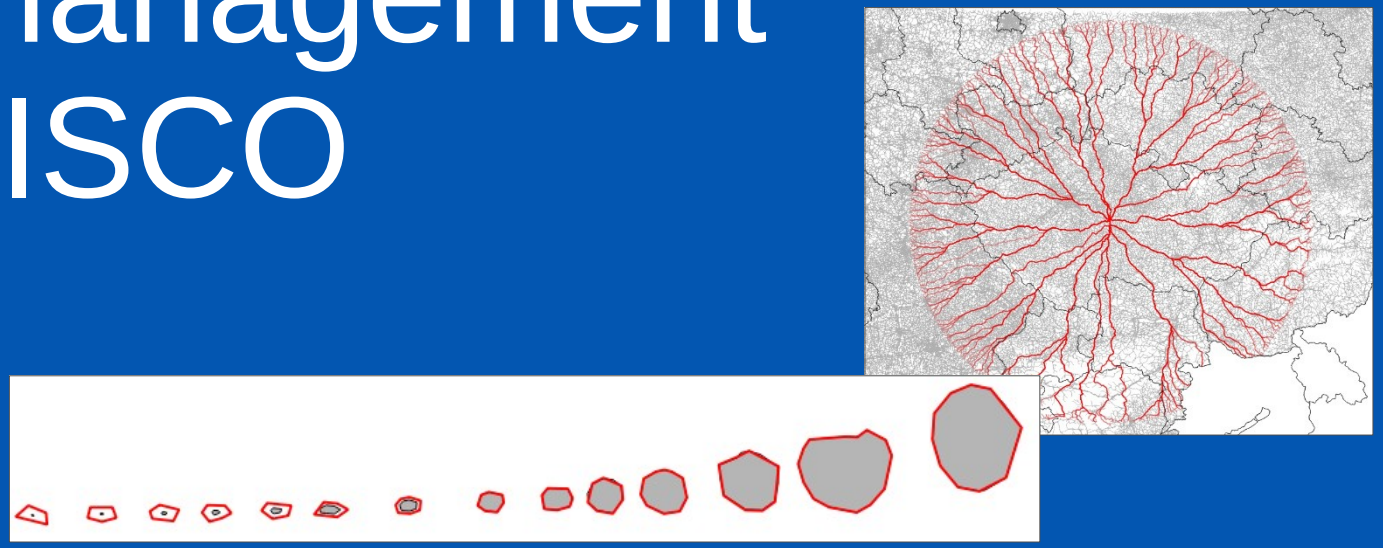
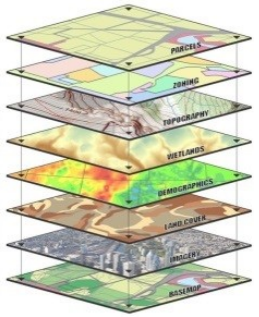
Some activities on geographical data quality management at Eurostat GISCO

Julien Gaffuri - Eurostat



4th International Workshop on Spatial Data Quality, 11-12 October 2023

GISCO – GIS at the Commission



GISCO is a permanent service of Eurostat that answers the common needs of Eurostat and the European Commission for geographical information at the level of the European Union (EU), its Member States and regions.

- Provision of GIS (reference) data, services and software,
- Support cartographic and spatial analysis activities,
- Stimulate the use of GIS to support commission activities,
- Support Eurostat activities on the integration of statistical and geospatial information.

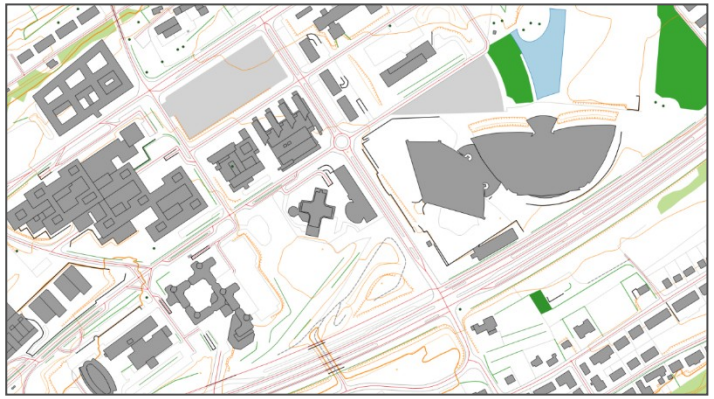
- https://ec.europa.eu/eurostat/web/gisco

Outline

1. Quality requirements
2. Quality control
3. Quality influence on spatial analyse

Quality requirements

Quality requirements



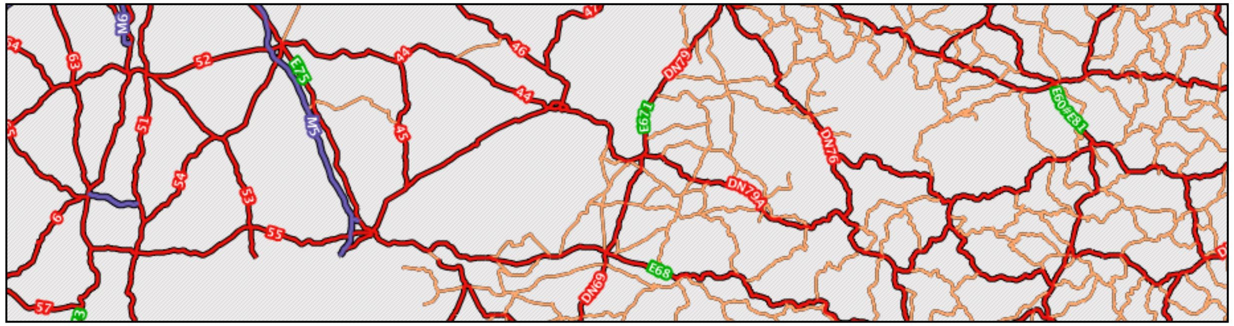
BD-L-TC, ACT Luxembou

- Thematic coverage: Base topographic datasets on buildings, ground infrastructure, transport networks, land use and cover, hydrography, orography, administrative boundaries, geographical names, POIs, etc.
- Requirements for completeness, positional accuracy, thematic accuracy, temporal quality (timeliness and update frequency, versioning with persistent identifiers).
- Call for tenders **ESTAT/2022/NP/0010-GISCO acquisition of topographic data layers**.

- EU context: Specific quality requirements

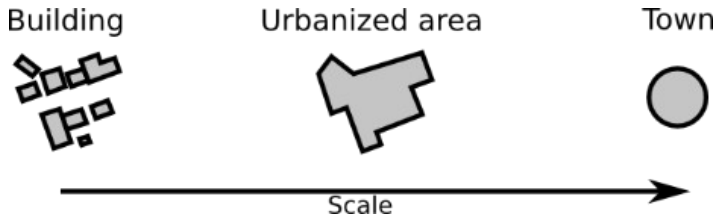
Quality requirements – Specific

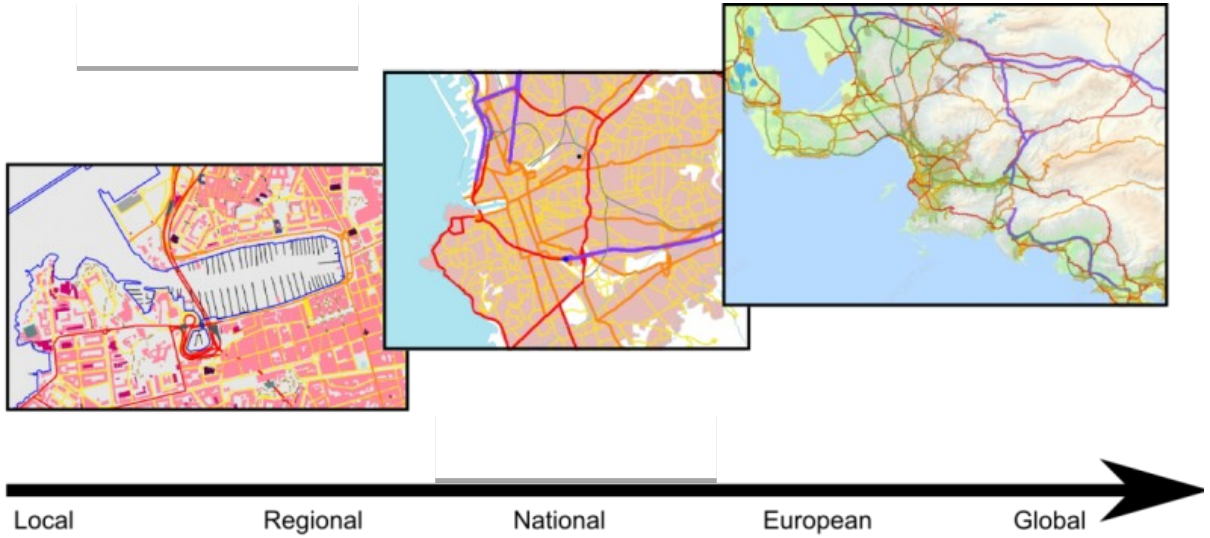
- Geographical extent, comparability
  - Several countries, cross-border regions or the entire European territories
  - Homogeneity of the quality across space
  - Topological consistency at border (with edge-matching)



Quality requirements – Specific

- Need for detailed data (1:10k) **and** generalised data (1:50k, 1:100k, 1:250)
- Need for **multi-scale** data – derived with AI-based automated generalisation.





Quality requirements – Specific

- Meta-quality
  - Quality must be known, measured, documented.
  - Stability of quality across versions.
- Sustainability – need for governance and resources.

Quality control

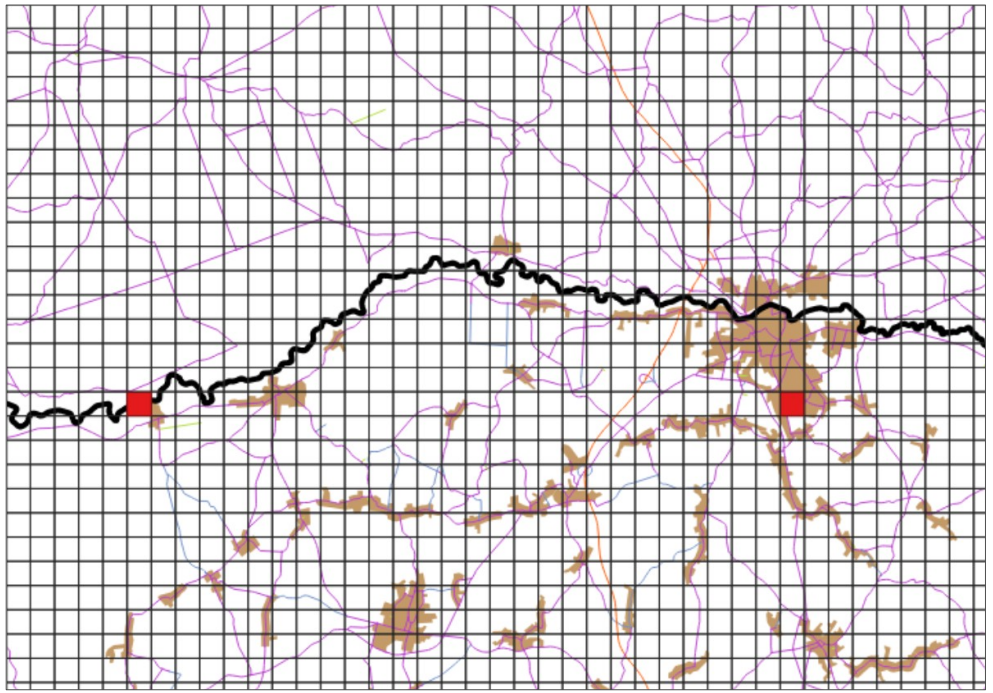
European Commission

Quality control at Eurostat-GISCO

- Structure and geometry validity
- Completeness
- Topological consistency – noding and edge-matching
- Timeliness
- Generalisation

