Introduction to Geospatial Raster and Vector Data with Python

netherlands eScience center

October 10 - 11, 2023

Instructing staff:

Francesco Nattino, Ou Ku, Pranav Chandramouli, Maurice de Kleijn

netherlands eScience center

Schedule

Day 1

Time	Topic
09:30	Welcome and icebreaker, setup check
09:45	Introduction to raster, vector, and CRS
10:30	**Coffee break**
10:40	Access satellite imagery using Python
11:30	**Coffee break**
11:40	Read and visualize raster data
12:30	**Lunch break**
13:30	Vector data in Python
14:30	**Coffee Break**
15:40	Align Raster and vector data
16:15	Wrap-up
16:30	END

Day 2

Time	Topic
09:30	Welcome and icebreaker, setup check
09:45	Raster Calculations in Python
10:30	**Coffee break**
10:40	Parallel raster computations using Dask
11:30	**Coffee break**
11:40	Parallel raster computations using Dask
12:30	**Lunch break**
13:30	Calculating Zonal Statistics on Rasters
14:30	**Coffee Break**
15:40	Calculating Zonal Statistics on Rasters
16:15	Post-workshop Survey
16:30	**Drinks**

Location logistics

- Coffee and toilets are in the hallway, just outside of the classroom.
- If you leave the building, be sure to be accompanied by someone from the eScience Center to let you back in through the ground floor door
- For access to this floor you might need to ring the doorbell so someone can let you in.
- In case of an emergency, you can exit our floor using the main staircase. Follow green light signs at the ceiling to the emergency staircase.
- Wifi: Eduroam should work. Otherwise use the "matrixbuilding" network, password will be printed out and available somewhere in the room.

The Netherlands eScience Center

netherlands eScience center

netherlands eScience center

Who are we?

- The Netherlands eScience Center is a national center for **innovative software solutions in academic research**.
- Established to bridge the gap between digital technologies and scientific and scholarly inquiry.
- Our Research Software Engineers
  - help researchers interpret results,
  - make tools and methods reusable for the wider research community,
  - co-author research and methodological publications.

netherlands eScience center

netherlands eScience center

Digital Skills Programme

- Hands-on courses, 2-3 days
- In person & online
- Based on The Carpentries & CodeRefinery, and
- In-house developed materials.

Topics cover:

- Open & Reproducible Research Software
- Intermediate and Advanced Research Software Skills
- Advanced Technologies (e.g. GPUs, Deep Learning)

Schedule 2023

January	• Data Analysis and Visualisation in Python for Researchers • Machine learning in Python with scikit-learn	**July**	• Parallel Programming with Python
March	• Reproducible research with R packages • Intermediate Research Software Development with Python • Astronomical Data Science with Python	**August**	• Introduction to Deep Learning
April	• Parallel Programming with Python • GPU programming	**September**	• Reproducible research with R packages
May	• Introduction to Geospatial Raster and Vector Data with Python • Introduction to Deep Learning	**October**	• Good Practices in Research Software Development (CodeRefinery • Introduction to Geospatial Raster and Vector Data with Python
June	• Image Processing with Python • Good Practices in Research Software Development (CodeRefinery)	**November**	• GPU programming • Machine learning in python with scikit-learn • Intermediate Research Software Development with Python

netherlands eScience center

netherlands eScience center

See upcoming workshops: esciencecenter.nl/events/?f=workshopsnl/events/?

To be notified about coming up workshops, subscribe to the Newsletter: eepurl.com/dtjzwP

netherlands eScience center

Do you want us to come and teach a custom workshop at your organisation?

Our Digital Skills Workshops are **free and open to researchers in the Netherlands**.

We also offer paid options of custom workshops dedicated to your organisation. To learn more, contact:

training@esciencecenter.nl

netherlands eScience center

Do you want to teach with our materials?

- All material that we use is open-source available
- We can help you setup workshops, contact training@esciencecenter.nl

netherlands eScience center

NL-RSE brings together the community of people writing and contributing to research software from Dutch universities, knowledge institutes, companies and other organizations to share knowledge, to organize meetings, and raise awareness for the scientific recognition of research software.

Website: https://nl-rse.org/
Email: info@nl-rse.org
Twitter: @nl_rse

netherlands eScience center

Let’s stay in touch

www.eScienceCenter.nl

info@esciencecenter.nl

+31 (0)20 460 4770

@eScienceCenter

Netherlands Escience Center

@eScienceCenter@akademienl.social

netherlands eScience center

netherlands eScience center

https://tinyurl.com/2023-10-10-geospatial-python

Let’s stay in touch

Case: Wildfires

Wildfires have been burning across Rhodes

https://news.sky.com/story/wildfires-on-rhodes-force-hundreds-of-holidaymakers-to-flee-their-hotels-12925583

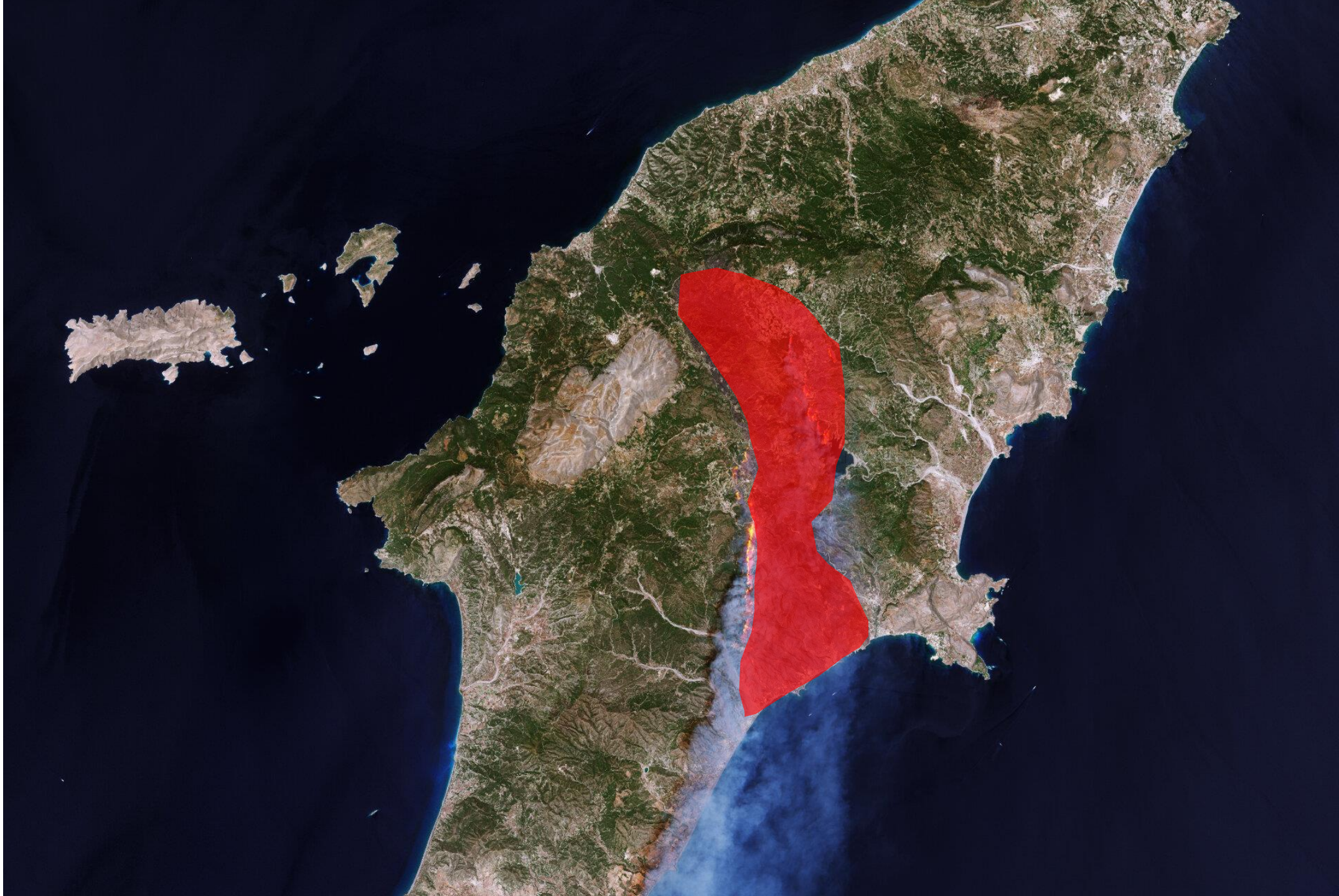
https://www.esa.int/ESA_Multimedia/Images/2023/07/Rhodes_wildfire_forces_thousands_to_flee

Question:

