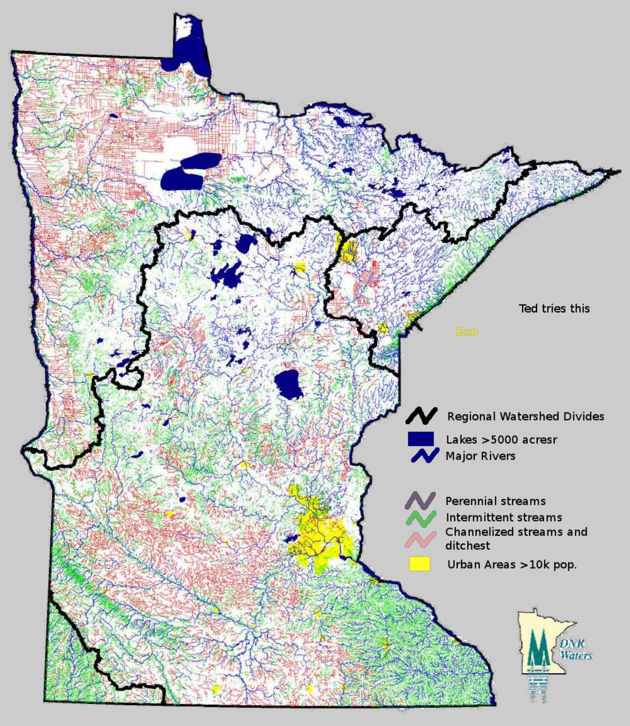
Efficient Algorithms for Geographic Watershed Analysis

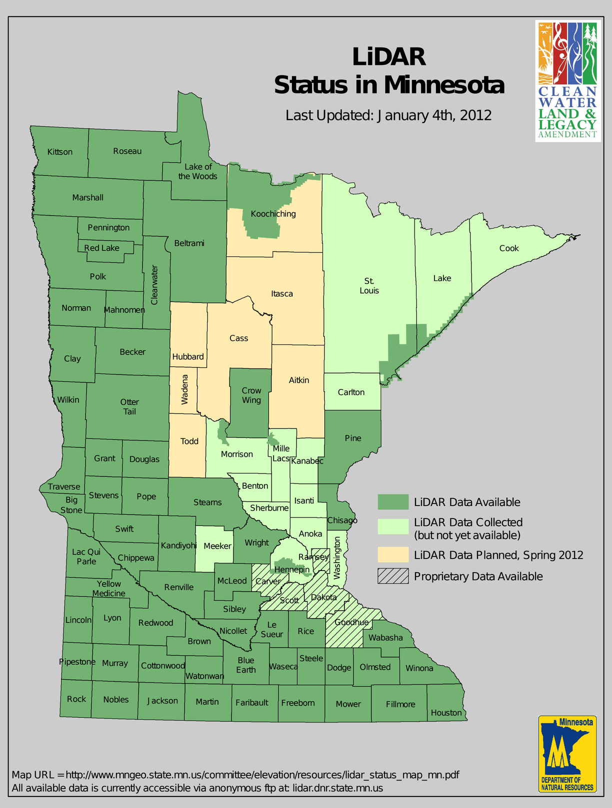
Richard Barnes, Clarence Lehman, David Mulla, Jacob Galzki, Haibo Wan, and Joel Nelson

Computation provided by the Minnesota Supercomputing Institute



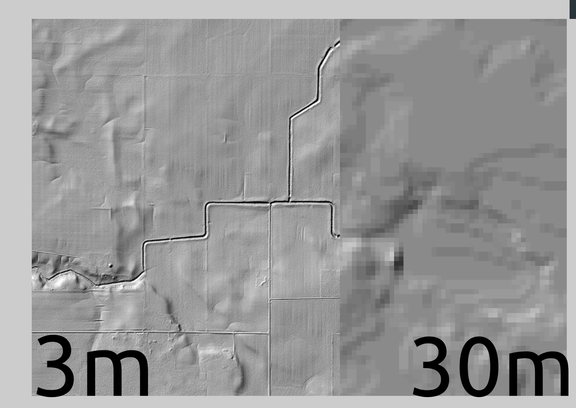
Minnesota has a vast natural watershed of lakes, ponds, rivers, and streams. Interwoven is a large artificial watershed of drain tiles and ditches. Just as household plants need drains in their pots, so agricultural crops need drain tiles to keep their fields from getting too damp. In today's world the drain-tiles carry not only water but also fertilizers, antibiotics, pesticides, hormones, and other chemicals into the natural watershed.

