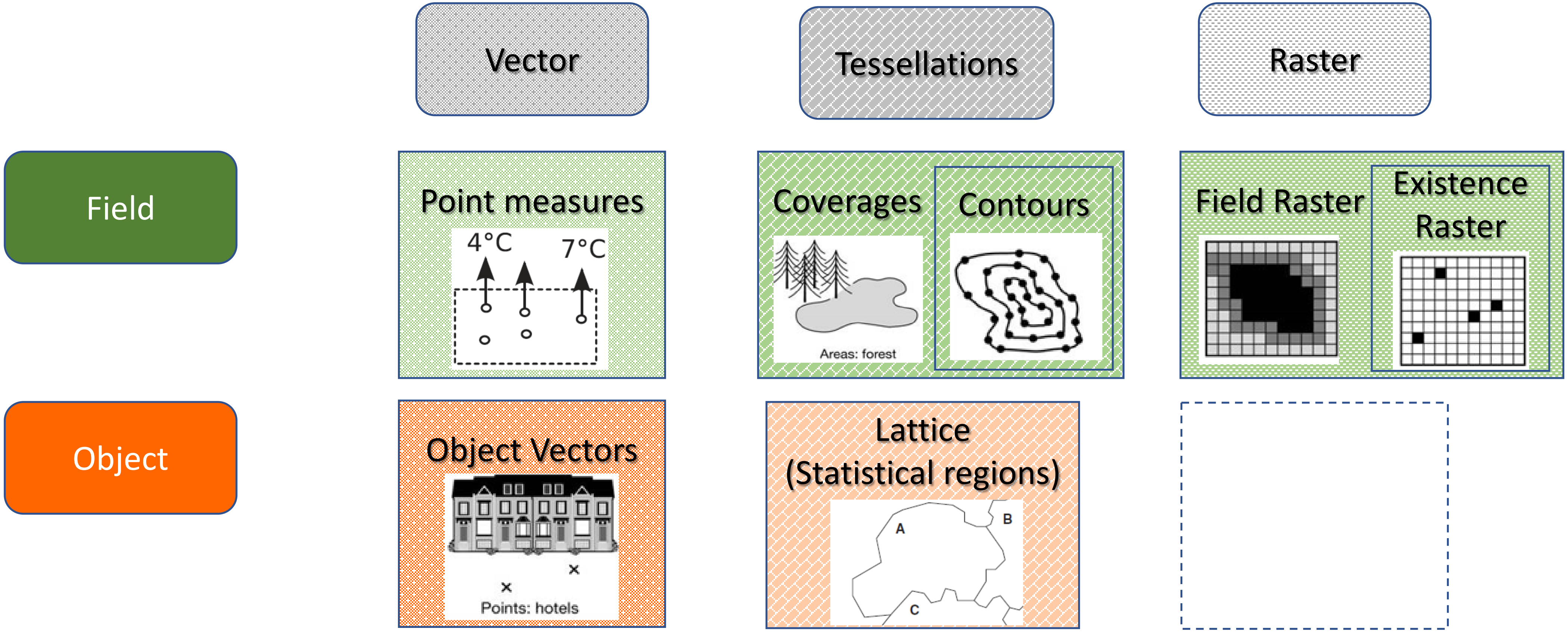
Distance based analysis

Spatial Data Analysis and Simulation modelling,
2020, Simon Scheider

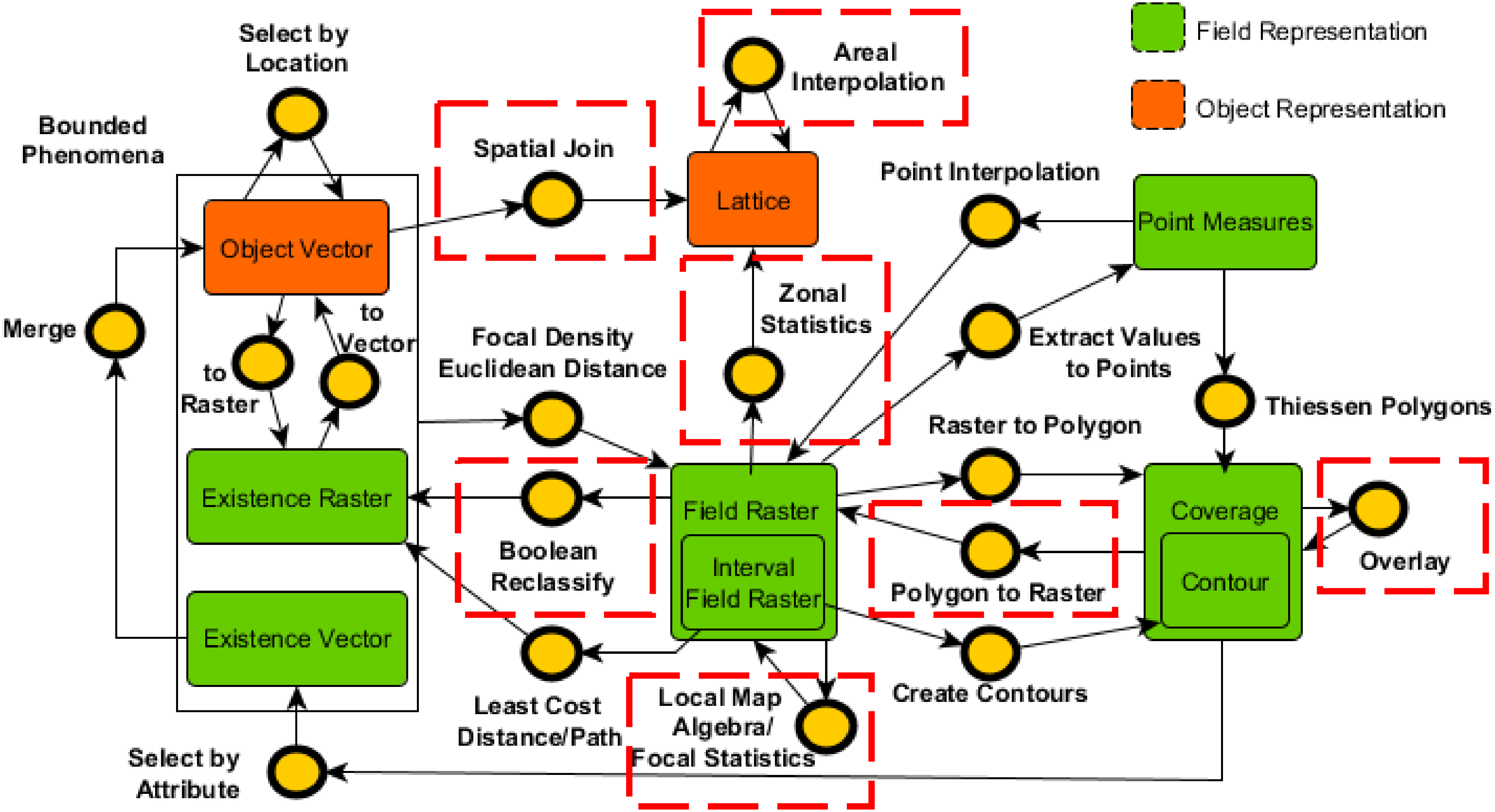
Outline

- Core concepts (recap)
- Vector Distance-Based Analysis:
  - Planar vs geodesic distances
  - Buffers
  - Nearest objects
  - Thiessen Polygons
- Raster Distance-Based Analysis:
  - Focal map algebra
  - Global map algebra (Proximity analysis)
  - Point Interpolation
- Core concept quiz

How geodata models represent core concepts

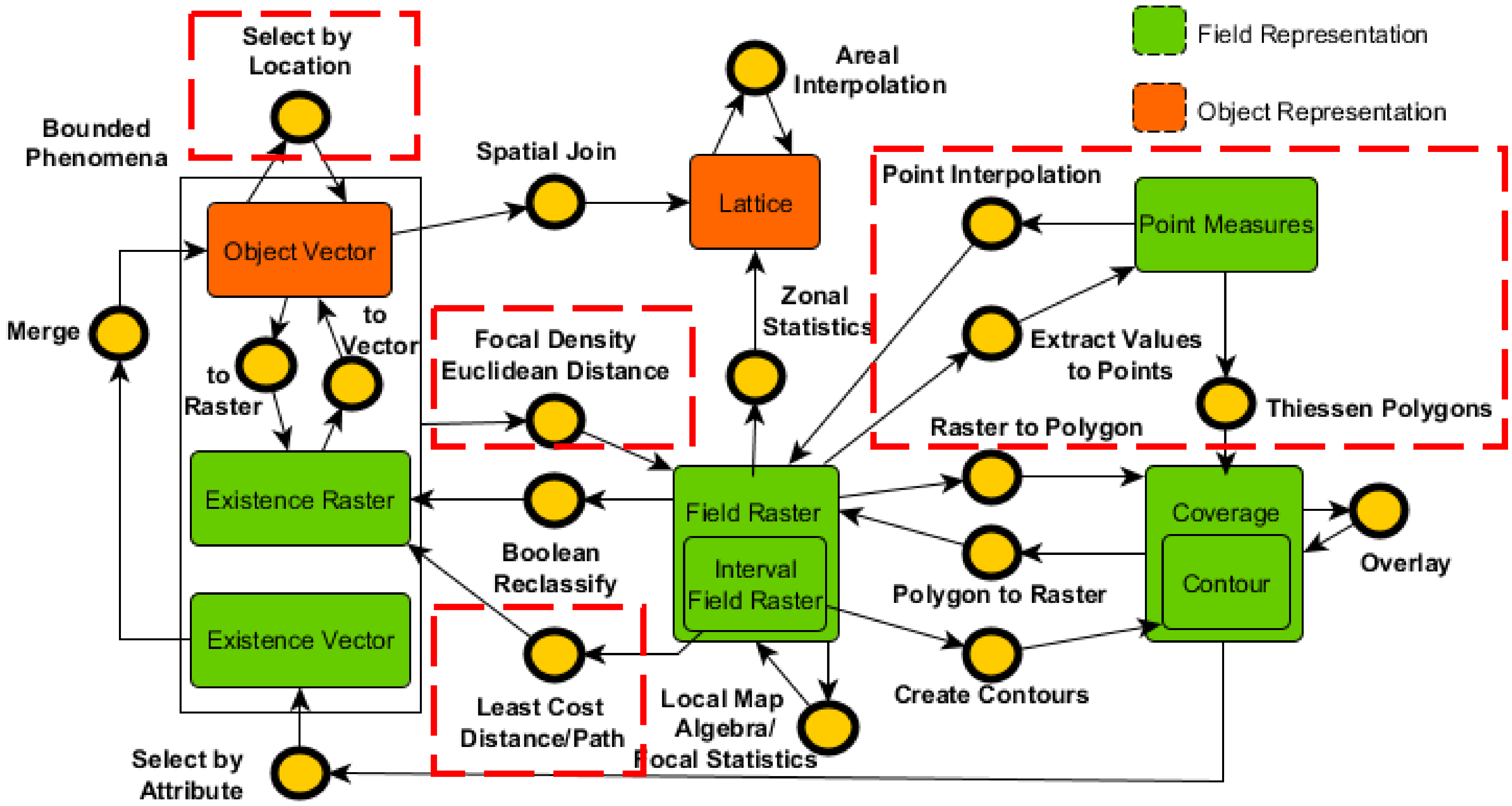


In which ways can geodata be transformed?



Overlay and aggregation (last lecture)

In which ways can geodata be transformed?



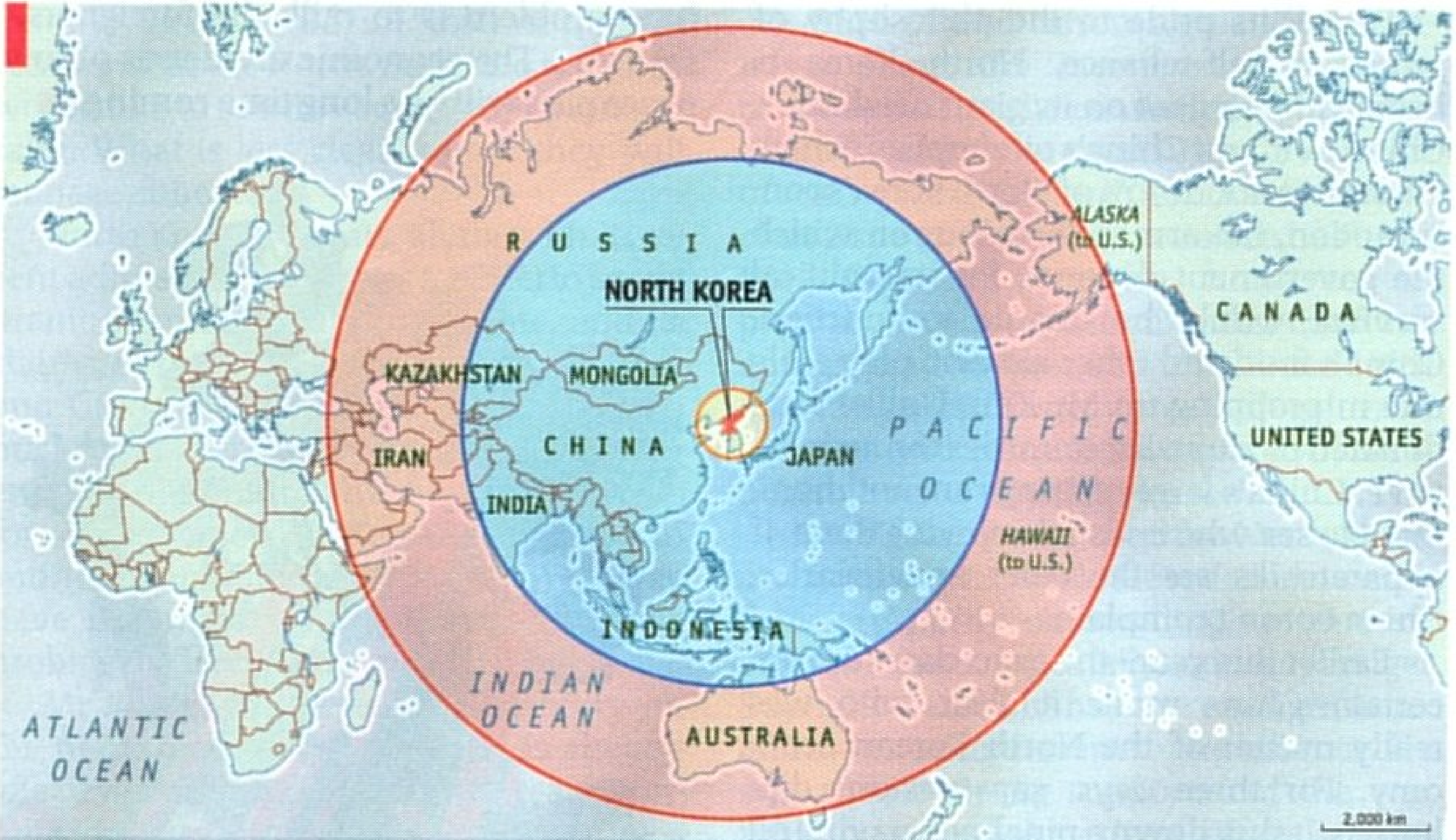
Distance based analysis (this lecture)

Core concept quiz

In order to make the workflows that you implement in the lab reusable on different data layers, think for a moment how they correspond to core concept transformations:

1) which core concepts correspond to which datasets in the lab?
2) which path in the computational diagram corresponds to which one of your workflows?

