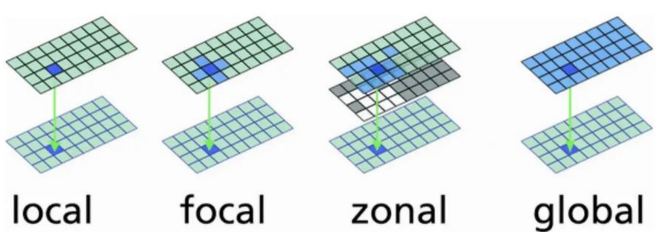
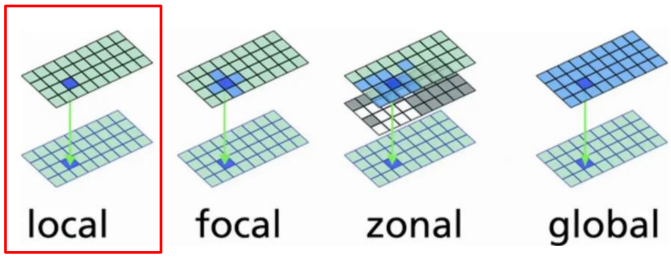
Geographic Information Systems 2024/2025

Lesson 15 - Raster data processing, raster alignment

Types of spatial operations using raster data



Types of spatial operations using raster data



Local Operations

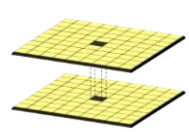
Let x_{ij}, y_{ij}, ... be the cell values row *i* and column *j* of raster gds *X*, *Y*, ..., all representing the same region and having identical resolution, and let Z be a new raster gds having identical resolution and number of rows and columns to X and Y

<u>LOCAL FUNCTIONS:</u> $z_{ij}=f(x_{ij})$ or $z_{ij}=f(x_{ij},y_{ij},\ldots)$

every cell value of *Z* only depends on the corresponding cell values of *X*, *Y* ...

examples: $z_{ij} = x_{ij} \times y_{ij}$, $z_{ij} = x_{ij}^2 \times y_{ij} + w_{ij}$ ... (Map Algebra)

$$\overset{X}{\begin{bmatrix} 1 & 4 & 5 \\ 5 & 3 & 2 \\ 2 & 5 & 2 \end{bmatrix}} + \overset{Y}{\begin{bmatrix} 5 & 1 & 3 \\ 1 & 2 & 1 \\ 1 & 4 & 2 \end{bmatrix}} = \overset{Z}{\begin{bmatrix} 6 & 5 & 8 \\ 6 & 5 & 3 \\ 3 & 9 & 4 \end{bmatrix}}$$



Images from https://gisgeography.com/map-algebra-global-zonal-focal-local/

Local Operations

- Performed on a **cell-by-cell** basis
- Output cell values depend only on the corresponding input cell(s)
- No consideration of neighboring cells

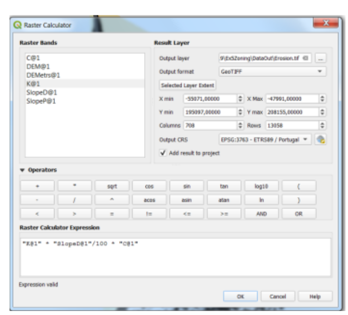
Function	Description	Example
Arithmetic	Add, subtract, multiply, or divide raster values.	Combine elevation and rainfall for flood risk.
Logical	Apply Boolean logic (AND, OR, NOT) or conditional statements.	Identify areas where elevation > 100m AND slope < 5°.
Reclassification	Assign new values based on ranges (e.g., 0–10 → 1, 11–20 → 2).	Convert continuous elevation into discrete classes.
Trigonometric	Calculate sine, cosine, or tangent of cell values.	Model solar radiation angles.

GIS Tools:
- **ArcGIS**: Raster Calculator, Reclassify
- **QGIS**: Raster Calculator, Reclassify by Table

Local Operations – Arithmetic

Arithmetic Overlay: given 2 gds *X* & *Y*, an arithmetic overlay returns a raster gds Z such as $z_{ij}=f(x_{ij},y_{ij})$ for every cell, row *i* and column *j*

- Function f can be `+`, `-`, `*`, `max`, `min`, a logic operation (`and`, `or`, `not`), ..., according to the domain of X and Y
- Arithmetic overlay functions are also called Map Algebra
- The ArcGIS/QGIS Raster Calculator is used to execute arithmetic overlay functions with any number of input gds



Local Operations – Arithmetic

Example

In a given region, $A_{ij}=h$ represents height and $B_{ij}=t$ represents temperature.

