Calculating a LiDAR-based "Erosion Score": An ArcGIS Toolbox Tutorial

Aaron Ruesch, Sarah Kempen, and Theresa NelsonWisconsin Department of Natural Resources,101 S. Webster St.Madison, WI 53707-7921<Aaron.Ruesch@wisconsin.gov>

This document is a tutorial describing the steps required to calculate a spatially explicit erosion score based on a LiDAR digital elevation model (DEM). The erosion score is a bivariate index describing a landscape’s relative vulnerability to sheet, rill, and gully erosion by incorporating information regarding topography, soils, rainfall, and land cover. The tools are intended for relatively small watersheds (less than 100 sq. km.) that have already been identified as watersheds that contribute higher non-point-source pollutant loads. For example, a hydrologic unit (HUC) that has been identified in a Total Maximum Daily Load (TMDL) model as a critical source area would be an appropriate place to use this tool. Presuming that the TMDL model has accurately identified the watershed as a pollutant source, we can further refine the prioritization to specific fields or sub-watersheds within the HUC12. With this framework, the intention is to prioritize the implementation of best-management practices on fields where pollutant loads are estimated to be the highest, thereby increasing the probability of improving downstream water quality using the fewest resources (i.e., "bang-for-buck" implementation).

# Preparation

The tools were built for Wisconsin landscapes, and thus use data specific to Wisconsin, including post-processed LiDAR DEMs distributed by [WisconsinView](http://relief.ersc.wisc.edu/wisconsinview/form.php)[[1]](#footnote-24)[wisconsinViewWebsite] and Wisconsin-specific coordinate systems (Wisconsin Transverse Mercator). In terms of software, the tools require ArcGIS 10.0 or later, the Spatial Analyst extension, and an optimized DEM fill tool that can be downloaded [here](http://tools.crwr.utexas.edu/OptimizedPitRemoval/Optimized-Pit-Removal-V1.5.1.zip)[[2]](#footnote-26). In terms of data, you will need LiDAR data downloaded from [WisconsinView](http://relief.ersc.wisc.edu/wisconsinview/form.php), a feature class polygon for the boundary of your area of interest (recommended smaller than 100 sq. km.), [NLCD land-cover data](http://www.mrlc.gov/nlcd06_data.php)[[3]](#footnote-28)[nlcdWebsite], [gSSURGO soils data](http://datagateway.nrcs.usda.gov/GDGOrder.aspx?order=QuickState)[[4]](#footnote-30)[nrcsWebsite], and optionally, the [USDA common land unit dataset (CLU)](http://www.fsa.usda.gov/FSA/apfoapp?area=home&subject=prod&topic=clu)[[5]](#footnote-32). To download the CLU dataset, you must be a FSA/NRCS/RD[[6]](#footnote-33) employee. If you are not employed by one of these agencies, you may use an alternate boundary for summarization (e.g., parcels, sub-watersheds). Running the tools will likely require a moderate-to-high level of familiarity with GIS concepts and ArcGIS Desktop software.

