

Summed Point Influence tool version 1.0 beta

Kim Fisher (kfisher@wcs.org) and Karl Didier (kdidier@wcs.org)

12.24.2011

The summed Point Influence (SPI) tool estimates cumulative influence on an area from a set of geographic features. It is meant as a simple, transparent, and widely applicable tool requiring a minimum of data inputs, GIS experience, or preprocessing. As such, it does not claim to model reality precisely for any particular application; more precise models will necessarily require more sophisticated tools and more detailed (and expensive!) input data. Nevertheless, our premise is that SPI offers useful results for a minimal investment.

Assumptions

SPI originated with a need to model the amount of hunting by local town residents across areas where WCS works. Gosia Bryja and Karl Didier wrote a script in Avenue Macro Language (AML) for ArcInfo that added the results of a Cost Distance calculation weighted by population for each town to approximate hunting pressure; SPI builds on this basic concept.

SPI depends on three basic assumptions, presented here in terms of hunting pressure but extensible to other scenarios involving the cumulative influence of a set of discrete source features on a region. Other use cases might include the spread of bees from a set of hives to pollinate a landscape, the effect of water or air pollution emanating from a set of point sources, or a model of patrol effort across a landscape by park rangers starting from patrol posts. The assumptions are:

1. Places near towns, in terms of the time a person requires to travel there from their town, tend to be hunted more than places farther away from towns. This is the basic concept behind 'cost distance'—linear ("as the crow flies") distance combined with a cost surface quantifying the relative ease/speed of travel in each location. Walking or driving on a road most of the way to a place will take less time than walking through forest.
2. Places near large towns tend to be hunted more than places near small towns. More generally the idea of a 'weight' attached to the influence features, this is typically a population attribute of the towns, which could be total population, population of known hunters, population of males 15-50 years old in a certain socioeconomic stratum, etc.
3. Places close to many towns tend to be hunted more than places close to only a few. This is the assumption that requires summing the calculated influence of each feature individually, rather than basing the result on a single cost distance calculation, which measures cost distance to the *nearest* feature only, not all features cumulatively.

Installation

Summed Point Influences requires ArcGIS 9.3 or above at an ArcView license level or above, and the Spatial Analyst extension. The most updated version of the tool is available for download from <http://conservationsupport.org/WhatWeDo/GISSpatialAnalysis/tabid/3735/Default.aspx>. After downloading the zip file, follow these steps to use the tool:

1. Unzip the file into a location on your machine that you can find later. If you don't have any directory already set up for managing scripts and tools, use the My Toolboxes location for your machine:

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=The_My_Toolboxes_folder_-_where_toolboxes_are_created

You should now have a directory called summedPointInfluences with a bunch of files inside. To run the tool, you need all these files, in the right places.

2. In ArcGIS, create or open an mxd for your project, and then add the Summed Point Influences toolbox to your list of toolboxes:

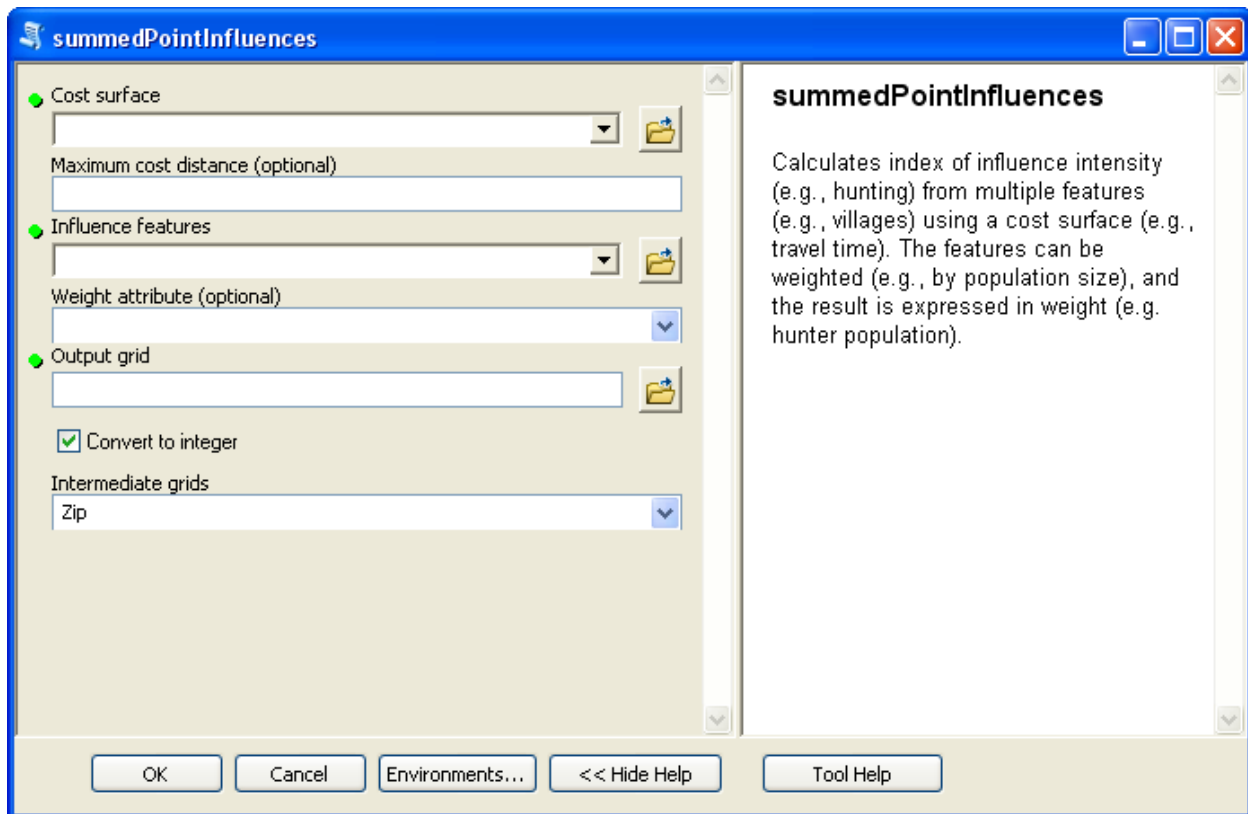
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Basic_toolbox_management

Note that if you save your mxd after adding the toolbox, the toolbox will appear the next time you open the mxd. You can also right-click on ArcToolbox > Save Settings > To Default. Otherwise, you will have to add the toolbox again.

3. Open the toolbox and double-click the Summed Point Influences tool to run it. Make sure the Help pane in the right of the tool dialog is expanded, so that you can see brief descriptions of the inputs and requirements.

Inputs

Summed Point Influences was designed to be run with the minimum input preparation required to operationalize the above assumptions. Each input to the tool is explained below.



As with any tool, it is best to make sure all your input datasets are in the same projection. In addition, SPI requires that all your data is in a projected coordinate system with a meaningful linear unit (i.e., meters, feet, etc.).

Cost surface

The input cost surface represents the relative cost or impedance of traveling planimetrically through each cell. For simplicity and because the cost is often measured in minutes, the cost value in each cell should be for the cell, and not per unit distance as the ArcGIS Cost Distance tool requires (e.g., if the cost raster cell size is 1 km, the cost units might be the number of minutes required to travel 1 km).

All cells in the raster must be positive (i.e. no negative or zero costs), and the maximum cost allowed is 10000. To avoid excessive run times, the tool limits the number of cells (rows * columns) to 100000000, or ~10000 x 10000. Consider reducing cell size if your cost surface has more than this many cells. Cell locations with NoData in the input cost raster act as barriers; any cell location that is assigned NoData on the input cost surface will receive NoData on all outputs. The input cost raster can be integer or floating-point, and in any compatible raster format (ESRI GRID, ERDAS Imagine, and TIFF have been tested).

The input cost surface is typically the most time-consuming input to prepare, involving various combinations of Spatial Analyst tools using land cover, transportation, hydrology, and border data. Cost values should accurately reflect local conditions, as measured by informant interviews or direct evidence. For cost surfaces reflecting human travel times, remember that similar features can have different effects on travel times in different parts of the world; for example, water features can speed

travel in landscapes dominated by navigable rivers, or be serious obstacles (high cost or even NoData barriers) in landscapes with a few non-navigable streams or lakes. Similarly, political borders might impose high travel costs in well-patrolled places but no travel costs in unmonitored backcountry areas. The cost surface should thus ideally assign realistic values to the most comprehensive range possible of particular local travel conditions.

As a guide, for a list of the kinds of data that might influence how a hunter moves through a landscape, and reference values for quantifying costs, see “Friction surface components” at <http://bioval.jrc.ec.europa.eu/products/gam/sources.htm>, by Andrew Nelson at the Global Environment Monitoring Unit of the European Commission’s Joint Research Centre.

