POLITECNICO DI MILANO

OSM seen from a GIS researcher: experiences & perspectives

Marco Minghini

Aizu-Wakamatsu, Fukushima, Japan | August 19, 2017

Who I am

- Postdoctoral Research Fellow at Politecnico di Milano, Italy
  - research topics: Volunteered Geographic Information (VGI), Citizen Science, (geo)crowdsourcing & OpenStreetMap
- Secretary of the ISPRS (International Society for Photogrammetry and Remote Sensing) WG IV/4 "Collaborative Crowdsourced Cloud Mapping (C³M)" since 2016
- Charter Member of the Open Source Geospatial Foundation (OSGeo) since 2015
- OSM contributor since 2014 [username: mingo23]
- OSM teacher and mapathon organizer since 2015
- Voting Member of HOT since 2017
- Faculty Advisor of PoliMappers – a YouthMappers chapter based at Politecnico di Milano, since 2016
  - https://wiki.openstreetmap.org/wiki/User:Mingo23

Research on OSM

- ✔ Over the last few years, OSM has become a research topic on its own
  - ➔ 5 core research areas (+ 50 research trends) were identified [1]:
    - ✗ quality assessment and analysis
    - ✗ assessment of contributors' behavior
    - ✗ application to navigation and disaster
    - ✗ traffic simulation and mobility
    - ✗ indoor navigation models
- ✔ This presentation will focus on 3 recent research works on OSM:
  - ➔ 1. quality assessment of OSM road networks
  - ➔ 2. analysis of OSM contribution patterns
  - ➔ 3. use of OSM to generate Land Use/Land Cover maps

[1] Sehra S.S., Singh J. & Rai H.S. (2017). Using Latent Semantic Analysis to Identify Research Trends in OpenStreetMap. *ISPRS International Journal of Geo-Information*, 6(7), 195.

1

Quality assessment of OSM road networks

OSM quality

- Increasing availability of open data from National Mapping Agencies and Commercial Mapping Companies usable as a source of comparison for VGI (and OSM) data, i.e. for extrinsic quality assessment
- Literature provides plenty of works assessing or comparing OSM quality against that of authoritative datasets:
  - strongly focused on road network
  - OSM compared to data from NMA (UK Ordnance Survey, French NMA, USGS TNM/TIGER, etc.) and CSC (Navteq, TeleAtlas, etc.)
  - semi- or fully-automated
- Comparison techniques are very strong and fit for purpose, but mostly application and dataset specific:
  - hard to replicate
  - difficult to extend to other dataset comparisons

Our methodology

- ✔ Novel methodology to compare OSM and authoritative road datasets:
  - ➔ fully automated
  - ➔ focused on spatial accuracy and completeness
  - ➔ flexible, i.e. not developed for a specific dataset
  - ➔ built with FOSS4G (Free and Open Source Software for Geospatial)
    - ✗ reusable and extensible in case of need

Our methodology – Overview

- Currently developed as 3 GRASS GIS modules:
  - written in Python
  - available with a Graphical User Interface (GUI)
- Comparison between the OSM & the reference road network datasets composed of 3 consecutive steps:
  - 1. Preliminary comparison of the datasets and computation of global statistics
  - 2. Geometric preprocessing of the OSM dataset to extract a subset which is fully comparable with the reference dataset
  - 3. Evaluation of OSM spatial accuracy using a grid-based approach
- Source code: https://github.com/MoniaMolinari/OSM-roads-comparison

Case study: Paris

data © IGN and © OpenStreetMap contributors

Step 1: Preliminary comparison of the datasets

- ✔ Compute the total length of the OSM and IGN datasets and their length difference, both in map units and percentage [*required*]
  - ➔ output values are returned in a text file

