Image from NASA:
https://commons.wikimedia.org/wiki/File:Earth_Western_Hemisphere_transparent_background.png#filelinks

Geospatial Problem Solving

Raster Analysis

WV VIEW

AmericaView
Empowering Earth Observation Education
AmericaView.org - Est. 2003

In this module, I expand upon the last module focused on vector-based spatial analysis by now exploring network analysis.

Module Topics

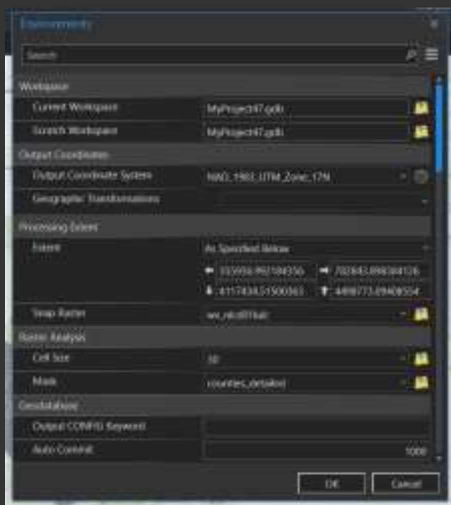
1. Processing environments
2. Extraction, mosaicking, and stacking
3. Pyramids
4. Reclassification
5. Moving windows, convolution, and focal statistics
6. Euclidean and cost distance
7. Raster math and logic
8. Conditional statements
9. Raster overlay
10. Categorical raster processing
11. Extract raster cell values at points
12. Summarize continuous raster data by zone
13. Summarize categorical raster data by zone
14. Density surfaces
15. Multidimensional raster data processing

2

These are the topics covered in this module.

Environments

Environment Settings



- Current Workspace
- Scratch Workspace
- Output Coordinate System
- Processing Extent
- Snap Raster
- Cell Size
- Mask

4

Before beginning a raster-based analysis, I would encourage you to define your environment settings. Defining these settings can make for cleaner analysis and save you time, as you will not need to define the settings every time you run a tool.

Setting a current workspace can minimize the need to search for files or search for folders or geodatabases to store outputs. If a current workspace is set, tools that honor the setting will look/go there first to find an input or save an output file. The scratch workspace will be the default location to save intermediate outputs or outputs that you do not intend to save permanently.

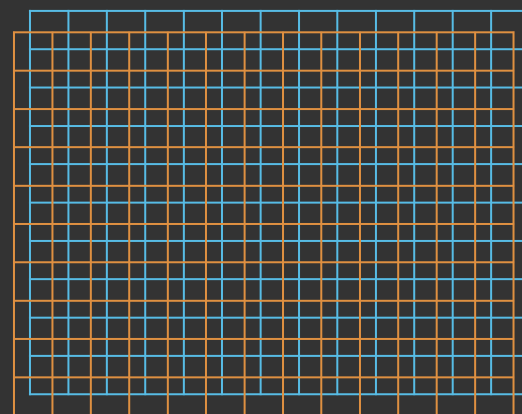
The output coordinate system defines the default coordinate system in which the output will be saved.

The processing extent defines the geographic extent over which an analysis will be performed. This is commonly defined relative to the extent of an input layer. Note that the processing extent does not cause the result to be masked to a polygon extent. It represents a rectangular extent. If you want to mask the result relative to an extent, you will need to define a mask.

You can also define a default cell size for the outputs. We will discuss the snap raster setting on the next slide.

Snap Raster

❖Makes sure output grid aligns with the selected snap raster



5

Setting a snap raster is used to force cell alignment. If a snap raster is defined, then the cells in the output will align with the cells in the defined snap raster.

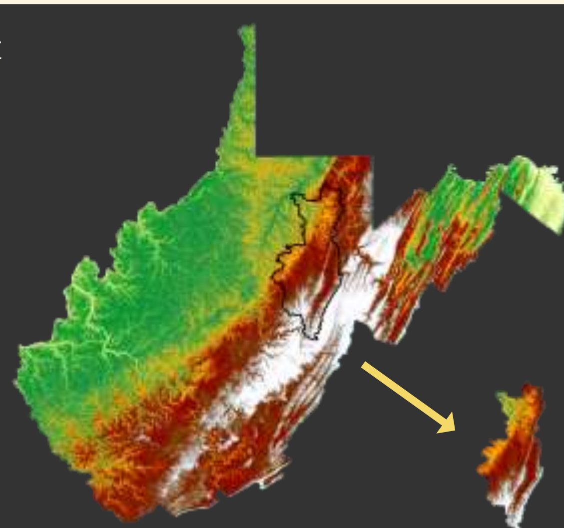
In the example above, the red and green grid cells do not align correctly. A snap raster could be used to force this alignment.

Extraction and Mosaic

6

Extraction Methods

- ❖Extract by Circle = extract cells that fall into a circle that is defined by a center coordinate and radius
- ❖Extract by Mask = extract cells that fall within a raster or polygon mask
- ❖Extract by Point = extract cells at point locations
- ❖Extract by Polygon = extract cells that fall within the extent of a polygon as defined by its vertices
- ❖Extract by Rectangle = extract cells that fall within a rectangular extent



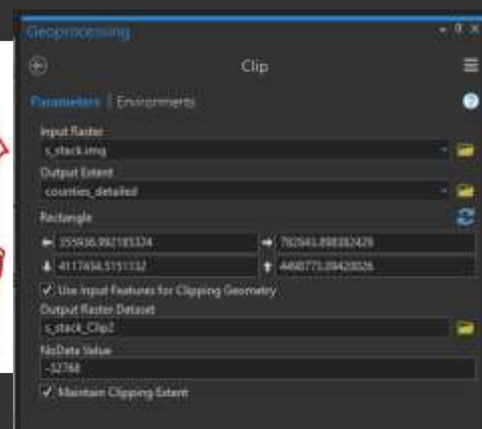
7

A common raster pre-processing task is to extract a portion of a raster from a larger dataset. In the example here, a portion of a DEM is being extracted from a larger DEM.

A variety of methods are available to extract a portion of a raster grid as defined on the slide. I commonly use Extract by Mask. To use this tool, a polygon mask or raster mask is used. Cells that fall within this mask will be written to a new file. That is how the example on the page was produced.

Take some time here to read through the descriptions of the different extraction methods.

Image Clip Tool in ArcGIS Pro



8

The Clip Raster Tool in ArcGIS Pro is also useful for subsetting or masking multiband raster data. Note that this is different from the Clip Tool used for extracting vector spatial data.

Mosaic to New Raster

- Combine single band or multiband raster grids into a single raster
- Can be used for all types or raster grids, including images
- Must specify file extension, pixel type, cell size, and number of bands

Instead of extracting a portion of a raster grid, you might be interested in combining multiple raster grids. This is a common task performed for image data. As another example, multiple digital elevation models could be merged to create an elevation surface that fills a larger spatial extent. This process can be accomplished using the Mosaic to New Raster Tool.

This tool will require you to define a pixel type (8-bit, 16-bit, float, etc.), a file extension (I commonly use .img or .tif), an output cell size, and the number of bands.

Mosaic to New Raster

❖ Assigning values at overlapping locations:

- FIRST = The output cell value of the overlapping areas will be the value from the first raster dataset mosaicked into that location.
- LAST = The output cell value of the overlapping areas will be the value from the last raster dataset mosaicked into that location. This is the default.
- BLEND = The output cell value of the overlapping areas will be a horizontally weighted calculation of the values of the cells in the overlapping area.
- MEAN = The output cell value of the overlapping areas will be the average value of the overlapping cells.
- MINIMUM = The output cell value of the overlapping areas will be the minimum value of the overlapping cells.
- MAXIMUM = The output cell value of the overlapping areas will be the maximum value of the overlapping cells.
- SUM = The output cell value of the overlapping areas will be the total sum of the overlapping cells.

NAIP Orthophotography

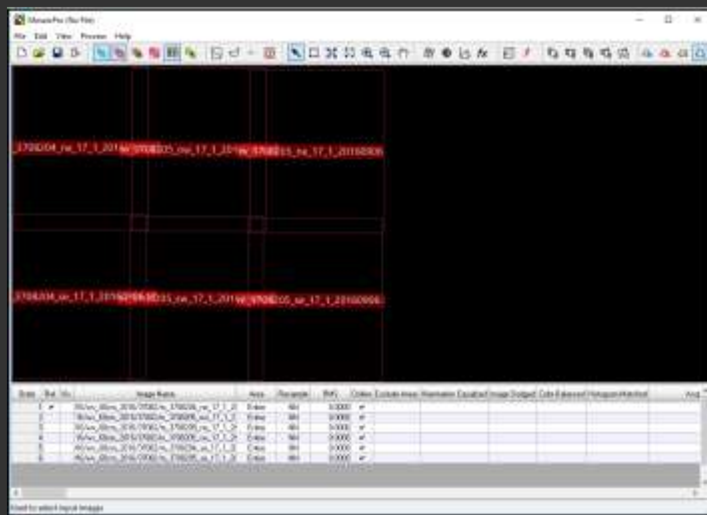
10

When you have overlapping values, a method must be applied to obtain the cell value in the overlapping areas. When there is no overlap, the decision is easy, as there is only one cell value.

For continuous data, many methods are available, including first, last, mean, maximum, and minimum. Since categorical data cannot be averaged, first or last are commonly used.

Take some time to review the methods described on the slide.

Mosaic Outside of ArcGIS

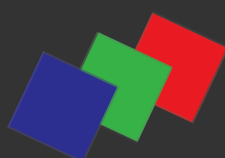


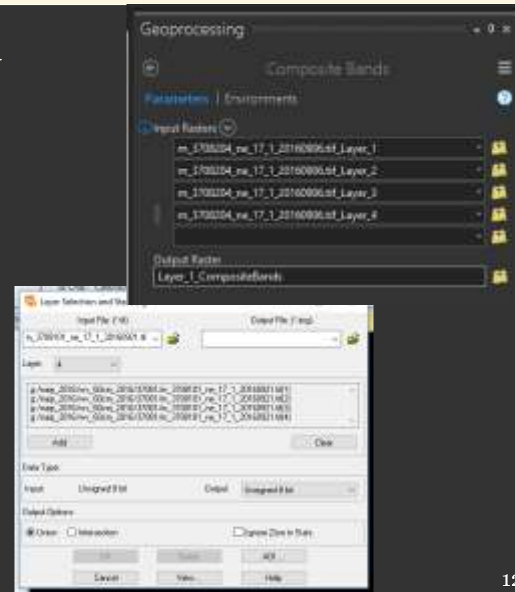
11

As a note, other tools are available to mosaic raster data. The Mosaic Pro Tool within Erdas Imagine is commonly used to mosaic image data. I have found this tool to be fast. It also offers more advanced color balancing methods that can be of great value when merging images. It also allows you to write the result to multiple files as opposed to a single, continuous dataset. This is valuable if the output will be very large.

Stack Grids

- ❖Stack or combine individual grids into a multiband raster
- ❖Commonly used for image data
- ❖Can be used for all raster data, not just images
- ❖Output bands or layers will all have the same pixel type
- ❖Bands will be layered in the order provided





12

You may have many raster grids that you would like to combine to a single file by stacking the separate grids or bands as a single multiband file.

As an example, when Landsat data are downloaded they are often provided as individual bands. However, you may want to use these data as a multiband image for analysis or symbology. So, you would need to combine the bands into a single raster with multiple bands.

This can be accomplished using the Composite Bands Tool in ArcGIS Pro or the Stack Tool in Erdas Imagine.

Note that it is possible to stack more than just image bands. For example, you could stack a set of topographic variables into a multiband file. I have found this to be useful when preparing data for modeling.

Note that all bands or layers in a stack must have the same pixel type. For example, if you stack three 8-bit grids and one 16-bit grid, all the data will be converted to 16-bit. If one of your input grids is float, then all your grids will be converted to float.

Pyramids

❖ Reduce spatial resolution versions of the data to aid in faster display

13

Raster datasets can be large files, especially if they cover a large spatial extent and/or have a small cell size. This can result in slow display.

Building pyramid layers provides a means to improve the display speed and performance of raster layers.

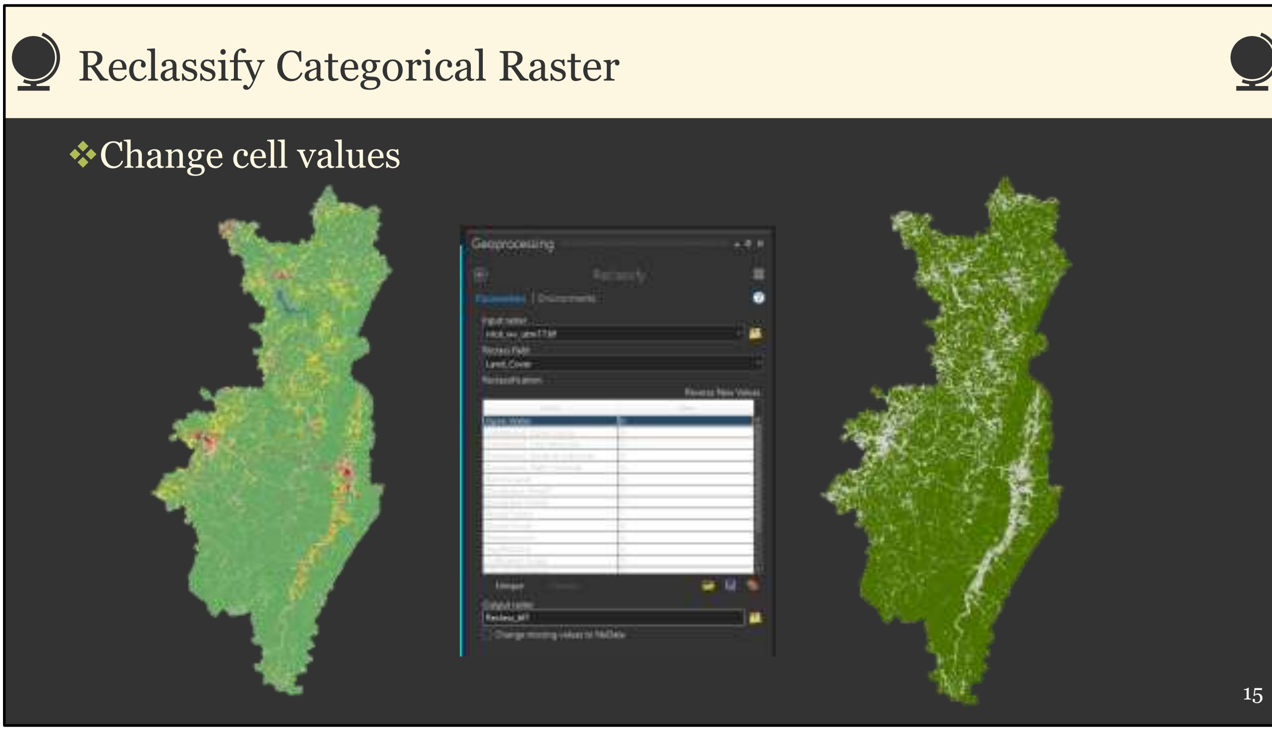
Pyramids are reduced spatial resolution versions of the data. When you are zoomed out, the reduced resolution data are drawn. The high spatial resolution data are not drawn until you zoom in.