Vector Distance Analysis



North Korea's missile threat

Type	Maximum range	Payload	Status
Nodong	1,300 km (810 miles)	700 kg (1,550 pounds)	Currently deployed
Taepodong-1	Up to 10,000 km	Several hundred kg	Test failed 1998, not yet operational
Taepodong-2	10,000-15,000 km	Several hundred kg	Not yet tested

Source: Task Force for US Korea Policy, Centre for International Policy

The Economist, May 3, 2003

Flat-earth thinking. Thank you to those readers who pointed out that, by superimposing concentric circles on a Mercator projection, the map in our May 3rd issue greatly underestimated the potential reach of North Korea's missiles. We stand corrected.

North Korea's missile threat

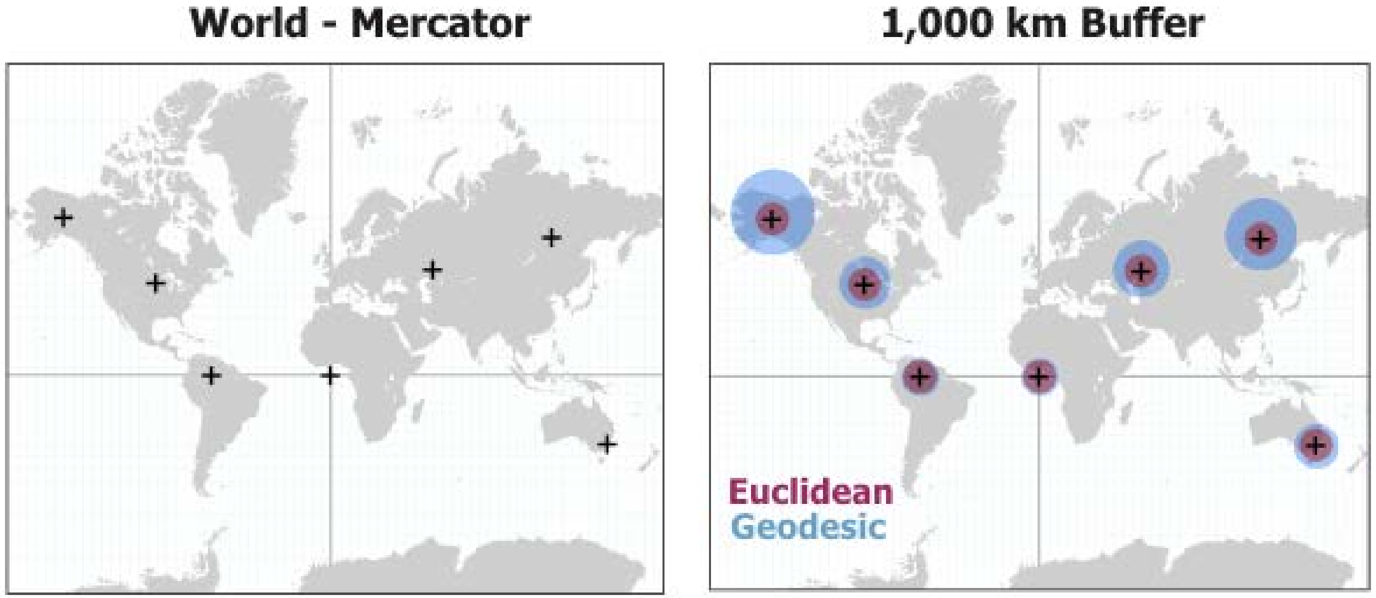
Type	Maximum range	Payload	Status
Nodong	1,300 km (810 miles)	700 kg (1,550 pounds)	Currently deployed
Taepodong-1	Up to 10,000 km	Several hundred kg	Test failed 1998, not yet operational
Taepodong-2	10,000-15,000 km	Several hundred kg	Not yet tested

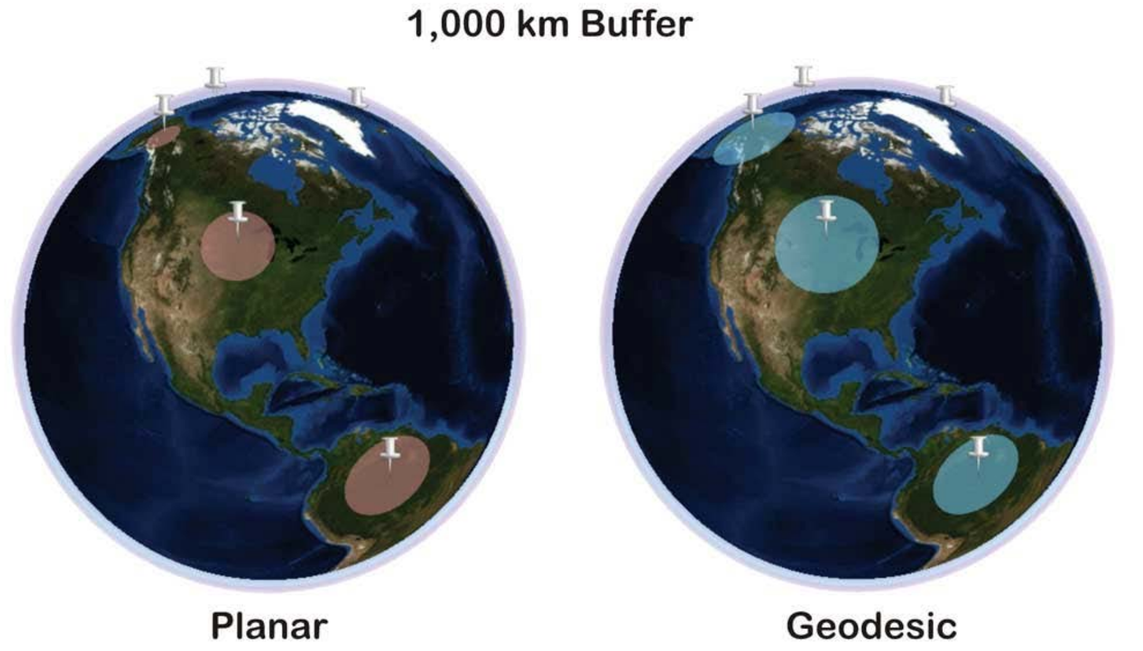
Source: Task Force for US Korea Policy, Centre for International Policy

The Economist, May 17, 2003

Planar vs. Geodesic Distances

- Planar distances are measured in a projected coordinate system (CRS) -> thus they are distorted on the (spherical) earth surface
- Geodesic distances are measured “on” a Spheroid model of the Earth



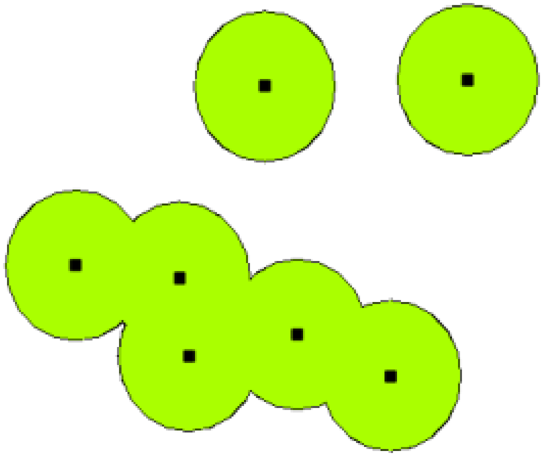


Proximity I: Buffering

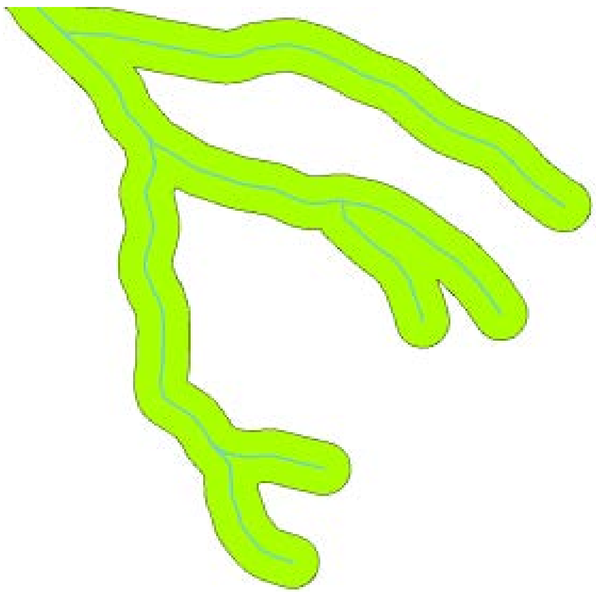
- **Buffering** usually creates two areas: one area that is **within** a specified distance to selected real world features and the other area that is **beyond**. The area that is within the specified distance is called the **buffer zone**.
- A **buffer zone** is any area that serves the purpose of keeping real world features distant from one another.
- Common types of buffer zones may be greenbelts between residential and commercial areas, border zones between countries

The border between the United States of America and Mexico is separated by a buffer zone. (Photo taken by SGT Jim Greenhill 2006).

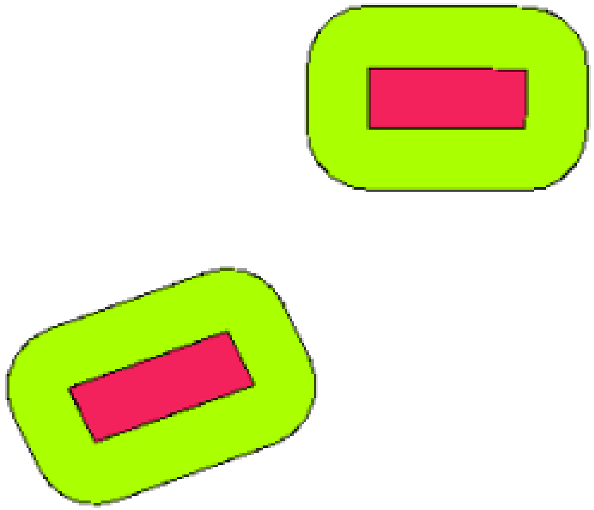
Proximity I: Buffering



A buffer zone around vector points.



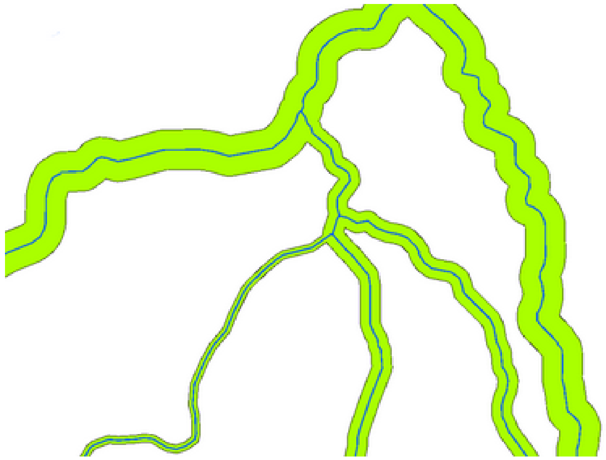
A buffer zone around vector polylines.



A buffer zone around vector polygons.

Proximity I: Buffering

- The **buffer distance** or buffer size **can vary** according to numerical values provided in the vector layer attribute table
- The numerical values have to be defined in map units according to the Coordinate Reference System (CRS) used with the data.
- For example, the width of a buffer zone along the banks of a river can vary depending on the intensity of the adjacent land use.

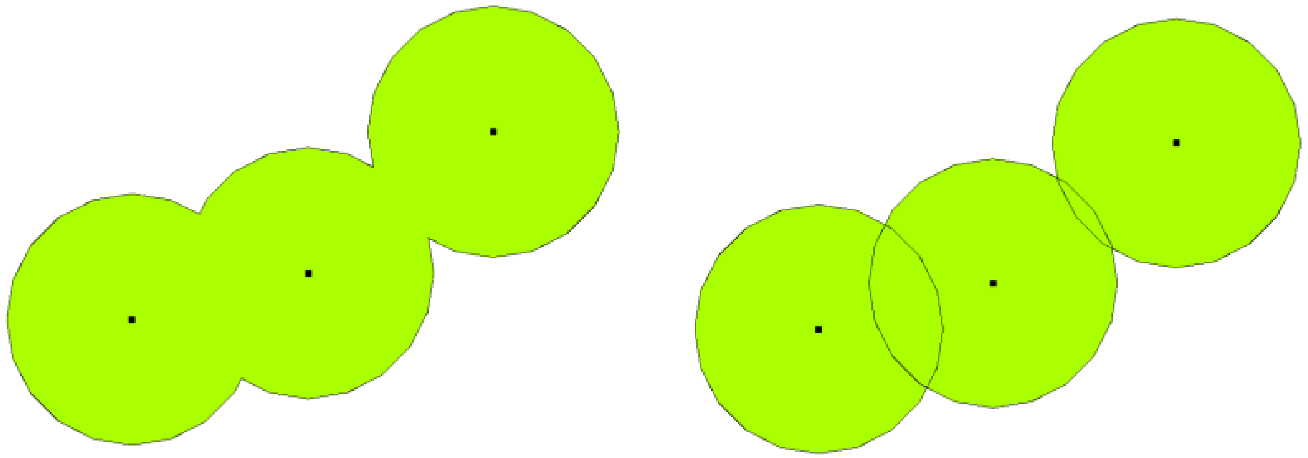


Buffering rivers with different buffer distances.

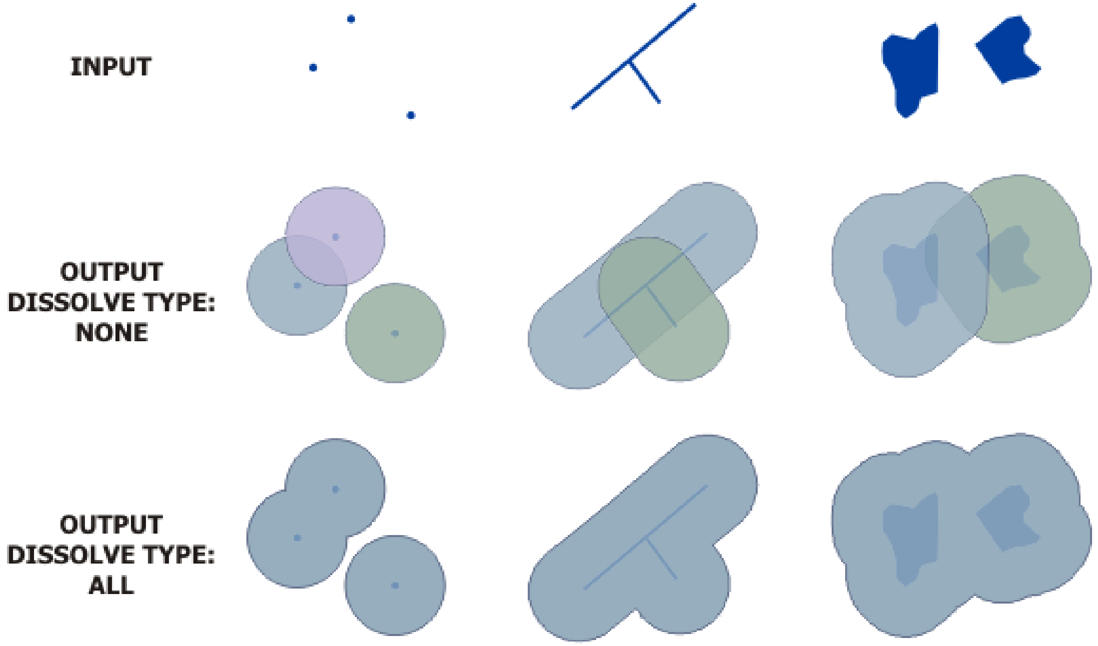
River	Adjacent land use	Buffer distance (meters)
Breede River	Intensive vegetable cultivation	100
Komati	Intensive cotton cultivation	150
Oranje	Organic farming	50
Telle river	Organic farming	50

