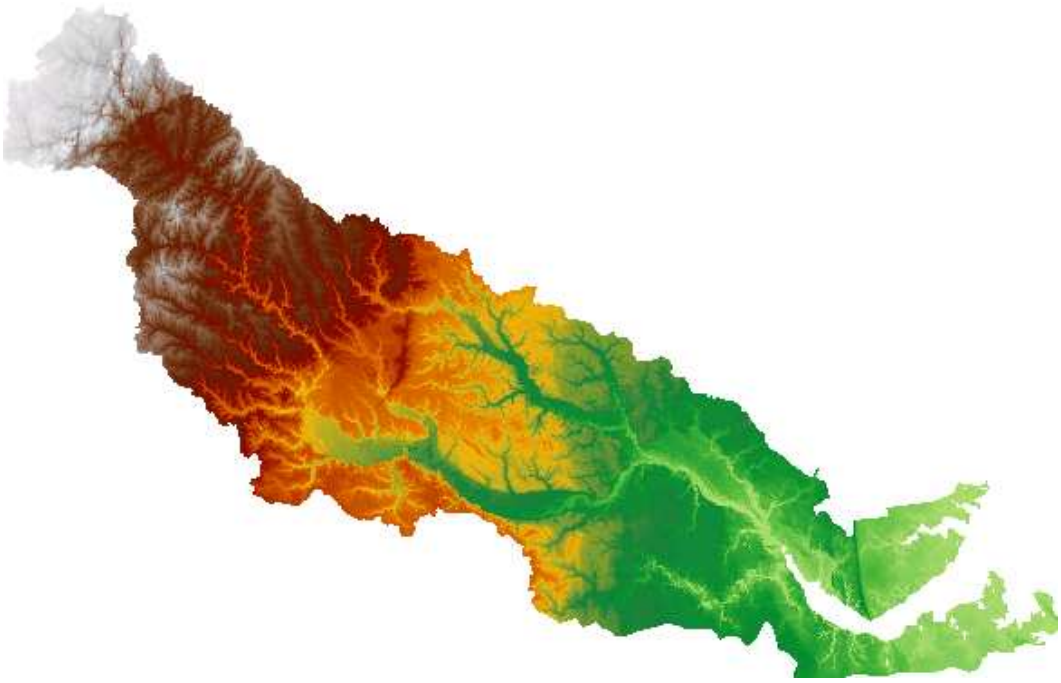


Figure 2.3: Neuse river basin data set

2.5.3 GRASS Implementation

The GRASS module `v.surf.rst` uses a quad-tree segmentation, but is not I/O-efficient in several key areas which we briefly discuss; constructing the quad tree, supporting a bit mask, finding neighbors, and evaluating grid cells. All data structures in the GRASS implementation with the exception of the output grid are stored in memory and must use considerably slower swap space on disk if internal memory is exhausted. During construction points are simply inserted into an internal memory quad tree using the incremental construction approach of Section 2.2. Thinning of points using the parameter ε during construction is implemented exactly as our implementation. The bit mask in `v.surf.rst` is stored as a regular grid entirely in memory and is accessed randomly during interpolation of segments instead of sequentially in our approach.

Points from neighboring quad-tree segment are not found in advance as in our algorithm, but are found when interpolating a given quad-tree segment q; the algorithm creates a window w by expanding q in all directions by a width δ and querying the quad tree to find all points within w. The width δ is adjusted by binary search until the number of points within w is between a user specified range $[n_{\min}, n_{\max}]$. Once an appropriate number of points is found for a quad-tree segment q, the grid cells in q are interpolated and written directly to the proper location in the output grid by randomly seeking to the appropriate file offset and writing the interpolated results. When each segment has a small number of cells, writing the values of the T output grid cells uses $O(T) \gg \text{sort}(T)$ I/Os. Our approach constructs the output grid using the significantly better $\text{sort}(T)$ I/Os.

2.6 Experiments

We ran a set of experiments using our I/O-efficient implementation of our algorithm and compared our results to prior GIS tools. We begin by describing the data sets on which we ran the experiments,

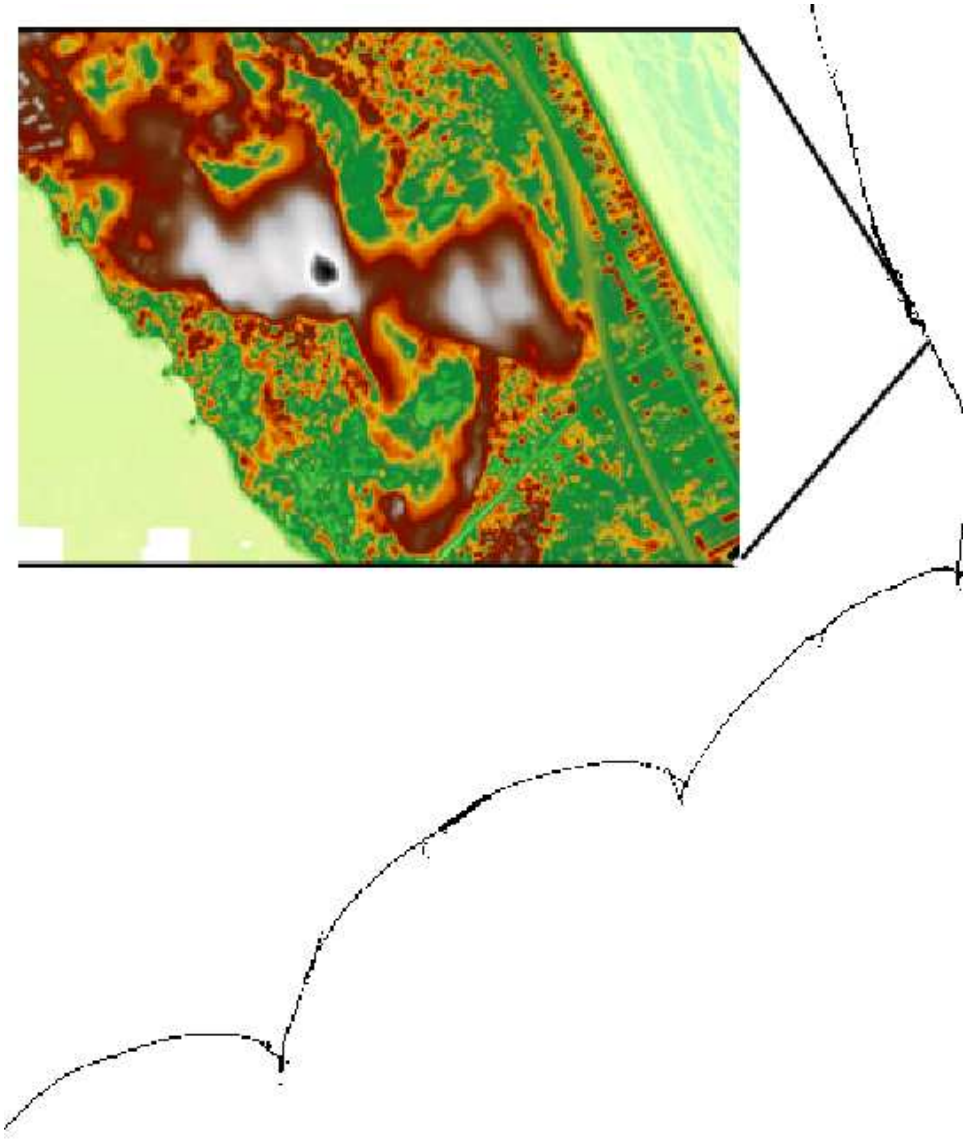


Figure 2.4: Outer Banks data set, with zoom to very small region.