This can greatly improve display speed as you zoom in and out or pan around.

If you plan to work with raster data, especially if you will need to zoom and pan around the data, such as when digitizing, I highly recommend building pyramids. This process can be slow, but it may save you time in the long run.

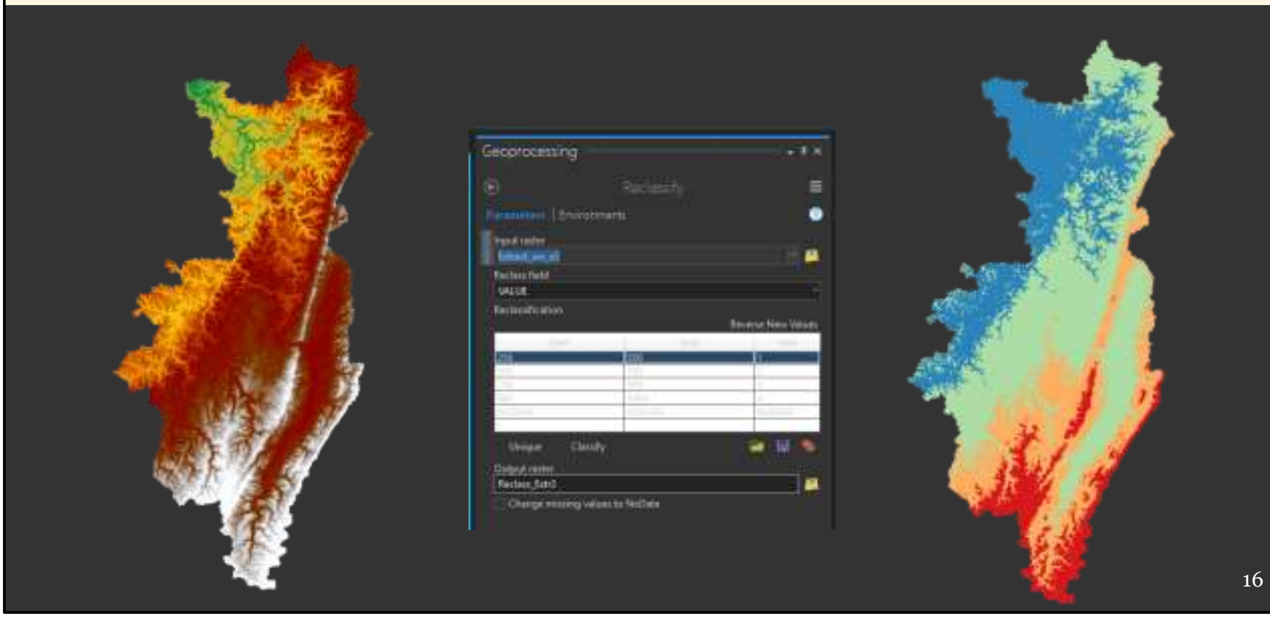
Reclassification

14



Reclassification allows you to change the cell values in a raster. The example on this slide demonstrates changing the cell values for a categorical raster grid, in this case a land cover grid from the National Land Cover Database. Here, all values corresponding to forest types were recoded to 1 and all other types were coded to 0. So, many categories were generalized to forest vs. not forest.

Reclassify Continuous Raster Data



16

It is also possible to reclassify continuous raster data. In this example, a digital elevation model is being reclassified into four categories. Each category represents a range of elevation measurements.

Reclassify NoData

- Can reclassify NoData to a new value

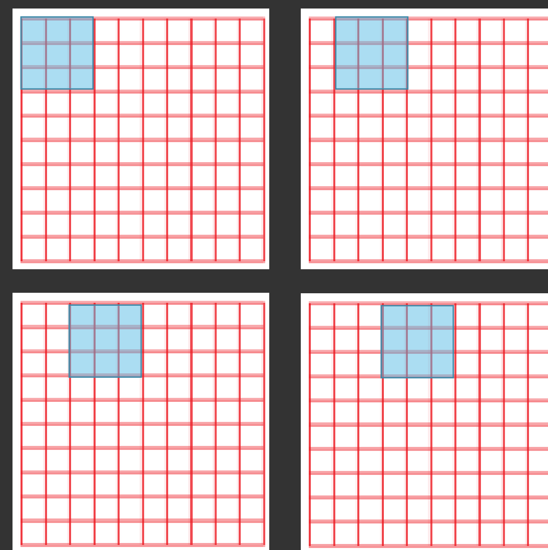
NoData cells can also be reclassified. In other words, other cell values can be reclassified to NoData or NoData can be reclassified to an actual value.

Neighborhood-Based Analysis

18

What is a neighborhood?

- Neighborhood analysis, moving window analysis, focal operation, kernel analysis
- Make calculation or compare cell values around a center cell

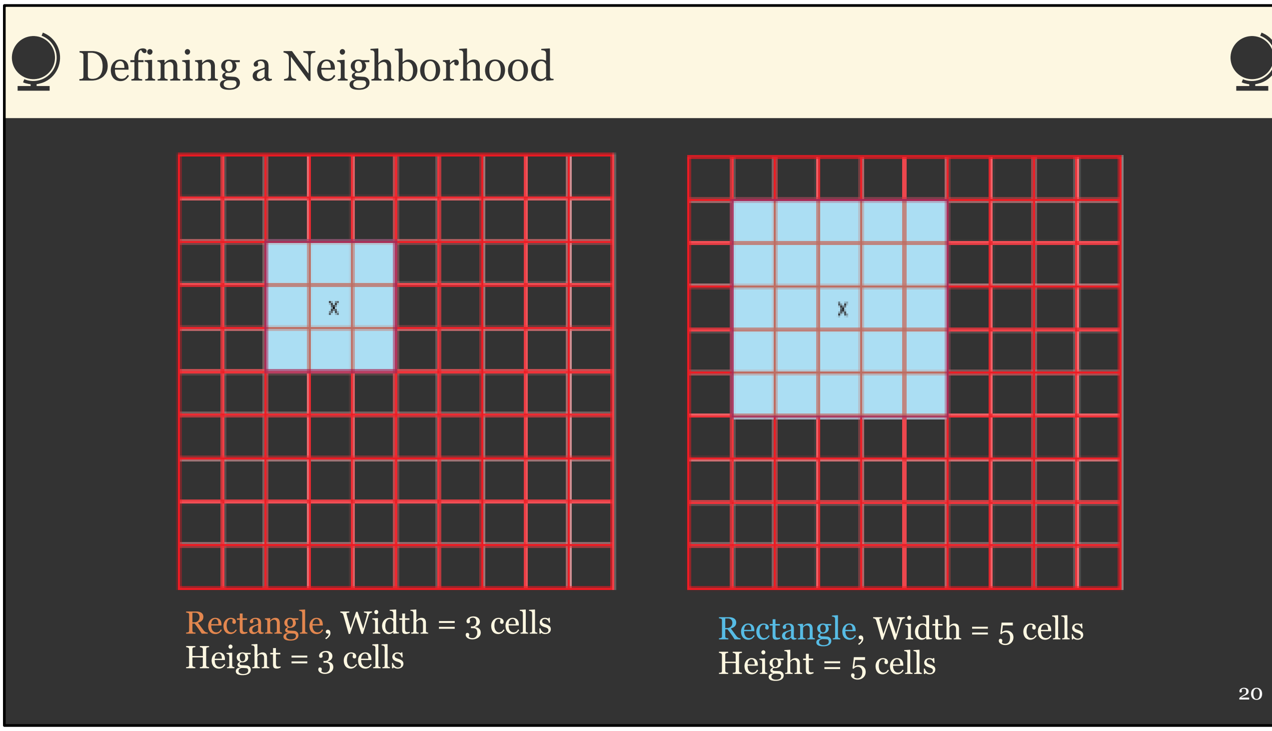


19

Many raster analysis tasks rely on comparing a raster grid cell value to the values in neighboring cells. These types of processes go by many names including neighborhood analysis, moving window analysis, focal operations, and kernel analysis. Further, these types of processes are not limited to geospatial analysis. Similar techniques are used to blur or sharpen photographs in digital photo editing software packages.

These methods function by moving a window over the raster surface. A calculation is performed based on the cell values within the moving window. The result is then written to a new raster cell at the location associated with the cell at the center of the window. The window then moves on to the next cell, using a different center cell and window. The calculation is performed again, and the process continues until all cells have been used as the center cell.

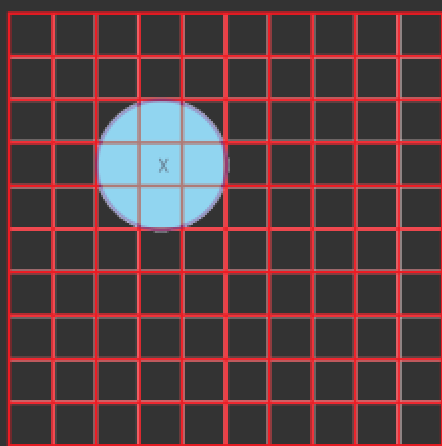
This slide conceptualizes a 3-by-3 cell window moving over a raster grid. For example, the mean of the cells within the window could be calculated then returned to a new cell at the location of the center cell in the grid. The window will then move on to the next cell and perform the calculation again.



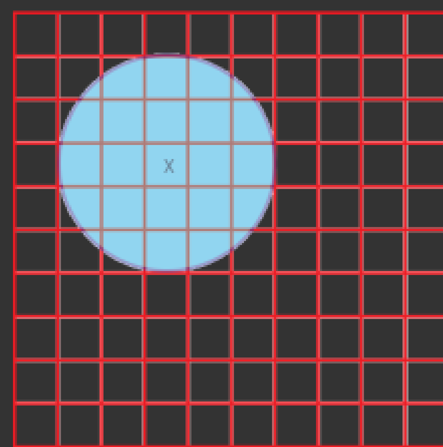
Moving windows can have different sizes and shapes. The examples here represent rectangular or square windows. The first example measures 3-by-3 cells and the second measures 5-by-5 cells.

Generally, an odd number is used so that there is a center cell.

Defining a Neighborhood



Circle, Radius = 1 cell

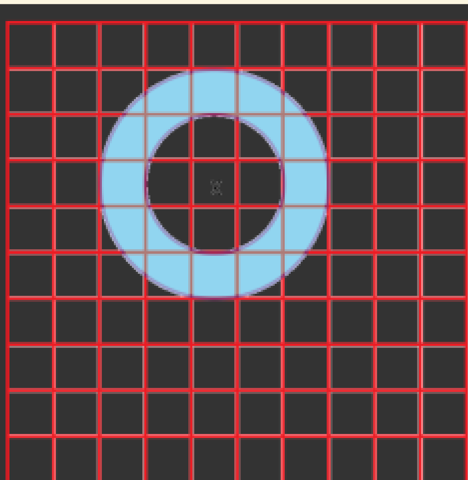


Circle, Radius = 2 cells

21

Circular windows can also be used. When using a circular window, a radius is defined relative to the map units or a number of cells. The examples here represent circular windows with radii of 1 and 2 cells.

Defining a Neighborhood



- Wedge = Wedge shape radiating from center cell that is specified using a start angle, end angle, and radius
- Irregular = Irregular shape specified by a window or kernel

Annulus, Inner Radius = 1 cell, Out Radius = 1 cell

22

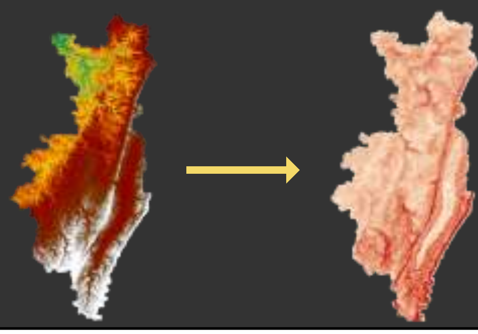
More complex moving window shapes can be used including an annulus, wedge, or irregular shape.

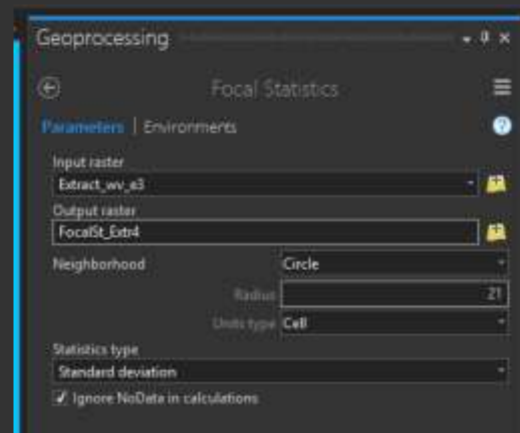
As an example, an annulus is defined by an inner ring that is not included in the window and an outer ring that is included. Such a shape allows you to make comparisons between a cell and cells that are not in direct contact with it.

Please read the description of wedge and irregular window shapes.

