Preprint:

A Robust Cellular Automaton Surface Drainage Preprocessing Algorithm for High Resolution Digital Elevation Models

Robert N. Eli, Ph.D., P.E., M.ASCE[1]

[1]Natural Resource Analysis Center, 4203 Agricultural Science Building, West Virginia University, Morgantown, WV 26506; Email: robert.eli@mail.wvu.edu

Published in:

Proceedings of World Environmental and Water Resources Congress 2025, held in Anchorage, Alaska, May 18–21, 2025. Sponsored by the Environmental and Water Resources Institute of ASCE.

Citation:

Eli, R. N. 2025. "A Robust Cellular Automaton Surface Drainage Preprocessing Algorithm for High Resolution Digital Elevation Models." 131–143. American Society of Civil Engineers. https://doi.org/10.1061/9780784486184.014.

A Robust Cellular Automaton Surface Drainage Preprocessing Algorithm for High Resolution Digital Elevation Models

Robert N. Eli, Ph.D., P.E., M.ASCE[1]

[1]Natural Resource Analysis Center, 4203 Agricultural Science Building, West Virginia University, Morgantown, WV 26506; Email: robert.eli@mail.wvu.edu

ABSTRACT

A robust cellular automaton (CA) algorithm has been developed that overcomes many of the difficulties in preprocessing high resolution raster format digital elevation models (DEMs) prior to creating the surface drainage network. The algorithm is unique in that a small fixed hydraulic gradient is generated using a parallel processing computation that progressively fills depressions and drains flat surfaces. The algorithm facilitates the location and inclusion of subsurface drainage infrastructure (e.g., culverts) and the re-establishment of building structures to the DEM via outside databases, such as Microsoft Building Footprints. The CA algorithm uses an iterative arithmetic computation method that never requires searching beyond the 3 by 3 raster cell neighborhood, which guarantees a robust algorithm that can establish a minimum hydraulic gradient under any condition, including flat surfaces, without failure. The algorithm and associated processing procedures are demonstrated on a small watershed with significant suburban development.

INTRODUCTION

A Digital Elevation Model (DEM) is a representation of the bare ground (bare earth) topographic surface of the Earth excluding trees, buildings, and any other surface objects ("What is a digital elevation model (DEM)? | U.S. Geological Survey" n.d.). LiDAR derived high resolution (1-meter grid resolution) bare earth DEM data sets are currently available over most of the continental US ("3D Elevation Program | U.S. Geological Survey" n.d.). These latter data sets can be downloaded in 10,000-meter square tiles to cover map areas of interest to the user. Elevations are in meters with a specified vertical accuracy of 10 cm RMSE ("Topographic Data Quality Levels (QLs) | U.S. Geological Survey" n.d.). The DEM data sets are GeoTIFF files which have a raster image format (with file extension: .tif) and include geographic metadata that locates the DEM in map space ("OGC GeoTIFF" n.d.). The map projection used for these DEM data is UTM ("Universal Transverse Mercator coordinate system" 2024). Each grid cell has an elevation in meters with coordinate locations also given in meters. As shown in Figure 1, raster DEMs are characterized by a square grid of cells in a two-dimensional (2-D) array of rows and columns. Conveniently, each grid cell elevation value can be accessed by its row number "n" and column number "m" when the DEM is stored as a 2-D array in computer memory.

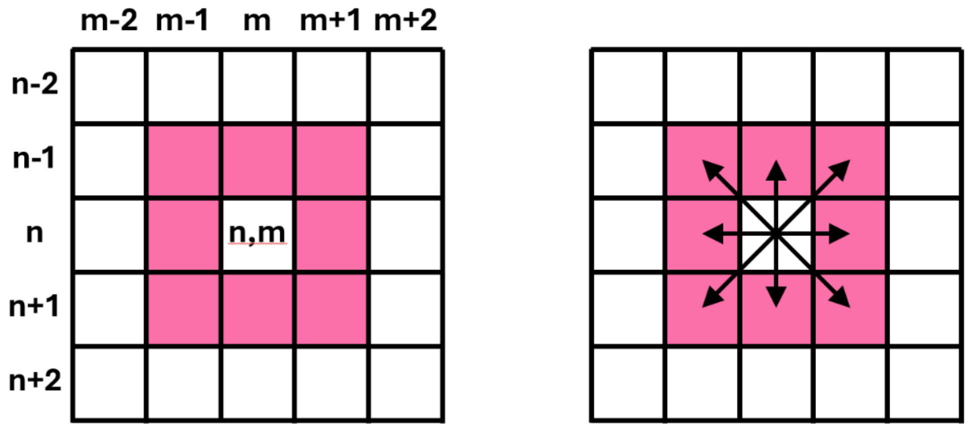

Figure 1. Illustration of the D8 grid cell neighborhood (red cells) and the eight possible flow directions from a focus cell (n,m) within a raster DEM.

The current wide availability of high resolution DEMs over the continental US is a rich spatial data resource for hydrologic modelers and hydraulic engineers interested in the design and performance analysis of drainage infrastructure. There is a long history of hydrologists utilizing DEMs to define surface drainage networks, beginning over 40 years ago. However, there is a big difference between what surface features are resolved using a 1-meter raster grid cell size as compared to the 5- and 10-meter grid cell sizes that were available years ago. When viewed as an image using GIS software, the richness of detail in the LiDAR-derived 1-m DEMs is readily apparent. Roadways, bridges, highway fills, stormwater drainage ponds, and other anthropogenic changes can be clearly identified, along with small first order stream channels. Building structures and larger bridges are removed during processing of the LiDAR point clouds to produce the raster 1-m DEMs. It is unfortunate that building structures are removed and replaced by fictitious surfaces since those can be a significant portion of the surface area of suburban watersheds and can impact hydrologic responses. Vegetation removal is more in keeping with the needs of modelers since surface elevations are more important and vegetated areas can be restored, if needed, using other data sets. Building structures can also be restored using building footprint data sets but those data can contain errors or omissions. The purpose of this paper is to present the details and performance of a new tool to assist the construction of fully connected drainage networks when using 1-m DEMs in moderate to high relief topography. Specifically, the focus will be on resolving barriers to flow connectivity that are made more numerous by the higher resolution DEM.