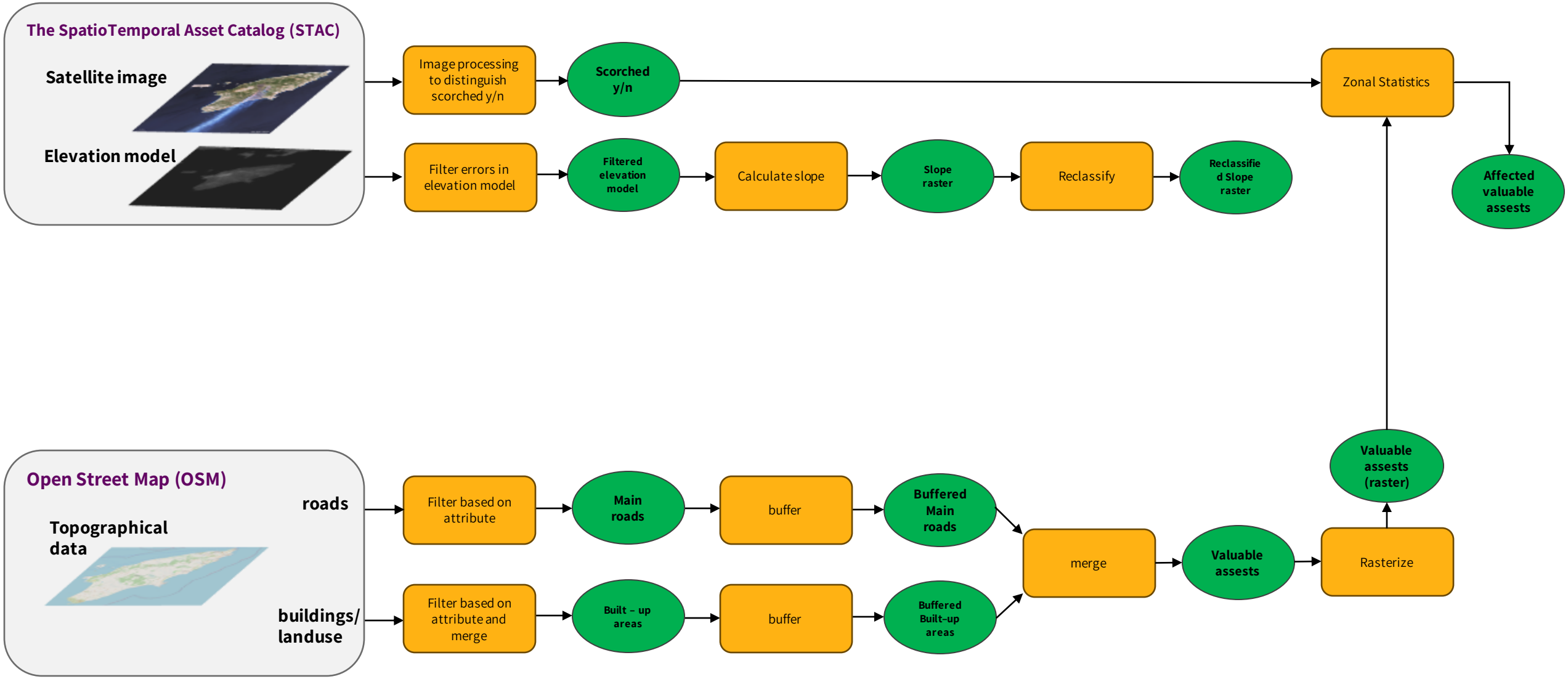
Which built-up areas are affected by the wildfire on Rhodos in the summer of 2023?

Aim:

We aim to generate a pipeline (recipe) which [1] identifies scorched areas based on satellite images and confronts this with built-up areas and infrastructures. Furthermore, **[2]** we research the relation between scorched areas and slopes.

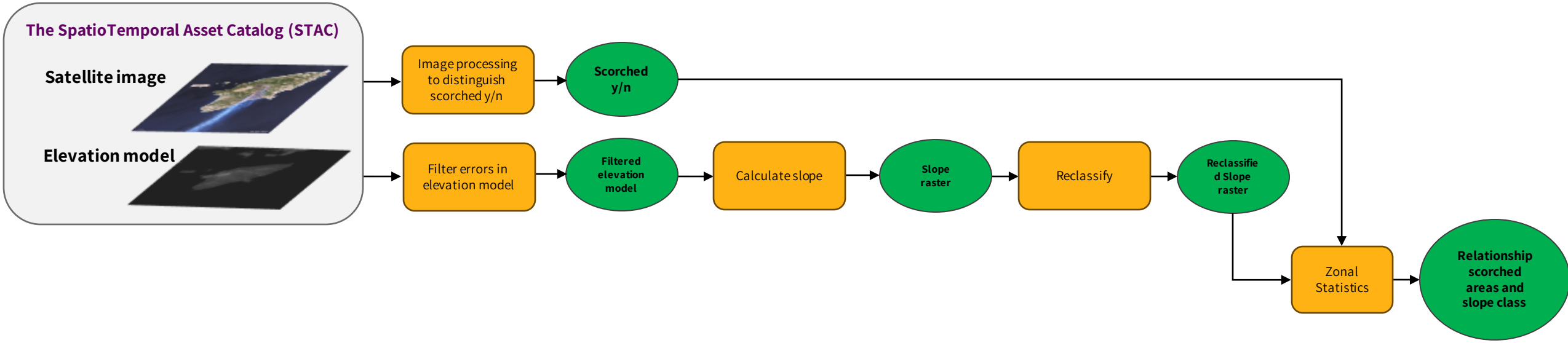


[1] “Recipe to identify wildfire affected areas “



DISCLAIMER!!! This pipeline is foremost meant for education purposes to get an understanding of using python for geospatial.

[2] “Recipe to research relationship between slope and scorched areas “



DISCLAIMER!!! This pipeline is foremost meant for education purposes to get an understanding of using python for geospatial.

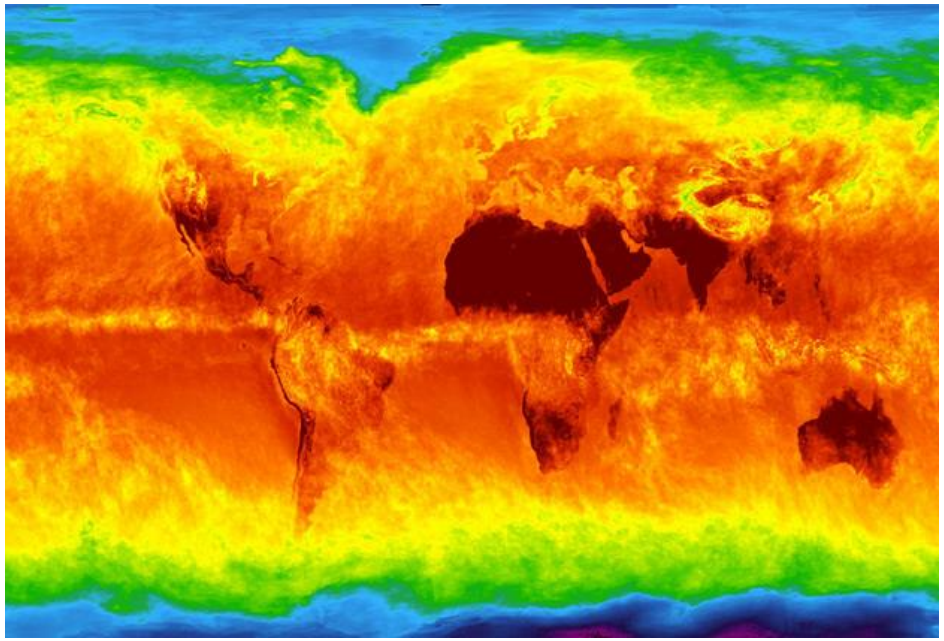
netherlands eScience center

A little theory about Geodata

A little theory about Geodata

discrete view and **continuous** view

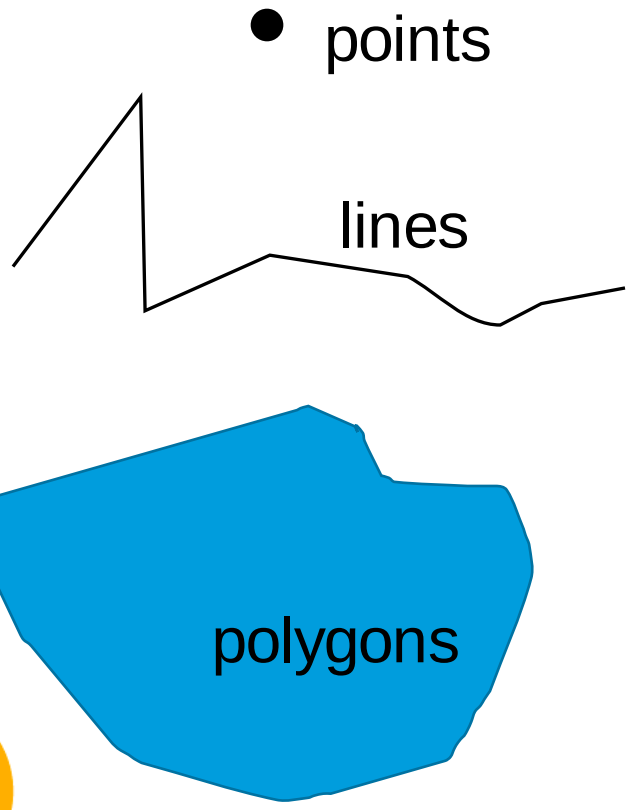
Discrete phenomena



Continuous phenomena

Geo Data Models

Vectordata



- ‘discrete objects’: points, lines and polygons (areas)
- Represented as sequences of coordinates (**vertices**)
- To every object an attribute table can be connected.

Geo Data Models

Rasterdata

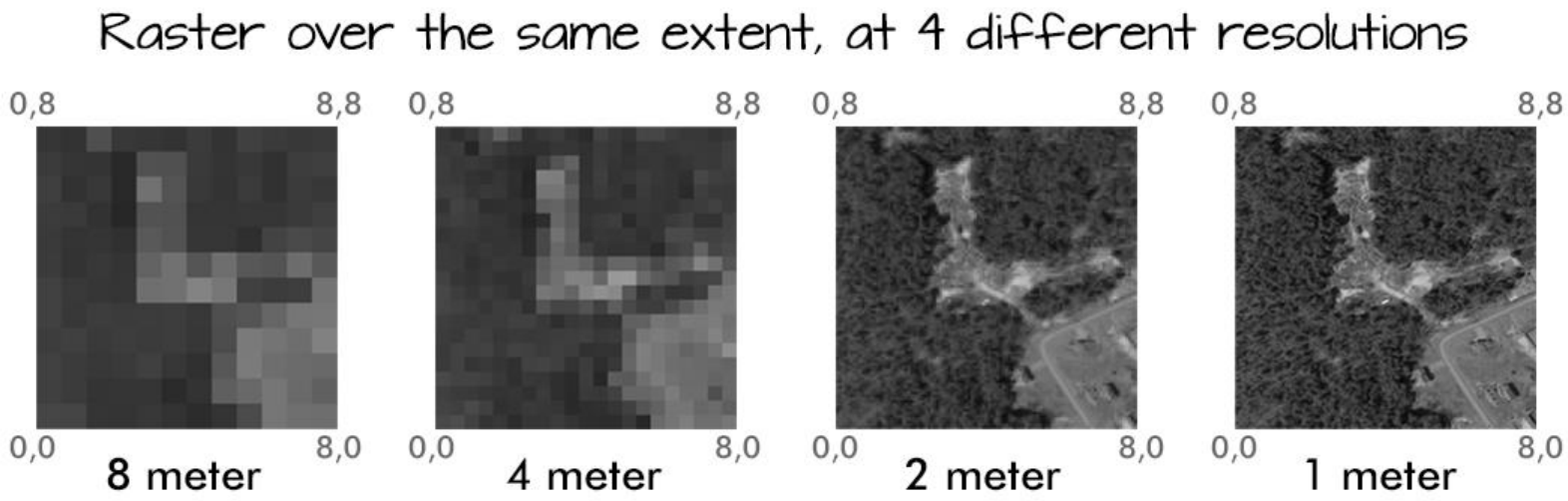
- rows and columns with numbers (matrix)
- Every cell in a raster has
  - a position
  - a value (or multiple)