1. *Load the tools into ArcGIS*: Open ArcMap. Navigate to ArcToolbox. Right-click on then click . Navigate to the file stored in the root directory of the distributed toolbox.
2. *Define an area of interest*: The area of interest, for the purpose of water quality improvement, will likely be a watershed boundary, and therefore will be referred to as the "watershed" within this document. Watershed boundaries are available on the internet through USGS—the finest-scale watershed boundaries that are available are [Hydrologic Unit Codes level 12 (HUC12s)](http://water.usgs.gov/GIS/huc.html)[[7]](#footnote-35). The DNR has an internal version of HUC16 delineations that is available upon request. If no watershed boundary exists, you can manually digitize a boundary using the LiDAR DEM or contours if they are availabe.
3. *Download the gSSURGO soils database*: Go to the NRCS [GeoSpatialDataGateway](http://datagateway.nrcs.usda.gov/GDGOrder.aspx?order=QuickState) ’Order by State’ website. Under , choose . You will be prompted to select a dataset—scroll down to and choose and click , then again. Fill in the form with your personal information and click and . An email will be sent to you when your order is ready. Find the FTP site in the email, download the associated zip file, and unzip it. There will be another zip file in the resulting directory—unzip that as well and note the location of .
4. *Download the NLCD2006 land-cover raster*: Go to the [Multi-Resolution Land Characteristics Consortium (MRLC)](http://www.mrlc.gov/nlcd06_data.php) download website for NLCD2006 and click the top link under labeled . The data should immediately begin downloading. When the download is complete, unzip the file and note the location of .
5. *Download LiDAR DEM(s)*: Go to the [WisconsinView Data Portal](http://relief.ersc.wisc.edu/wisconsinview/form.php) (see link on page . If this is your first time visiting the WisconsinView Data Portal, fill in the form at the bottom of the page with your personal information, click , accept the disclaimer, and click . Click the county(s) where your watershed is located. This will direct you to an FTP site—click the zip file in the top level directory. Unzip the file and note the location where the resulting directory was stored.

# Tools

## Condition the DEM

To properly describe how water moves across your watershed, we must first "condition" the DEM (methods based on ), which means to remove presumed errors in elevation so that the GIS model of hydrology can correctly infer the direction of flow based on changes in elevation.

One of the major barriers in this process is correctly modeling the flow of water through culverts and bridges. Therefore, we have to first manually digitize all culverts and bridges in our watershed where flowpaths are not properly reprsented by the DEM. The majority of these cases will occur where road berms were built. These topographies create "digital dams" that prevent the hydrologic model from appropriately describing downslope flow.

To eliminate digital dams, we will digitize each culvert as a polyline feature following these steps (see Figure  [fig:digitizingCulverts]):

1. Create a "filled" DEM using the tool in ArcGIS Spatial Analyst and subtract the original DEM from the filled DEM to create a raster of the depression depths. Alternatively, you can type this snippet of code into the ArcGIS Desktop Python window (image):from arcpy.sa import \*dem = "C:/path/to/dem"depressionDepth = "C:/path/to/depressionDepth"depressionDepth = Fill(dem) - Raster(dem)depressionDepth.save(depressionDepth)replacing "C:/path/to/dem" and "C:/path/to/depressionDepth" with the file paths for your LiDAR DEM and the output raster for inferring where depressions lie.
2. Deep depressions near road berms are places where a culvert is probably located. Locate one of these depressions using the depression-depth raster from above, then zoom to it.
3. To visualize digital dams, open the tab in the window for the original DEM. Choose the option using within the box, and choose a color gradient that spans multiple hues (e.g., a rainbow color scheme).
4. Create an empty polyline feature class.
5. Start an editing session on the empty polyline feature class.
6. Each polyline feature should only have two nodes. The nodes need to be drawn in order, from upstream to downstream, across the digital dam.
7. Use the rainbow-colored DEM to locate the lowest elevation DEM gridcell on the upstream side of the digital dam. Digitize the upstream node on the center of that cell.
8. Next, locate a DEM gridcell on the downstream side of the digital dam that has lower elevation than the gridcell underlying the upstream node. Digitize the downstream node on the center of that cell and finish the feature.
9. Repeat until all digital dams have culvert features spanning across them.

A correctly digitized polyline will inform a DEM interpolation routine that the digital dam should be "cut" along the geometry of the culvert feature according to the slope of the elevation between the DEM gridcells on the up- and downstream side of the digital dam.

After all culverts have been digitized, open the tool in the toolbox. Add the polyline feature class, the polygon feature class, and the raster dataset that you downloaded from [WisconsinView](http://www.wisconsinview.org/). The tool also asks for the ; after decompressing the downloaded [optimized fill software](http://tools.crwr.utexas.edu/OptimizedPitRemoval/Optimized-Pit-Removal-V1.5.1.zip)[[8]](#footnote-39), choose the file from the resulting directory. Finally, define an and an . The will be a raster file that has minor errors and digital dams removed. It will have a coarsened cellsize of 3 feet (unless the cell size was manually changed in the ). The will be a DEM that has all "pits" (depressions) removed where water would theoretically pool. That is, the resulting elevation model has no seepages or wetlands that store water—all water runs off.