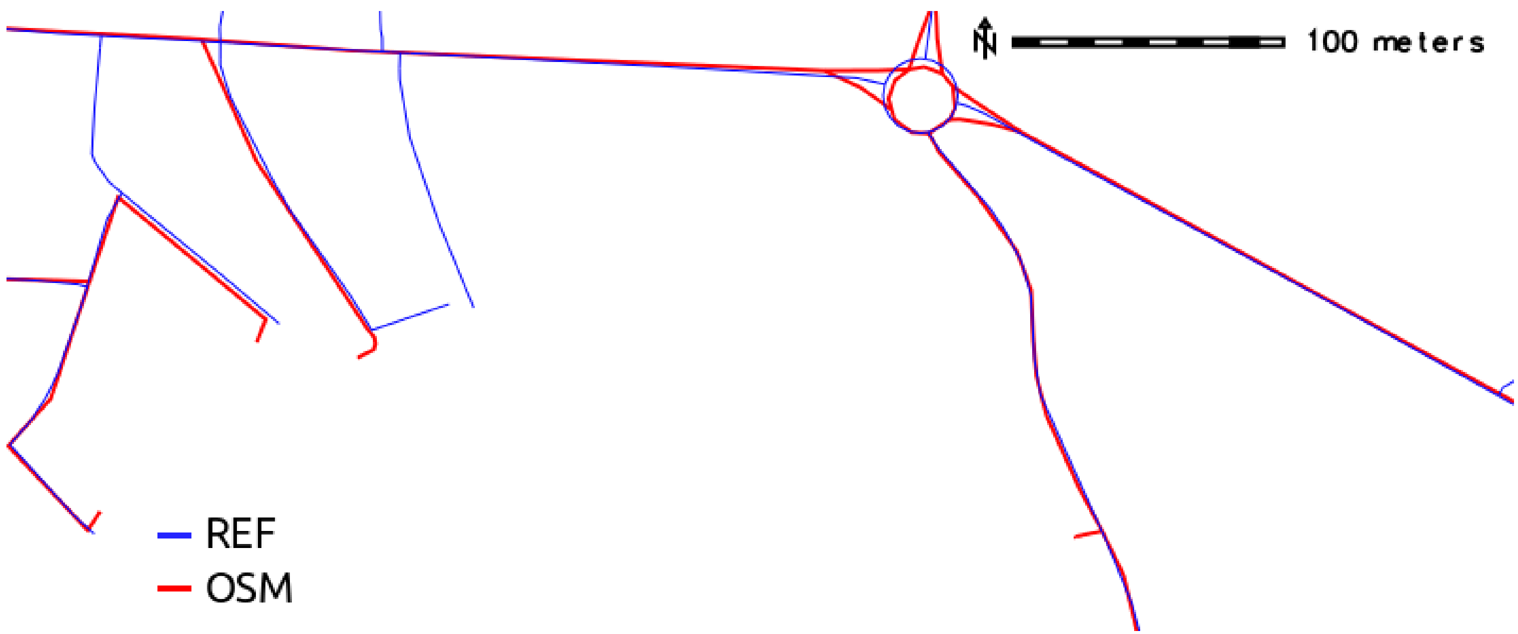
```
REF length: 2686373.1 m
OSM length: 3124627.0 m
REF-OSM difference: -438253.9 m (-16.3%)
```

