NHDPlus V2 Accumulation Tools and High-Resolution Data

Sarah McDonald, USGS Contractor

Chesapeake Bay Program Office

Lower Mississippi-Gulf Water Science Center

Medium Resolution Data (1:100k Scale)

- NHDPlusV2 is medium resolution National Hydrography Dataset (NHD) data.
  - Catchments
  - Flowlines
  - PlusFlow
  - PlusFlow AR
    - AR: Addition/Removal events
- SAS Routing database (Mike Wieczorek)
  - Improved NHDPlusV2 PlusFlow table
  - Includes attributes such as:
  - https://www.sciencebase.gov/catalog/item/5669a79ee4b08895842a1d47

NHDPlusV2 Accumulation Tools - Overview

- StreamCat
  - USEPA; Marc Weber, Rick Debbout, Ryan Hill, …
  - https://github.com/USEPA/StreamCat

- Stream Connectivity with a Routing Database
  - USGS; Sarah McDonald
  - https://github.com/smcdonald125/NHD-Accumulation
  - Goal:
    - Translate Mike Wieczorek's SAS code to an open-source, Python tool

StreamCat

- Required data to build upstream network:
  - NHDPlusV2 PlusFlow
  - NHDPlusV2 flowlines dbf
  - interVPU.csv
- The PlusFlow table is altered by removing records from the PlusFlow table:
  - that do not flow to another segment (TOCOMID == 0)
  - where the FROMCOMID is type “CoastLine” in the NHD flowlines dbf
  - where FROMCOMID is listed in the Remove column of InterVPU.csv
  - where the FROMCOMID is not in the current VPU zone and is not listed as the thruCOMID in InterVPU.csv

StreamCat Continued

- Uses ArcPy to calculate catchment statistics
  - Continuous Raster – ZonalStatisticsAsTable
  - Classified Raster – Tabulate Area
- ArcPy cell size environment is set to 30m for all data
- Accumulating area
  - The results of the catchment statistics include a “raster area” (excludes NoData)
  - Raster area is divided by the NHD catchment attribute “AreaSqKM” to produce the “PctFull” variable.
  - AreaSqKM is accumulated like all variables (summing all values in the network)
  - PctFull is accumulated by taking a weighted average of the PctFull data, using the catchment areas (AreaSqKM) as the weights.

StreamCat Continued

- Final accumulated values can be processed using one of the following three methods:
  - Mean, Density, Percent
  - Density and Percent use the accumulated AreaSqKM value multiplied by the weighted average of the PctFull value to calculate “total accumulated raster area”

Stream Connectivity with a Routing Database

- Open source Python 3.7 tool; Libraries listed below:
  - gdal, geopandas , fiona, numpy, pandas, rasterio, sas7bdat , shapely
- Run Time: ~80 minutes for continuous raster covering the CBW
  - Can be optimized
- Required data to build upstream network:
  - SASS routing database (CONUS)
- The Routing Database is altered by:
  - Reducing the database to a Vector Processing Unit (VPU), if a zone is specified (recommended)
    - The reduced table will be produced as a CSV for future use
  - Removing attributes that are not needed for connectivity

Stream Connectivity with a Routing Database Continued

- Catchment statistics are calculated using open-source numpy functions
  - Continuous Rasters – One stat can be picked per run:
    - Maximum: np.amax()
    - Minimum: np.amin()
    - Mean: np.nanmean()
    - Median: np.nanmedian()
    - Sum: np.sum()
  - Classified Rasters:
    - Count of pixels for each unique class: np.unique()

Stream Connectivity with a Routing Database Continued

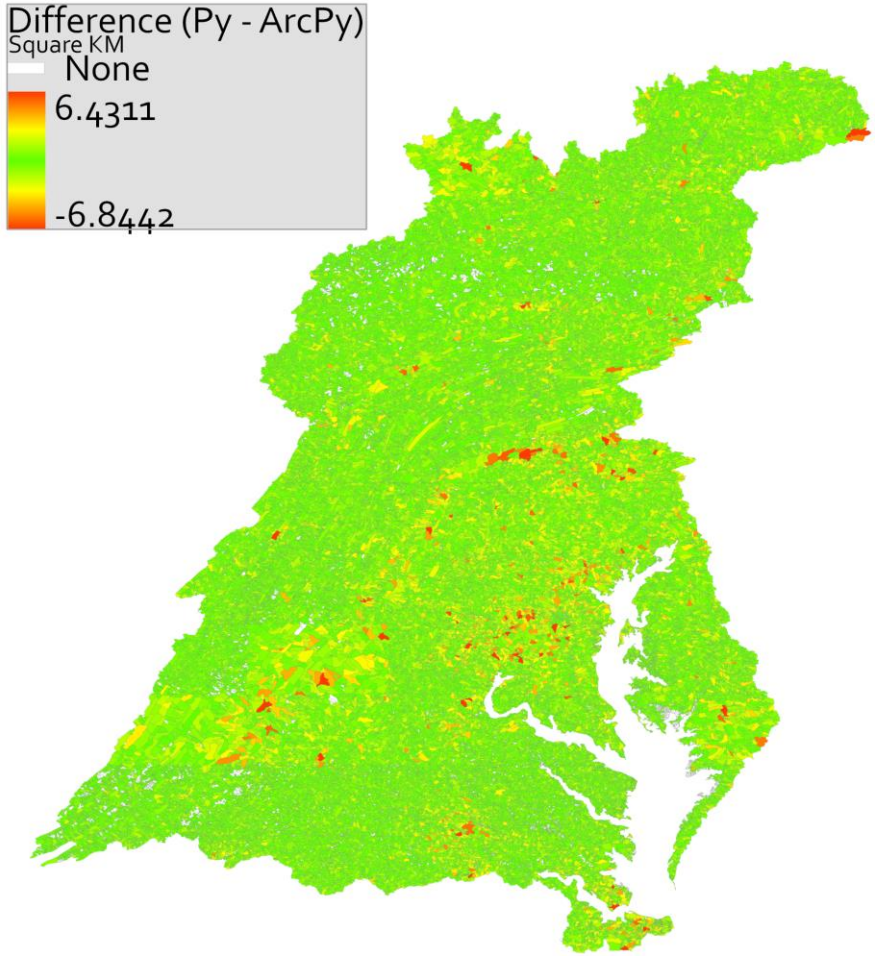
- Accumulating Area:
  - Continuous Raster
    - Raster area covering the catchment that is not NoData is summed for all upstream catchments
  - Classified Raster
    - Pixel count per class is summed for all upstream catchments and multiplied by a user-defined conversion factor