RESOLUTION OF PITS, DEPRESSIONS, AND FLAT AREAS

Pits, depressions, and flat areas may be physically present on the earth's surface but can also occur because of the lack of accuracy or resolution in the DEM data. In this study the notation used will be as follows: if the focus cell (n,m) in figure 1 has an elevation lower than the 8 surrounding neighbor cells, then it is labeled a "pit"; and if the focus cell (n,m) is equal in elevation to one or more of the 8 surrounding neighbor cells, but lower than the rest, it is labeled a "flat". Pits and flats may also exist within a "depression", which are larger areas of contiguous lower elevations that have no drainage outlet. Additionally, a LiDAR data processing procedure called "hydro-flattening" is typically used to interpolate elevations used for water surfaces based on bank elevations at the water's edge ("Elevation-Derived Hydrography Data Acquisition Specifications: Positional Assessment Requirements | U.S. Geological Survey" n.d.). Hydro-flattening is necessary since LiDAR wavelengths used for terrestrial mapping are scattered or absorbed by water, yielding little to no usable data over water surfaces. Hydro-flattening can introduce large flat areas in the DEM that drainage network processing algorithms must be capable of handling.

Drainage Network Algorithms. The D8 raster DEM drainage network algorithm was first introduced by (O'Callaghan and Mark 1984). It is the simplest method of defining surface flow direction and has been investigated by many others, including (Band 1986), (Jenson and Domingue 1988), (Fairfield and Leymarie 1991), (Martz and Garbrecht 1992). As shown in figure 1, the D8 algorithm allocates all surface water outflow from each individual grid cell (n,m) to one of the 8 neighboring grid cells. The direction selected is based on the direction of the steepest downslope. The D8 algorithm requires that the DEM be preprocessed in some manner to deal with any pits, depressions, and flat areas in the DEM that would otherwise prevent a complete surface drainage network from being defined. (Kenny et al. 2008) presents an excellent overview of the many different DEM preprocessing procedures used by other investigators. Most preprocessing procedures can be broadly categorized as one of three methods: 1. DEM filling, 2. Outlet breaching, and 3. Imposing relief across flat areas. DEM filling is most often used to resolve the pits but can also be used to fill and overtop depressions. Outlet breaching drains depressions by cutting through barriers such as roadway fills. Imposing relief across flat areas adds small elevation increments to selected grid cells within the flat area to create a minimal downslope gradient in the direction of an outlet point.

The limitations of the D8 flow direction algorithm have been thoroughly discussed in the literature cited above, which are principally due to the 8 possible flow directions at increments of 45 degrees. This limitation can result in unrealistic drainage network maps in upslope areas corresponding to unconcentrated overland flow, producing parallel straight flow paths downslope before entering a stream channel. The direction of these parallel flow paths also tends to be biased by the raster grid orientation relative to the terrain surface. The parallel flow paths still occur when the D8 is applied to 1-m DEMs, but they are pushed well upslope since the higher resolution can resolve the first order stream channels which terminate the parallel flow lines. Multiple flow direction (mfd) algorithms have been developed that give a more realistic result but are more complex and add dispersion to the drainage network by distributing the downslope flow to more than one of the neighboring grid cells (Qin et al. 2007). The primary focus of this study is to present a new DEM preprocessing algorithm, so the D8 algorithm is used to illustrate the resulting drainage networks due to its simplicity.

PREPROCESSING USING A CELLULAR AUTOMATON ALGORITHM

A two-dimensional cellular automaton that operates on a two-dimensional array of values is an iterative process that computes updates to each array value from time level t to time level $t+1$ as a function of some fixed neighborhood of values at time level t (Packard and Wolfram 1985). If applied to the raster DEM illustrated in figure 1, using the D8 neighboring grid cells, the resulting functional relationship is given by:

$$z_{n,m}^{t+1} = f\left[z_{n-1,m-1}^{t},\, z_{n-1,m}^{t},\, z_{n-1,m+1}^{t},\, z_{n,m-1}^{t},\, z_{n,m+1}^{t},\, z_{n+1,m-1}^{t},\, z_{n+1,m}^{t},\, z_{n+1,m+1}^{t} \right] \quad (1)$$

Although time is not a variable in equation 1, it is useful to think of the iterative process as having a positive direction in time. Equation 1 requires a parallel computation where all array elevation values are updated simultaneously. This can be accomplished using parallel computing, but most personal computers (PCs) are serial processors which require two parallel DEM arrays to avoid mixing the two different time levels. The updated elevation value z at time level $t+1$ is stored in the second array leaving the z value at time level t in first array unchanged. After all the array values are updated using equation 1, the two arrays are then swapped before beginning the next iteration. The need for two parallel DEM arrays would appear to double computer memory requirements. Fortunately, the number of pits and flats in most DEMs is a small percentage of the total number of values in the DEM array, so a programming trick can be employed to greatly reduce memory size and improve the execution speed when using a serial processing algorithm.

Cellular Automata (CA) with Integer Arithmetic. The 1-m DEMs, as downloaded, have elevations provided in single precision (32-bit) floating point format with three or four decimal place precision. Since the elevation data is specified to have a 10-cm RMSE accuracy, there is little justification for the extra precision, but the extra decimal places are useful in preprocessing the elevation data to resolve pits, flats and depressions. A new algorithm named HydroFill has been developed to preprocess the DEM data by using integer arithmetic in conjunction with a simple CA rule to resolve the pits, flats and depressions and provide an unbroken downslope drainage path from all grid cells to the DEM outside boundary. HydroFill requires that the raster DEM be first copied and converted from the 32-bit floating-point format to a 32-bit integer format. The integer copy (iDEM) should retain most, if not all, of the significant figures used in the original DEM. The format conversion requires applying an “integer precision multiplier” (IPMult) whose magnitude determines how many significant figures are retained. The integer precision that seems to work well in moderate relief topography is a millimeter, hence the corresponding IPMult value used is 1,000. The integer precision selected is equal to the “unit depth” which is the depth increment added to all grid cells classified as flat during each iteration of equation 1. The unit depth also determines the minimum hydraulic gradient (1-mm per meter, in this study) which results when flat areas are processed by the CA algorithm. In areas of low relief topography, a tenth of a millimeter might be a more appropriate unit depth since a smaller minimum hydraulic gradient may be needed.

CA Rule Definition. After the iDEM is created, the preprocessing can proceed by first filling all pits to a level exactly equal to the lowest elevation found in the surrounding D8 neighborhood grid cells. This initial step converts all pits into flats prior to application of the CA rule at the beginning of the first iteration. The CA rule can be stated very simply as: "Raise all flats within the iDEM by one unit depth during each iteration of equation 1". It is important to emphasize that the CA rule only functions correctly with use of an integer DEM (iDEM) which has an integer precision that is exactly equal to the unit depth. Additionally, it is important to recall that application of the CA rule via equation 1 is a parallel computation that updates all flats within the iDEM simultaneously. Iterative application of the CA rule will eventually resolve all the flats in the iDEM, and the iterations cease when there are no flat grid cells remaining.