## Identify Non-contributing Areas

Because the eroded soil that flows into non-contributing areas (also known as "Internally Drained Areas") does not contribute to water quality problems in terms of suspended solids, these areas will eventually be removed from the erosion score (ES) model. But first, the extent of non-contributing areas must be identified.

We assume that non-contributing areas are areas where the volume of the depression on the landscape is large enough to store the runoff produced from a precipitation event as large as a specified storm event. To model runoff, we use a "curve number" approach in which the proportion of precipitation that results in runoff is based on land cover and hydrologic soil group.

### Prepare Precipitation Data

To identify non-contributing areas, we need raster precipitation data describing a depth of rain that will be used to define the volume threshold (i.e., capacity) of each non-contributing area. The depth of rain is specified in the tool by the frequency of occurence (in units of years) of a specified duration of storm (in units of hours). For example, a frequency of value 10 and duration of value 24 represents the 10-year 24-hour rainfall event. The tool will automatically download raster data from NOAA’s [Precipitation Frequency Data Server](http://dipper.nws.noaa.gov/hdsc/pfds/)[[9]](#footnote-43). The resulting precipitation raster will be in thousandths of inches

In ArcToolbox, expand the toolbox and double-click the script tool. Define the and of storm. Then choose a (usually the DEM for the watershed) which will define the cell size, grid, extent, and mask of the resulting precipitation depth raster. Finally, define an .

### Create a curve number raster

Open the tool. For , find the database file (extension .dbf) called —it will be in the root directory of the folder where the distributed tools are located. This table includes curve numbers for each unique combination of NLCD class and SSURGO hydrologic soil group (see Table  [table:curveNumberTable] taken from ). For the , browse for the version of NLCD2006 that you downloaded. For the , point to the file geodatabase that you downloaded from the NRCS GeoSpatialDataGateway (see Preperation). Finally, add the from tool 1 (to use as a grid template) and define an output .

We will use the NLCD land cover data to create a raster surface of curve numbers. First, clip your version of NLCD to the extent of the from tool 1, . Then, add a field to the clipped NLCD raster to store a curve number value associated with each land cover class. There is some debate over what curve numbers should be used for each land cover type (and generally whether curve numbers can effectively be applied to satellite-based land-cover classes).

### Run the tool

After you have created a curve number raster, you are ready to run the tool for identifying non-contributing areas. Open the tool in the toolbox. Add the and from tool 1, the from tool 2a, the you just created, and the polygon. Finally, define an raster (a raster with unique, but arbitrary, IDs for each non-contributing watershed) and an . The is simply the conditioned DEM with the extent of non-contributing areas set to *NoData*.

## Recondition DEM for Non-contributing Areas

Creating a hydrologically valid DEM is dependent on the extent of DEM cells. Often, the DEM cells are coerced to model "outward" flow on the edge of the DEM raster, presuming that the edge of the raster defines an area outside of the watershed. To properly model non-contributing areas, we have to recondition the DEM to force the direction of hydrologic flow *toward* the non-contributing area. Thus, accumulating flow will not be propogated downstream of the non-contributing area. That is, flow that enters a non-contributing area will "dissappear."

To hydrologically correct for non-contributing areas, open the tool. Add the from tool 2c. You may also optionally add , which could include any features on the landscape that could eliminate runoff, such as grass waterways, strip cropping, or any other best-management practices that would eliminate runoff that can be identified spatially. These features should be rasterized to the cellsize, extent, and grid (i.e., "snapped") of the . Next, find the path to the as you did in tool 1. Finally, define an . This DEM will be the final DEM used to estimate the ES.

