Circuity Factor Documentation

James Jin

Created: 8/4/2025

Within this ArcGIS Pro Project, there are three tools. Each of these tools needs to be run in a specific order. Some tools can be run multiple times since they use the Random Python library. Use the tools in this order:

1. [Calculate Straight Line Distances](#617ajox9yn4a)

2.  [Circuity Factor](#xdlpl8hr2sp6)

3. [Bootstrap Analysis](#obw3qe7fh339) (Currently, not necessary to run.)

For each of the tools, there is an output directory parameter. For each tool, there will be CSV files and/or text files created inside of the output directory. The output directory can be manually set by the user or if the parameter is left blank, an output directory will be created.

**Purpose and description:**

The purpose of these tools is to find a circuity factor that estimates a quadratic relationship between straight line distance and shortest road distance from a harvest site (or FIA plot) to a sawmill.

The workflow for these tools is based on the approach used in "Market Definitions for Hardwood Timber in the Southern Appalachians" (Prestemon et al., 1999). There are some changes to the original methodology, particularly to how sampling is done and the types of sawmills used. For the Circuity Factor tool, Neyman allocation was used to determine optimal sample sizes for a sawmill type, with a sample size floor of 30. In these tools, the sawmill types used are:

* Lumber/Solid Wood
* Pellet
* Chip
* Pulp/Paper
* Composite Panel/Engineer Wood Product
* Plywood/Veneer

The basic workflow finds straight-line distances from each harvest site to each type of sawmill. Then, at least 30 (can be set to more) road distance samples are calculated for each sawmill type. Using these samples, an optimal sample size is calculated. If it is greater than 30 (or whatever the user sets it at), the sample size is increased to the value that was calculated and the remaining distance values are calculated. Then using the samples for each type, Ordinary Least Squares (OLS) regression is used to find a circuity factor for each sawmill type and for the total combined samples. Code for relevant algorithms will be included at the bottom.

**Important:** Any feature class input data should be put inside the project’s File GDB

## 1. Calculate Straight Line Distances

Description:

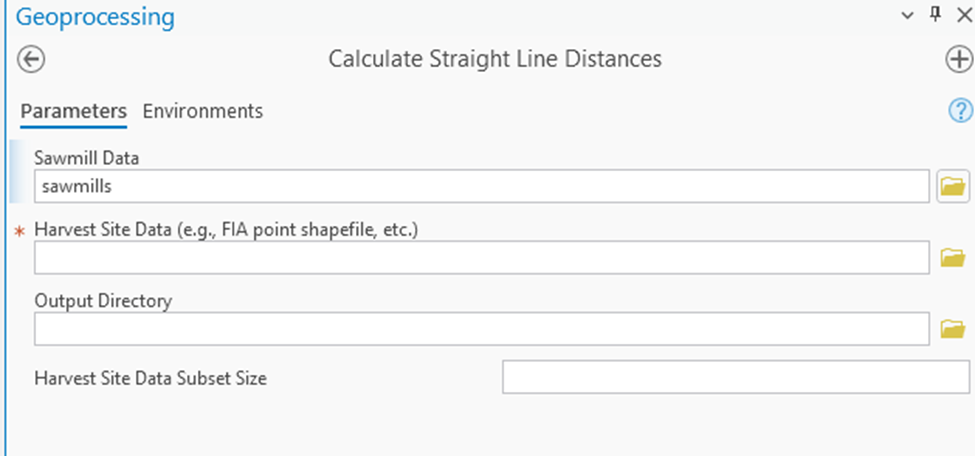
This script tool calculates the straight line distance (Euclidean distance) for every harvest site point to [every type of sawmill](#v1gctea9t5te). The maximum distance to search for is 120 miles. This script only needs to be run once for any given set of harvest site data and sawmill data. The result will be the same unless inputs are changed.

**Outputs:**

* CSV file containing straight line distances. There will be a message that gives the file path to the directory the CSV is in. The CSV file will have date and time in its name. Columns are:
  + Sawmill Type
  + Harvest Site ID
  + Sawmill ID
  + Straight Line Distance

**Inputs:**

* Sawmill Data
  + points feature class
* Harvest Site Data
  + can be points or polygons
* Output Directory
  + Optional, will create one if left blank
* Harvest Site Data Subset Size
  + optional, will not be used if left blank
  + Randomly selects a number of sites from the harvest site data equal to the subset size input



## 2. Circuity Factor

Description: Uses Neyman Allocation to select sample sizes for each sawmill type. For each sawmill type, a number of site-to-sawmill pairs, equal to the calculated sample size (or user set sample size, default 30), is selected for road distance calculations. Once road distances are calculated for each type, circuity factor is calculated using OLS regression. Can select for a specific sawmill type to be calculated or for every sawmill type to be calculated at once.

This script can be run multiple times to get different results as both sample size calculations and final road distance calculations are reliant on random sampling.