Admitting that in the region there is a relation among height, temperature & precipitation (p) such as:

$p = h * t / 0.6 + 123$ (precipitation model)

it is possible to create a raster gds representing the estimated precipitation $C_{ij} = p$ executing

$C = A * B / 0.6 + 123$

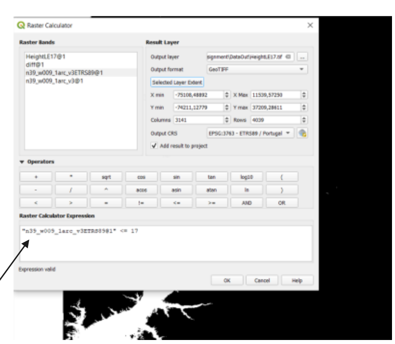
Local Operations – Boolean

Boolean gds creation: given one raster gds X and a logical expression involving X pixels, say $e(x_{ij})$, returns a raster gds Z

such as $$z_{ij} = \begin{cases} 1(True) & if \quad e(x_{ij}) = True \\ 0(False) & if \quad e(x_{ij}) = False \end{cases}$$

Boolean gds can be created using :

- **ArcGIS**: Raster Calculator, Reclassify
- **QGIS**: Raster Calculator, Reclassify by Table



"n39_w009_1arc_v3ETRS89@1" <= 17

QUIZ

Raster gds *A* represents a given study area; if A_{ij} is irrigated then $A_{ij}=1$, otherwise $A_{ij}=0$.
Raster gds *B* represents the same study area; if B_{ij} is a corn area then $B_{ij}=1$, otherwise $B_{ij}=0$.

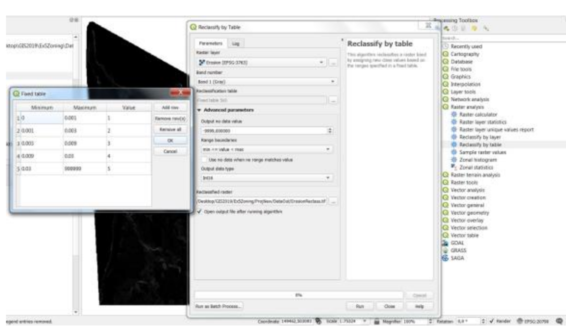
- What operation can be used to identify irrigated **and** corn areas?
  "A@1" × "B@1“
- What operation(s) can be used to identify irrigated **or** corn areas?
  ("A@1" + "B@1) = 2 OR ("A@1" + "B@1”) = 1
- What operation can be used to identify irrigated areas that **are not** corn areas?
  "A@1" x (NOT "B@1”)

Local Operations – Reclassify

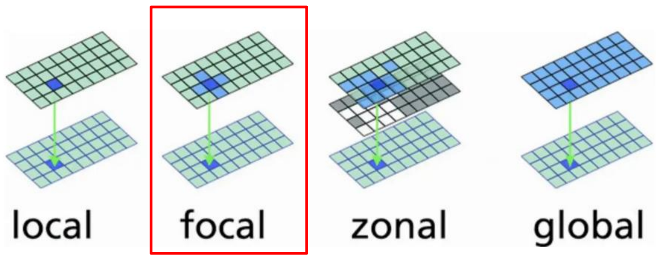
Reclassify: Given gds *X* and one function of one variable f, the new gds *Z* is defined as $z_{ij}=f(x_{ij})$

- Function f may be defined by a table assigning one value to every interval of X domain

  Example:



Types of spatial operations using raster data

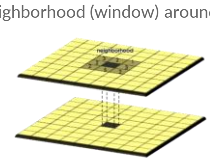


Focal (Neighborhood) Operations

Let x_{ij}, y_{ij}, ... be the cell values row *i* and column *j* of raster gds *X*, *Y*, ..., all representing the same region and having identical resolution, and let Z be a new raster gds having identical resolution and number of rows and columns to X and Y

FOCAL FUNCTIONS: $z_{(ij)} = f(x_{(i+p,j+k)})$ or $z_{(ij)} = f(x_{(i+p,j+k)}, y_{(i+p,j+k)}, \ldots)$ where $p,k \in N$, typically with $-n \leq p,k \leq n$ for some $n \geq 1$

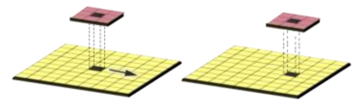
every cell value of Z depends on the values within a defined neighborhood (window) around the corresponding cell in X, Y, ...