Maximum cost distance

The optional maximum cost distance input represents the maximum total accumulated cost value from each input feature to be considered. For example, if your cost is measured in minutes (to cross a cell), and your influence is a hunter population that does not travel more than 8 hours away from a village to hunt, you could enter 480 ($8 * 60$ minutes) here. For some applications, limiting maximum cost distance may not be appropriate; usually, however, there is some upper bound beyond which the influence from input features cannot reasonably be measured. Recall the first assumption of the tool, which is that areas farthest away are areas of least influence. All areas beyond the maximum cost distance will receive an influence value of 0 in the tool results.

To understand better how maximum cost distance affects results requires a brief summary of the tool’s logic. First, a cost distance for each input feature is calculated, with a minimum value of 0 at the feature, and a maximum value of the greatest cost distance (e.g., travel time) within the cost surface. If a maximum cost distance is specified for the tool, the largest of all features’ cost distance maximums will be the specified maximum cost distance value. If a maximum cost distance is not specified for the tool, the largest of all features’ cost distance maximums will be whatever happens to be the largest maximum cost distance from an individual feature.

The difference is important because this value, the largest of all maximum cost distances, is used to normalize and invert each feature’s cost distance. The influence values depend on where they fall relative to this greatest maximum cost distance. They will tend to be higher at any given point if no maximum cost distance is entered into the tool, because they are relative to the maximum cost distance (lowest influence) of the feature with the greatest cost distance.

Influence features

The second required data input, influence features represent the sources from which to measure influence, and can be either points, lines, or polygons. These are often human settlements from which some type of human influence—for example, hunting—emanates. To avoid excessive run times, SPI limits the number of features to 1000. Also, make sure that all your features fall within cells on the cost surface that have values.

The tool respects feature selections, so you can run it with only relevant points selected, or run it multiple times with different subsets of influence features and then add the results. For example, if people from towns with vehicles influence the landscape differently than people from towns without vehicles, you could run the tool twice, once using a selection of towns with vehicles and using a cost surface appropriate to vehicles, and once with towns without vehicles and using a cost surface appropriate to foot travel.

Weight attribute

The influence features can be treated equally, if you have no proxy attribute for the relative amount of influence from each feature. If this is the case, leave the weight attribute field blank. Otherwise, use the pulldown menu to specify the field containing your features' influence weight, often some measure of population. If no field is chosen, the influence features will all be considered to have equal weight, and the unit of the results will be the number of influence features with access to each cell. The mean of the weights cannot exceed 100000; if your weight (e.g. population) exceeds this limit, consider creating a new field equal to some order of magnitude less (e.g., divide by 1000).

When modeling human influence, remember to be specific about what your population values mean. If you have nothing more specific than total population, then your results will be a similar proxy. If you can estimate the number of people relevant to the analysis (rather than the total), you will get results with realistic numbers of people with access. Ideally you would have the actual hunter population of each town as measured by household surveys; failing that, you could try to estimate the number of hunters from demographic data (males 15-50 years old, e.g.). Similarly, if you are measuring the combined effects of sport and subsistence hunting from different socioeconomic strata, estimate the number of each in different attributes, run the tool twice using these attributes, and then add the results.

Other factors can similarly be handled with data preprocessing. For example, a common factor in hunting analyses is distance to market: towns close to market towns will be likely to field a greater proportion of their population, or to send the same number of hunters out more often. You might handle these cases by first finding the distance of each town to the nearest market, and then applying a proportionate multiplier to each town's total population to achieve either more realistic approximations of hunter numbers or a number expressed in person-hours (rather than people).

Output grid

Specify the location and name of the output raster. The path cannot have any spaces, and remember that ESRI GRID raster names cannot have more than 13 characters. Other formats (TIFF, Imagine) are also acceptable.

Convert to integer

Check to make the output an integer raster. Uncheck to make the output a floating-point raster. For most likely applications of SPI, the increased precision of a floating-point result will not be warranted by the precision of the input data; nor will the units be as useful (e.g., fractions of people). Furthermore,

integer rasters are much smaller. Nevertheless, the floating-point option is included for unusual applications.

