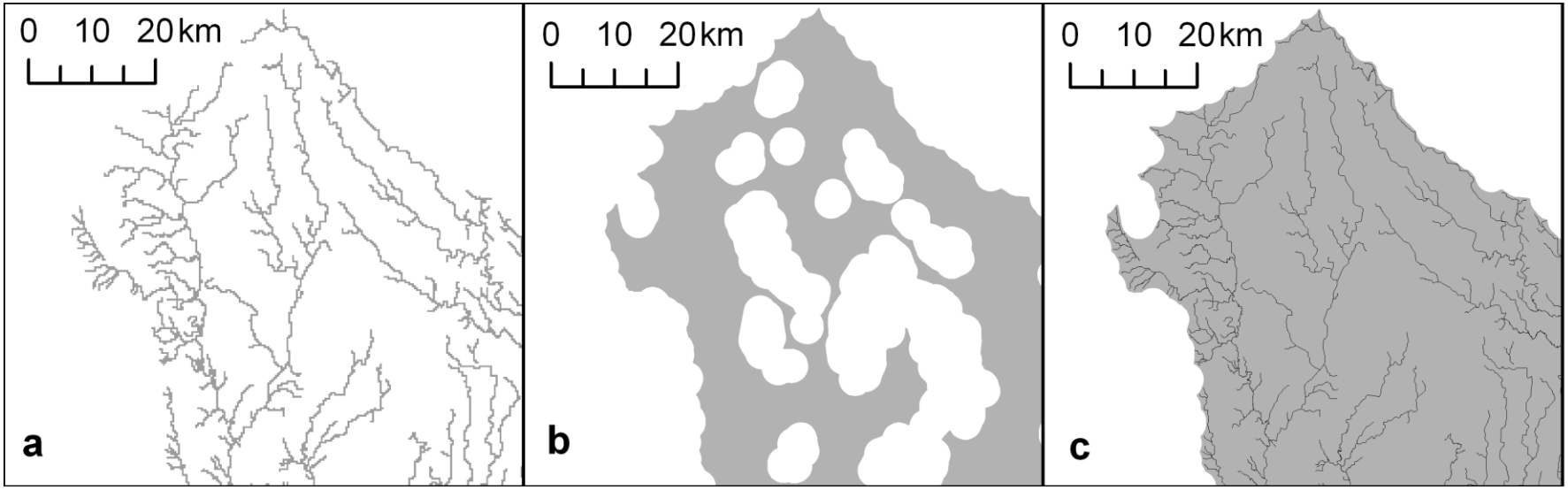
1. Delineate the area of interest (AOI)
2. Delineate line density partitions
3. Compute average line density of each partition

(2) Protect Local Density: Delineate the area of interest (AOI)

a) Rasterize stream network w/ 300 meter (m) cell size.

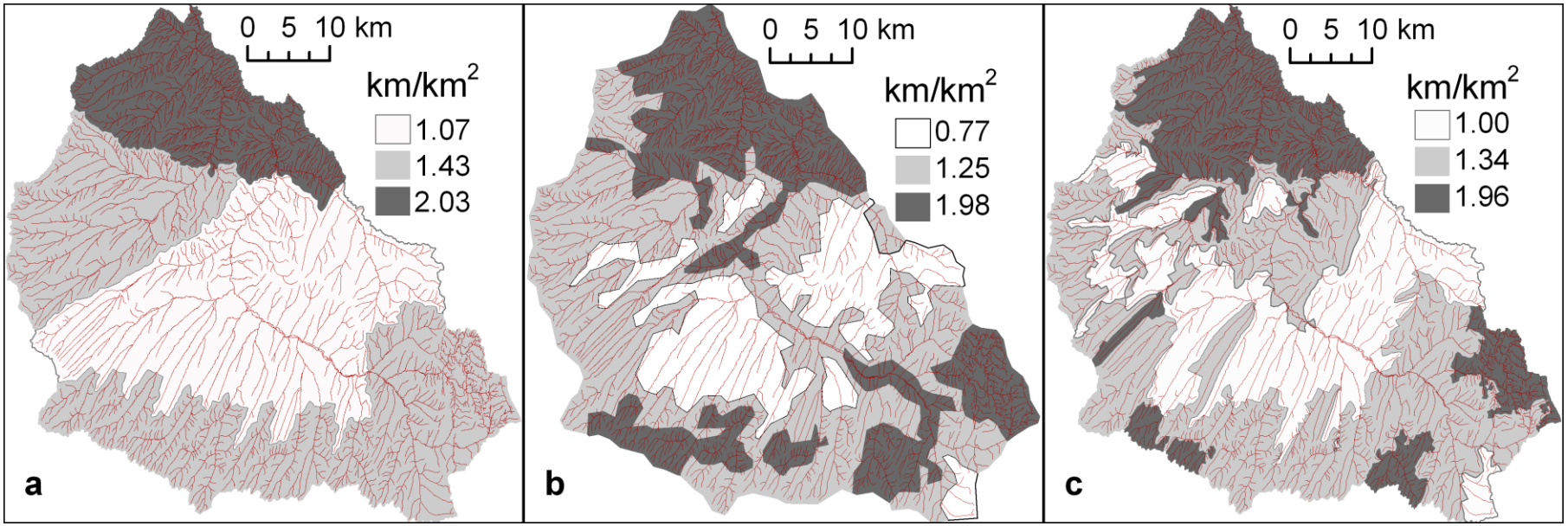
b) Convert raster lines back to polygons, buffer 2800m, negative buffer 2600m to generate a boundary polygon that extends about 200-300m beyond the extent of the source flowlines.

c) Dissolve to remove interior holes.



(2) Protect Local Density: Delineate density partitions

manual **raster-based** vector-based catchment clustering



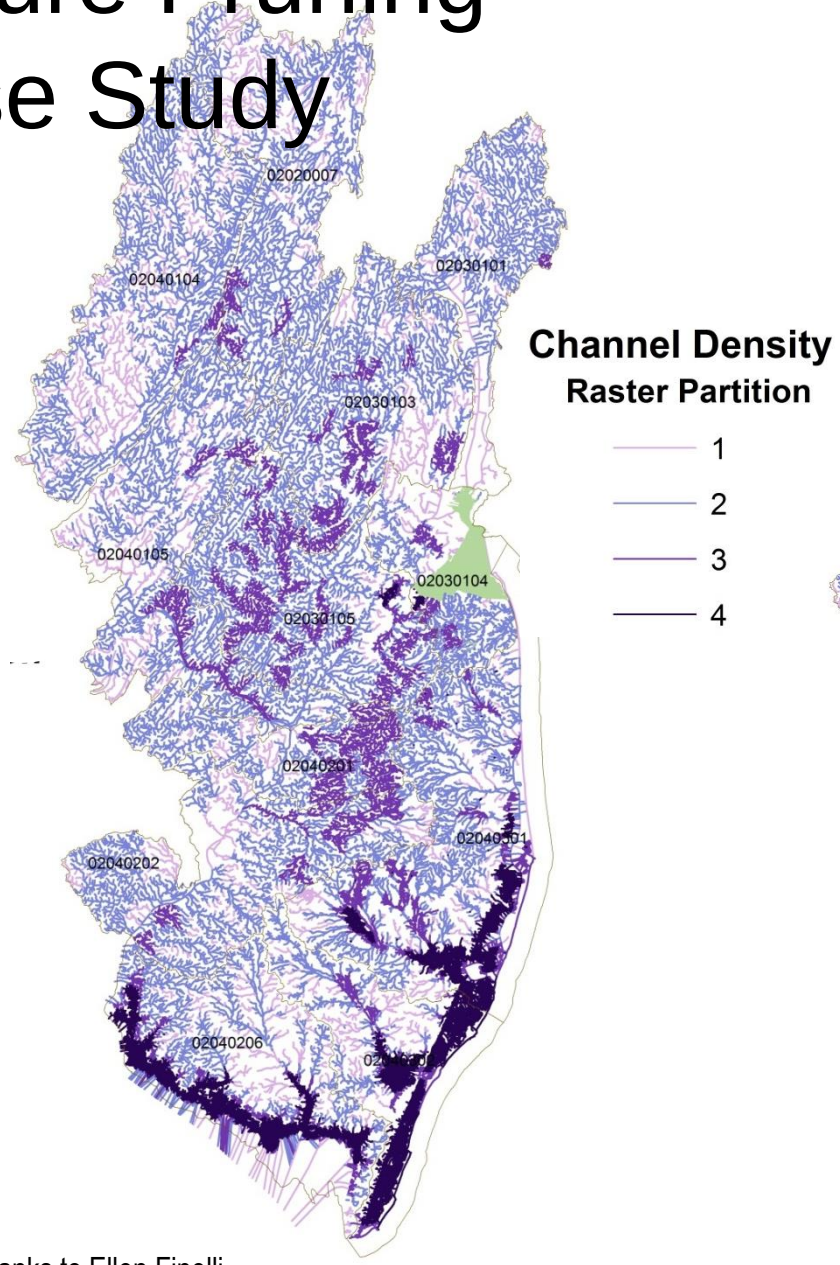
Compute actual average density for each partition

(3) Conditional Feature Pruning
New Jersey Case Study

Channel Density Partitions

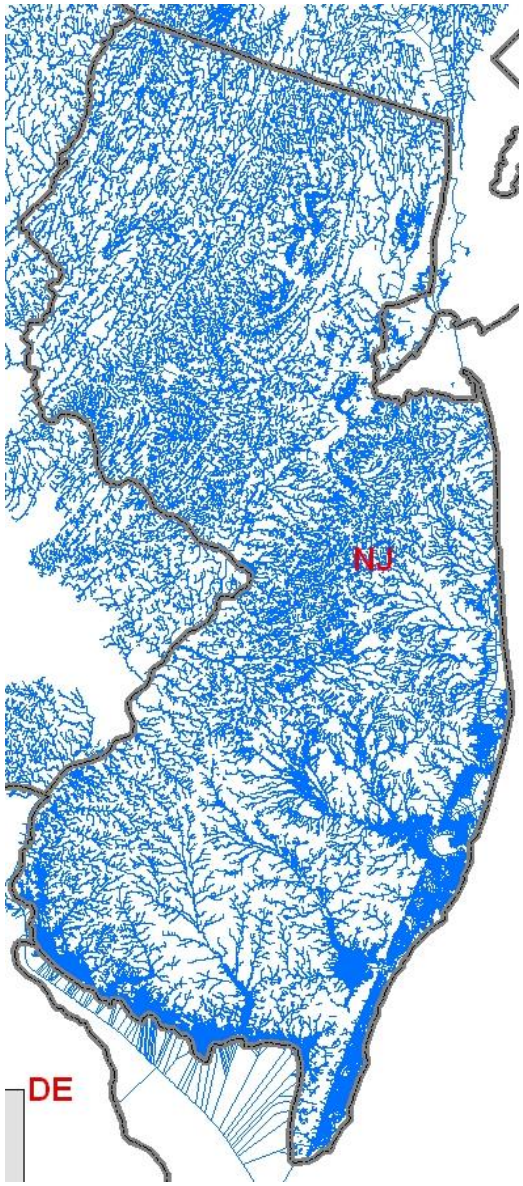
Local Resolution Catchment Partition Density				
NJ HUCs	**Part 1**	**Part 2**	**Part 3**	**Part 4**
02020007	0.78138	1.27962	2.29771	
02030101	0.53154	1.22715	2.57289	
02030103	0.69940	1.36554	2.59074	
02030104	0.34825	1.23386	2.42316	2.21904
02030105	0.60090	1.47131	2.34448	5.70696
02040104	0.77651	1.23280	2.15372	
02040105	0.71360	1.30721	2.36065	
02040201	0.77530	1.30006	2.47314	
02040202	0.67128	1.28302	2.65832	
02040206	0.626974	1.224485	2.248854	7.457076
02040301	0.45869	1.26256	1.74257	6.15804
02040302	0.606237	0.993774	1.300671	3.928845