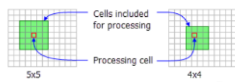


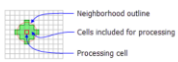
Images from https://gisgeography.com/map-algebra-global-zonal-focal-local/

Focal (Neighborhood) Operations

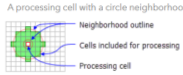
- Output cell values depend on a **neighborhood** of input cells (e.g., 3x3 window)
- It commonly uses a moving window
  - arrangement of cells (e.g., circular, rectangular, etc) that shifts in position



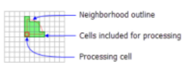




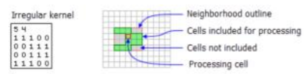
A processing cell with a circle neighborhood (radius = 2 cells).



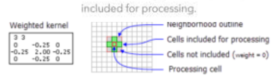
A processing cell with the default annulus neighborhood (inner radius = 1 cell, outer radius = 3 cells).



A processing cell with the default wedge neighborhood (radius = 3 cells, start angle = 0, end angle = 90).



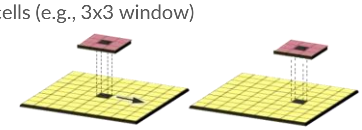
An irregular neighborhood kernel is shown with an illustration of which cells will be included for processing.



A weighted neighborhood kernel is shown with an illustration of which cells will be included for processing.

Focal (Neighborhood) Operations

- Output cell values depend on a **neighborhood** of input cells (e.g., 3x3 window)
- It commonly uses a moving window
  - arrangement of cells (e.g., circular, rectangular, etc) that shifts in position
  - applying a moving window operation to each cell smooths values in a raster
- Used for spatial filtering, smoothing, or edge detection



Function	Description	Example
Focal Statistics	Mean, median, sum, standard deviation within a window.	Smooth DEM noise with a 3x3 mean filter.
Edge Detection	Kernels (e.g., Sobel, Laplacian) highlight boundaries.	Detect roads in satellite imagery.
Morphological	Erosion/dilation to modify feature shapes.	Clean up classified land cover data.

GIS Tools:

- **ArcGIS**: Focal Statistics, Kernel Density
- **QGIS**: SAGA Morphological Filter, GDAL Proximity

Focal (Neighborhood) Operations – Examples

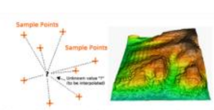
Sum Statistic:

- rectangular 3 by 3 cell neighborhood shape is specified
- Ignore No Data in calculations parameter

Focal (Neighborhood) Operations – Examples

Inverse Distance Weighted (IDW)

- this method assumes things that are close to one another are more alike than those farther apart (true in many geographic phenomena under some constraints)
- estimates cell values based on the weighted average of values from nearby points
- weights are inversely proportional to the distance between the estimation location (the cell) and the known points
- The greater the distance weight parameter, the less the effect points will have if they are far from the estimated point.

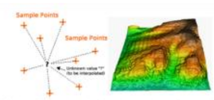


Inverse Distance Weighted interpolation based on weighted sample point distance (left). Interpolated IDW surface from elevation vector points (right). Image Source: Mitas, L., Mitasova, H. (1999).

Focal (Neighborhood) Operations – Examples

Inverse Distance Weighted (IDW)

- this method assumes things that are close to one another are more alike than those farther apart (true in many geographic phenomena under some constraints)
- estimates cell values based on the weighted average of values from nearby points
- weights are inversely proportional to the distance between the estimation location (the cell) and the known points
- The greater the distance weight parameter, the less the effect points will have if they are far from the estimated point.
- Estimated values are always between the minimum and maximum values of the sample
- Estimations are better inside the convex hull enclosing all the sample points

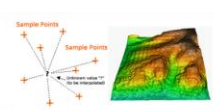


Inverse Distance Weighted interpolation based on weighted sample point distance (left). Interpolated IDW surface from elevation vector points (right). Image Source: Mitas, L., Mitasova, H. (1999).