StreamCat VS Stream Connectivity with a Routing Database

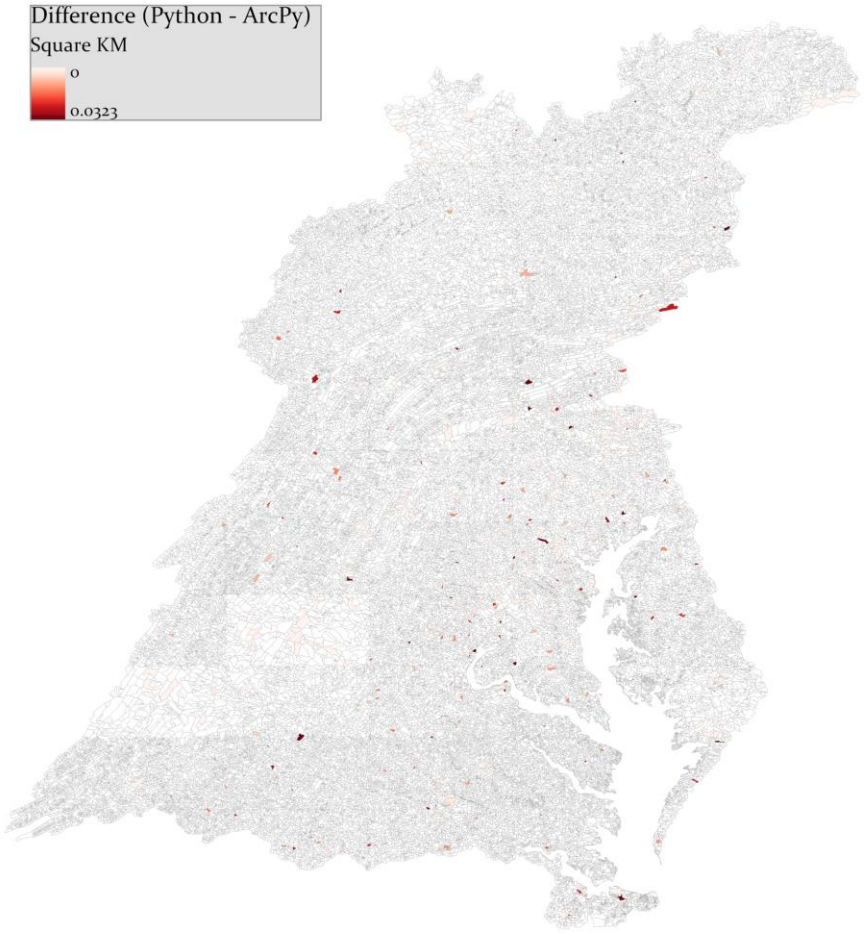
- Comparing area calculation methods:
  - For the Chesapeake Bay Watershed, using a 10m raster that covers all catchments (no NoData)
    - PctFull * Accumulated NHD Area method yielded the exact values as the accumulated raster area
- Enforcing a 30m cell size in the ArcPy environment with finer scale data (10m)
  - Resulted in difference in catchment statistic values

StreamCat VS Stream Connectivity with a Routing Database Cont

Catchment Impervious Cover Difference:
ArcPy ZonalStats vs Python numpy
ArcPy Environment Cell size = 30