Simplify @ 35m 30m 25m 20m

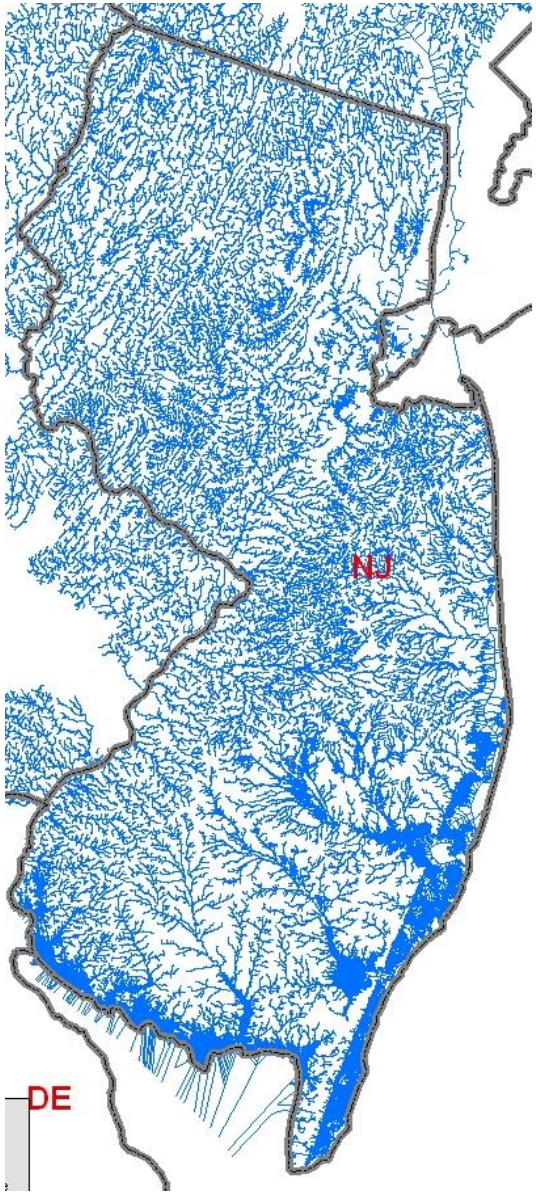


Thanks to Ellen Finelli

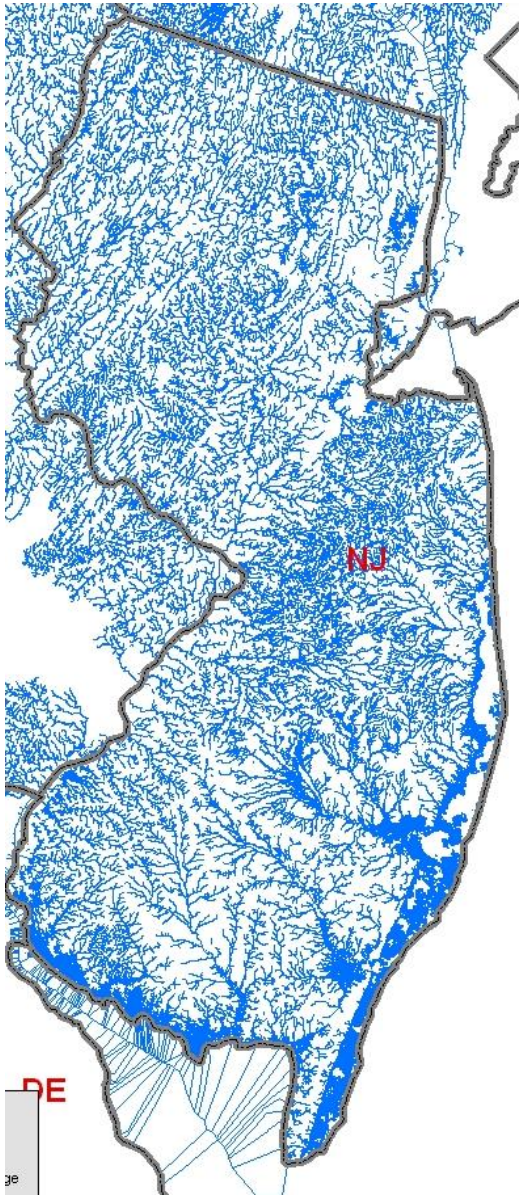
Local Res (densified to 5K)



Pruned / Simplified to 24 K



Archive 24K



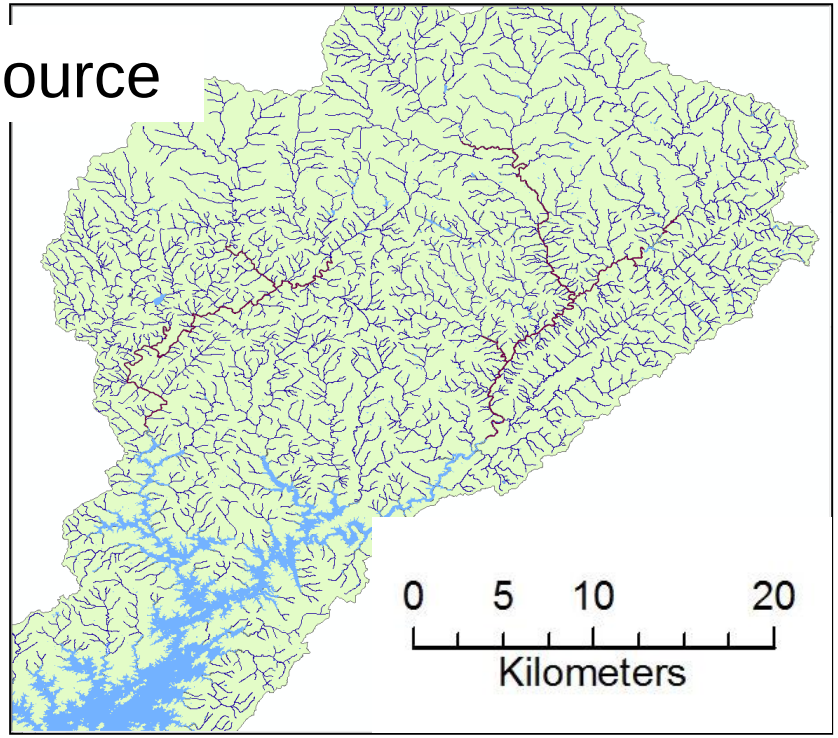
Local Res (densified to 5K)

Pruned / Simplified to 24 K

Archive 24K

(3) Another Advantage of Conditional Feature Pruning

24k Source

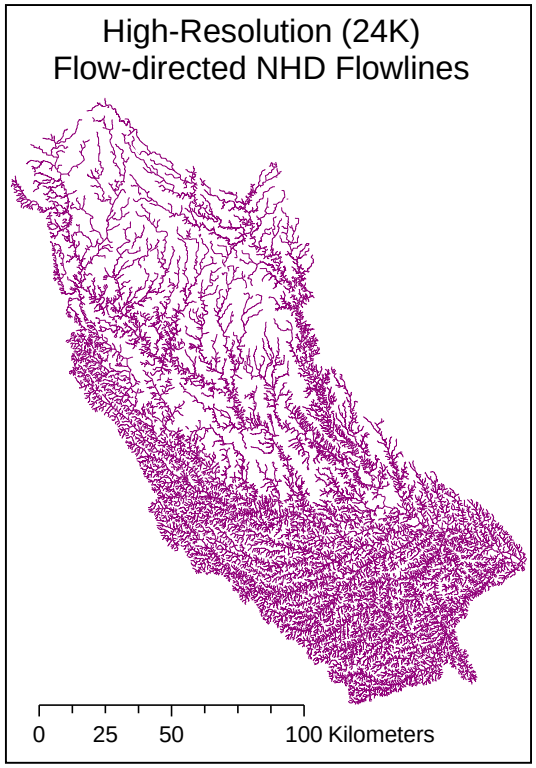


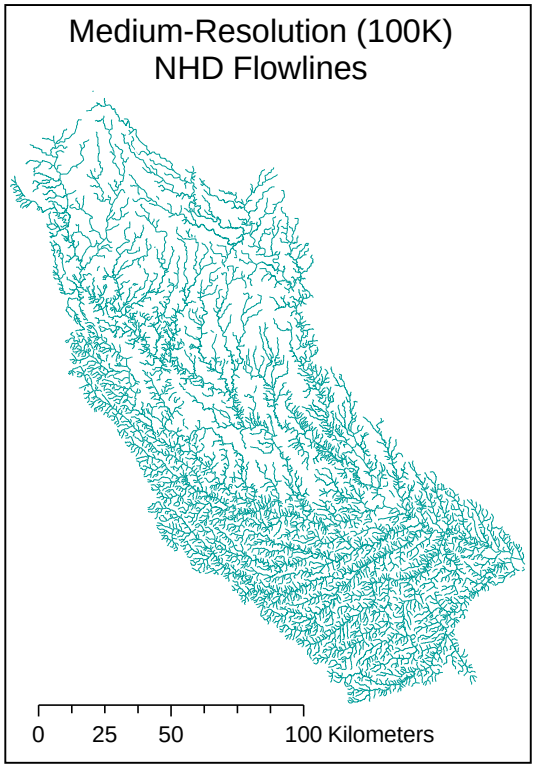
Automatically eliminate inconsistent compilation issues…

Atlanta (HUC 03010001)

(3) Another Advantage of Conditional Feature Pruning

…while protecting density differences that reflect meaningful geographic process (glaciation) which is evident at large and intermediate scales.