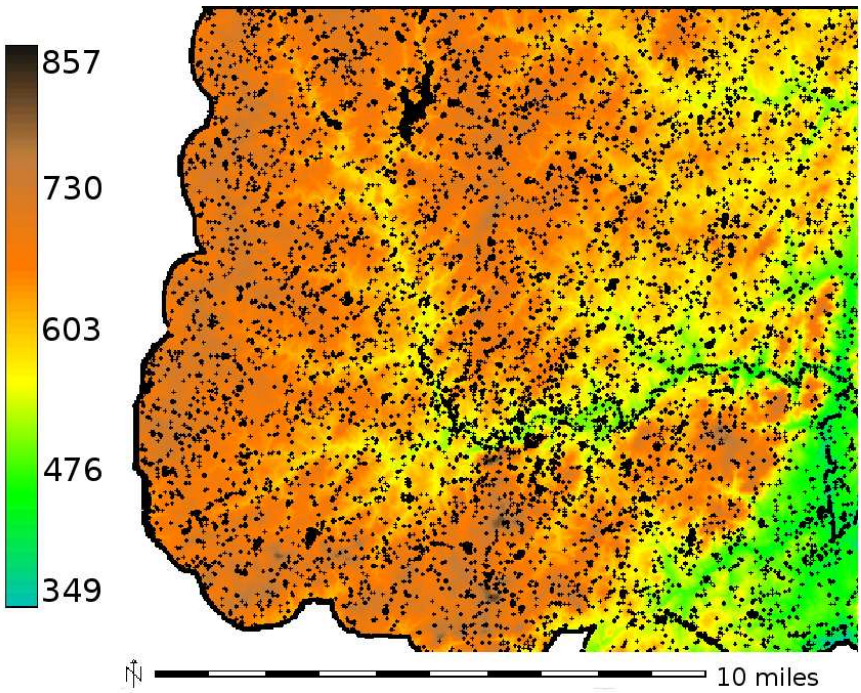
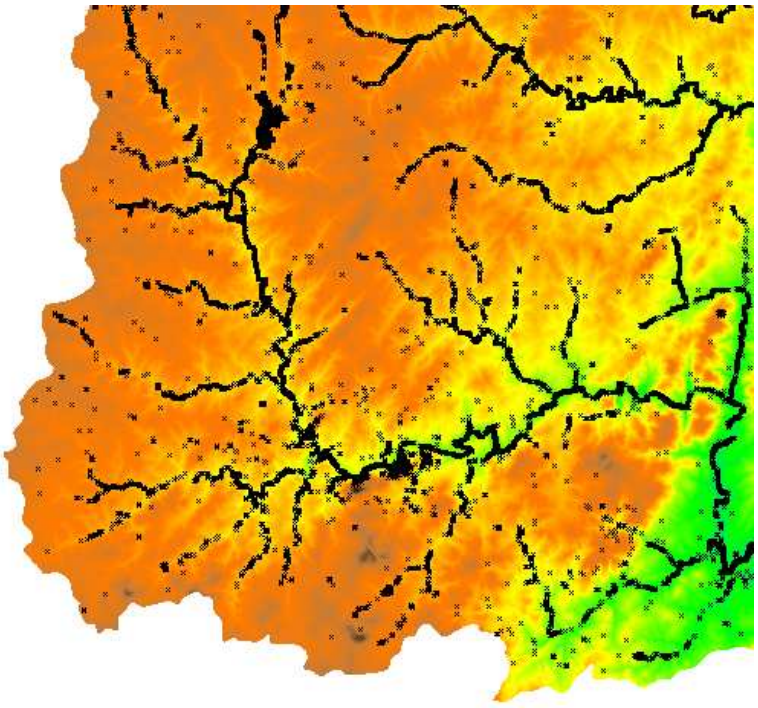


Figure 2.6: Interpolated surface generated by our method. Black dots indicate cells where the deviation between our method and `v.surf.rst` is greater than three inches.

DEMs are often constructed from multiple sources, including lidar points and supplemental breaklines where feature preservation is important. Future work will examine methods of incorporating multiple data sources into DEM construction. Finally, the ability to create large scale DEMs efficiently from lidar data could lead to further improvements in topographic analysis including such problems as modelling surface water flow or detecting topographic change in time series data.



(a)

Figure 2.7: Interpolated surface generated by our method. Black dots indicate cells where the deviation between our method and ncfloodmap data is greater than two feet.

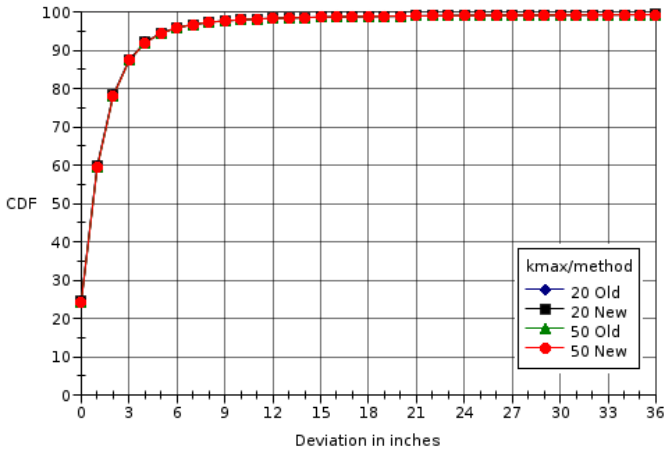
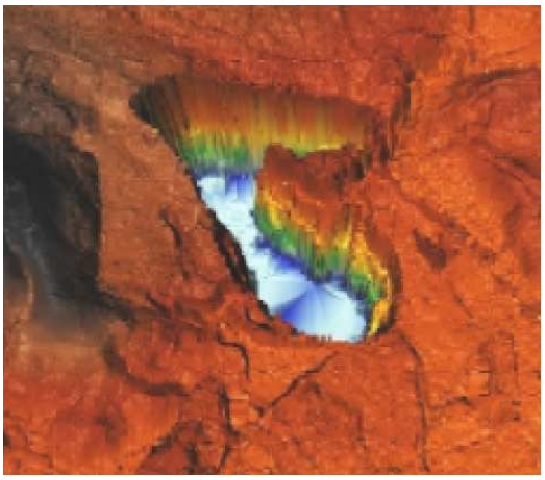
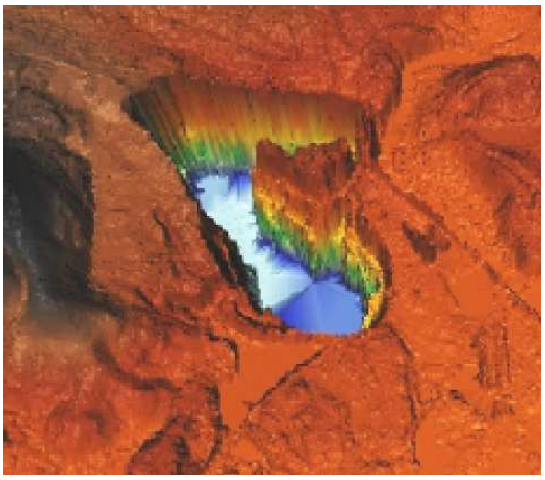


Figure 2.8: Cumulative distribution of deviation between interpolated surface and data downloaded from ncfloodmaps.com. Deviation is similar for both our method and `v.surf.rst` for all values of $k_{\max}$.



