Development of a tool to calculate distance for veterinary epidemiological applications

Mirco Cazzaro, Francesca Scolamacchia, Paolo Mulatti and Nicola Ferré

Istituto Zooprofilattico Sperimentale delle Venezie – GIS Office – Legnaro, Italy
Department of Information Engineering – University of Padova – Italy

27th Association of Geographic Information Laboratories in Europe 2024 (AGILE 2024)

Glasgow, 6th June 2024

TABLE OF CONTENTS

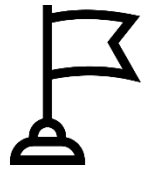
01. INTRODUCTION

PROBLEM

Control measures for animal diffusive diseases are generally based on euclidean distances, which might not be optimal in certain situations (e.g. mountain areas)

TASK

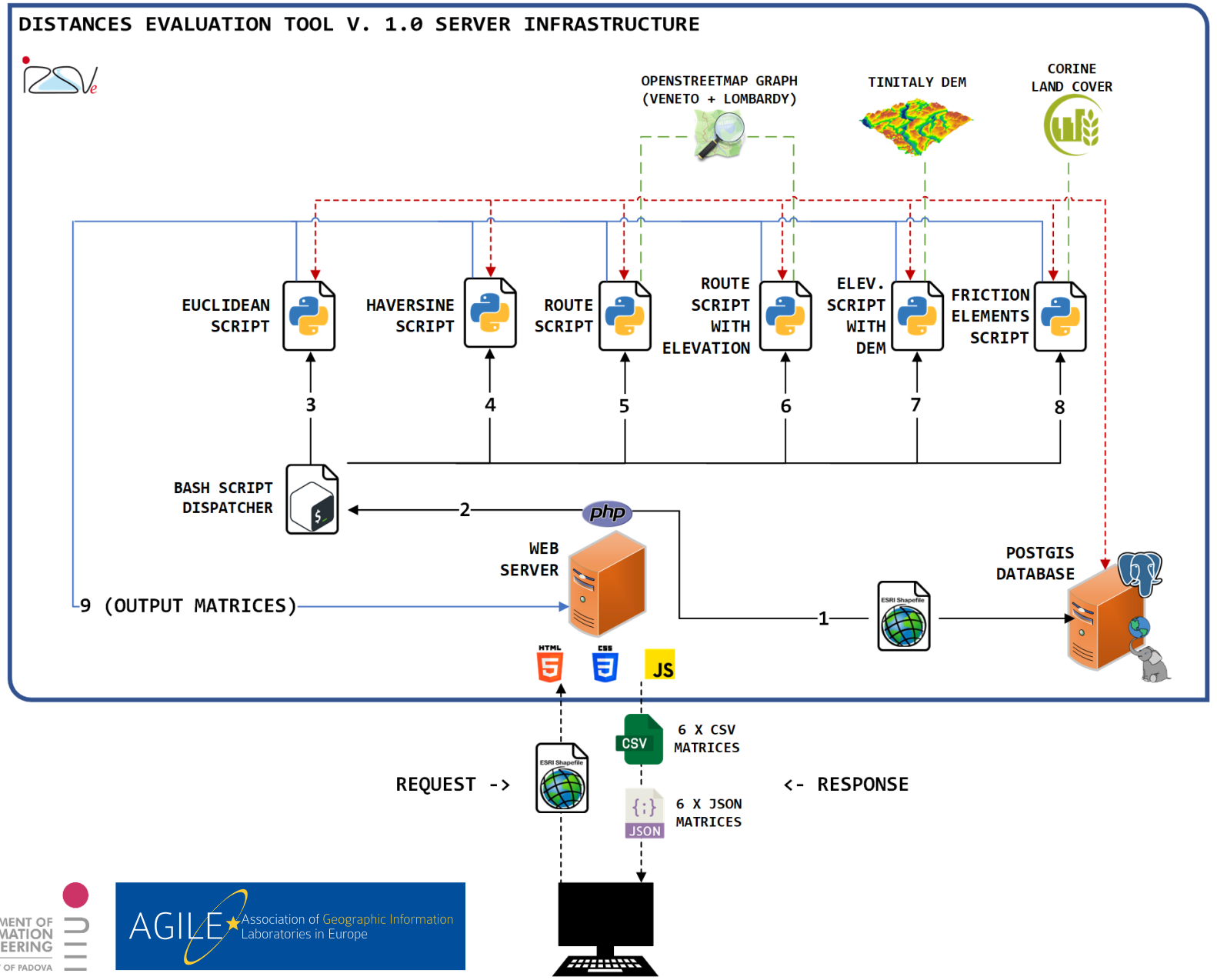
Develop an effective and efficient system capable of computing different distance matrices between farms, accessible through a user-friendly web interface