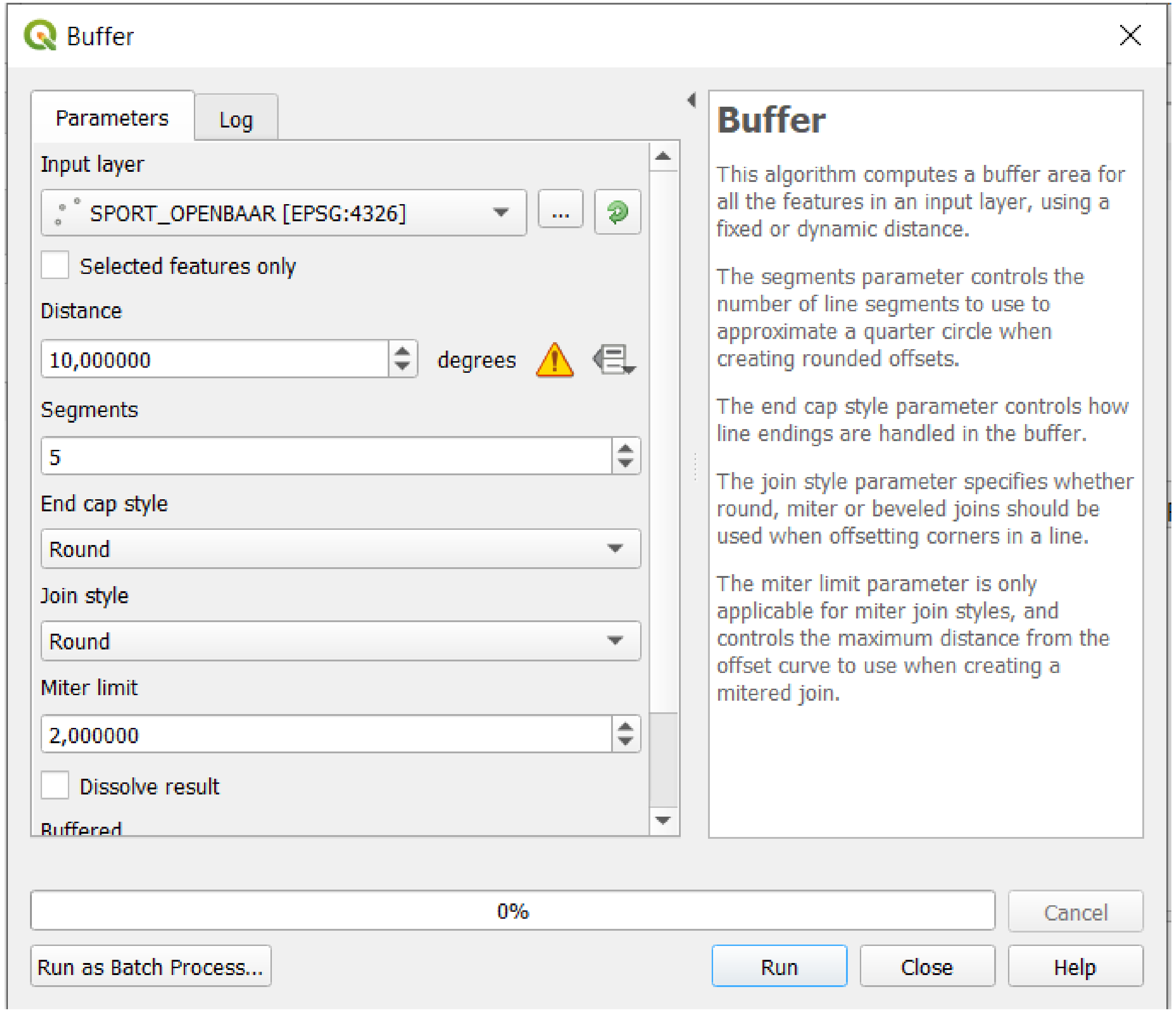
Proximity I: Buffering

- Buffers can be **merged/dissolved** into a single geometric object to avoid overlapping areas



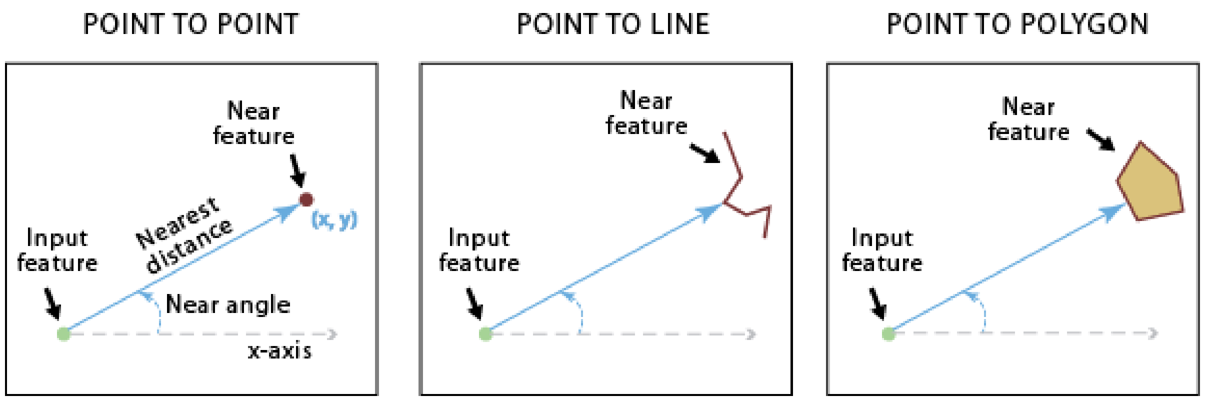
Buffer

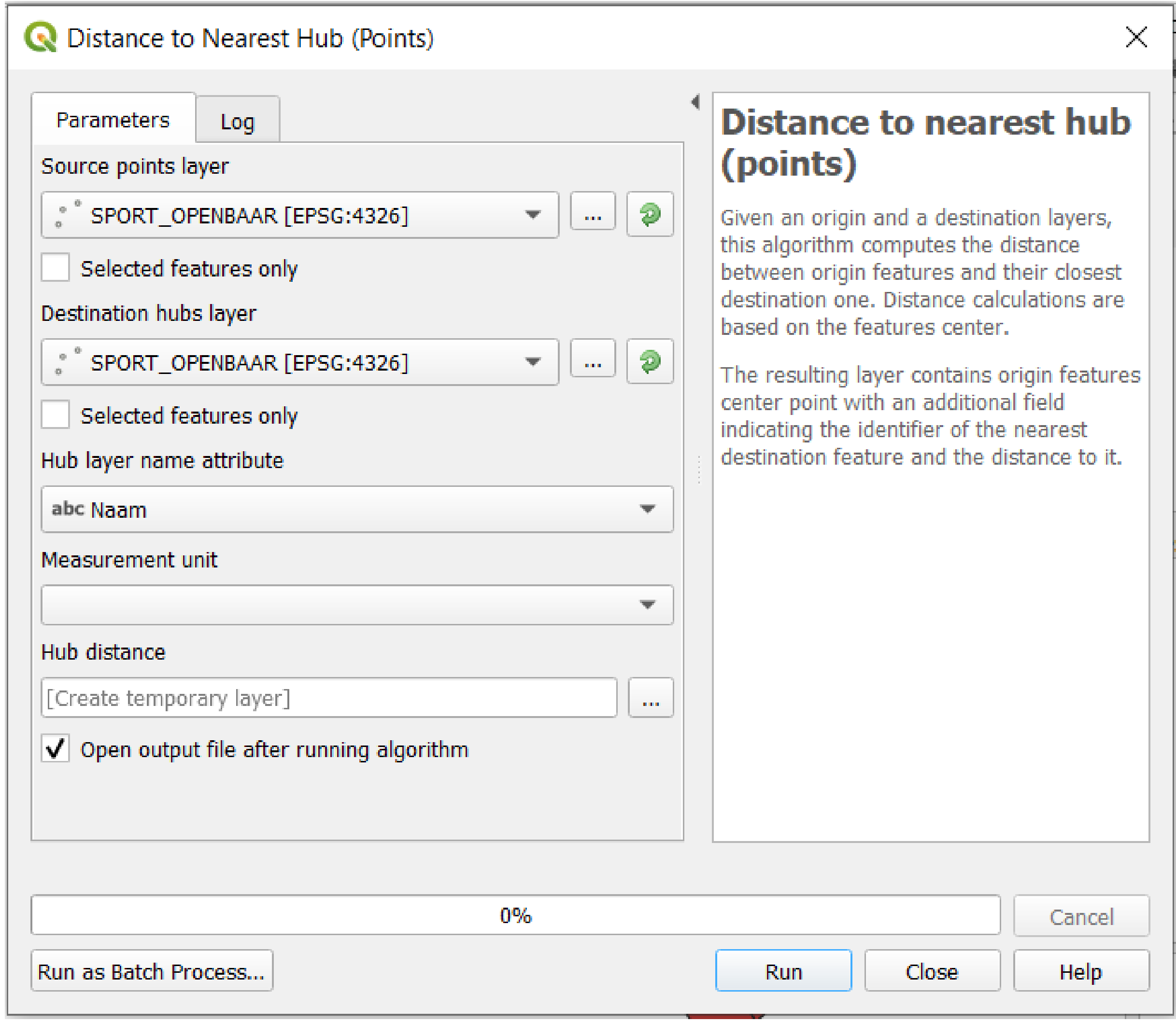

INPUT
OUTPUT
DISSOLVE TYPE:
NONE
OUTPUT
DISSOLVE TYPE:
ALL


Buffer
Parameters
Log
Input layer
SPORT_OPENBAAR [EPSG:4326]
...
Selected features only
Distance
10,000000
degrees
Segments
5
End cap style
Round
Join style
Round
Miter limit
2,000000
Dissolve result
Buffered
Buffer
This algorithm computes a buffer area for all the features in an input layer, using a fixed or dynamic distance.
The segments parameter controls the number of line segments to use to approximate a quarter circle when creating rounded offsets.
The end cap style parameter controls how line endings are handled in the buffer.
The join style parameter specifies whether round, miter or beveled joins should be used when offsetting corners in a line.
The miter limit parameter is only applicable for miter join styles, and controls the maximum distance from the offset curve to use when creating a mitered join.
0%
Cancel
Run as Batch Process...
Run
Close
Help

Proximity II: Distance to nearest

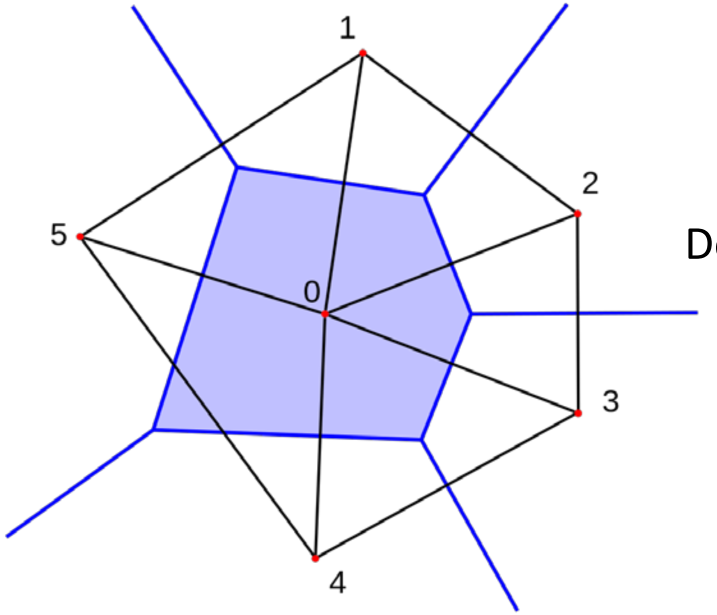
- Calculates distance and additional proximity information between the source features and the **closest** feature in another layer.



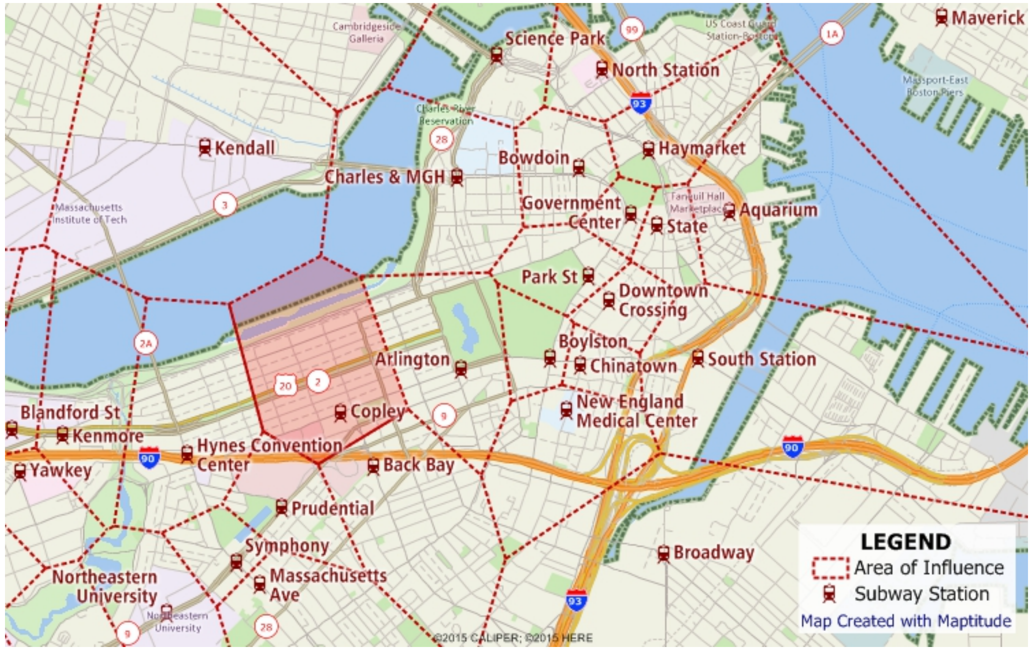


Proximity III: Thiessen Polygons

- Thiessen Polygons (=Voronoi Diagrams/Dirichlet tessellations) exactly bisect distances between points using a Delaunay Triangulation
- Thus all points within a polygon are closest to its centre point
- Can be used to construct catchment areas or to interpolate measures