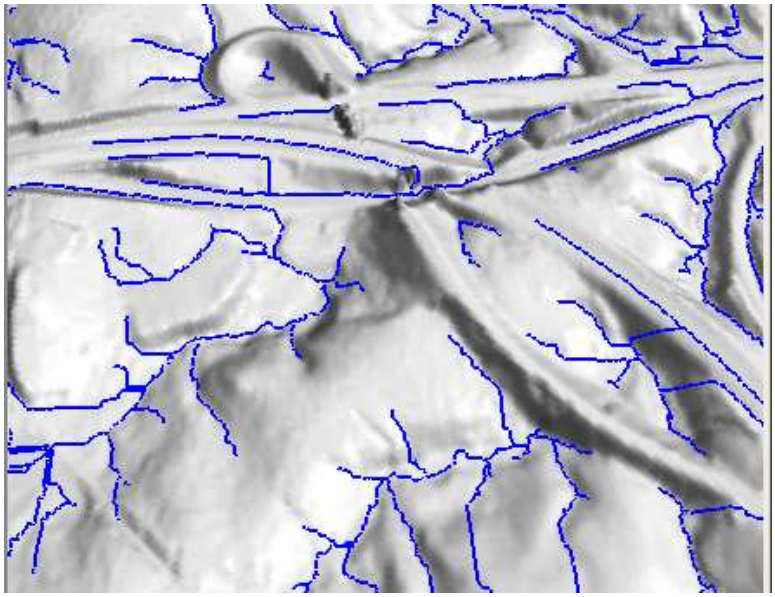
(a)



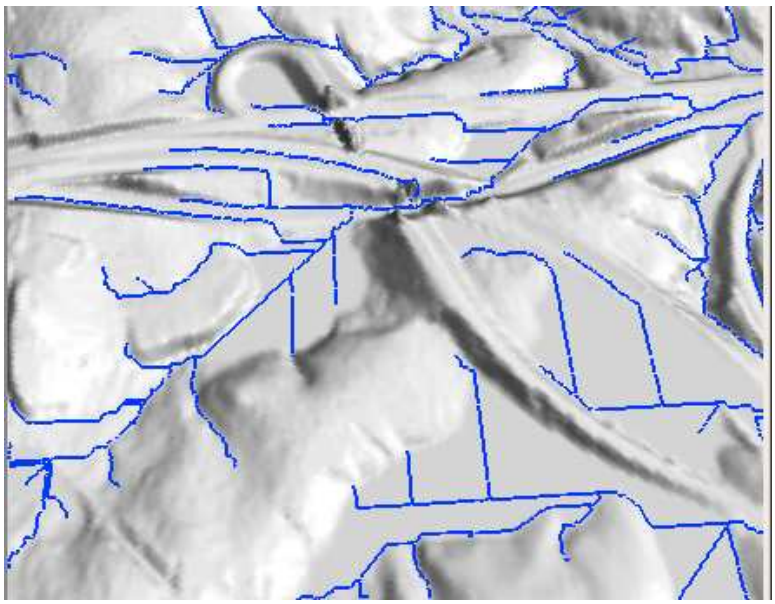
(b)

(c)

Figure 3.4: (a) Original terrain. (b) Terrain flooded with persistence threshold $\tau = 30$. (c) Terrain flooded with $\tau = \infty$.

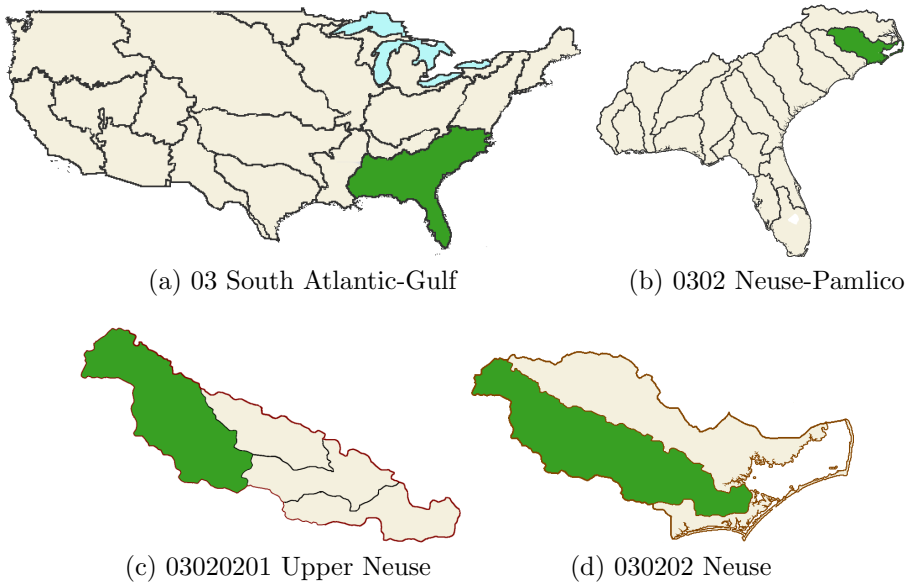


(a)



(b)

Figure 3.9: (a) Terrain and flow graph edges shown in blue with flooding of only low persistence sinks (b) Terrain and flow graph edges with flooding of all sinks.



(a) 03 South Atlantic-Gulf (b) 0302 Neuse-Pamlico

(c) 03020201 Upper Neuse (d) 030202 Neuse

Figure 4.1: A region, sub-region, basin and sub-basin in the USGS Hydrologic Unit System.

elevation model. As the quality and resolution of digital elevation models improve, the published HUC boundaries may not exactly match the boundaries suggested by the data. Second, HUCs at the sub-basin level may be too large for a particular application. Further sub-levels are in development but are not complete at this time. Third, HUCs are only available for the United States. Other countries and organizations have other coding methods [90]. Finally, the digits chosen for a particular HUC are, for the most part, arbitrary. Given two HUCs, it is often difficult or impossible to determine if water from one HUC flows into the other based on their numbering alone. Because finding the hydrological units upstream and downstream from a given location is a common task, a numbering scheme that allows a computer or user to relate hydrological units, without the need for visual inspection, would be helpful.

4.1.2 Introduction to Pfafstetter labels

The Pfafstetter labeling method described by Verdin and Verdin [90] addresses several disadvantages of the USGS Hydrologic Unit System. As mentioned earlier, the method can automatically divide a terrain into a hierarchy of arbitrarily small hydrological units, each with a unique label. Furthermore, Pfafstetter labels encode the basic topological connectivity of the hydrological units, allowing users to determine if one basin is upstream or downstream of another by examining the labels.

We present a conceptual definition of Pfafstetter labels here and will give a more formal definition in the context of grid DEMs in Section 4.2. Before defining Pfafstetter labels, we define a *river* $\mathcal{R}$ to be a directed path of monotonically non-increasing height over a terrain. The highest and lowest points on the river are the *source* and *mouth* (or *outlet*), respectively. The *basin* of $\mathcal{R}$ consists of the contiguous area of the terrain whose water flows, or *drains*, into the given river at a point between the source and mouth. All water in $\mathcal{R}$ eventually flows through the outlet. Within a basin corresponding to a river $\mathcal{R}$, other rivers exist that flow into $\mathcal{R}$. These rivers are called *tributaries* of $\mathcal{R}$, and the *confluence* of a river $\mathcal{R}$ and a tributary is the mouth of the tributary of $\mathcal{R}$, that is, the point where the tributary joins $\mathcal{R}$. Each tributary has a corresponding basin that is a sub-region of the basin of $\mathcal{R}$.