Focal Statistics

- Calculates a statistic within the window
- Stores result on center cell in new output
- Completes operations for all cells





23

Focal statistics is a type of moving window analysis. As the name implies, this method is used to calculate a statistic within a moving window.

As the window progresses over the raster grid, the resulting statistic is written to a new cell at the center cell location. The process continues to perform the calculation using all the cells in the original grid as the center cell.

Focal Statistics

- MEAN = Calculates the mean (average value) of the cells in the neighborhood.
- MAJORITY = Calculates the majority (value that occurs most often) of the cells in the neighborhood.
- MAXIMUM = Calculates the maximum (largest value) of the cells in the neighborhood.
- MEDIAN = Calculates the median of the cells in the neighborhood.
- MINIMUM = Calculates the minimum (smallest value) of the cells in the neighborhood.
- MINORITY = Calculates the minority (value that occurs least often) of the cells in the neighborhood.
- RANGE = Calculates the range (difference between largest and smallest value) of the cells in the neighborhood.
- STD = Calculates the standard deviation of the cells in the neighborhood.
- SUM = Calculates the sum (total of all values) of the cells in the neighborhood.
- VARIETY = Calculates the variety (the number of unique values) of the cells in the neighborhood.

24

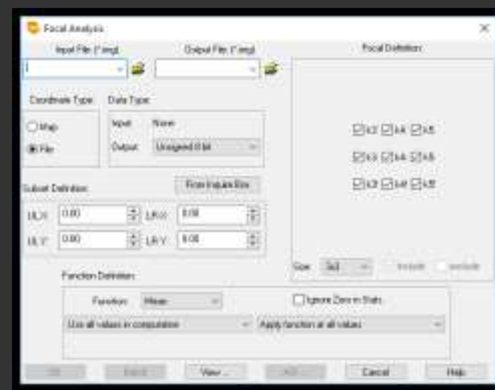
A variety of statistics can be calculated using focal statistics.

For a continuous surface, such as a digital elevation model or a topographic slope grid, the mean, maximum, median, minimum, range, standard deviation, or sum may be calculated. Many statistical values don't make sense for categorical raster grids, such as land cover. However, the majority, minority, or variety could be calculated. As an example, the variety measure would calculate the number of unique categories within the moving window.

Take some time to review the available statistics presented on this slide.

As a final note, larger windows sizes are generally used to represent more coarse or generalized patterns while smaller windows are used to represent more local patterns. There is not necessarily a correct window size. Instead, this depends on the use case. If you are not sure what window size to use, I would recommend experimenting with multiple window sizes. It is also possible to make a calculation using different window sizes then take an average.

Focal Statistics in Remote Sensing

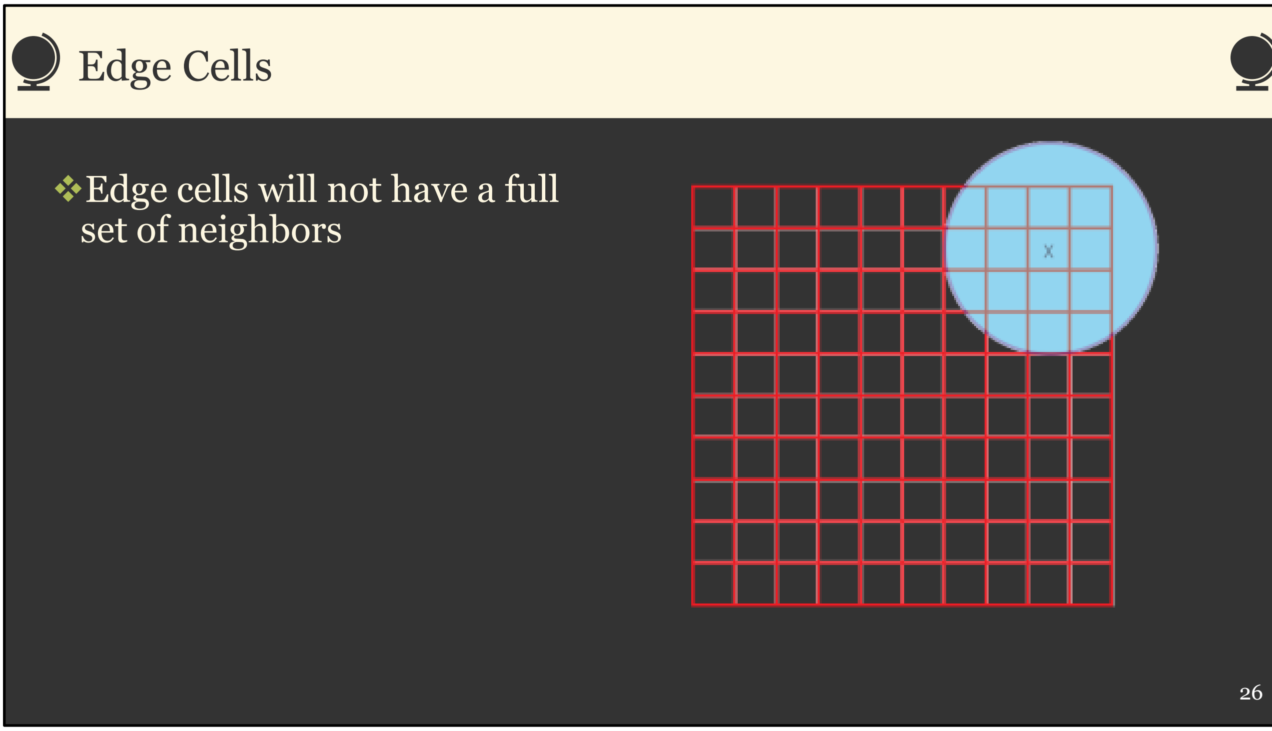


25

Moving window analyses also have applications in remote sensing. This slide shows some tools that are available in the Erdas Imagine software.

Example applications in remote sensing include high pass filters, low pass filters, and edge detection.

If you are interested in these techniques, I would recommend taking a remote sensing class.



It may have occurred to you that cells near the edge of the raster may not have a full set of neighbors as demonstrated in this slide.

In these situations, two options are available. First, the result could be calculated using only the available cells. Second, results could only be calculated for cells that have a full set of neighbors.

If the second option is used, the size of the raster grid, or the number of rows and columns, will be reduced.

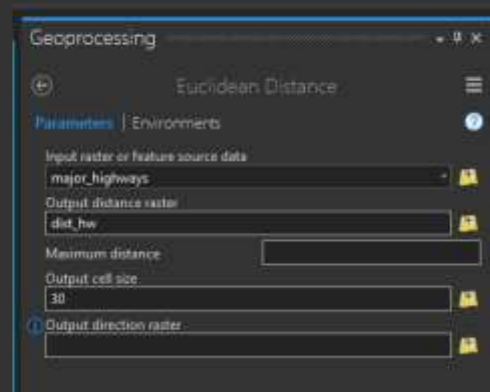
If you are concerned with this issue, I would recommend running an analysis using a larger extent than is required. That way, you will always have a full set of cells to cover the extent of interest.

Distance

27

Euclidean Distance

❖Straight line or Euclidean distance from a source to the cell



28

Euclidean distance simply represents the straight-line distance from a cell to the nearest input feature. As an example, the surface shown here represents the straight-line distance from the cell to the nearest major road.

Each cell is coded with the distance to the nearest input feature.

It is also possible to define a maximum distance. If a cell is further than the maximum distance from the nearest input feature, then it will not be calculated.

Is Euclidean distance meaningful?

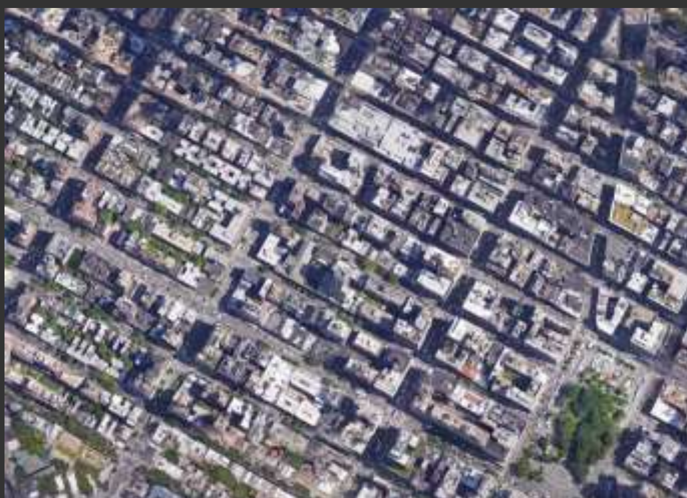


Image from Google Earth

29

Euclidean distance is generally easy to calculate. As a result, it might be overused. Before using this measure, you should ask yourself whether the Euclidean or straight-line distance has real meaning.

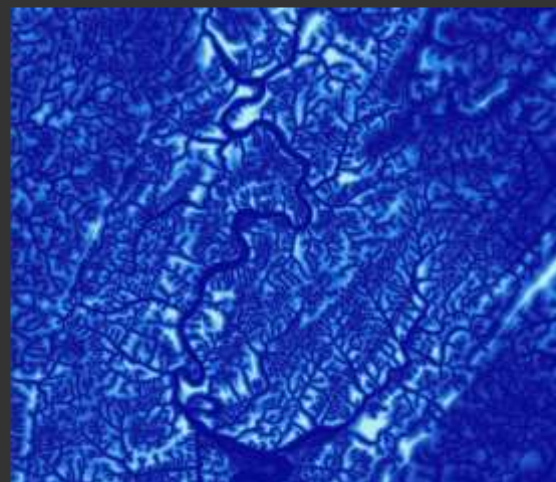
For example, in the image above the straight-line distance may have little meaning for a pedestrian using the sidewalks. In this case, calculating distance along the sidewalks might be of greater value. Such a task could be performed using vector-based network analysis.

So, make sure that Euclidean distance is of value before using it in an analysis.

Cost Distance

❖ Euclidean distance but with a cost or impedance applied

Example: Distance from streams weighted by slope



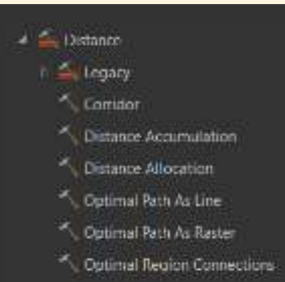
30

Cost distance is similar to Euclidean distance. However, it allows for a measure of cost or impedance to be applied.

The example in this slide represents distance from streams weighted by slope. So, topographic slope is being used as a measure of cost or impedance to travel away from streams.

This specific result could be useful for delineating floodplains or riparian areas.

Distance Tools in ArcGIS Pro



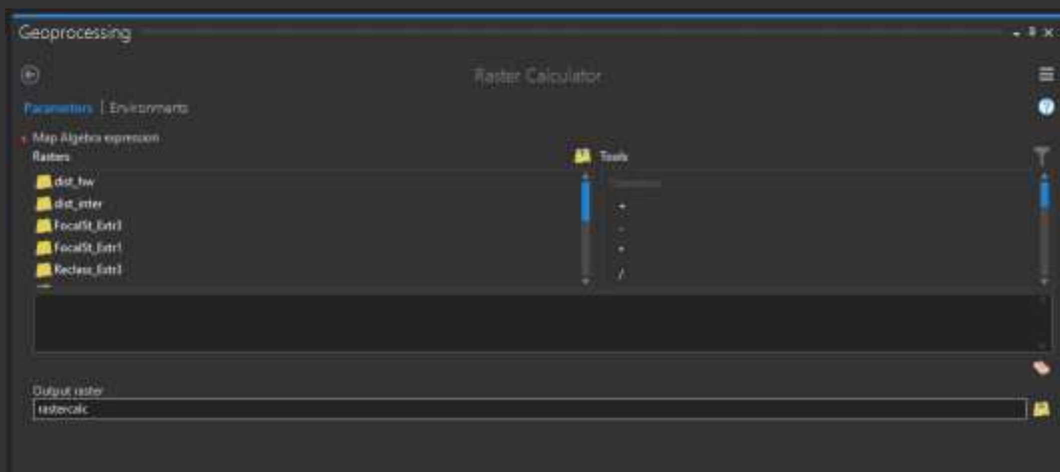
31

ESRI has been introducing some new tools into ArcGIS Pro that relate to distance calculations and analysis. The older tools have been maintained as legacy tools.

Raster Math

32

Raster Calculator



33

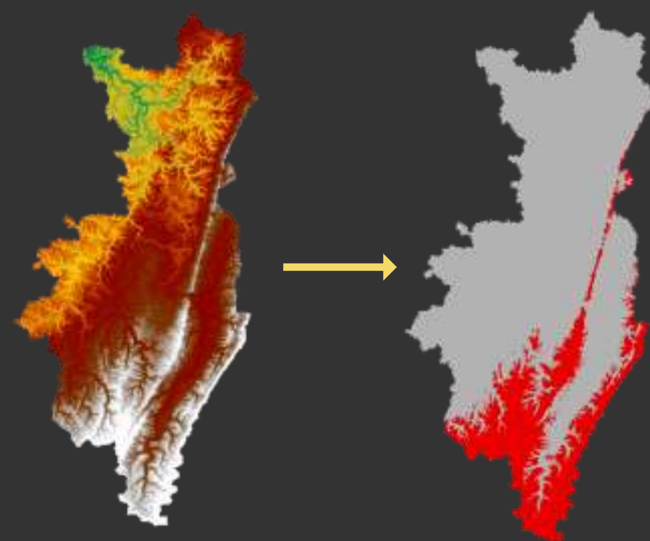
The Raster Calculator Tool allows you to perform mathematical calculations on a cell-by-cell basis. This tool is very flexible and can be used to perform a wide variety of raster-based calculations. Examples are provided on the following slides.

Produce Binary Raster Grid

- "Elevation" > 800
- 1 = Yes = True
- 0 = No = False

850	700	820
600	820	900
NoData	650	675

1	0	1
0	1	1
NoData	0	0



34

Raster Calculator can be used to produce binary, 0-1, or yes-no surfaces. In these outputs, cells that met the query criteria are coded as 1 and cells that do not meet the criteria are coded as zero. 1 always indicates true while 0 will indicates false.