1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	3	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	1	2	2	2	3	3	3
1	1	1	1	1	1	3	3	3	3
1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	1	1	3	3

netherlands eScience center

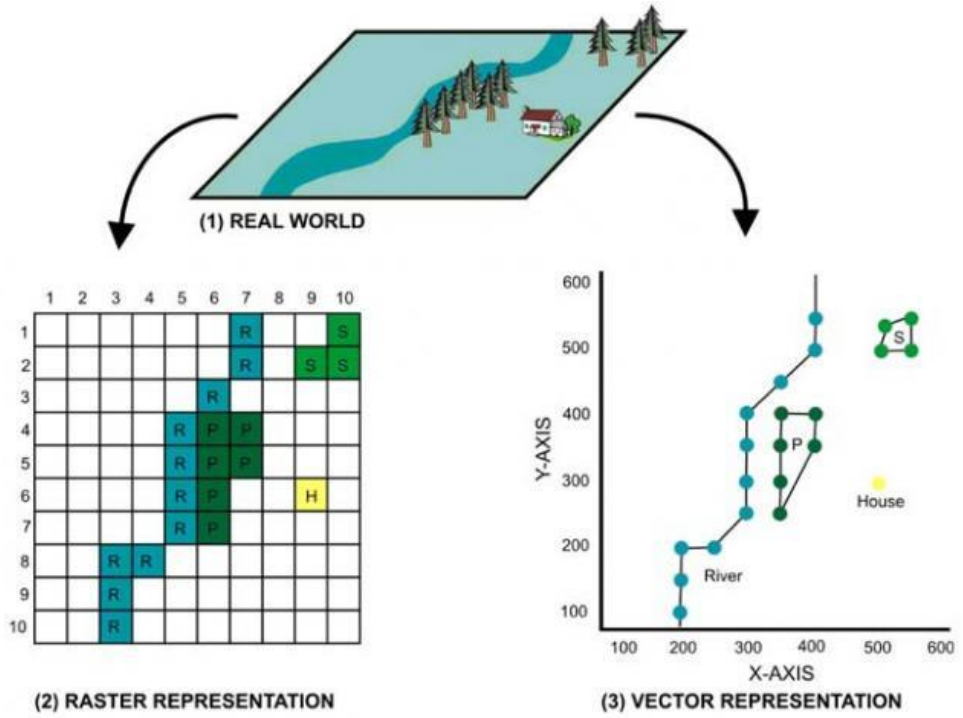
Geo Data Models

Rasterdata



Geo Data Models

Raster



Vector

You can represent many phenomena as raster and vector

(although representing continuous phenomena as vector is very uncommon)


But wait ...
the Earth is not flat ...

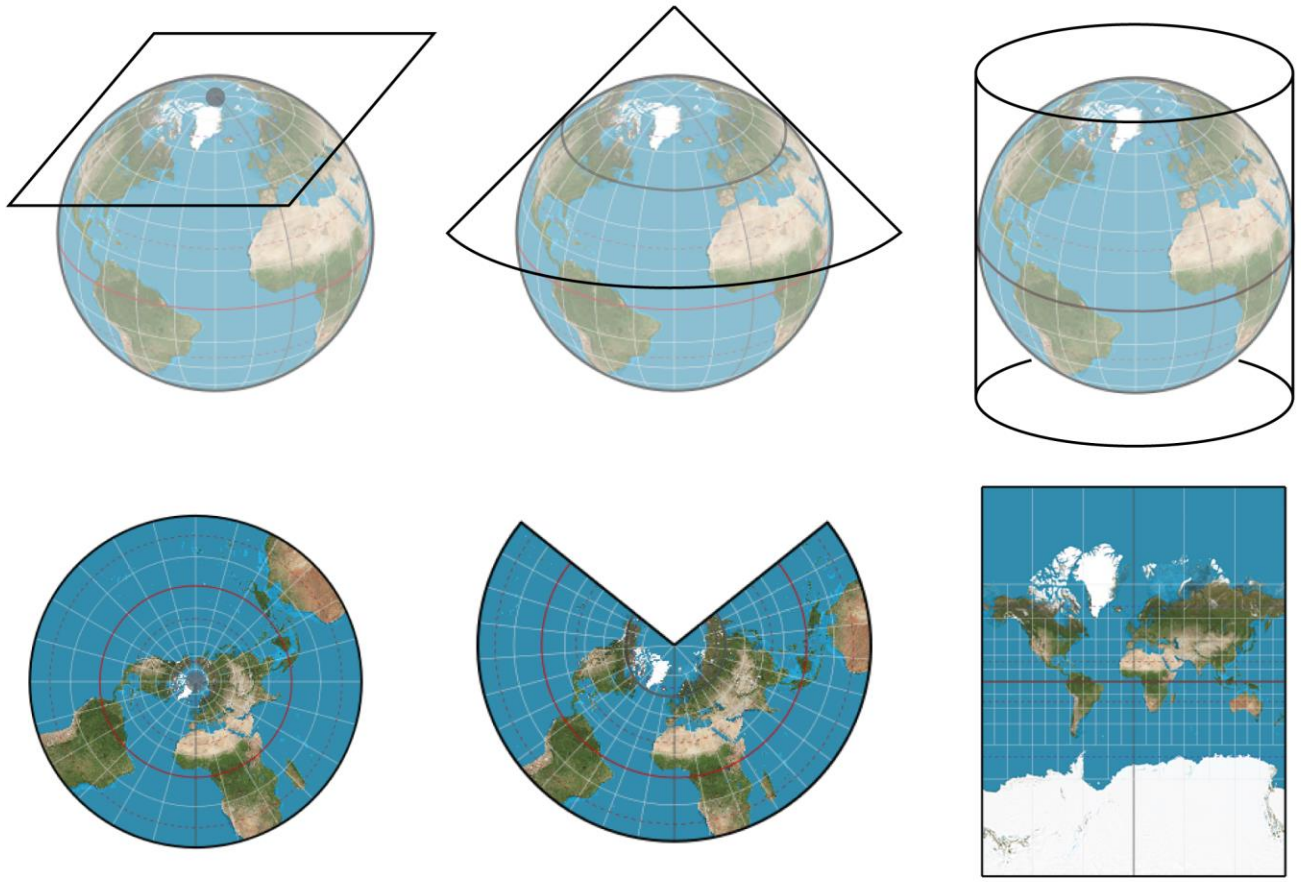
Coordinate Reference System (CRS)

Ellipsoid: is Earth an orange? Or a lemon?

Projection:
How can I peel this orange/lemon?

Coordinate Reference System (CRS)

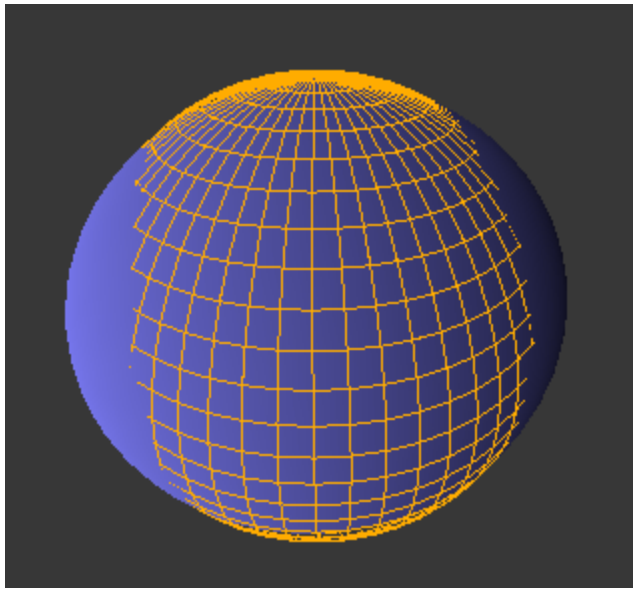
Projected coordinate systems

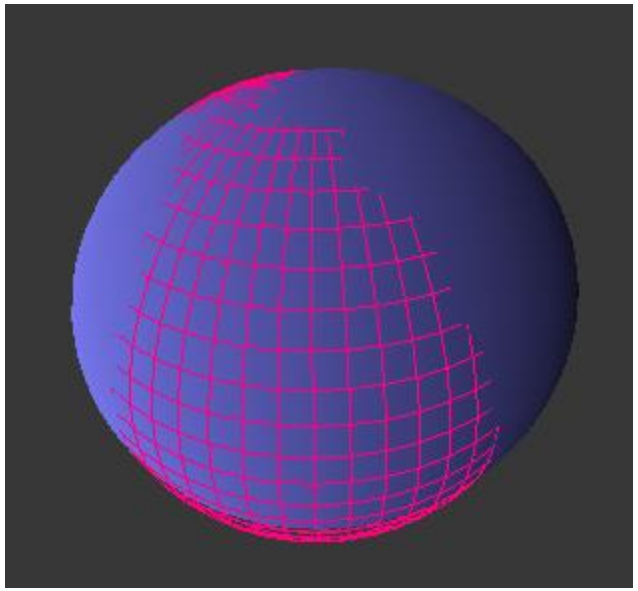


https://gistbok.ucgis.org/bok-topics/2018-quarter-04/map-projections

Coordinate Reference System (CRS)

Ellipsoid





Geographic coordinate system defines the ellipsoid.

http://www.georeference.org/doc/the_earth_as_an_ellipsoid.htm

https://www.thetruesize.com/

Coordinate Reference System (CRS)

A CRS mainly contains :

- **Ellipsoid:** A model of the shape of the earth
- **Projection:** A mathematical transformation from a global earth to a flat surface.