Given a river $\mathcal{R}$ along with its corresponding basin and tributaries, the Pfafstetter method [90]

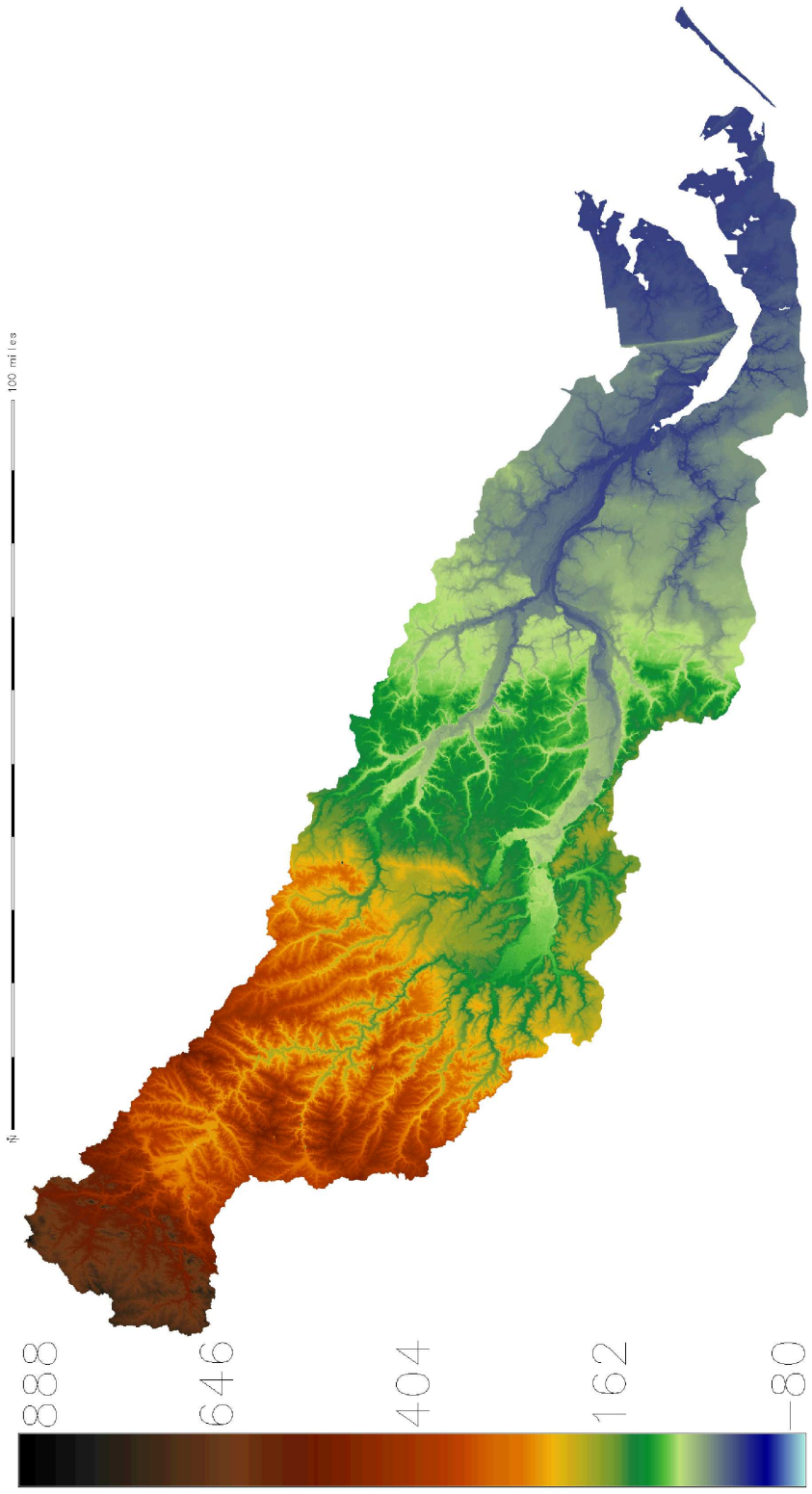
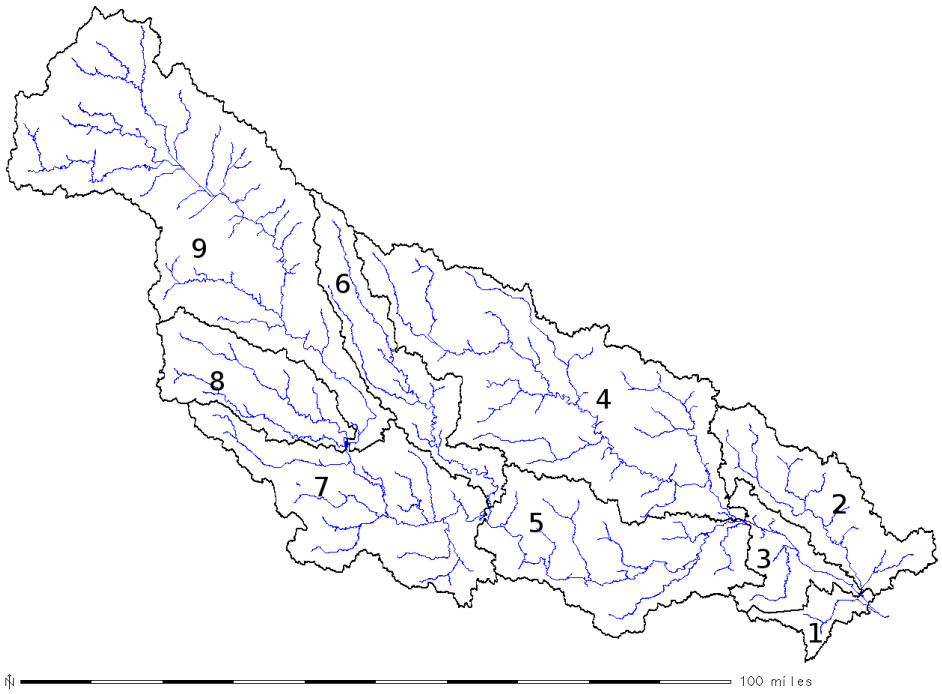
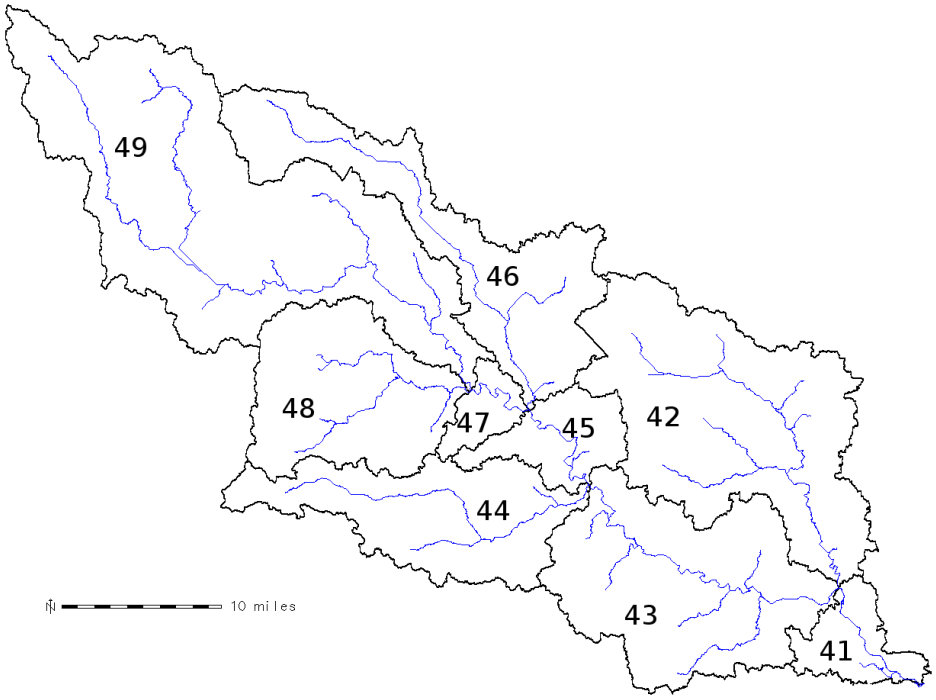


Figure 6.4: DEM of Neuse river basin derived from lidar points



(a)



(b)

Figure 6.5: (a) First level of Pfafstetter watershed labels for largest basin in Neuse. (b) Recursive decomposition of basin four.