Focal (Neighborhood) Operations – Examples

Why is IDW a Focal Operation:

- output value for a given cell depends on the values of surrounding cells within a defined neighborhood
- extent of the neighborhood depends on the search radius, number of points, or other settings specified in the IDW algorithm
- calculation involves a moving window that shifts across the raster
- at each location, values within the window influence the value of the central cell



Inverse Distance Weighted interpolation based on weighted sample point distance (left). Interpolated IDW surface from elevation vector points (right). Image Source: Mitas, L., Mitasova, H. (1999).

Focal (Neighborhood) Operations – Examples

Slope (a common GIS definition):

- slope for a cell in a raster is the steepest slope of a plane defined by the cell and its eight surrounding neighbors

DEM

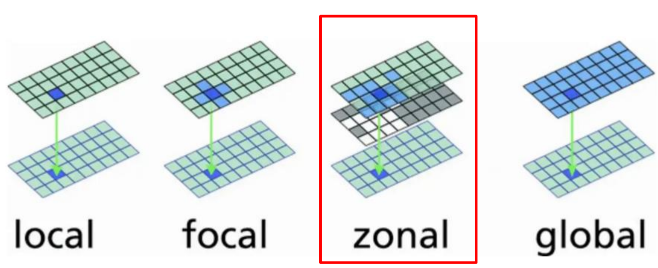
19	15	10
17	10	12
16	11	14

Slope (%)

…	…	…
…	S	…
…	…	…

$S = (19 - 10)/x \times 100$
where $x = \sqrt{2} \times r$
(r is the raster resolution)

Types of spatial operations using raster data



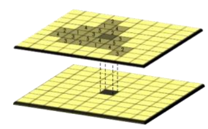
Zonal Operations

Let x_{ij}, y_{ij}, ... be the cell values row *i* and column *j* of raster gds *X*, *Y*, ..., all representing the same region and having identical resolution, and let Z be a new raster gds having identical resolution and number of rows and columns to X and Y

ZONAL FUNCTIONS: given a neighbourhood *V* of cell *(i,j)*, let X_{ij} be the set of cell values of gds *X* in the neighbourhood *V*, then $z_{ij}=f(X_{ij})$ or $z_{ij}=f(X_{ij},Y_{ij},...)$

every cell value of *Z* depends on the cell values in neighbourhood *V* of the corresponding cell of *X*, *Y*, ...

examples: $z_{ij} = avg\ (X_{ij})$, $z_{ij} = slope$



Images from https://gisgeography.com/map-algebra-global-zonal-focal-local/

Zonal Operations

- Apply a math function to a group of cells within a specified zone
- Output values depend on **zones** defined by a second raster or vector layer
- Aggregates statistics for cells within the same zone

Function	Description	Example
Zonal Statistics	Mean, sum, max/min for each zone.	Calculate average rainfall per watershed.
Zonal Geometry	Compute area, perimeter, or centroid of zones.	Measure forest patch sizes.
Zonal Majority	Identify the most frequent value within a zone.	Determine dominant land cover type per region.

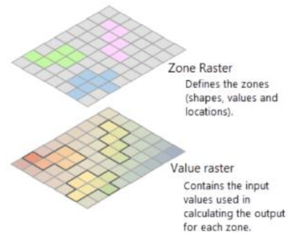
GIS Tools:

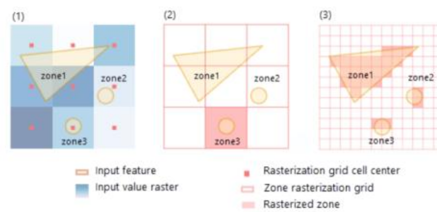
- **ArcGIS**: Zonal Statistics, Zonal Geometry
- **QGIS**: Zonal Statistics, Raster Analysis

