CBP LULC Data Project: 2022 Release: Lessons Learned

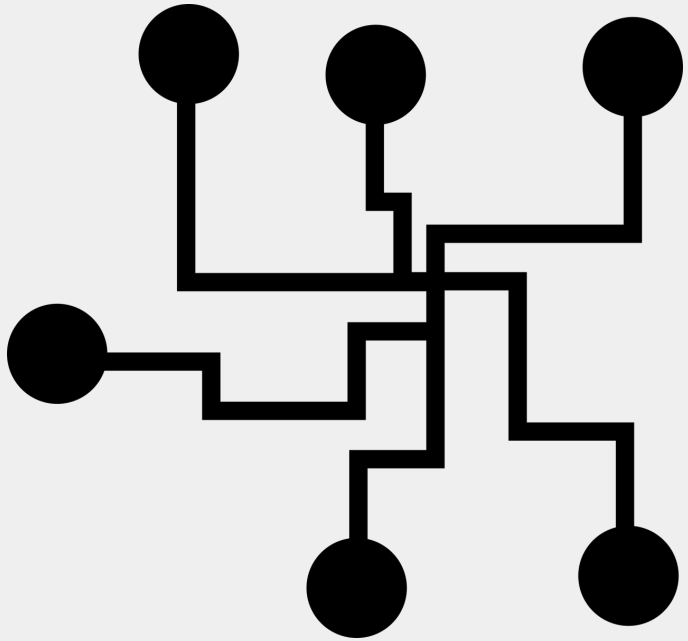
Katie Walker, Chesapeake Conservancy
kwalker@chesapeakeconservancy.org
(443) 345-5985

Overall Timeline

- Establish clear and timely “pencils down” deadlines for defining classification scheme and roll-ups
- External deadlines significantly impacted project timeline, requiring additional product versioning and rushed QC
- Presenting project updates to stakeholders when there are not decisions or requested feedback can distract from timeline
- Alter county boundary workflows to eliminate LC mosaic bottleneck
- Utilize previous LU data to test model revisions
- Better division of labor, with increased staffing, and increased redundancies in expertise

Local + Ancillary Data

- Create standardized metadata documentation for all ingested data when downloaded
- Mosaic local data at regional scale for uniformity and broad application
- Revisit workflow for re-classifying and using local data based on latest LULC schema
    - Host a workshop with localities to better prepare local data for ingestion into LC and LU models (i.e. Creating LU and zoning masks based on desired outcomes instead of utilizing data as is) - ?



Computation and Processing

- Revise all modules for consistency and efficiency
- Revisit distributed processing solutions
- Prioritize workflow design to simplify coding of the modules
    - Diagram modeling workflow with full team early in process for more transparency on what rules are made and when
    - Finalize class list, names, and lookup table for cleaner coding structure
    - Reduce redundant code/functionality with increased programming style rules and function calls over one way control transfers
- Revisit object-oriented coding solutions and data encapsulation methods
- Long-term goal to revisit some functions to process in raster format for improvements to processing time
- Improve logging for more efficient troubleshooting and planning

Data Prep

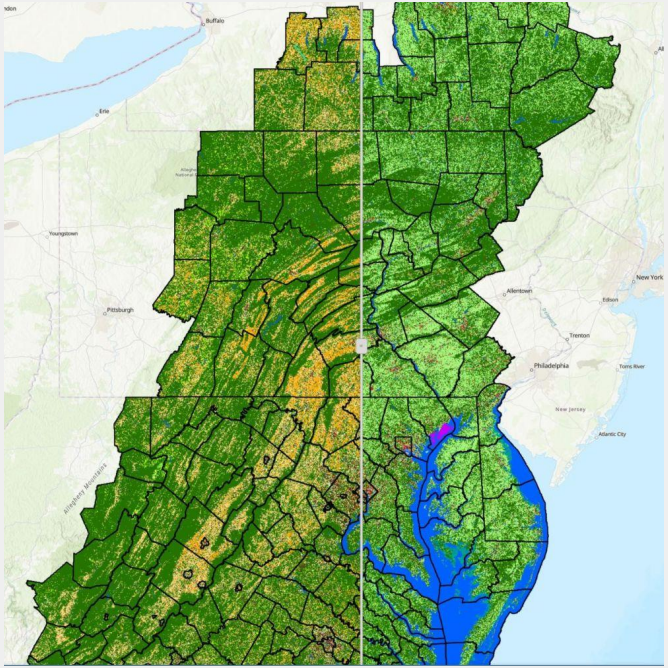
- If ANY input data changes, the model will need significant revision and testing; prepare as such
- Issues with inconsistent roads in parcel data, workflow needs to be revised for uniform application
- There are more data processing that can be added to data prep workflow to front load processing and reduce redundancies in later modules
- Need to re-evaluate the use of geopackages and SQL for improvements to processing time across all modules

Modeling Decisions

- Establish clear and timely “pencils down” deadlines for rule changes to allow adequate time for processing and QC
- Review implications of current classification scheme for known use cases (i.e. mapping pasture and hay together)
    - Solicit specific feedback from advisors, workgroups, and stakeholders for class improvements
- Consider integration of LC Change into LU production (i.e. tree canopy change to improve calls for harvested forest
- Host more collaborative working sessions to develop definitions and make rule change decisions to increase confidence
- Establish more dynamic, scientific thresholds for rule decisions to reduce qualitative decision points (i.e. explore use of random forest modeling)

Data Outputs + QAQC

- Establish a QAQC lead across all data products, not per product
    - Establish tangible and measurable targets for QC to reduce qualitative “thumbs up”
- Increase consistency in application of methods, calls, and fixes between counties (LC)
- Reduce high visibility errors (esp LC) by applying a more consistent QC process
- Improve specific flagged classes (i.e. barren LC, solar fields, wetlands, animal operations)
- Include additional time for iterative QC
- Validate outputs using quantifiable metrics from ancillary data (i.e. ag census, FIA, local zoning/LU)
- Plan for stakeholder QA of land use products for priority counties prior to full run



Data Hosting/Roll-out

- Standardized nomenclature for data products
  - Ensure model outputs align with nomenclature decisions to reduce post-processing
- Plan ahead for minimizing size and data type (bit depth)
- Develop comprehensive plan for version control and depreciating old versions of data
- Better estimate time needed for zipping/publishing
- Don’t migrate your physical server the same time you’re upgrading your Enterprise GIS system