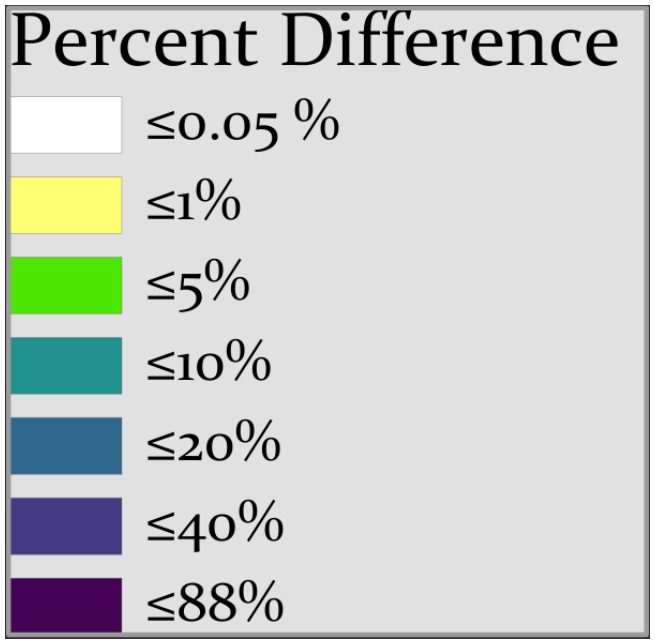


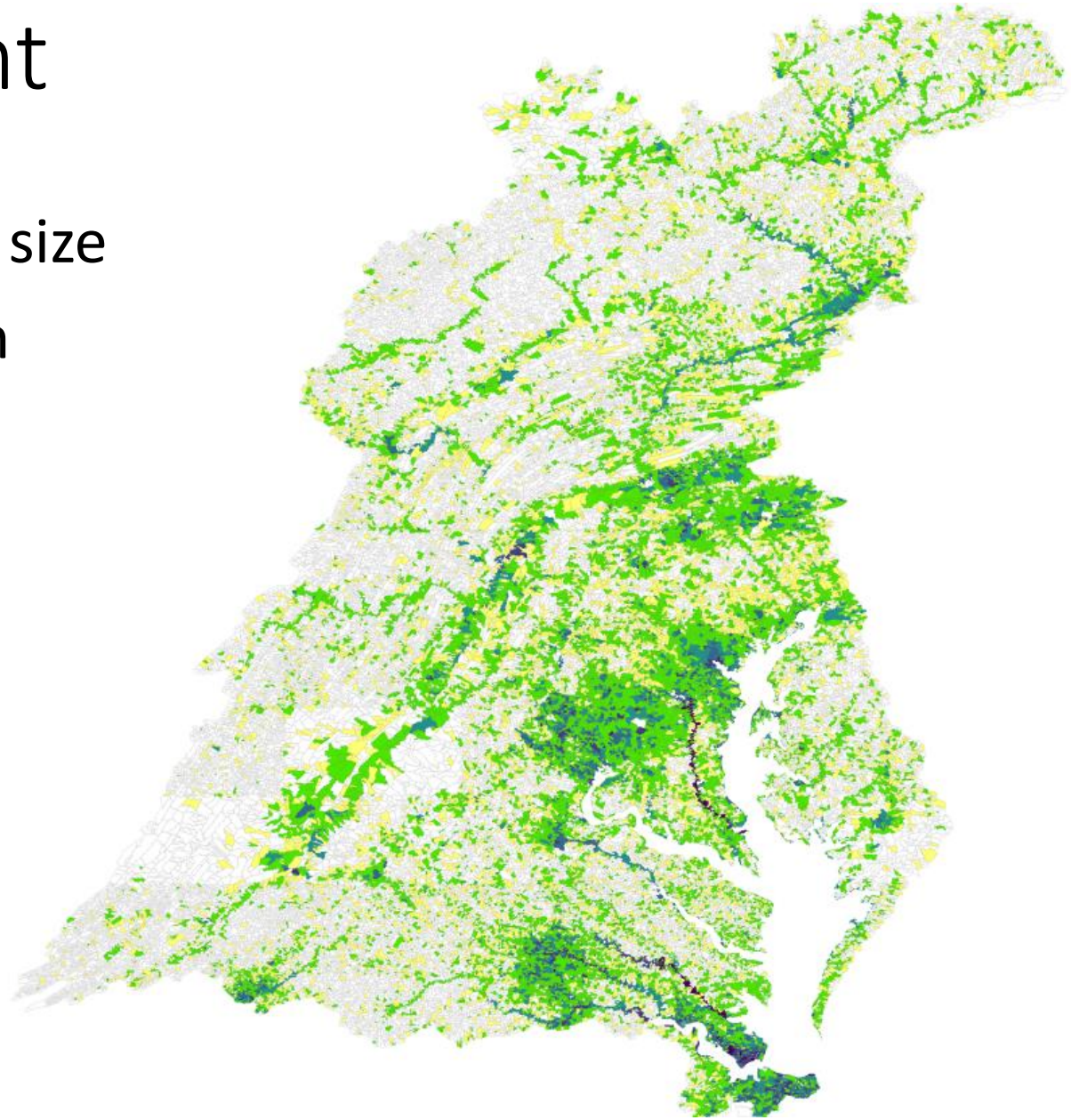
Catchment Impervious Cover Difference:
ArcPy ZonalStats vs Python numpy
ArcPy Environment Cell size = 10



StreamCat VS Stream Connectivity with a Routing Database Cont

- Zonal Statistics using 30m cell size results accumulated in Python code. Differenced with StreamCat Data