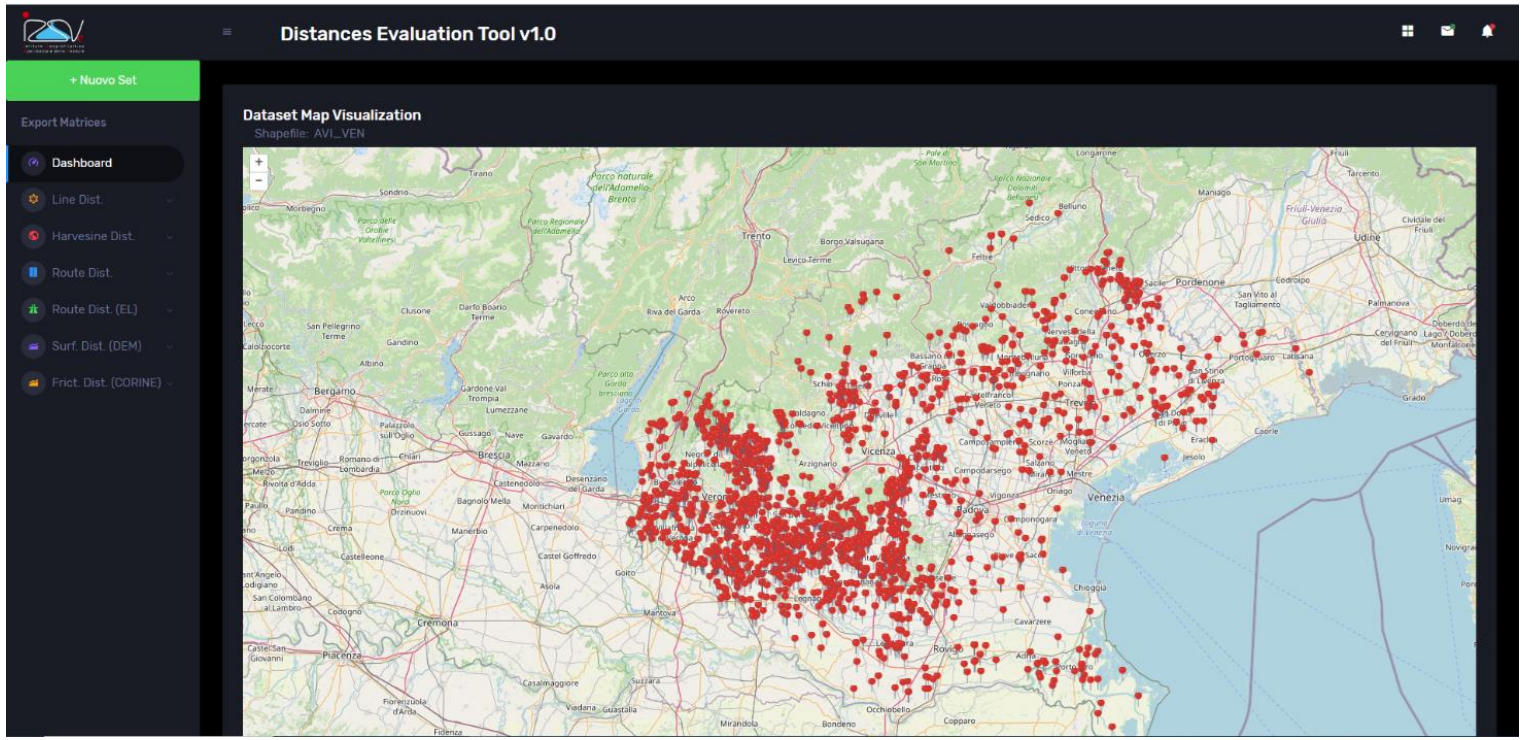
GOALS

Test both output correctness and temporal system performance with respect to available CPUs