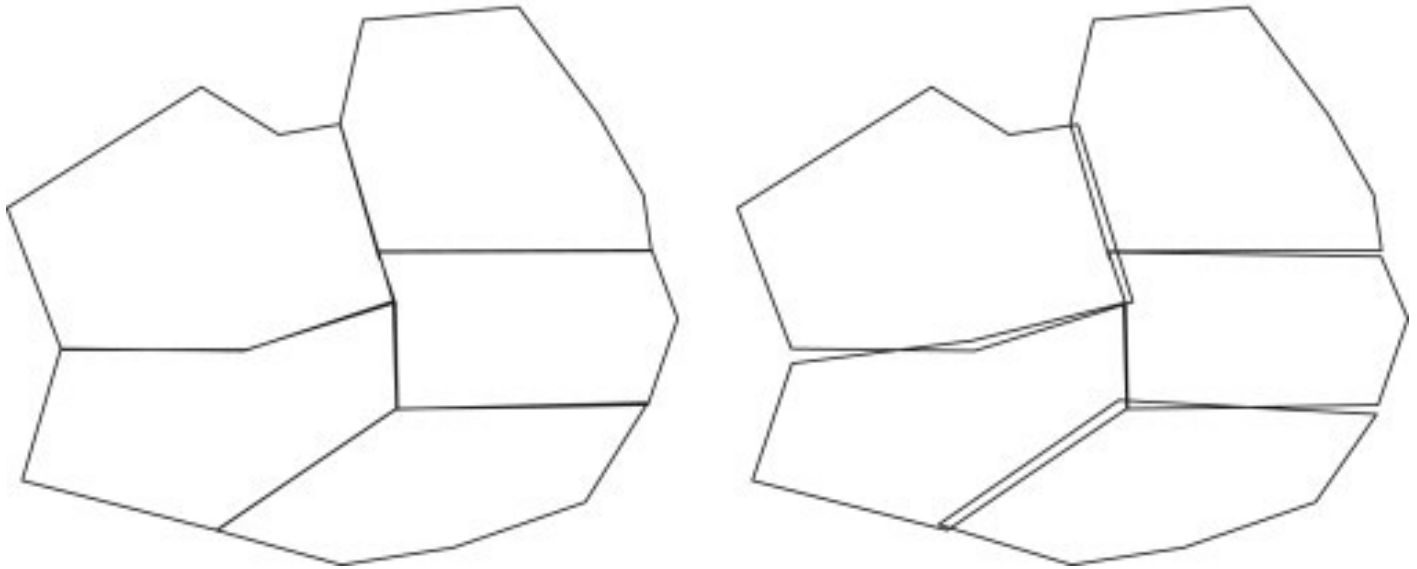
Quality control – Completeness

- Assess the number of omissions and commissions for some feature classes.
- Pseudo-random sampling – of 1km grid cells:
  - 30 grid cells (one per country)
  - 17 cross-country grid cells
- For each grid cell:
  - Cross-source comparison



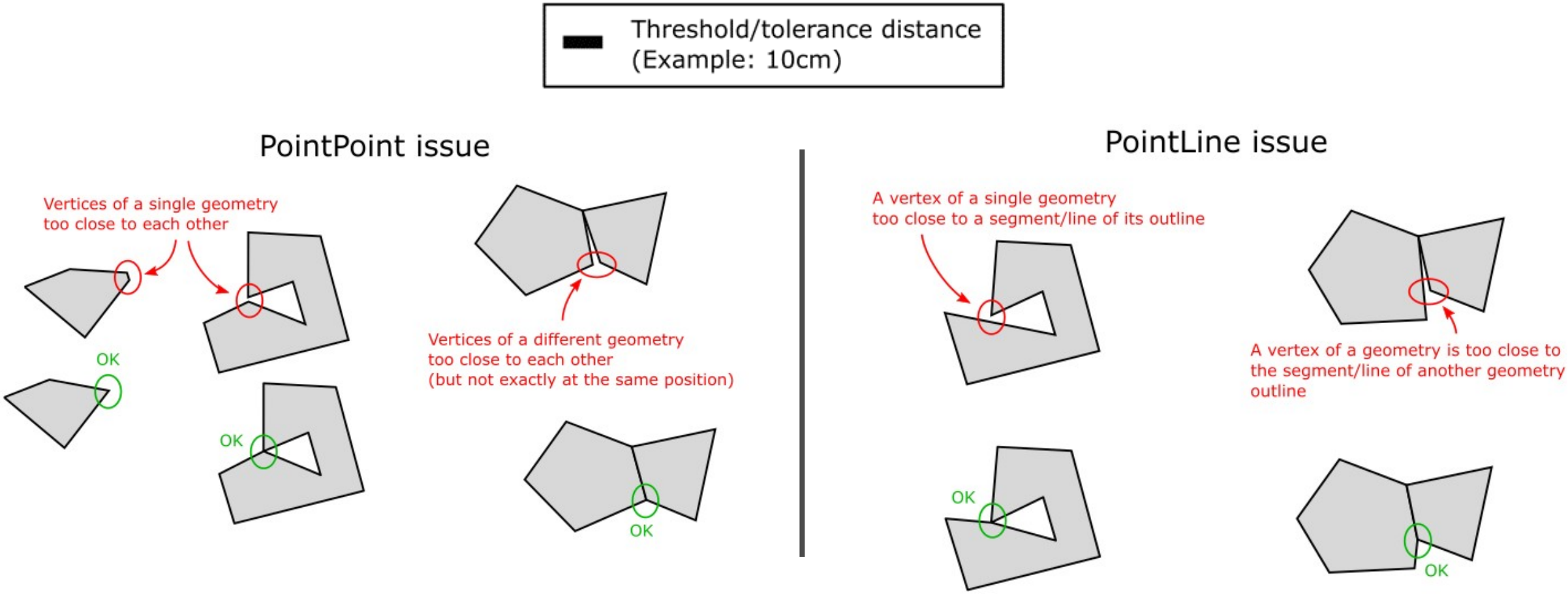
Quality control – Topology – Tessellations

- Topology validation: No gap – no overlap



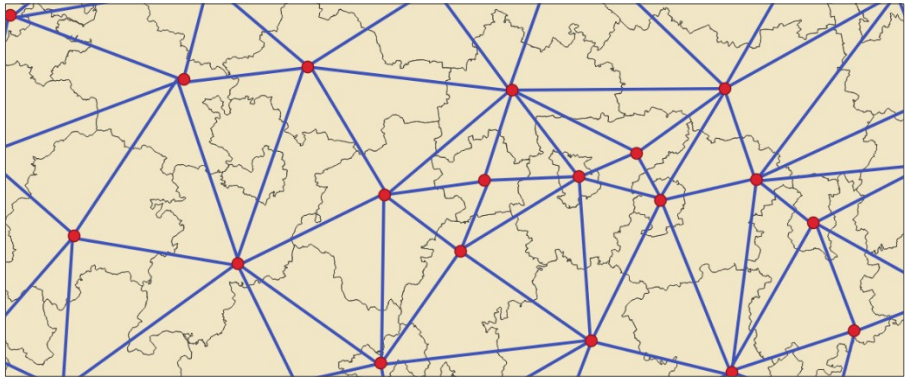
Quality control – Topology – Tessellations

- Topology validation: Strict noding



Quality control – Topology – Road network

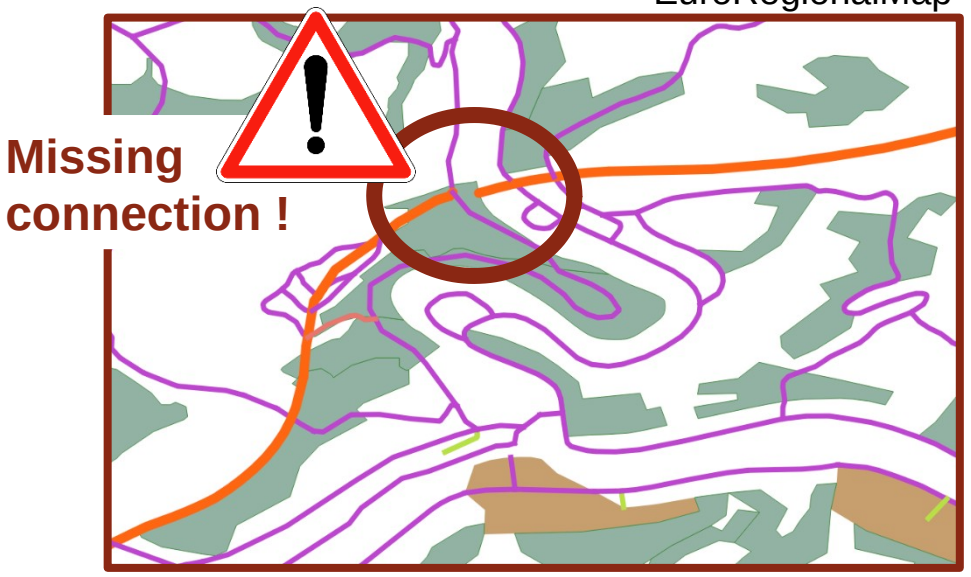
- Origin-distance matrix on a sample of 2053 locations.
- Driving time for Delaunay triangulation segments.



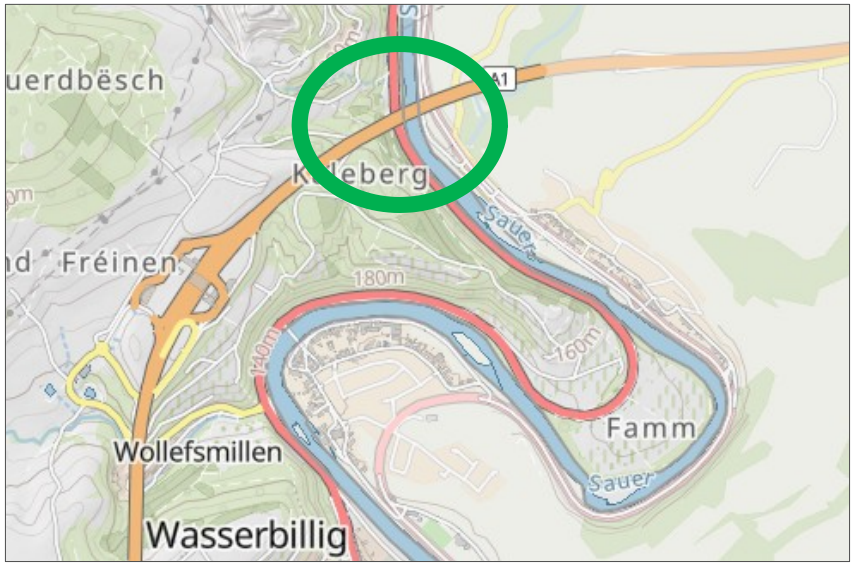
- Comparison of results based on different datasets and versions of them.
- Outlier detection
- Give insight on: Edge-matching issue, thematic accuracy, completeness.

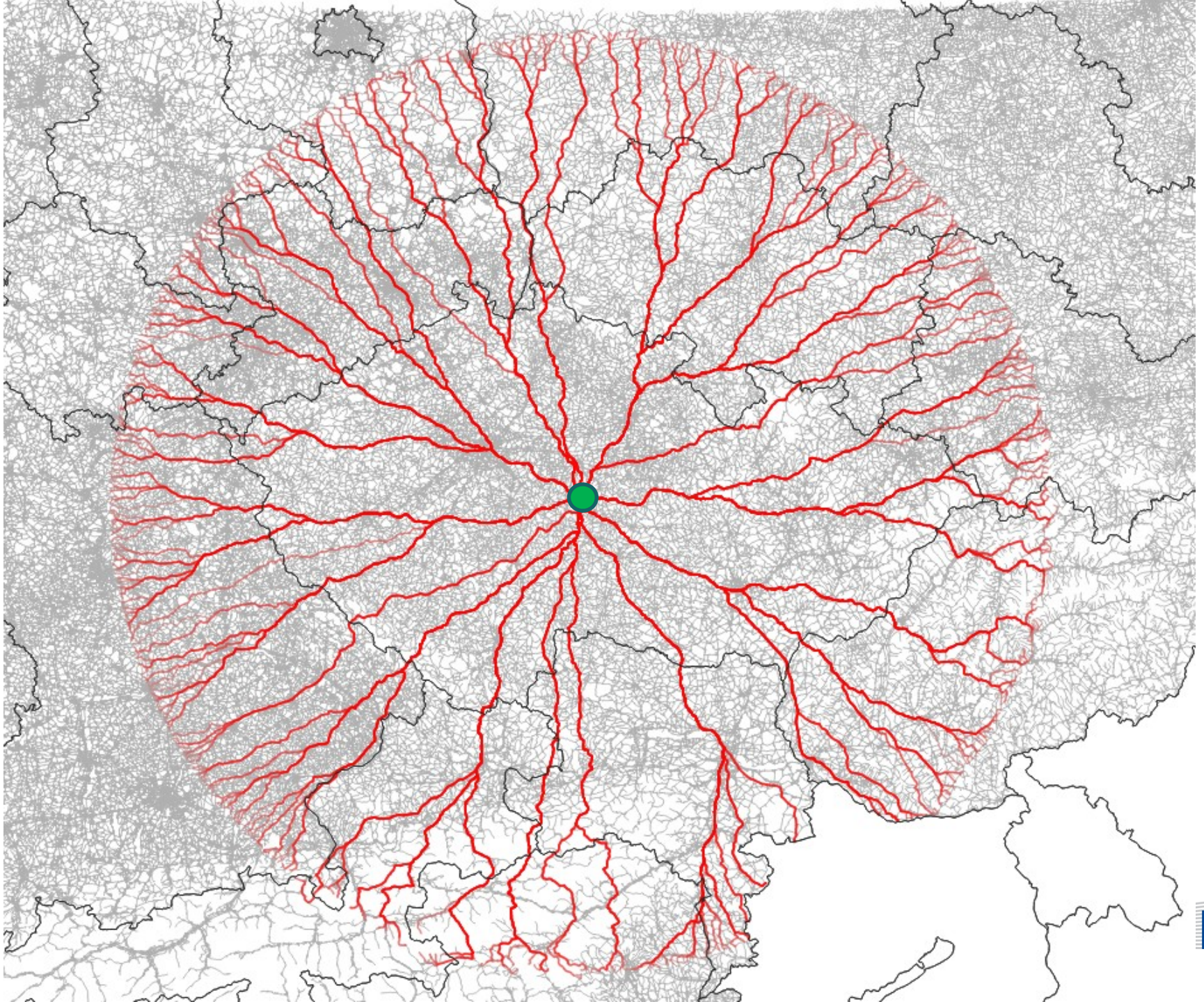
Quality control – topology – Road network