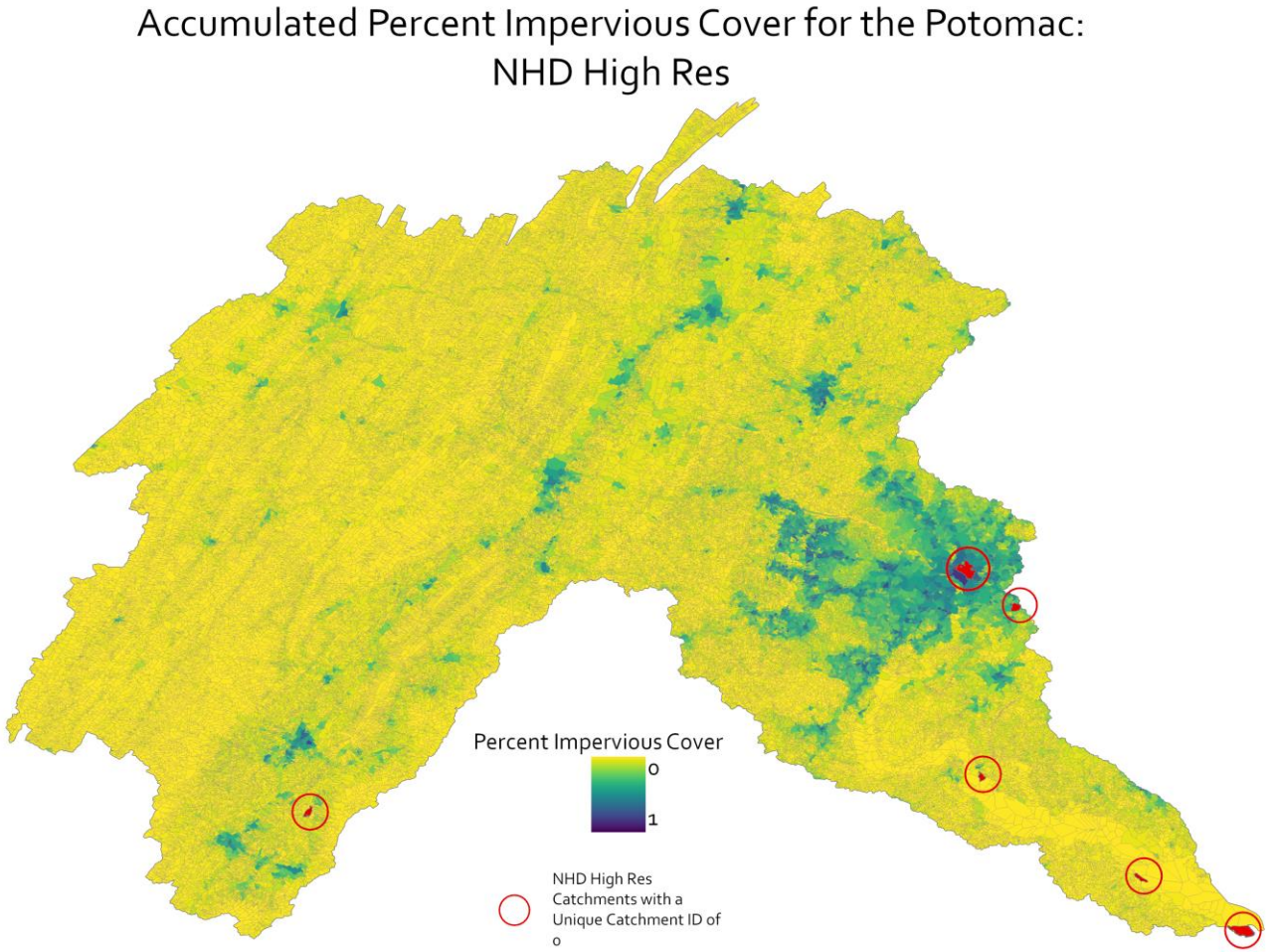
High Resolution Data: 1:24k Scale

- NHD-HR is high-resolution National Hydrography Dataset (NHD) data; Can be downloaded at the HUC4 and HUC8 subbasin level
  - Catchments
  - Flowlines
  - PlusFlow

- ecoSHEDs is improved NHD-HR data
  - Catchments
  - Truncated flowlines
  - http://conte-ecology.github.io/shedsGisData/