02.0. SYSTEM ARCHITECTURE



02.1. SYSTEM ARCHITECTURE – WEB APPLICATION

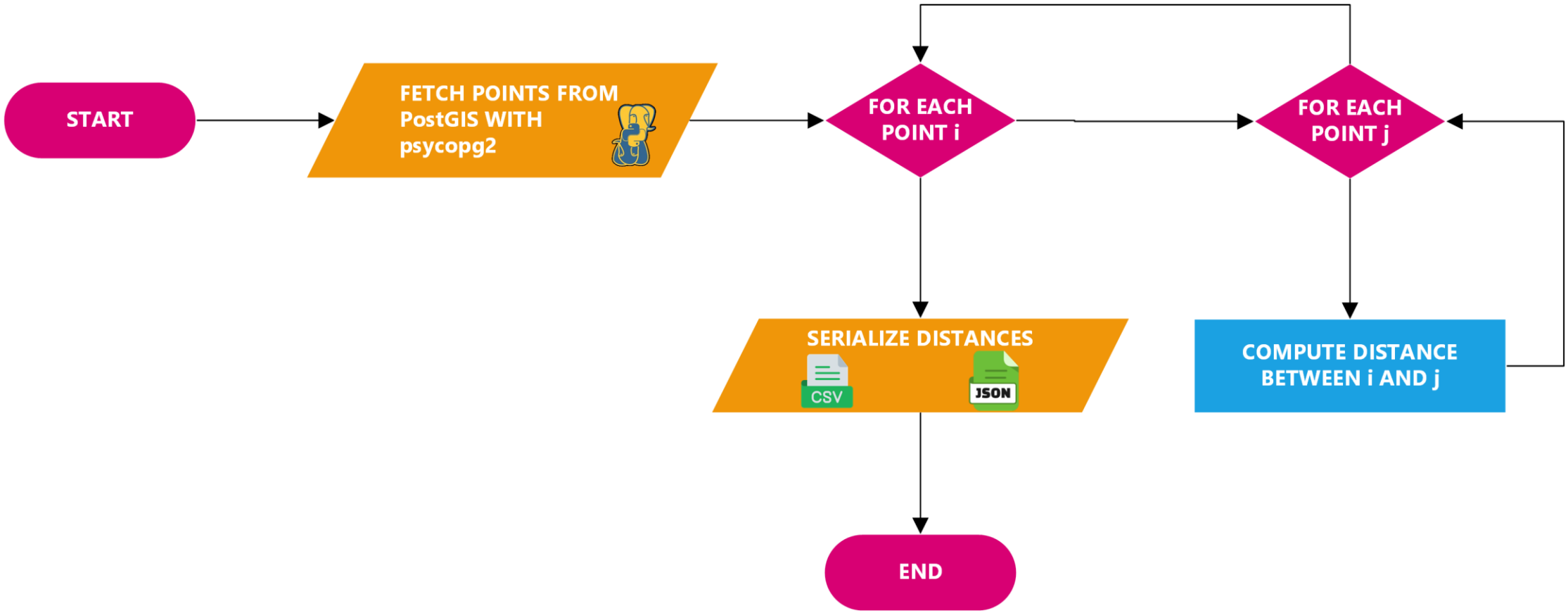


Through the Web Application, users can trigger the system execution.

By clicking on «+ Nuovo Set» the user is required to upload both the .shp and .shx files containing points representing farms.