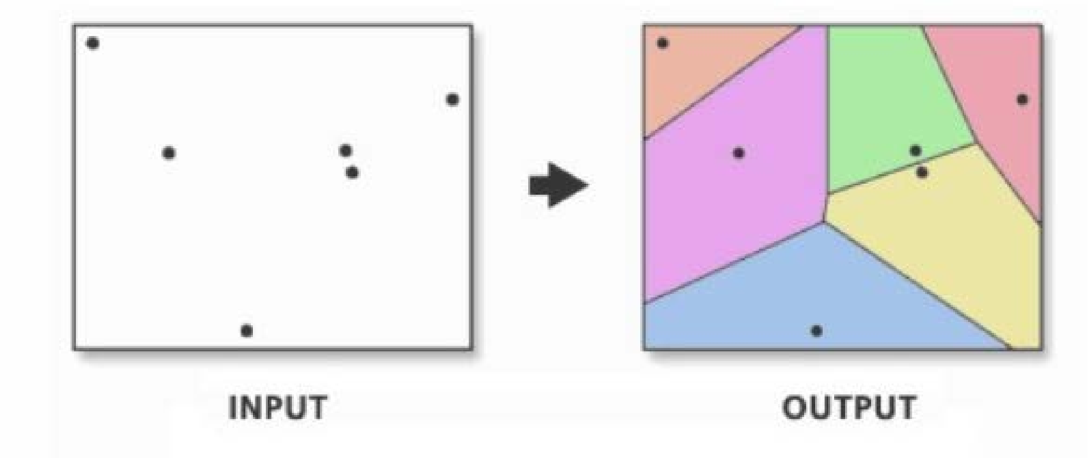


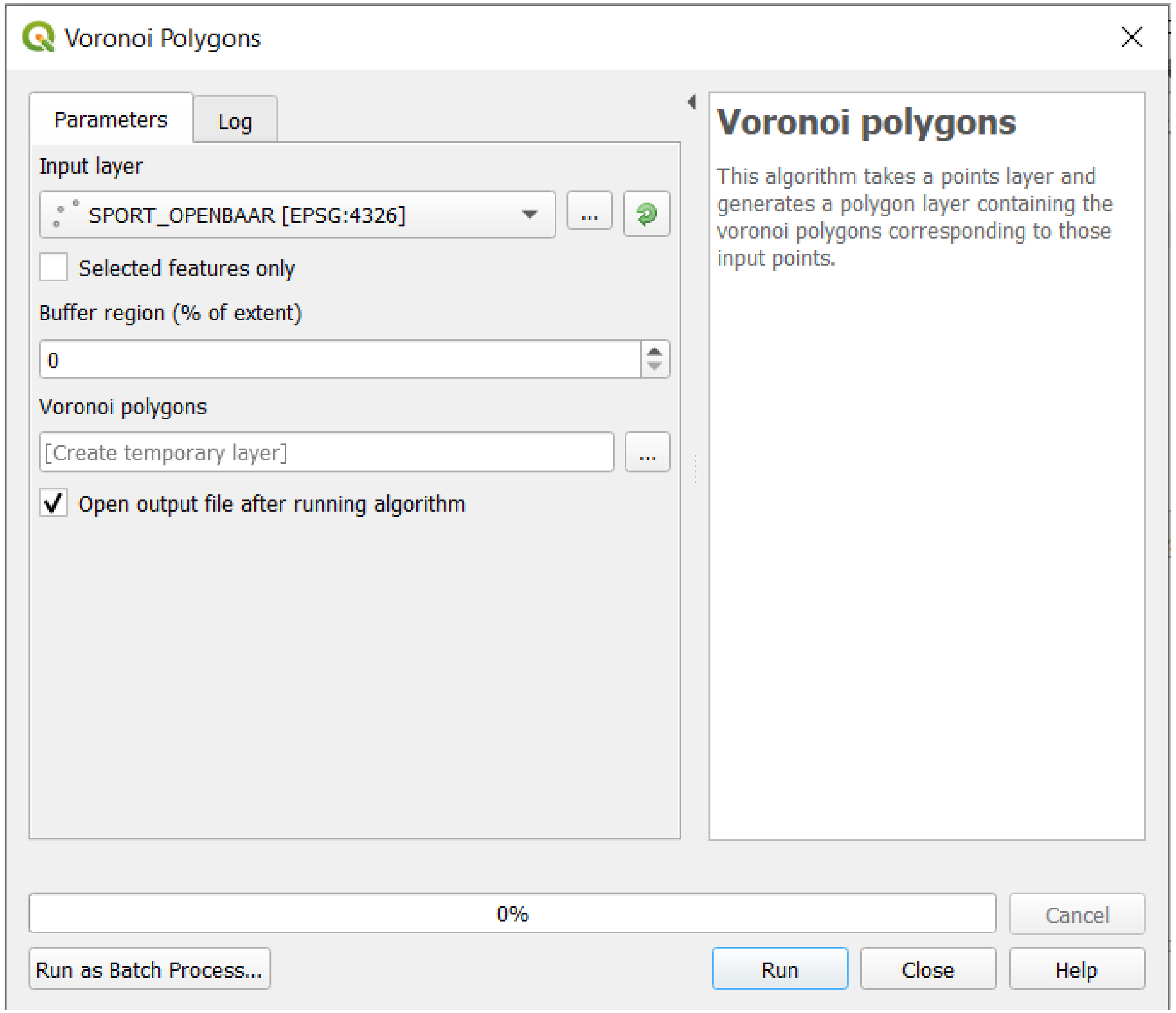
Delaunay-Triangulation



Voronoi polygons

- Creates polygons from point layer

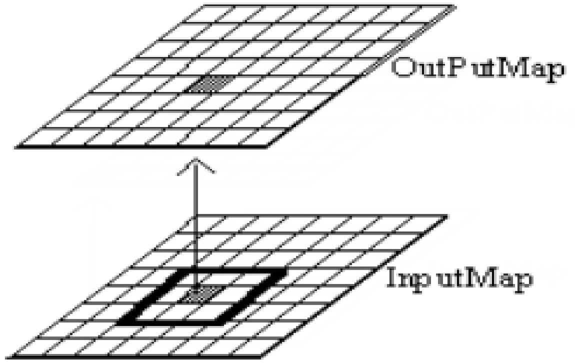




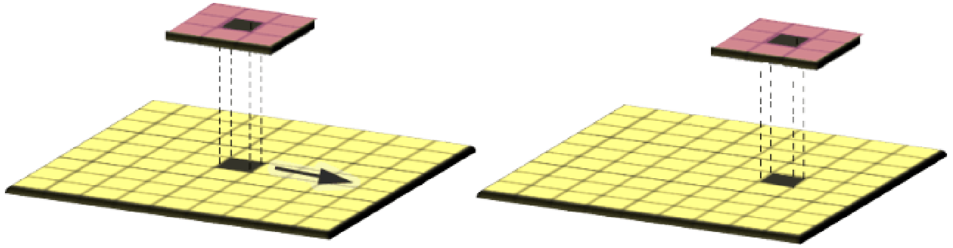
Raster Distance Analysis

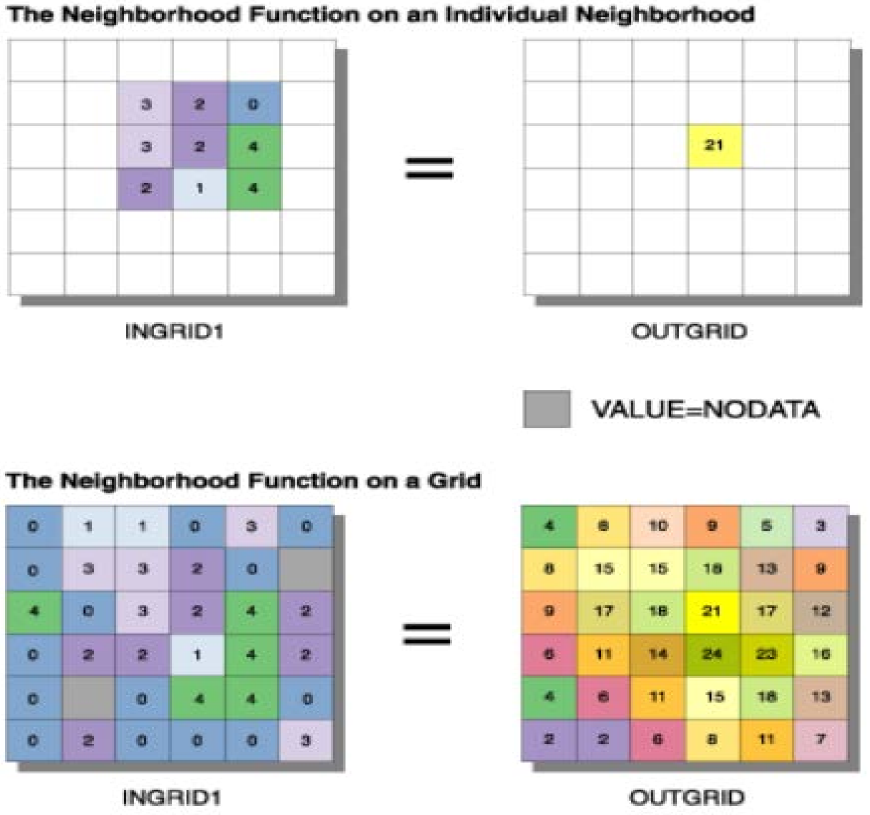
Map Algebra functions: Focal

- Focal functions: on cells within a cell neighbourhood
- Quiz: What arithmetic operator is used in this focal function example?
- focal sum



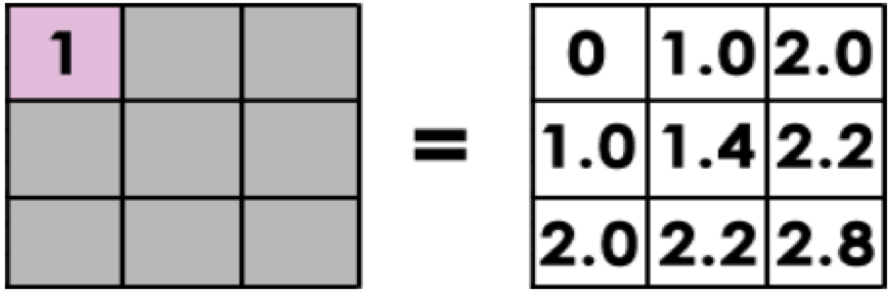
a) Immediate Neighborhoods

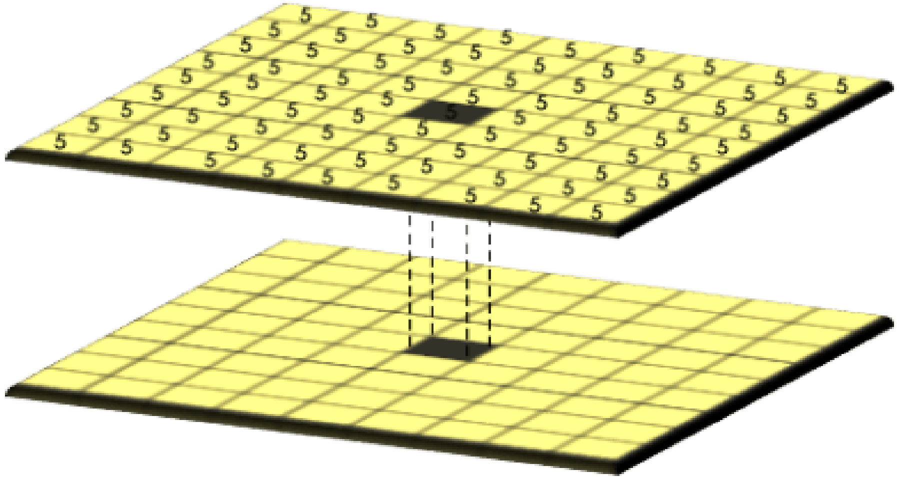




Map Algebra functions: Global

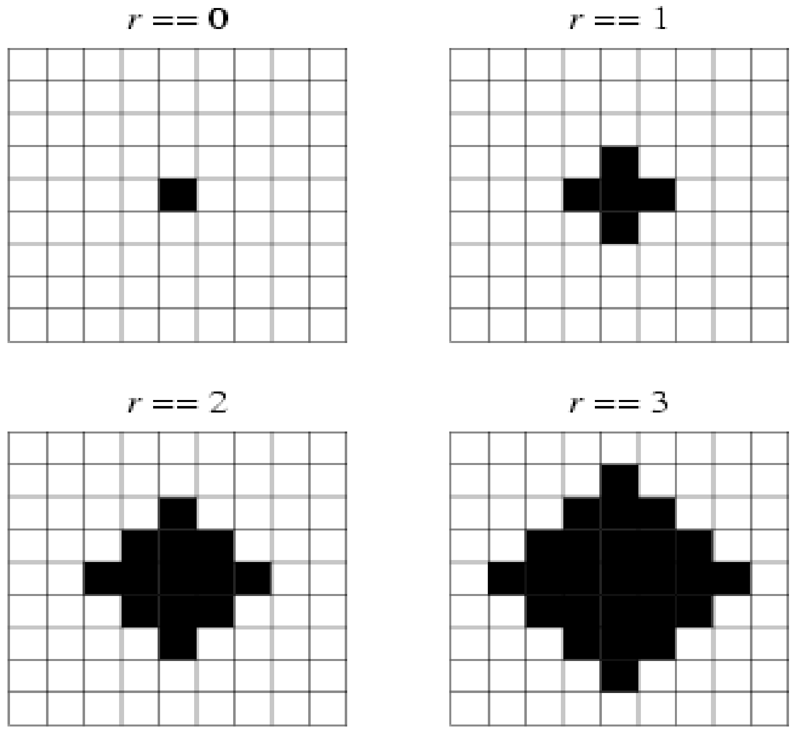
- Global functions: on all cells
- Quiz: What operator is used in this global function example?
- Euclidean distance to a source cell



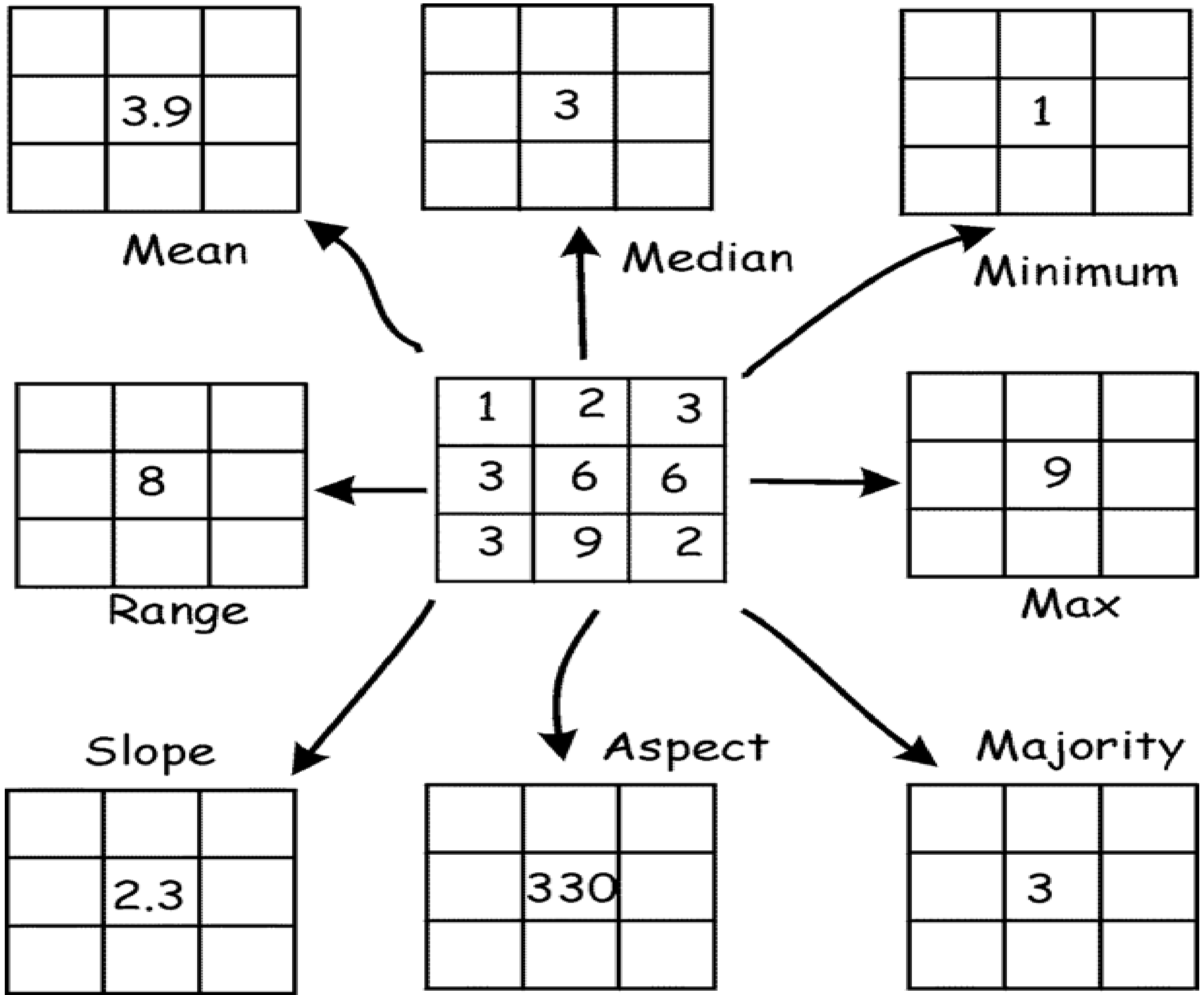


Focal functions:
Von Neumann Neighborhood

- **Diamond-shaped**
- To define a set of cells surrounding a given cell
- Ranges r = 0,1,2,3
- N = 2*r(r+1)+1
- centered square number”