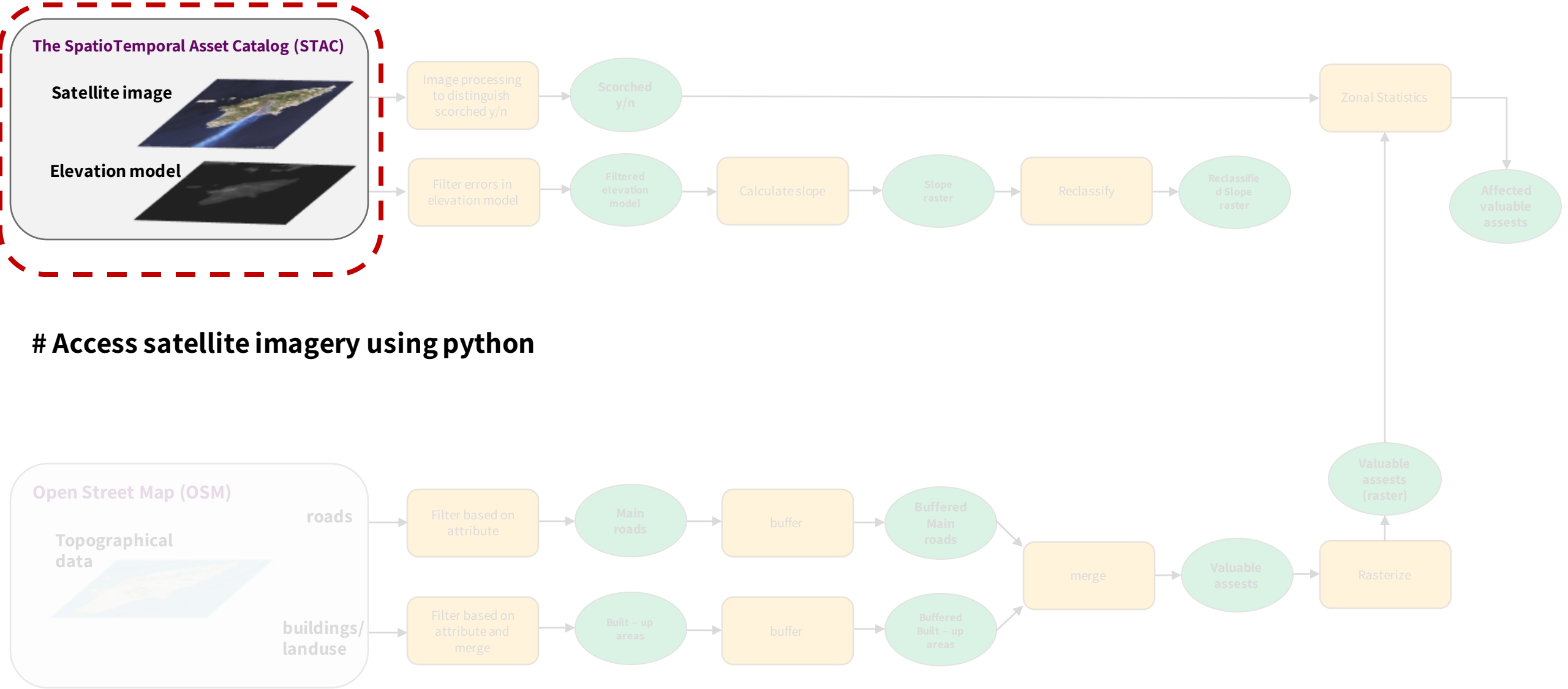
netherlands eScience center

The Geospatial **(PYTHON)** Landscape

- Open satellite data access: ***pystac_client***, ***pystac***
- Work with raster: ***rioxarray***, ***xarray-spatial***
- Work with vector: ***geopandas***

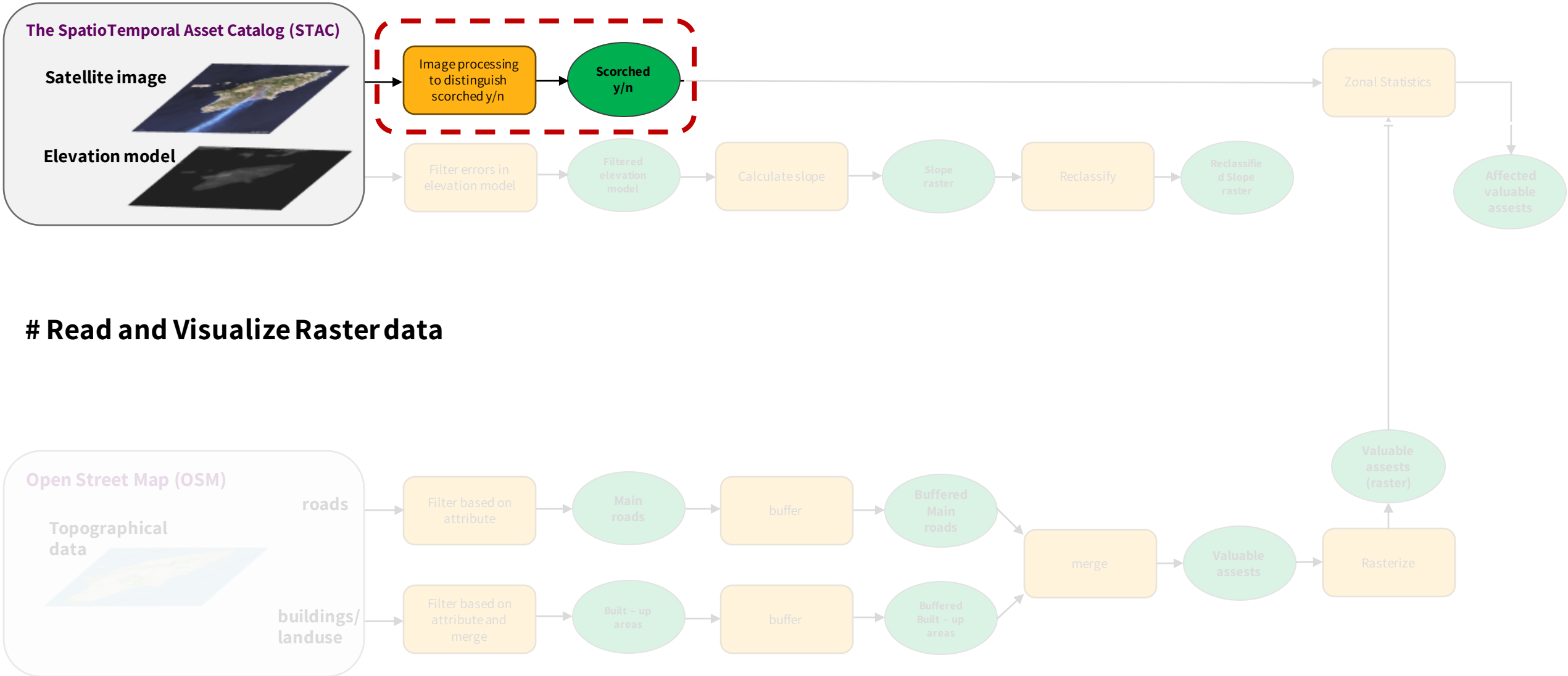
- ~~*Proprietary software:* ***arcpy*** *~ both raster and vector*~~

[1] “Recipe to identify wildfire affected areas “



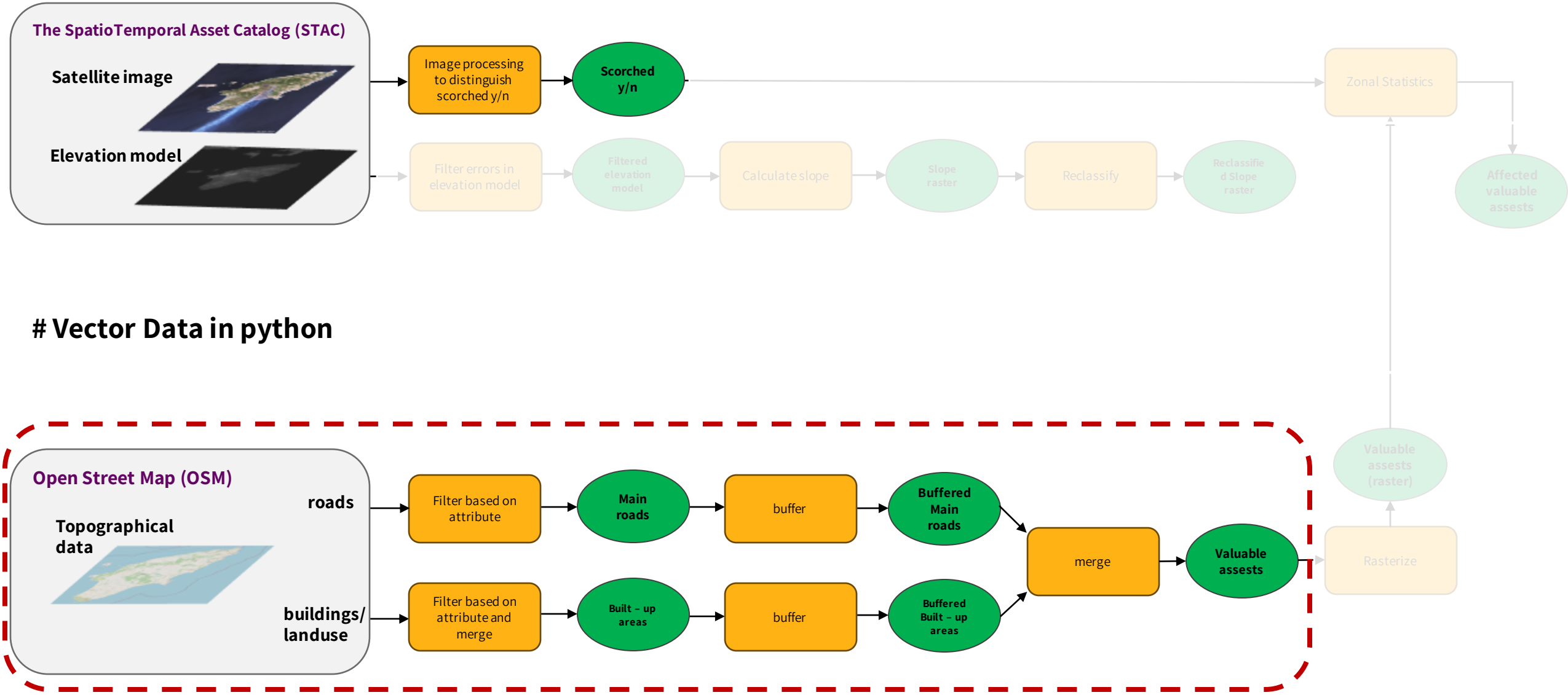
 # Access satellite imagery using python