Zonal Operations - Example

Zonal statistics

- In 1):
  - Zone 1 is larger than an individual cell
  - Zone 2 and zone 3 are smaller than a cell
  - A cell center falls outside zone 2 but within zone 3
- During the zone rasterization process in 2):
  - since no cell centers fall within zone 1 and zone 2 these disappear
  - only zone 3 is rasterized
- To avoid this, create more cell centers by specifying a smaller cell size in the environment

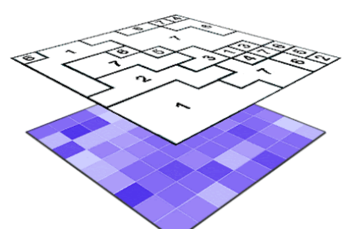


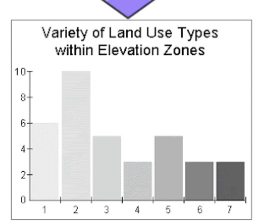


Zonal Operations - Example

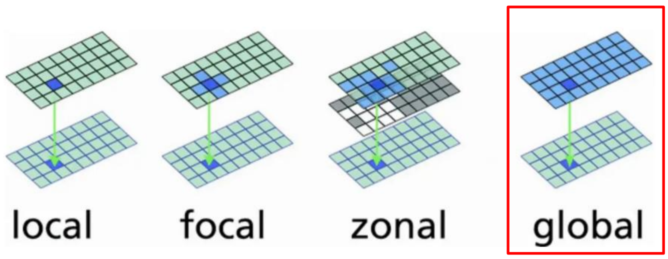
Zonal statistics – Variety

- In 1):
  - Zone 1 is larger than an individual cell
  - Zone 2 and zone 3 are smaller than a cell
  - A cell center falls outside zone 2 but within zone 3
- During the zone rasterization process in 2):
  - since no cell centers fall within zone 1 and zone 2 these disappear
  - only zone 3 is rasterized
- To avoid this, create more cell centers by specifying a smaller cell size in the environment





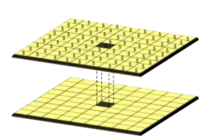
Types of spatial operations using raster data



Global Operations

Let x_{ij}, y_{ij}, ... be the cell values row *i* and column *j* of raster gds *X*, *Y*, ..., all representing the same region and having identical resolution, and let Z be a new raster gds having identical resolution and number of rows and columns to X and Y

<u>GLOBAL FUNCTIONS:</u> every z_{ij} value depends on all cell values of *X*, *Y*, ... and on their coordinates.



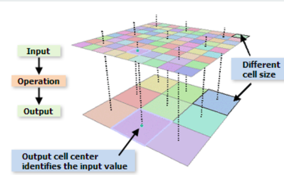
Images from https://gisgeography.com/map-algebra-global-zonal-focal-local/

Global Operations

- Applies a bulk change to all cells in a raster
- Evaluate relationships across the full extent of the data
- Involve iterative algorithms that consider the entire dataset

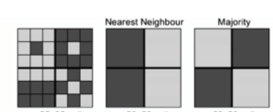
Operation	Description	Input	Output
Euclidean Distance	Calculates the straight-line distance from each cell to the nearest source.	Source raster.	Raster showing the distance from each cell to the closest source cell.
Cost Distance	Calculates the least accumulative cost distance from each cell to the nearest source.	Source raster, cost raster.	Raster showing the least accumulative cost distance from each cell to the source.
Viewshed Analysis	Identifies the areas visible from a given point or a set of points.	DEM, observer location(s).	Raster showing which cells are visible from the observer location(s).
Flow Accumulation	Calculates the accumulated flow of water over each cell.	Flow direction raster.	Raster where each cell contains the accumulated weight of all cells.
Least-Cost Path	Determines the path with the minimum cumulative cost between two points.	Cost raster, starting point, ending point.	Polyline feature representing the least-cost path.