✗ ≅450 km more in OSM than IGN dataset!

```
BUFFER(m)|OSM_IN(m)|OSM_IN(%%)|OSM_OUT(m)|OSM_OUT(%%)|REF_IN(m)|REF_IN(%%)|REF_OUT(m)|REF_OUT(%%)
1.0|1374755.9|44.0|1749871.2|56.0|1366471.0|50.9|1319902.1|49.1
2.0|2014259.9|64.5|1110367.2|35.5|1982713.7|73.8|703659.4|26.2
3.0|2298072.4|73.5|826554.6|26.5|2223153.5|82.8|463219.6|17.2
4.0|2464185.0|78.9|660442.0|21.1|2329270.3|86.7|357102.8|13.3
5.0|2582784.2|82.7|541842.9|17.3|2387687.7|88.9|298685.4|11.1
6.0|2671758.8|85.5|452868.2|14.5|2424463.5|90.3|261909.6|9.7
7.0|2738327.0|87.6|386300.0|12.4|2451476.9|91.3|234896.2|8.7
8.0|2792053.8|89.4|332573.2|10.6|2471557.1|92.0|214816.0|8.0
9.0|2828903.0|90.5|295724.1|9.5|2488514.1|92.6|197859.0|7.4
10.0|2859512.1|91.5|265114.9|8.5|2501974.7|93.1|184398.4|6.9
11.0|2886190.1|92.4|238436.9|7.6|2513592.9|93.6|172780.2|6.4
12.0|2908071.9|93.1|216555.1|6.9|2523138.5|93.9|163234.6|6.1
13.0|2925602.0|93.6|199025.1|6.4|2532070.5|94.3|154302.6|5.7
14.0|2941922.8|94.2|182704.2|5.8|2540322.9|94.6|146050.2|5.4
15.0|2956112.7|94.6|168514.3|5.4|2548274.0|94.9|138099.1|5.1
16.0|2967813.5|95.0|156813.5|5.0|2555431.5|95.1|130941.6|4.9
17.0|2977318.7|95.3|147308.3|4.7|2562238.1|95.4|124135.0|4.6
18.0|2986371.8|95.6|138255.2|4.4|2568276.5|95.6|118096.6|4.4
19.0|2994833.4|95.8|129793.7|4.2|2574052.2|95.8|112320.9|4.2
20.0|3001796.0|96.1|122831.1|3.9|2579434.1|96.0|106939.0|4.0
```

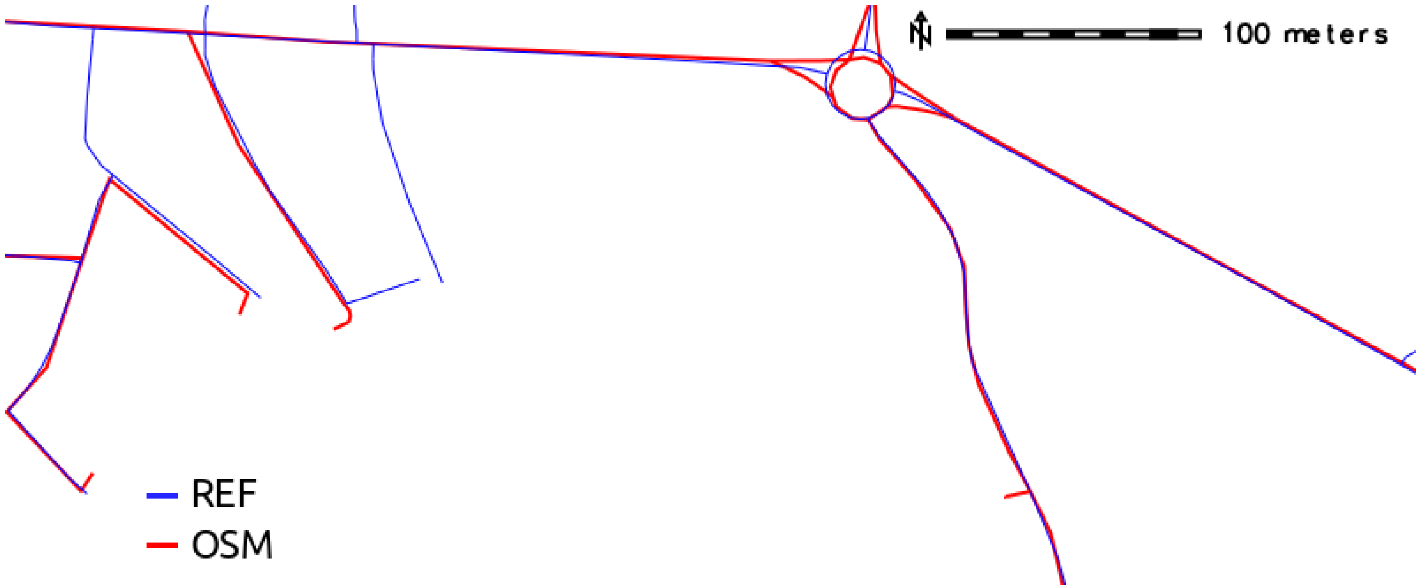
Step 2: preprocessing of the OSM dataset