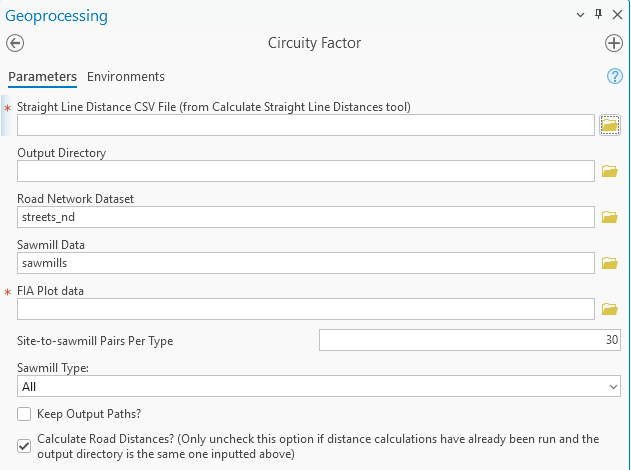
**WARNING:** This script tool may crash when run in ArcGIS Pro. For better stability, run this tool one time for each type, then once more for all types with the “Calculate Road Distances?” option unchecked. More detailed explanation in the [Recommended Usage](#yaiwj2i085y7) Section.

**Outputs:**

* There will be a message once the script tool completes that gives the file path of the directory that contains the output files. The directory name contains date and time.
* CSV files contain distance calculations. Columns are:
  + Harvest Site ID
  + Sawmill ID
  + Straight Line Distance
  + Road Distance
* Circuity factor summary text files. One for each type and one for the combined samples.

**Inputs:**

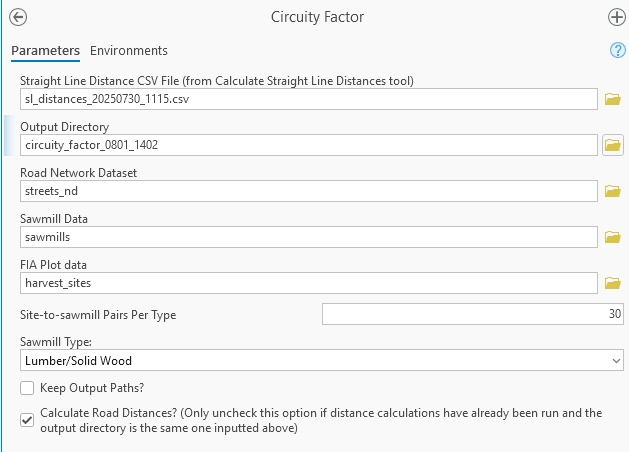
* Straight Line Distances CSV file (output from [tool 1](#617ajox9yn4a))
* Output Directory
  + Optional, will create one if left blank
* Network Dataset
  + Called “streets\_nd” in the project geodatabase
* Sawmill Data (same data in [tool 1](#617ajox9yn4a))
* Harvest Site Data (same data in [tool 1](#617ajox9yn4a))
* Site-to-Sawmill Pairs Per Sawmill Type
  + Default set to 30, not recommended to go lower
  + Sets the minimum sample size for sampling
* Sawmill Type
  + A list of options of [each sawmill type](#v1gctea9t5te) and an option for all sawmill types
  + Selecting “All” will do the complete calculation, including sampling, individual circuity analysis, and combined circuity analysis
  + Selecting a single type will only do the sampling and circuity analysis for that type
* Keep Generated Site-to-Sawmill paths?
  + Default unchecked
  + A line feature class for the route path is created with each road distance calculation. By default, they are deleted after creation. If the user desires to keep the path feature classes, check this option.
* Calculate Road Distances?
  + Default checked
  + By default, the script tool calculates road distances every time. If previously calculated road distances are desired to be used for recalculating circuity factor, uncheck this option.
  + If this option is unchecked, an existing output directory must be provided with the necessary road distance CSV files in it.



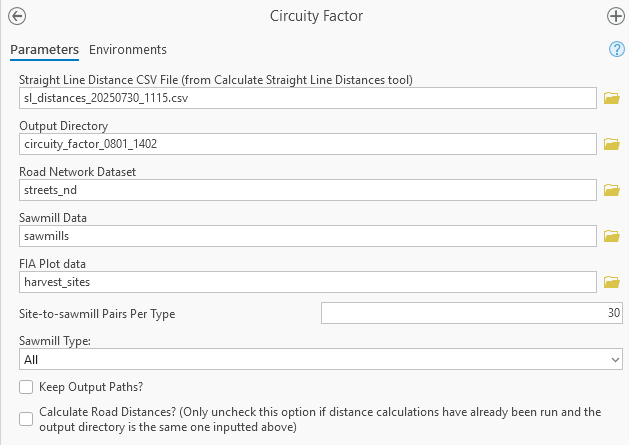
**Recommended Usage in ArcGIS Pro**

This script is prone to crashing or locking up in ArcGIS Pro. Crashes will cause ArcGIS Pro to close and a “[Send Report](https://pro.arcgis.com/en/pro-app/latest/get-started/report-software-errors-and-improvements.htm)” window to open up. Occasionally the tool can “lock up” where it simply runs on forever. While the script tool runs, there will be messages sent by the tool updating the user on its progress. If the message hasn’t been updated for 5 minutes and the tool hasn’t completed running, it is likely the tool has locked up and will not complete. It may be necessary to completely close ArcGIS Pro.