EuroRegionalMap



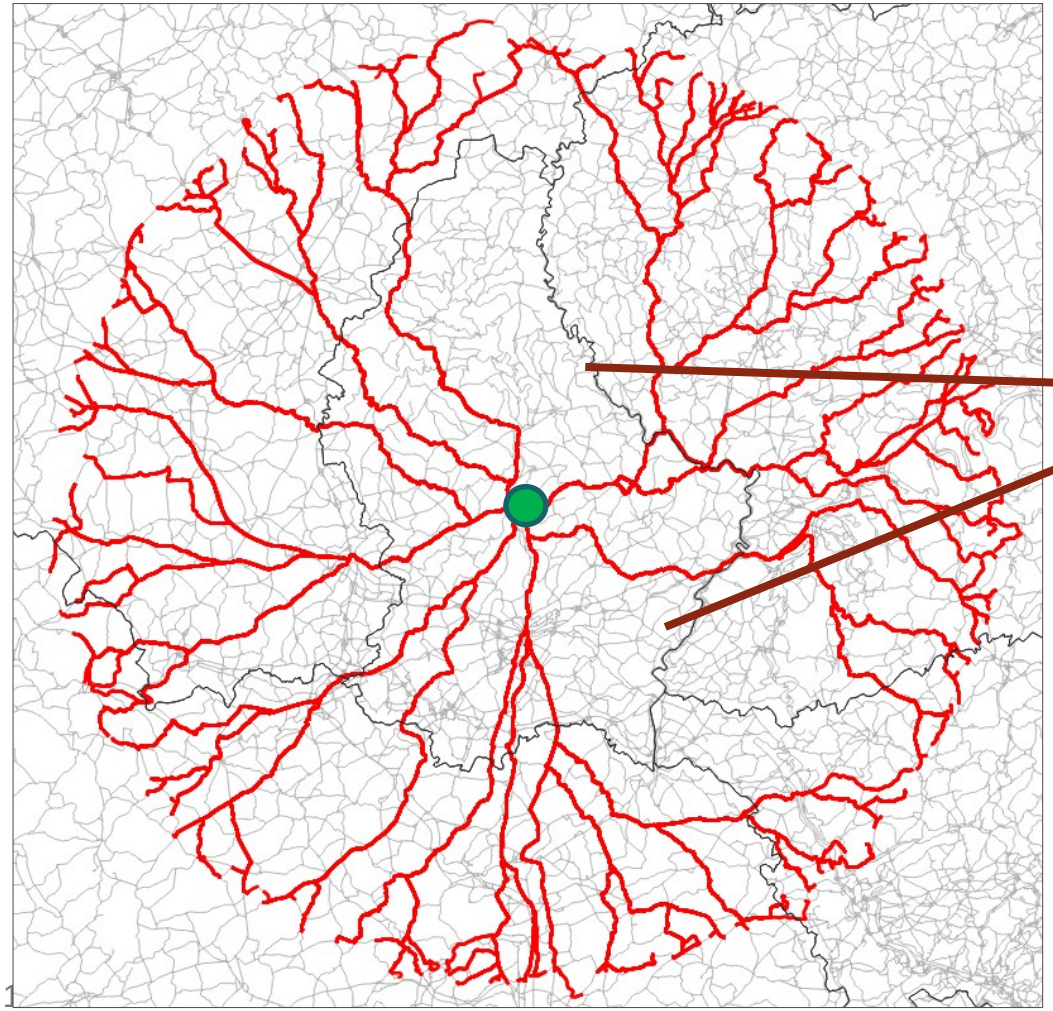
BD-L-TC, ACT Luxembourg





Quality control – topology – Road network

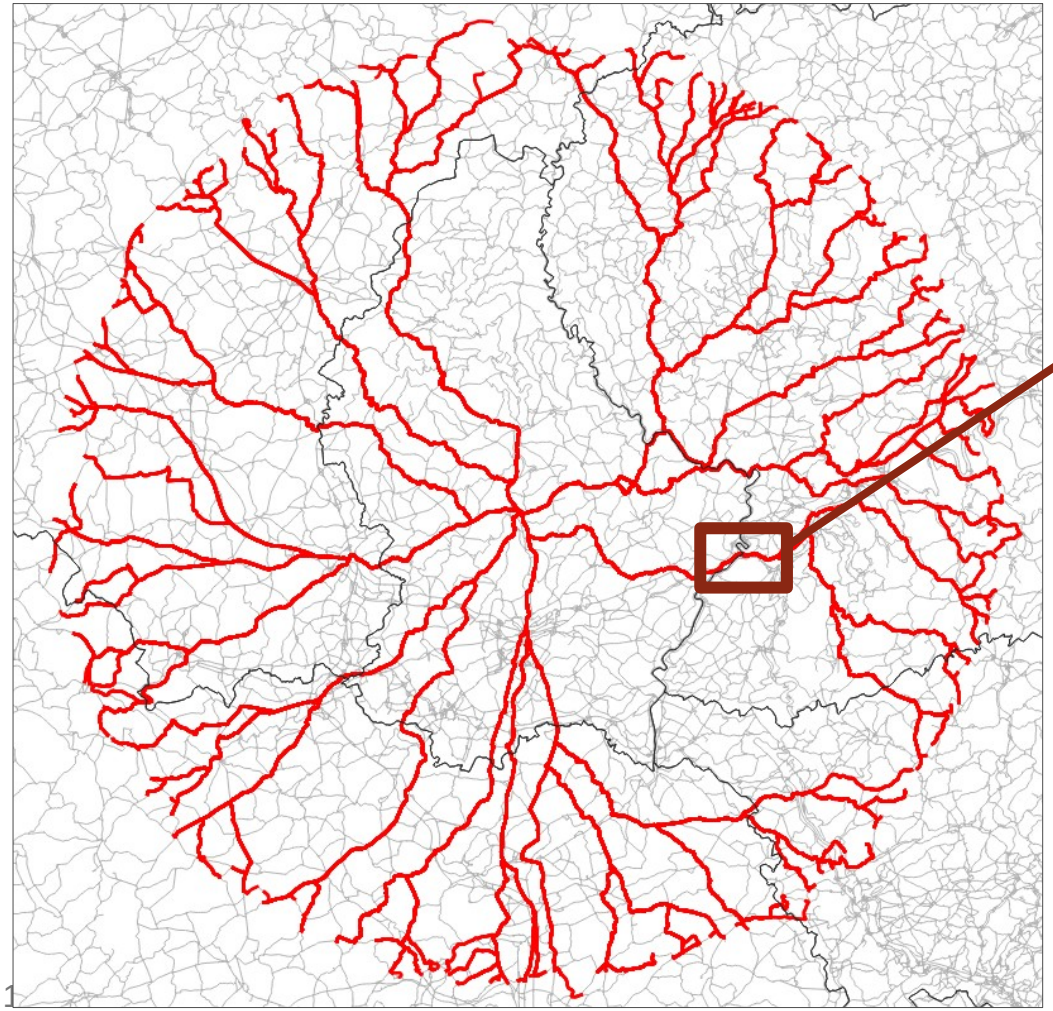
EuroRegionalMap



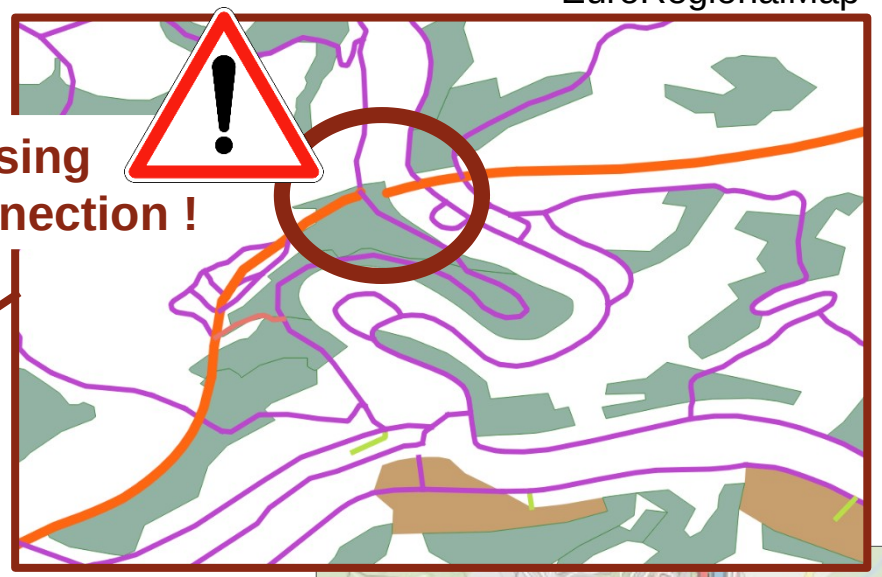
Missing cross-border connections !