Preprocessing Depressions and Flat Areas Using the CA Rule. Individual flat grid cells must have at least one D8 neighbor cell at the same elevation, by definition, but none that are lower in elevation. If at least one of the neighboring D8 cells at the same elevation is also a flat, then a "flat area" is defined. Flat areas contain contiguous grid cells that are individually defined as flat. Flat areas can occur on the surface of the DEM, particularly when hydro-flattening is used to assign elevations to water surfaces. Flat areas are also created during the iterative application of the CA rule within depressions. All depressions have at least one pit at the beginning of preprocessing. That pit is initially filled to create a flat. A single flat within a depression is sufficient to start the process of filling the depression as the CA rule is iteratively applied to add elevation increments equal to the unit depth. Importantly, the requirement that the iDEM precision exactly matches the unit depth prevents any new pits from being formed in depressions as the simulated water surface elevation increases. New flat grid cells are created as the water depth increases, thus expanding the flat area as if the depressions are being filled with water. The flat areas within depressions continue to expand and grow in depth during each iteration until the water surface elevation rises one unit depth above the lowest grid cell (or cells) on the depression boundary, allowing overflow to occur. When overflow occurs, the flat grid cells neighboring the boundary grid cells at the overflow point are no longer defined as flat at the beginning of the next CA iteration since they have a newly acquired downslope flow direction. When flats are resolved due to acquiring a downslope to a neighboring cell during the CA iterative process, the downslope gradient is always equal to the minimum hydraulic gradient of 1-mm per meter. Once a flat cell is resolved next to an overflow point, it no longer receives the unit depth elevation increase during the following iterations. But its immediate neighboring flat grid cells upstream within the flat area still receive an elevation increase during the next iteration, thus resolving them also. This process is repeated during each following iteration and the minimum hydraulic gradient is propagated backward into the depression and eventually consumes all the flat grid cells, eliminating the depression. Other flat areas that are not within depressions are resolved in the same way by propagating the minimum hydraulic gradient inward from the area boundary until all flat grid cells are consumed. If the flat area boundary is closed, as with buildings with flat roofs or a plateau in the topography, the minimum hydraulic gradient propagates inward from the boundary simultaneously on all sides until meeting in the geometric center of the flat area.

The HydroFill Algorithm. As explained above, equation 1 requires that the CA rule be applied as a parallel computation that updates all flat grid cells simultaneously from time level t to time level t+1 (one iteration of the CA rule). This normally requires two parallel iDEM arrays to avoid mixing the two time levels when using a serial processing algorithm. HydroFill (a serial processing algorithm) avoids the need for the second parallel iDEM array and significantly reduces the execution time by making use of the fact that the number of flat grid cells in a typical DEM undergoing preprocessing is a very small fraction of the total (typically less than 5%). The algorithm utilizes three small one-dimensional (1-D) memory stacks, labeled “a”, “b”, and “f”, as shown in figure 2. The “a” stack stores the grid cell row-column locations of all flat grid cells being processed in the current iteration and the “b” stack stores those to be processed in the next iteration. The “f” stack avoids the need for a second parallel iDEM array by temporarily storing the water depths added as the flats are raised by the unit depth during each iteration. At the end of each iteration the “f” stack depths are added to the iDEM array. The HydroFill algorithm (currently coded in MATLAB script) follows the computational sequence outlined below:

1. First, scan the iDEM by row-column and locate all sink and flat grid cells but skip those cells identified as culvert entrances.
    a. Fill sinks to an elevation equal to the lowest D8 neighbor grid cell (converts sinks to flats).
        i. Store depths added in the “f” stack.
    b. Store row-column locations of all flats in the “a” stack.
2. Begin the iterative process by pulling row-column locations of flats from the “a” stack.
    a. Raise each flat grid cell elevation by adding 1 unit depth.
        i. Store depths added in the “f” stack.
    b. Check to see if the processed flat is already stored in the “b” stack.
        i. If “no”, add the row-column location to the “b” stack.
    c. Check the surrounding D8 neighborhood grid cells for those equal to or one unit depth greater in elevation than the focus cell.
        i. If “yes”, add those grid cells to the “b” stack.
3. The current iteration is completed after all grid cells stored in the “a” stack have been processed.
    a. Add the fill depths stored in the “f” stack to the iDEM array.
    b. Check to see if the “b” stack is empty.
        i. If “yes”, terminate the program.
        ii. If “no”, swap the memory pointers to the “a” and “b” stacks and return to step 2 above (stack “b” becomes stack “a” for the next iteration).