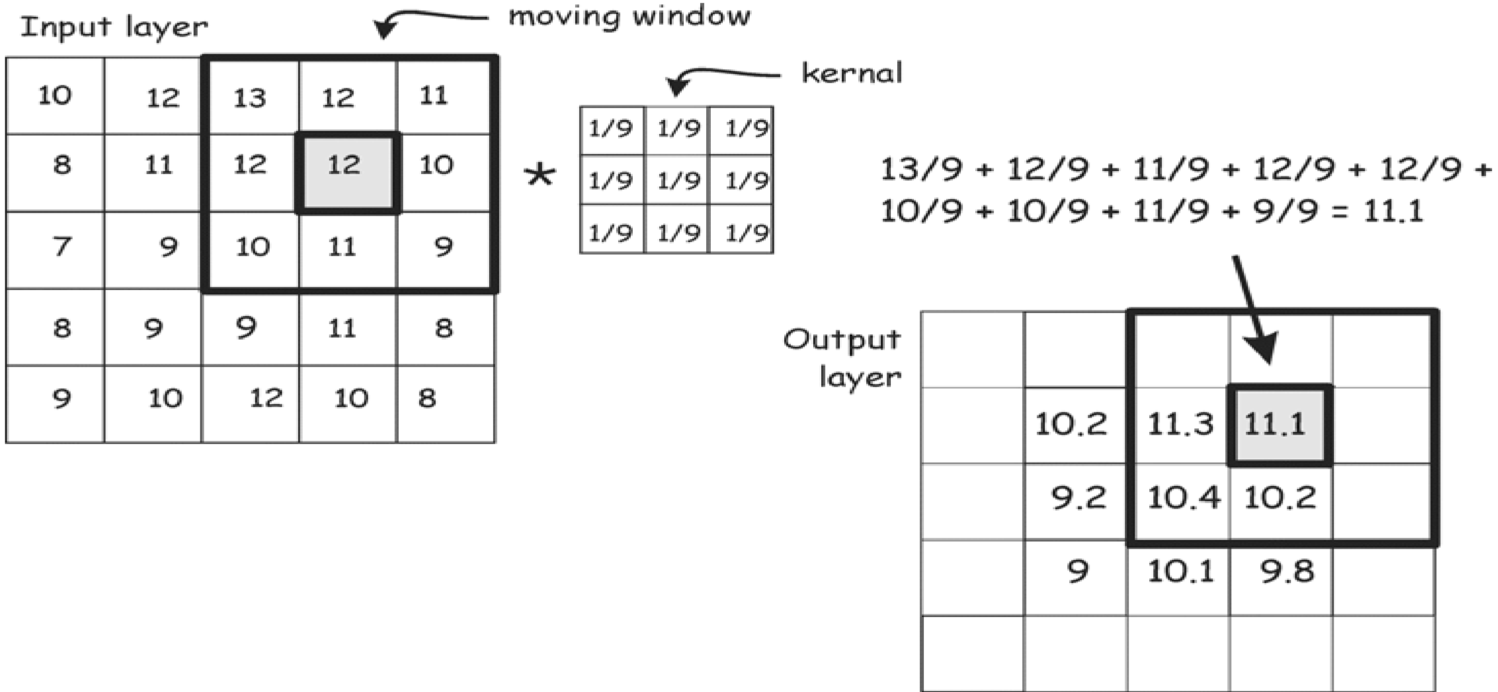


Examples of Moving Window (focal) functions



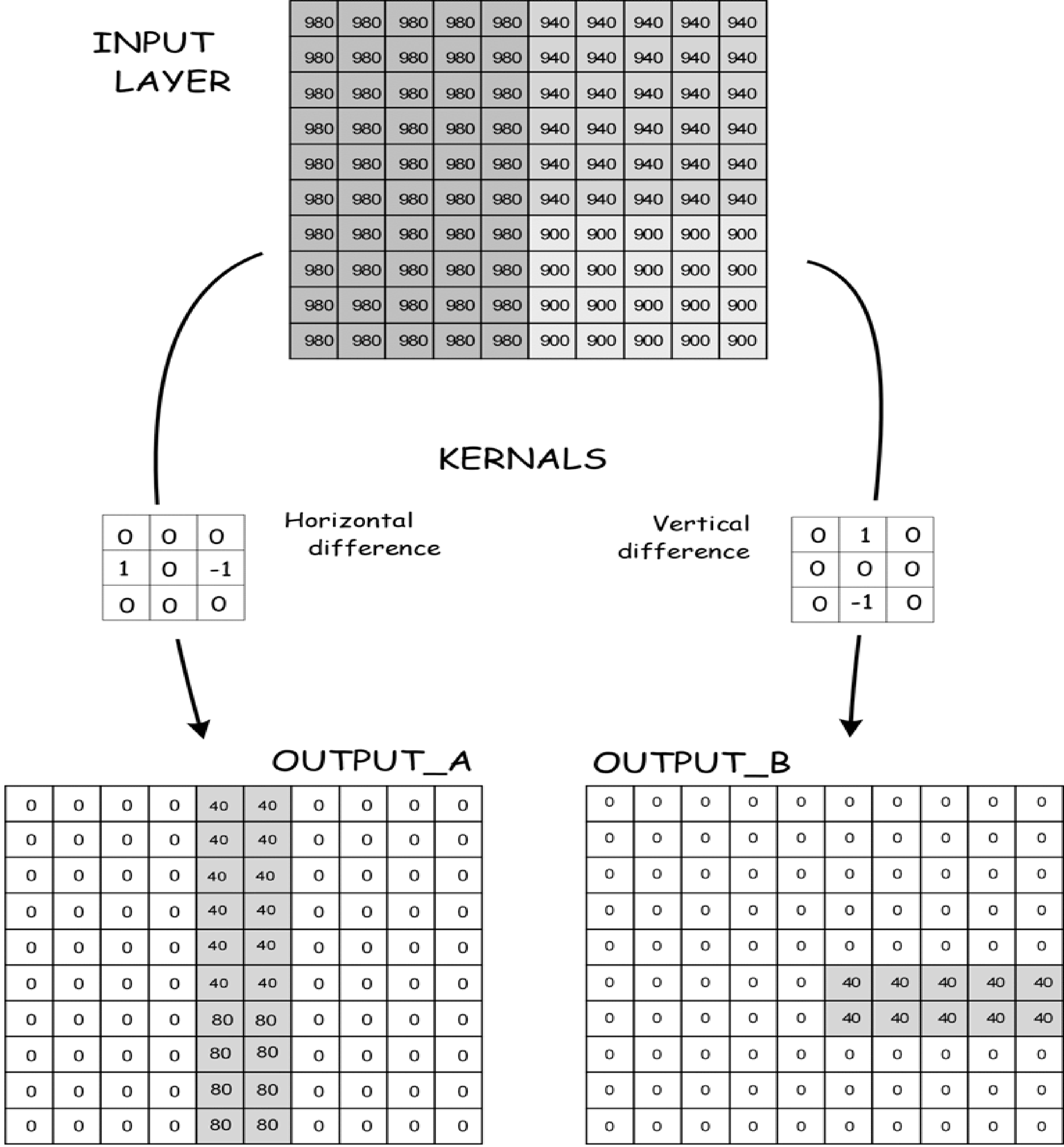
Moving Windows and Kernels

- Kernel : Set of **constants** for multiplying values within a given window
- What can you see at the **margins** of the output grid?

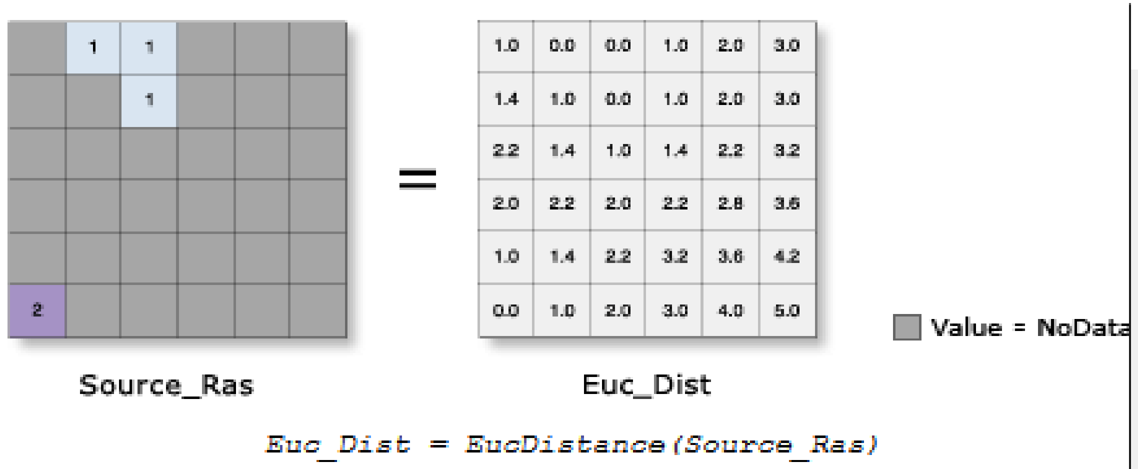


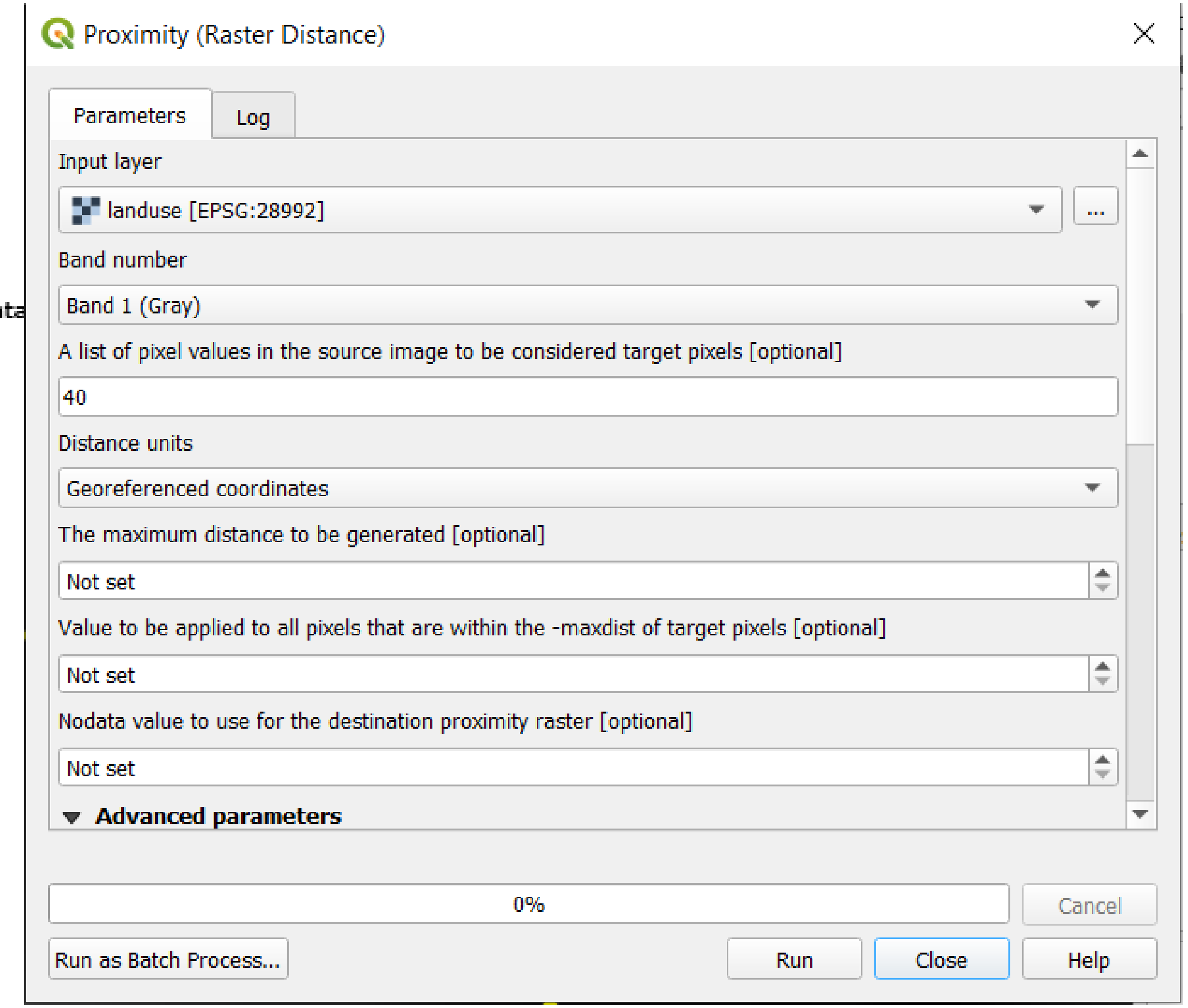
Edge Detection using Kernels

- Concentration changes; elevation changes,…
- Discovering **contrasts / differences** within the focal neighborhood