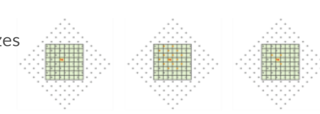
Global Operations - Example



Resampling – Adjusts pixel size of a raster dataset to match a new resolution or coordinate system

- required because the source & target grid cells do not match
- involves interpolating new pixel values based on existing ones
    - can also be considered as interpolation methods
- original cell values will affect what will be the next cells values
- directions:
    - Upsampling - converting to higher resolution/smaller cells.
    - Downsampling - converting to lower resolution/larger cell sizes
- involves decisions about interpolation methods





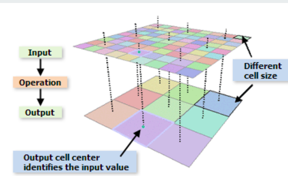
Global Operations - Example

Resampling interpolation methods:

- **Nearest Neighbor**: the new pixel will have the value of the nearest pixel in the input raster (the distances to pixel centres are used to identify the nearest pixel in the input raster)
- **Bilinear** (2x2 kernel): a method to interpolate functions of 2 variables
- **Cubic** (4x4 kernel): Cubic Convolution Approximation
- **Cubic B-Spline** (4x4 kernel): Cubic B-Spline Approximation
- **Lanczos** (6x6 kernel): Lanczos windowed sinc interpolation
- **Average**: computes the average of all non-NODATA contributing pixels
- **Mode:** selects the value which appears most often on all the sampled points
- **Maximum**, **Minimum**, **Median**, **First Quartile** (Q1) or **Third Quartile** (Q3) of all non-NODATA contributing pixels

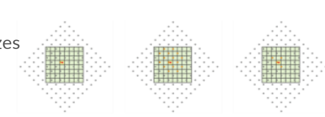
* Underlined, methods that can only be applied to continuous variables

Global Operations - Example



Resampling – Adjusts pixel size of a raster dataset to match a new resolution or coordinate system

- required because the source & target grid cells do not match
- involves interpolating new pixel values based on existing ones
  - can also be considered as interpolation methods
- original cell values will affect what will be the next cells values
- directions:
  - Upsampling - converting to higher resolution/smaller cells.
  - Downsampling - converting to lower resolution/larger cell sizes
- involves decisions about interpolation methods
- some resampling methods account for the entire dataset
- resampling is required as part of reprojection



Global Operations - Example

CRS Conversion – Process of transforming spatial data from 1 CRS to another

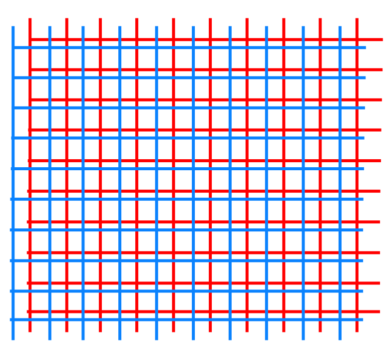
- CRS defines how the 2-dimensional, projected map in your GIS relates to real places on the earth (includes information about projection, datum, & ellipsoid)
- transforms raster data from 1 CRS to another, encompassing both geometric transformation & datum adjustments
- involves:
  - changing underlying mathematical model (projection, datum, ellipsoid) that interprets coordinates
  - may involve datum transformation (e.g., NAD83 to WGS84), projection transformation (e.g., geographic to UTM), or both
- simple: assigning a new CRS to data (if the data is unprojected but you know the correct CRS)
- complex: transforming coordinates from one CRS to another, which may require mathematical calculations & possibly additional transformation files

Reprojection

Raster operations

Operations between different raster layers ideally requires them to have the same structure:

- same resolution (cell size)
- same extent

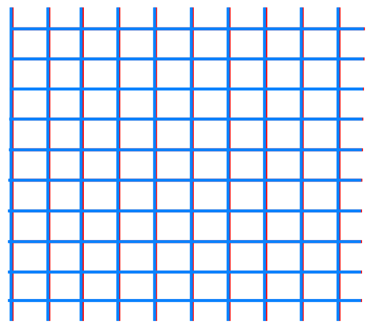


Raster operations

Operations between different raster layers ideally requires them to have the same structure:

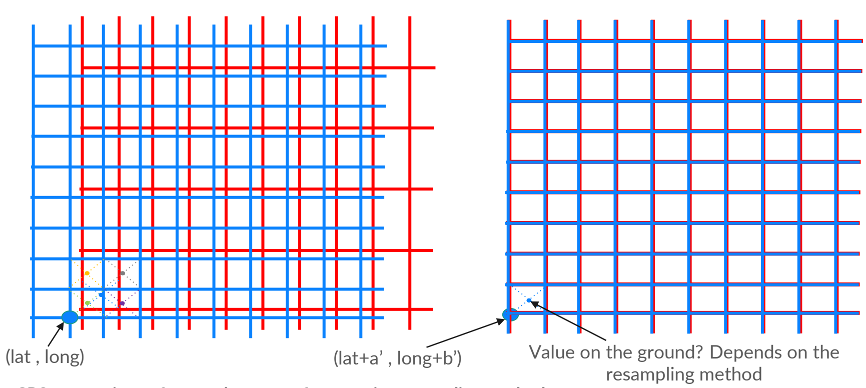
- same resolution (cell size)
- same extent