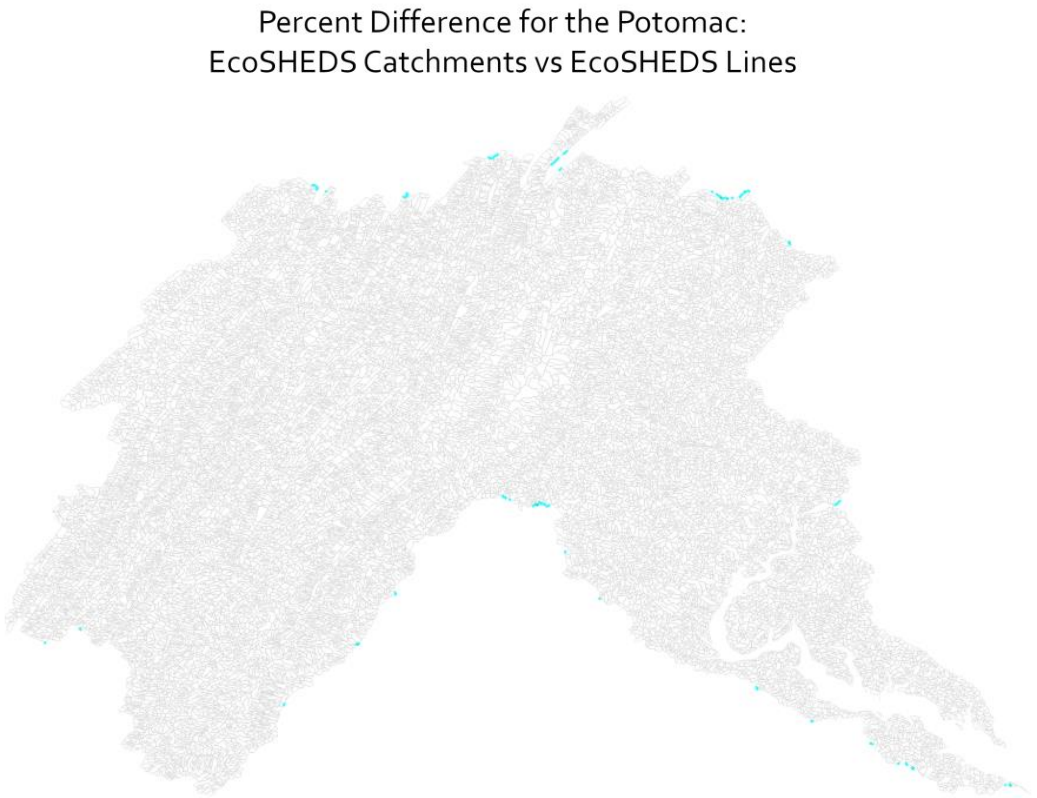
NHD-HR

- The NHD High Res catchments contained some errors



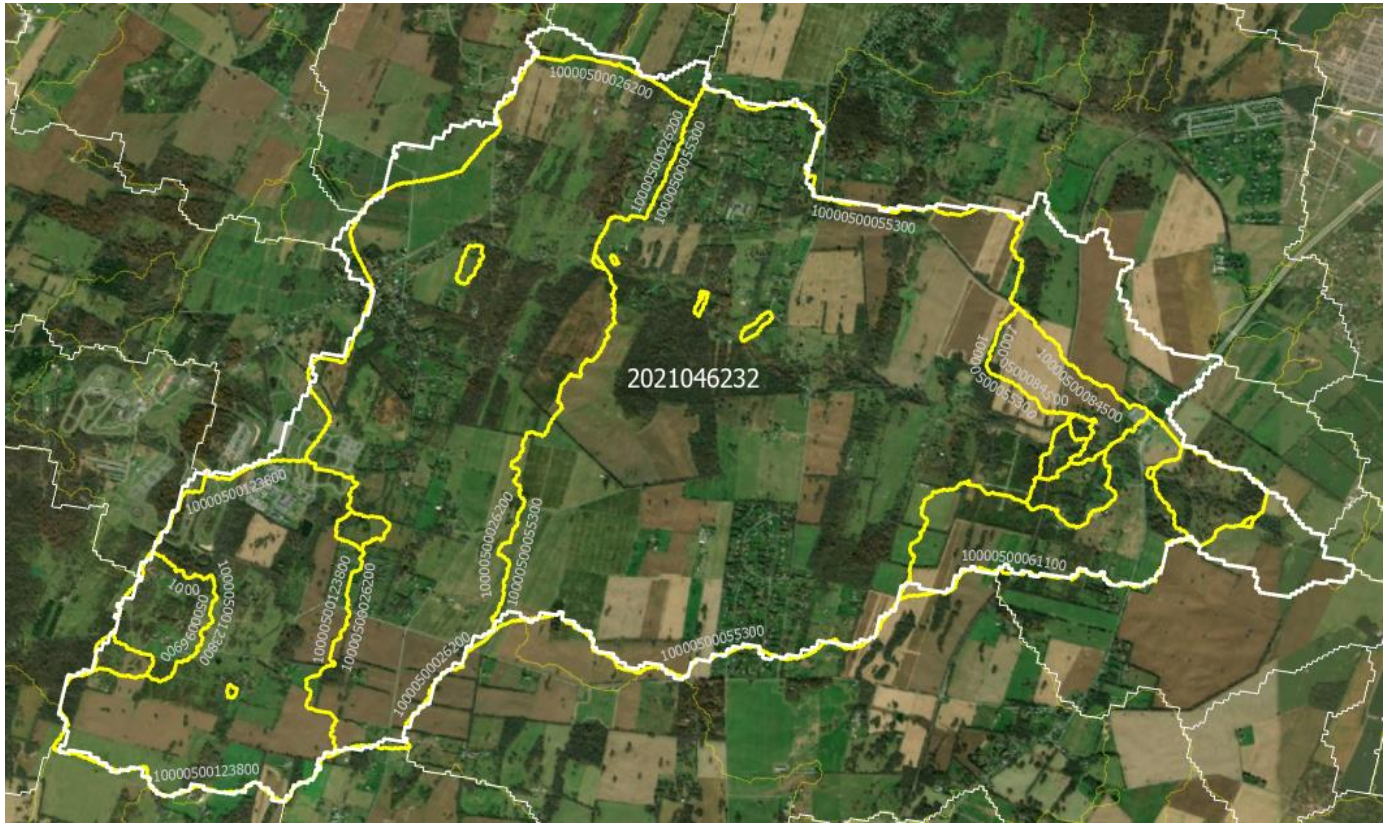
ecoSHEDs

- Both the catchments and truncated flowlines contain the “NextDownID” data that can be used to build the upstream network
- The upstream networks are different depending which dataset was used