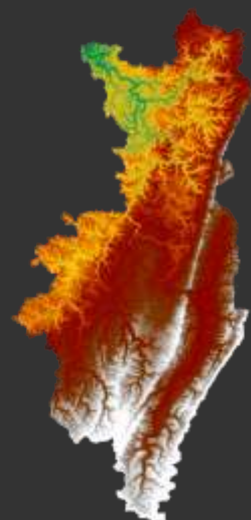
In this example, I am querying the raster to find all cells in a digital elevation model that have an elevation greater than 800. Cells that hold an elevation value greater than 800 will be coded to 1 in the output. All other cells will be coded to 0. The result is a binary output.

Operations Involving A Single Raster

- SquareRoot("Elevation")

850	700	820
600	820	900
NoData	650	675

29.2	26.5	28.6
24.5	28.6	30
NoData	25.5	26.0



35

Operations can be performed that only require one raster input.

In this example, the square root of the elevation values in the input raster grid will be returned to the output raster.

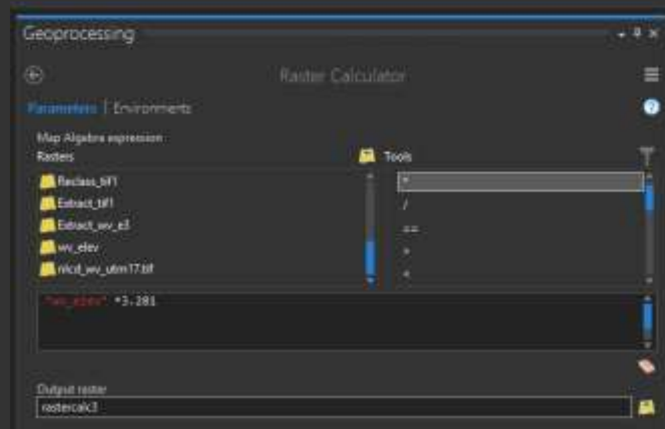
Other examples include converting grid cells from float to integer, taking the tangent of the cell values, or taking the natural log of the cell values.

Operations Involving Raster Grids and Constants

❖"Elevation" * 3.281

850	700	820
600	820	900
NoData	650	675

2789	2297	2690
1969	2690	2953
NoData	2133	2215



36

Operations can also be performed that require raster grids and constants.

In the example, elevation values in meters are being multiplied by 3.281 to output elevation values in feet.

Unit conversion is a common use of such techniques.

Operations Involving Multiple Raster Grids

1	0	1
0	0	1
NoData	1	1

X

1	NoData	1
0	1	0
NoData	1	NoData

=

1	NoData	1
0	0	0
NoData	1	NoData

1	0	1
0	0	1
NoData	1	1

+

1	NoData	1
0	1	0
NoData	1	NoData

=

2	NoData	2
0	1	1
NoData	2	NoData

It is also possible to perform calculations using multiple raster grids.

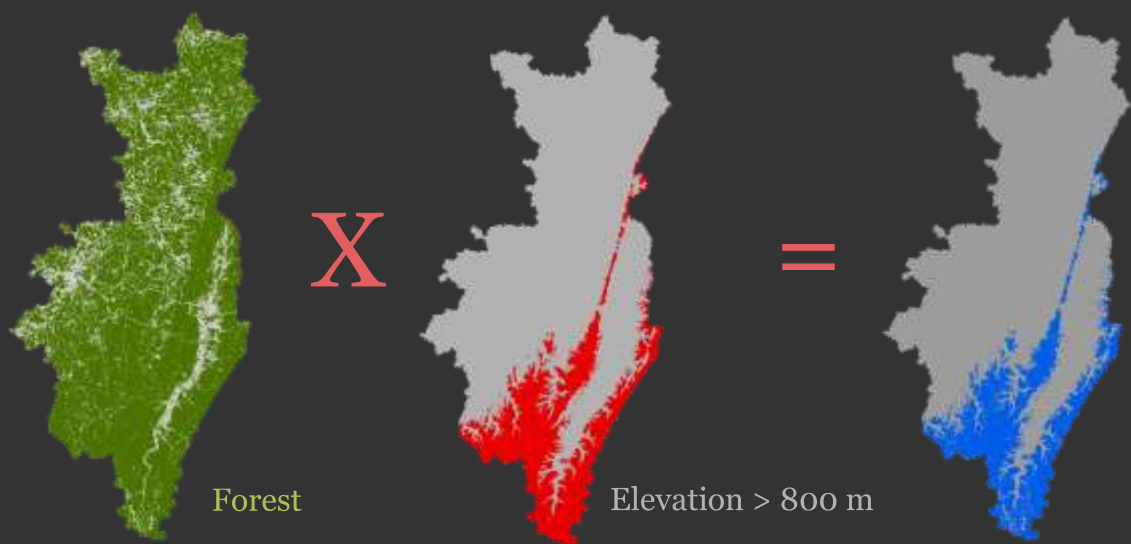
When multiple raster grids are used, cells that overlap will be involved in calculations.

In the example above, cells are being multiplied and added together.

Note that a NoData result will be returned any time at least one of the grids holds a NoData value at that location. Calculations can only be performed when values are available for all grids at that location.

When cells do not align properly, how the grids overlay to perform these calculations can be ambiguous. This is one reason to use a snap raster.

Operations Involving Multiple Raster Grids



This is another example of raster math involving two grids. A binary surface separating forest and not forest is being multiplied by a binary grid representing locations with elevations above 800 meters. The result will return 1 only where the forest and elevation criteria are both true. So, values of 1 in the output represent locations that are forest at an elevation above 800 meters.

Some Useful Functions

+ Add

- Subtract

* Multiply

/ Divide

== Equal to

!= Not equal to

> Greater than

>= Greater than or equal to

< Less than

<= Less than or equal to

& Boolean AND

| Boolean OR

^ Boolean XOR

~ Boolean NOT

39

Many functions are available for use within Raster Calculator. Take some time to review the available functions presented here. Note that these are not the only function available. These are just examples.

Some Useful Functions

Abs Absolute value

Exp Base e exponential

Exp10 Base 10 exponential

Exp2 Base 2 exponential

Float Convert from integer to float

Int Convert from float to integer

Ln Natural log

Log10 Base 10 log

Log Base 2 log

Square Calculates square

SquareRoot Calculates square root

IsNull Identifies NoData cells

40

Here are some additional functions available in Raster Calculator. Take some time to look over these functions.

Conditional Statements

- Con(set condition, what to return if true, what to return if false)

Example: Con(“Land_Cover" = 41, 1, 0)

- If the land cover raster grid is assigned a value of 41 (deciduous forest), return 1. If the land cover grid is assigned any other value, return 0.

41

Conditional statements can also be integrated into Raster Calculator syntax. In a conditional statement, the first argument is the condition, the second argument defines what to return if the condition is true, and the last argument defines what to return in the condition is false.

In this example, the condition is that the land cover code must be equal to 41. This code represents deciduous forest. If the condition is met, or the location is coded as deciduous forest, the result will be 1. If the condition is not met, or the result is not deciduous forest, the result will be 0. So, the output here will be a binary output in which 1 indicates deciduous forest and 0 indicated not deciduous forest.

Impact of NoData on Raster Math

- NoData cells return NoData!!!!!!
- Operations are only completed where all cell values are available

1	0	1
0	0	1
NoData	1	1

X

1	NoData	1
0	1	0
NoData	1	NoData

=

1	NoData	1
0	0	0
NoData	1	NoData

42

As a final note on raster calculations, a result will only be returned when values are available in all grids.

As an example, if you multiply 100 grids together, and they all have cell values at a location except one grid, which has a NoData assignment at that location, the result will be NoData.

Values must be available in all grids at a cell location in order for a calculation to be performed at that cell location.

Raster Generalization

43

Majority Filter

- Replace center cell with majority value in 3 x 3 cell window
- FOUR vs. EIGHT = use just orthogonal cells or orthogonal and diagonal cells
- MAJORITY vs. HALF = Must have a majority or only half to change value
- This is used to generalize categorical raster data

30	20	30
30	20	20
40	30	30

44

The raster generalization methods that we will discuss here are only applicable to categorical raster grids, such as land cover.

First, we will discuss majority filtering. Majority filtering can be used to replace a cell value with a different value based on neighboring cells. Generally, this is performed using a 3-by-3 moving window, so this is another example of a method that makes use of neighborhood analysis. Specifically, this method is used to replace a cell value with the majority cell value of the neighboring cells.

If the FOUR method is used, then only orthogonal cells will be considered. These are the cells that share a side with the center cell. If the EIGHT method is used, then all eight neighboring cells will be considered, or the four orthogonal and four diagonal cells.

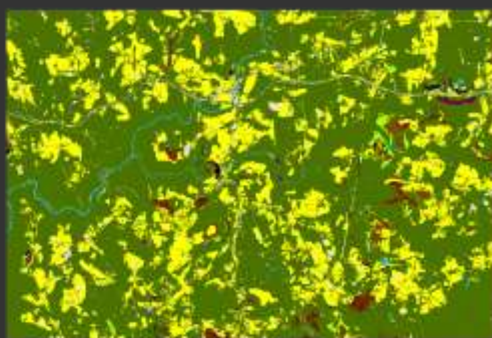
If the Majority setting is applied, then the cell value will only be changed if a majority or more than half of the neighbors contain the new cell value. If the Half setting is used, then at least half of the neighboring cells must contain the cell value.

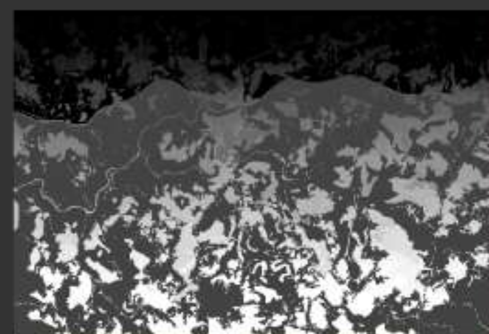
In the example provided here, if EIGHT and Majority were used then the cell value would be changed to 30. This is because 30 is the majority value in the eight neighboring cells, as five of the eight cells contain this value. If the FOUR and majority method were used, then the center cell value would be maintained as 20, since there are two 30 and two 20 cells in regards to the orthogonal neighbors, so there is no majority. Regardless of the settings used, the-3 by-3 moving window will move through the grid to perform this calculation using each cell as the center cell.

A common application of majority filtering is to reduce the salt-and-pepper effect that commonly results from image classification. Pixels that are different from their neighbors can be recoded to the majority value of their neighbors, thus reducing the salt-and-pepper issue.

Region Group

❖ Group cells with the same code in a categorical raster and code them with a unique identifier

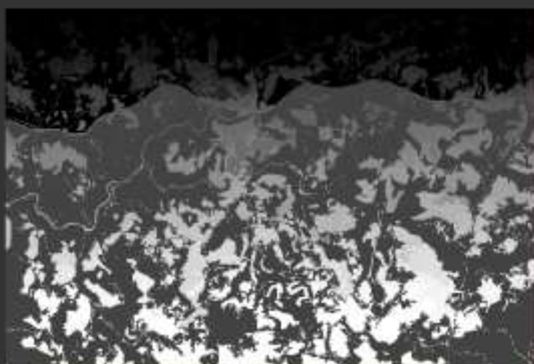




45

Region group is used to code all regions in a categorical raster to a unique code. So, any contiguous set of pixels that have the same code will be grouped, and each of these unique regions will be assigned a unique code. The goal here is to identify contiguous regions of the same category.

Region Group



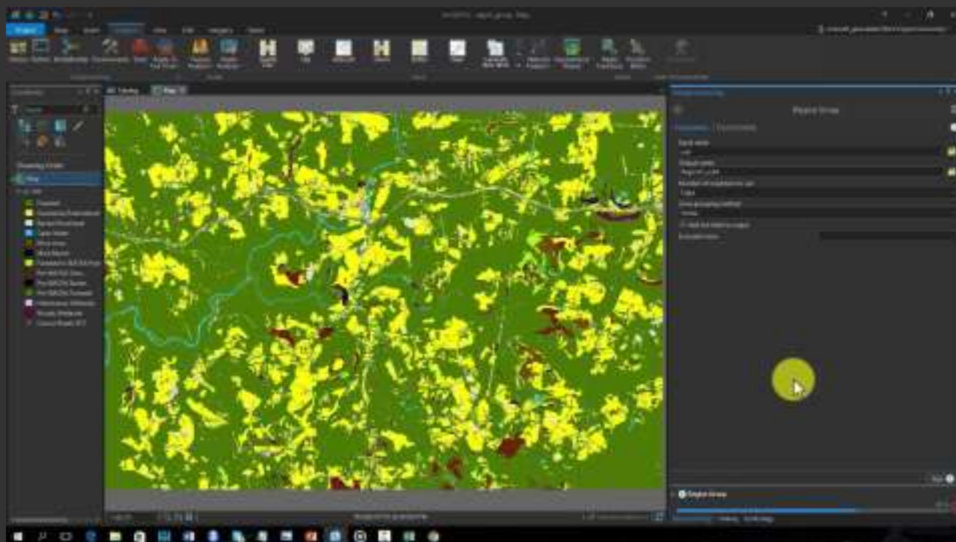
- Must decide between the FOUR or EIGHT method
- Can convert to polygons after regions are grouped
- Will needed to decide if you want to simply the polygons

46

When performing region grouping, similar to majority filtering, the FOUR or EIGHT methods can be applied. If the EIGHT method is applied, an area of the same code will be considered contiguous even if the pixels only touch diagonally. In contrast, when using the FOUR method an area will only be considered contiguous if the cells share a side, or are orthogonal. Generally, the FOUR method will yield more unique regions than the EIGHT method.

Once regions have been assigned unique codes, they can be converted to polygons with or without simplification. If no simplification is applied, then the polygons will follow the boundaries of the cells.