Intermediate grids

During execution, SPI generates an influence raster for each feature. These intermediate influence rasters can be illuminating for interpretation of results—i.e., examining the contribution of any single feature to the overall influence results. However, they can also take up large amounts of disk space. Use this pulldown to choose what to do with the intermediate influence rasters generated by the script:

- "Zip" (default) will compress the intermediate rasters into a zip file in the same directory as the output raster
- "Retain" will retain the intermediate rasters, uncompressed, in the same directory as the output raster
- "Delete" will remove the intermediate rasters altogether

If "Delete" is not chosen, intermediate rasters are stored in a directory or zip file named 'SPIworking' (or SPIworking1, SPIworking2, etc.) in the same directory specified for the output grid (or in the directory containing the geodatabase specified for the output grid).

Results

SPI reports results in the same unit used for the weight of the influence features. If the weight used is hunter population, then the modeled output will show the number of hunters that have access to and could potentially hunt in each location. In a sense, the output reflects the 'risk' that an animal that people prefer to hunt would have of being shot everywhere in the landscape. If no input weight is used, then the output units will likewise not reflect weight; the results will show the number of input features (e.g., towns, rather than hunters in the towns) with access to each location.

When used to model hunting pressure, SPI makes two important assumptions about how hunters behave that, in reality, are probably not true but are difficult to incorporate into models. First, the model assumes that the amount of hunting in a place is not dependent on the numbers of animals hunters believe they will encounter. In other words, the model assumes that when a person decides where to go hunting, the only thing they consider is how long it will take to get there, and not how many animals they will find. In reality, people are constantly collecting information, based on their previous experiences and what other people say, about how many animals are in which places, and using this information to decide where to go.

Second, SPI currently assumes that the amount of hunting that takes place in a particular location is *linearly* related to the amount of time it takes to get to that place from a village. This assumption could very well be false for a particular landscape. It might be the case that most hunters prefer to stay close to home, while a few very dedicated hunters are willing to go a very long distance on their hunting trips. The ability to specify a different normalization function is under consideration for future versions of SPI (see below); in the meantime, similar results may be achieved by considering different groups of hunters as different weights to the influence features, and running the tool for each group separately using different maximum cost distance values each time.

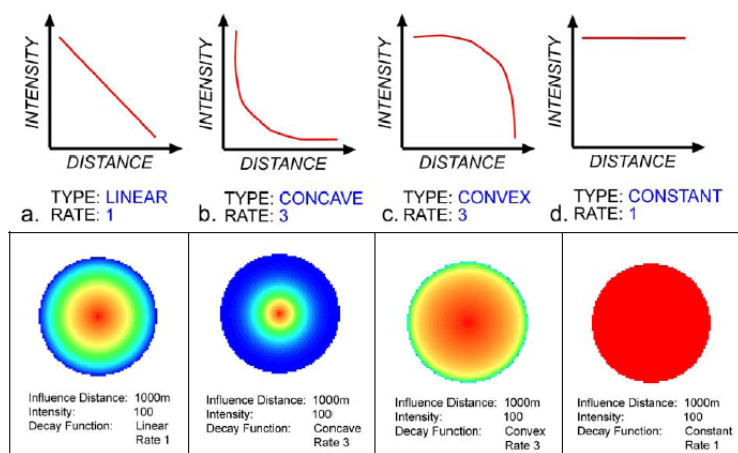


Possible future enhancements

Please report all bugs and comments on this version 0.7 of SPI to Kim Fisher (kfisher@wcs.org). In addition, we are considering a version 2.0 of the tool, if time allows. Please let us know if these enhancements would be useful enough to you to warrant further development, or if you have other enhancements you would like to see:

- *Arcpy-only version:* With some inner refactoring, the tool could be rewritten to use the arcpy module introduced with ArcGIS 10. This might make the tool faster; it would however mean dropping support for versions of ArcGIS prior to 10.

- *Choice of normalization functions:* As above, an ability to specify different, non-linear normalization functions to control the relationship of cost distance to influence might produce more realistic results. The diagrams at right (reproduced from PAT_v3_Tutorial_6Aug09.pdf) show the options offered by TNC's Protected Area Tools suite, and could serve as a



Examples of distance decay functions and the resulting output based on a single point occurrence and a 1000m influence distance. Red and orange hues indicate higher intensity values while green and blue hues represent lower intensity values.

model.

- *'Uphill' incorporation:* The individual feature cost distance calculations could also output a backlink raster. The tool could offer an optional slope raster input that could be used in conjunction with this backlink raster to further alter the individual influence raster by weighting cells that are uphill or downhill as measured from the influence feature. However, this additional calculation for every feature would add significant computing time; also, since many applications of SPI (such as hunting pressure) require consideration not only of outbound travel but return trips, it is unclear how useful this would be.