Understanding analysis in Spatial Analyst

The easiest way to understand cell-based modeling is from the perspective of an individual cell (the worms eye approach) as opposed to the entire raster (the birds eye approach). To do so, think of yourself as a cell in a raster dataset. You represent a location, and you have a value. All ArcGIS Spatial Analyst operators and functions will ask you to manipulate or retain your value based on a defined series of rules.

Calculating an output value for each cell

To calculate an output value for your specified location (cell) using any Spatial Analyst operation or function, there are three things you need to know:

- The value of your specified location (cell)
- The manipulation of the operator or function
- Which other cell locations and their values to include in your calculations

How do you determine these three things?

- You automatically know what the value is for your location (your input cell value).
- Each operator and function in Spatial Analyst manipulates the value at your location in different ways. Based on the operator or function being applied, from knowledge built into Spatial Analyst, you will know how to manipulate your value.
- With some Spatial Analyst operations and functions, you can calculate an output value by knowing only the value of your location, such as raising your value by a specified power (a local function). To complete other operations and functions, you need to know the values of other locations within the raster dataset to which your specified location belongs, such as looking in a neighborhood around you (a focal function) or including cell locations and their values defined by other raster datasets (such as zonal functions).
- This three-step process occurs for each location (cell) in the raster dataset within any Spatial Analyst function. All operators and functions work on a cell-by-cell basis, and each calculation for each cell requires the value of the cell, the manipulation that is being applied, and other cell locations to include in the calculations. The Spatial Analyst operators and functions are grouped into categories based on how they manipulate values—you only need to understand how the cell values are manipulated in the different categories.
- For many functions, you can refine how the manipulation (the calculations) will be performed through user-defined parameters. For example, the cells to include in each calculation for a focal function may vary based on the neighborhood that is defined.

The three-step process considering several functions:

- Cos function: When applying the Cos function to your raster dataset, you must know the value of your specified location and how to take the cosine of that value to return an output value.
- Neighborhood statistics function: When applying this function (a focal function) to determine the maximum value within a 3-by-3 neighborhood, you must know your location's value and the values of the eight immediate neighbors around your location. You will assign the highest of the nine values to your location in the output raster dataset.
- Zonal statistics function: When applying this function with the mean option, you must know your location's value and you must take the mean of all the values of the cells that belong to the same zone as your cell which is defined by a zone raster dataset.
- Addition operator: When applying this operator to your raster dataset and to two other raster datasets, you must add your location's value to the values of the same location your cell represents in the two other raster datasets to return an output value for your location.
- Euclidean Distance function: When applying this function, you must know your location, and determine how far your location is from the closest source (which is defined by a source dataset) to return an output value for your location.

The above logic occurs for each cell in the raster.

# Operators and functions of Spatial Analyst

The functions associated with raster cartographic modeling can be divided into five types:

- Those that work on single cell locations (local functions)
- Those that work on cell locations within a neighborhood (focal functions)
- Those that work on cell locations within zones (zonal functions)
- Those that work on all cells within the raster (global functions)
- Those that perform a specific application (for example, hydrologic analysis functions)

Each of these categories can be influenced by, or based on, the spatial or geometric representation of the data and not solely on the attributes that the cells portray. For example, a function that adds two layers together (via single cell locations) is dependent on the cell's location and the value of its counterpart in the second layer. Functions applied to cell locations within neighborhoods or zones rely on the spatial configuration of the neighborhood or zone as well as the cell values in the configuration.
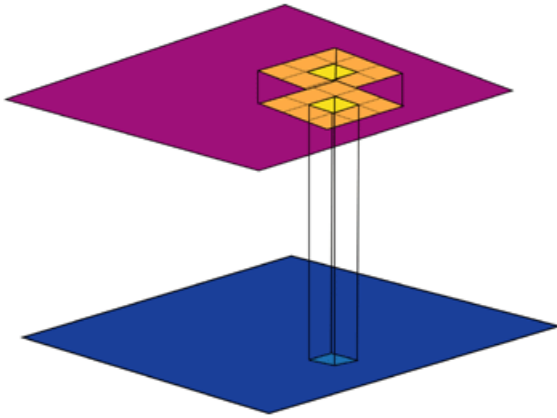
Local functions