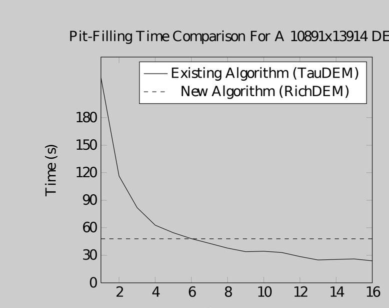
This project is to analyze where wetlands and other vegetated buffers can be placed on the landscape to intercept drain waters and help purify them before they reach the natural watershed. The computational problem comes because new LIDAR images have expanded the resolution of geographic digital elevation models (DEMs) up to a thousandfold or more. This in turn has taxed the ability of existing algorithms to process the expanded datasets. Here we explain the project and present new efficient algorithms for parallel and scalar processing that reduce run-times from days on ordinary computers to minutes or seconds using the new algorithms in a parallel supercomputing environment.



DEMs are two-dimensional floating-point arrays representing terrain elevation, the data for which is generally acquired via LIDAR. DEMs are used extensively to model hydrologic processes and properties including soil moisture, terrain instability, erosion, stream power, and flooding. These properties are derived from the slope, flow direction, and catchment area metrics of each cell of the DEM.

The DEM's resolution limits the accuracy of these metrics and any property derived from them. Resolutions have grown from 30m in the recent past to <1m resolutions today. This has led to increased data sizes: 45GB for a 3m DEM of one-third of Minnesota. While computer performance has increased appreciably, legacy equipment and inefficient algorithms make processing these improved data sources costly, if not impossible.

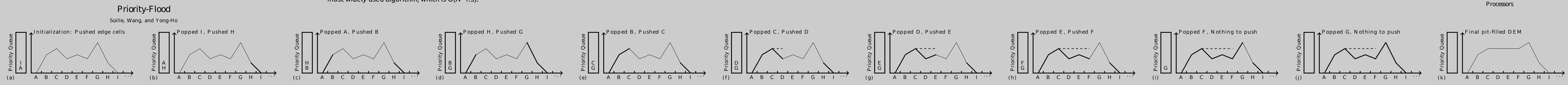




Faster Pit-filling

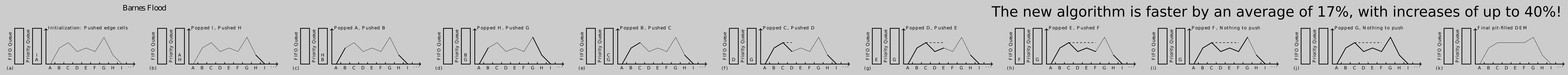
Pits are inwardly-draining regions of the DEM which have no outlet. Sometimes representative of natural terrain, they may also result from technical issues in the DEM's collection and processing, such as from biased terrain reflectance or conversions from floating-point to integer precision. Pits may be resolved either by breaching or by filling. Here, an existing O(n log n) algorithm is improved to O(m log m), where m is less than or equal to n, and both the new and old algorithms are compared against the most widely-used algorithm, which is O(N^1.5).

Priority-Flood
Soille, Wang, and Yong-Ho



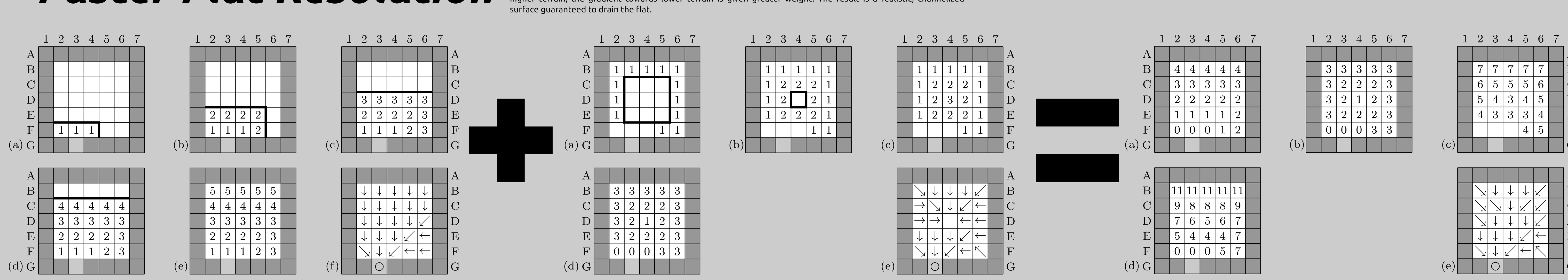
Barnes Flood

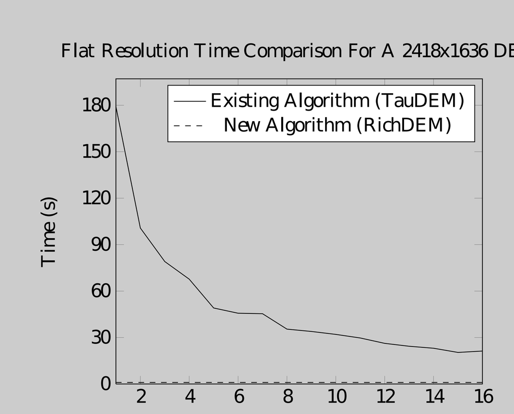
The new algorithm is faster by an average of 17%, with increases of up to 40%!