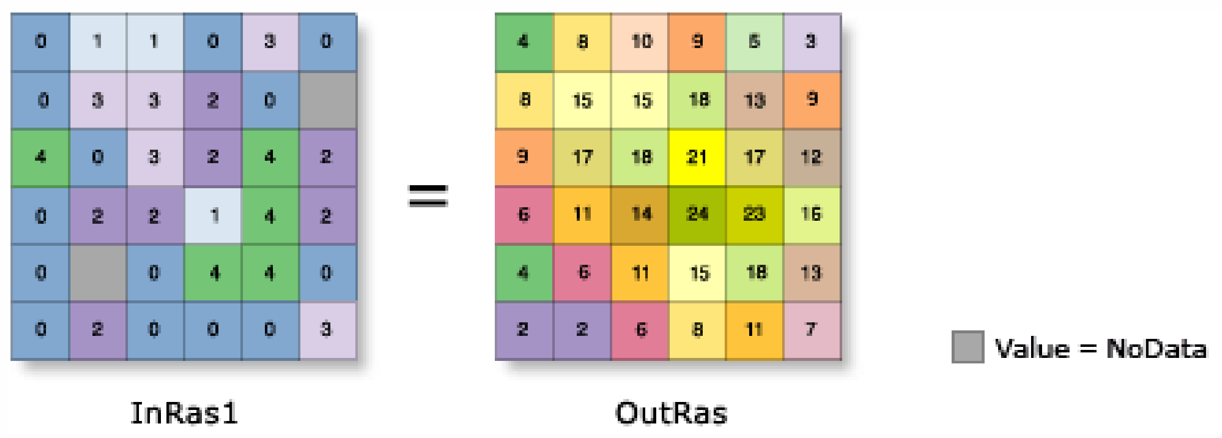
Proximity raster



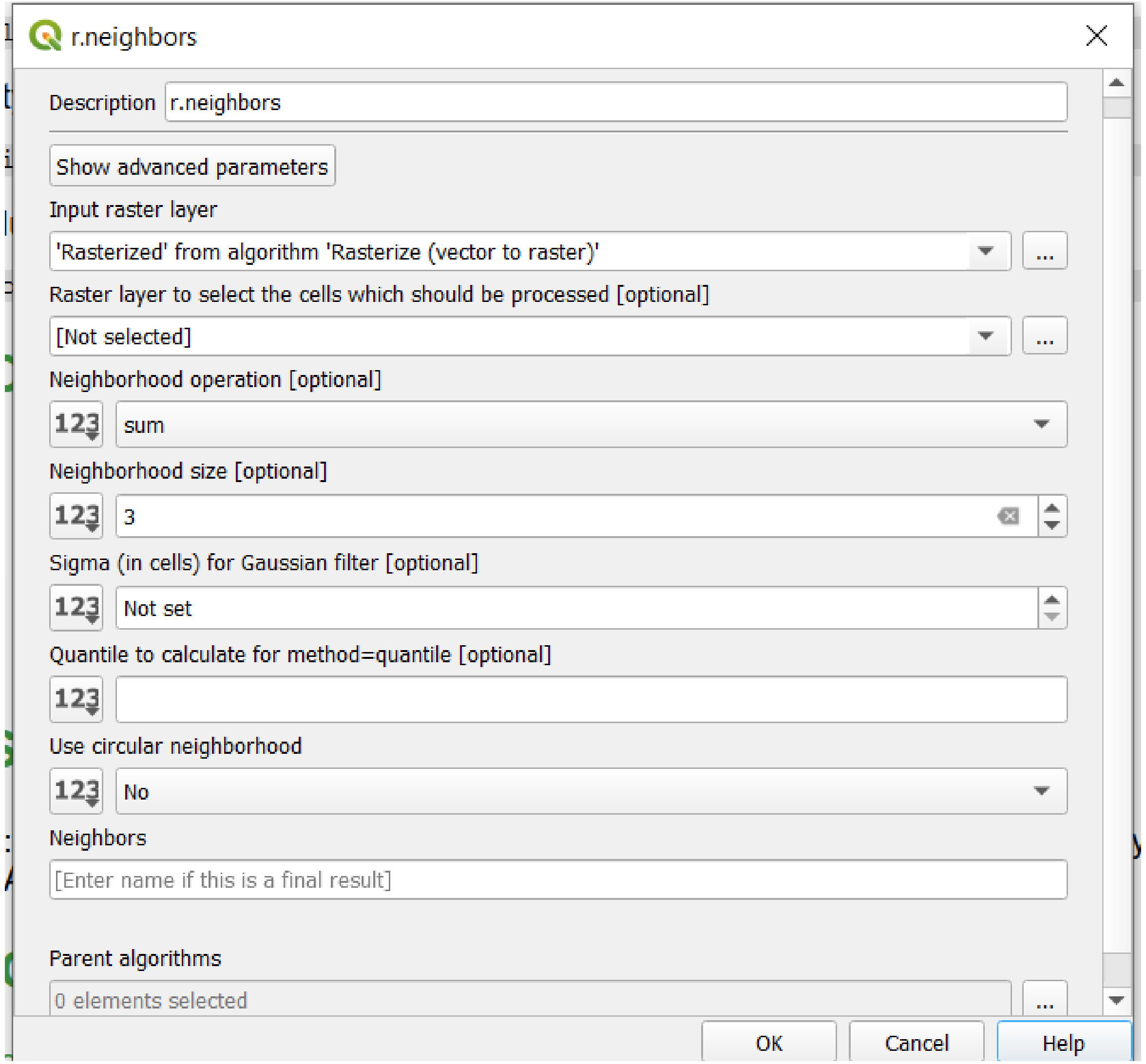


Focal statistics

Calculates for each input cell a neighborhood statistics.

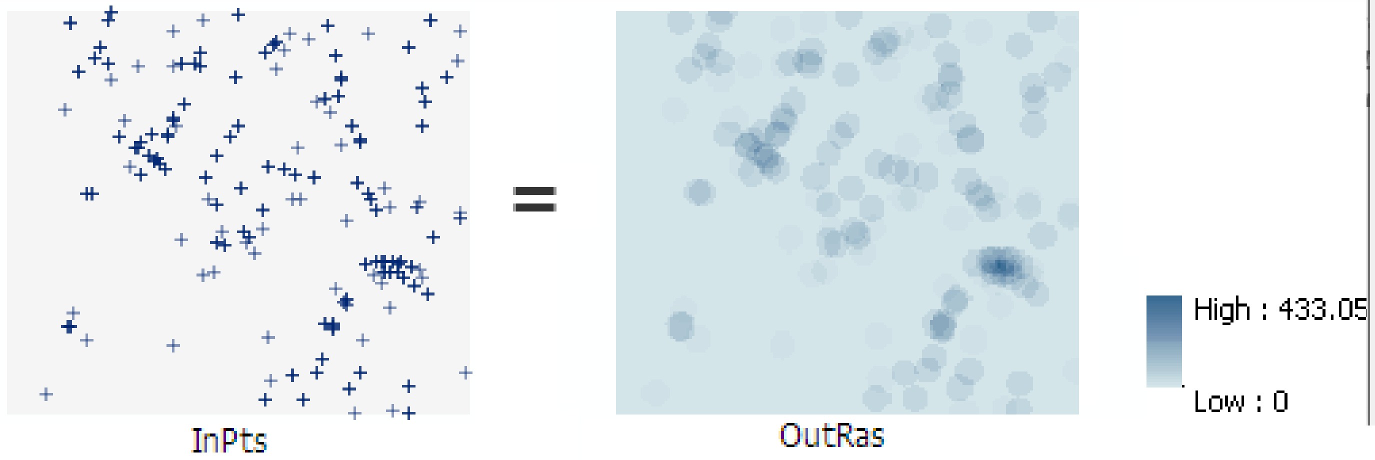



```
OutRas = FocalStatistics(InRas1, NbrRectangle(3,3,MAP), "SUM", "")
```

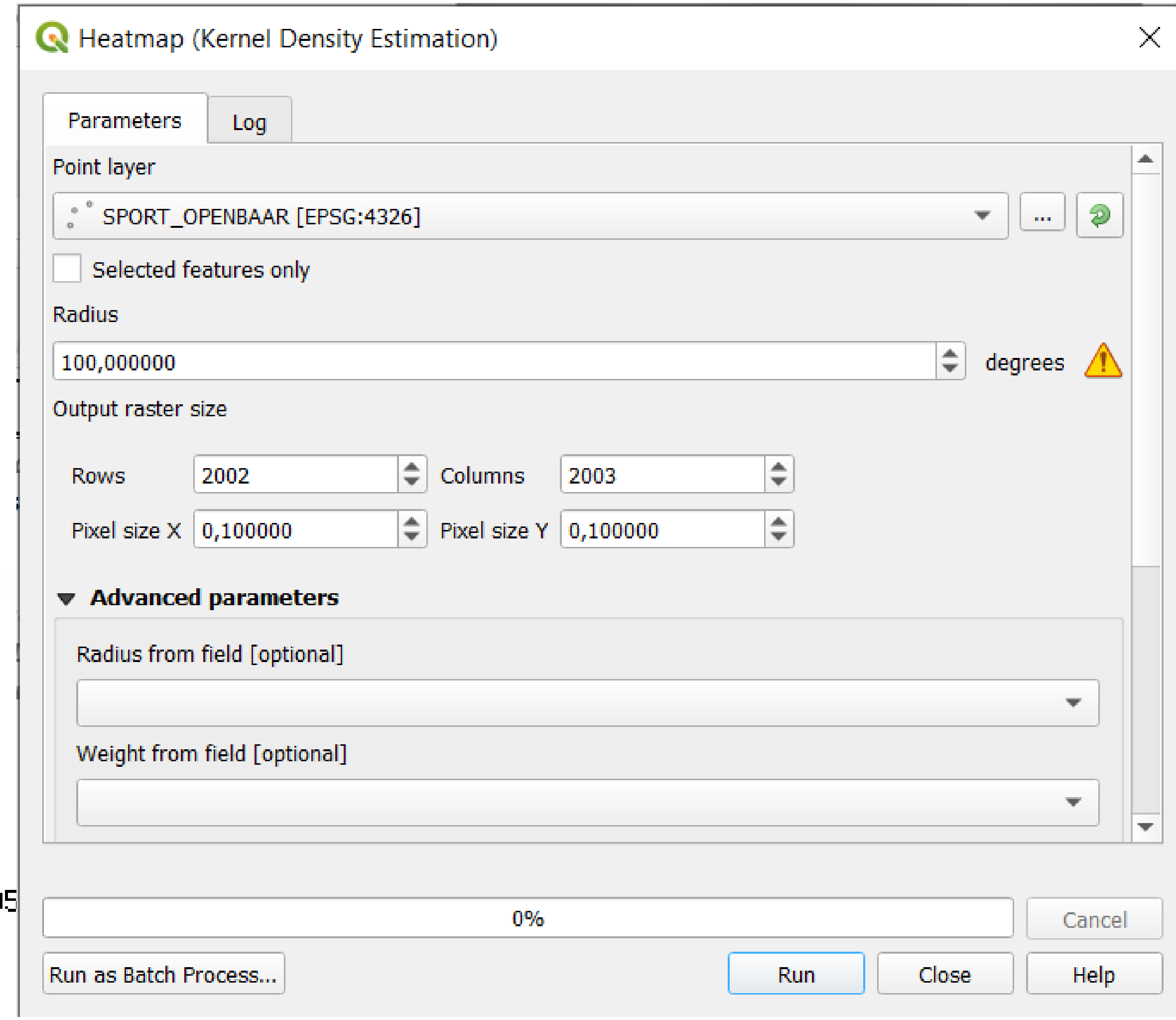


Kernel density

Calculates a magnitude-per-unit area from point features that fall within a neighborhood around each cell.

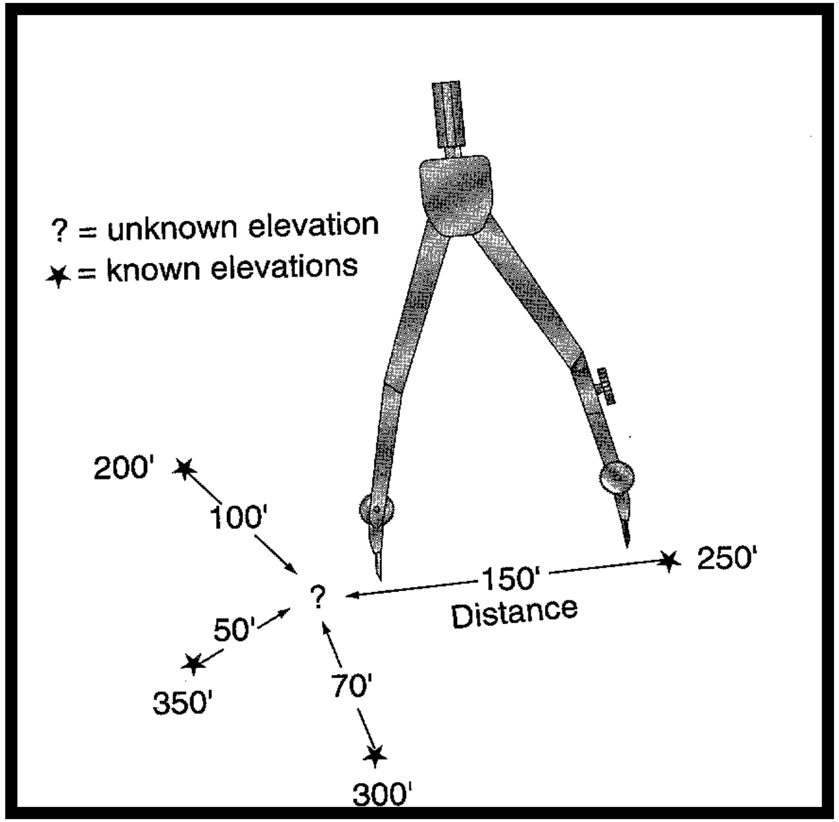


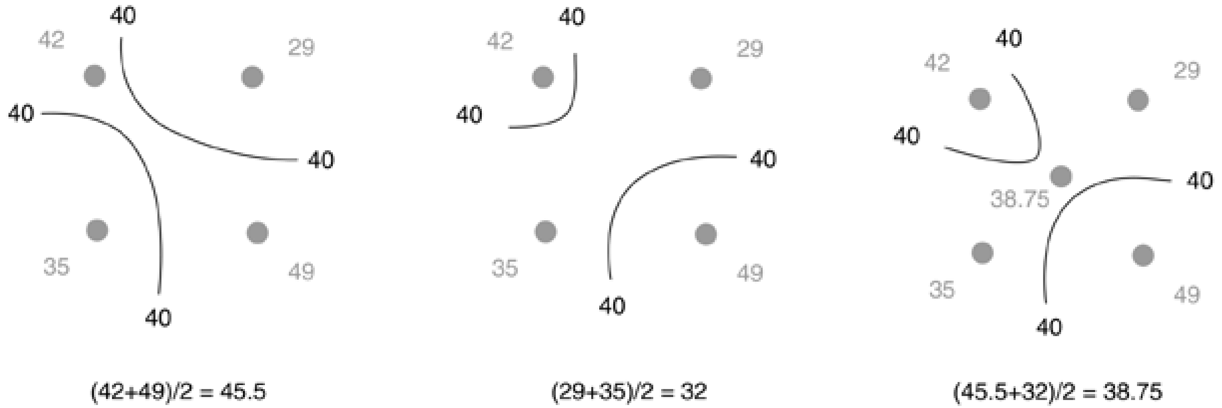

```
OutRas = PointDensity(InPts, None, 30)
```



Point interpolation

- Point sample of (field) measurements
- How to estimate the field values that were not sampled?
- How to draw isolines when linear interpolation does not work (saddle problem)?