[1] “Recipe to identify wildfire affected areas “

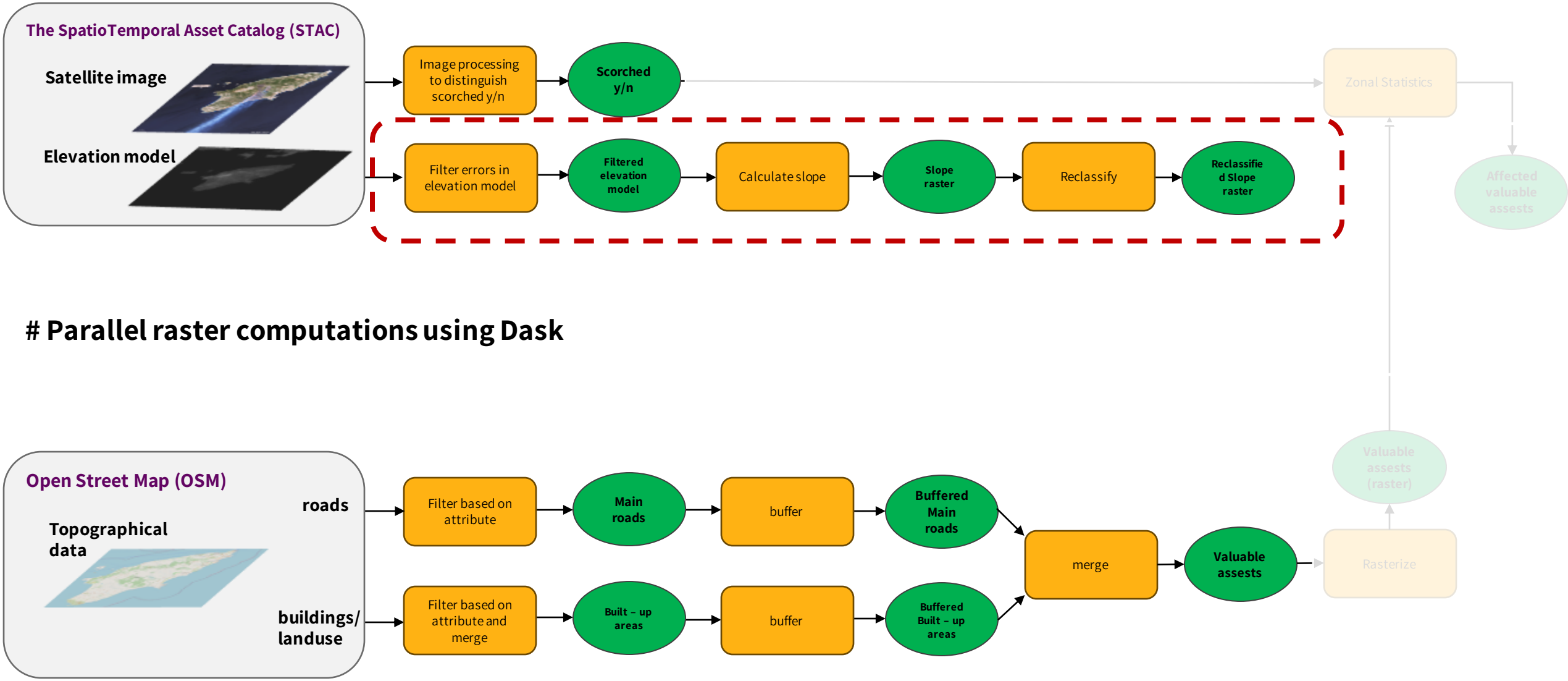


Read and Visualize Raster data

[1] “Recipe to identify wildfire affected areas “

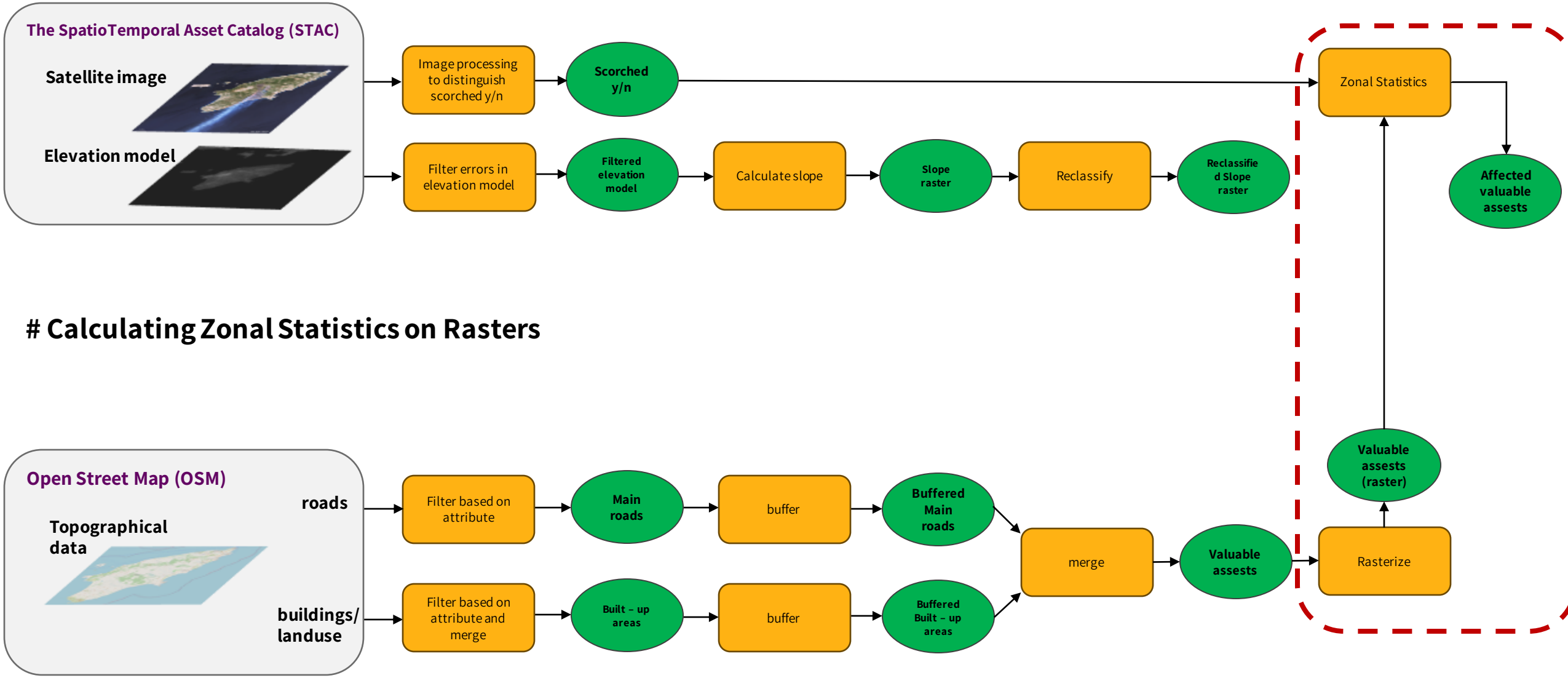


[1] “Recipe to identify wildfire affected areas “



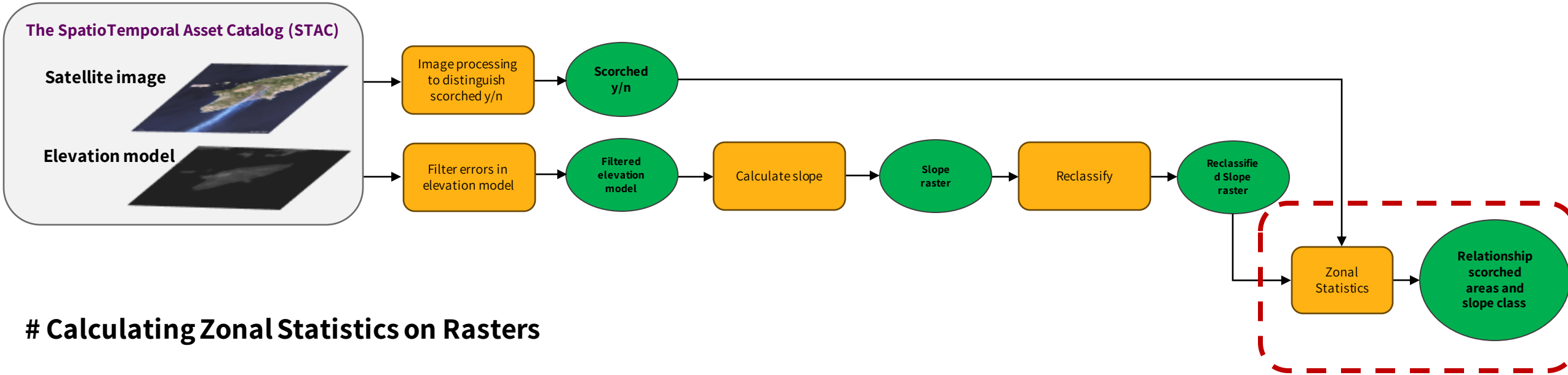
Parallel raster computations using Dask

[1] “Recipe to identify wildfire affected areas “



Calculating Zonal Statistics on Rasters

[2] “Recipe to research relationship between slope and scorched areas “



Calculating Zonal Statistics on Rasters

netherlands eScience center

Raster Analysis

- Map algebra / cartographic modelling
  - Sum (+), subtract (-), multiply (x), divide (:)
  - functies (sin, cos, tan, ln, log)
  - Reclassify
  - Logical operators screening (AND, OR, XOR) and ranking (MIN, MAX)