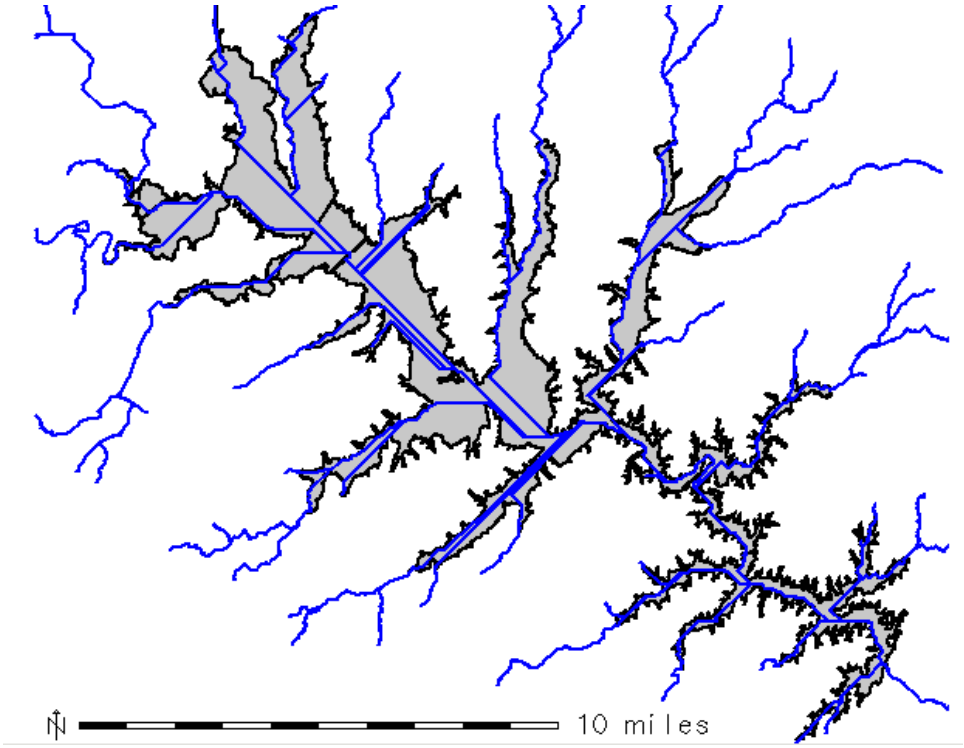


Figure 6.6: Falls lake, with a dam located near the Southeast corner of the figure. The boundary of the Falls lake flat is outlined in black while blue lines show rivers entering the reservoir and routed across the flat area.

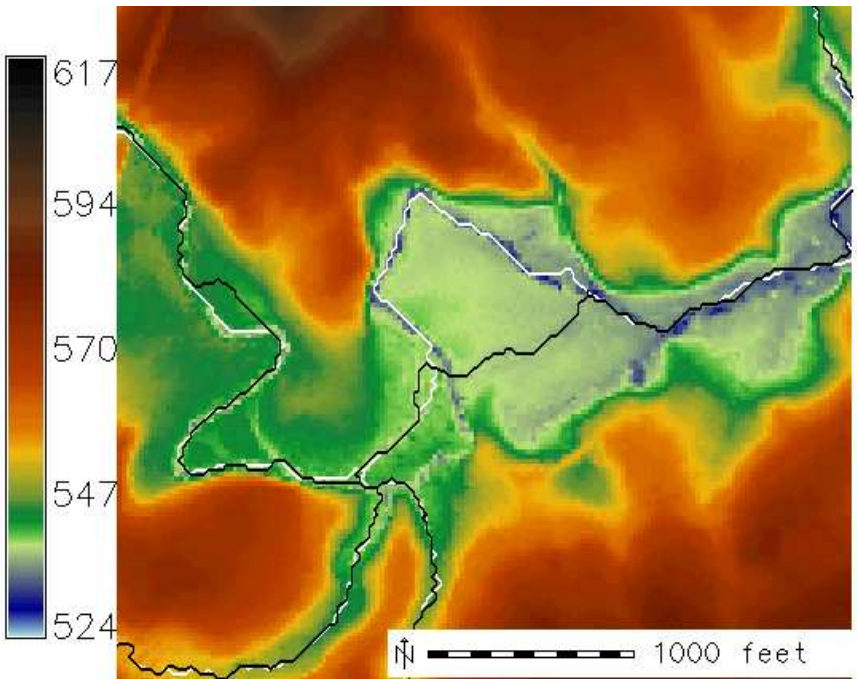
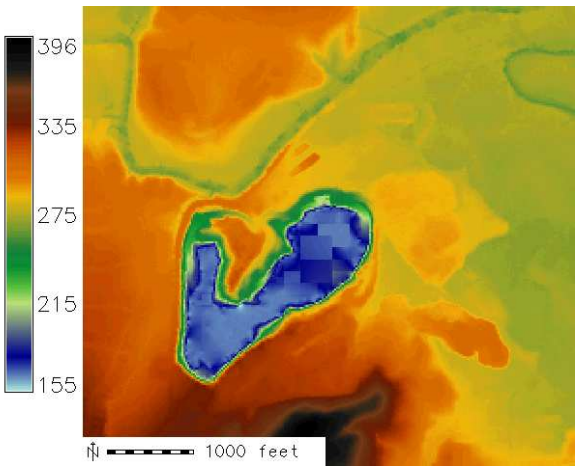


Figure 6.7: Rivers extracted using ε =10 ft (white) and 20 ft (black).

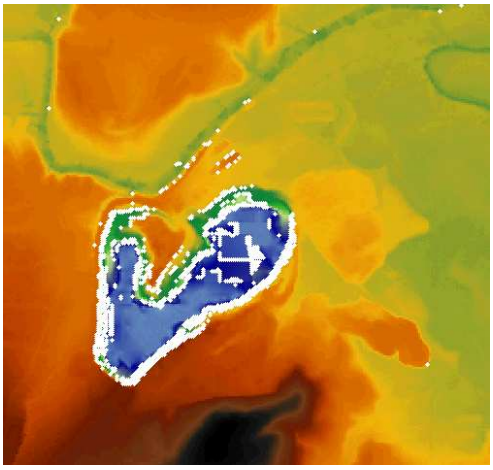
eter while using $k_{\max} = 8$ and ε =10ft. A smoothing parameter of 0 results in a interpolated surface that passes exactly through the input points. For a non-zero smoothing parameter, the algorithm constructs an approximation surface in which the input points can deviate from the constructed surface. The default smoothing parameter is 0.1. Smoothing only effects the interpolation routine, and not the quad-tree construction. By increasing the smoothing parameter, we can decrease the number of sinks in the constructed terrain. Our results are summarized in Table 6.5. We compute the RMS deviation by comparing the grids to the base grid with the default smoothing of 0.1. For a smoothing parameter of 5, the RMS deviation increased significantly. We also observed some strange blocky edges in the terrain with this high smoothing parameter that suggested that such high smoothing values should be avoided. While increasing smoothing can decrease the number of sinks somewhat, many sinks still remain even after significant smoothing. By looking at the persistence of the sinks created, we noted that increasing smoothing typically eliminates sinks with very small persistence while occasionally reducing the persistence of other sinks by a foot or less. Since smoothing did not remove larger sinks, smoothing had little effect on the hydrologically conditioned DEM, the river network, or watershed boundaries as the small differences in smoothing where negligible compared to the extensive flooding performed by the hydrological conditioning stage. Thus, we found that we could just use the default smoothing. For other terrain applications, such as topographic analysis, Mitasova et al. [68] describe the benefits of tuning the smoothing parameter.

smoothing	0	0.1	1	5
# sinks (thousands)	186.4	178.0	129.7	67.9
RMS deviation (ft)	0.0131	N/A	0.089	0.300

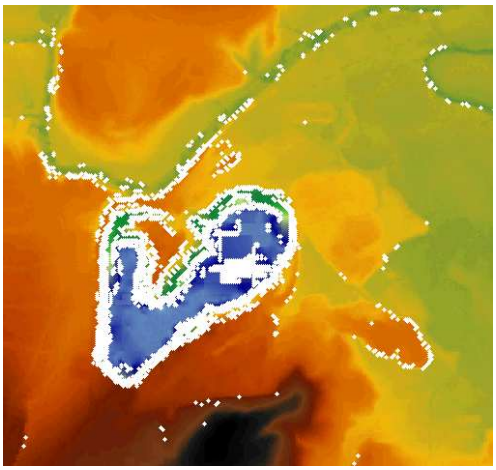
Table 6.5: Impact of smoothing parameter on number of sinks, and RMS deviation.