The Web Application first stores these files in a PostGIS database, reloads the main page to display the points on a map, and then triggers a bash script that runs all the Python scripts in sequence.

02.2. SCRIPTS LOGIC



02.2.1. EUCLIDEAN & HAVERSINE

- **Euclidean Distance:**

$$dist_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

- **Haversine Distance**

$$a = sin(\frac{x_i - x_j}{2})^2 + cos(x_j) \cdot cos(x_i) \cdot sin(\frac{y_i - y_j}{2})^2$$

$$c = 2 \cdot atan2(\sqrt{a}, \sqrt{1-a})$$

$$dist_{i,j} = 6.373 \times 10^6 \cdot c$$

02.2.2. PATH DISTANCES (WITH AND WITHOUT ELEVATION)

GraphHopper + OpenStreetMap GEOFABRIK downloads

GraphHopper is an open-source routing library and server written in Java and provides a routing API over HTTP.

Geofabrik is a for-profit company based in Karlsruhe, Germany. Geofabrik offers many free services to the OpenStreetMap community, including daily data extracts.

Given 2 points `i` and `j`:

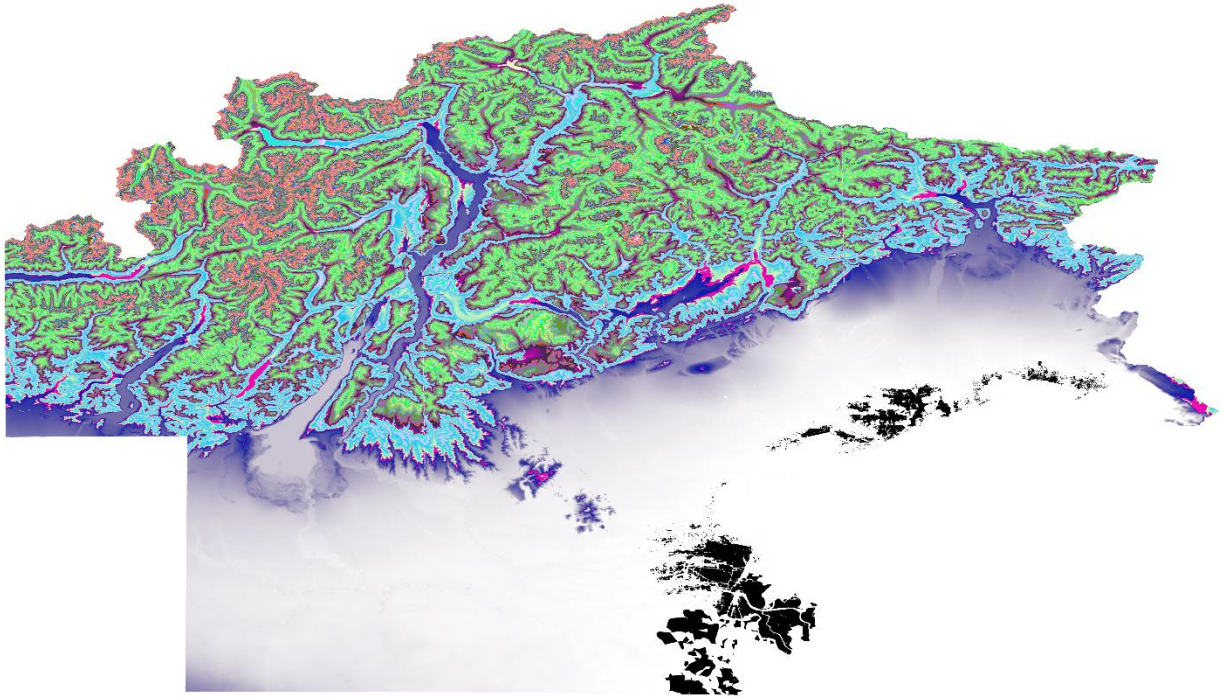
- Path distance without elevation

```
/route?point={i}&point={j}&points_encoded=false&elevation=false
```

- Path distance with elevation

```
/route?point={i}&point={j}&points_encoded=false&elevation=true
```