Local functions, or per-cell functions, compute a raster output dataset where the output value at each location (cell) is a function of the value associated with that location on one or more raster datasets. That is, the value of the single cell, regardless of the values of neighboring cells, has a direct influence on the value of the output. A per-cell function can be applied to a single raster dataset or to multiple raster datasets. For a single dataset, examples of per-cell functions are the trigonometric functions (for example, sin) or the exponential and logarithmic functions (for example, exponential log).
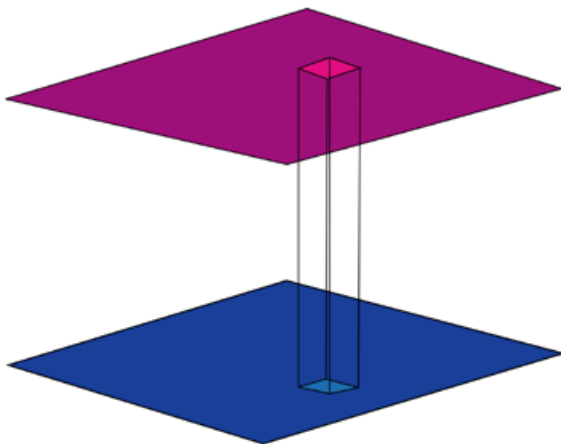
Examples of local functions that work on multiple raster datasets are functions that return the minimum, maximum, majority, or minority value for all the values of the input raster datasets at each cell location.
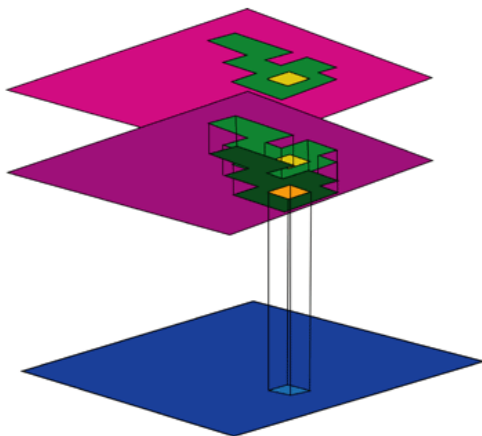
Focal functions

Focal, or neighborhood, functions produce an output raster dataset in which the output value at each cell location is a function of the input value at a cell location and the values of the cells in a specified neighborhood around that location. A neighborhood configuration determines which cells surrounding the processing cell should be used in the calculation of each output value.

Neighborhood functions can return the mean, standard deviation, sum, and range of values within the immediate or extended neighborhood.

Zonal functions

Zonal functions compute an output raster dataset where the output value for each location depends on the value of the cell at the location and the association that location has within a cartographic zone. Zonal functions are similar to focal functions except that the definition of the neighborhood in a zonal function is the configuration of the zones or features of the input zone dataset, not a specified neighborhood shape. However, zones do not necessarily have any order or specific shapes. Each zone can be unique. Zonal functions return the mean, sum, minimum, maximum, or range of values from the first dataset that fall within a specified zone of the second.
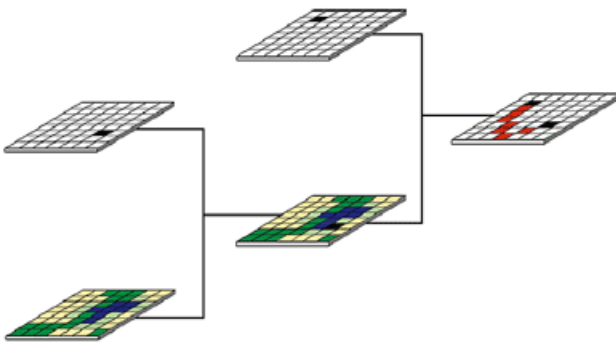
Global functions

Global, or per-raster, functions compute an output raster dataset in which the output value at each cell location is potentially a function of all the cells combined from the various input raster datasets. There are two main groups of global functions: Euclidean distance and weighted distance.

Euclidean distance global functions

Euclidean distance global functions assign to each cell in the output raster dataset its distance from the closest source cell (a source may be the location from which to start a new road). The direction of the closest source cell can also be assigned as the value of each cell location in an additional output raster dataset.