Video: Majority Filter



47

Expand

- Expand out a specific category in a categorical raster by a specified number of cells

Green expanded by 5 30 m pixels

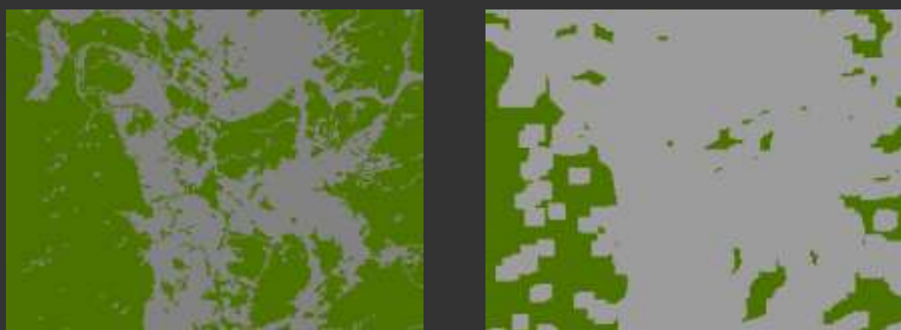
48

Expand is used to extend a certain code or category out by a defined number of cells.

In the example provided, the green cells represent forest, and the gray cells represent all other categories. Using Expand, the forest areas have been expanded by 5 cells or 150 meters, reducing the area of not forest.

Shrink

- Shrink a specific category in a categorical raster by a specified number of cells by replacing them with the values of the cell that is most frequent in its neighborhood



Green shrank by 5 30 m pixels and replaced with gray

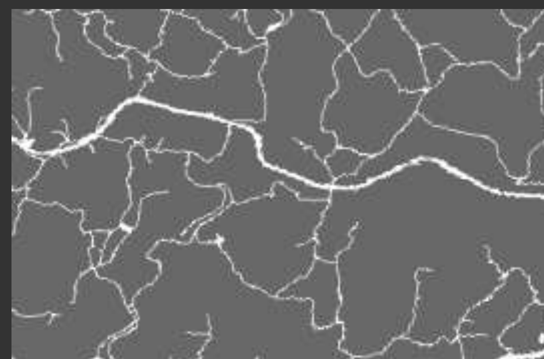
49

Shrink is the opposite of expand. Instead of expanding areas with a certain code or value, that code or value is reduced. The areas that are reduced for this class are replaced with the most frequent value or code in its neighborhood.

In the example above, the forest class has been reduced by 5 cells or 150 meters. Cells in these areas have been recoded as not forest.

Nibble

❖ Replace cells of a raster corresponding to a mask with the values of the nearest neighbor



50

Nibble is used to replace cells values with the nearest cell value that is not masked out.

In the example above, the red areas indicate some level of development. We would like to remove these areas and replace them with the nearest non-developed category. To accomplish this task, a mask is produced. This mask is used to mask out all developed areas. Using this mask, areas that are developed will be recoded as the nearest code that is not developed, since all developed areas were masked out.

Progress to the next slide to see the result.

Nibble

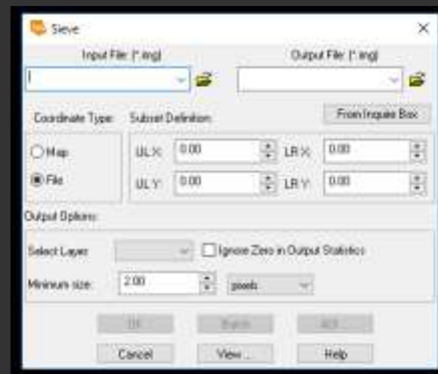
❖Replace cells of a raster corresponding to a mask with the values of the nearest neighbor

51

As you can see, the developed cells have now been replaced by the nearest non-developed category, effectively removing this class from the dataset.

Sieving

- Replace regions smaller than a certain size in a categorical raster with the surrounding class
- Can be used to approximate a minimal mapping unit (MMU)



52

As a last example of raster generalization methods, sieving can be used to remove areas of the same category that are smaller than a certain size. The user must define a minimal size area to maintain as a land area or number of cells. The tool will then find all contiguous regions that are smaller than this defined size. Once these areas are found they will be removed and assigned to the surrounding class.

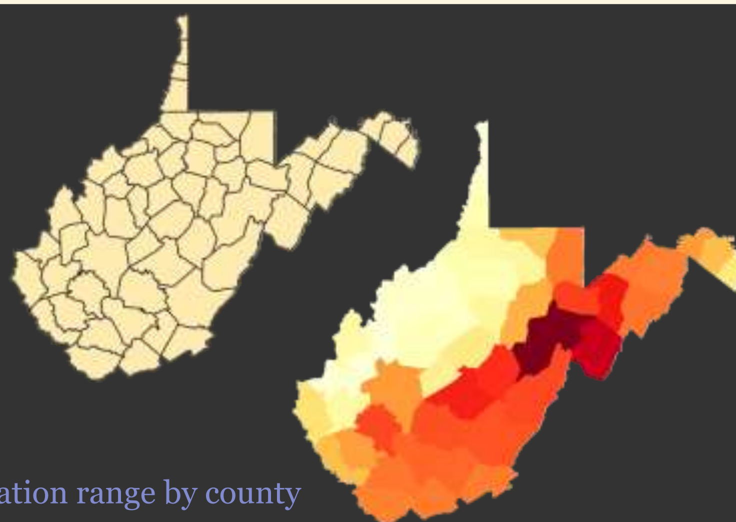
As an example, this tool could be used to remove small areas in a land cover classification that are smaller than a defined size. For example, small barren areas within a forested extent could be removed and coded as forest. This method can be used to remove the salt-and-pepper effect common in land cover classifications.

Additionally, this method can be used to approximate a minimal mapping unit (MMU). Areas smaller than the minimal mapping unit can be recoded and assigned to the surrounding class.

Raster Summarization

Zonal Statistics

- Used to obtain statistics from a continuous raster within an area or zone as defined by another raster or a polygon vector layer
- Can obtain results as a raster surface for a single statistic



Elevation range by county

54

A common spatial analysis task is to summarize raster data relative to vector data or categorical raster zones. For example, you may want to summarize continuous raster data relative to polygons. This can be accomplished using the Zonal Statistics Tool. This method will calculate a statistical measure, such as the mean, median, standard deviation, minimum, or maximum, from the cell values within a zone.

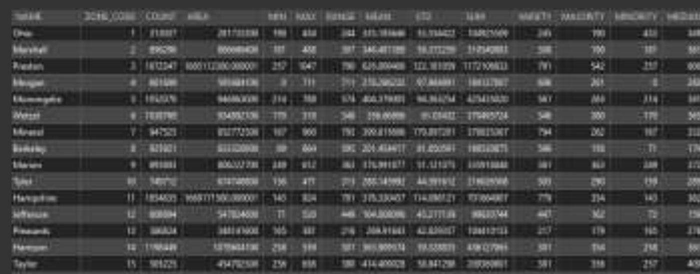
The zones can be defined using a categorical raster or vector polygons. If vector polygons are used, then they will be converted to raster zones prior to the summarization. As a result, this method generally will not work correctly if overlapping polygons are used, since only one code can be assigned to each cell. The result will be a statistical measure that will be the same for each cell within each zone.

The example above shows the range of elevation for each county in West Virginia. Since the zones were defined as counties, all cells in the same county were assigned the same value, in this case the elevation range.

Since a singe band raster is produced, only one statistical measure can be generated at a time.

Zonal Statistics as Table

- Used to obtain statistics from a continuous raster within an area or zone as defined by another raster or a polygon vector layer
- Can obtain multiple statistics as a table: mean, majority, maximum, median, minimum, minority, range, standard deviation, sum, and variety
- Can use table join to join the results back to the zone data



Elevation statistics by county

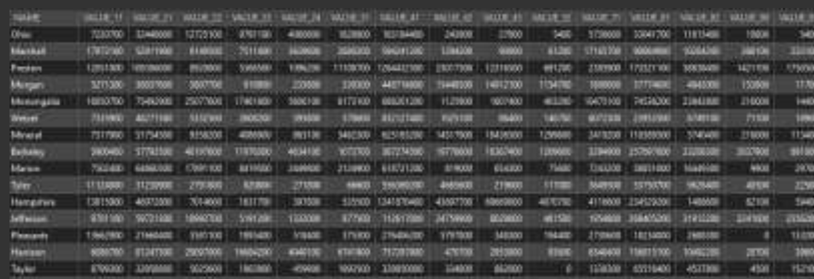
55

If you would like to calculate multiple statistics in a table format, this can be accomplished using Zonal Statistics as Table. Instead of producing a raster output for a single statistical measure, a table output is generated that can provide multiple statistical measures. Each row in the table will correspond to a zone, in this example a specific county.

If you would like to associate these results with the original zones, this can be accomplished using a table join. Here, this table could be joined back to the county vector file using the shared NAME field.

Tabulate Area

- Used to obtain areas for categories in a categorical raster within an area or zone as defined by another raster or polygon vector layer
- Areas will be in square map units
- Can use table join to join the results back to the zone data



Area of land cover by county

56

The Zonal Statistics tools are designed to provide summary statistics relative to zones and are generally more applicable for summarizing continuous raster data, such as a digital elevation model. If you would like to summarize categorical raster data, such as land cover, this can be accomplished using the Tabulate Area Tool. The output will be a table that provides the land area of each category within each zone.

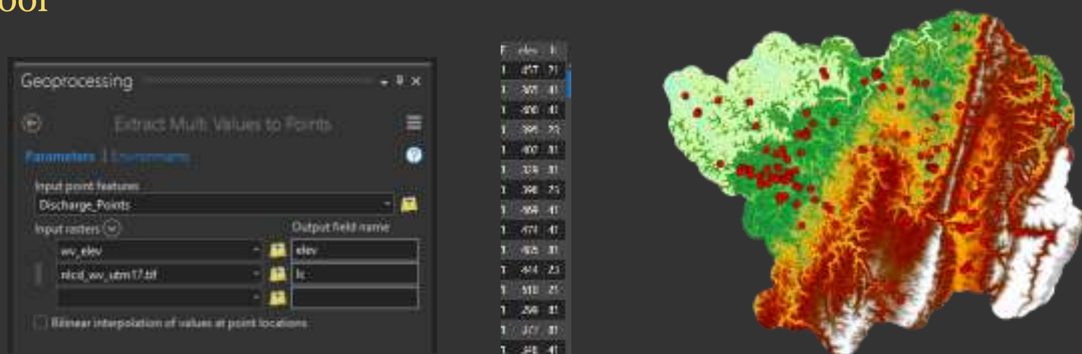
If you would like to obtain the percentage of the land area as opposed to the area, then you could perform table calculations in new fields.

This table could also be joined to the original zone using a table join for further analysis.

Similar to the Zonal Statistics tools, vector polygon zones will be converted to raster zones to perform the summarization. So, incorrect results will be produced for overlapping polygons.

Extract Values to Points

- Extract raster values at point locations and add the data to the attribute table
- Can use Extract Multi Values to Points Tool or Extract Values to Points Tool



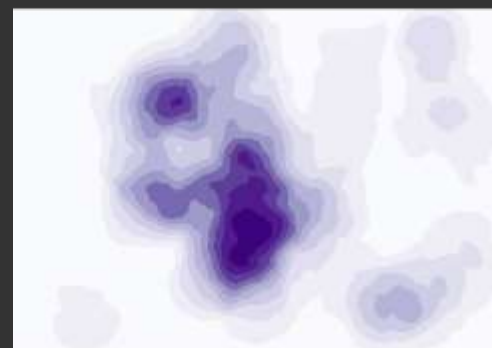
It is also possible to extract values at points. The result will be a new field in the attribute table for the point layer that holds the cell value from a raster grid occurring at that location. This is an easy way to extract cell values at point locations in preparation for later analysis.

Estimating Densities

58

Point Density/Line Density

- Calculates the magnitude per unit area for point or line features that fall within a neighborhood around each cell
- Line length/area
- Number of points/area
- Can incorporate weights
- Must specify neighborhood shape and size



59

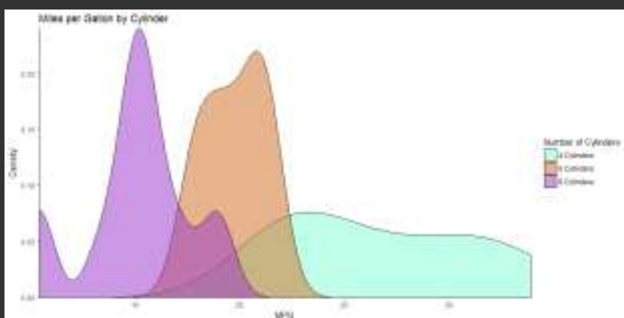
You may be interested in obtaining a raster-based measure of density for point or line features. This can be accomplished using the Point Density or Line Density tools. These tools are yet another example of methods that make use of neighborhood analysis or moving windows.

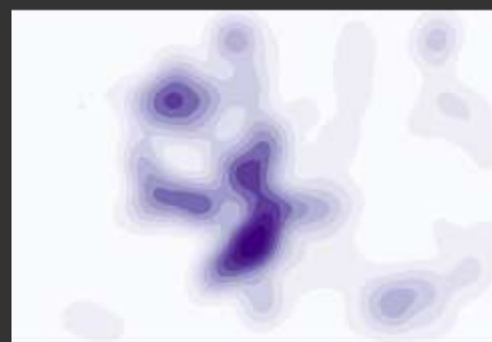
The user must define the size and shape of the moving window. The tool then will calculate the number of points or length of lines within the window. On the center cell for each window, the tool will return a measure of density. For points this will be the number of points per unit area. For lines, this will be the length of lines per unit area.

The size and shape of the window can have a large impact on the output. More global or generalized patterns will be produced with larger windows whereas more local patterns will be observed if a smaller window is used. There is not necessarily a correct window size, as this depends on the goals of the analysis.

Kernel Density

- Fits a kernel density function to estimate density





60

Kernel density is similar to point or line density, except that a different method is used to calculate density. Specifically, this method makes use of a kernel function. The next slide provides an explanation of this method. Please take the time to read through this explanation.