(a)

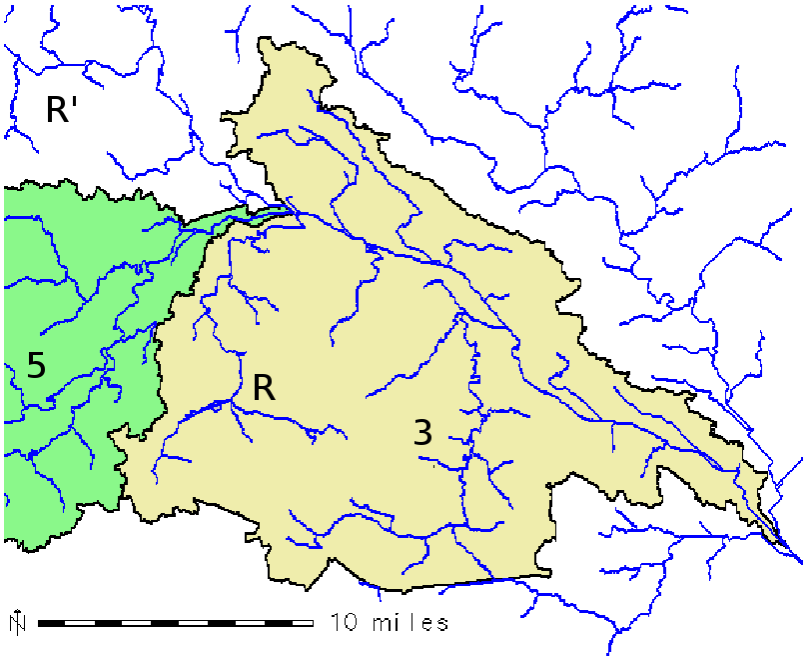


(b)

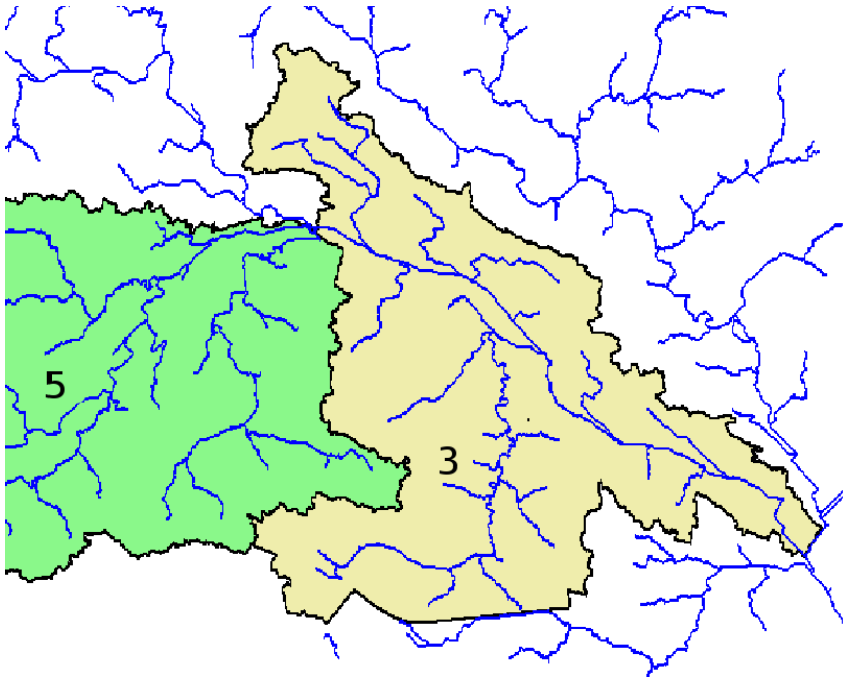


(c)

Figure 6.9: Spatial distribution of deviations from (a) 20ft grid elevations for (b) 10ft grid (c) 40ft grid. White points indicate spots where vertical elevation deviation exceeds 5ft.



(a)



(b)

Figure 6.10: Two watershed regions labeled by Pfafstetter algorithm have quite different boundaries in the (a) 10ft grid and (b) 40ft grid.

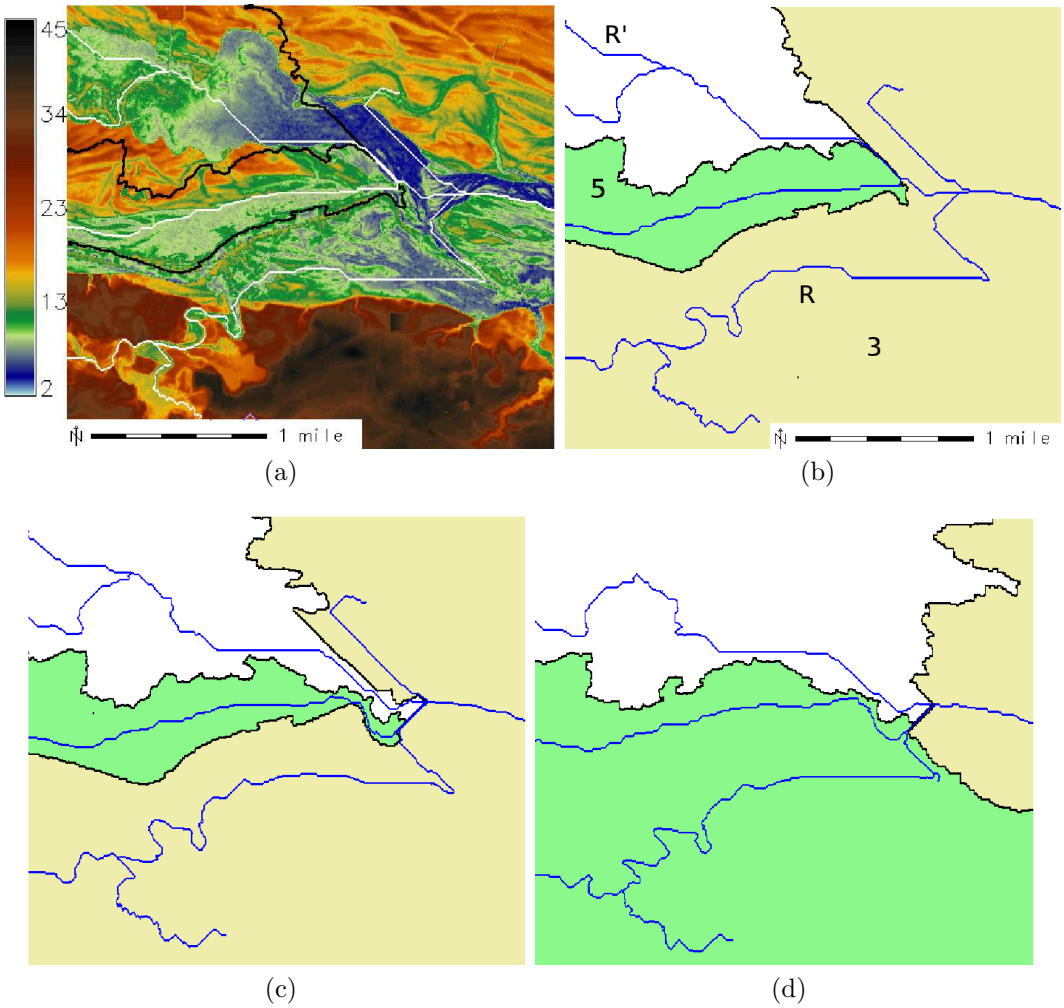


Figure 6.11: A detailed view of Figure 6.10. (a) Base terrain shown at 10ft resolution. Watershed boundaries at (b) 10ft, (c) 20ft, and (d) 40ft grid resolutions