NHD-HR and ecoSHEDs Comparison

- NHD-HR includes more catchments than ecoSHEDs
  - NHD-HR has a total of 106,729 catchments in the Potomac
  - ecoSHEDs has a total of 25,860 catchments in the Potomac

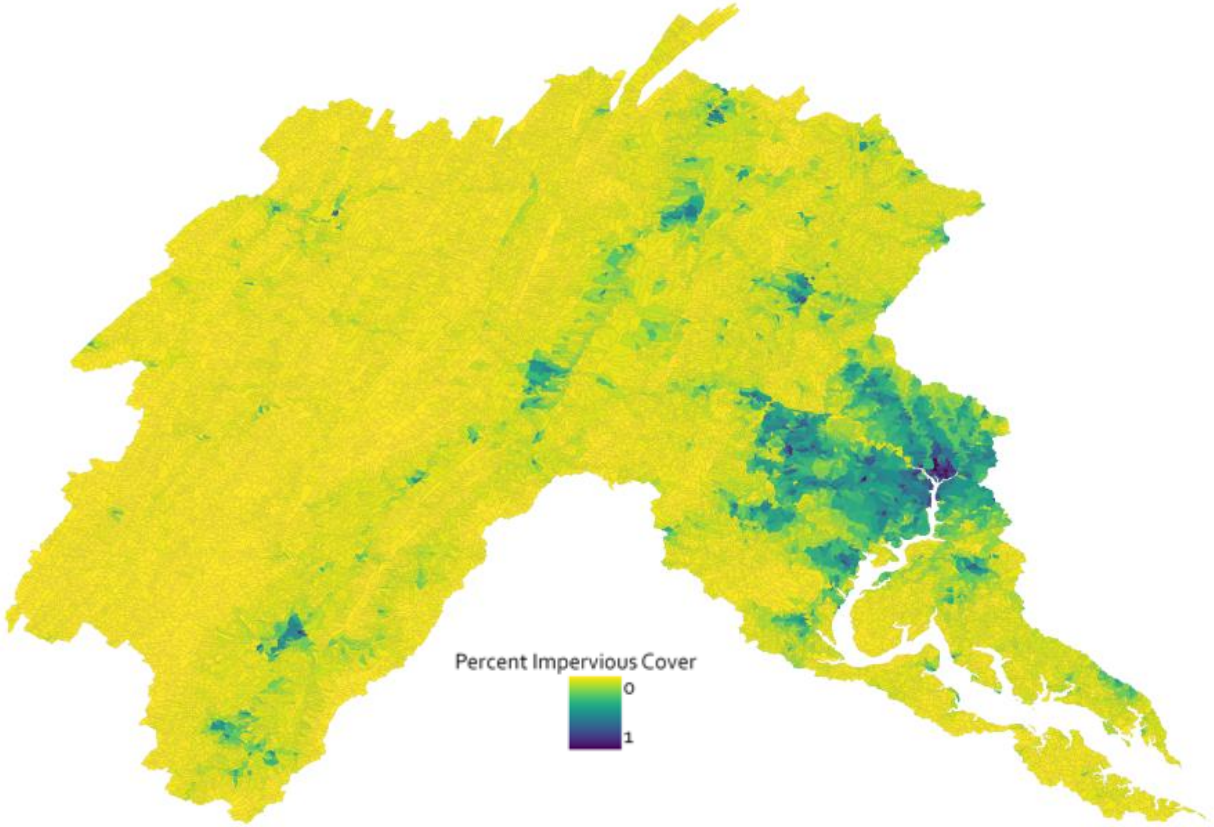


NHD-HR Catchments

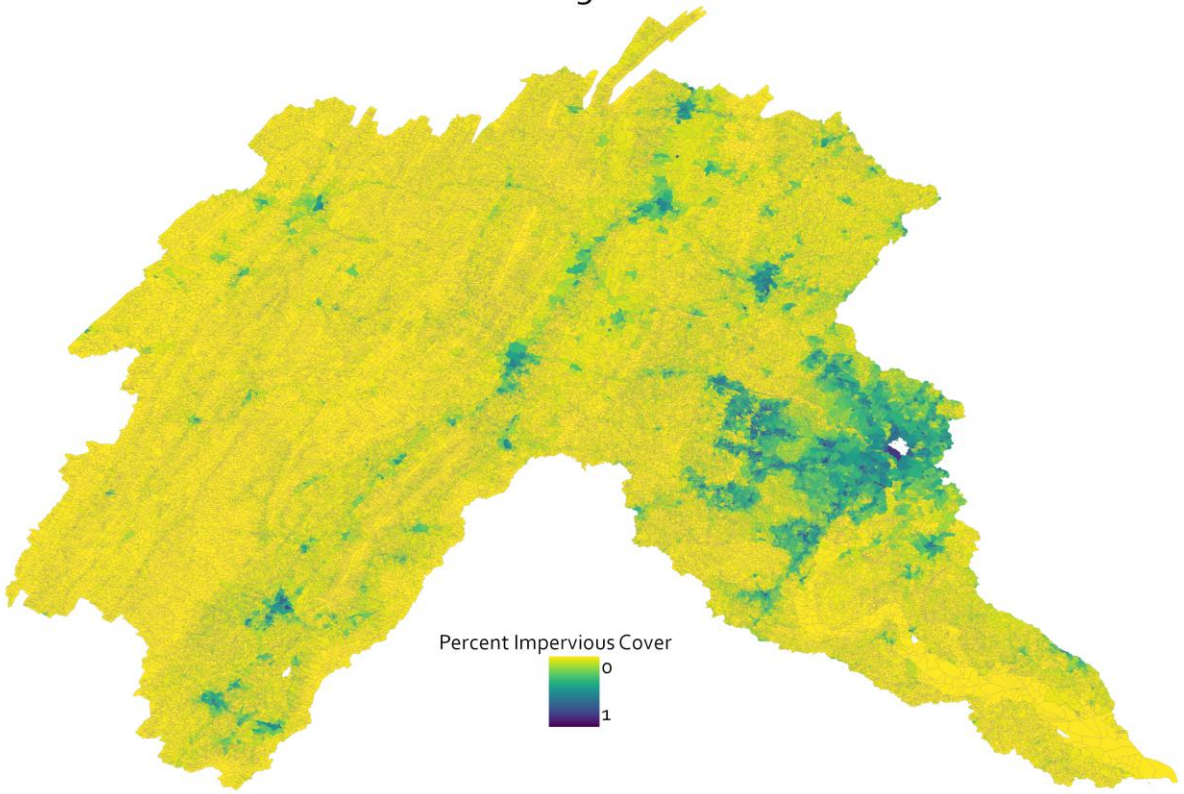
ecoSHEDs Catchments

NHD-HR and ecoSHEDs Comparison

Accumulated Percent Impervious Cover for the Potomac:
EcoSHEDS Catchments

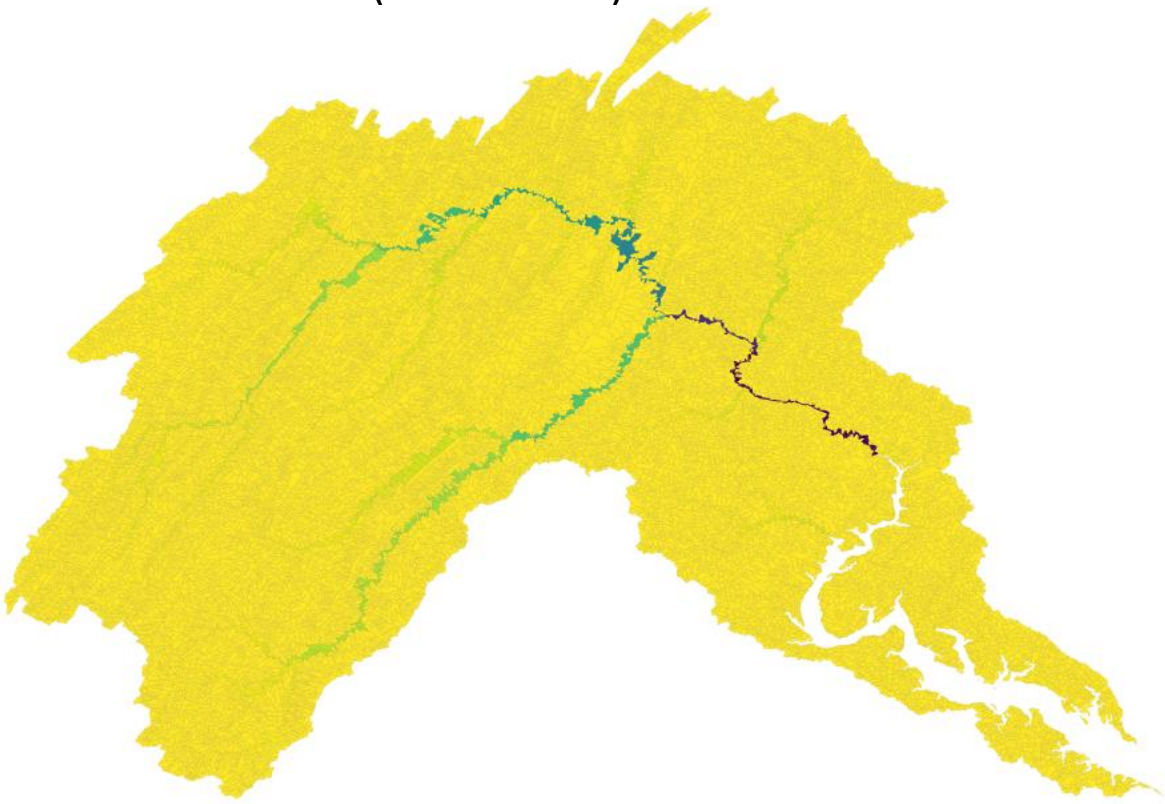


Accumulated Percent Impervious Cover for the Potomac:
NHD High Res

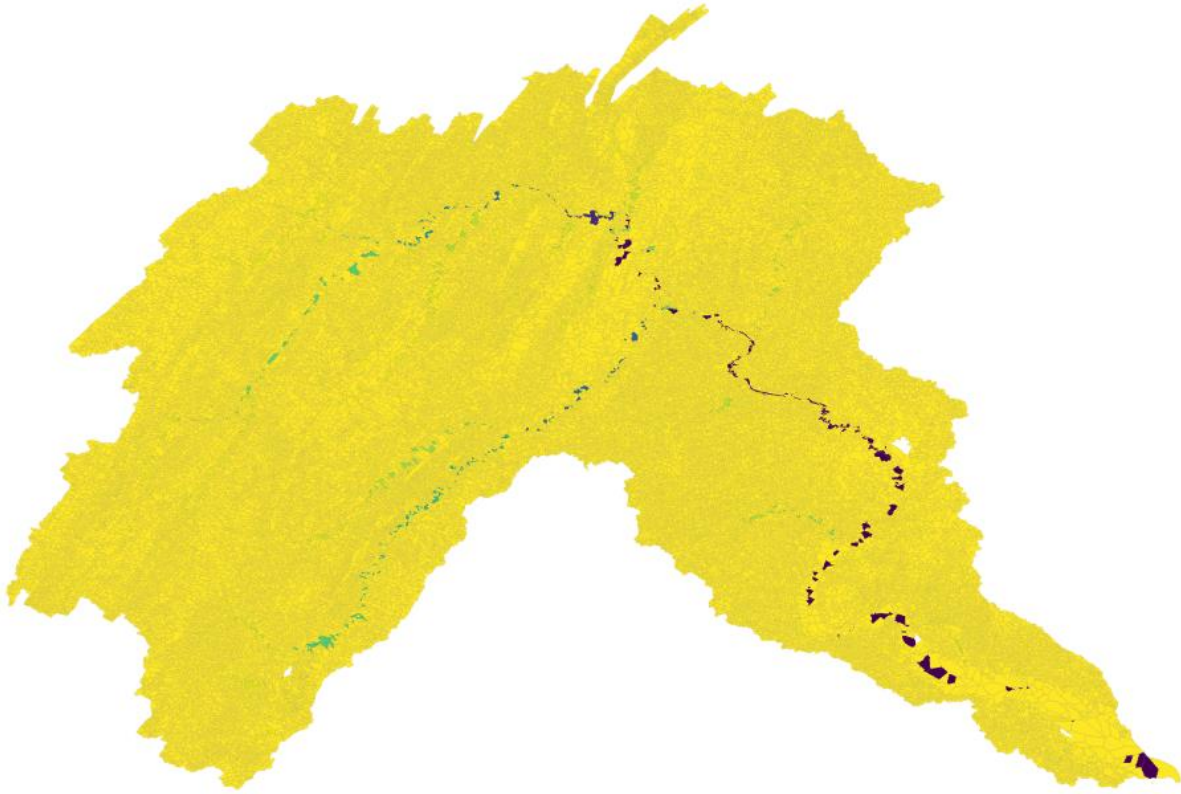


NHD-HR and ecoSHEDs Comparison Cont

EcoSheds (catchments)



NHD High Res