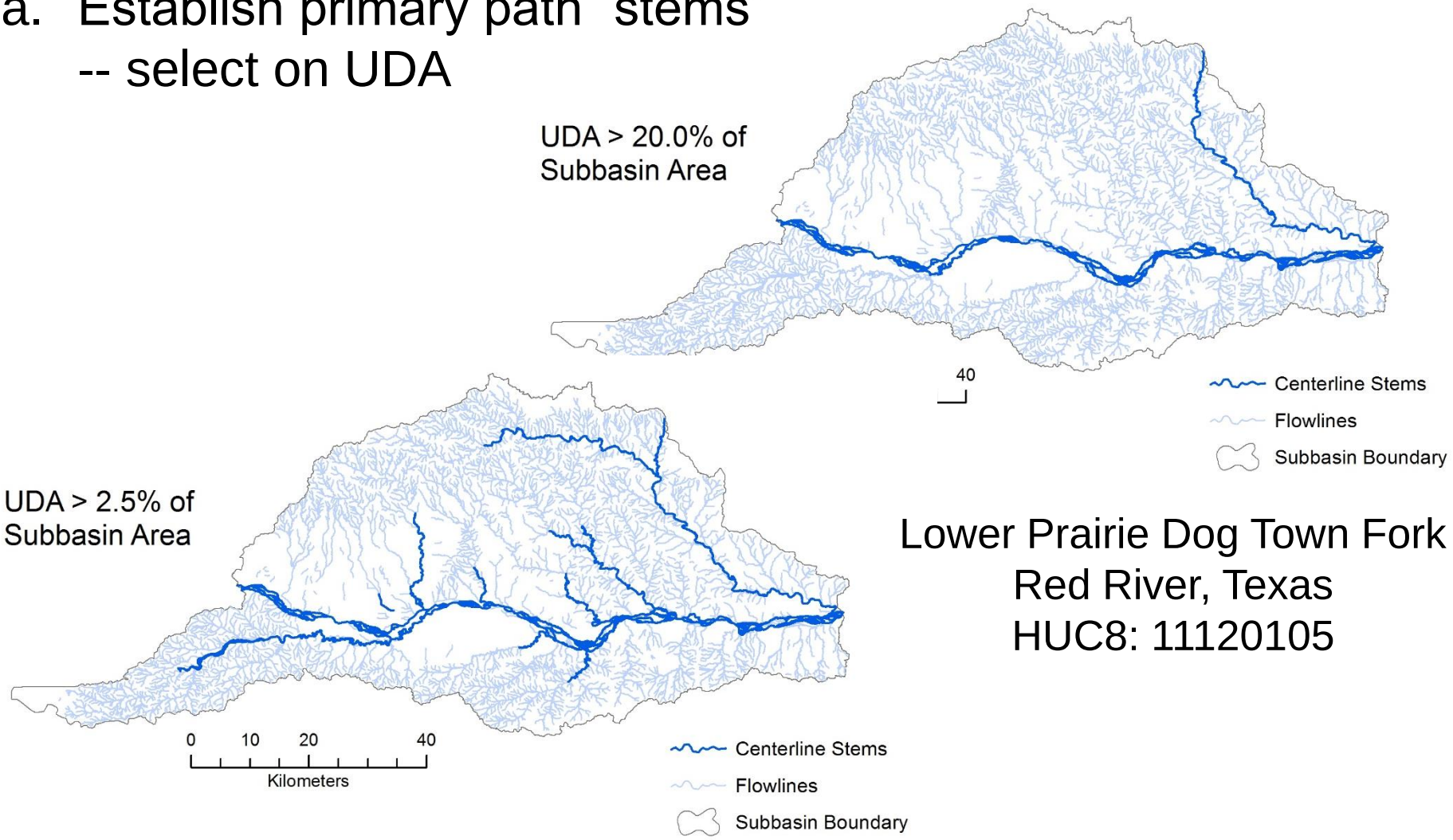
4 HUC 8s near Des Moines Iowa (0710000…4, 6, 7, 8)

(5) Establish relative feature prominence

- Primary channels, main stems, thalwegs, cartographic centerlines
  - Contain the most water – preserve at all scales b/c useful for topographic base mapping and hydrological analysis
  - Automatic methods often based on stream order
- Automated delineation difficult with 24K NHD
  - Stream order not explicitly coded, due to database size and irregular update cycle
  - Prioritize stream channels by estimating Upstream Drainage Area (UDA) for each channel in the network and running advanced node trace

(5) Basic Delineation Algorithm – Stage 1

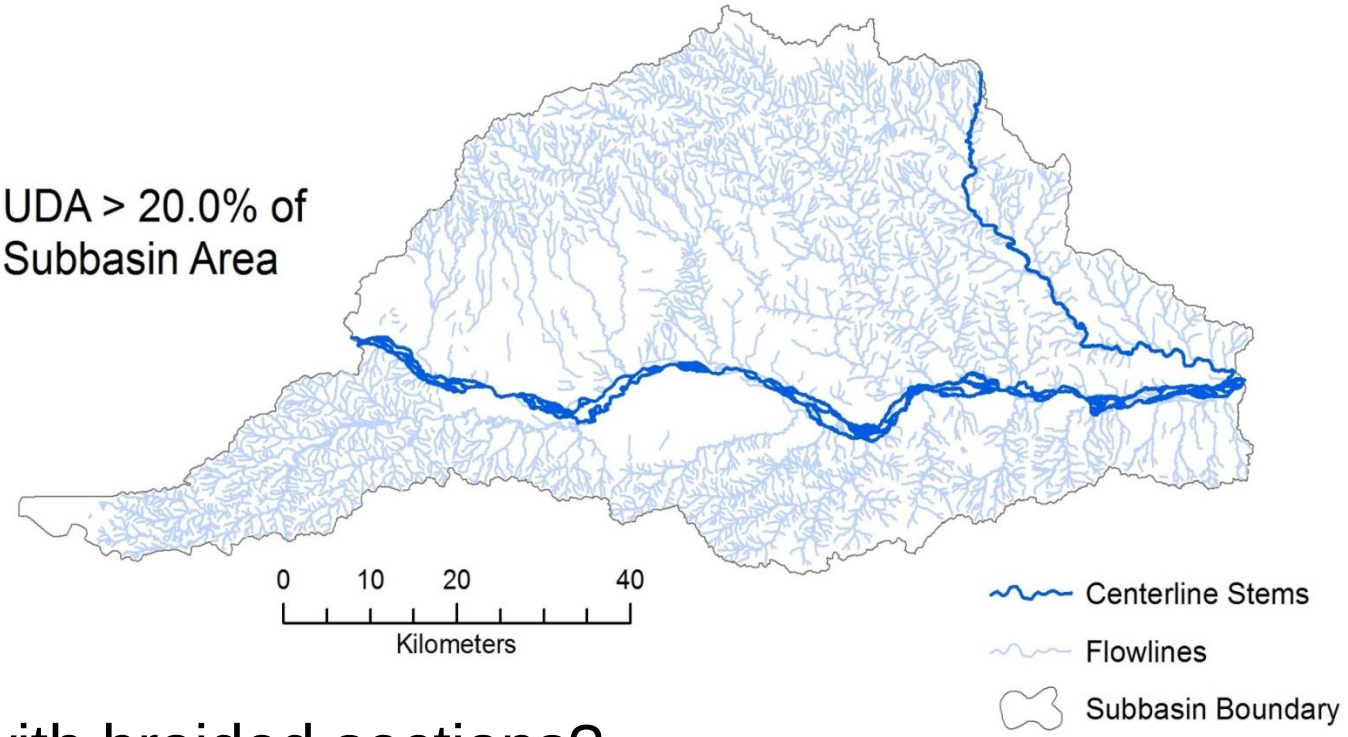
a. Establish primary path “stems”
   -- select on UDA



Lower Prairie Dog Town Fork
Red River, Texas
HUC8: 11120105

(5) Basic Delineation – Node Trace

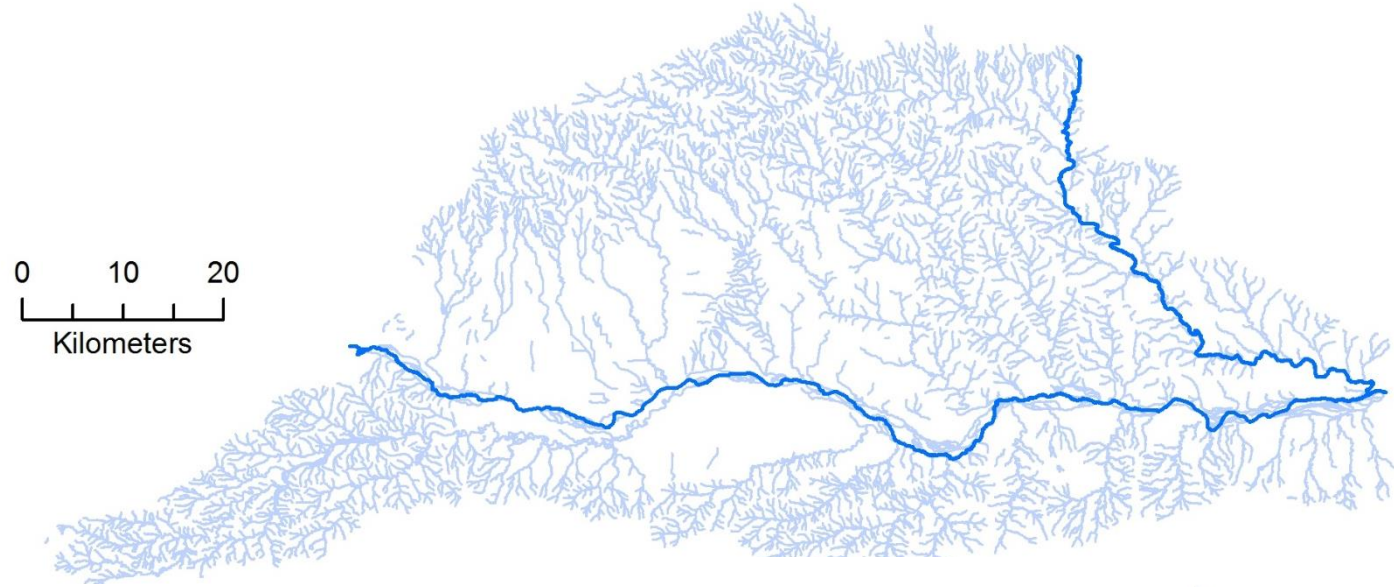
b. Confluence-to-confluence check to follow Reachcode, GNIS name, UDA estimate