Quality control – topology – Road network

EuroRegionalMap

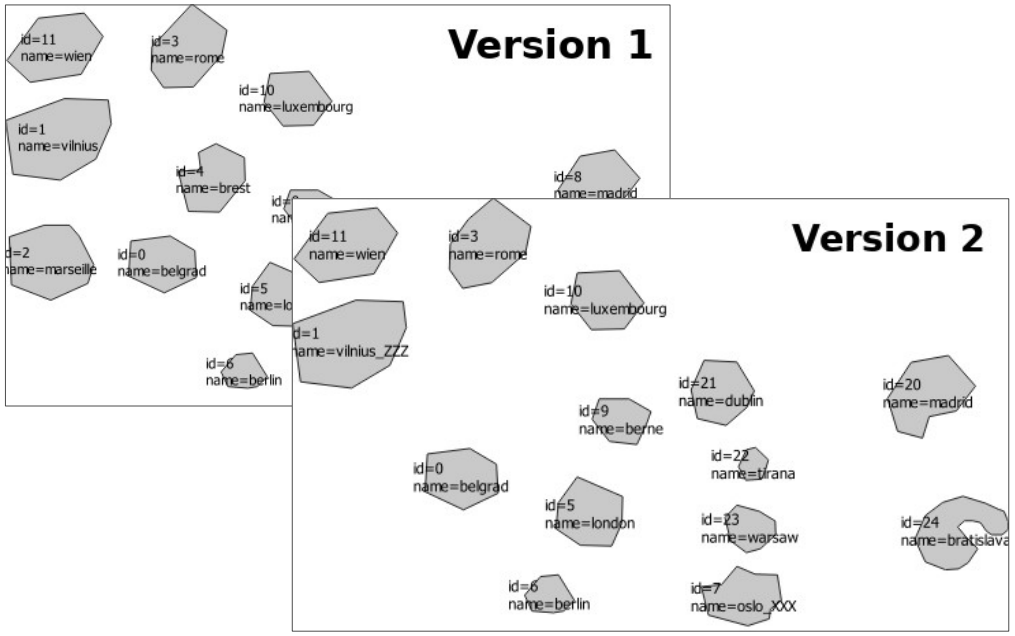


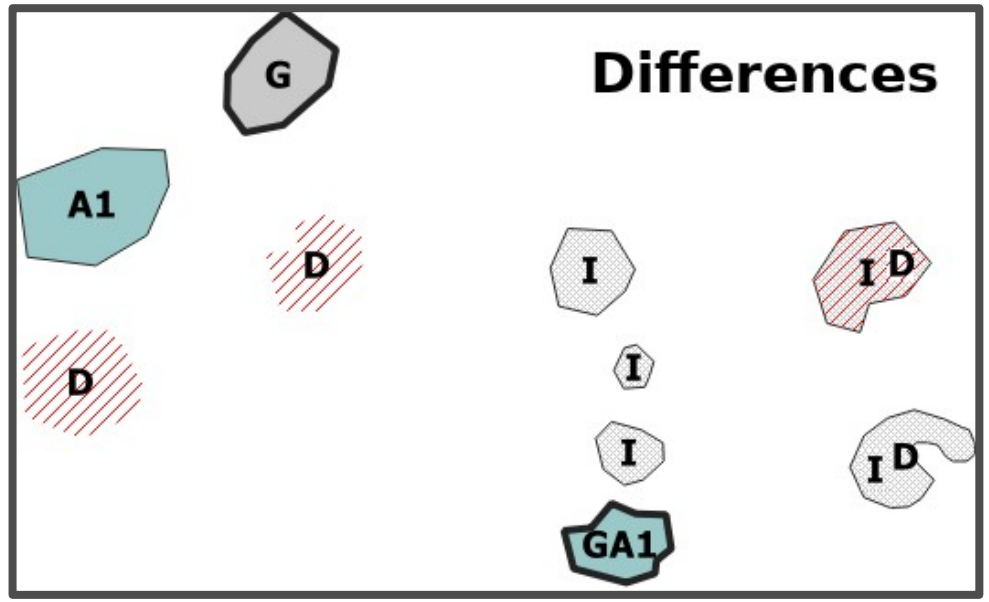
EuroRegionalMap



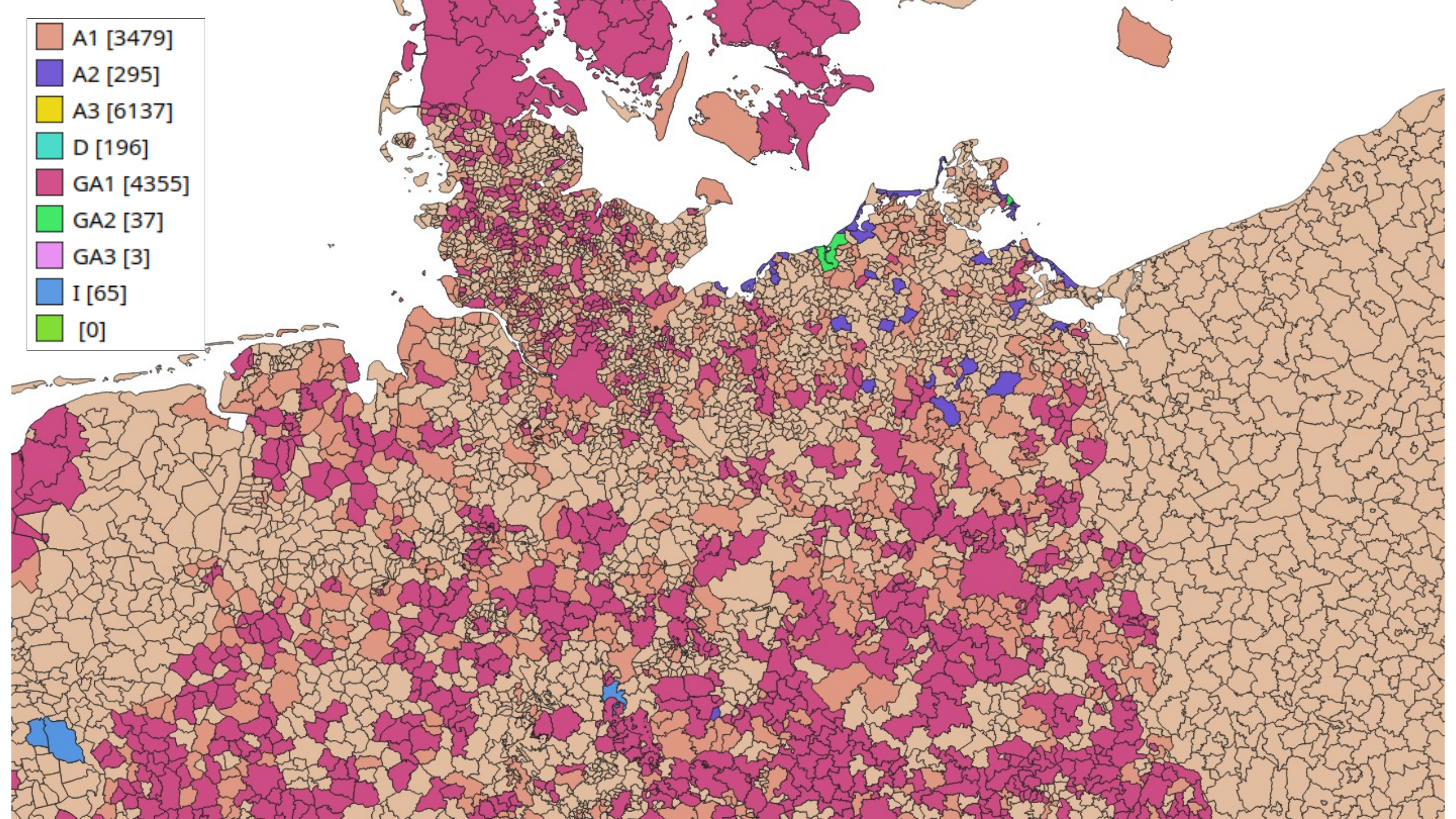
Quality control – Timeliness

- Compare two consecutive versions of a dataset – analyse the changes.
- Based on **GeoDiff** tool (https://github.com/eurostat/GeoDiff)
- Check identifier stability



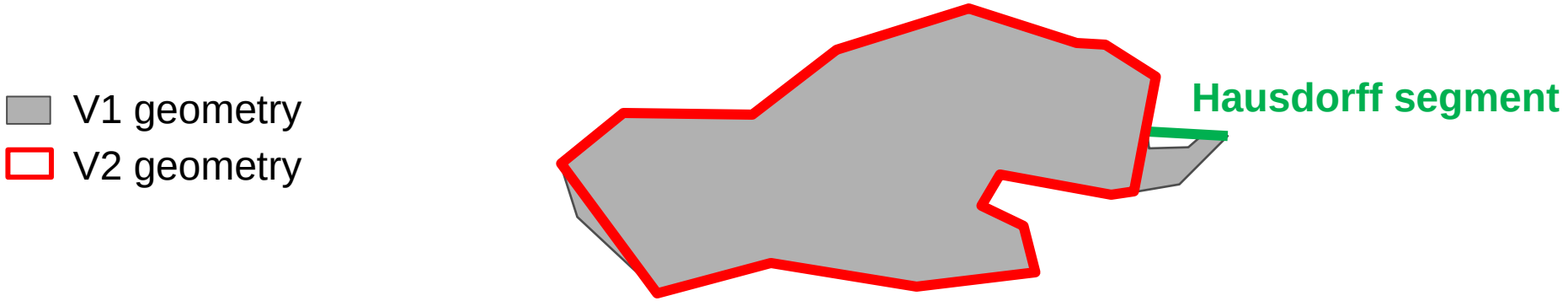


European Commission

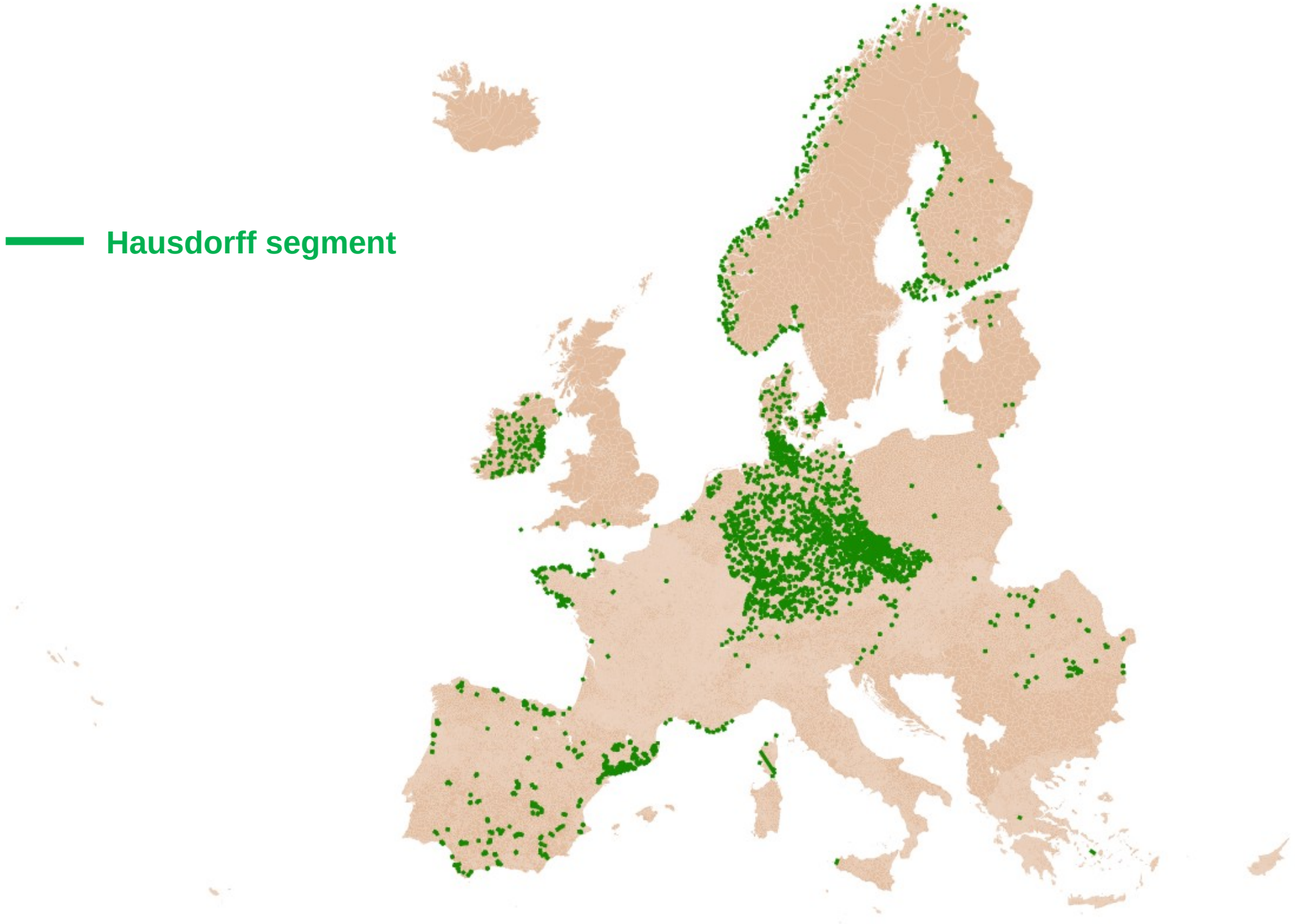

A1 [3479]
A2 [295]
A3 [6137]
D [196]
GA1 [4355]
GA2 [37]
GA3 [3]
I [65]
[0]

Quality control – Timeliness

- Analyse geometrical changes, based on Hausdorff distance.