Immediate Applications

Stream Connectivity with Routing Database

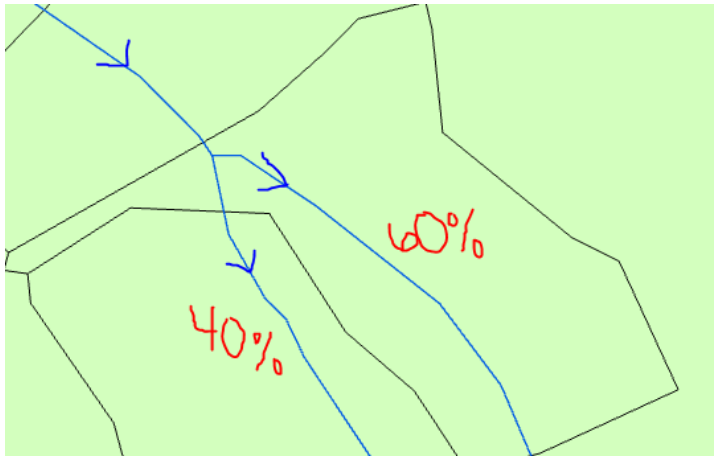
- Accumulate Chesapeake bay Healthy Watersheds landscape metrics
- Accumulate simulated impervious cover from the Chesapeake Bay Land Change Model (CBLCM)
- Update sediment simulation for the Chesapeake Bay Watershed Model

Next Steps

- Update to allow user to accumulate more than one raster
- Include a weighted average by % flow for divergences option
  - Wieczorek's SAS code does this
- Add functionality to mask by region(s) to accumulate within
- Make code user-friendly
  - Jupyter notebook?
  - Configuration file?
  - What is easiest to use and set up locally?

Questions

- What attributes should be included for routing?
  - Medium resolution uses ctonode and cfromnode in the routing database
  - High res can use
    - NHD's ToNHDPID and FromNHDPID
    - ecoSHED's NextDownID
- Most effective way of sharing?
  - Jupyter Notebook?
  - GitHub?



- Is it useful to accumulate values based on the % flow divergences?
  - Divergences split flow, should the accumulation account for this?
  - Should all metrics be accumulated this way, or should it be optional?
  - Should area be accumulated this way?
- Which of these questions is most important?
- Others?

Contact

- smcdonald@chesapeakebay.net