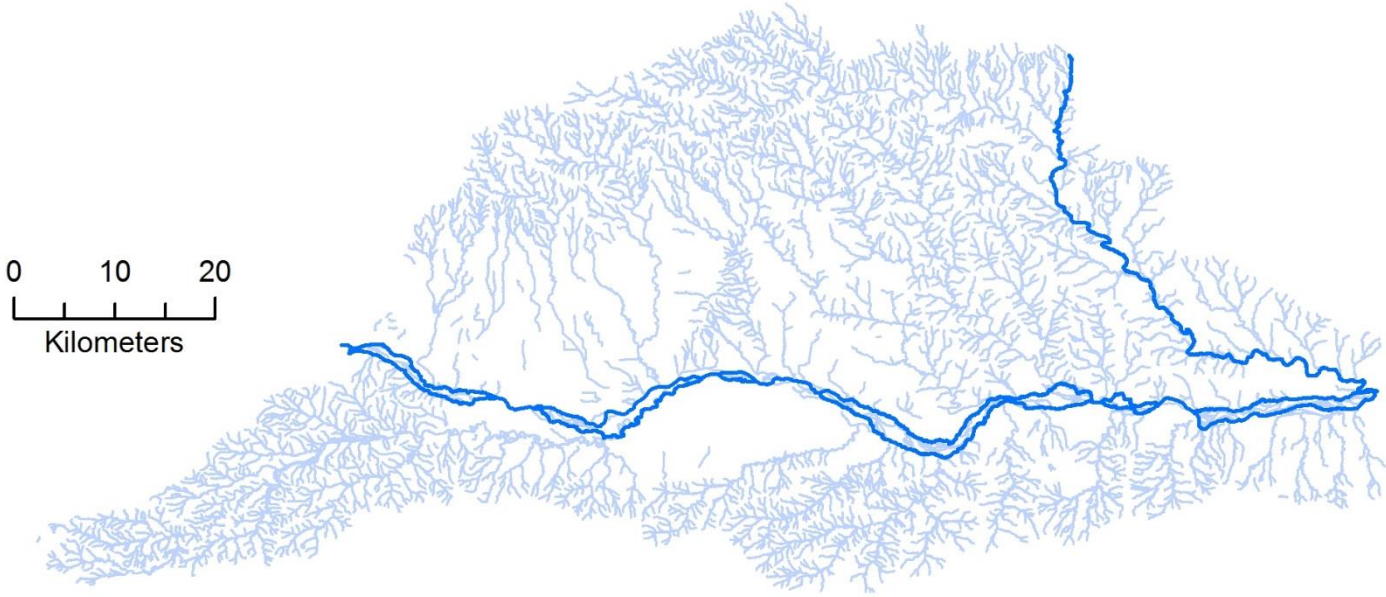
How to deal with braided sections?

Following basic identification, delineate one or several paths through the braid to prioritize channels or to clarify map display