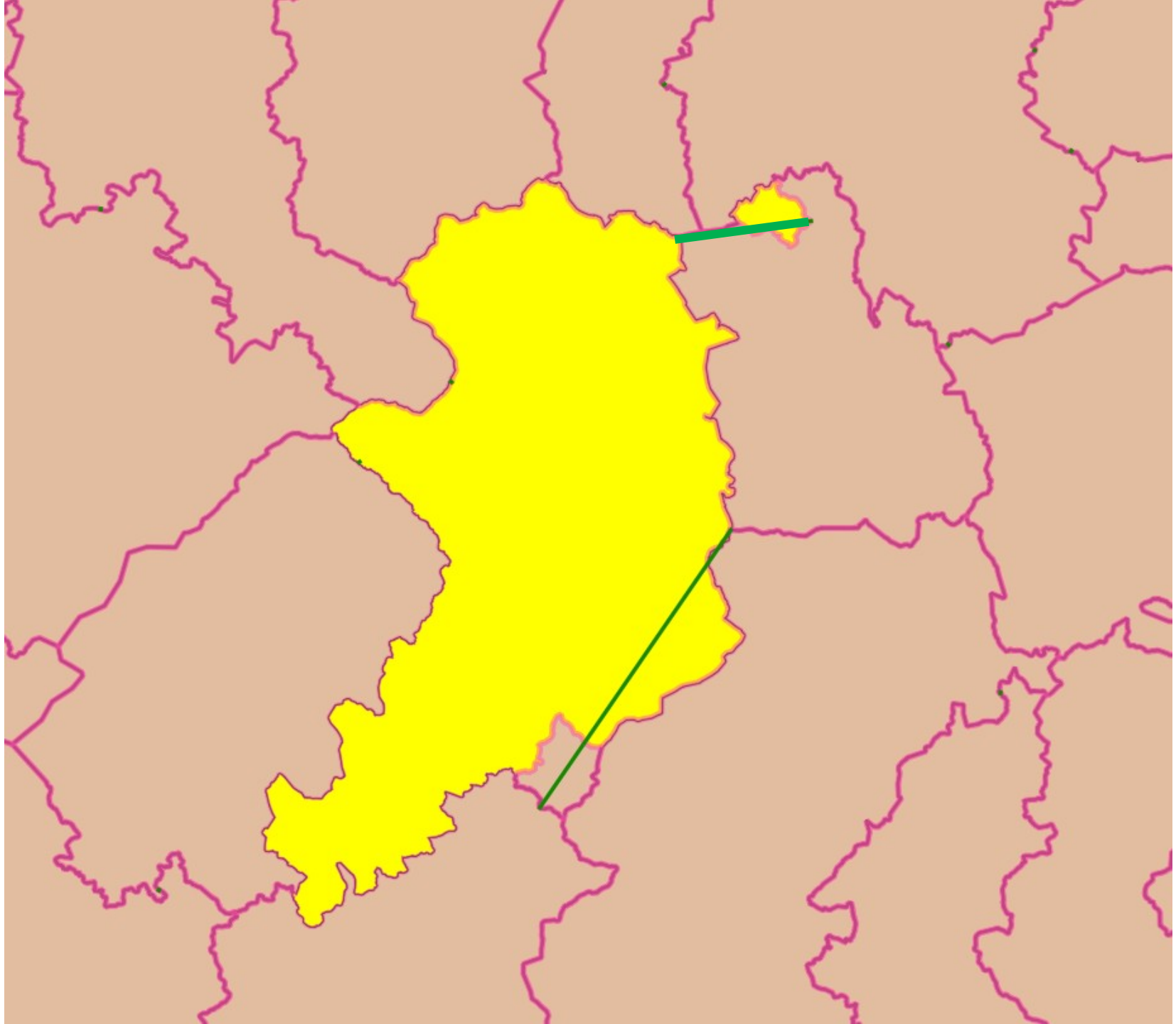
Hausdorff segment



Hausdorff segment

Version 1

Version 2

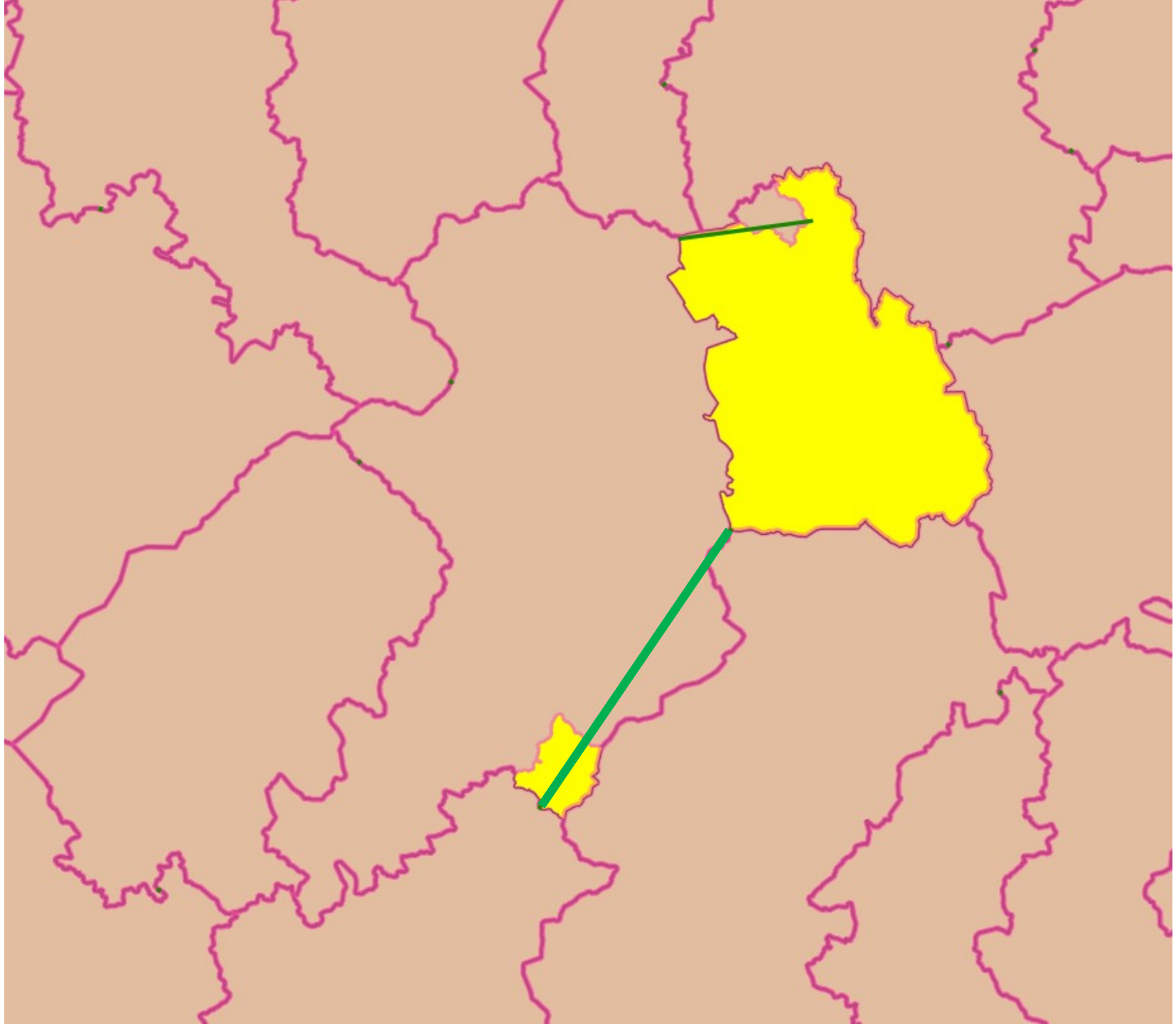


European Commission

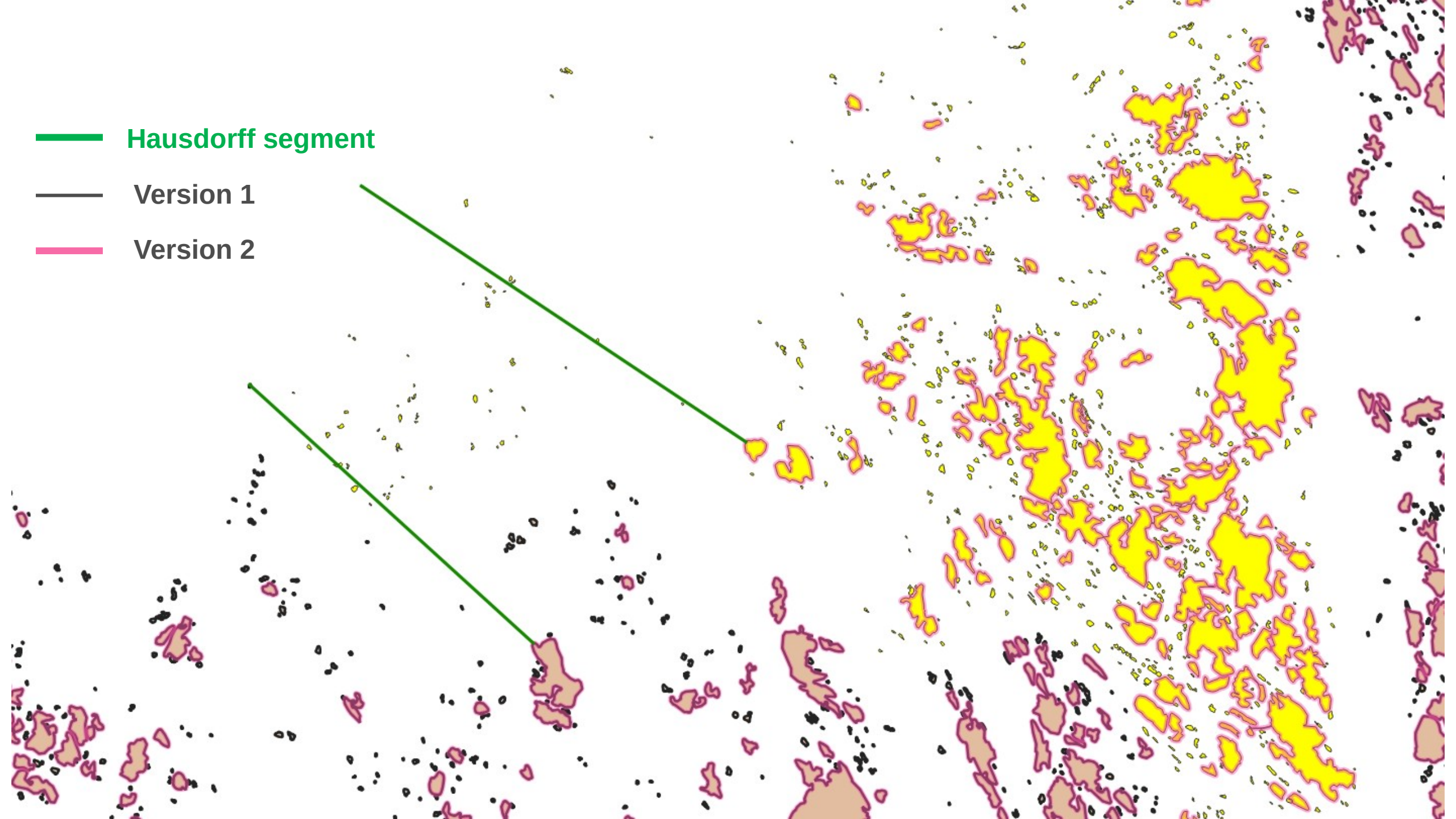
Hausdorff segment

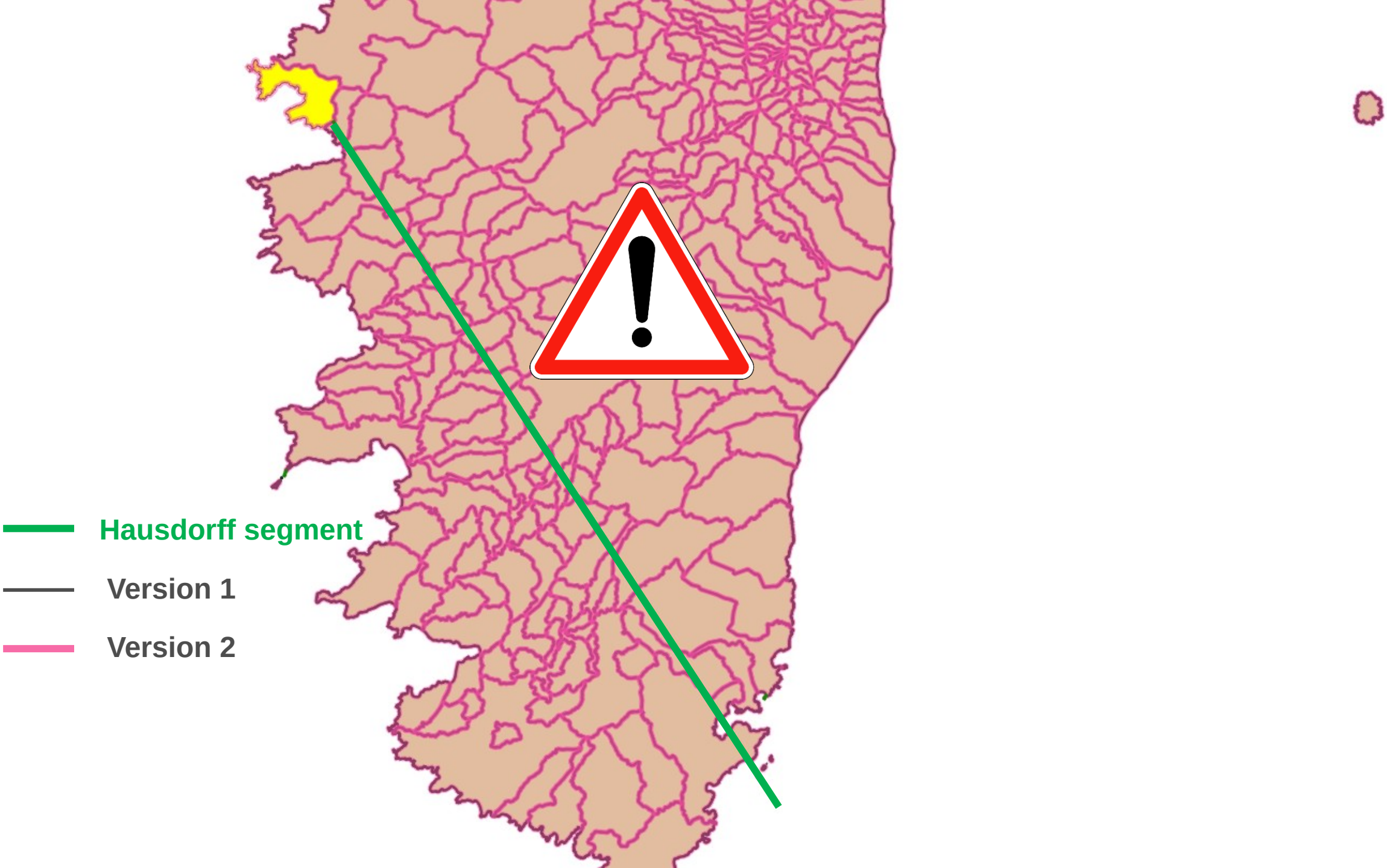
Version 1

Version 2



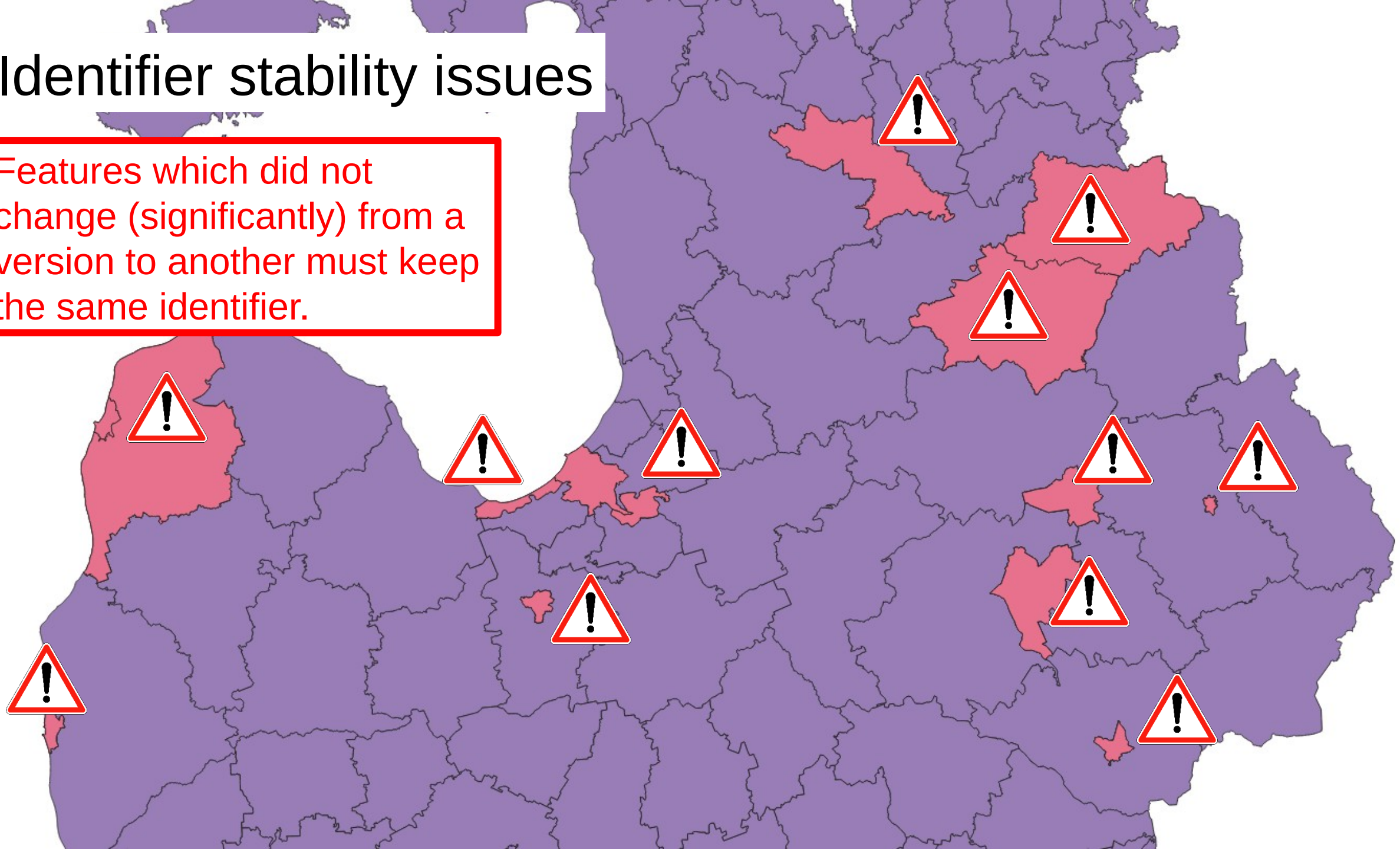
European Commission


Hausdorff segment
Version 1
Version 2


Hausdorff segment
Version 1
Version 2

Identifier stability issues

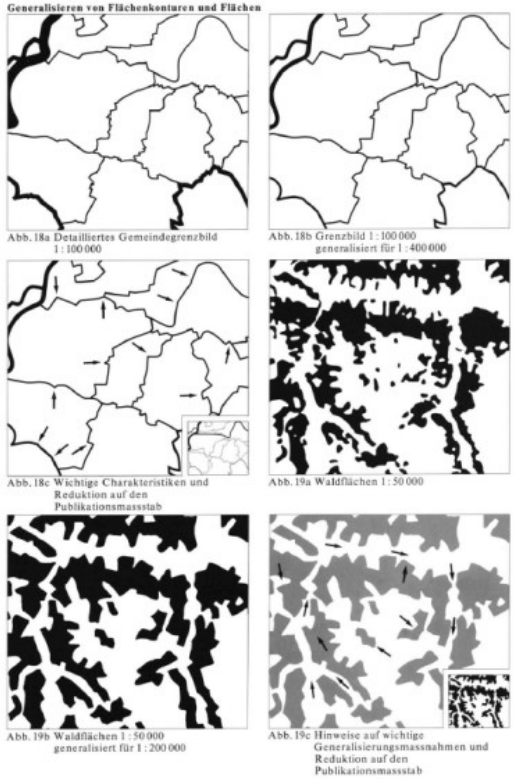
Features which did not change (significantly) from a version to another must keep the same identifier.