Point Density vs. Kernel Density

- In point and line density, a neighborhood is specified that calculates the density of the population around each output cell.
- Kernel density spreads the known quantity of the population for each point out from the point location.
- The resulting surfaces surrounding each point in kernel density are based on a quadratic formula with the highest value at the center of the surface (the point location) and tapering to zero at the search radius distance. For each output cell, the total number of the accumulated intersections of the individual spread surfaces is calculated.

61

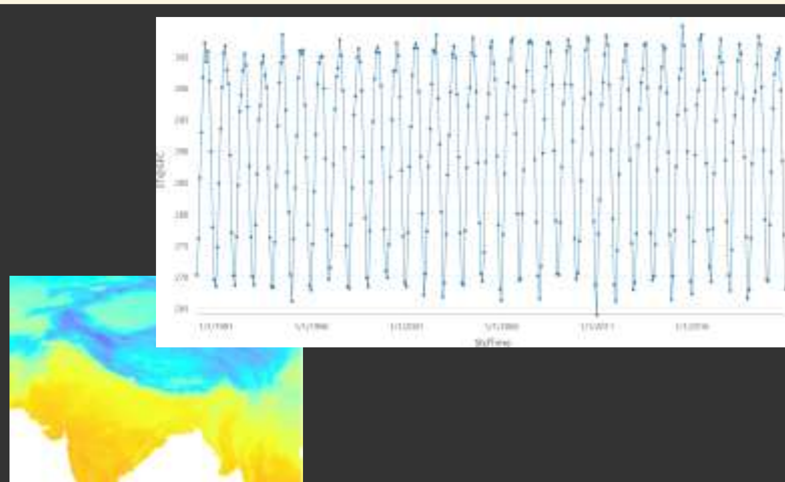
This slide provides a comparison of the point density and kernel density methods.

Multidimensional Raster

62

Multidimensional Raster File Formats

- Store more than three dimensions
- Commonly used in climate science
- Time
- Depth
- Multiple Variables
- Common Formats
  - GRIP, HDF, NetCDF, CRF



63

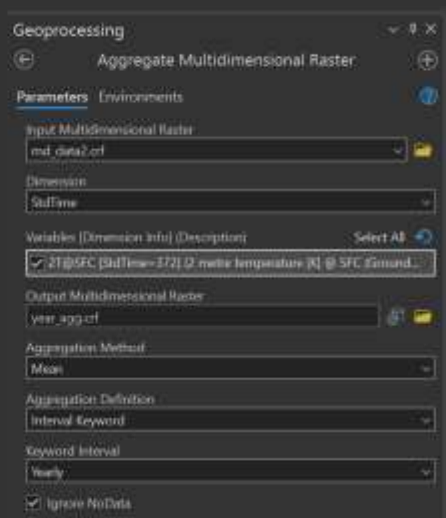
Multidimensional raster grids further extend the raster data model to incorporate more than three dimensions (i.e., height, width, and channels). For example, a set of climate variables could be stored as a single file with height, width, variable, time, and altitude dimensions. Such models are sometimes referred to as data cubes. A series of multispectral images representing different dates, such as a series of Landsat images, could be stored as a single file with height, width, channel, and time dimensions. Some common file formats for multidimensional raster grids include:

GRIB = General Regular-Distributed Information in Binary
HDF = Hierarchical Data Format
NetCDF = Network Common Data Form.

ESRI has developed a raster file type for multidimensional raster data called Cloud Raster Format (CRF).

Multidimensional raster grids expand the types of data that can be represented and also the types of analyses that can be performed. Specifically, multidimensional raster data support time series analysis, anomaly detection, aggregation over time or other domains, and trend analysis. Climate scientists and modelers have been using these types of file structures for some time because they undertake the types of analyses just listed.

Subset or Aggregate



- Relative to different dimensions
  - Space
  - Time
  - Altitude/depth
  - Bands

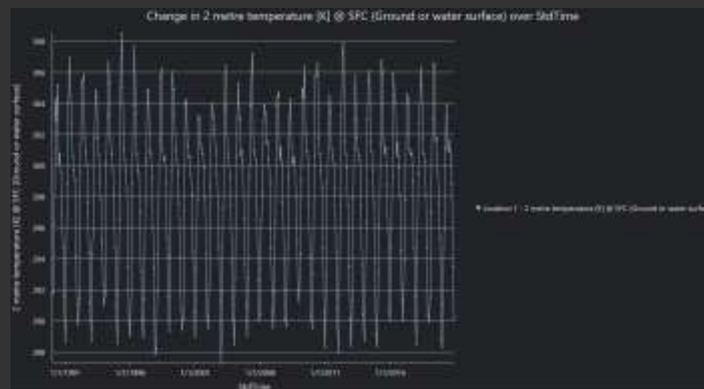
64

It is possible to extract spatial subsets of data based on spatial extents, time periods, or ranges of altitude or depth. It is also possible to extract out specific variables or bands.

Data can also be aggregated using statistical measures. For example, daily temperature measurements could be aggregated to monthly means or medians.

Time Series Analysis

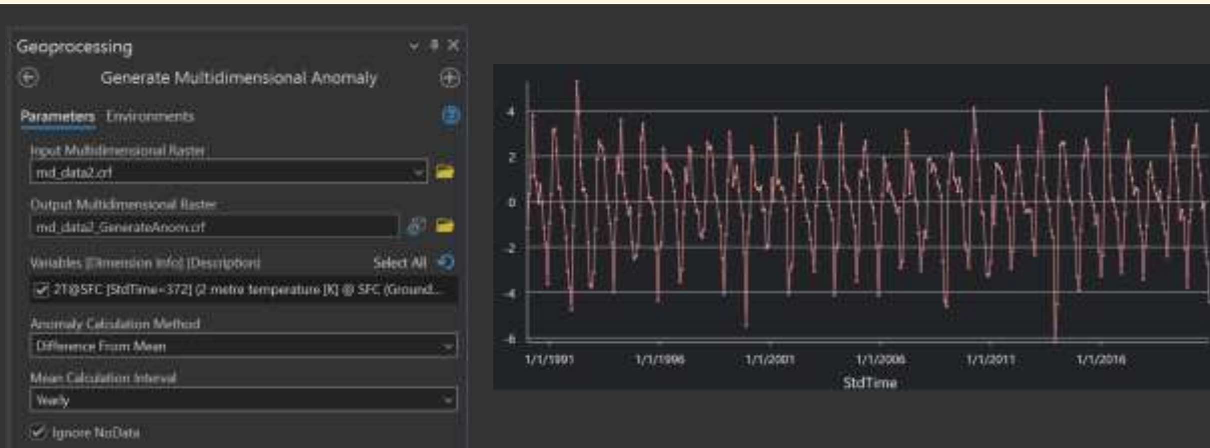
- Seasonal patterns
- Oscillating patterns
- Longer-term trends



65

It is common for multidimensional raster grids to be used to store time series data, such as climatological data. This supports the analysis of patterns over time, such as seasonality, oscillating patterns, and long-term trends, such as changes in climate. Different statistical methods have been developed to support these types of analyses. A discussion of time series analysis is outside the scope of this course.

Anomaly Detection



66

Multidimensional raster data also support anomaly detection. For example, very large storm events or periods of drought may be detectable in a time series. Different algorithms have been developed to detected anomalies. A discussion of these methods is outside of the scope of this course.

Spatial Analysis Tools

67

In this last section, I will briefly introduce some other tools/environments in which to process and analyze raster data.

Software Options

- ArcGIS Pro
- ArcMap
- Erdas Imagine
- ITT ENVI
- QGIS/GRASS/SAGA
- R

Logos were obtained from organization websites.

68

There are many other environments that are available in which to perform spatial analysis and work with raster data.

Although we mainly work with ArcGIS Pro and QGIS in this course, I wanted to take the time to note some other options.

There are several open-source and free software packages that can be used to perform spatial analysis, including R.

Additionally, spatial analysis methods focused on imagery are provided by remote sensing software packages, such as Erdas Imagine and ENVI.

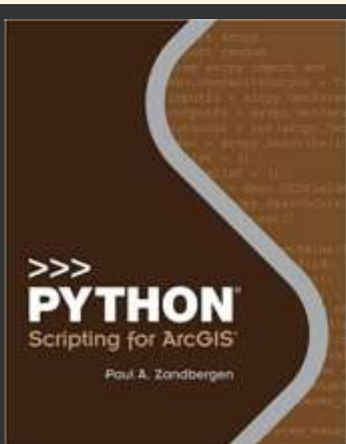
Raster Functions

- Different way to analyze raster data
- Results not saved to disk
- Virtual raster data in memory
- Recalculate with changing scales and/or extents

69

A recent addition to ArcGIS Pro are Raster Functions. Raster Functions offer a different means to process raster data. Instead of writing results to disk as a file, results are dynamically produced in memory or RAM. As the user pans or zooms around the map space, the output will update to the appropriate extent and scale. The key benefit of this method is that it allows for faster processing/calculation. I anticipate that Raster Functions will grow in popularity with time.

ArcPy Python



https://esripress.esri.com/display/index.cfm?fuseaction=display&websiteID=276&moduleID=0

```
class LicenseError(Exception):
    pass

class SpatialRefProjError (Exception):
    pass

import arcpy
from arcpy import env
from arcpy.sa import *
import os
import arcpy.mapping as mapping

try:
        #Check for spatial analyst license
    if arcpy.CheckExtension("Spatial") == "Available":
        arcpy.CheckOutExtension("Spatial")
    else:
        raise LicenseError
    demRaster = arcpy.GetParameterAsText(0)
    dscRaster = arcpy.Describe(demRaster)

    if dscRaster.spatialReference.type != "Projected":
        raise SpatialRefProjError

    # Set overwrite option
    env.overwriteOutput = True
    #Set extent based on users input
    ext = arcpy.GetParameterAsText(2)
    if ext == "DEFAULT" or ext == "MAXOF" or ext == "MINOF":
        env.extent = dscRaster.extent
    else:
        env.extent = ext

    analysisWindow = arcpy.GetParameterAsText(1)
```

Script by Jeff Evans

70

Spatial analysis can be performed in a code-based environment. Using ESRI products, including ArcGIS Pro and the associated ArcPy library, Python can be applied to undertake a variety of tasks. Python can be especially useful if you need to loop through data layers or automate a sequence of steps.

Open-Source Python

https://numpy.org/

https://pypi.org/project/rasterio/

https://docs.xarray.dev/en/stable/index.html

71

There are also open-source Python libraries for working with raster grids and data arrays more generally. These libraries include NumPy, Rasterio, and Xarray.

R

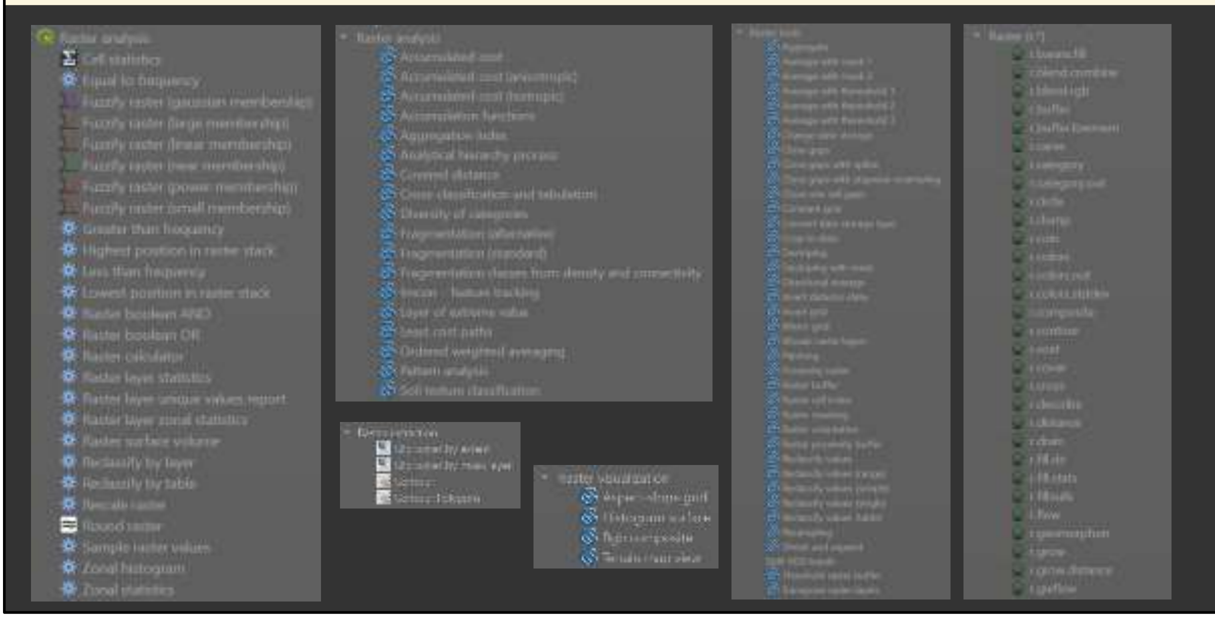
https://github.com/rspatial/raster

https://github.com/rspatial/terra

72

In the R data science environment, geospatial raster data have traditionally been read and analyzed using the raster package. However, this package has more recently been superseded by the terra package.

QGIS

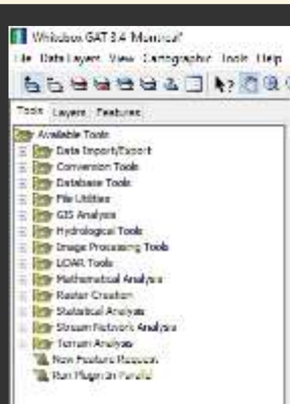


73

Similar to analyzing vector data, QGIS and associated tools (i.e., GRASS, GDAL, and SAGA) offer a variety of tools for working with and analyzing raster data. SAGA specifically focuses on raster-based analysis.