(5) Automatic Delineation of Braids

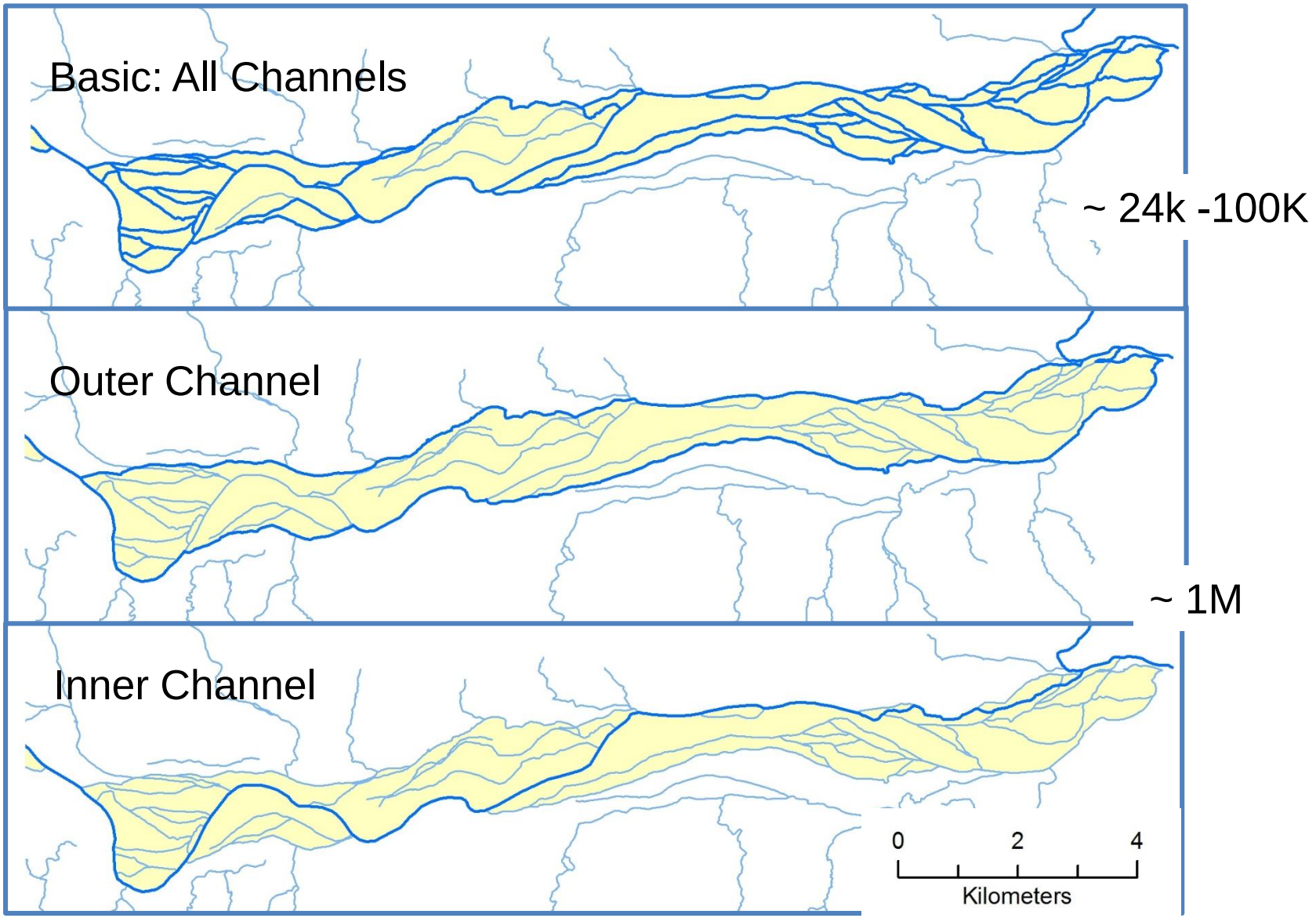


Solution 1
Inner Channel



Solution 2
Outer Channel

Algorithm Comparison

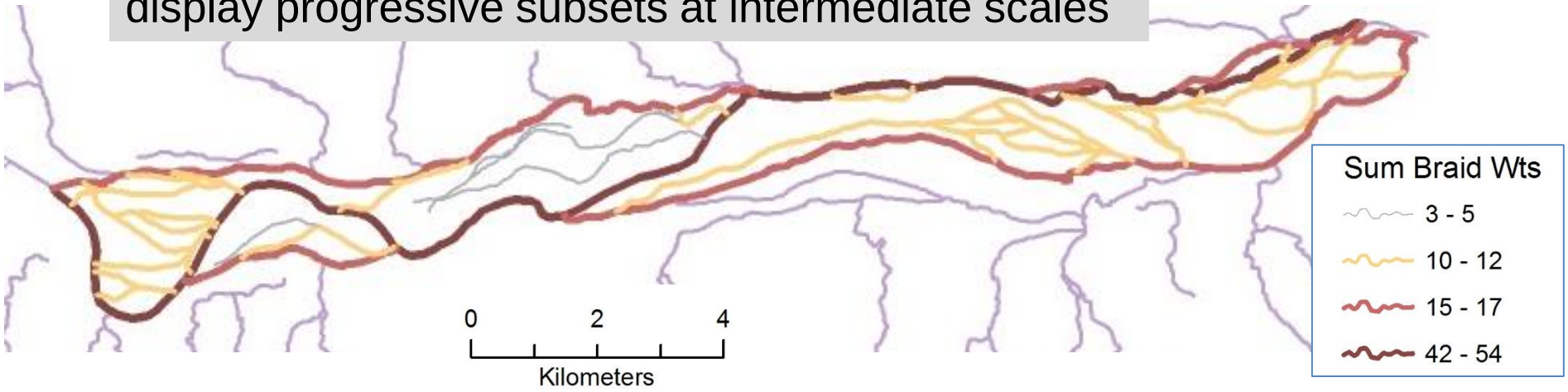


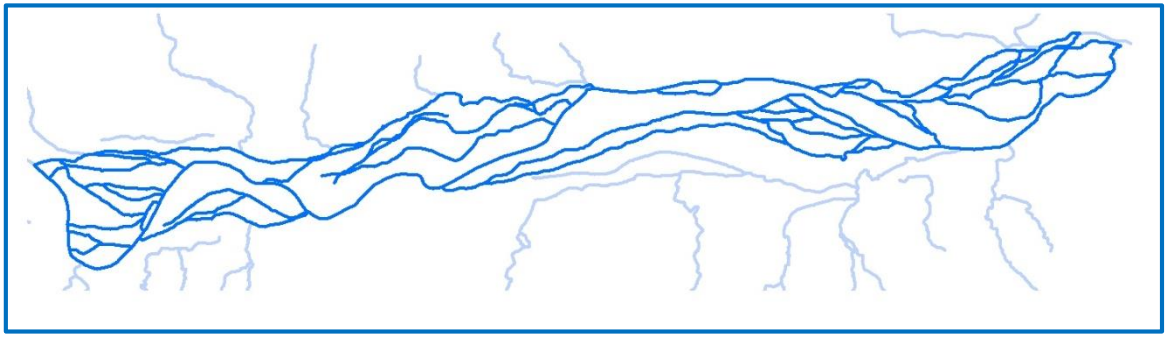
Solution 3: Weighted Channel Algorithm

flow_weights_braid_poly2

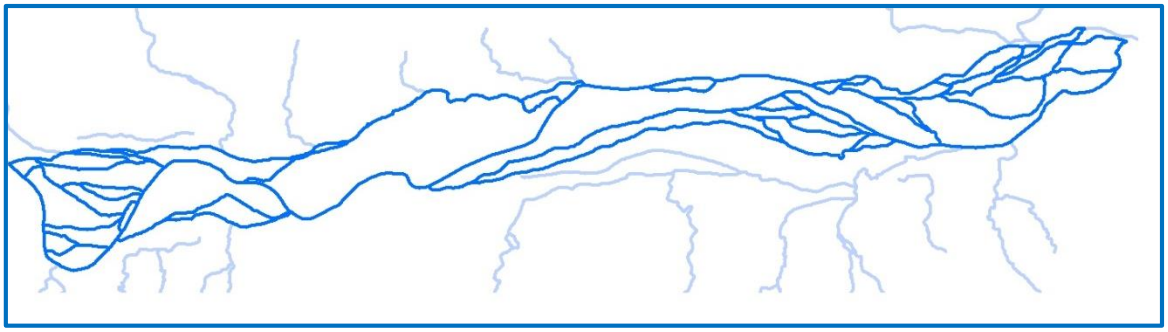
GNIS_Name	ReachCode	WBAreaC	FType	TOT_CM_SQ_KM	inr_pa	all_pat	outr_p	Braid_wt	GNIS_wt	Fcode_wt	inFcode_wt	IP_wt	OP_wt	UDA_wt
<Null>	11120105001474	91860479	558	3002273.48103	FLW	CTR1	CTR1	17	0	3	0	0	10	4
<Null>	11120105001475	91860479	558	3002299.484245	FLW	CTR1	CTR1	17	0	3	0	0	10	4
Prairie Dog Town For	11120105001533	<Null>	558	3002227.905993	CTR1	CTR1	CTR1	52	20	3	0	15	10	4
<Null>	11120105001534	<Null>	460	3002243.379563	FLW	CTR1	FLW	10	0	0	1	0	0	4
<Null>	11120105001534	<Null>	558	3002243.365024	FLW	CTR1	FLW	10	0	0	1	0	0	4
Prairie Dog Town For	11120105001535	<Null>	558	3002252.921182	CTR1	CTR1	FLW	42	20	3	0	15	0	4
Prairie Dog Town For	11120105001536	<Null>	558	3002253.17834	CTR1	CTR1	FLW	42	20	3	0	15	0	4
Prairie Dog Town For	11120105001536	<Null>	558	3002252.956609	CTR1	CTR1	FLW	42	20	3	0	15	0	4
Prairie Dog Town For	11120105001537	<Null>	558	3002254.119431	CTR1	CTR1	FLW	42	20	3	0	15	0	4
<Null>	11120105001539	<Null>	460	0.376885	FLW	FLW	FLW	3	0	0	1	0	0	2
<Null>	11120105001539	<Null>	460	0.370465	FLW	FLW	FLW	3	0	0	1	0	0	2
<Null>	11120105001539	<Null>	460	0.405738	FLW	FLW	FLW	3	0	0	1	0	0	2
<Null>	11120105001540	<Null>	558	3002244.278496	FLW	CTR1	FLW	10	0	0	1	0	0	4
<Null>	11120105001540	<Null>	460	3002244.275736	FLW	CTR1	FLW	10	0	0	1	0	0	4
Prairie Dog Town For	11120105001542	<Null>	558	3002268.803883	CTR1	CTR1	FLW	42	20	3	0	15	0	4
<Null>	11120105001543	<Null>	558	1.120726	FLW	FLW	FLW	5	0	3	0	0	0	2

Rank all flowline channels in braid polygon;
display progressive subsets at intermediate scales

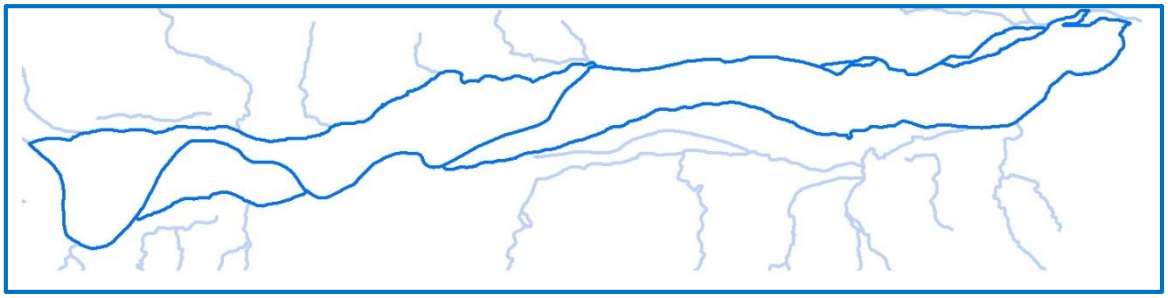




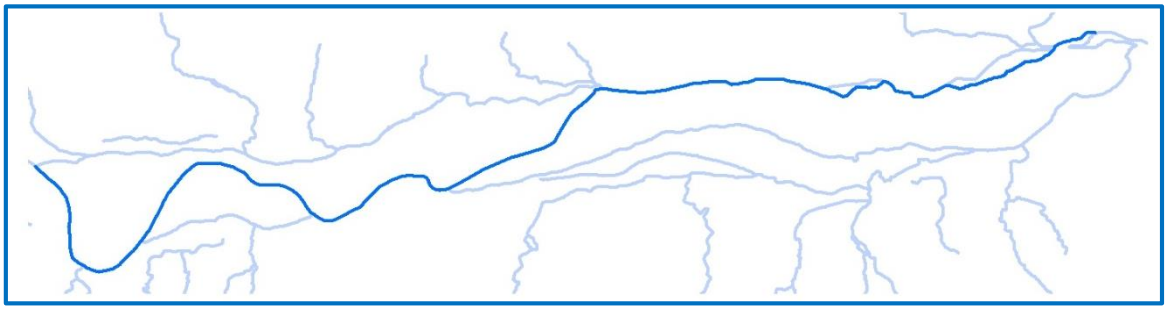
(wts 1-60)
All Paths



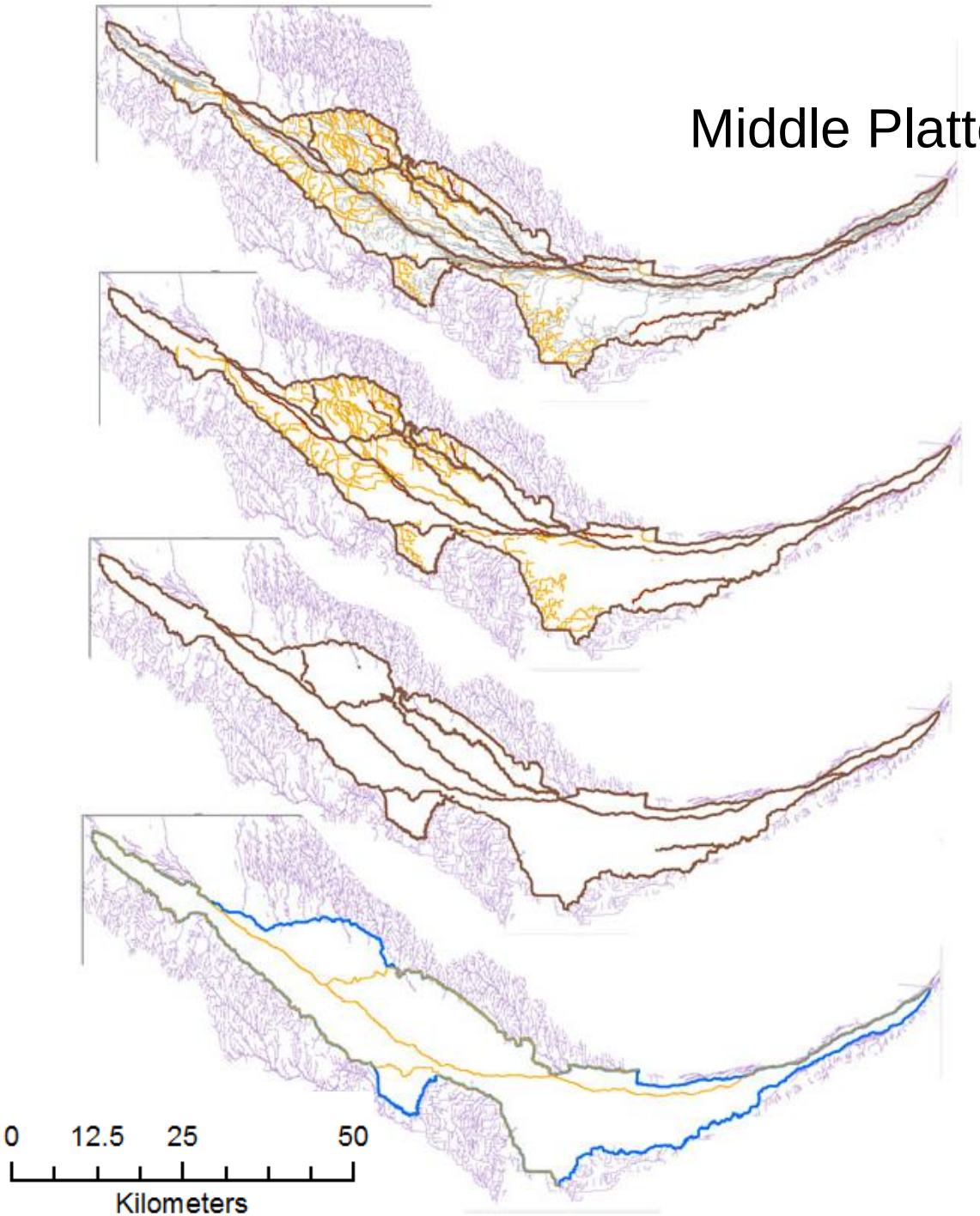
(wts 6-60)
Lose terminal channels



(wts 13-60)
Lose most internal connector but keep outer channel



(wts 18-60)
Identical to Inner Channel solution



NHD 10200101
Middle Platte-Buffalo Rivers, Nebraska

Sum Braid Wts
- 2 - 8
- 9 - 12
- 13 - 16
- 17 - 59

(wts 2-60)

All Paths

(wts 9-60)

Lose terminal channels

(wts 13-60)