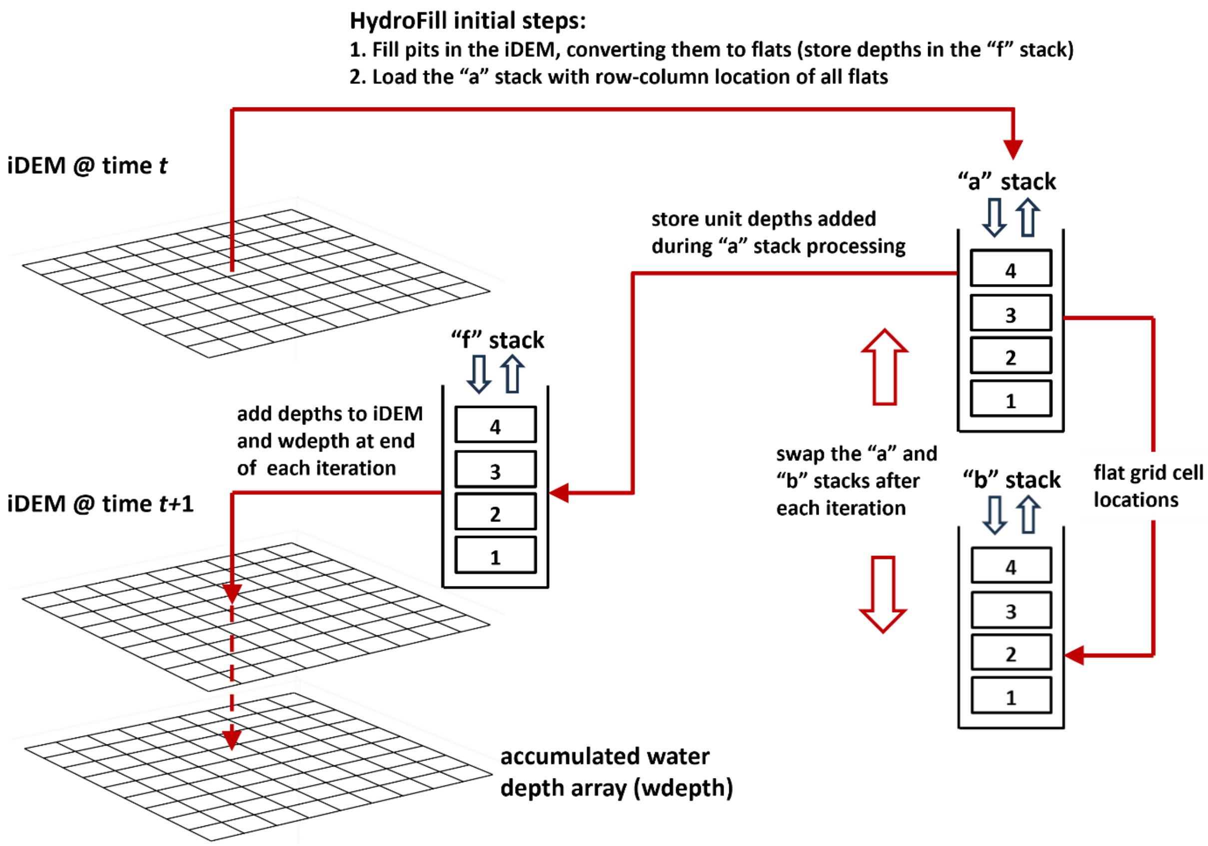

Figure 2. HydroFill iDEM preprocessing flow chart.

JOINT HYDROFILL AND QGIS APPLICATIONS

The HydroFill algorithm is currently implemented in MATLAB script to facilitate research and development ("MATLAB" n.d.). As discussed in detail above, the algorithm preprocesses DEM data to resolve pits, flats, and depressions prior to using the D8 flow accumulation algorithm to generate the completed drainage network for a DEM of interest. A suitable geographic information system (GIS) is needed to process and display the DEM, along with additional map layer data produced HydroFill. QGIS is a free and open-source GIS that was selected for this study ("Spatial without Compromise · QGIS Web Site" n.d.). As shown in figure 3, the first step is to download the DEM tiles along with the building structure footprints. In the examples that follow, 1-m DEM tiles and the associated building footprints are downloaded for an area of interest (northern West Virginia). The building footprints used are generated by Microsoft (MS) and can be downloaded from GitHub ("GitHub - microsoft/USBuildingFootprints: Computer generated building footprints for the United States" n.d.). The MS building footprints file is a vector polygon layer with a .shp (shapefile) format. Since HydroFill only works with raster DEM files, the building footprints vector file is rasterized and clipped in size within QGIS to exactly match the DEM. When the building footprints are added to the DEM, they must be given an arbitrary flat roof elevation above the surrounding terrain so that HydroFill can generate the minimum hydraulic gradient on the roofs and produce gradients that route surface flow around the building structures. The building roofs are processed the same way as any other flat area within the DEM. Each flat roof elevation is set to a value that is above its foundation elevation using the equation: $H_{roof} = z_{\max} + \log_{10}(A_{bldg})$, where H is the computed roof elevation, z is the maximum building footprint elevation on the terrain surface, and A is the building footprint area in square meters.

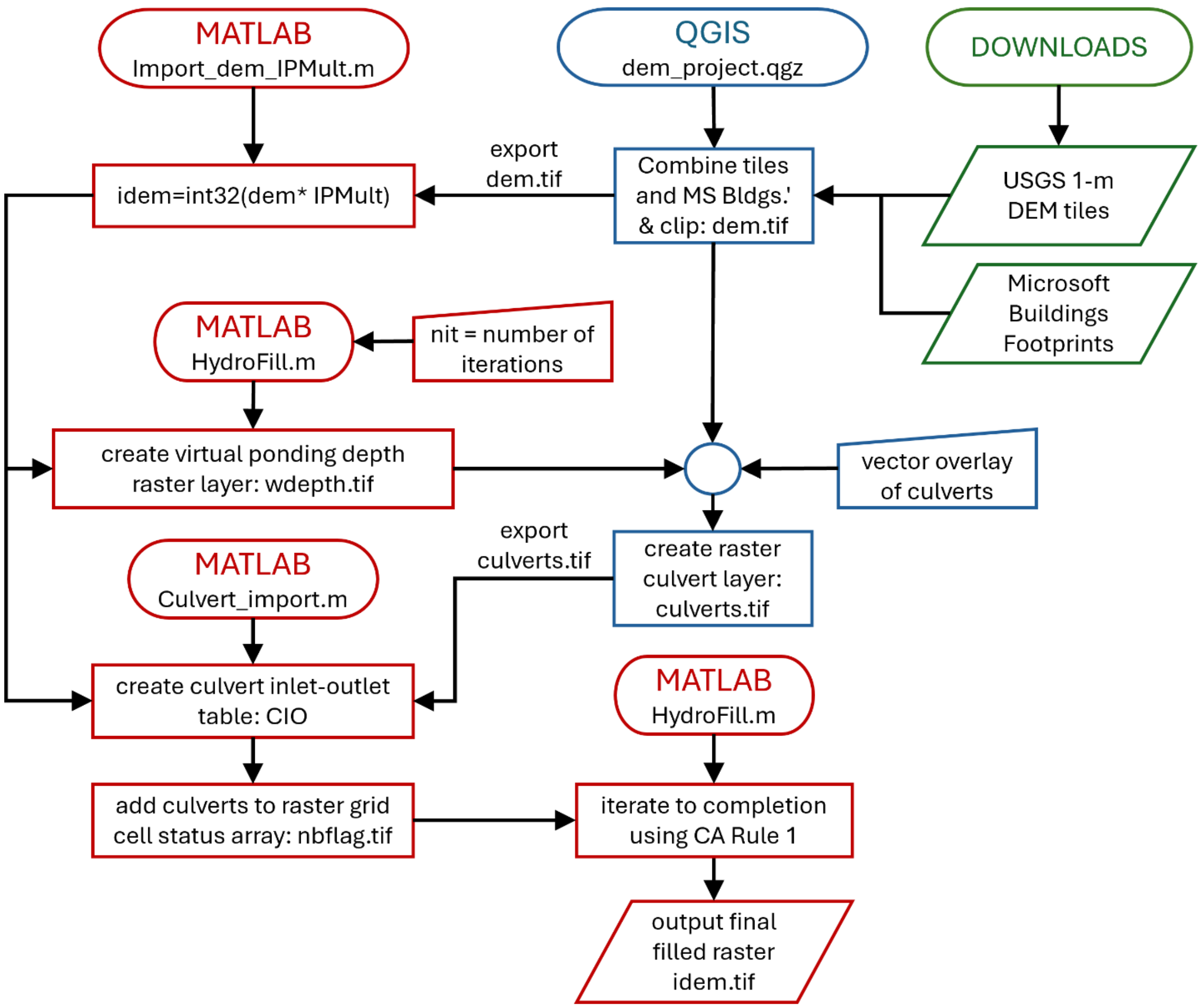

Figure 3. Flow chart illustrating the joint processing of DEM data and drainage infrastructure to output the fully drainable integer raster DEM (idem.tif).