Weighted distance global functions

A global weighted-distance function determines the cost of moving from a destination cell (the location where you want to end the road) to the nearest source cell (the location where you want to start the road) over a cost surface (cost being determined by cost schema, such as cost of construction). To take this one step further, the shortest or least-cost path over a cost surface can be calculated over a non networked surface from a source cell to a destination cell using the global least-cost path function. In all the global calculations, knowledge of the entire surface is necessary to return the solution.

Application functions

There are a wide series of cell-based modeling functions developed to solve specific applications. An application function performs an analysis that is specific to a discipline. For example, hydrology functions create a stream network and delineate a watershed. The local, focal, zonal, and global functions are general functions and are not specific to any application. There is some overlap in the categorization of an application function and the local, focal, zonal, and global functions (such as the fact that even though slope is usually used in the application of analyzing surfaces, it is also a focal function). Some of the application functions are more general in scope, such as surface analysis, while other application functions are more narrowly defined, such as the hydrologic analysis functions. The categorization of the application functions is an aid to group and understand the wide variety of Spatial Analyst operators and functions. You may find that

a specific application function can manipulate cell-based data for an entirely different application from its category. For example, calculating slope is a surface analysis function that can be useful in hydrologic analysis as well.
Application functions include the following:

- Density analysis
- Surface generation
- Surface analysis
- Hydrologic analysis
- Geometric transformation
- Generalization
- Resolution altering

An overview of Spatial Analyst

ArcGIS Spatial Analyst provides a rich set of tools to perform cell-based (raster) analysis. Of the three main types of GIS data (raster, vector, and TIN), the raster data structure provides the most comprehensive modeling environment for spatial analysis.
Cell-based systems divide the world into discrete uniform units called cells, based on a grid structure. Every cell represents a certain specified portion of the earth, such as a square kilometer, hectare, or square meter. Each cell is given a value to correspond to the feature or characteristic in which it is located, or describes the location, such as an elevation value, soil type, or residential classification. In a cell-based system, geographic location is not defined as an attribute but is inherent in the storage structure, known as the locational perspective.
The locational perspective allows ArcGIS Spatial Analyst to store continuous data (for example, elevation, oil concentration, and sound) more effectively. In continuous data, each location has a quantity, magnitude, or intensity assigned to it and the values are meaningful relative to one another. The locational perspective also allows for greater diversity in spatial analysis for both discrete (for example, land use and vegetation type) and continuous data, which will become apparent in the wide variety of discussions accompanying each toolset.

Toolset	Description
Color Models	Color modeling tools allow you to control the display of rasters and raster composites using the red, green, blue (RGB) and hue, saturation, value (HSV) color models. These tools allow you to define which raster will be assigned to each component of each model. That is, you can assign which rasters will represent the red, green, and blue colors in the RGB model and the hue, saturation, and value components in the HSV model. The group of rasters for each model can then be displayed in a composite.
Conditional	The conditional tools allow for control of the output values based on the conditions placed on the input values. The conditions that can be applied are either attribute queries or a condition that is based on the position of the conditional statement in a list. A simple attribute query might be: If a cell value is greater than 5, multiply it by 10; otherwise, assign 1 to the location.
Conversion	When you need your feature data to be converted into raster data, or if you have raster data that needs to be converted into another format, you can use the Conversion tools. When converting your data, make sure you use the appropriate conversion tool for the desired output format.

Density	By calculating density, you spread point values out over a surface. The magnitude at each sample location (line or point) is distributed throughout a landscape, and a density value is calculated for each cell in the output raster. For example, density analysis will take population counts assigned to town centers and distribute the people throughout the landscape more realistically.
Distance	There are two main ways to perform distance analysis in ArcGIS Spatial Analyst: Euclidean distance and cost distance. The Euclidean distance functions measure straight-line distance from each cell to the closest source, (The source identifies the objects of interest, such as wells, roads, or a school.) The cost distance functions (or cost-weighted distance) modify Euclidean distance by equating distance as a cost factor, which is the cost to travel through any given cell. For example, it may be shorter to climb over the mountain to the destination, but it is faster to walk around it.
Extraction	The extraction tools allow you to extract a subset of cells by either the cell's attributes or their spatial location. Extracting cells by attribute is accomplished through a where clause. For example, your analysis may require an extraction of cells higher than 100 meters in elevation from an elevation raster. You can also extract by a specified shape. For example, extract all cells that fall inside (or outside) a specified circle, rectangle, or polygon.
Generalization	Sometimes a raster dataset contains data that is erroneous or irrelevant to the analysis at hand or is more detailed than you need. For example, if a raster dataset was derived from the classification of a satellite image, it may contain many small and isolated areas that are classified incorrectly. The generalization functions assist with identifying such areas and automating the assignment of more reliable values to the cells that make up the areas.
Geometric Transformation	The Geometric Transformation tools are used to manage and manipulate the geometry of rasters. There are three main groups of Geometric Transformation tools: • those that change the geometry of the dataset through projections and georeferencing (geometric transformation) • those that change the orientation of the raster • those that combine several adjacent rasters into a single raster
Groundwater	The groundwater tools can be used to perform rudimentary advection-dispersion modeling of constituents in groundwater.