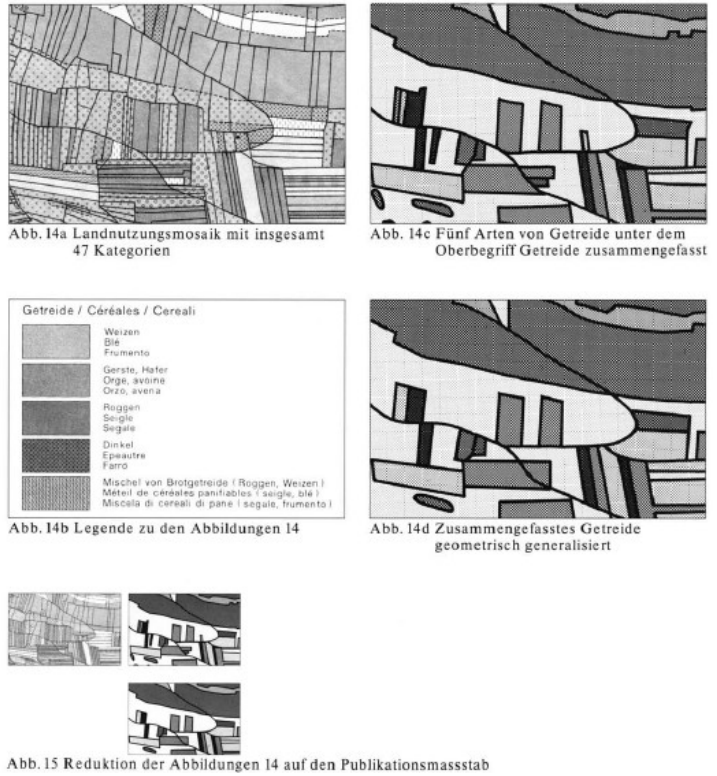


Quality control – Generalisation

- Geometric level of detail: Validation for administrative units.
- Minimum mapping unit / distance

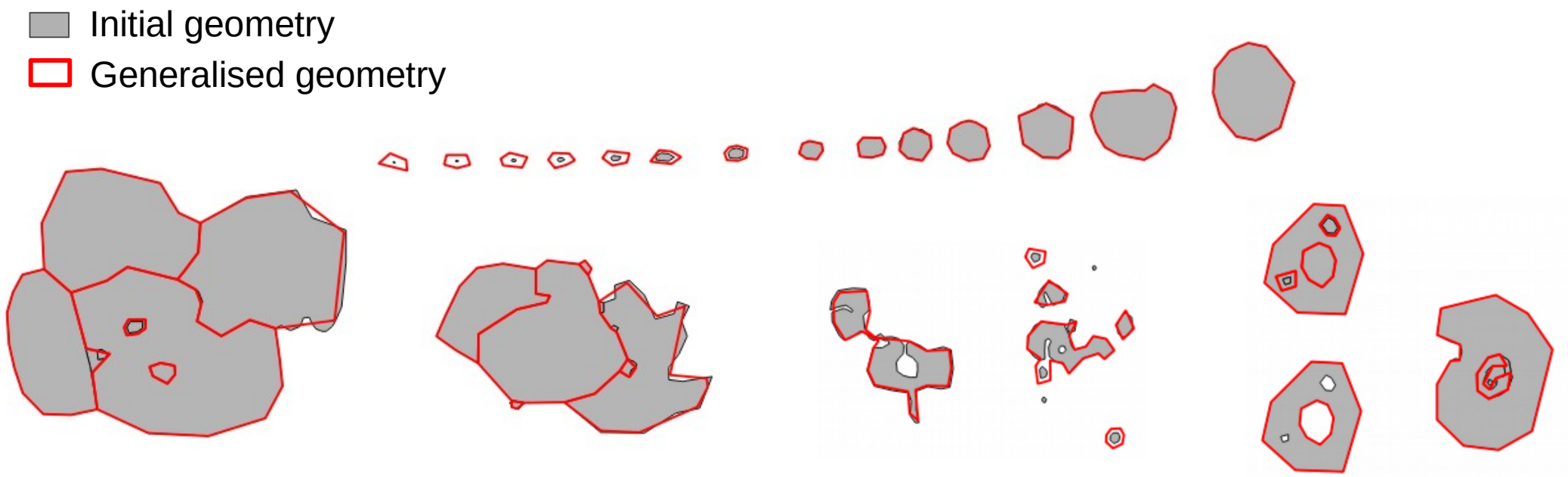
Scale	Resolution in map mm	Resolution in ground meter
1:1M	0.2mm	200m
1:3M	0.2mm	600m
1:10M	0.2mm	2km
1:20M	0.2mm	4km
1:60M	0.2mm	12km





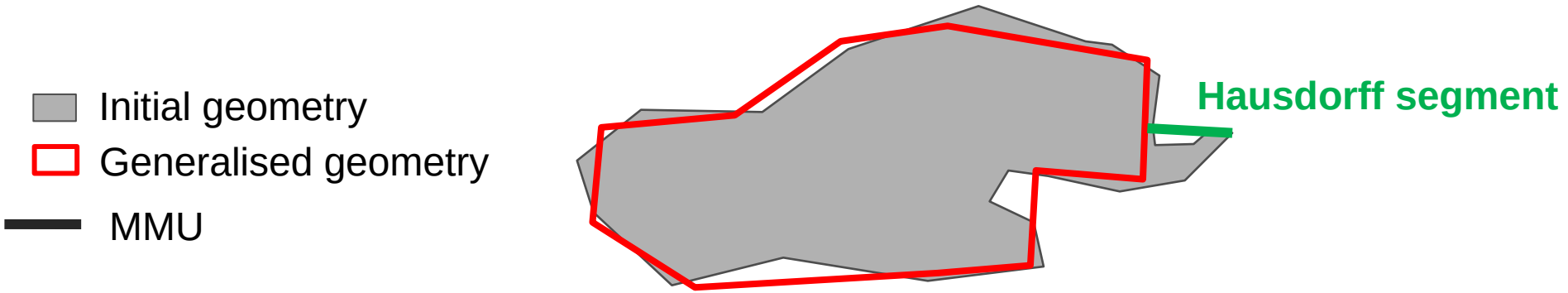
Quality control – Generalisation

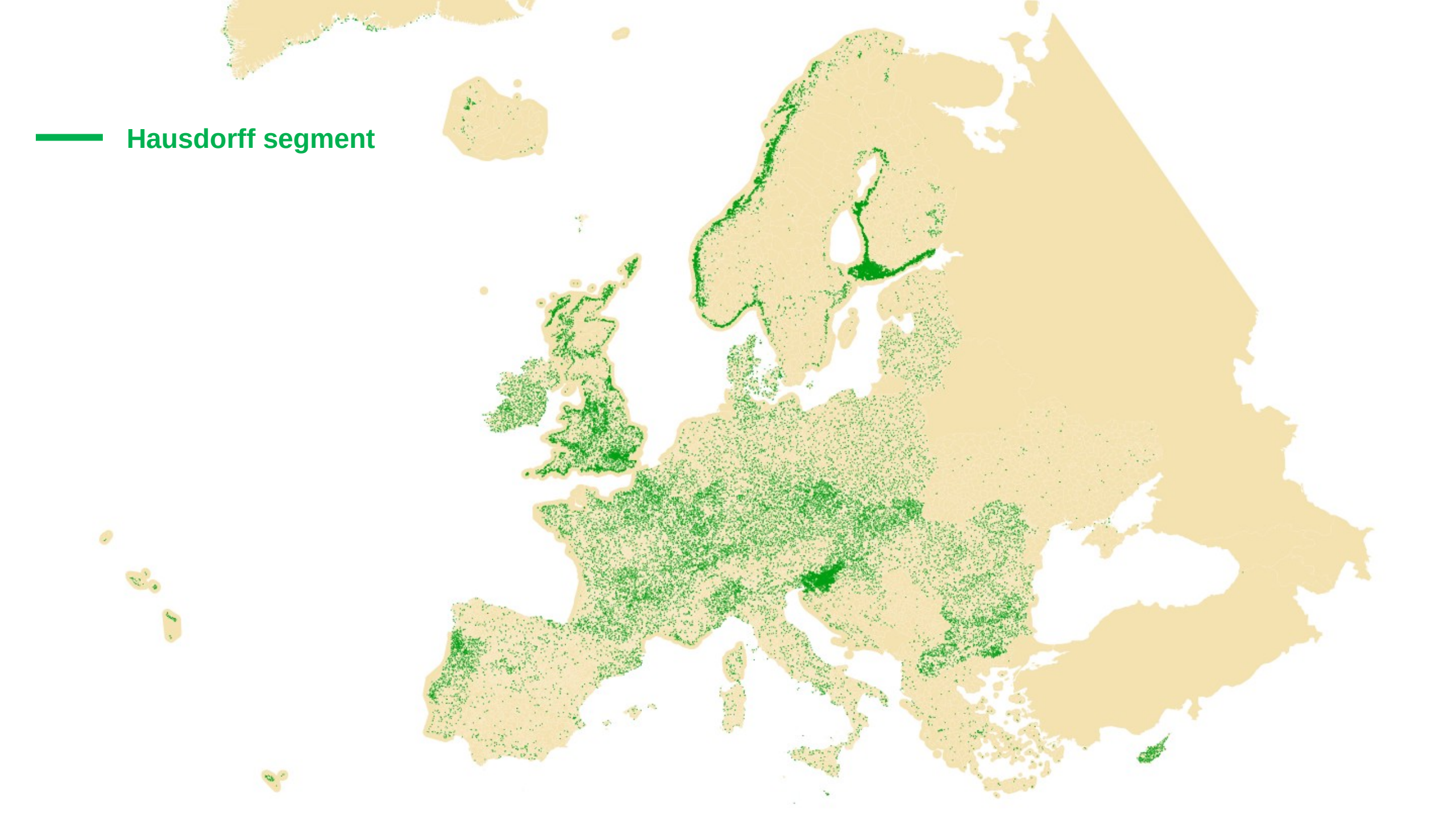
- Geometric level of detail: Validation for administrative units.
- Based on **RegionSimplify** tool (https://github.com/eurostat/RegionSimplify/)