Culvert Location Using Virtual Ponding. A common problem encountered when generating drainage networks using high resolution DEMs is artificial depressions created by elevated barriers that interrupt downslope flow connectivity. For example, an artificial depression is created by a roadway fill that crosses a stream channel. A common solution is to use a suitable algorithm to implement a “breaching” of the fill in such a way that reconnects the upstream drainage network to a point downstream of the roadway (Poppenga et al. 2010). The breaching method seems to work well for natural stream crossings, but it is not well suited for more complex subsurface drainage infrastructure that can be found in a typical developing suburban area. HydroFill provides a “virtual ponding” capability that can be useful in locating culverts and other subsurface drainage structures. As explained above, HydroFill fills depressions using a parallel computation that maintains a flat-water surface that rises 1-mm per iteration of the CA rule. Rather than let the algorithm iterate until it completes the drainage hydraulic gradient by overflowing the depressions, it can be stopped after a fixed number of iterations, leaving ponded water in the depressions at a known maximum depth. Ground truth survey information can then be combined with this virtual ponding to locate culverts and other subsurface drainage within the iDEM and thus complete the drainage connectivity.

Figure 4 illustrates placement of a network of culverts with the help of virtual ponding and ground truth information. The image is focused on an interstate highway fill which is clipped from a 20.25 sq. km, 1-m DEM of a suburban area near Morgantown, WV. Referring to figures 2 and 3, the virtual ponding depth “wdepth” is created by HydroFill. The wdepth layer is parallel to iDEM and has matching 1-m grid cells to store the water depths being added to each cell. For this application, HydroFill was programmed to execute for 2,500 iterations, which results in a wdepth array with a maximum depth of 2.5-m. Wdepth was then overlaid on the DEM, using QGIS, producing the left image in figure 4. The culverts were then digitized using the wdepth virtual ponding (plus ground truth information) to guide their precise location on the map layers, as shown in the right image in figure 4. The culvert map layer is a line segment vector shapefile (.shp file extension) that locates each culvert entrance and exit and assigns a culvert identification (id) number. Other descriptive information, such as culvert length, size, and slope, can be added to the attribute table that is created as the culverts are digitized. The culvert endpoints are extracted from the culvert line segment shapefile, producing a point shapefile with the points assigned a positive id number for the culvert inlet and a negative id number for the corresponding outlet. This latter point shapefile is then burned into a raster grid cell layer named “culverts.tif” so that the culvert inlet and outlet id numbers are then located at specific 1-m grid cells, with all remaining grid cells initialized to zero. With the culvert inlet and outlet id numbers now being assigned to specific grid cells, the row-column locations of the culvert inlet and outlet are identified and can be related to the raster iDEM. The culverts.tif file is then read into the MATLAB script program “Culvert_import.m” which creates the culvert inlet-outlet table “CIO”. CIO stores the culvert id number along with inlet and outlet grid cell row-column locations. The culvert inlet and outlet id numbers are pulled from the CIO table and stored at their grid cell locations in the raster grid cell status array “nbflag” (see figure 3). The nbflag array is read into the HydroFill program, bringing the culvert ids and locations along with it. The grid cells in nbflag are assigned integer digits during execution of HydroFill to signify whether that grid cell is already stored in the “b” stack or is a new addition to the “b” stack (figure 2). The purpose of nbflag is to assure that the same grid cell is not added to the “b” stack in multiples. The nbflag array can also be used to store the culvert id numbers since the culvert inlets are not to be added to either the “a” or “b” stacks and thus are not processed by the CA algorithm during the iterative computations. The culvert inlets then become sinks, which results in all the other grid cells within the depression having a downslope gradient toward the culvert inlet. The culvert exits are not treated differently than other grid cells during the iterative computation but do receive all the flow that enters the inlet when the drainage area accumulation is calculated. Figure 5 shows the result of letting HydroFill run to completion with no culverts added (left image), as compared to having the culverts added (right image). The resulting CA iteration statistics are plotted in figure 6. The left plot is for the case with no culverts and the right plot is for a total of 50 culverts installed on the entire DEM area of 20.25 sq. km. The plots illustrate the large difference in the number of iterations required to completely preprocess the iDEM and create a downslope flow path for all grid cells to the array boundary. The filling of several large depressions in the left image (no culverts) creates the “sawtooth” pattern of increasing numbers of flat grid cells that peak at the point where an individual depression is overtopped and then drops as the minimum hydraulic gradient is propagated backward into the depression, eliminating the flats. The no-culverts case required 17,846 iterations to complete while 4,662 iterations were required with the 50 culverts installed. The execution time was 10 times faster with the culverts installed (128 seconds on a Dell Precision 5820 desktop computer).

APPLICATION OF HYDRODRAIN TO COMPLETE THE D8 DRAINAGE NETWORK

After HydroFill preprocessing is completed, all iDEM grid cells have an uninterrupted downslope flow path to the outer boundary of the iDEM array. HydroFill only provides a downslope direction for each grid cell, so a second MATLAB script program (HydroDrain) is required to accumulate the drainage area downslope using the D8 nearest neighbor algorithm. As a first step, the HydroDrain algorithm reformats the iDEM, returning it to a 32-bit floating point format with 3 decimal place precision. HydroDrain then determines which D8 neighborhood cell corresponds to the steepest downslope and assigns that neighbor cell as the downslope cell receiving flow in the form of drainage area accumulation. The slope and distance to the downslope cell is also stored, along with the downslope cell row-column location within the DEM. Since the DEM data used here has 1-meter square grid cells, it is convenient to just count the number of grid cells that are tributary to each cell in the DEM array. The number of tributary grid cells accumulated to each cell are added to the focus cell and then passed on to the next cell in the downslope direction. This process yields a continuing flow path downslope with accumulating flow area. A convenient way to visualize the downslope flow paths is to plot the accumulated grid cell area above a selected threshold. These flow paths are overlaid on the DEM and wdepth layers shown in figure 5, with a red color ramp on the left image and a blue color ramp on the right image (higher accumulated areas have a darker color shade). The area threshold used for plotting the flow paths is 2,000 sq. m.