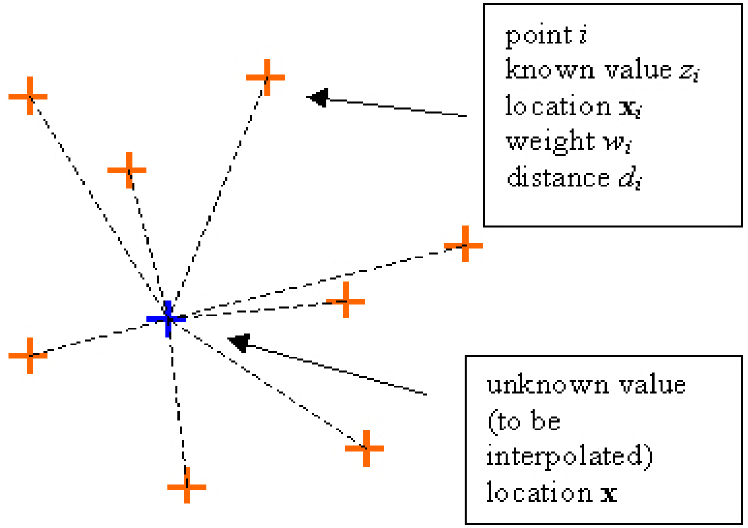


Interpolation: Inverse Distance Weighting (IDW)

Each input point has local influence that diminishes with distance

Estimates are weighted averages of values points within window R

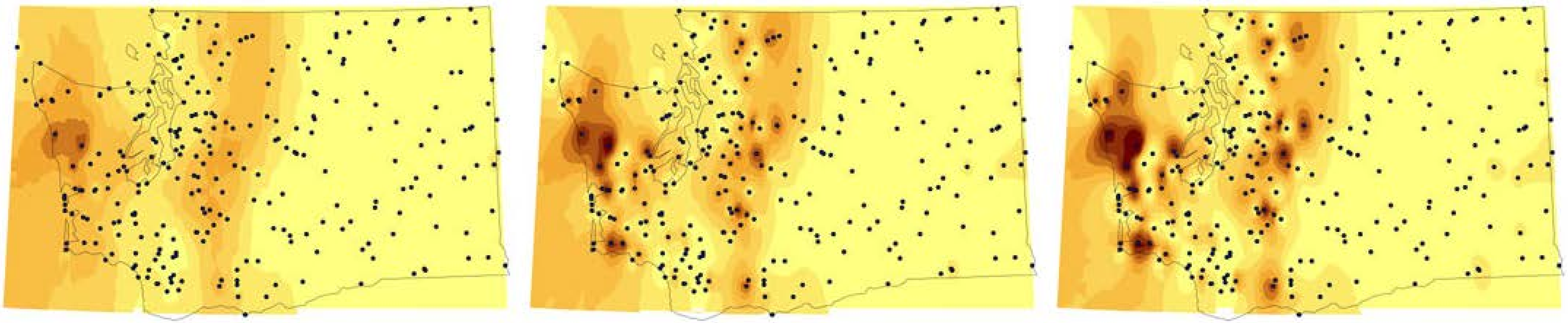
$$Z_{i,j} = \frac{\sum_{p=1}^{R} z_p d_p^{-n}}{\sum_{p=1}^{R} d_p^{-n}}$$



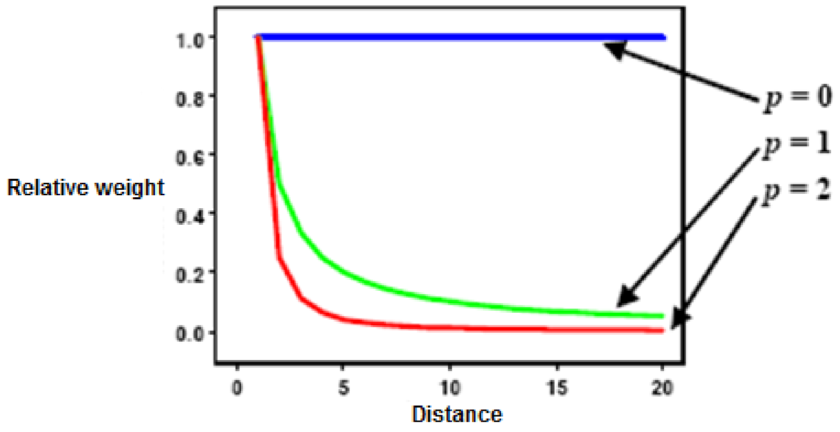
where weight is some function of distance (e.g., w = d^{-n})

IDW parameters: exponents

d^{-1} d^{-2} d^{-3}



The lower n, the more distant points have a greater effect on the overall pattern, the greater n, the more local points have a greater effect

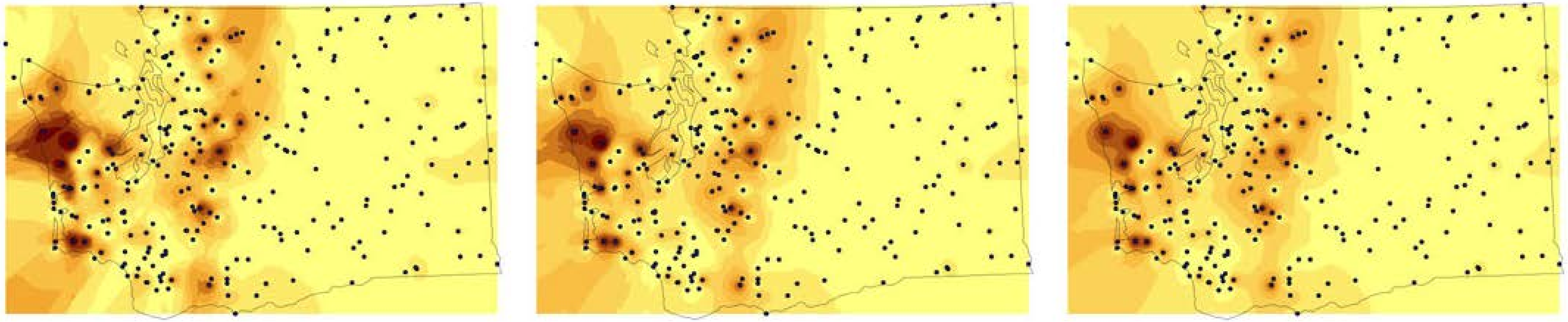


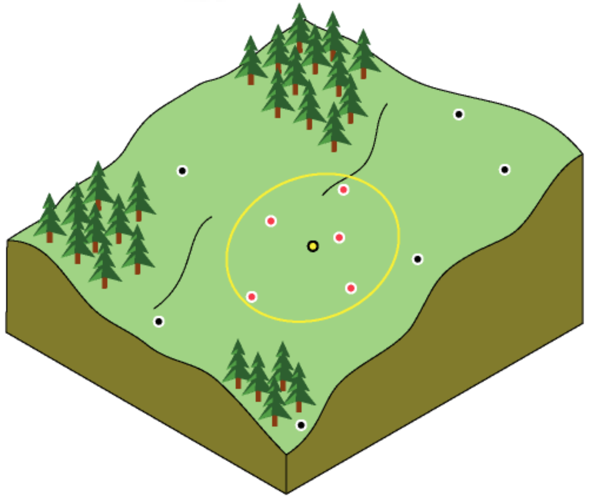
IDW parameters: search neighborhood

5NN

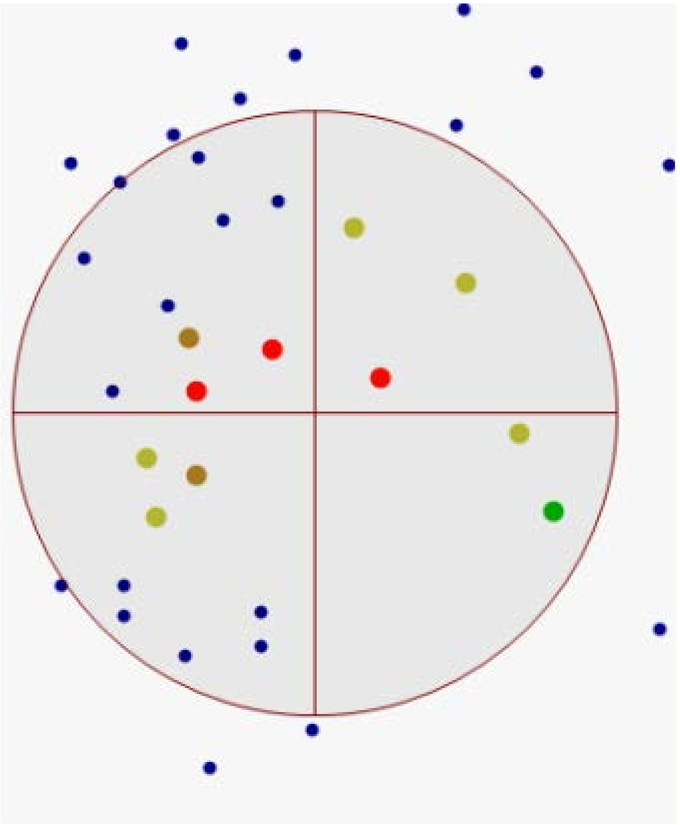
12NN

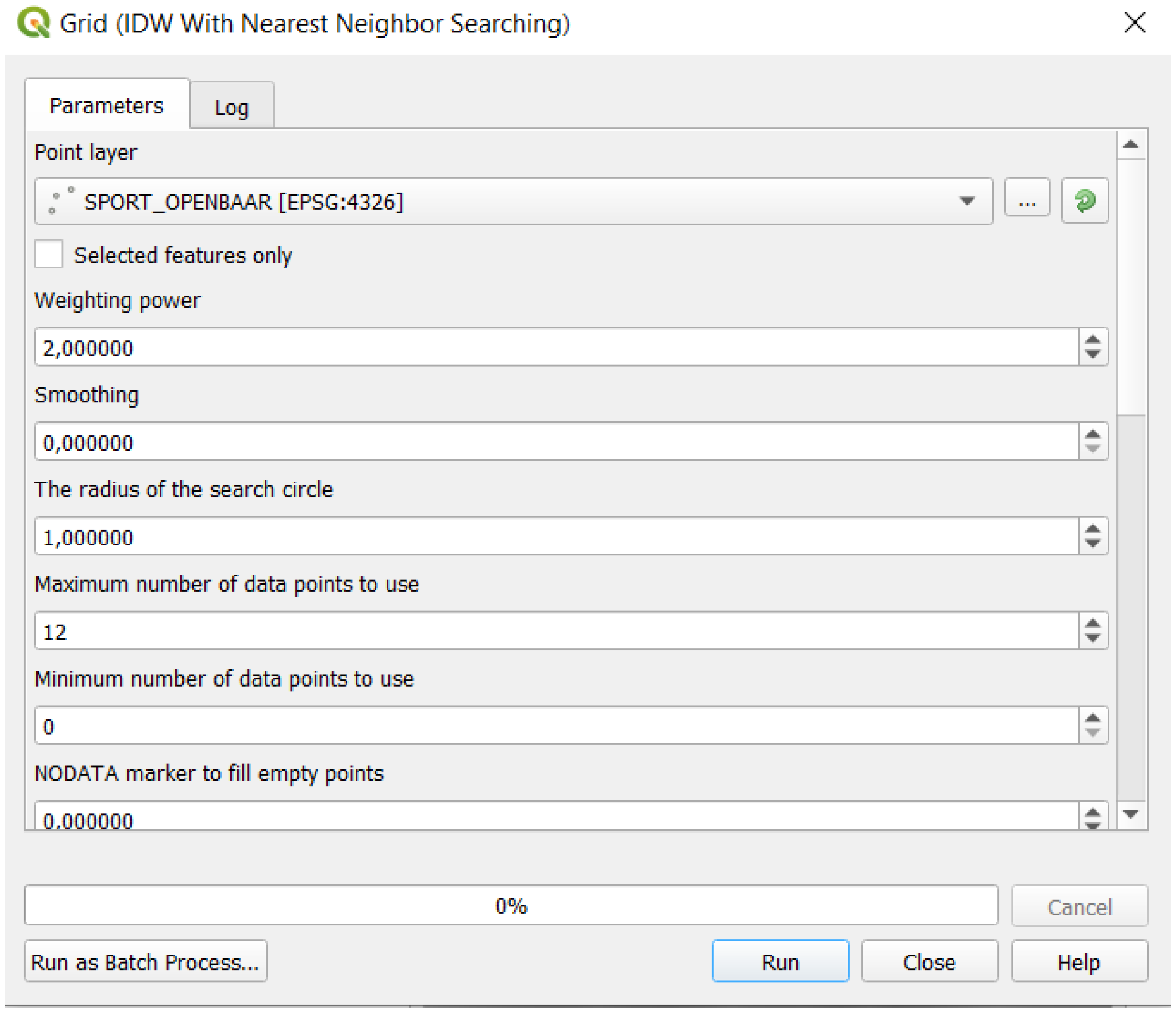
25 NN





IDW

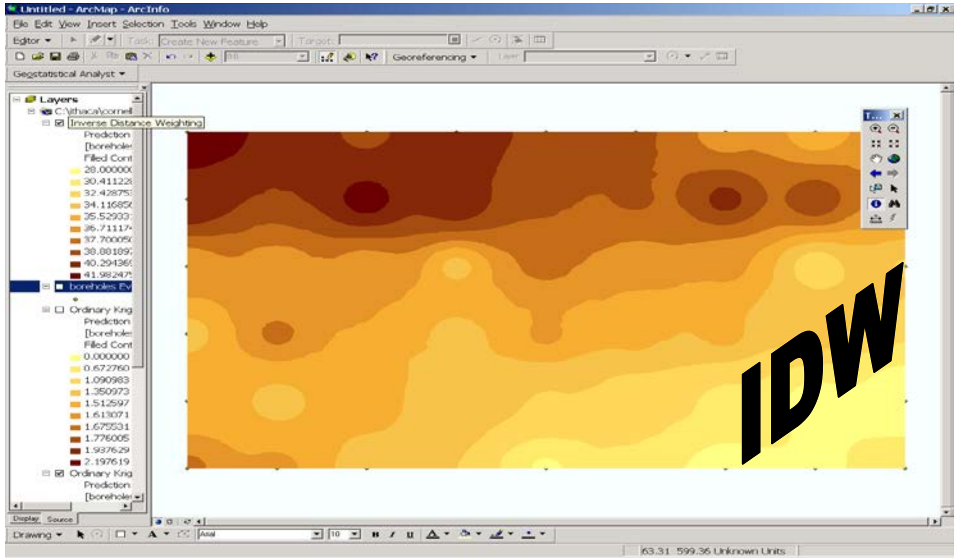


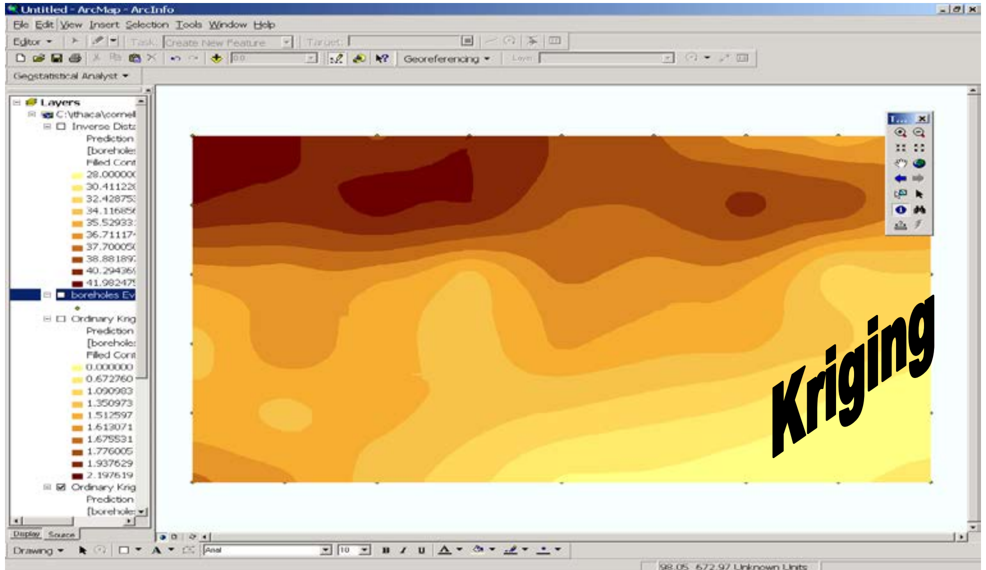

Grid (IDW With Nearest Neighbor Searching)
Parameters
Log
Point layer
SPORT_OPENBAAR [EPSG:4326]
...
Selected features only
Weighting power
2,000000
Smoothing
0,000000
The radius of the search circle
1,000000
Maximum number of data points to use
12
Minimum number of data points to use
0
NODATA marker to fill empty points
0,000000
0%
Cancel
Run as Batch Process...
Run
Close
Help

IDW vs. Kriging (Geostatistics)

- Interpolated values limited by range of the data
- How many points should be included?
- What about irregularly distributed points?
- What about the map edges?