## Calculate Stream Power Index

The first metric in the bivariate ES is the "stream power index." The stream power index includes information on slope and upstream drainage area—it is used for estimating the locations of gully erosion:

where is the stream power index at gridcell , is the upstream drainage area (flow accumulation at gridcell multiplied by gridcell area), and is the slope at a gridcell in radians.

Open the tool in the toolbox. Add the from tool 1. This input will be used to calculate slope, so you do not want to confuse this input with any "filled" DEM—filled DEM are used only for modeling hydrologic flow, and thus should not be used to calculate other topographic metrics such as slope. Add the from tool 3. This input will be used to calculate drainage area using a "flow accumulation" raster. Assign a —this value sets the number of upstream gridcells for which the flow is assumed to transition to surface water[[10]](#footnote-48). The stream power index is intended to only model landscape erosion, not instream or bank erosion, and thus gridcells assumed to be surface water should be excluded. Finally, define an .

## Calculate Soil Loss Using the Universal Soil Loss Equation (USLE)

### Overview of the USLE

The second metric in the bivariate ES is soil loss as estimated by the USLE. The USLE estimates sheet and rill erosion across fields using:

where at each gridcell , is soil loss in , is rainfall erosivity in , K is soil erodibility in , is slope/slope-length (dimensionless), is a land cover factor (unitless), and is a "practice" factor (unitless metric describing the capacity of a land-management practice to prevent erosion) which is beyond the scope of this tool and thus will assumed to be 1. All these data need to be acquired prior to running the tool, and will take some time to prepare. Preparation steps are outlined below in detail.

### Prepare GIS datasets

Rainfall erosivity (or R-factor) is the cumulative force of rainfall on soil. In Wisconsin where topographic effects to rainfall intensity are minimal, erosivity varies at very large spatial scales. Because this tool is intended for small watersheds, erosivity can safely be assumed to be constant. For an appropriate constant value of erosivity, see Fact Sheet 3.1 page 5 for a map of erosivity in U.S. customary units. Multiply this value by 17.02 to convert to international (SI, or metric) units. Alternatively, if you know *a priori* that erosivity varies at a scale within the extent of your watershed, you can define an . If you choose to use an , resample the grid to the extent and resolution of the from tool 1—ensure that the resampled grid matches the grid of the by setting the to the in the of the tool.

Erodibility, or "K-factor," is an attribute that can be found in the gSSURGO soils database. Open the tool in the toolbox. Under , point to the file geodatabase that you downloaded from the NRCS GeoSpatialDataGateway (see Preperation). Keep the field as the default . Add the to use as a template for the resulting K-factor raster and the polygon used to clip the DEM (these are the same two inputs from tool 1). Finally, define an and run the tool.

The last step in preparing the GIS datasets for calculating soil loss is creating a . The C-factor is a unitless metric that adjusts the amount of soil loss for different land cover types. To create the C-factor raster, we will simply reclassify the NLCD2006 land cover codes to a meaningful C-factor value. Example C-factors are listed in Table  [table:cFactorTable] (taken from , Appendix F). Open the tool in the . Define the as the NLCD2006 and the as . Manually input the old-to-new reclassification in the provided interface. The new values can only be integers, so enter the desired C-factors after multiplying by 1000. Define an and run the tool. The resulting C-factor raster will need to be reconverted to a decimal number. Open the tool in . In the box, enter the following command:Float("C:\cFactorTimes1000") / 1000where "C:\cFactorTimes1000" is the path to the reclassified raster from above. The resulting will be used as an input to the tool.

### Run the tool

Open the tool in the toolbox. Add the from tool 1, and the from tool 3. If you decide to use a raster version of erosivity, add that dataset to the parameter. On the contrary, if erosivity can be considered constant, add that value to the box. In either case, use an erosivity value in SI units of . You will notice that both erosivity parameters are considered optional—if no value is given, an erosivity value of 1 will be used. If an erosivity value of 1 is used, the ES will be the same as if you had entered a , however the resulting soil loss raster cannot be interpreted on an absolute scale. Next, add both the calculated from gSSURGO and the calculated from NLCD. The is considered optional—if no value is given, the C factor is given a value of 1. As you did in tool 4, define a to set the number of upstream gridcells for which the flow is assumed to transition to surface water. Finally, define an output soil loss raster. If all inputs were in SI units, and all optional inputs were set to a meaningful value, the output will be in units of .