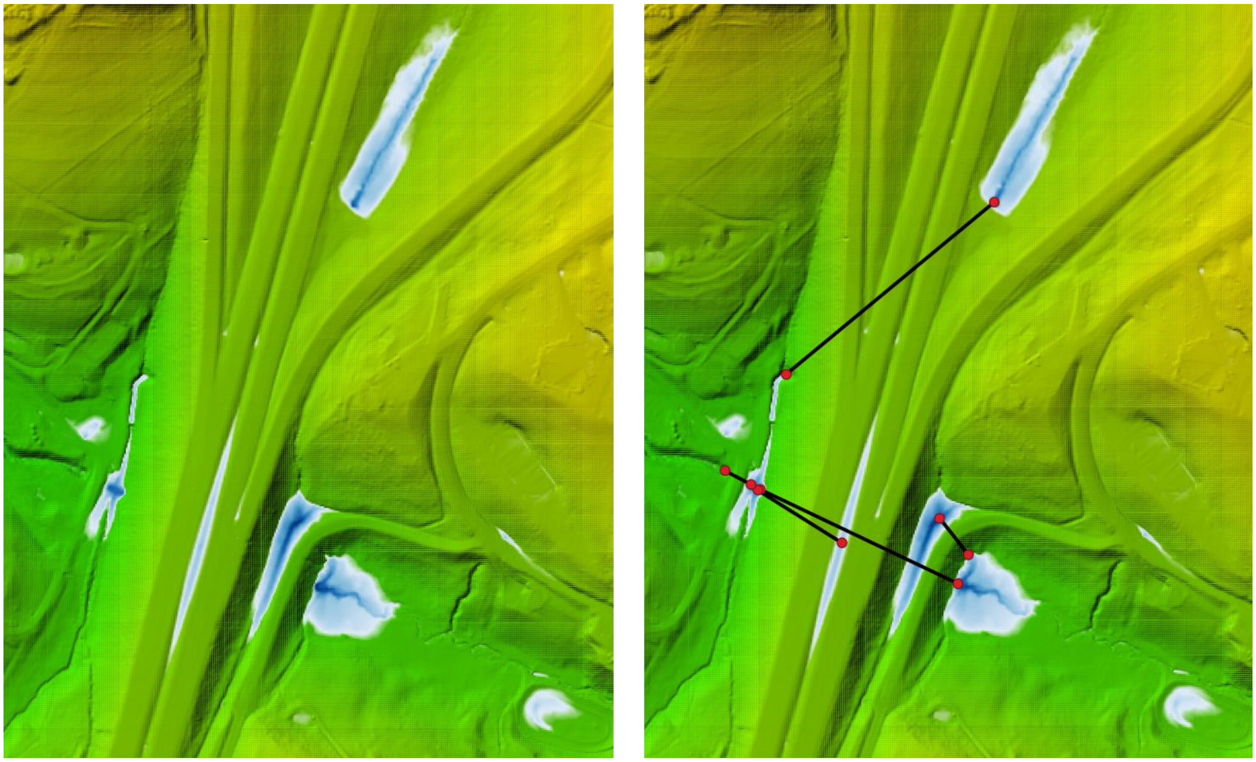

Figure 4. I-68 divided highway fills over West Run near Morgantown, WV. Left: HydroFill virtual ponding to a maximum depth of 2.5 m (using a blue color ramp). Right: Culvert network placement using virtual ponding, combined with ground truth information.

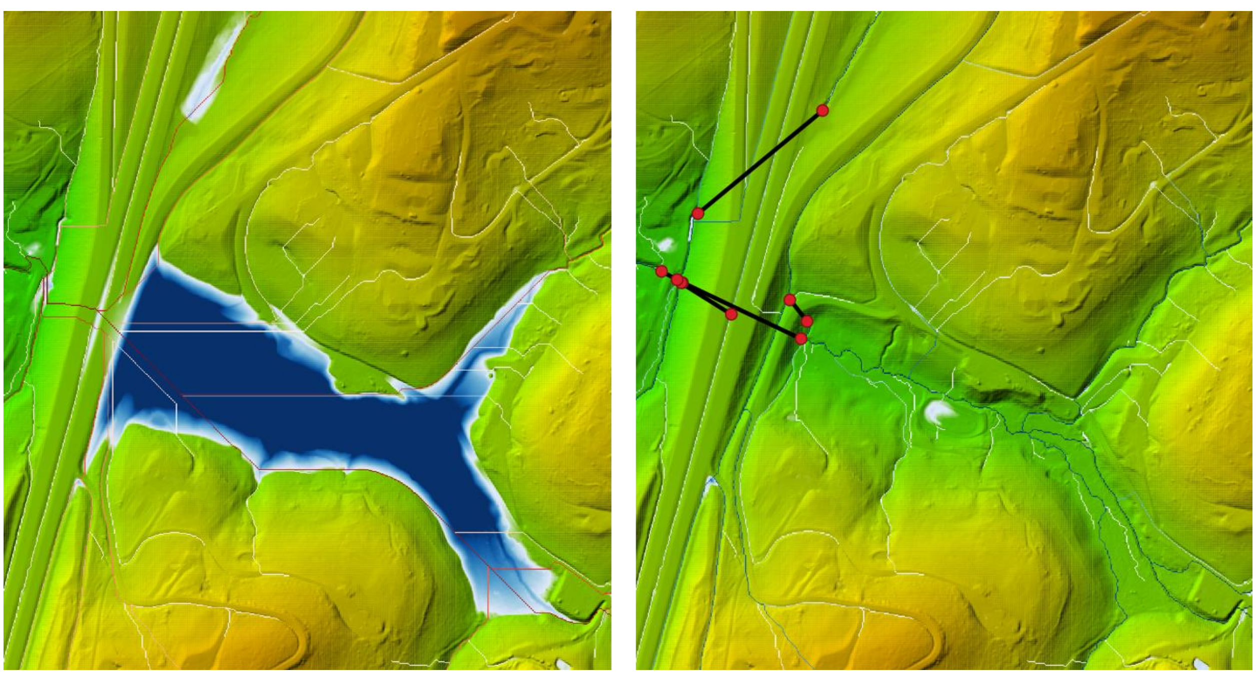

Figure 5. Comparison of ponding depths with HydroFill allowed to run to completion for the case of no culverts (left image) as compared to culverts added (right image).