Quality control – Generalisation

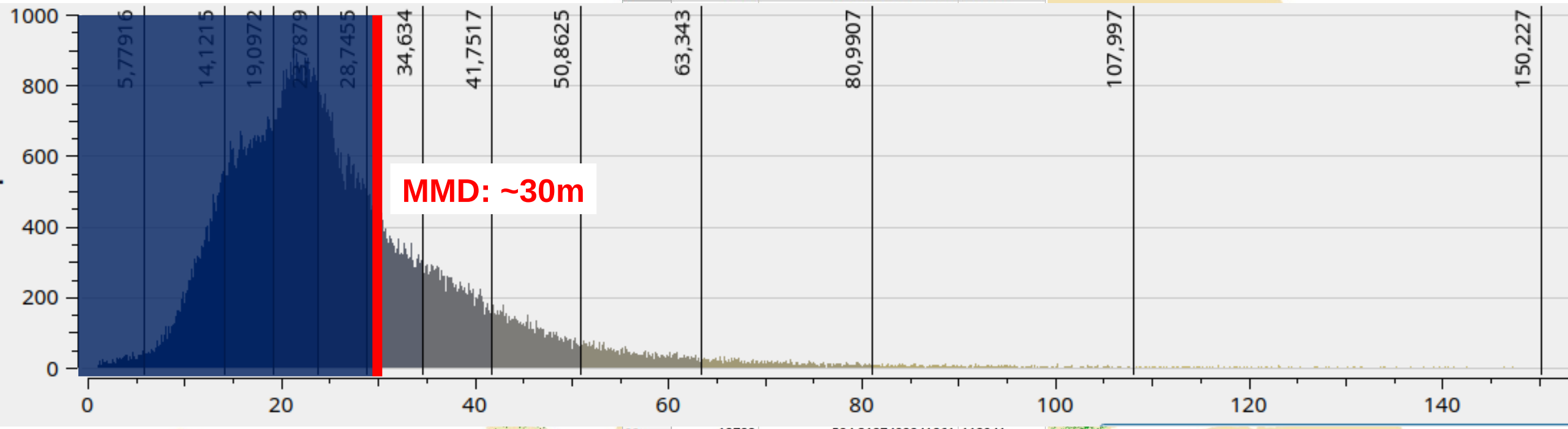
- Geometric level of detail: Validation for administrative units.
- Process:
  - Run **RegionSimplify** on the dataset to be validated.
  - Compare the outcome with the dataset to be validated, with **GeoDiff**, based on Hausdorff distance.



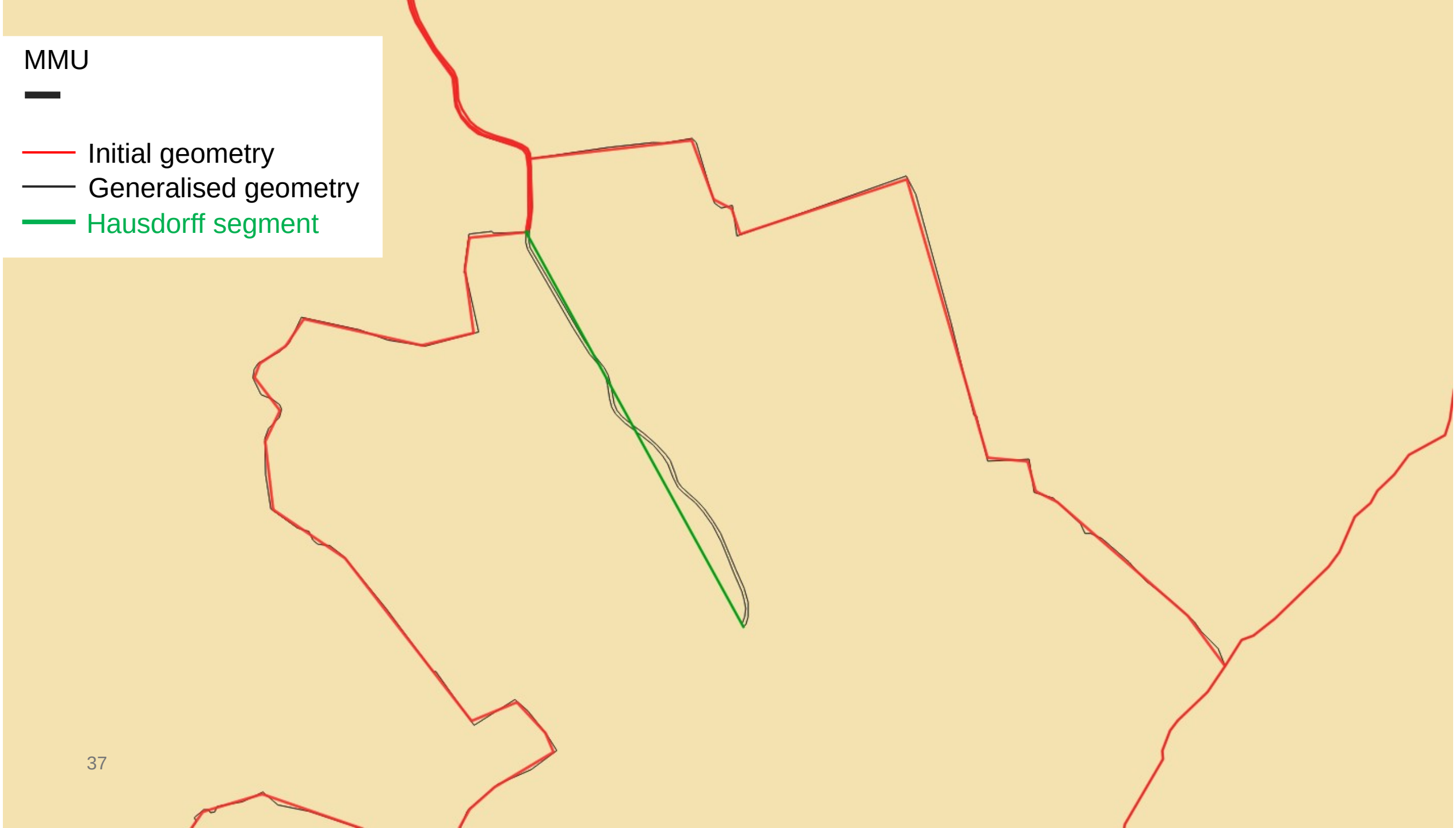

Hausdorff segment

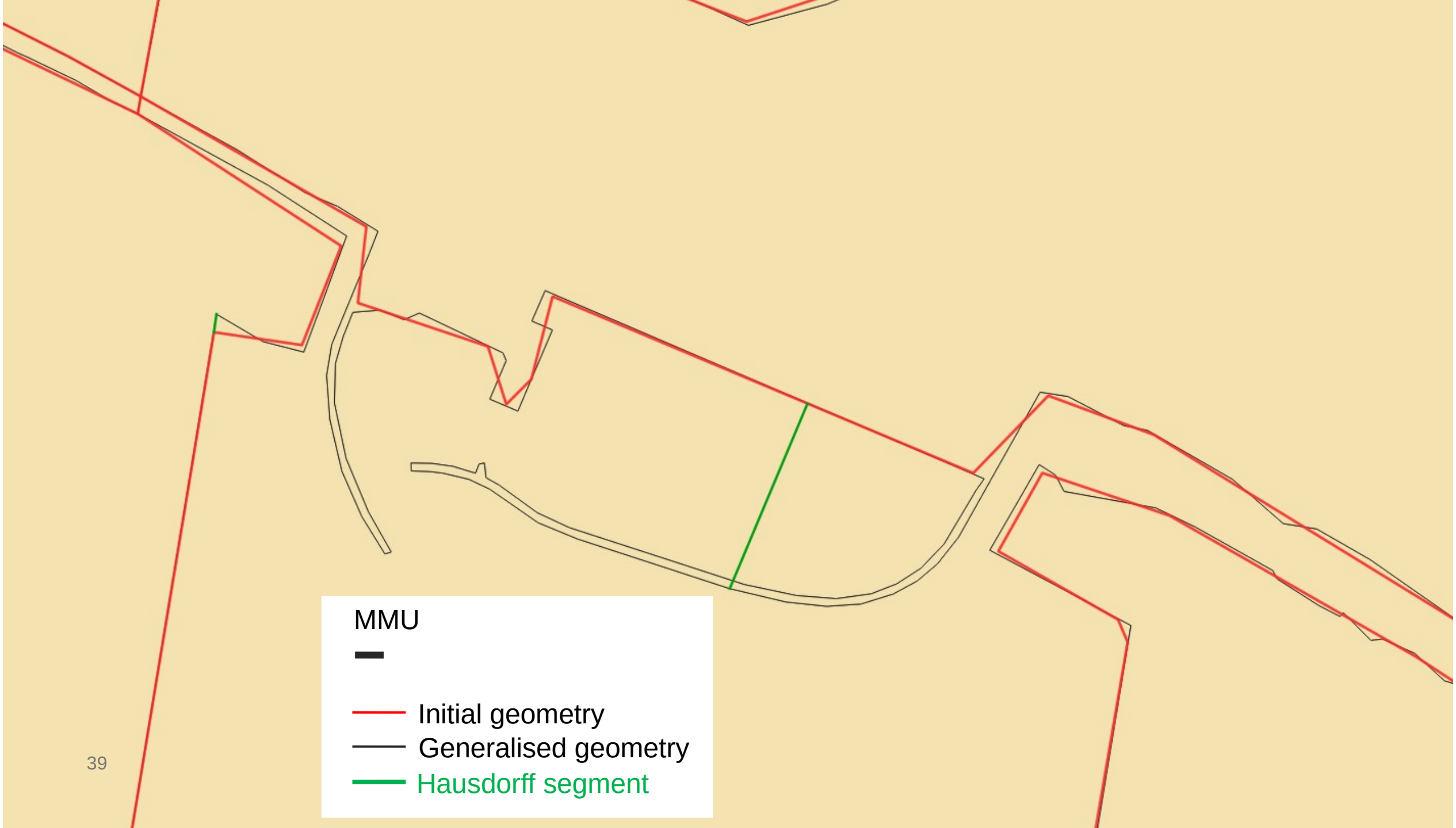
Hausdorff segment

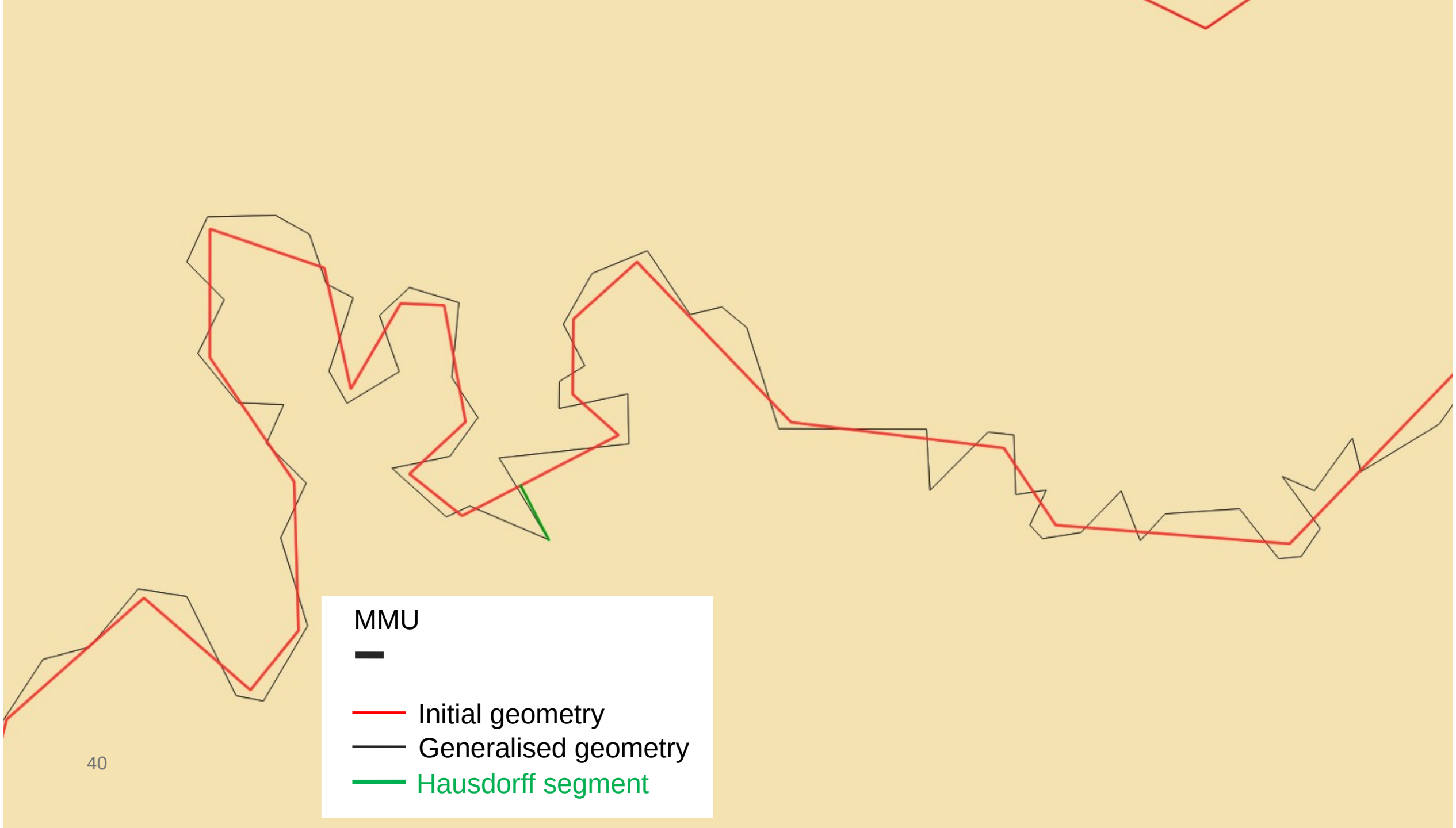
	fid	hdist	id
1	9044	1892,6123010229633	109042
2	2121	1402,8158542999822	102377
3	4951	1060,3006586555134	105047
4	3316	892,2731983726761	103487
5	4550	866,1210098743366	104648
6	13157	795,8259960238428	113294