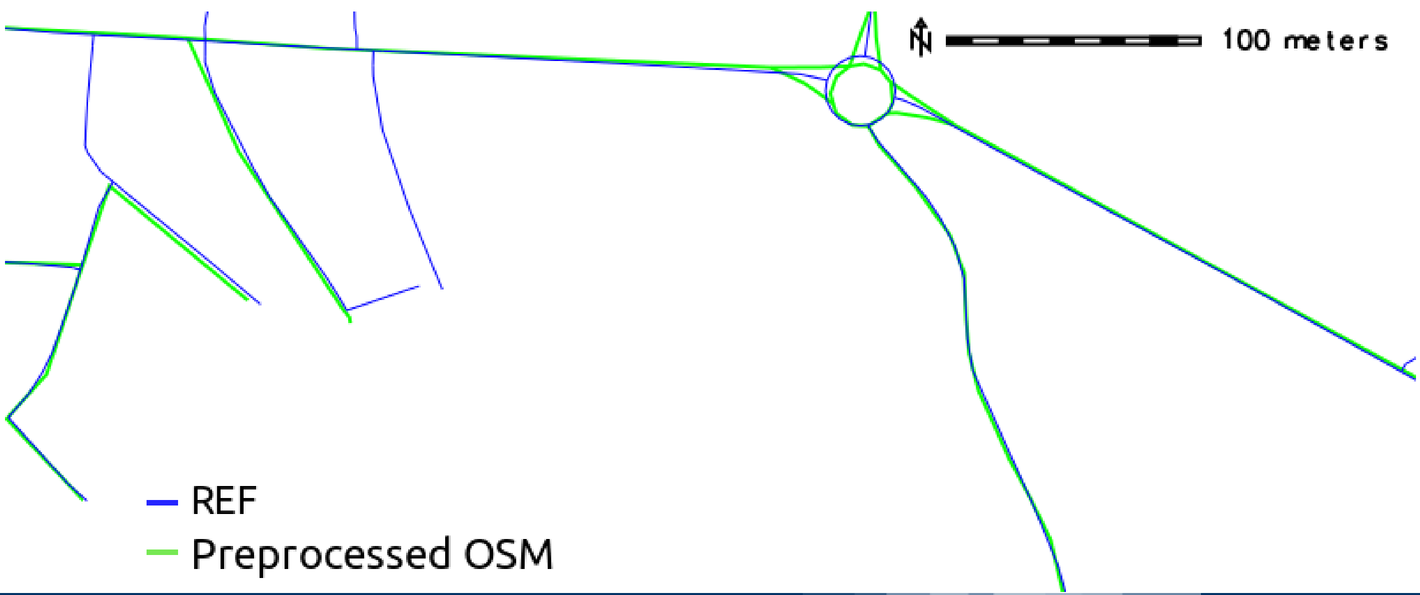
- ✔ Cleaning of OSM dataset to make it comparable with IGN dataset
- ✔ Apply a buffer of user-specified width around the IGN dataset
  - ➔ suitable buffer width derived from Step 1
  - ➔ delete all the OSM roads falling outside the buffer