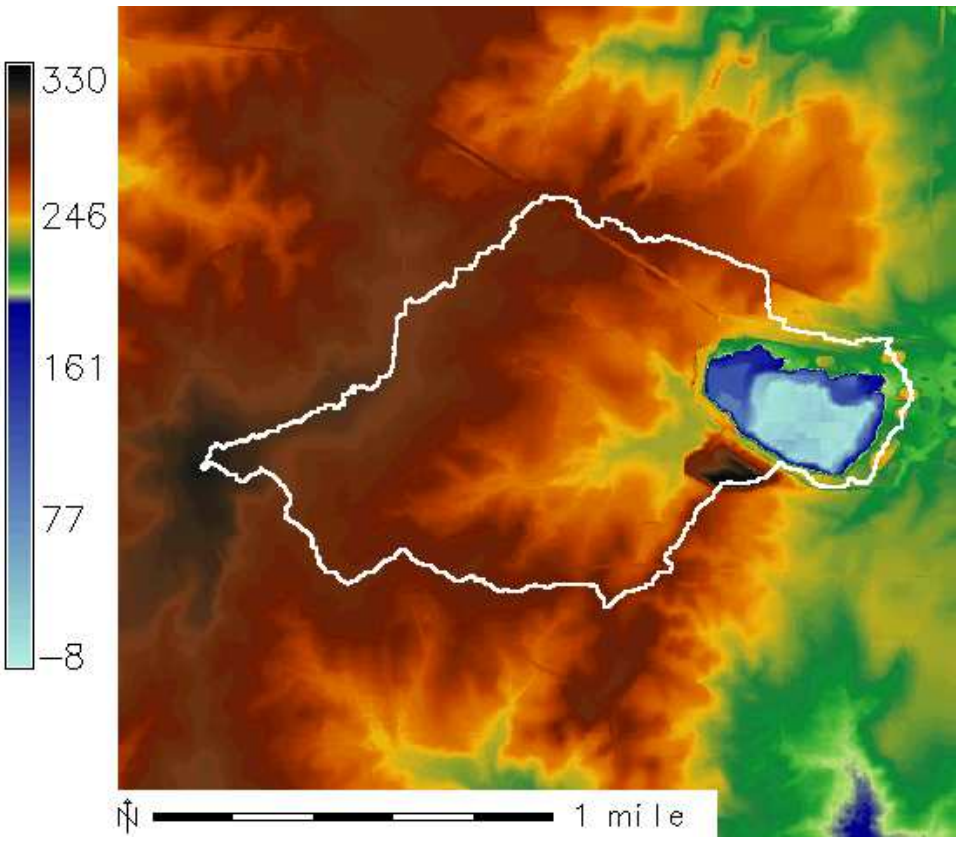


Figure 6.12: A quarry and its 600 acre watershed is preserved with a persistence threshold of 220 feet or less.

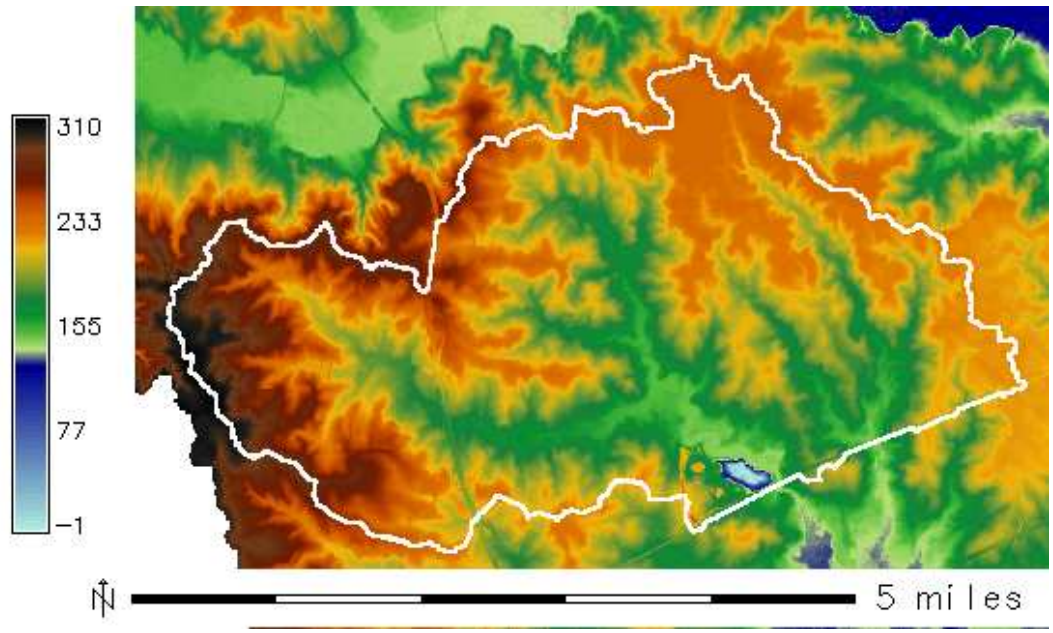
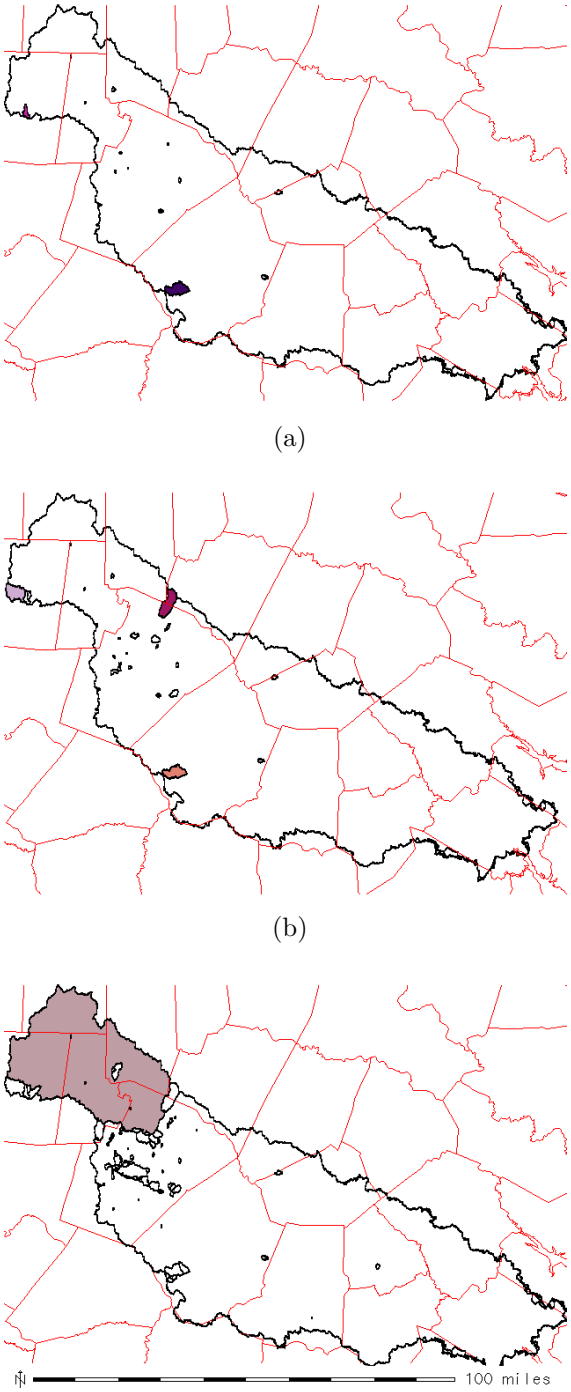


Figure 6.13: The largest (7300 acres) incorrectly computed closed basin, shown in white, for a persistence threshold of 50ft. A bridge in the southeast blocks flow.