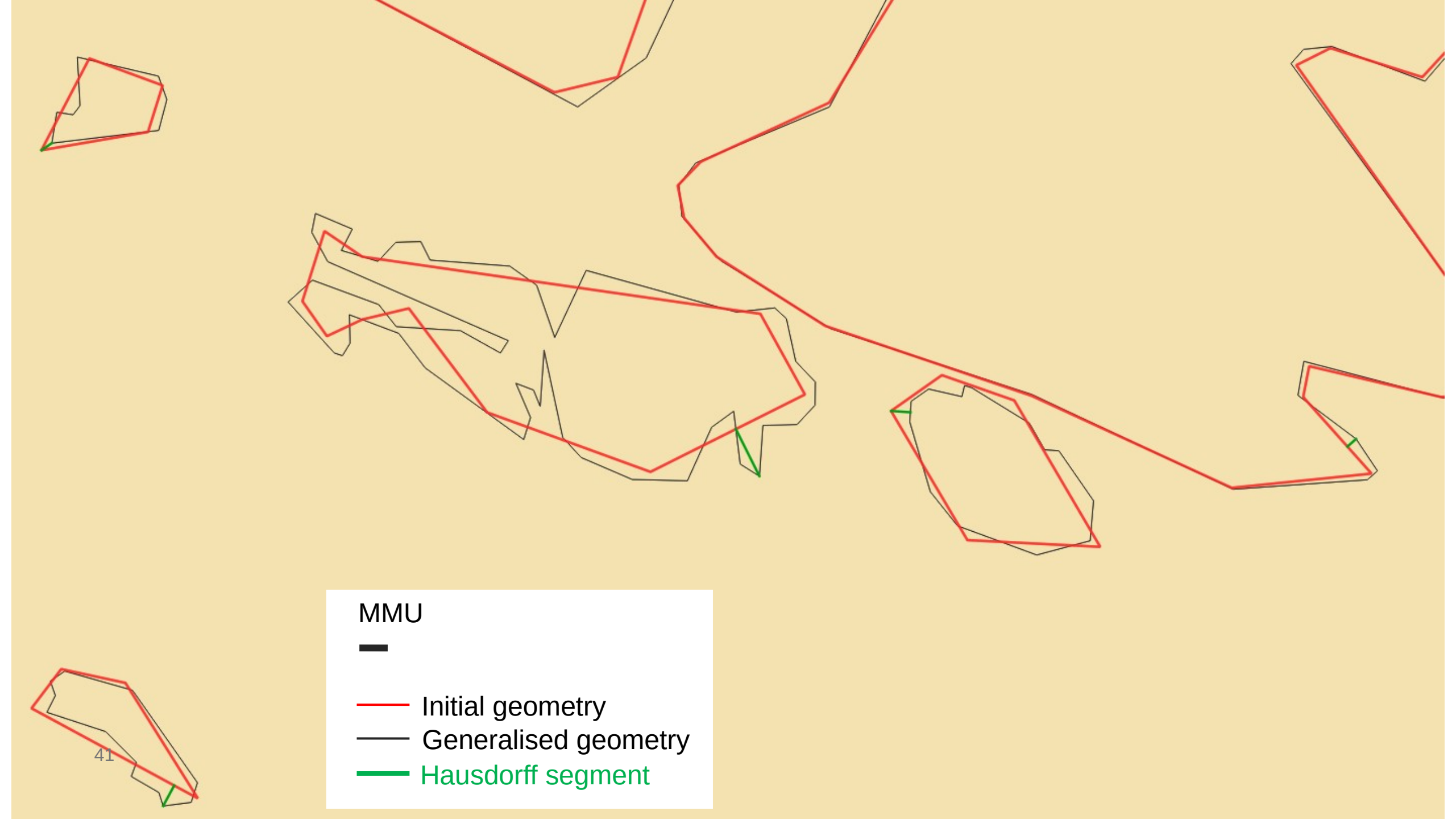


23	12788	524,3127492241861	112941
24	2482	524,1883964872062	102713
25	7206	521,6452775909871	107218
26	5024	521,3755933575284	105128
27	13806	517,4104045124304	113935
28	13217	517,1339594360225	113353
29	93575	506,14852000075825	5
30	1545	501,5832131373726	101832
31	18081	489,5032129636285	122696


MMU
Initial geometry
Generalised geometry
Hausdorff segment


MMU
Initial geometry
Generalised geometry
Hausdorff segment


MMU
Initial geometry
Generalised geometry
Hausdorff segment

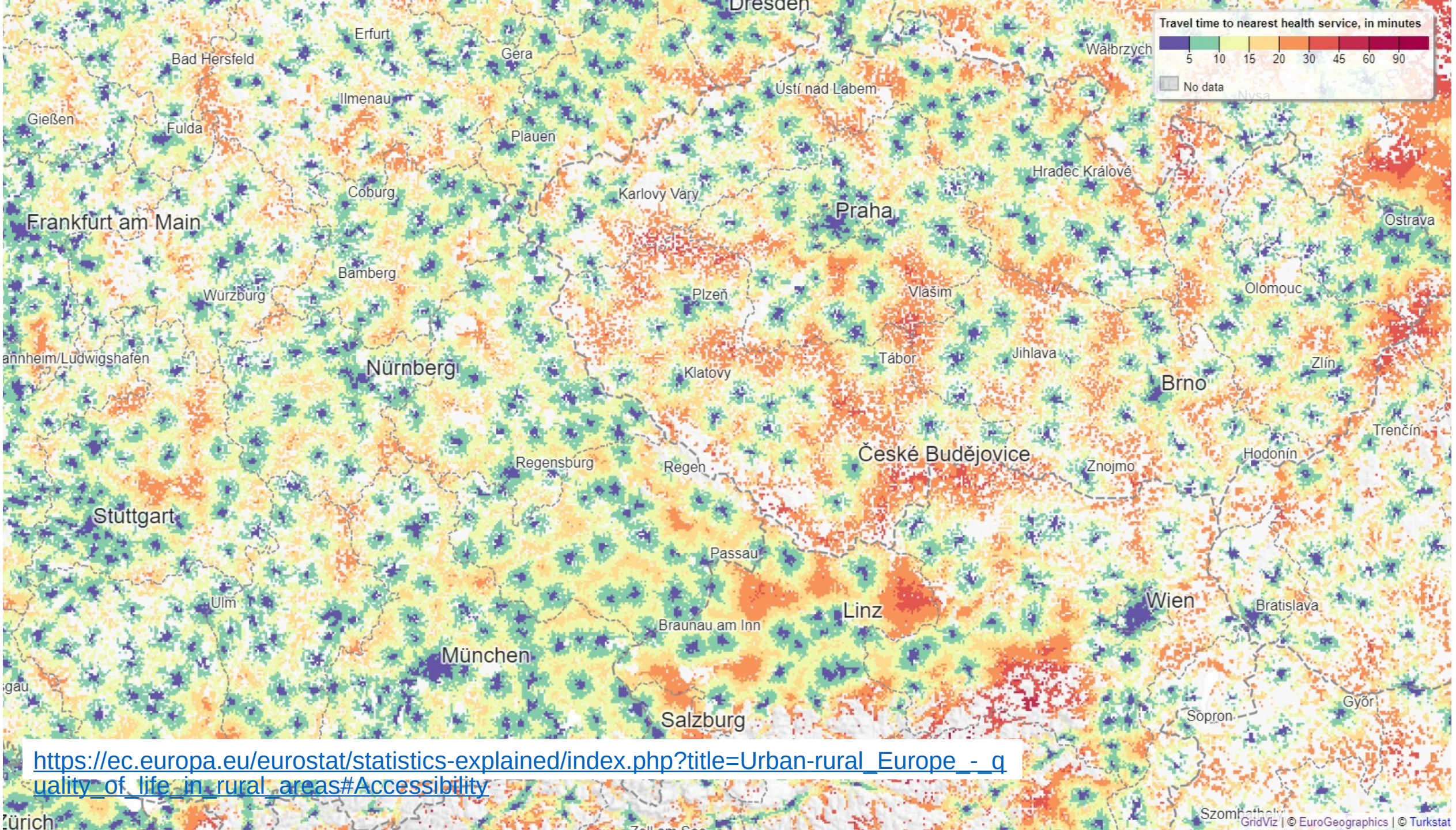

MMU
Initial geometry
Generalised geometry
Hausdorff segment

Quality influence on spatial analyses

European Commission

Data quality and spatial analysis

- Accessibility analysis to healthcare and education services.
- 1km resolution population grid.
- Travel time to nearest service by road transport.



https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Urban-rural_Europe_-_quality_of_life_in_rural_areas#Accessibility

Data quality and spatial analysis

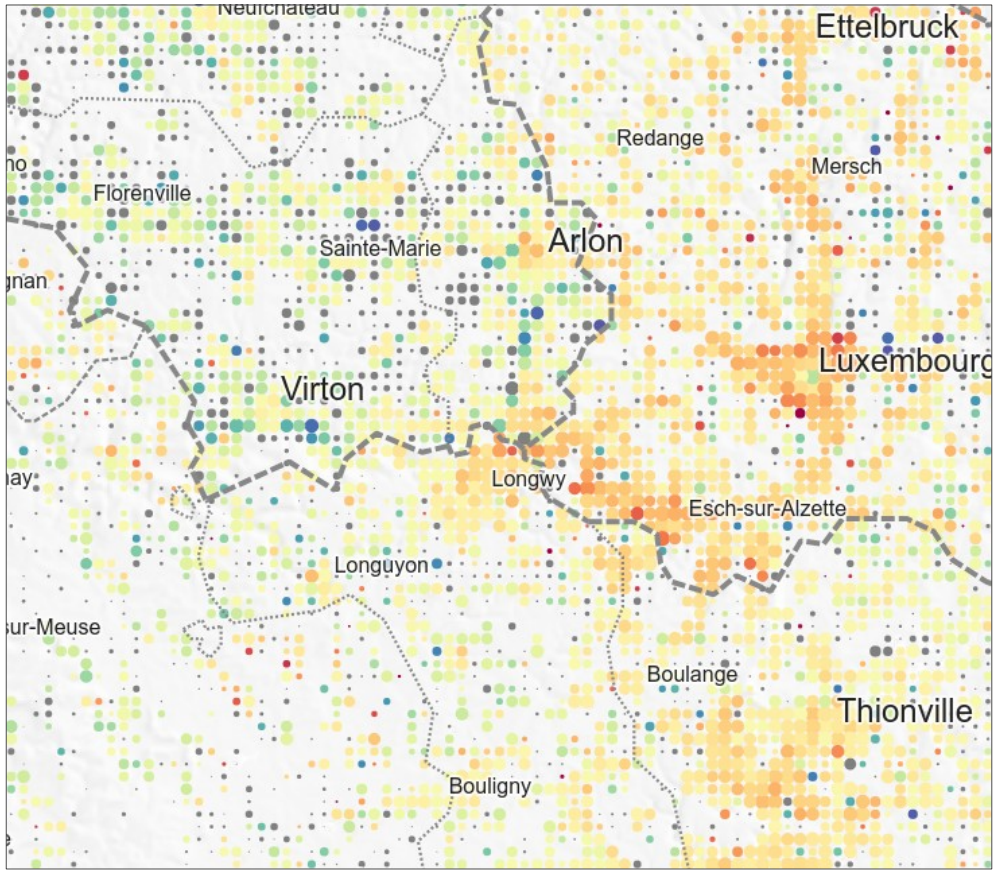
- Tests based on 3 data sources:
  NMCA data, TomTom multinet, OpenStreetMap.

Comparison case	RMS error – Healthcare (min)	RMS error – Education (min)
NMCA vs. TomTom multinet	**3.15**	**1.32**
TomTom multinet vs. OpenStreetMap	**13.27**	**4.37**
OpenStreetMap vs. NMCA vs.	**12.05**	**4.11**

Data quality and spatial analysis

- Building demography: Various indicators on building surface, by use (residential, industrial, etc.)
- 1km resolution grid
- NMCA data, OpenStreetMap.

Comparison case	RMS error – Residential (m²)	RMS error – Industrial (m²)	RMS error – Commercial & services (m²)
NMCA vs. OpenStreetMap	**85 142**	**20 548**	**15 927**



Outline

1. Quality requirements
2. Quality control
3. Quality influence on spatial analyse

Thank you

European Commission