To best avoid this from happening, it is recommended to run this tool for each of the 6 [sawmill types](#v1gctea9t5te) like so:



The first tool run does not need an output directory but the other 5 runs need to have the same output directory as the first run. Alternatively, the outputs can be moved into one directory. Then after each of the 6 sawmill types are completed, run the Circuity Factor tool once more **with “Sawmill Type” set to “All” and “Calculate Road Distances?” unchecked** like so:

****

It also helps to close the Map View in ArcGIS Pro while running this script tool.

## 3. Bootstrap Analysis

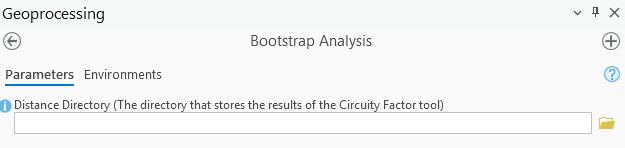
Description: This tool repeatedly samples from the same samples calculated in [tool 2](#xdlpl8hr2sp6) by randomly selecting samples with replacement. Instead of sampling from pairs of straight line and road distance, this tool converts these two values into a multiplier value (road distance / straight line distance. The multipliers for a given sample are averaged. Each sawmill type that is sampled gets its own average multiplier and the combined sawmill type sample also gets its own average multiplier. These values for each of the repeated resampling is output to a CSV file. The resampling occurs 10000 times.

**Outputs:**

* The CSV file is the only output file with the columns the same as the [sawmill types](#v1gctea9t5te), as well as a combined average column. The CSV file name contains date and time.

**Inputs:**

* Distance Directory
  + The output directory from [tool 2](#xdlpl8hr2sp6) where all of the road distance CSV files are stored.



## Calculations and Script Pseudocode

The formula for sample size (*n*) when estimating a population mean that was used:

calculate\_straight\_line\_distances.py:

Uses calculate\_sl\_distances() in distance\_calculator.py

DEF calculate\_sl\_distances(hv\_sites, sawmill\_data, sawmill\_types, output\_dir):

SET output CSV file path using date and time

CALL open() on CSV file, write enabled

Create SearchCursor on hv\_sites, "OBJECTID" for fields

FOR each entry in search cursor for hv\_sites:

Select harvest site based on OBJECTID

CALL Near() from each sawmill in sawmill\_data to selected harvest site

with 120 mile search radius

FOR each sawmill\_type in sawmill\_types:

GET all sawmills of sawmill\_type and near distance > 0

GET sawmill id and near distance of the sawmill with the lowest near distance

Write out sawmill\_type, harvest site ID, sawmill ID, and near distance to output CSV file

ENDFOR

ENDFOR

ENDDEF

circuity\_factor.py:

Algorithm for sample and Neyman Allocation:

Create dictionary to store straight line distances and ids, keys being sawmill types, values being dictionaries with keys as ids and values as straight line distances

Read into dictionary from straight line distances CSV file

SET pairs\_per\_type to 30 (default value)

SET z = 1.96 and E = 0.1

FOR each sawmill type in SL distances dictionary:

CALL open() on output CSV file

GET a randomized list of object IDs from SL distances dictionary

SET sample\_size to pairs\_per\_type

SET count to 0

FOR each ID in randomized ID list:

IF count == pairs\_per\_type:

CALCULATE standard deviation

CALCULATE n = (z\*\*2 \* std\_dev\*\*2) / E \*\* 2

Round n up

IF n > sample\_size:

SET sample\_size to n

ENDIF

ENDIF

IF count is equal to sample\_size:

CALL break

ENDIF

TRY:

Select harvest site using Random ID

Select sawmill using sawmill ID from SL distances dictionary

CALCULATE road distance using the harvest site and sawmill

Write both IDs, straight line distance, and road distance to CSV

IF road distance is greater than 120:

RAISE exception

ENDIF

EXCEPT failure to find a route or road distance too long:

IF not last ID in random ID list:

CALL continue

ELSE:

CALL break

ENDIF

FINALLY:

DELETE all temporary layers, datasets, and feature classes

INCREMENT count by 1

ENDFOR

ENDFOR

This line: CALCULATE road distance using the harvest site and sawmill

calls to a user defined function that uses ArcGIS Network Analyst to find the shortest path to the sawmill.

References

Prestemon, J. P., Pye, J. M., Abt, K. L., & Wear, D. N. (1999). Market definition for hardwood timber in the Southern Appalachians. In *P. 91-98 in LA. Munn, SH Bullard, SC Grado, and DL Grebner (eds.), Proceedings of the 1999 Southern Forest Economics Workshop, April 18-1999, Biloxi, MS.*.