WhiteboxTools and Whitebox GAT

- Over 440 tools for working with spatial data
- Whitebox GAT has a GUI
- WhiteboxTools accessed via QGIS, R, Python, or a command-line interface
- Created by Dr. John Lindsay (Geomorphometry & Hydrogeomatics Research Group; Geography, Environment & Geomatics; University of Guelph)



Whitebox

Geospatial Analysis Tools

https://jblindsay.github.io/ghrg/Whitebox/

WhiteboxTools

https://jblindsay.github.io/wbt_book/intro.html

https://jblindsay.github.io/ghrg/index.html

74

WhiteboxTools and Whitebox Geospatial Analysis Tools (GAT) provide more than 440 tools for analyzing geospatial data, including tools for data preparation, vector analysis and overlay, raster and image analysis, geomorphometric analysis, hydrological analysis, and LiDAR processing.

WhiteboxTools can be accessed in Python, R, QGIS, and via a command-line interface. They can also be accessed via a graphical user interface (GUI) known as Whitebox GAT. This toolset was developed by Dr. John Lindsay of the Geomorphometry & Hydrogeomatics Research Group at the University of Guelph.

Raster-Based Analysis In ArcGIS Pro

75

The following exercises explore raster-based analysis in ArcGIS Pro.

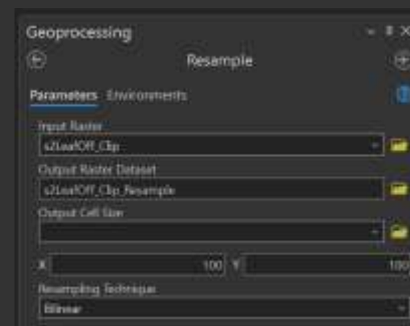
Clip Raster Grid

1. Navigate to the rasterSummarize map.
2. use the Clip Raster Tool to clip the s2LeafOff layer relative to the ws polygon layer using the following settings:
   - **Input Raster** = s3LeafOff
   - **Output Extent** = ws
   - **Use Input Feature for Clipping Geometry** = *selected*
   - **Maintain clipping Extent** = *not selected*

Data Management Tools → Raster → Raster Processing → Clip Raster

Resample Raster Grid

1. Navigate to the rasterSummarize map.
2. use the Resample Tool to coarsen the clipped version of the s2LeafOff layer to a 100 m spatial resolution using the following settings:
   - **Input Raster** = s3LeafOff (clipped version)
   - **Ouput Cell Size** = *100 m*
   - **Resampling Technique** = *Bilinear*

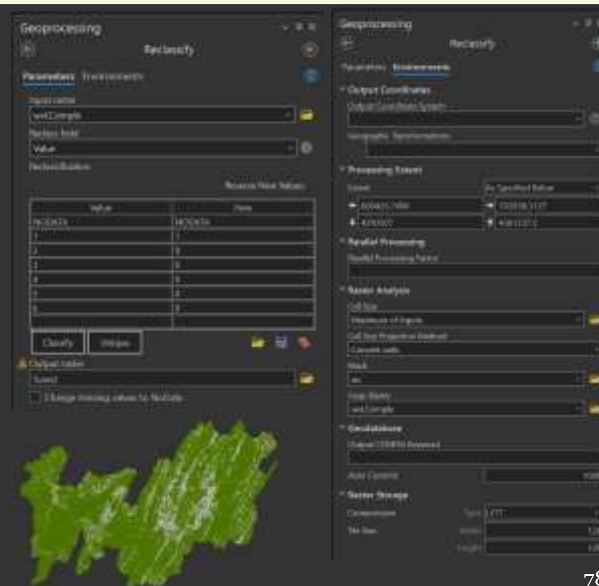


Reclassify Raster

1. Navigate to the rasterSummarize map.
2. Use the Reclassify Tool to create a binary forest raster using the following settings:

Parameters
- **Input Raster** = wvLCsimple
- **Reclass Field** = *Value*
- **Table** = see screen capture
- *Output Raster* = forest

Environments
- **Extent** = *Same as layer* ws
- **Mask** = ws
- *Snap Raster* = wvLCsimple



Spatial Analyst Tools → Reclass

78

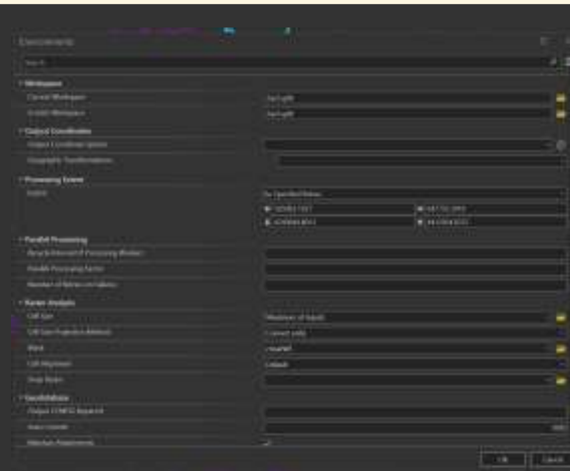
Extract Study Area Extent

1. Navigate to the rasterMathOverlay map.
2. Select the Cheat River Watershed from the huc8WS layer (see image to right).
3. Right-click on the huc8WS layer in the Contents Pane. Select Data followed by Export Features.
4. In the Export Feature Window, save the new layer as cheatWS. Click OK to execute the tool. The result will be added to your map.

79

Set Environments

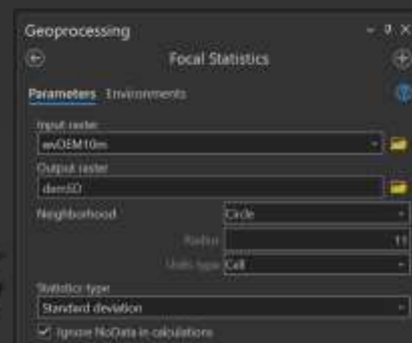
So that you don't have to specify environment settings for each tool execution, you will now set these globally.



1. Navigate to the Analysis Tab.
2. Click the Environments icon. Environments
3. Make the following changes:
   - **Processing Extent** = *Same as layer* cheatWS
   - **Mask** = cheatWS
   - **Snap Raster** = wvDEM10m
4. Make sure to click OK to save the changes.

Focal Statistics

1. Navigate to the rasterMathOverlay map.
2. Use the Focal Statistics Tool to create a raster of the standard deviation of elevation within a 21 m radius circular window using the following settings:
   - **Input Raster** = wvDEM10m
   - **Output Raster** = demSD
   - **Neighborhood** = *Circle*
   - **Radius** = *11*
   - **Unit Type** = *Cell*
   - **Statistic Type** = *Standard Deviation*
   - **Ignore NoData in calculation** = *checked*



Spatial Analyst Tools → Neighborhood → Focal Statistics

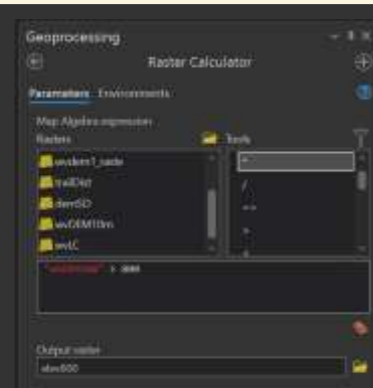
Distance Grid

1. Navigate to the rasterSummarize map.
2. Use the Distance Accumulation Tool to create a raster of the Euclidean distance from trails using the following settings
   - **Input Raster or Feature Source Data** = trails
   - **Output Distance Accumulation Raster** = trailDist
   - Use the default or for all other settings
   - In the environments for the tool, change the **Cell Size** to *Same as Layer* wvDEM10m

Spatial Analyst Tools → Distance → Distance Accumulation

Raster Math 1

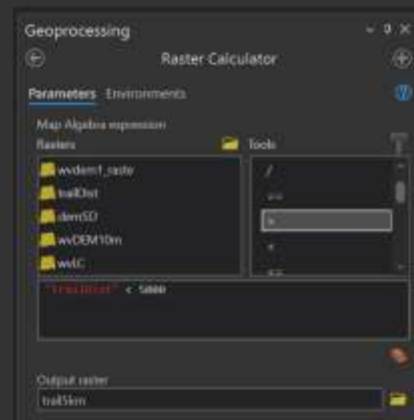
1. Navigate to the rasterMathOverlay map.
2. Use Raster Calculator and the following syntax to create a binary grid of all elevations greater than 800 meters: “wvDEM10m” > 800
3. Name the output elev800.



Spatial Analyst Tools → Map Algebra → Raster Calculator

Raster Math 2

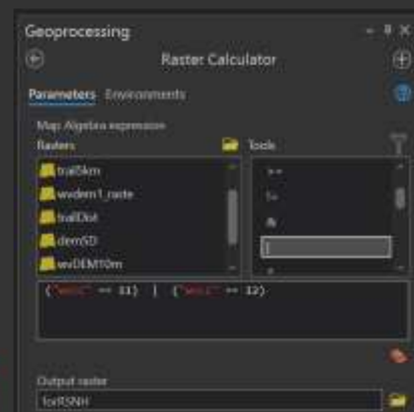
1. Use Raster Calculator and the following syntax to create a binary grid of all areas that are within 5 kilometers of a trail: “trailDist” < 5000
2. Name the output trial5km.



Spatial Analyst Tools → Map Algebra → Raster Calculator

Raster Math 3

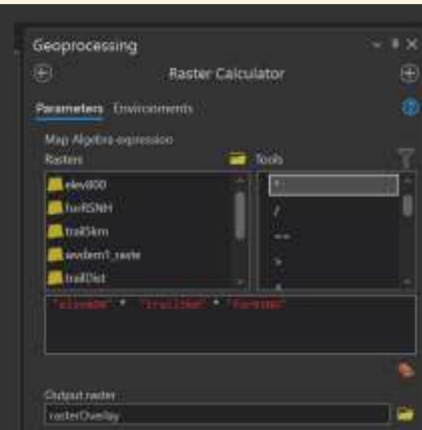
1. Use Raster Calculator and the following syntax to create a binary grid that codes specific forest types (red spruce and northern hardwood) to 1:
   ("wvLC" == 11) | ("wvLC" == 12)
2. Name the output forRSNH.



Spatial Analyst Tools → Map Algebra → Raster Calculator

Raster Math 4

1. Use Raster Calculator to combine the last three results to find all areas that are above 800 m in elevation, within 5 km of a trail, and either a Red Spruce or Northern Hardwood Forest: "elev800" * "trail5km" * "forRSNH"
2. Name the output rasterOverlay.

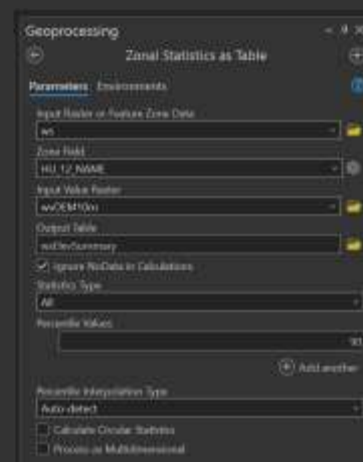


Spatial Analyst Tools → Map Algebra → Raster Calculato

Summarize Continue Raster Grid by Zone

1. Use the Zonal Statistics as Table Tool to obtain elevation summary statistics for each watershed using the following settings:
   - **Input Raster or Feature Zone Data** = ws
   - **Zone Field** = “HU_12_NAME”
   - **Input Value Raster** = wvDEM10m
   - Output Table = wvElevSummary
   - Use the default arguments for all other parameters.

The results will be added to the Contents Pane as a standalone table. Which watershed had the highest mean elevation?

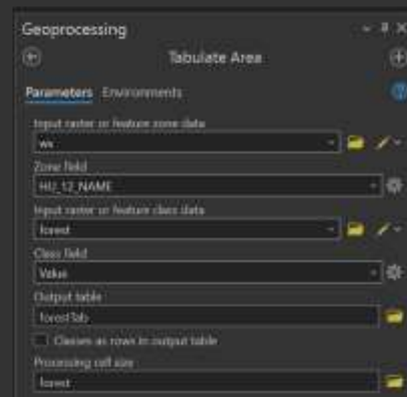


Spatial Analyst Tools → Zonal → Zonal Statistics as Table

87

Summarize Categorical Raster Grid by Zone

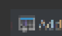
1. Use the Tabulate Area Tool to obtain land areas of forest and not forest cover for each watershed
   - **Input Raster or Feature Zone Data** = ws
   - **Zone Field** = “HU_12_NAME”
   - **Input Raster or Feature Class Data** = forest
   - Output Table = forestTab
   - Use the default arguments for all other parameters.

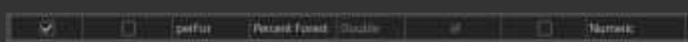


The results will be added to the Contents Pane as a standalone table. Which watershed had the largest land area of forest?

Spatial Analyst Tools → Zonal → Tabulate Area

Convert to Percentage

1. Right-click on the forestTab table in the Contents Pane. Click <u>Open</u> to open a view of the table.
2. In the Table View, click <u>Add</u>. 
3. In the Fields View for the table, set the **Field Name** for the new column to “perFor”, provide an **Alias** of “Percent Forest”, and set the **Data Type** to *Double*. Click <u>Save</u> in the Standalone Table Pane to save the new field.



4. Back in the Table View, right-click on the “Percent Forest” column name then select <u>Calculate Field</u>. In the Calculate Field Window, provide the following syntax: ((!VALUE_1!)/(!VALUE_1! + !VALUE_0!))*100. Click <u>OK</u> to calculate the field.

Which watershed had the highest percentage of forest cover?

Raster-Based Analysis In QGIS

90

The following slides discuss raster analysis using QGIS.

Mask/Clip

- Processing Toolbox → GDAL → Clip Raster by Extent
- Processing Toolbox → GDAL → Clip Raster by Mask Layer
- Processing Toolbox → SAGA → Features – Features-Raster Tools → Clip Raster with Polygon
- Processing Toolbox → SAGA → Features – Raster-Tools → Clip Rasters

91

Multiple tools are available for extracting out spatial subsets of raster data.

Stack and Merge

- Processing Toolbox → GDAL → Raster Miscellaneous → Merge

- To stack or composite bands = select “Place each input into a separate band”
- To merge or mosaic = Do not select “Place each input into a separate band”

92

The Merge tool can be used to both merge multiple raster grids (similar to the Mosaic to New Raster tool in ArcGIS Pro) and composite or stack image bands into a multi-band or layer file (similar to the Composite Bands tool in ArcGIS Pro).