- Different Operations

Local Operations performed on a cell by cell basis

Focal Operations performed using a moving group of cells

Zonal Operations performed using zones (groups of cells having the same value)

Global Operations performed using the whole grid

Local: Calculations

- Multiply

3	0	4
1	2	2
5	1	3

x 3 =

9	0	12
3	6	6
15	3	9

- Multiply with other raster

3	0	4
1	2	2
5	1	3

x

2	5	1
6	3	1
2	4	2

=

6	0	4
6	6	2
10	4	6

netherlands eScience center

Local: exclusionary ranking

1	2	4
0	1	2
1	3	1

MAX

4	2	1
1	3	1
0	2	2

=

4	2	4
1	3	2
1	3	2

1	2	4
0	1	2
1	3	1

MIN

4	2	1
1	3	1
0	2	2

=

1	2	1
0	1	1
0	2	1

netherlands eScience center

Local: weighted addition

1	2	4
0	1	2
1	3	1

+

4	2	1
1	3	1
0	2	2

=

5	4	5
1	4	3
1	5	3

3 x

1	2	4
0	1	2
1	3	1

+ 2 x

4	2	1
1	3	1
0	2	2

=

11	10	14
2	9	8
3	13	7

netherlands eScience center

Local: Reclassification

One-to-one change – input raster cell value is replaced with new value in the output raster (integer rasters only)

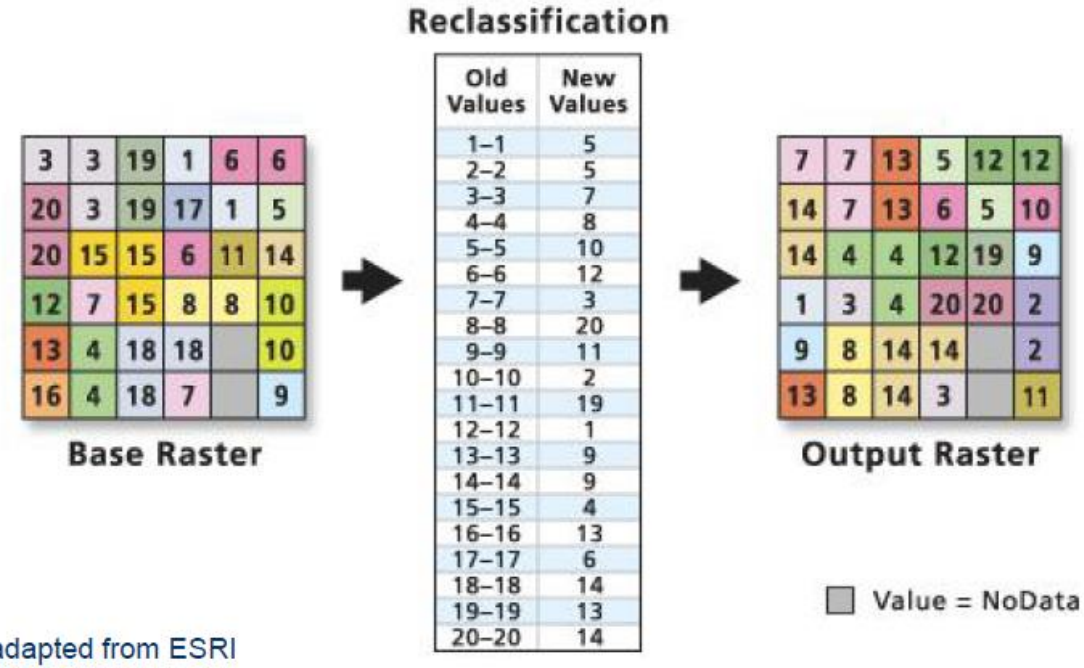


Diagram adapted from ESRI ArcGIS 9.3 Helpdesk

netherlands eScience center

Focal

- ‘kernel’ or neighbourhood
- functions
  - sum, average, maximum, minimum, deviation

netherlands eScience center

Focal: Slope class

$$\tan S = \sqrt{\left(\frac{\delta Z}{\delta X}\right)^2 + \left(\frac{\delta Z}{\delta Y}\right)^2}$$

δZ/ δX =

-1Z	0Z	1Z
-2Z		2Z
-1Z	0Z	1Z

/8ΔX

δZ/ δY =

1Z	2Z	1Z
0Z		0Z
-1Z	-2Z	-1Z

/8ΔY

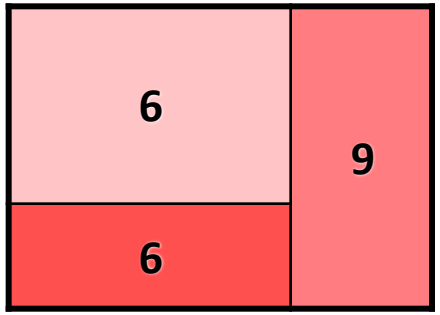
netherlands eScience center

Zonal

- Calculations for “zones”
- functions
  - sum, average, maximum, minimum, deviation

3	0	4
1	2	2
5	1	3

A	A	B
A	A	B
C	C	B



ZONALSUM

netherlands eScience center

Global

- Calculation for whole grid
- distance, viewshed, hydrological models etc.