Raster alignment



Raster alignment

Grids with **same CRS** but **different spatial resolution**: Geographic translation of grid in blue & interpolation



CRS conversions of raster data sets often require resampling methods

Raster alignment & resample tools

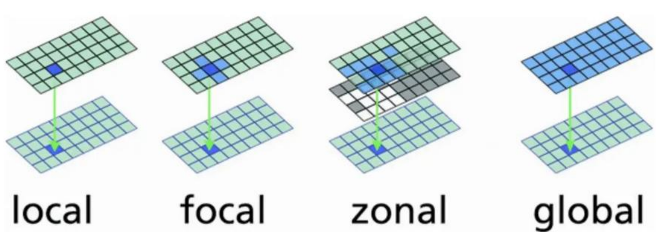
QGIS

- The tool **Align Raster** can take several rasters as input and to **align them perfectly**, which means: to resample to the **same cell size** and **offset** (or location) in the grid.

ArcGIS

- The tool **Resample** will create new rasters with a new cell resolution. The Environment parameters can be used to set the extent, output CRS.
- Other operations include **Project** and **Extract**.

Types of spatial operations using raster data



Types of spatial operations using raster data

Detailed Summary

Characteristic	Local Operations	Focal Operations	Zonal Operations	Global Operations
Spatial Scope	Individual cells only	Defined neighborhood around each cell	Predefined zones (regions)	Entire raster dataset
Mathematical Expression	z(i,j) = f(x(i,j), y(i,j), ...)	z(i,j) = f(x(i+p,j+k), y(i+p,j+k), ...) where -n≤p,k≤n	z(i,j) = f(all x,y in zone Z)	z(i,j) = f(all x,y in raster)
Input Dependencies	Values at same position only	Values within a moving window	Values within the same zone	Potentially all values in raster
Window/Pattern	No window (point)	Typically 3×3, 5×5, circular, annular, or wedge	Irregular shapes defined by zones	Entire raster extent
Edge Handling	No special handling needed	Requires special treatment for edges	No edge issues (zones define boundaries)	Handles entire extent

Types of spatial operations using raster data

Detailed Summary

Characteristic	Local Operations	Focal Operations	Zonal Operations	Global Operations
Common Functions	Arithmetic, Boolean, reclassification	Mean, median, max/min, standard deviation, slope	Mean, sum, majority, variety per zone	Distance, cost path, viewshed
Example Applications	Land suitability modeling, combining datasets	Image smoothing, edge detection, terrain analysis	Watershed analysis, administrative statistics	Distance to features, least-cost paths
Output Type	Raster with same dimensions	Raster with same dimensions	Raster or table with one value per zone	Raster or single summary values
Processing Complexity	Low (O(n) where n=cells)	Medium (O(n×w) where w=window size)	Medium-high (O(n) but with zone tracking)	High (O(n²) for some algorithms)

Types of spatial operations using raster data

Detailed Summary

Characteristic	Local Operations	Focal Operations	Zonal Operations	Global Operations
Typical GIS Tools	Raster Calculator	Focal Statistics, Filter	Zonal Statistics	Euclidean Distance, Cost Distance
Visualization Example	Cell-by-cell math (A+B)	Smoothed surface, edge enhancement	Regional averages, statistics by boundary	Distance surfaces, viewsheds
Key Limitations	Cannot detect patterns	Limited to local patterns	Requires zone definition	Computationally intensive
Data Requirements	Single or multiple aligned rasters	Single raster + neighborhood definition	Value raster + zone definition	Single or multiple rasters

Credits

The structure and content of the course, including most of the slides and exercises was developed by Professor Graça Abrantes (SIG 2021_2022) and Professor Rui Figueira (SIG 2023_2024)