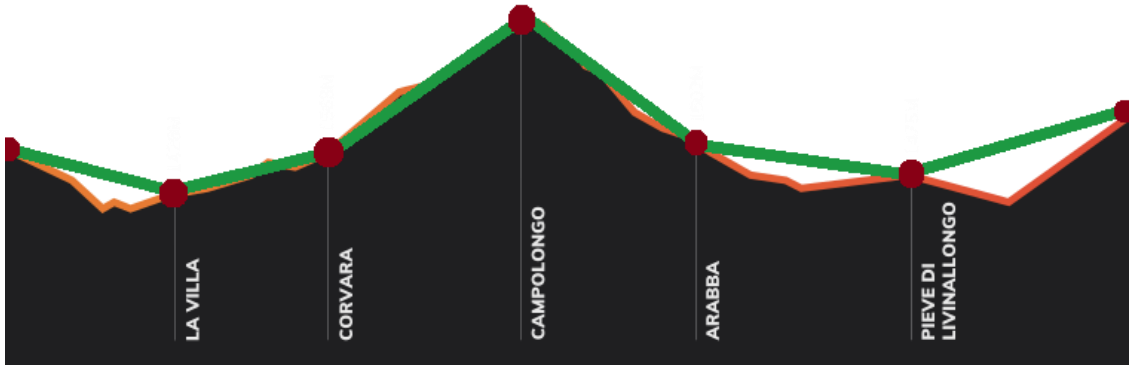
02.2.3. ELEVATION DISTANCES (WITH D.E.M.)



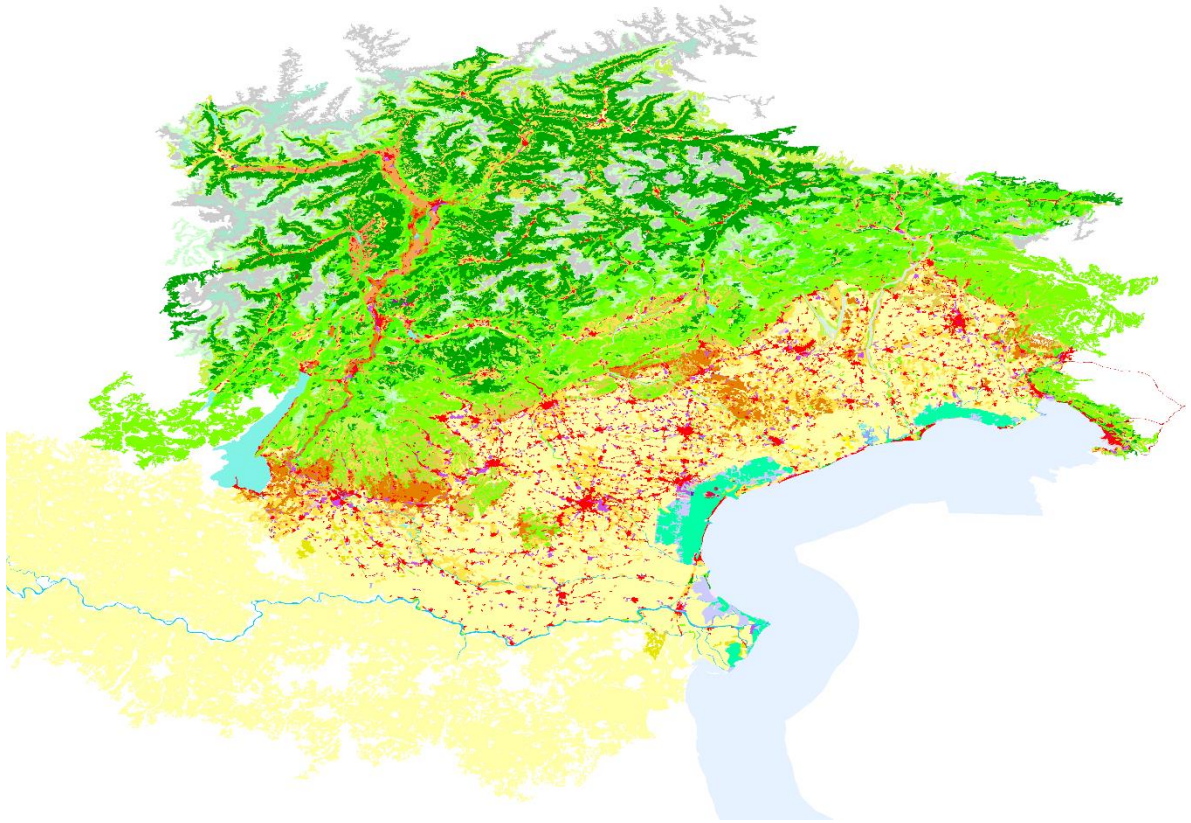
10-meters precision Digital Elevation Model currently used in our system