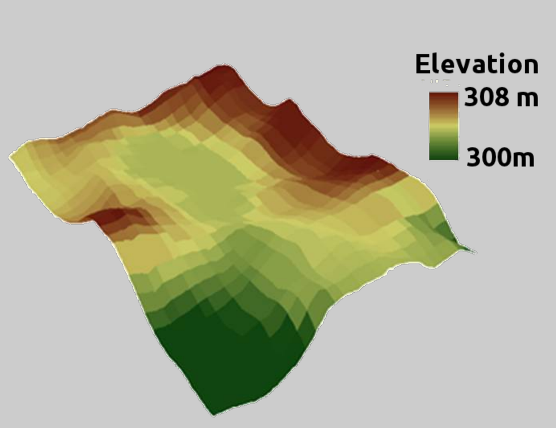
Faster Flat Resolution

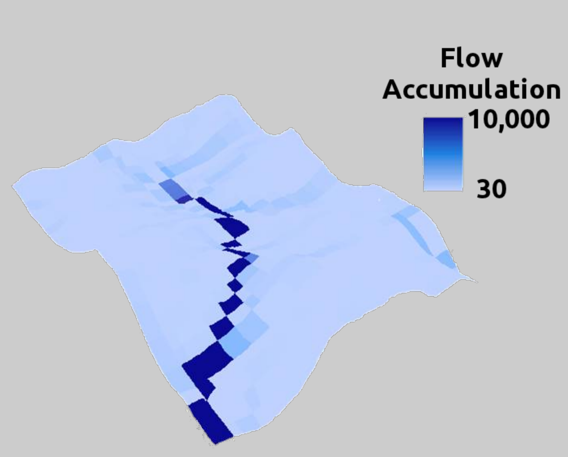
Unfortunately, pit-filling often results in flats. These are regions of the DEM where every cell is at the same level, preventing the definition of a flow direction. Here, we present an O(N) replacement for an older O(N^2) algorithm by Garbrecht and Martz. The method superimposes a gradient towards lower terrain with a gradient away from higher terrain; the gradient towards lower terrain is given greater weight. The result is a realistic, channelized surface guaranteed to drain the flat.