Lose canals, ditches, irrigation pipelines

But not quite identical to Inner / Outer Channel solutions

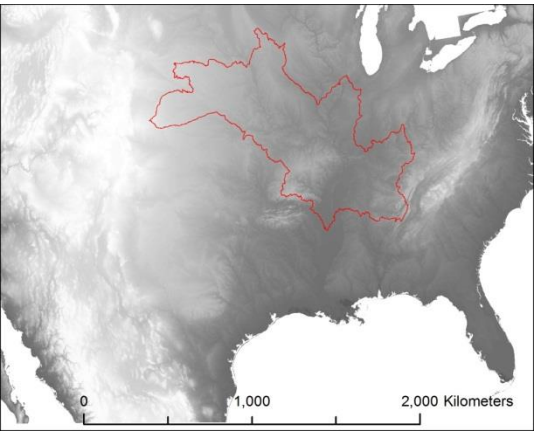
Case Study

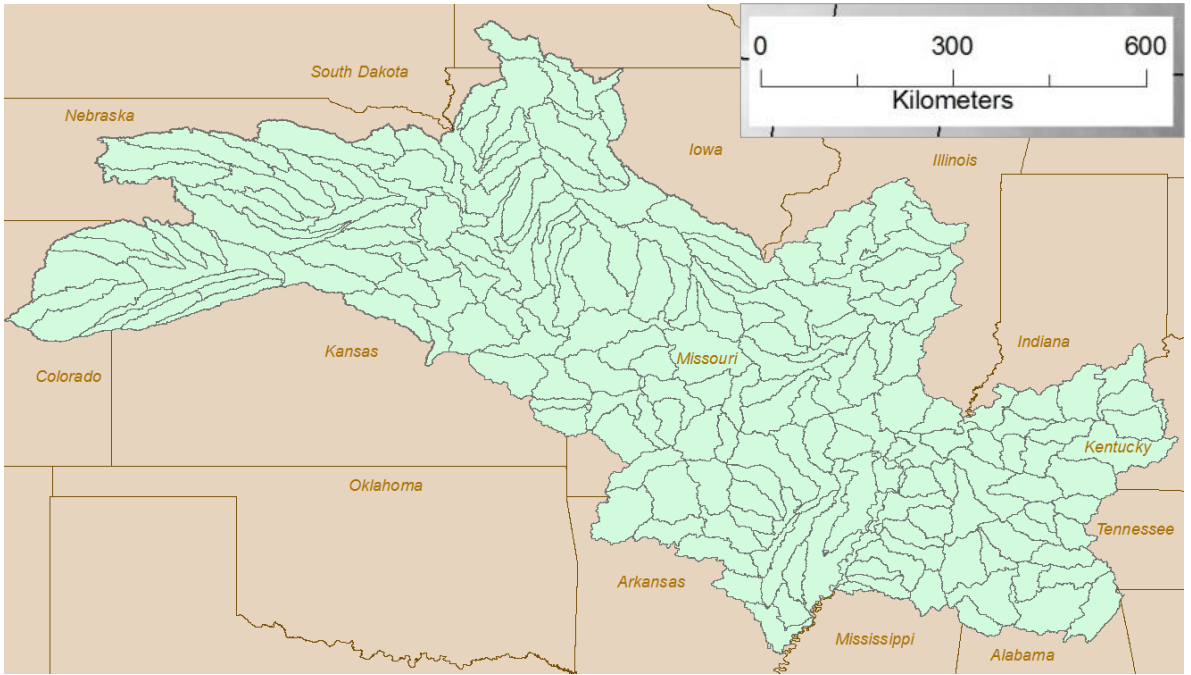
Generalize 1:1M Hydrography from 24K NHD

204 NHD subbasins, different physiographic conditions

- Eastern end more humid, flatter, largely agricultural land use
- Western end drier, more hilly, rangeland and prairie

Mimic parameters used for current version of National Atlas

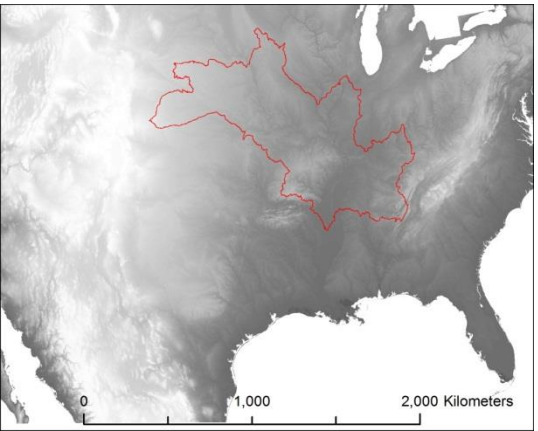


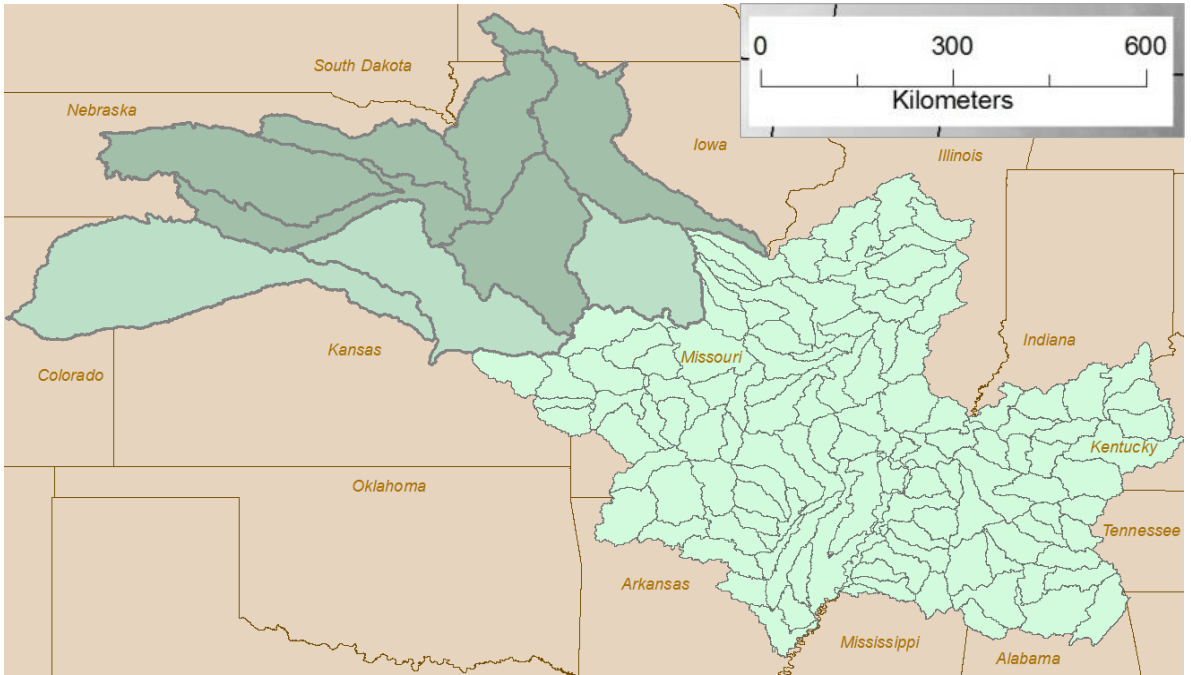


Results Shown Today

24k NHD generalized to 1M in 83 subbasins (9 HUC4s)

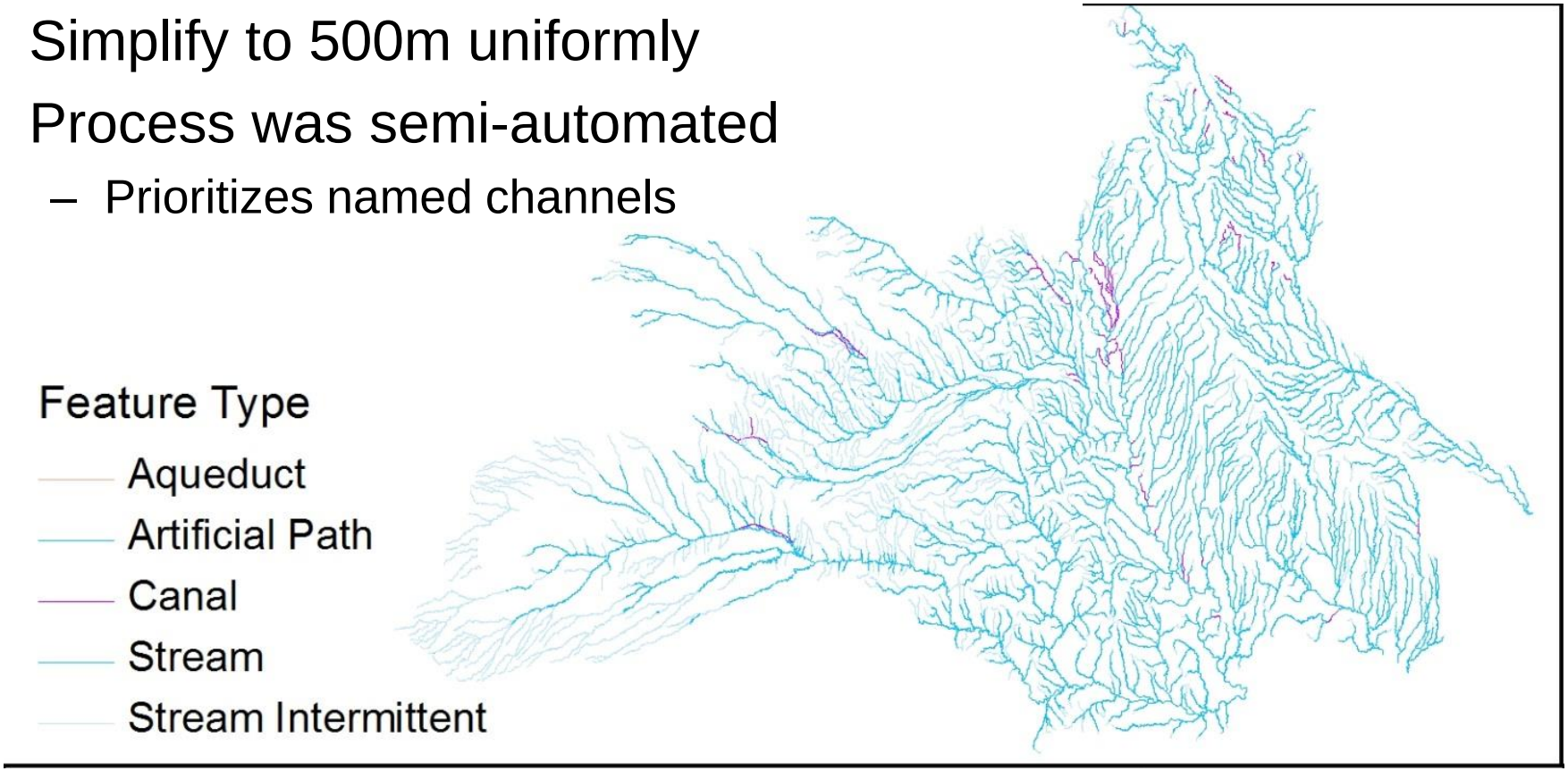
Some results shown only for northern 36 subbasins (6 HUC4s)