Rasterio

Given 2 points, we take the straight line connecting them and sample the altitude every 10 meters. The final distance is calculated as the sum of all segments connecting the 3-dimensional samples.



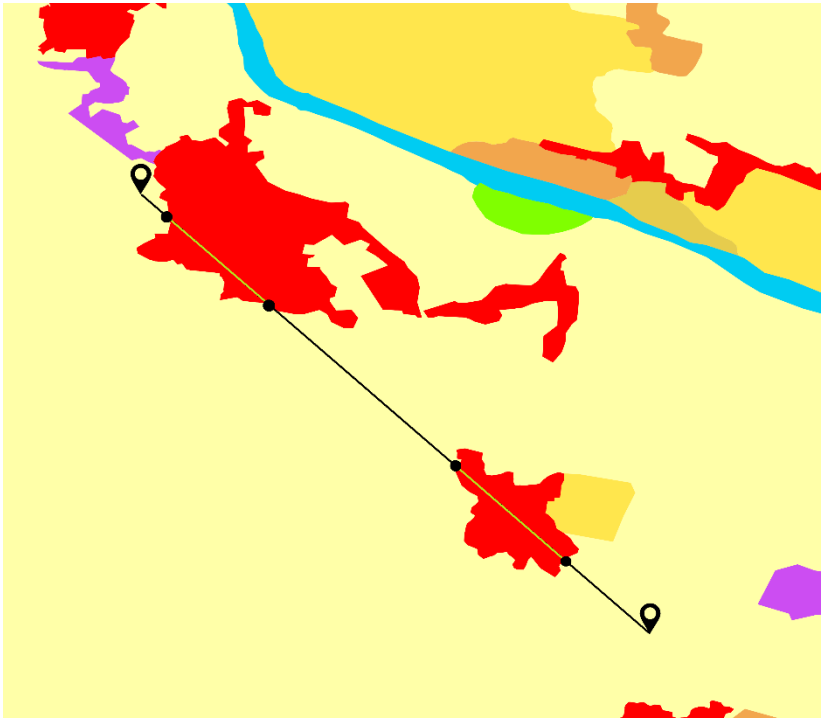
02.2.4. FRICTION ELEMENTS DISTANCES (WITH C.L.C.)



Corine Land Cover vector dataset used in our system (2018 v.20)

GeoPandas

Given 2 points, we draw a straight line connecting them and find the intersection of this line with the geometries that compose the Corine dataset. We then sum up the lengths of these intersections, grouped by the CLC 1st level.



Istituto Zooprofilattico Sperimentale delle Venezie

DEPARTMENT OF INFORMATION ENGINEERING UNIVERSITY OF PADOVA

AGILE Association of Geographic Information Laboratories in Europe

03. RESULTS

Temporal asymptotic performances of the algorithms:

- Euclidean and Haversine: $\Theta\left(\frac{n^2}{2}\right)$ (where each of the n operations involves using NumPy or math libraries operations);
- Path distances with and without elevation: $\Theta\left(\frac{n^2}{2}\right)$ (where each of the n operations involves an HTTP call);
- Elevation distances with DEM: $\Theta\left(\frac{n^2}{2} \cdot \frac{d}{10}\right)$ (where d is the average Euclidean distance among all pairs of points);
- Friction elements distances with CLC: $\Theta\left(\frac{n^2}{2} \cdot k\right)$ (where k is the number of different vector features in the model).