## Calculate Erosion Score

The ES is a bivariate index based on the stream power index and USLE-based soil loss. The ES gives equal weight to both by first, normalizing the distribution of soil loss by log-transforming (the stream power index is already normalized), second, centering their distribution on 0 and forcing the standard deviation to 1 by calculating a Z-score[[11]](#footnote-54), and finally taking the average of the Z-scores of the soil loss and the stream power index. Therefore, ES can be interpreted as the number of standard deviations from the mean. As a standard benchmark, an ES above 2 is above the 95th percentile of all pixels.

Open the tool. Add the from tool 5b and the from tool 4. You may optionally add a if you choose to summarize the ES—the natural boundary in this case would be the CLU field boundaries. If you choose to summarize by a zones, you must choose a (if the is a string/text data type, the zone values will be coerced to sequential integers). Finally, add the and optional outputs. The is required if a was defined.

# Interpretation and Limitations

Congratulations! If you have read this far, it hopefully means that you have an ES raster layer and likely an ES summarized by agricultural field. These results can help you prioritize where to focus your efforts for improving water quality in your watershed downstream.

If you defined an in the tool, you will have a table of summary statistics of erosion scores that you can join to your to locate priority features by creating color maps of different statistics using quantitative symbology on your boundary feature class. The summary table will have several statistics to choose from including the mean, maximum, and sum of all erosion score values within the boundaries of each feature. If you are most concerned with highly concentrated erosion, perhaps the most valuable statistic would be the mean erosion score, however the mean erosion score may reflect anomolous high values that skew the average—this is particularly a problem in small fields. If you are most concerned with total sediment runoff, perhaps the best statistic would be the sum, however the sum erosion score may reflect simply the size of the field, which may obscure otherwise priority fields that are smaller, but have more intense runoff events. If you are interested in the highest intensity runoff areas, you may be interested in the maximum erosion score, however this statistic may have the same problem with interpretation as the mean statistic—it may reflect anomolous high pixels, particularly in small fields. There are several ways to look at the resulting data, and no single statistic gives the correct answer. We advise the user to explore each of these statistics for different insights regarding erosion processes in your area of interest.

This dataset is best interpreted in combination with other determining factors of erosion and/or nutrient runoff. For example, as a watershed manager, you may be interested in what cropping practices are present—if information regarding specific practices are availabe and a P-factor can be inferred, these practices can be used to modify the USLE by multiplying a P-factor raster by the USLE raster. Additionally, you may be interested in fields that are on or near surface water—the ES summarized by field can be overlaid on a feature layer with information regarding the location of perennial or intermittent streams to gain insight into the potential for sediment delivery and transport.

Although the ES data product can be useful for prioritizing areas for implementation of best management practices, there are several key limitations that decision-makers should be aware of when interpreting the data. First, the ES is a unitless index. Although the output of the USLE raster is intended to have value in terms of total soil loss in , the stream power index raster does not. Because the two are combined into a single index, the resulting raster cannot be interpreted absolutely; differences from field-to-field can only be interpreted ordinally. Therefore, this tool cannot be used to make inferences regarding the water quality improvement associated with the implementation of best-management practices, but only to make decisions about where to prioritize implementation. Second, the ES itself has not been rigorously validated. Calculating erosion using a GIS is a relatively new science—there are many case studies investigating soil loss using a GIS-based USLE and a few studies identifying gullies using the stream power index. The results of these studies are mixed in terms of prediction accuracy, and there are no meta-analyses to date that broadly investigate the utility of either metric. Because of the above limitations, this tool cannot be used to predict erosion *rates*, but rather the relative probability of a field to have *more* erosion problems than its neighbors.