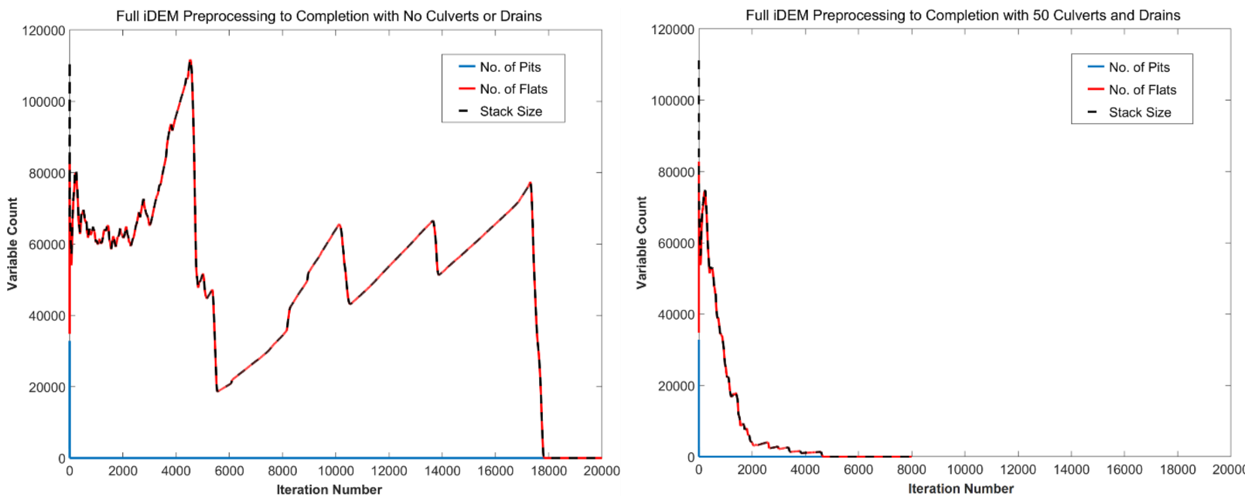

Figure 6. HydroFill iDEM preprocessing completion comparison between no culverts on the left and 50 culverts on the right for the full 20.25 sq. km iDEM.

ADDITION OF MICROSOFT BUILDING FOOTPRINTS DEM

The HydroFill and HydroDrain algorithms were also applied to the Whites Run Watershed, a small watershed contained within the full-size 20.25 sq. km 1-DEM data set used previously. The Whites Run DEM was clipped from the full-size DEM with a resulting array dimension of 3168 x 3259 meters (10.3 sq. km). The two algorithms were first applied in sequence to the DEM without the MS building footprints and then repeated with the buildings added. Figure 7 shows a comparison of the results using a hill shaded portion of the watershed that contains several drainage infrastructure features, including culverts and drainage ponds with outflow piping. As before, the drainage network can be visualized by producing a raster image with only the grid cell flow accumulation areas above a selected threshold of 2,000 sq. m. Six culverts are shown in figure 7 (in red) that are active in carrying significant flow accumulations through road fills. Three ponds can be seen that contain some virtual ponding. Two of the ponds have outflow riser pipes and the remaining pond was found to drain through overflow only. The dark red flow network lines are for the non-building case while the dark blue network lines are for the buildings being present. There are many cases where the dark red flow accumulation lines intersect building footprints, and some of those cases are shown with the yellow ovals. The dark blue flow accumulation lines can be seen to correct those flow lines to go around the building structures.

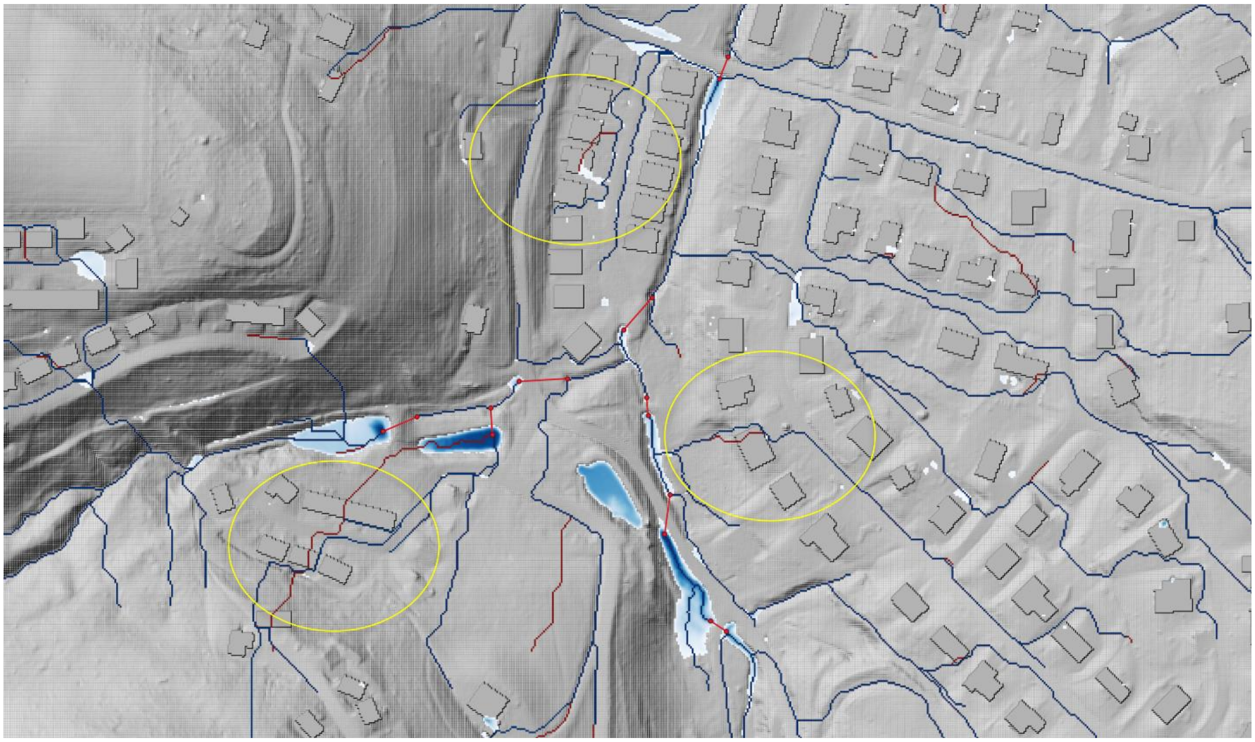

Figure 7. Comparison of drainage networks generated by HydroDrain for the no-buildings case (red lines) and the case with building structures being added to the DEM (blue lines).