[Hydrology](#)	Hydrology functions simulate the flow of water over an elevation surface and creates either a stream network or a watershed.
[Interpolation](#)	Surface interpolation functions create a continuous (or prediction) surface from sampled point values. The continuous surface representation of a raster dataset represents height, concentration, or magnitude (for example, elevation, pollution, or noise). Surface interpolation functions make predictions from sample measurements for all locations in a raster dataset whether or not a measurement has been taken at the location.
[Local](#)	In a local function, the value at each location on the output raster is a function of the input values at the location. When computing a local function, you can combine the input rasters, calculate a statistic, or evaluate criteria for each cell in an output raster based on the values of each cell from multiple input rasters. For example, you can find the mean precipitation for a 10 year period or find in how many years the precipitation exceeded 0.5 meters.
[Math](#)	The ArcGIS Spatial Analyst provides a full suite of mathematical operators and functions. These operators and functions allow for the arithmetic combining of the values in multiple rasters, the mathematical manipulation of the values in a single input raster, the evaluation of multiple input rasters, or the evaluation and manipulation of values in the binary format.
[Multivariate](#)	Multivariate statistical analysis allows for the exploration of relationships between many different types of attributes. There are two main types of multivariate analyses available in Spatial Analyst: • supervised and unsupervised classification • principal component analysis (PCA) A third multivariate analysis, regression, is available in ArcGrid Workstation. Accompanying these analyses is a series of tools to evaluate each step in the analysis process. These tools can be used, for example, to predict the biomass (the dependent variable) at each location given the quantities of precipitation, soil type, aspect, and temperature (the independent variables).

Neighborhood	Neighborhood functions create output values for each cell location based on the value for the location and the values identified in a specified neighborhood. The neighborhood can be of two types, moving or a search radius. Moving neighborhoods can either be overlapping or nonoverlapping. Overlapping neighborhood functions are also referred to as focal functions and generally calculate a specified statistic within the neighborhood. For example, you may want to find the mean or maximum value in a 3x3 neighborhood. The nonoverlapping neighborhood functions, or the block functions, allow for statistics to be calculated in a specified nonoverlapping neighborhood. Search radius functions perform various calculations based on what is in a specified distance from point and linear features.
Raster Creation	The Value Creation functions create new rasters in which the output values are based on a constant or a statistical distribution. The Create Constant Raster function creates an output raster of constant values within a specified map extent and cell size. The Create Normal Raster function assigns values to an output raster so the values produce a normal distribution. The Create Random Raster (or the Map Algebra Rand) function randomly assigns values to cells on an output raster.
Reclass	Reclassifying your data simply means replacing input cell values with new output cell values. There are many reasons why you might want to reclassify your data. Some of the most common reasons are, replace values based on new information, group certain values together, reclassify values to a common scale (for example, for use in a suitability analysis or for creating a cost raster for use in the Cost Distance function), set specific values to NoData, or set NoData cells to a value. There are several approaches to reclassifying your data: by individual values, by ranges, by intervals or area, or through an alternative value.
Surface	You can gain additional information by producing a new dataset that identifies a specific pattern within an original dataset. Patterns that were not readily apparent in the original surface can be derived, such as contours, angle of slope, steepest downslope direction (aspect), shaded relief (hillshade), and viewshed.
Zonal	Zonal functions take a value raster as input and calculatesfor each cell some function or statistic using the value for each cell and all cells belonging to the same zone. The zonal functions are grouped by how the zones are specified, by a single input value raster or by a second zone raster. You can use the zonal tools to locate the number of endangered species (the value raster) within each parcel (the zone raster) or to find the area or perimeter of each zone in a raster.