Faster Results



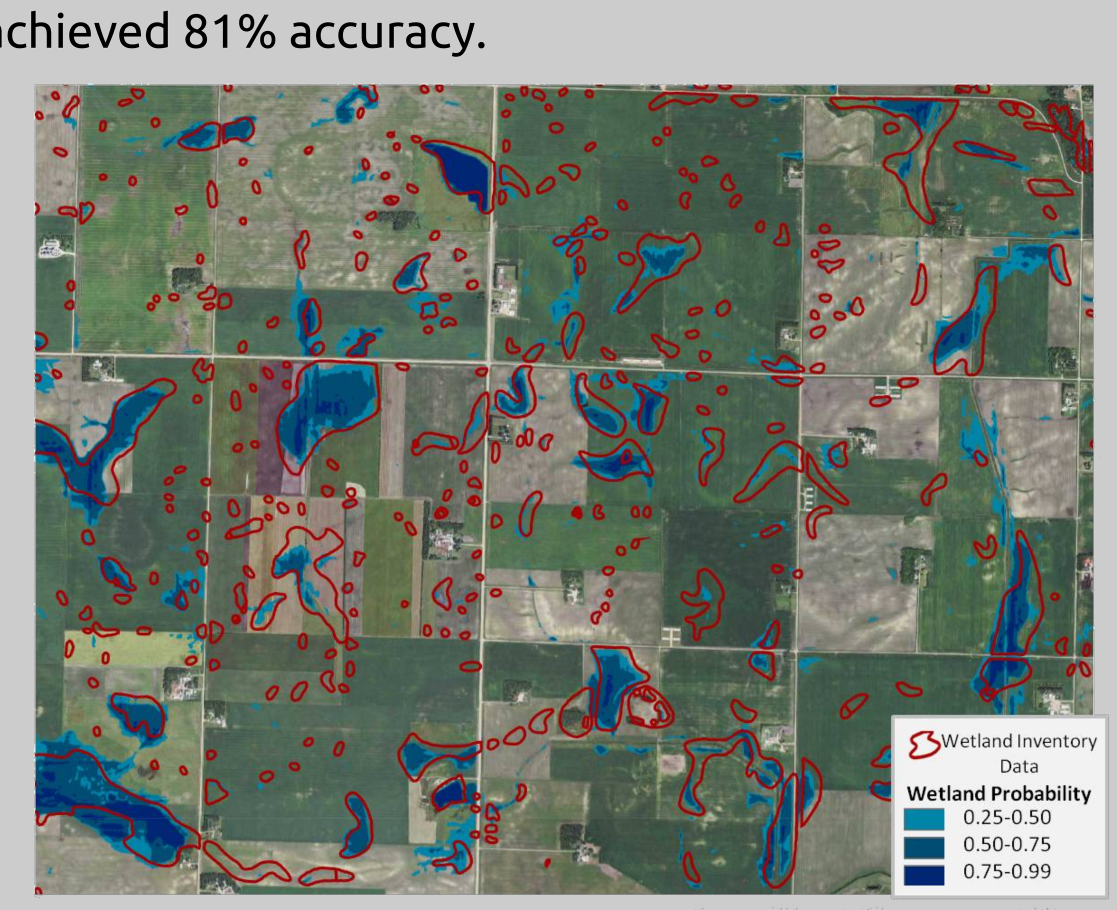


Terrain attributes (compound topographic index, hydric soil presence, slope, and profile curvature) are used in a linear regression model...

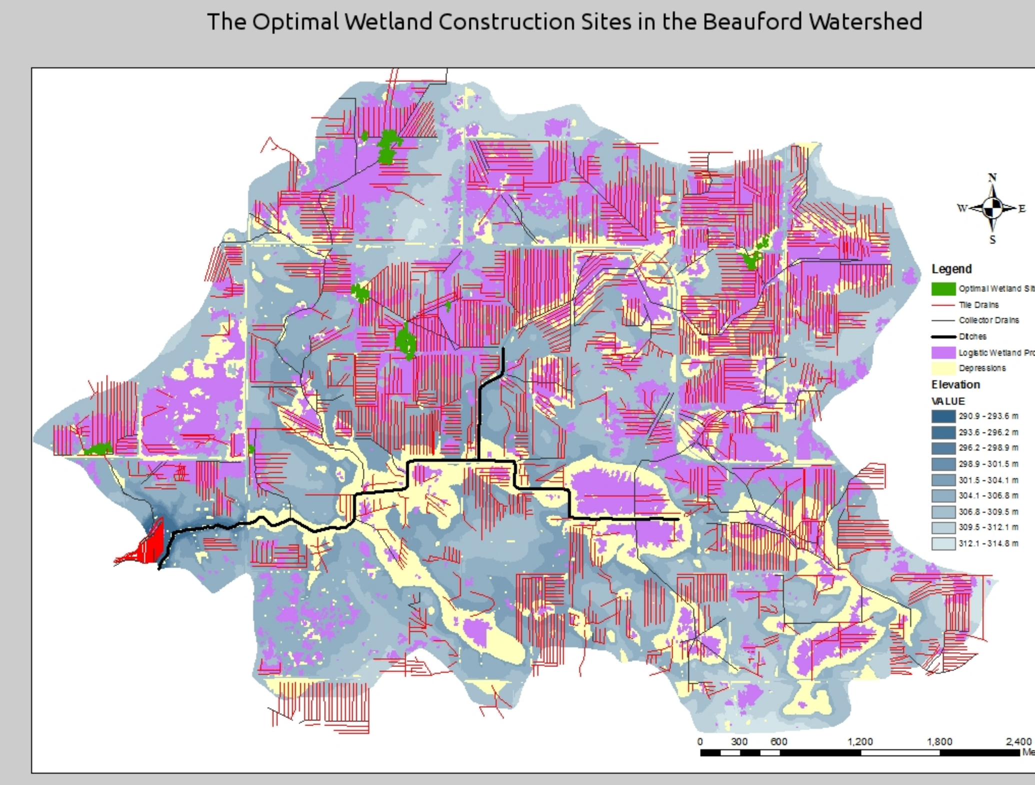
$$P = \frac{1}{1+e^{-z}}$$

$$z = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_3) + \ldots$$

...to determine the locations of wetlands, both filled and unfilled. The beta factors are trained using known wetlands and a threshold chosen on P. The process has achieved 81% accuracy.



The Optimal Wetland Construction Sites in the Beauford Watershed



If an identified wetland is (1) within a depression, has (2) appropriate properties (hydric soil, soil wetness, slope, profile curvature and so on) defined by a logistical model, (3) there are few tile drains within the site, (4) the site is close to a collector drain or ditch, and (5) the site is large enough for 14 days' storage. Then it is remediable!