Computation timings:

Environments:
· **VM Server**: 4 GB of RAM, 2 physical cores of an Intel® XEON CPU, 80 GB HDD;

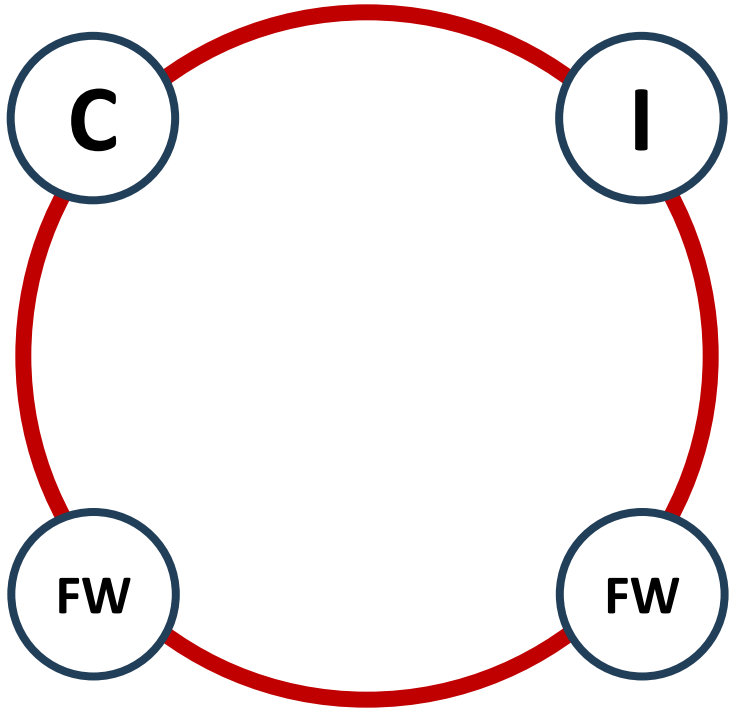
· **Lenovo® Thinkpad P15v**: 32 GB of RAM, Intel® Core i9-12900H 20 cores, 512 GB SSD M2 Micron® MTFDKBA512TFK.

Script	VM Server	Lenovo® Thinkpad
Euclidean	30 secs	3 secs
Haversine	35 secs	5 secs
Route	4 hrs	25 mins
Route with elevation	4 hrs	25 mins
Elevation with DEM	*10 hrs*	45 mins
Friction elements CLC *	*5 days*	12 hrs

* NB: Timing costs for friction element (CLC level no. 5, water) need to be multiplied by 5 to compute all the 5 levels, increasing them to 25 days or 60 hours, based on the system used.

04. CONCLUSION, IMPROVEMENTS & FUTURE WORKS

As demonstrated in this study, various models and technologies for extracting different types of distances can be integrated into a unified framework.



While working with matrices imposes a lower bound on performance in the order of n^2, many improvements can be made by optimizing factors within the algorithms' recurrence.

In research, we plan to compare the different matrices to assess how the different types of distances might impact epidemiological analyses.

In the application of our results, we plan to use this system as a foundation for a new software, capable of defining the most convenient path, from a risk perspective, to connect different points of interest (e.g. contaminated animal byproducts from an outbreak to authorized disposal plants).

Thanks for your attention!
Any question?

Mirco Cazzaro

mcazzaro@izsvenezie.it

Research funded by: *Italian Health Ministry RC IZSVE 05/21 titled "Indagine sulla prevalenza e distribuzione di patogeni trasmessi da zecche in allevamenti caprini e valutazione del rischio legato alla presenza del virus della Tick Borne Encephalitis (TBEV) nel latte crudo e in prodotti derivati. IXORISK".*

Istituto Zooprofilattico Sperimentale delle Venezie

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

AGILE Association of Geographic Information Laboratories in Europe