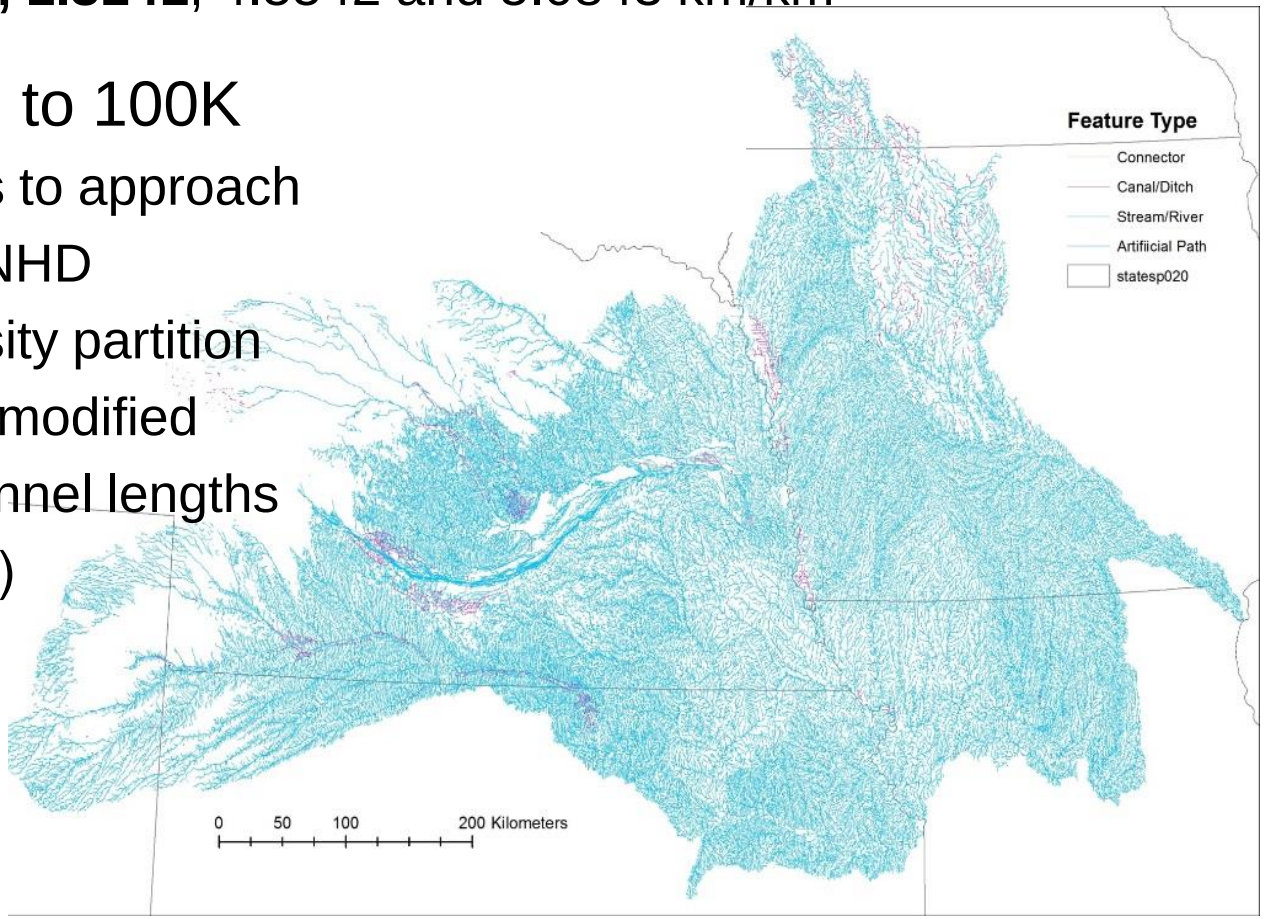
How was National Atlas Data Processed?

- Channel selection from headwaters
  - Headwaters selected on basis of inclusion in USGS 2M DLG data, VMAP_0, and Int'l Map of the World (IMW)
  - Downstream channels by node trace, GNIS name
- Simplify to 500m uniformly
- Process was semi-automated
  - Prioritizes named channels



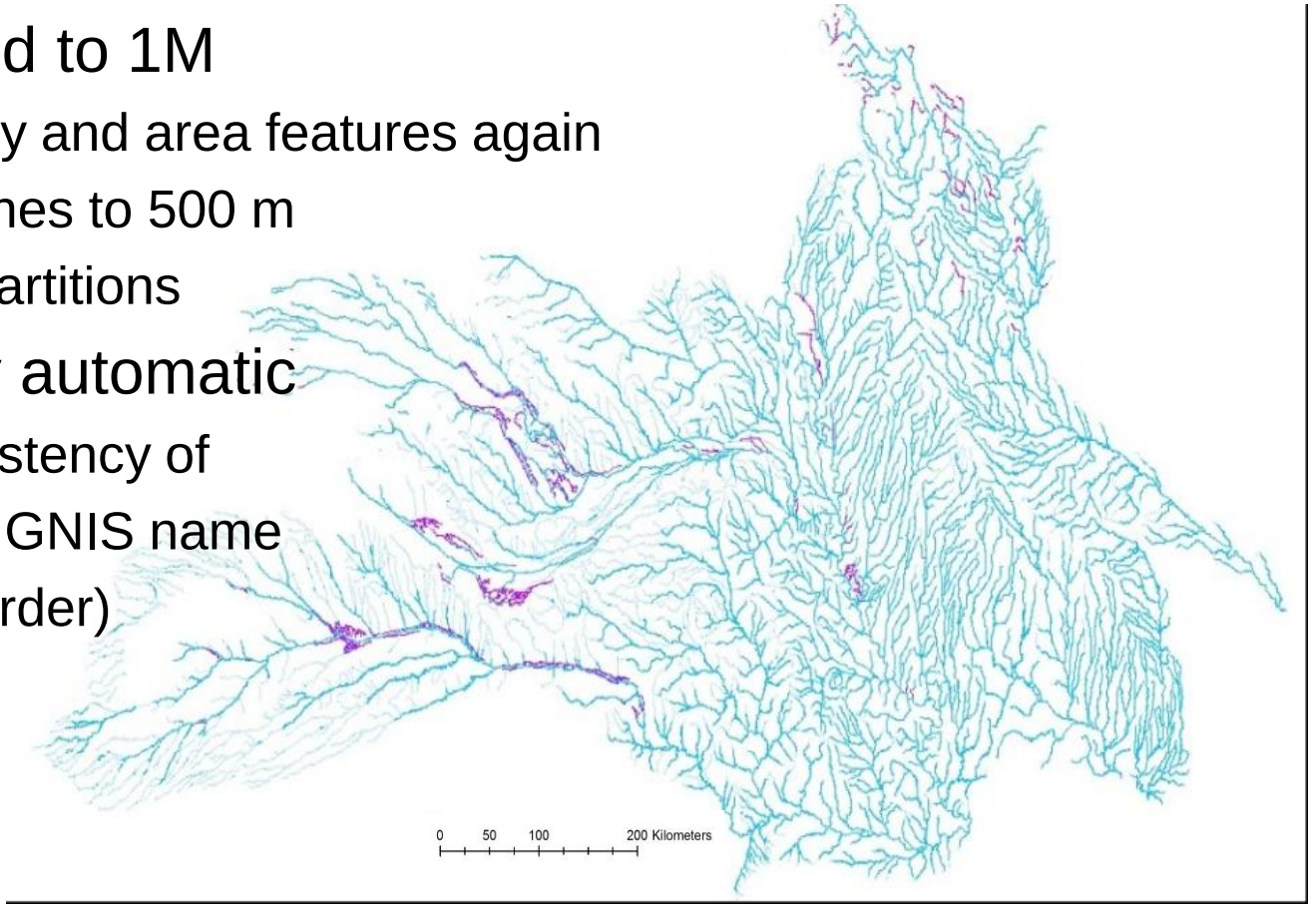
1M LoD Processing (phase 1 = 100k LoD)

- Enrich 24K NHD
  - Estimate catchment areas, UDA
  - Establish local density partitions to guide adaptive pruning
    - **0.0774, 0.7515, 1.3141**, 4.5542 and 5.0843 km/km^2
- 24K generalization to 100K
  - Prune line features to approach densities in 100K NHD
  - Simplify each density partition differentially using modified Radical Law – channel lengths (70m, 50m, 50m…)
  - Prune and simplify water polygons



1M LoD Processing (phase 2 = 1M LoD)

- Enrich 100K with Primary Path (inner channel)
  - Select on UDA >1.5% drainage area to mimic 1M NA flowline network
  - Primary path acts as a surrogate for pruning; protects headwaters
- 100K generalized to 1M
  - Prune waterbody and area features again
  - Bend Simplify lines to 500 m for all density partitions
- Process is fully automatic
  - Prioritizes consistency of channel selection, GNIS name and UDA (in that order)



Results and Comparison

Processing time ~1 hour per subbasin (includes enrichment) but fully automated

As expected, more detail

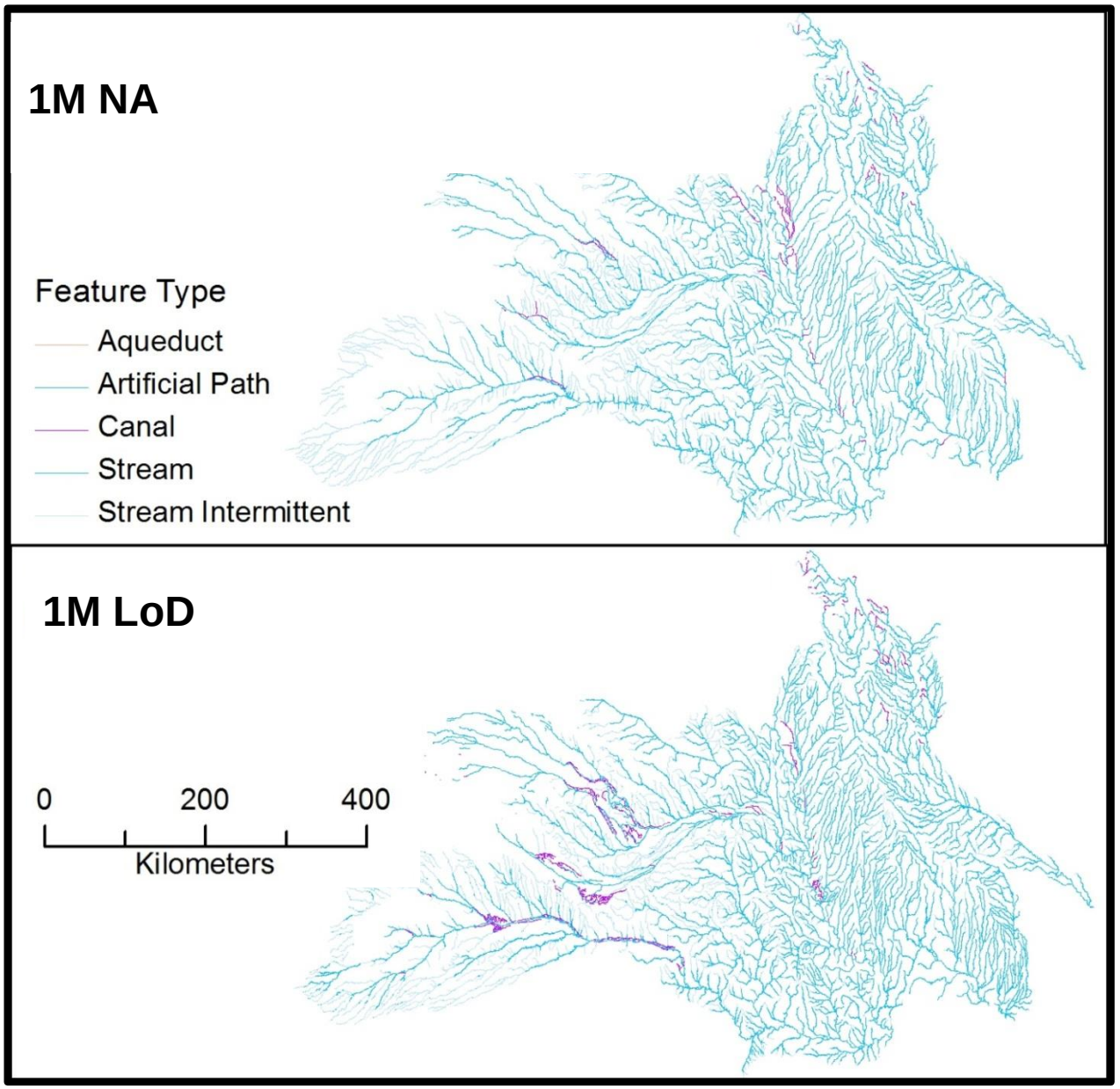
~70,000 LoD features

vs.