Resample

- Processing Toolbox → GRASS → Raster → r.resample
- Processing Toolbox → SAGA → Raster-Tools → Resampling

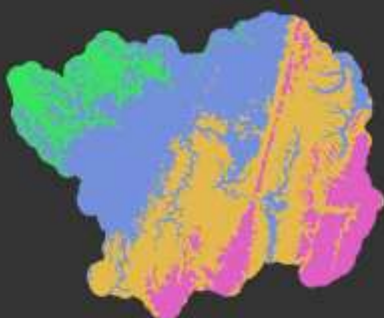
93

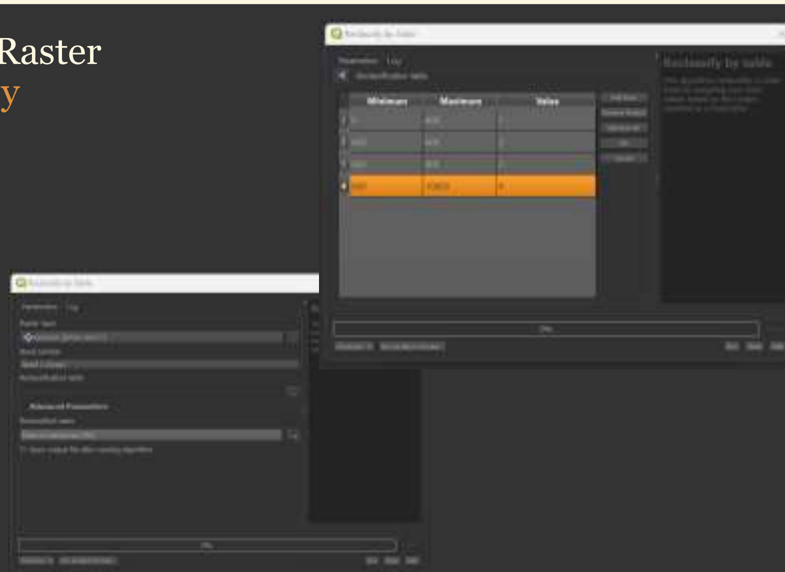
In QGIS, raster data can be resampled using several tools. The most common is the Warp/Reproject tool, which can change a raster's coordinate reference system, cell size, extent, and resampling method. The Align Rasters tool is useful when multiple raster grids need to share the same resolution, extent, CRS, and grid alignment. The Translate tool or Export → Save As option can also be used to create a new raster with a different resolution.

Methods are also made available through GRASS and SAGA.

Reclassify

❖ Processing Toolbox → Raster Analysis → Reclassify by Table





94

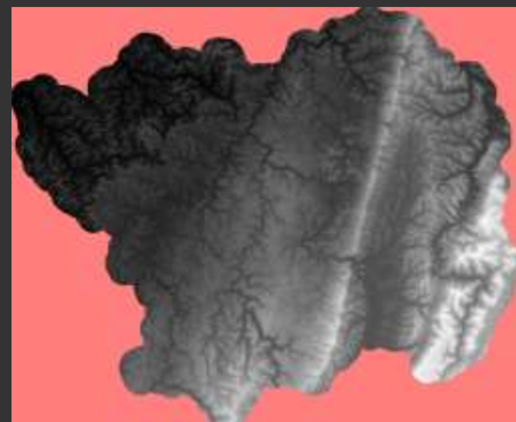
Reclassification is performed using the Reclassify by Table tool.

NoData

- Check NoData Value:

  Right-Click → Properties → Transparency

- Can define NoData values at export

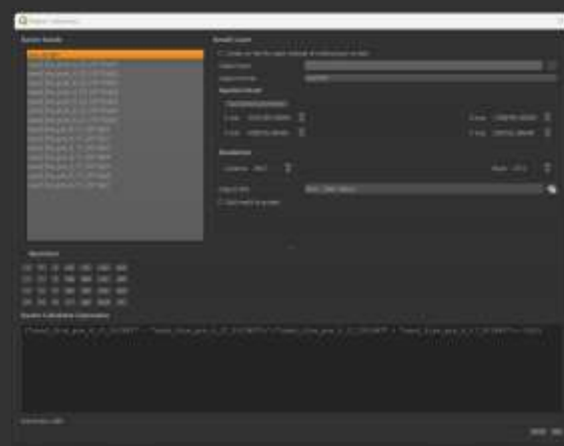


95

This slide notes how to visualize NoData values in QGIS. You can also define a NoData value when exporting raster data.

Raster Calculator

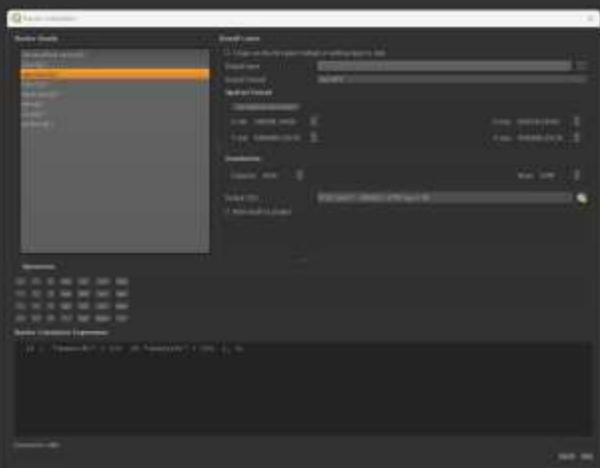
- Processing Toolbox → Raster Analysis → Raster Calculator
- Bands accessed via @Band Number
- Operators: +, -, *, /, >, <, >=, <, =, !=, ^, abs(), IF(), AND, OR, sqrt(), cos(), sin(), tan(), acos(), asin(), atan(), log10(), ln()



96

QGIS, similar to ArcGIS Pro, has a Raster Calculator tool. This slide notes how to access individual bands of a multiband raster grid and also some available operators or functions.

Conditional Statements



- Can use IF for conditional statement implementation
- IF(Condition, What to Return if TRUE, What to Return if FALSE)

97

This slide notes now to implement Raster Calculator conditional statements in QGIS using the if() function.

QGIS Conditional Statement

- Use the utah project.
- If the land cover is a forest type (41, 42, or 43), recode the values to 40. Otherwise, return the original land cover.

98

The task on this slide can be completed using the Raster Calculator tool and a conditional statement. It could also be answered using Reclassify by Table.

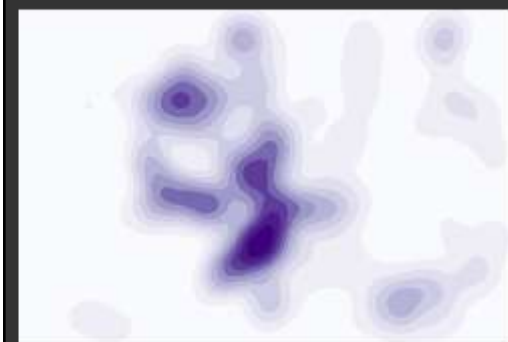
Euclidean Distance

- Processing Toolbox → GDAL → Proximity (Raster Distance)
- Source vector data must be rasterized first
- Processing Toolbox → GDAL → Rasterize (vector to raster)

99

These tools support calculation of Euclidean distance.

Kernel Density



- Processing Toolbox → Interpolation → Heatmap (Kernel Density Estimation)
- Processing Toolbox → SAGA → Raster – Rasterize → Kernel Density Estimation

100

These tools support calculation of kernel density grids.

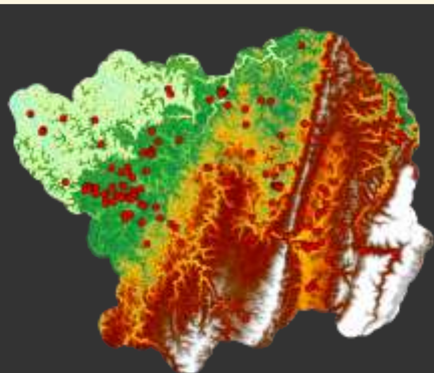
Moving Windows

- Processing Toolbox → SAGA → Spatial and Geostatistics – Raster → Focal Statistics
- Calculate a variety of statistics within a moving window
- Can change window size
- Can change window shape
- Can apply weights in window

101

This slide explores the Focal Statistics tool for using moving windows to calculate local statistical or summary measures.

Extract Pixel Values at Points



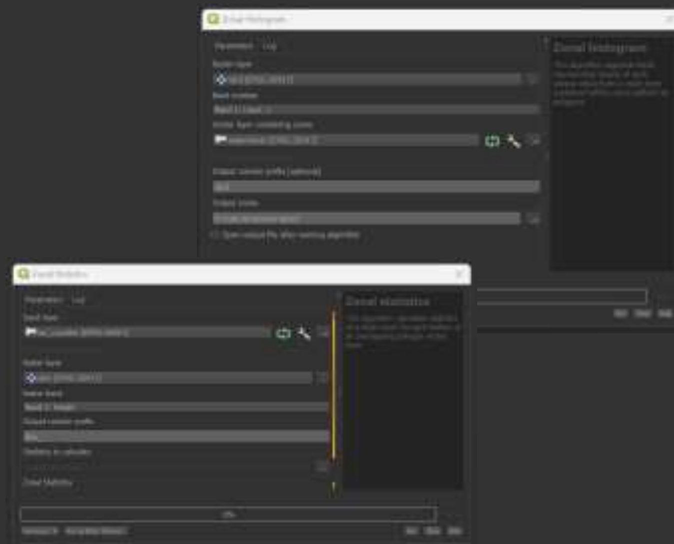
- Processing Toolbox → Raster Analysis → Sample Raster Values at Points
- Works for multiband/image data

102

The Sample Raster Values at Points tool allows for the extraction of raster cell values at point locations to the point layer's attribute table.

Zonal Statistics

- For Continuous Data: Processing Toolbox → Raster Analysis → Zonal Statistics
- For Categorical Data: Processing Toolbox → Raster Analysis → Zonal Histogram
- Results written to polygon attribute table



103

Zonal Statistics is used to summarize continuous raster data by zone while Zonal Histogram summarizes categorical raster data by zone. Both tools write the results to the polygon layer's attribute table.

QGIS Raster Calculator Exercises

- Use the utah project.
- Create a binary raster to differentiate areas with an elevation greater than 2,000 meters.
- Create a binary raster to differentiate areas that are have an Evergreen Forest land cover type.
- Create a binary raster of areas that are above 2,000 meters in elevation and have Evergreen Forest land cover type.

The questions on this slide can be answered using raster math.

QGIS Raster Summary at Points

- Use the utah project.
- Which town has the highest elevation?

105

This question can be answered using the Sample Raster Values at Points tool.

Continuous Raster Grid Summary in QGIS

- Use the utah project.
- Which county has the lowest mean elevation?

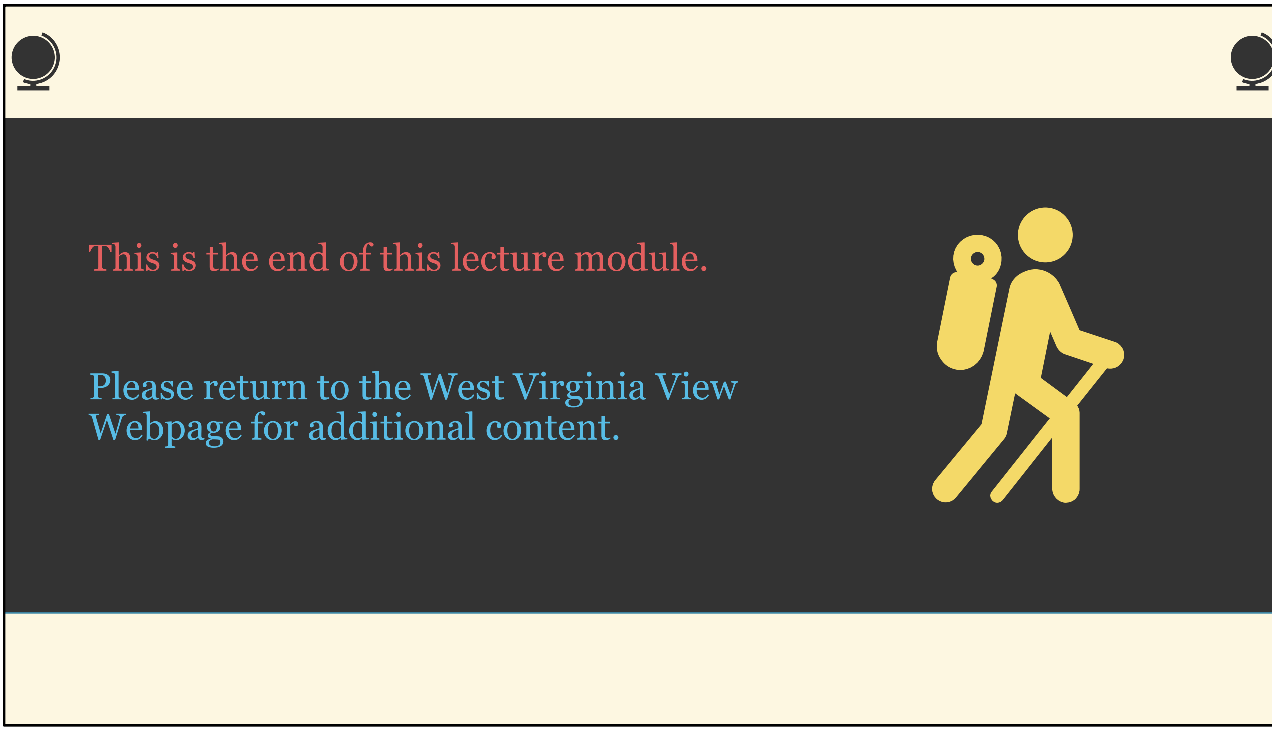
This question can be answered with the Zonal Statistics tool.

Categorical Raster Summary in QGIS

- Use the utah project.
- Which county has the highest percentage land area of forested land cover?

107

This question can be answered with the Zonal Histogram tool.



Thanks! Hope you found this useful.