		1

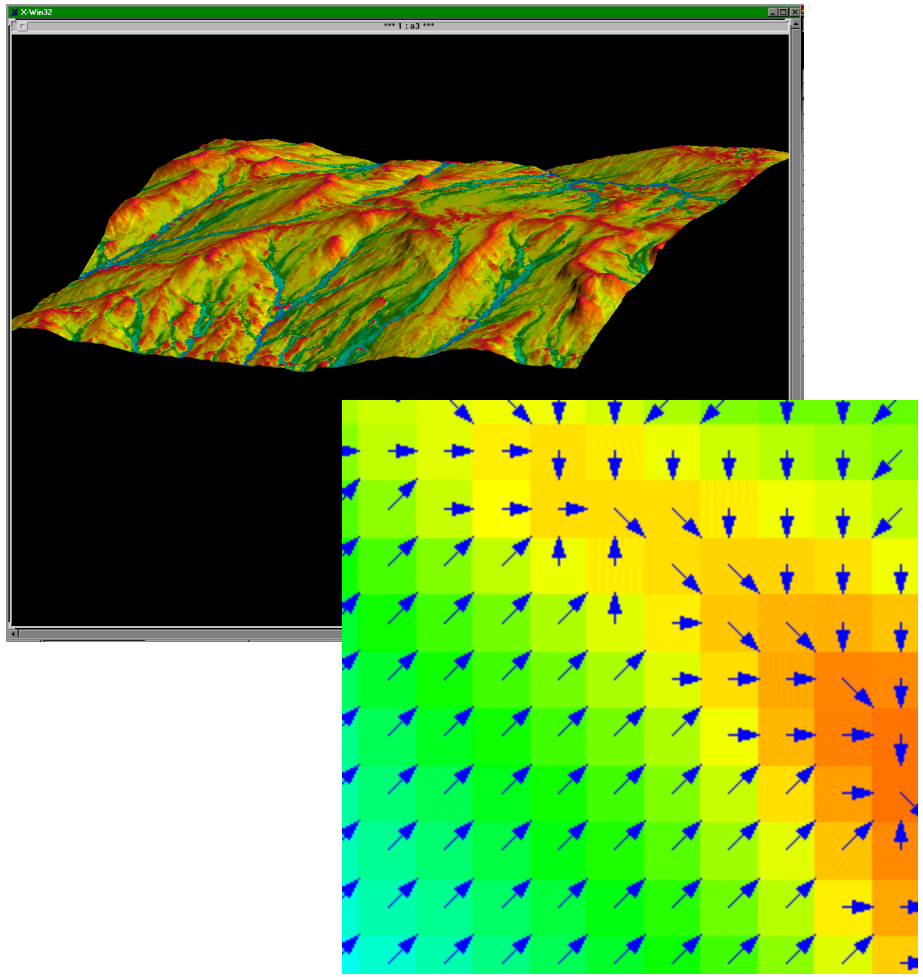
EUCLIDEAN DISTANCE

2	1	0
2,4	1,4	1
2,8	2,4	2

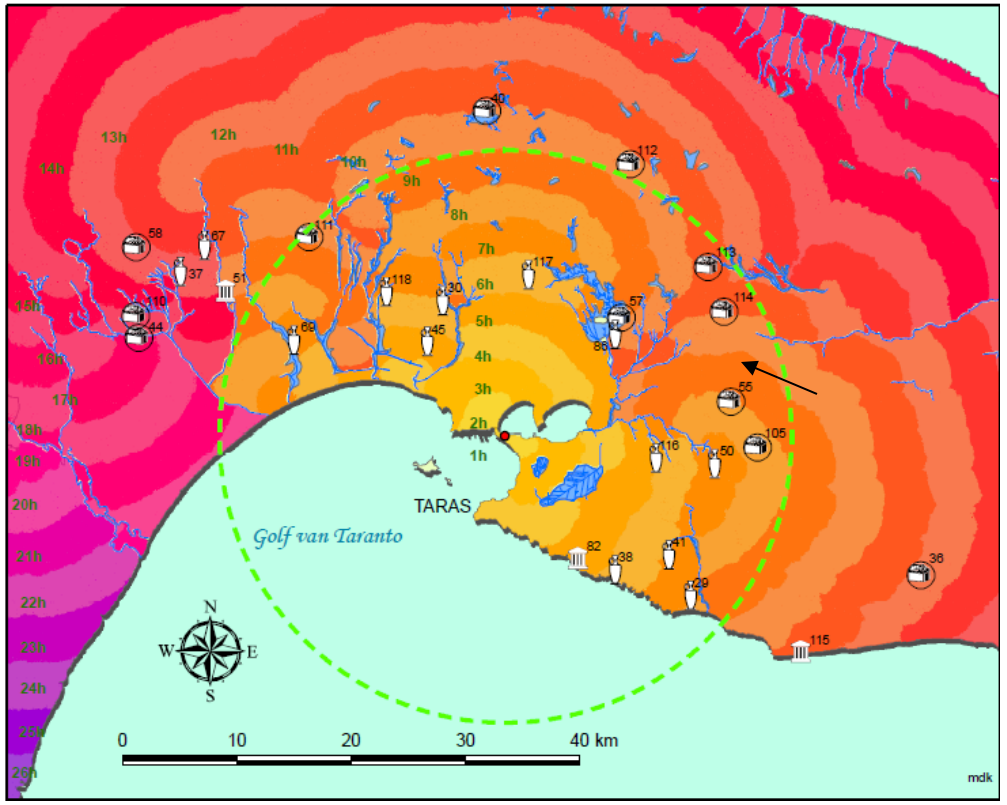
netherlands eScience center

Global: examples

Hydrological modelling

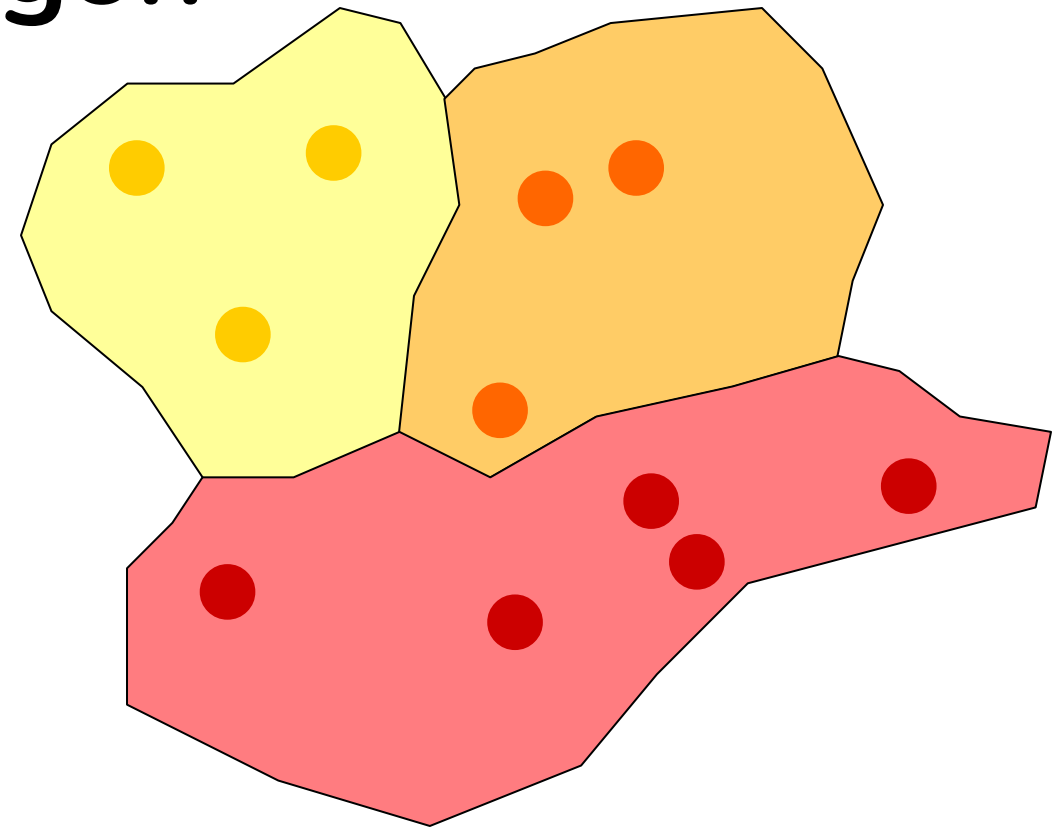


Travel time (on foot)