# Tables

[h]

|  |  |  |
| --- | --- | --- |
| NLCD Code | Description | C-Factor |
| 11 | Open Water | 0 |
| 21 | Developed, Open Space | 0.003 |
| 22 | Developed, Low Intensity | 0.002 |
| 23 | Developed, Medium Intensity | 0.001 |
| 24 | Developed, High Intensity | 0 |
| 31 | Barren Land (Rock/Sand/Clay) | 0 |
| 41 | Deciduous Forest | 0.003 |
| 42 | Evergreen Forest | 0.003 |
| 43 | Mixed Forest | 0.003 |
| 52 | Shrub / Scrub | 0.004 |
| 71 | Grassland / Herbaceous | 0.004 |
| 81 | Pasture / Hay | 0.004 |
| 82 | Cultivated Crops | 0.29 |
| 90 | Woody Wetlands | 0.003 |
| 95 | Emergent Herbaceous Wetlands | 0.004 |

[table:cFactorTable]

[h]

lcccc &NLCD Code and Description & A & B & C & D11 Open Water & 100 & 100 & 100 & 10021 Developed, Open Space & 89 & 92 & 94 & 9522 Developed, Low Intensity & 57 & 72 & 81 & 8623 Developed, Medium Intensity & 61 & 75 & 83 & 8724 Developed, High Intensity & 89 & 92 & 94 & 9531 Barren Land (Rock/Sand/Clay) & 77 & 86 & 91 & 9441 Deciduous Forest & 36 & 60 & 73 & 7942 Evergreen Forest & 36 & 60 & 73 & 7943 Mixed Forest & 36 & 60 & 73 & 7952 Shrub / Scrub & 35 & 56 & 70 & 7771 Grassland / Herbaceous & 49 & 69 & 79 & 8481 Pasture / Hay & 49 & 69 & 79 & 8482 Cultivated Crops & 67 & 78 & 85 & 8990 Woody Wetlands & 36 & 60 & 73 & 7995 Emergent Herbaceous Wetlands & 49 & 69 & 79 & 84

[table:curveNumberTable]

# Figures

[h] image [fig:digitizingCulverts]

1. http://relief.ersc.wisc.edu/wisconsinview/form.php [↑](#footnote-ref-24)
2. http://tools.crwr.utexas.edu/OptimizedPitRemoval/Optimized-Pit-Removal-V1.5.1.zip [↑](#footnote-ref-26)
3. http://www.mrlc.gov/nlcd06\_data.php [↑](#footnote-ref-28)
4. http://datagateway.nrcs.usda.gov/GDGOrder.aspx?order=QuickState [↑](#footnote-ref-30)
5. http://www.fsa.usda.gov/FSA/apfoapp?area=home&subject=prod&topic=clu [↑](#footnote-ref-32)
6. Farm Service Agency, Natural Resources Conservation Service, Rural Development [↑](#footnote-ref-33)
7. http://water.usgs.gov/GIS/huc.html [↑](#footnote-ref-35)
8. http://tools.crwr.utexas.edu/OptimizedPitRemoval/Optimized-Pit-Removal-V1.5.1.zip [↑](#footnote-ref-39)
9. http://dipper.nws.noaa.gov/hdsc/pfds/ [↑](#footnote-ref-43)
10. This value will vary depending on the cell size of your DEM and the type of hydrology in your area of interest. The default 50,000 was chosen by surveying a watershed in Southern Wisconsin and determining the number of pixels on a 3-meter resolution flow accumulation raster where hydrology transitioned to surface water. The same approach was used on a watershed in Northern Wisconsin using a 10-meter resolution flow accumulation raster and the flow accumulation threshold was determined to be 5,000. Assuming a linear relationship between cell size and flow accumulation threshold, we can derive a relationship using these two cases as a guide for setting the threshold: where is the flow accumulation threshold in thousands of pixels and is the cell size of the DEM in meters. [↑](#footnote-ref-48)
11. where is the pixel value of the pixel, is the mean of the raster dataset, and is the standard deviation of the raster dataset [↑](#footnote-ref-54)