- Kriging appears to give a more “smooth” look to the data
- Kriging avoids the “bulls eye” effect
- Kriging gives us a standard error





…back to the core concepts

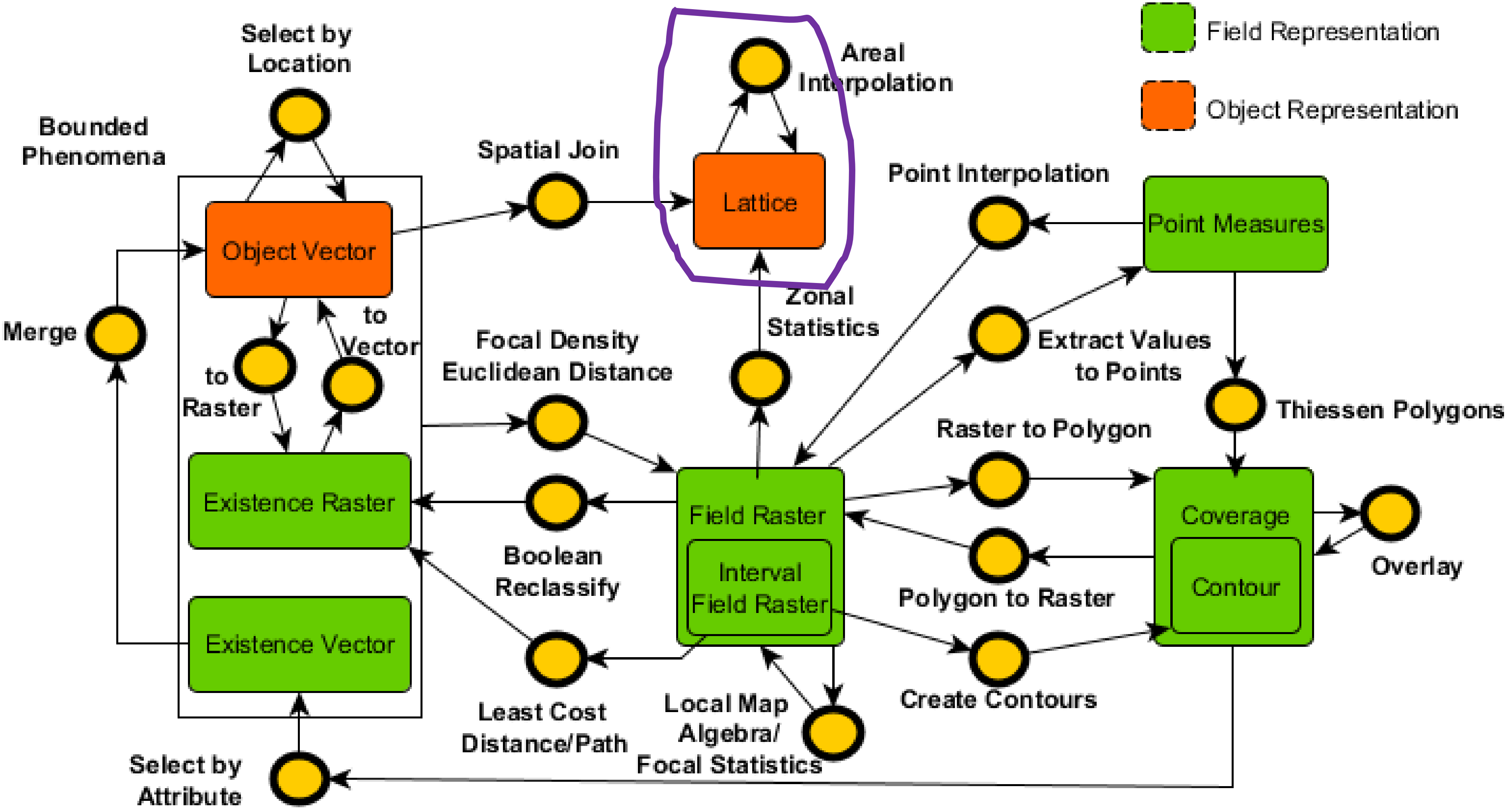
Core concept quiz 1)

Which core concepts correspond to which datasets in the practical?

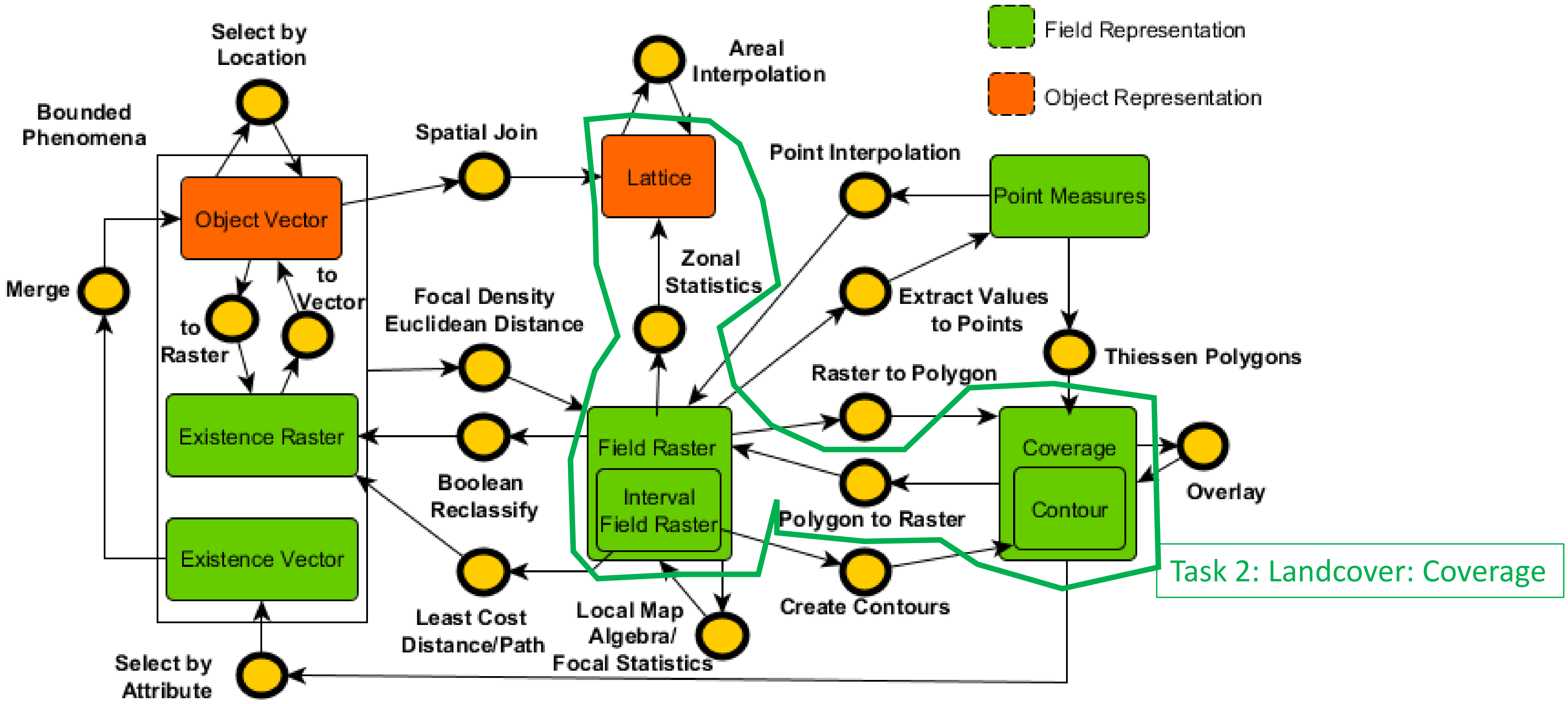
1) CBS Buurt statistics:	Lattice
2) Grondgebruik:	Coverage
3) Grondgebruik:	Coverage
4) Sport facilities:	Vector (Point) Object

Core concept quiz 2)

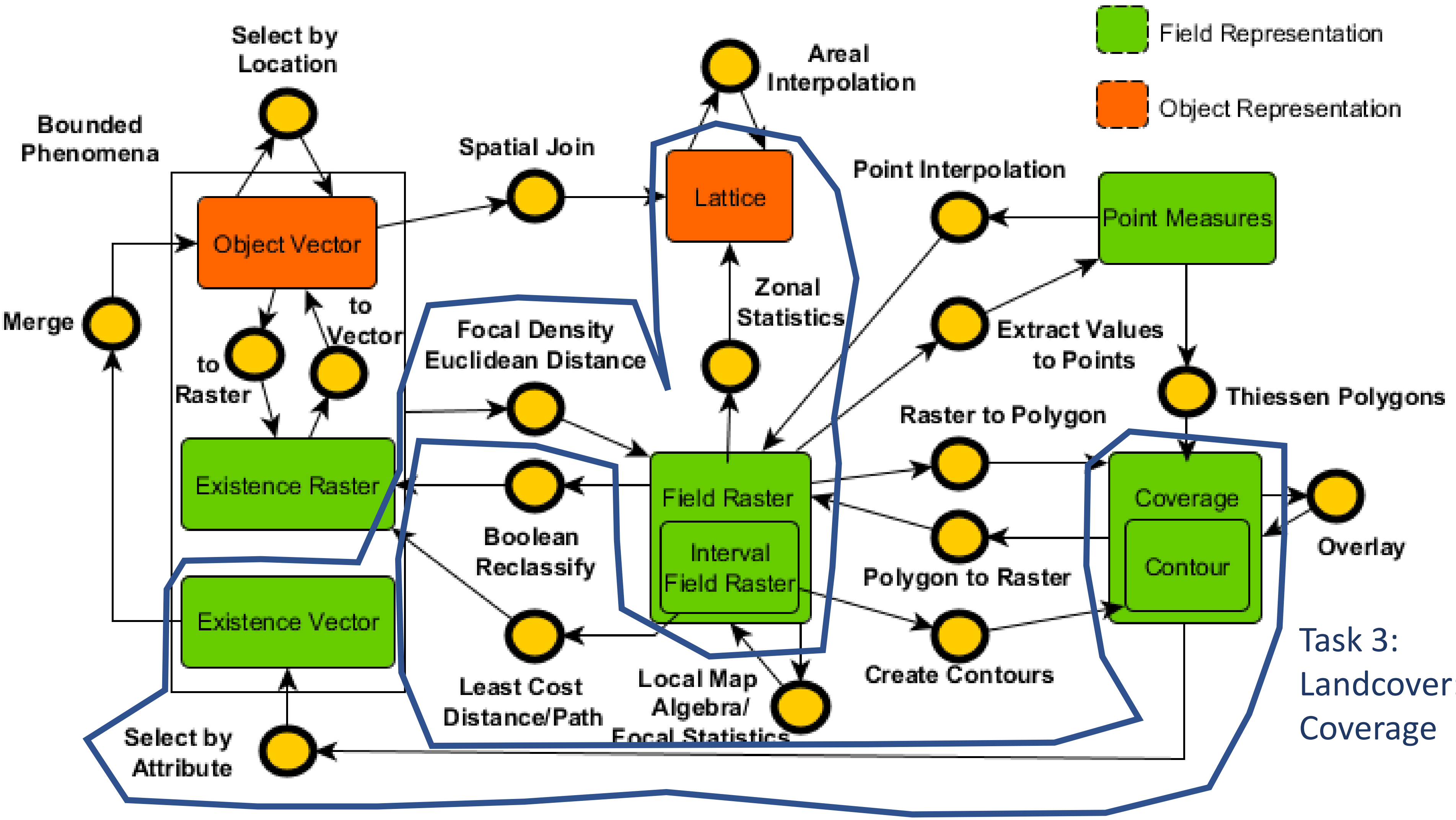
Task 1: CBS Buurt statistics: Lattice



Core concept quiz



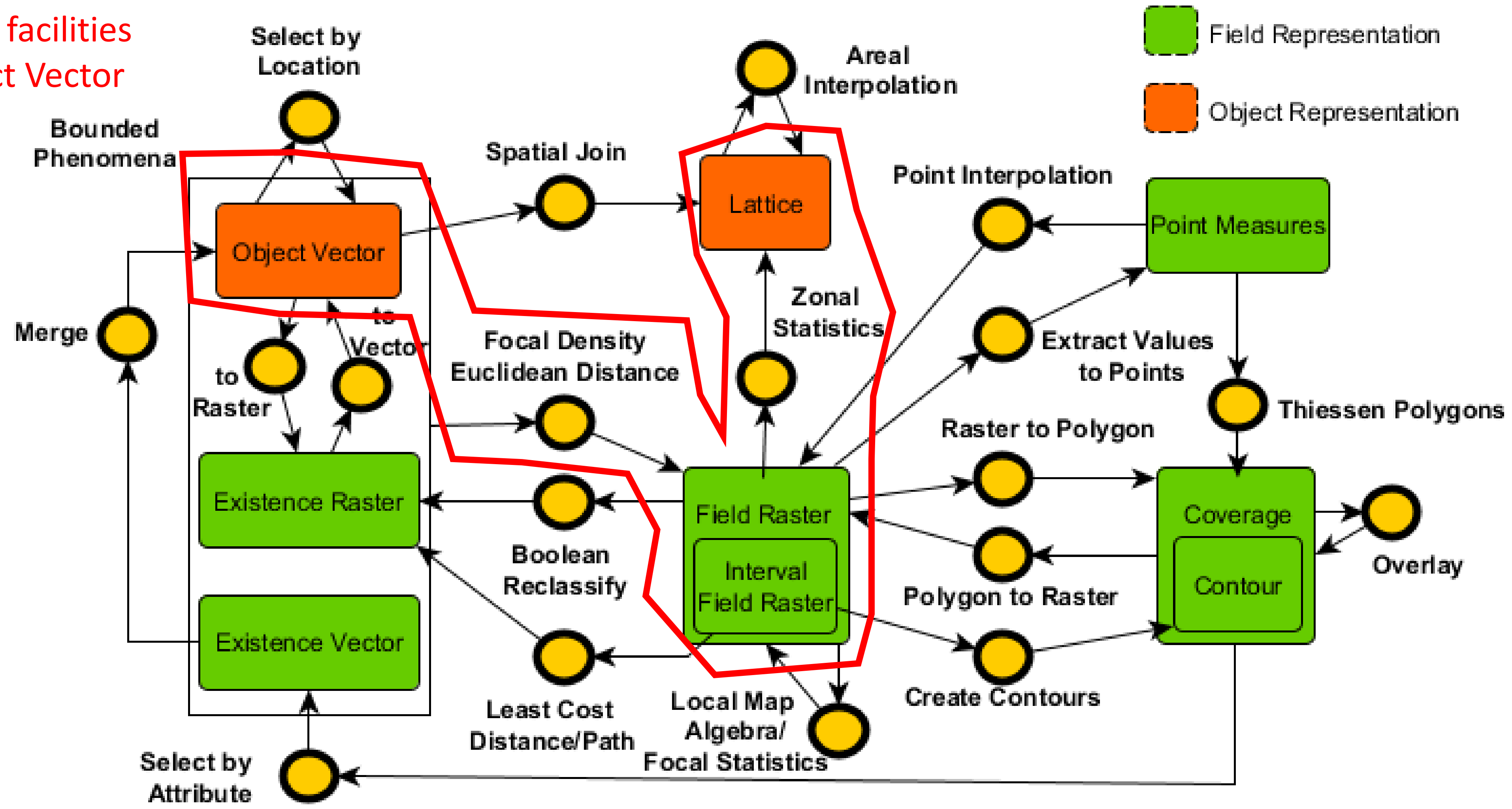
Core concept quiz



Task 3:
Landcover:
Coverage

Core concept quiz

Task 4:
Sport facilities
Object Vector



Questions?
(Q&A session)

References

- Chrisman 2002: Exploring Geographic Information systems, 2nd edition, Chapter 6 “Distance Relationships” (153-167)
- Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. International Journal of Geographical Information Science, 26(12), 2267-2276.
- Scheider, S., Meerlo, R., Kasalica, V., & Lamprecht, A. L. (2020). Ontology of core concept data types for answering geo-analytical questions. Journal of Spatial Information Science, 2020(20), 167-201 (http://www.josis.org/index.php/josis/article/viewArticle/555)