As we lower the persistence threshold, more sinks are kept and fewer are removed. At a persistence threshold of 40ft, 13 additional sinks appear. The drainage area boundaries are shown in Figure 6.14(b). These additional sinks are all examples of small streams being blocked by bridges, but the drainage area of the additional sinks is small and does not dramatically effect the watershed boundaries or the river network. In particular, most of the Neuse river basin drains to a single outlet along the coast in the southeast corner of the figure. However, if we lower the persistence threshold to 30ft, we see dramatic changes in the number and drainage area of the preserved sinks as illustrated in Figure 6.14(c). The most obvious observation is that water upstream of the Falls lake dam, shown in pink shading, is disconnected from the rest of the basin. Also, many more minima appear, especially in urban areas such as Wake county, just South of the disconnected Falls lake basin. A brief inspection of a number of these sinks in Wake county revealed 62 total sinks, 47 of which were caused by bridges blocking rivers, 10 of which were around quarries and five whose source was not obvious.

Because a large portion of the Neuse river basin is detached from the main basin with a persistence threshold of 30ft, we expect the watershed boundaries to be significantly different for persistence thresholds of 50ft and 30ft. In Figure 6.15, we see that while basin 9 loses a significant fraction of its drainage area when lowering the persistence to 30ft, it is still larger than basin 8 and the ordering of the Pfafstetter basins (indicated by color) is unchanged. If a further downstream area lost a significant fraction of its total area by lowering the persistence threshold, re-ordering and re-labeling of basins would be much more likely.

If we look at the component that was cut off from the Neuse river basin in Figure 6.15, we can see by the watershed labels in Figure 6.16, that flow has been routed in an unrealistic way. Note that under the Pfafstetter label method, water in even numbered basins and basin 9 flow into lower numbered odd basins. The figure shows a region surrounded by even number basins plus basin nine. Thus, this region has no outlet to any other region. This is consistent with the observation that the region is disconnected from the main Neuse river basin in the terrain model. The odd numbered basins are tightly clustered in the center where a sink collects all of the water. If we looked closely



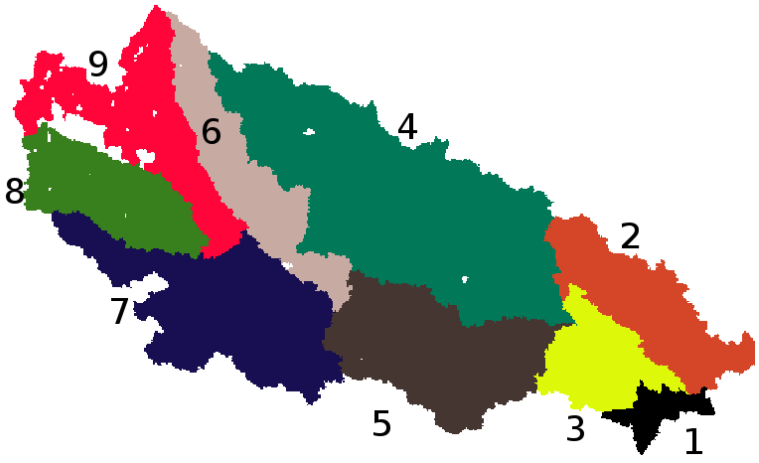
(a)

(b)

(c)

Figure 6.14: Drainage area boundaries of sinks shown in black with overlay of North Carolina county boundaries. A persistence threshold of (a) 50ft removes almost all sinks caused by bridges and creates one large primary basin. A threshold of (b) 40ft results in 28 remaining sinks, but the primary basin is intact. For a threshold of (c) 30 ft, the Neuse river basin becomes disconnected at the Falls lake dam (Northwest/shaded), and 96 sinks remain, most of which are due to bridges.

(a)



(b)

Figure 6.15: Pfafstetter basins for (a) persistence threshold of 50ft and (b) 30ft. Even though the headwaters are disconnected in the 30ft case, the ordering of the basin remains unchanged.

at the flow directions in this basin, we would see flow from basin 4 being directed Northwest towards the center, when in the real terrain water flows Southwest towards the Falls lake dam. Thus lowering the persistence below 30ft will not yield a good hydrologically conditioned DEM.

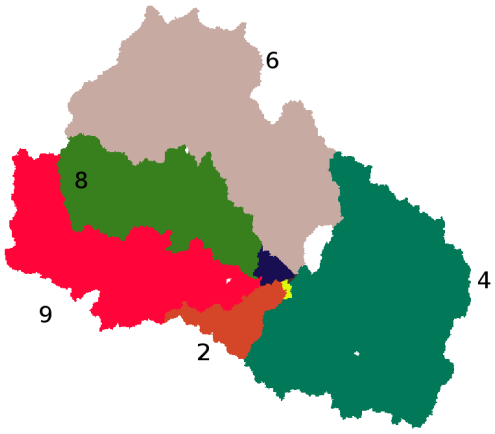


Figure 6.16: Watershed of Falls Lake area when persistence threshold is 20ft. Rivers computed in the southeast region eventually drain to a sink in the center of the image, instead of flowing under the dam which is to the the southeast

In the three persistence values tested in this section, we found that persistence can indeed be used to preserve real terrain features such as quarries, but that many bridges and an occasional dam create sinks with a moderately high persistence that should be removed. The gap in persistence values between the bridge with the highest persistence and the quarry with the lowest persistence is over 20ft. Thus, a persistence threshold of 55ft in this case study preserves all the quarries while routing flow across bridges. This new method of scoring and removing sinks below a threshold score could prove to be a valuable tool for many hydrological studies.

6.8 Conclusions

In this Chapter, we demonstrated that the algorithms presented in this thesis form a scalable and flexible pipeline that efficiently process massive amounts of data derived from modern remote sensing methods. Our primary emphasis in this thesis was on scalable algorithms, but we have seen that our tunable design allows us to explore interesting modeling issues as well. While lidar provides many potential benefits to the GIS community, our experiments highlighted the need for additional work in some areas. Bridges are particularly problematic for hydrological flow routing. We have seen in this Chapter that the topological persistence of most sinks blocked by bridges have a high persistence value, but this value is much lower than features such as quarries. We believe further improvements in th GIS modeling using topological persistence can help identify bridges effectively and lead to improved hydrological conditioning models that can automatically make small local cuts through bridges. This will significantly reduce the extent of terrain modification via flooding and dramatically reduce the size of flat areas.

Bridge removal is also important for accurate watershed extraction. As discussed in Section 6.6, subtle changes in the order of river mouths joining a main channel can significantly change watershed boundaries computed using the Pfafstetter method. Often times, odd flow routing paths are the result of poor flat routing models on areas that have been hydrologically conditioned by flooding sinks. Flooding sinks caused by bridges in not an ideal approach, but future work in bridge removal, hydrological conditioning alternatives to flooding, and improved flow routing on flat areas could help