Vector Analysis

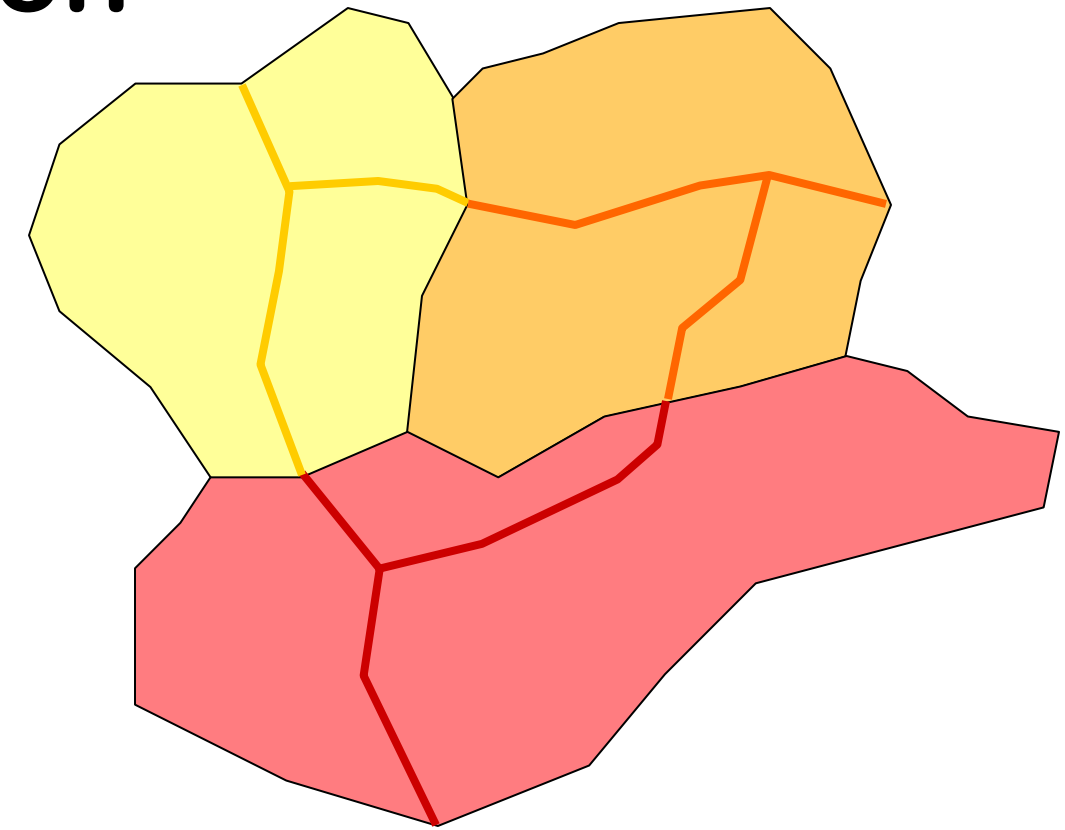
point-in-polygon



netherlands eScience center

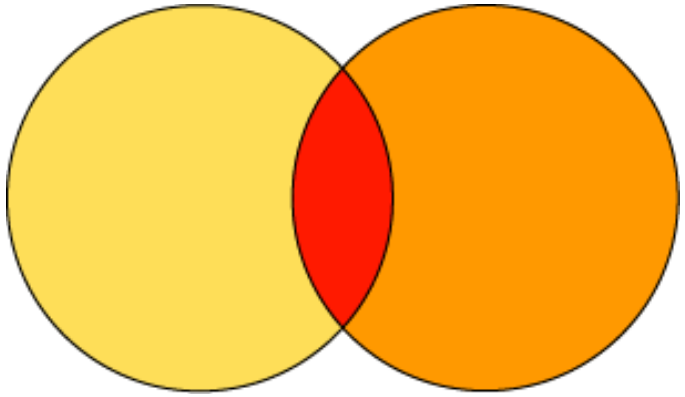
Vector Analysis

line-in-polygon

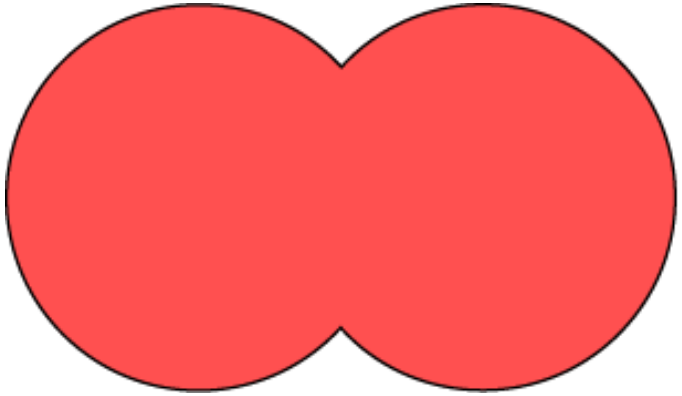


Vector Analysis

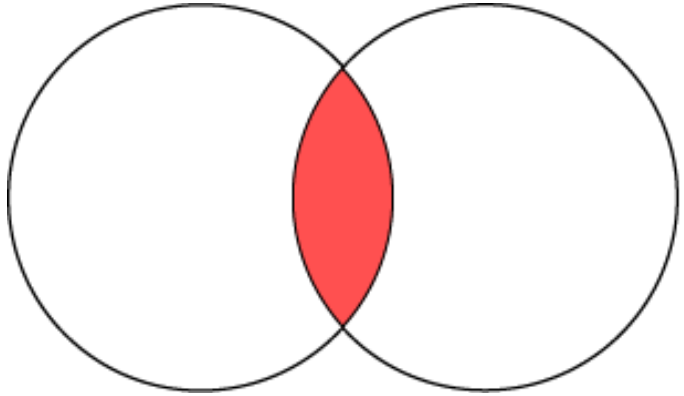
overlay



UNION



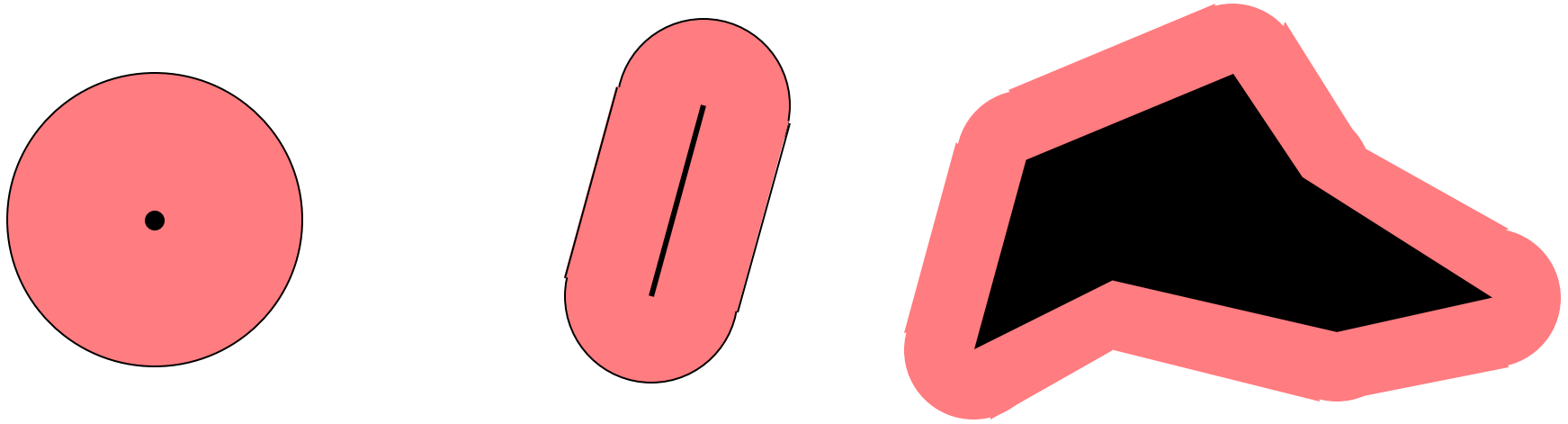
COMBINE/DISSOLVE



INTERSECT/CLIP

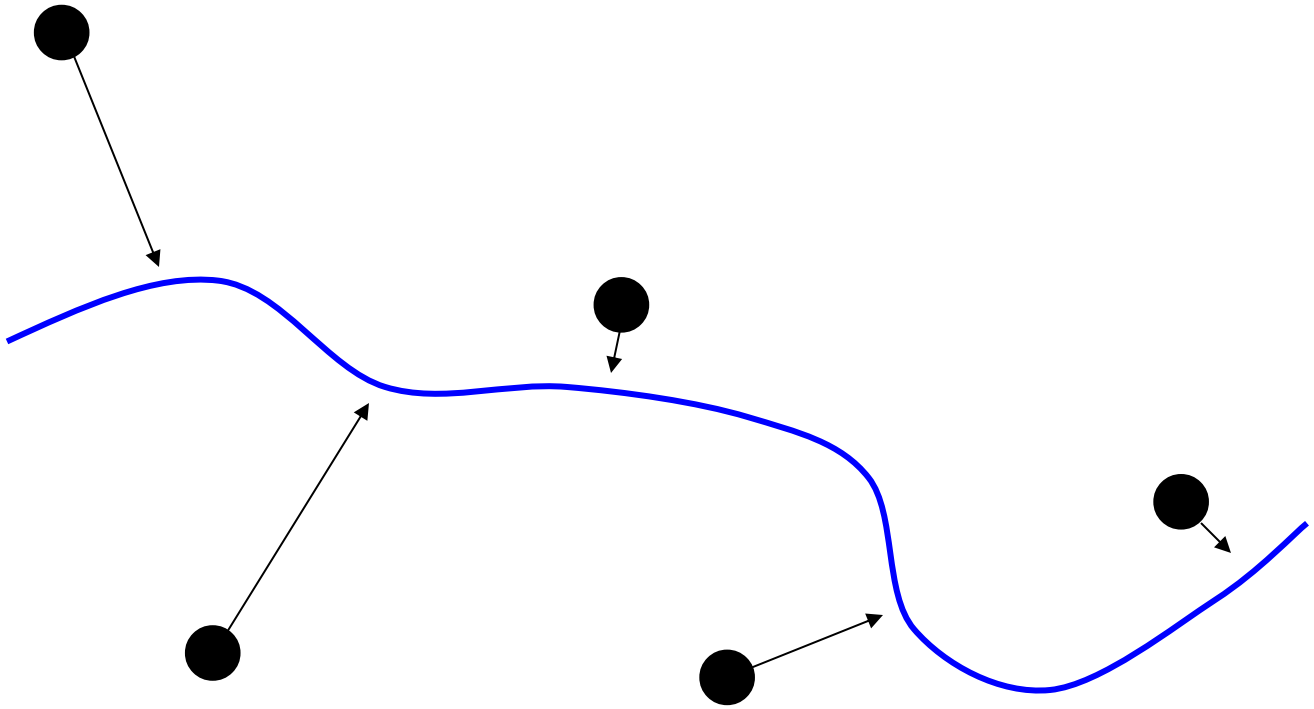
Vector Analysis

buffer



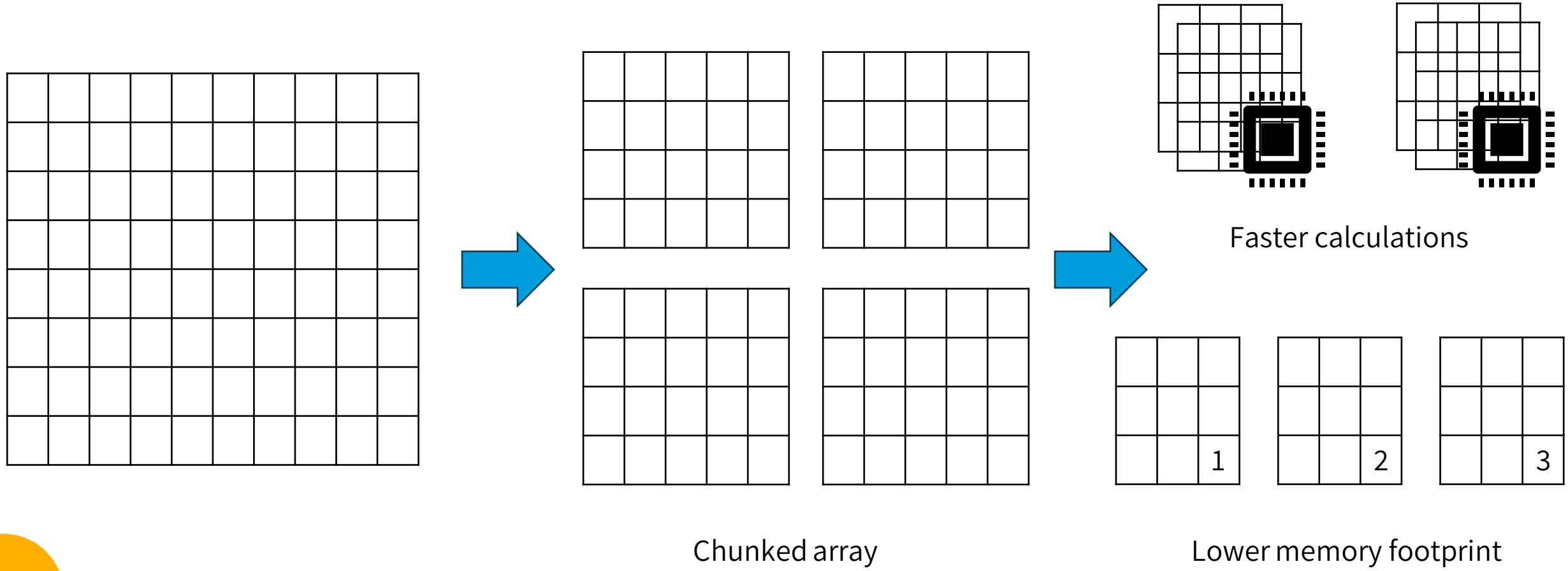
Vector Analysis

Nearby



netherlands eScience center

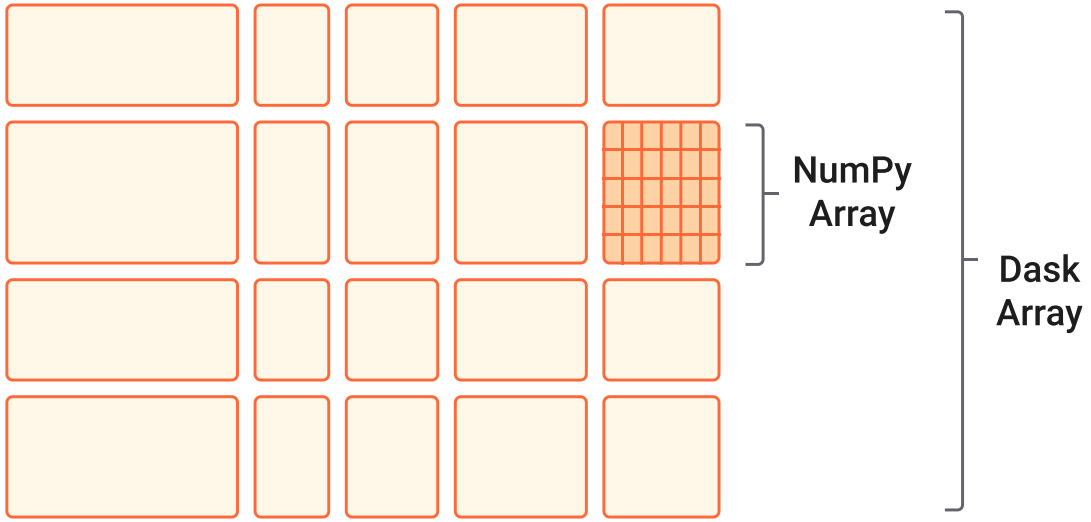
Parallel Raster Calculations



netherlands eScience center

Dask

- Python library for parallel and distributed computing.
- Different data structures and abstractions – we'll use **Dask Arrays.**
- Well integrated with (rio)xarray.



https://www.dask.org/

netherlands eScience center

Parallel or serial … ?

- Performance gain highly depends on:
  - Operations/algorithms
  - Size of the dataset
  - "Parameters", e.g. size (and shape) of the chunks
- Parallelization brings overhead – **serial calculations can be faster!**
- Always start from serial calculations, time profile