~28,000 NA features

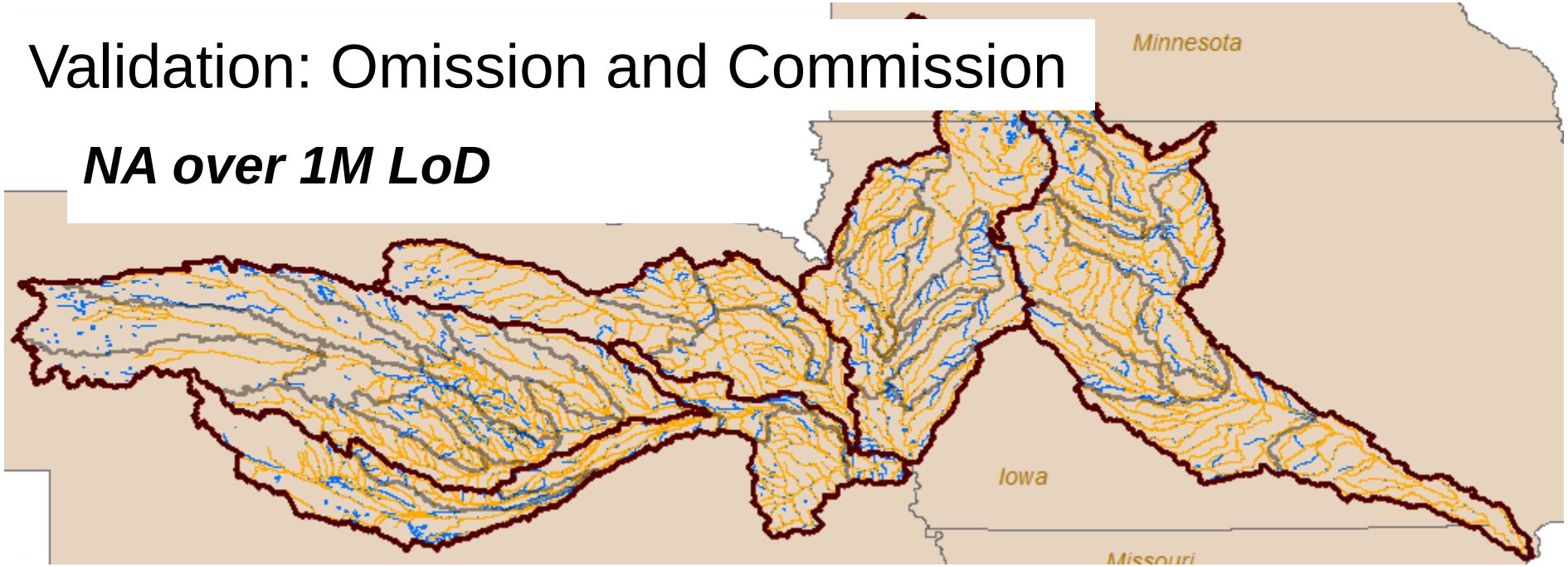
Issues / Problems

- Extra canals in LoD
- Working on this…

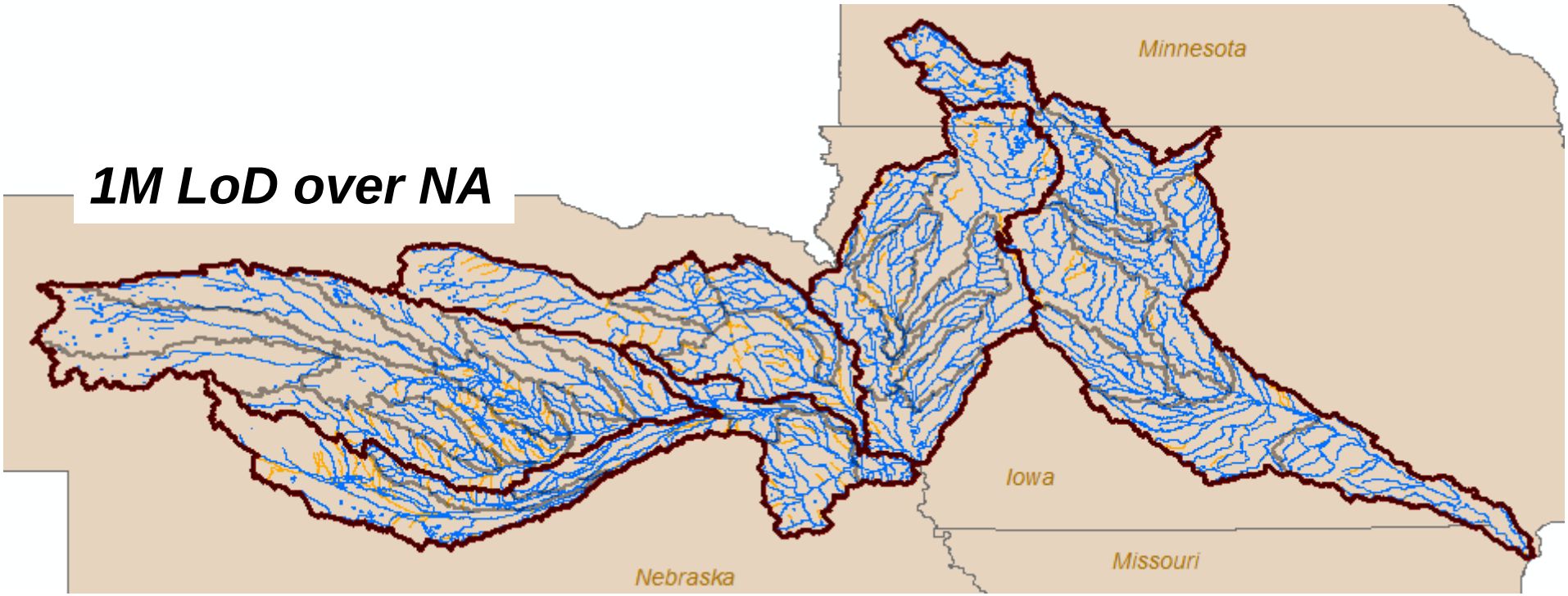


Validation: Omission and Commission

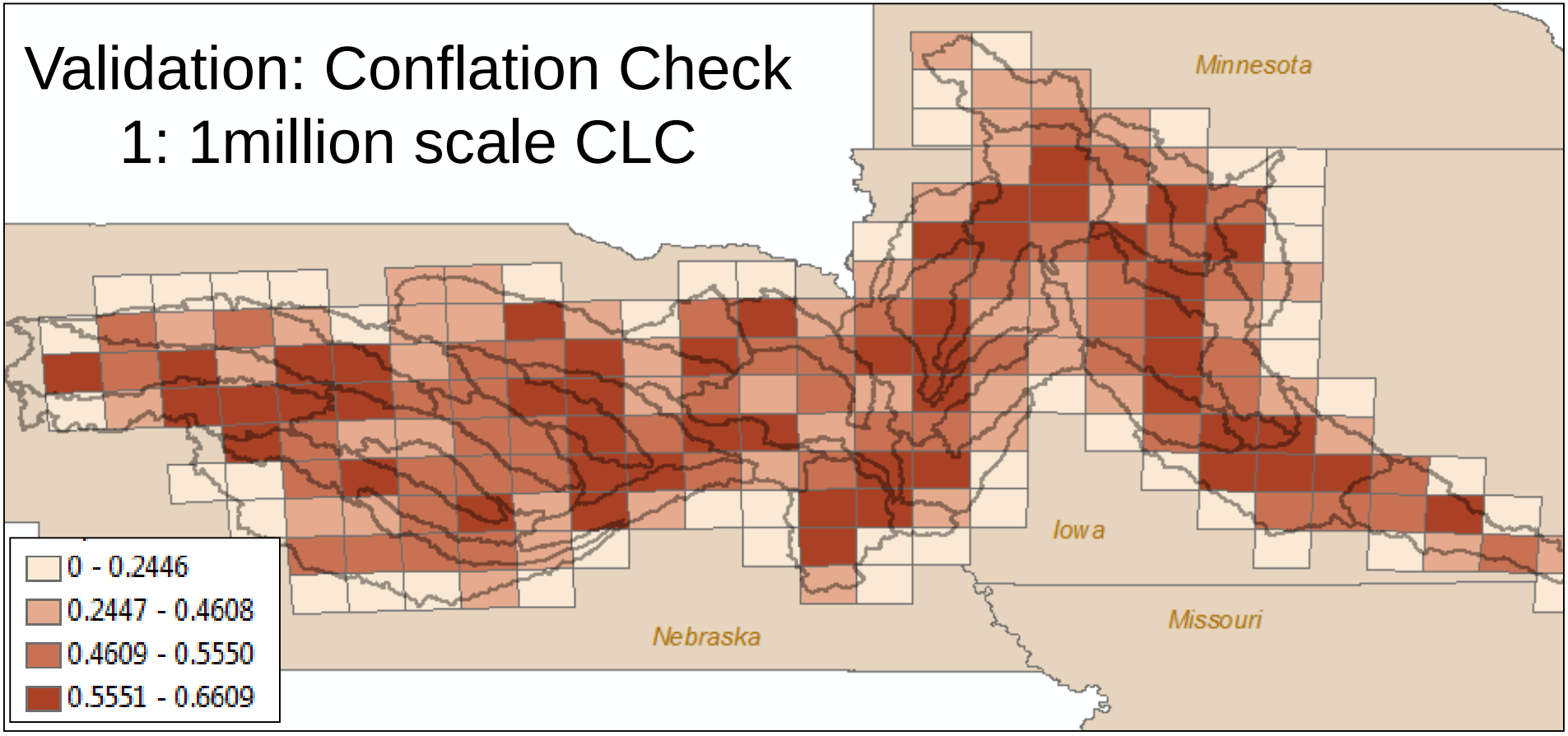
NA over 1M LoD



1M LoD over NA



Validation: Conflation Check 1: 1million scale CLC



Ratio of lengths (CLC) : $\frac{\text{matching channels}}{\text{omissions + comisssions + matches}}$

Values are percentages

- Value for Tier 1 (36 subbasins) = 0.7760 shown here
- Value for 204 subbasins = 0.7856

Summary

Fully automated generalization is achievable and offers several benefits for national mapping agency multi-scale data production:

- Protect local variations in stream density to improve cartographic aesthetics and geographic validity
- Replace 3 versions of NHD with integrated MRDB – automatic insertion of feature level linkages between 24K, 100K, 1M scales
- Explore primary path as alternative to data pruning to protect headwaters consistently for very large scale jumps