CONCLUSIONS AND FUTURE DIRECTIONS

The HydroFill CA algorithm is the essence of simplicity since it uses a single simple CA rule to build a minimal hydraulic gradient in flat areas found within raster grid cell DEMs. The algorithm requires the DEM to contain valid elevations in all grid cells and that the DEM be copied and reformatted with integer elevation values that have a precision that matches a chosen unit depth. Pits in the DEM are identified first and are filled and converted into what are defined as flat grid cells. The CA rule is simply to increase the elevation of flat grid cells by the unit depth during each iteration of the algorithm. Flat grid cells are defined based on the elevations of the surrounding D8 neighborhood cells. No searching is required beyond the D8 neighborhood which ensures robust characteristics that will provide an unbroken downslope path for each grid cell contained withing the DEM. The iterative application of the CA algorithm propagates information about the surrounding terrain characteristics at a velocity of one grid cell distance per iteration. Thousands of iterations result in surrounding terrain characteristics becoming available to the flat grid cell locations being currently processed by influencing the elevations of the neighbor D8 grid cells. Culverts and other drains are incorporated by simply not processing the grid cell corresponding to the inlet location, which results in that inlet grid cell becoming a sink. After the DEM is preprocessed by HydroFill, HydroDrain completes the accumulation of drainage area downslope to the DEM boundary. Culverts and drains have inlets located as sinks within depressions and HydroDrain connects those inlet grid cells to the corresponding outlet grid cells to complete an unbroken drainage network. Application of HydroFill and HydroDrain to the 20.25 sq. km 1-m DEM demonstrated that the drainage networks can be completed in the presence of a vast variety of terrain conditions, including depressions created by road fills and drainage ponds, and added building structures with flat roofs.

The current HydroFill and HydroDrain algorithms have been developed and refined over a 20-year period, beginning with a poster displayed at the NKS 2003 conference held in Boston, MA. A photo of the author presenting the poster is available online ("A New Kind of Science" n.d.). A copy of the photo is attached at the end of this paper. Continued development of these algorithms is planned in conjunction with hydrologic modeling and programming language conversion to Python. The software algorithms referenced in this paper will be made available via GitHub prior to publication ("elirober/BasinMetrics" n.d.).

REFERENCES

"3D Elevation Program | U.S. Geological Survey." n.d. Accessed November 7, 2024. https://www.usgs.gov/3d-elevation-program.

"A New Kind of Science: NKS 2003 Conference: Photos." n.d. Accessed December 11, 2024. https://www.wolframscience.com/conference/2003/photos/.

Band, L. E. 1986. "Topographic Partition of Watersheds with Digital Elevation Models." *Water Resources Research*, 22 (1): 15–24. https://doi.org/10.1029/WR022i001p00015.

"Elevation-Derived Hydrography Data Acquisition Specifications: Positional Assessment Requirements | U.S. Geological Survey." n.d. Accessed November 17, 2024. https://www.usgs.gov/ngp-standards-and-specifications/elevation-derived-hydrography-acquisition-specifications-1.

"elirober/BasinMetrics." n.d. Accessed December 11, 2024. https://github.com/elirober/BasinMetrics.

Fairfield, J., and P. Leymarie. 1991. "Drainage networks from grid digital elevation models." *Water Resources Research*, 27 (5): 709–717. https://doi.org/10.1029/90WR02658.

"GitHub - microsoft/USBuildingFootprints: Computer generated building footprints for the United States." n.d. Accessed November 22, 2024. https://github.com/microsoft/USBuildingFootprints.

Jenson, S., and J. Domingue. 1988. "Extracting Topographic Structure from Digital Elevation Data for Geographic Information-System Analysis." *Photogramm. Eng. Remote Sens.*, 54 (11): 1593–1600. Bethesda: Amer Soc Photogrammetry.

Kenny, F., B. Matthews, and K. Todd. 2008. "Routing overland flow through sinks and flats in interpolated raster terrain surfaces." *Computers & Geosciences*, 34 (11): 1417–1430. https://doi.org/10.1016/j.cageo.2008.02.019.

Martz, L. W., and J. Garbrecht. 1992. "Numerical definition of drainage network and subcatchment areas from Digital Elevation Models." *Computers & Geosciences*, 18 (6): 747–761. https://doi.org/10.1016/0098-3004(92)90007-E.

"MATLAB." n.d. Accessed November 27, 2024. https://www.mathworks.com/products/matlab.html.

O’Callaghan, J. F., and D. M. Mark. 1984. “The extraction of drainage networks from digital elevation data.” *Computer Vision, Graphics, and Image Processing*, 28 (3): 323–344. https://doi.org/10.1016/S0734-189X(84)80011-0.

“OGC GeoTIFF.” n.d. *Open Geospatial Consortium*. Accessed November 9, 2024. https://www.ogc.org/publications/standard/geotiff/.

Packard, N. H., and S. Wolfram. 1985. “Two-dimensional cellular automata.” *J Stat Phys*, 38 (5): 901–946. https://doi.org/10.1007/BF01010423.

Poppenga, S. K., B. B. Worstell, J. M. Stoker, and S. K. Greenlee. 2010. *Using Selective Drainage Methods to Extract Continuous Surface Flow from 1-Meter Lidar-Derived Digital Elevation Data*. *Scientific Investigations Report*. U.S. Geological Survey.

Qin, C., A. -X. Zhu, T. Pei, B. Li, C. Zhou, and L. Yang. 2007. “An adaptive approach to selecting a flow-partition exponent for a multiple-flow-direction algorithm.” *International Journal of Geographical Information Science*, 21 (4): 443–458. Taylor & Francis. https://doi.org/10.1080/13658810601073240.

“Spatial without Compromise · QGIS Web Site.” n.d. Accessed November 27, 2024. https://www.qgis.org/.

“Topographic Data Quality Levels (QLs) | U.S. Geological Survey.” n.d. Accessed November 8, 2024. https://www.usgs.gov/3d-elevation-program/topographic-data-quality-levels-qls.

“Universal Transverse Mercator coordinate system.” 2024. *Wikipedia*.

“What is a digital elevation model (DEM)? | U.S. Geological Survey.” n.d. Accessed November 7, 2024. https://www.usgs.gov/faqs/what-a-digital-elevation-model-dem.

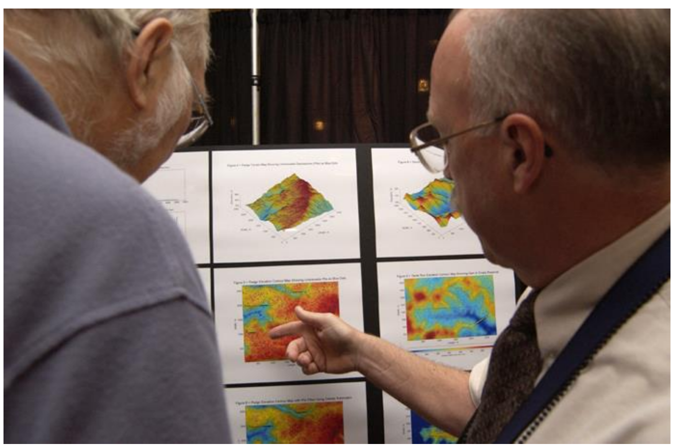

Photo 1. Poster presentation at NKS-2003 (“A New Kind of Science” n.d.)