Step 2: preprocessing of the OSM dataset

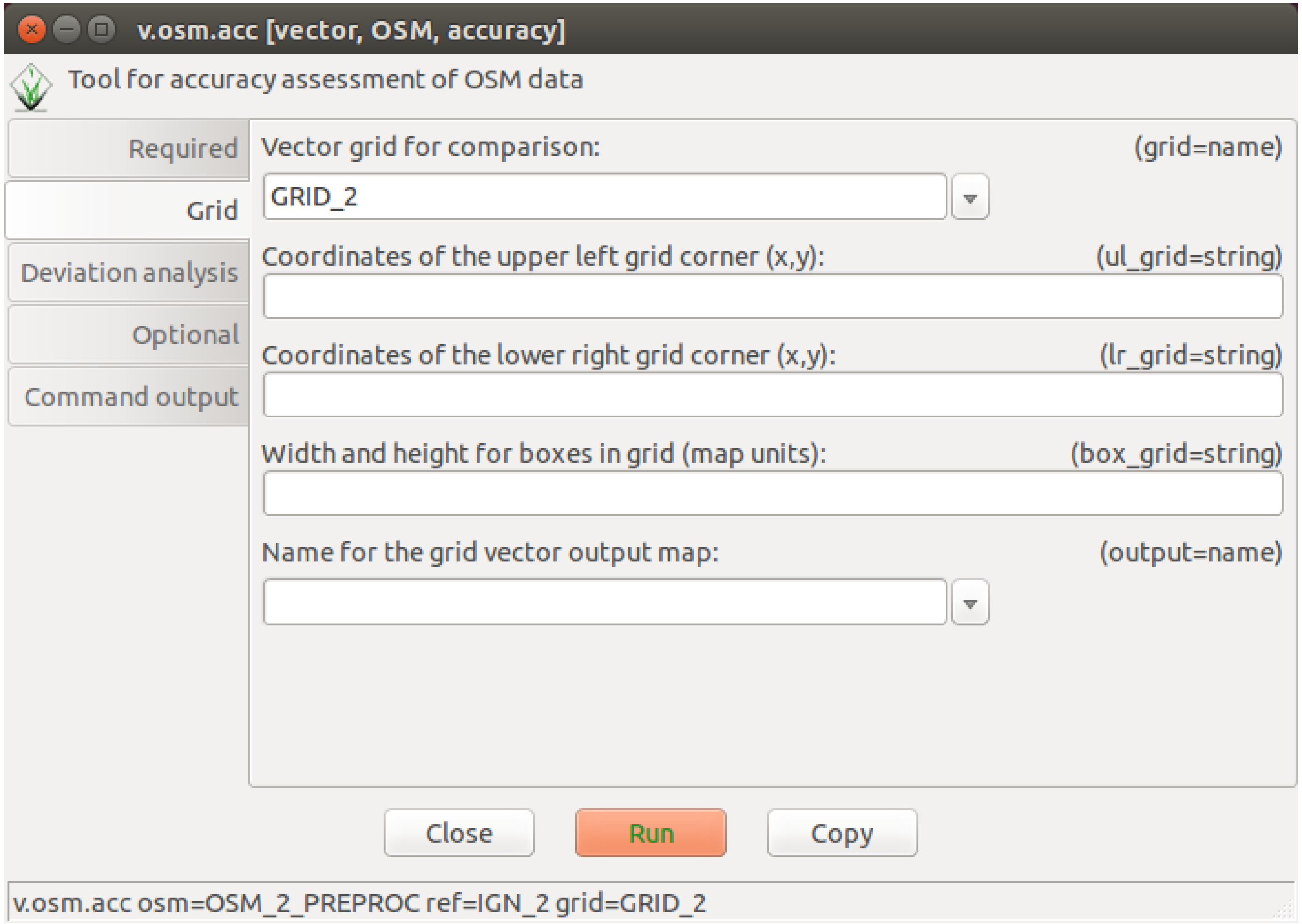
✔ Further clean the OSM dataset:





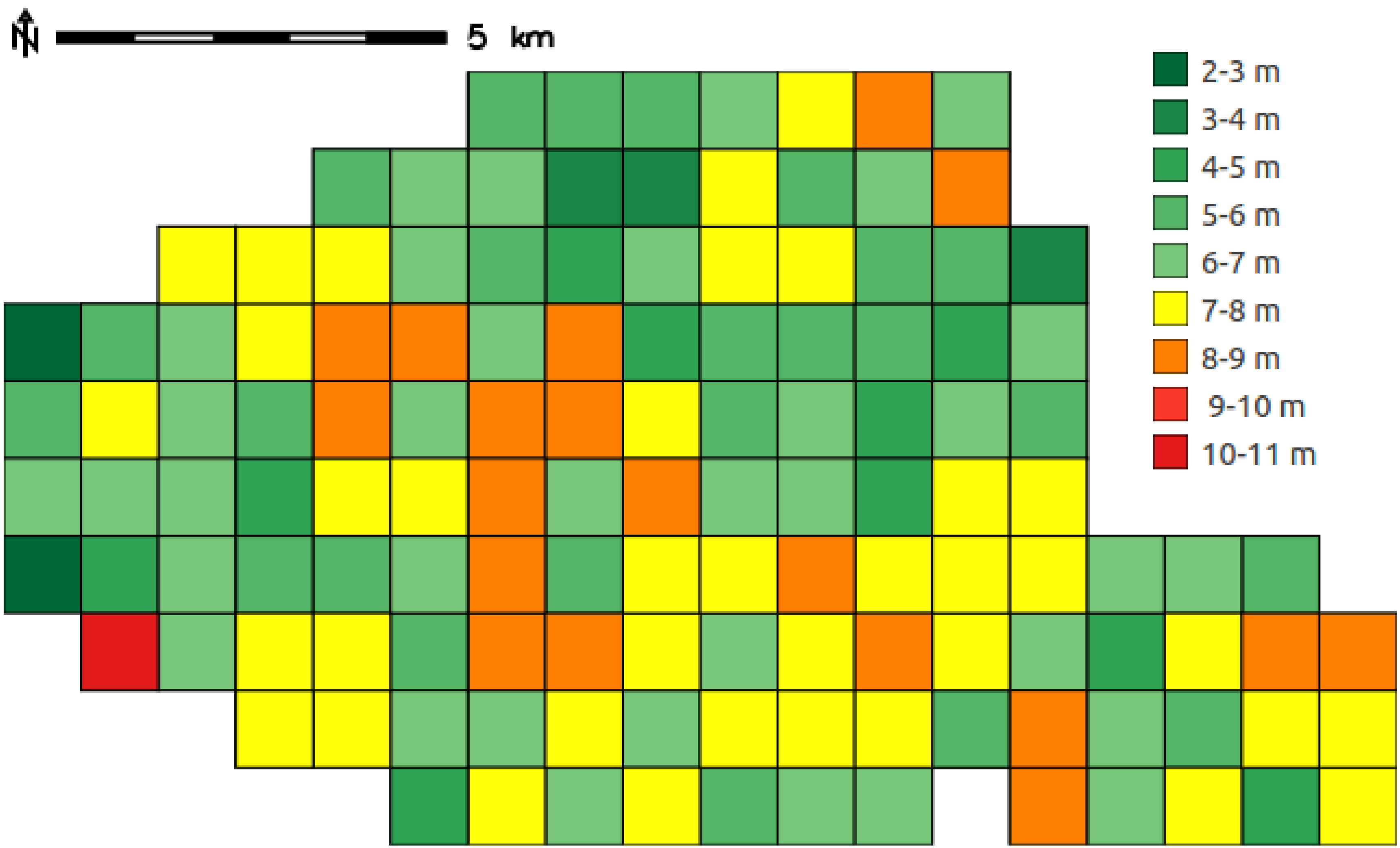
Step 3: grid-based evaluation of OSM accuracy

- ✔ Use a grid to take into account OSM heterogeneous nature:



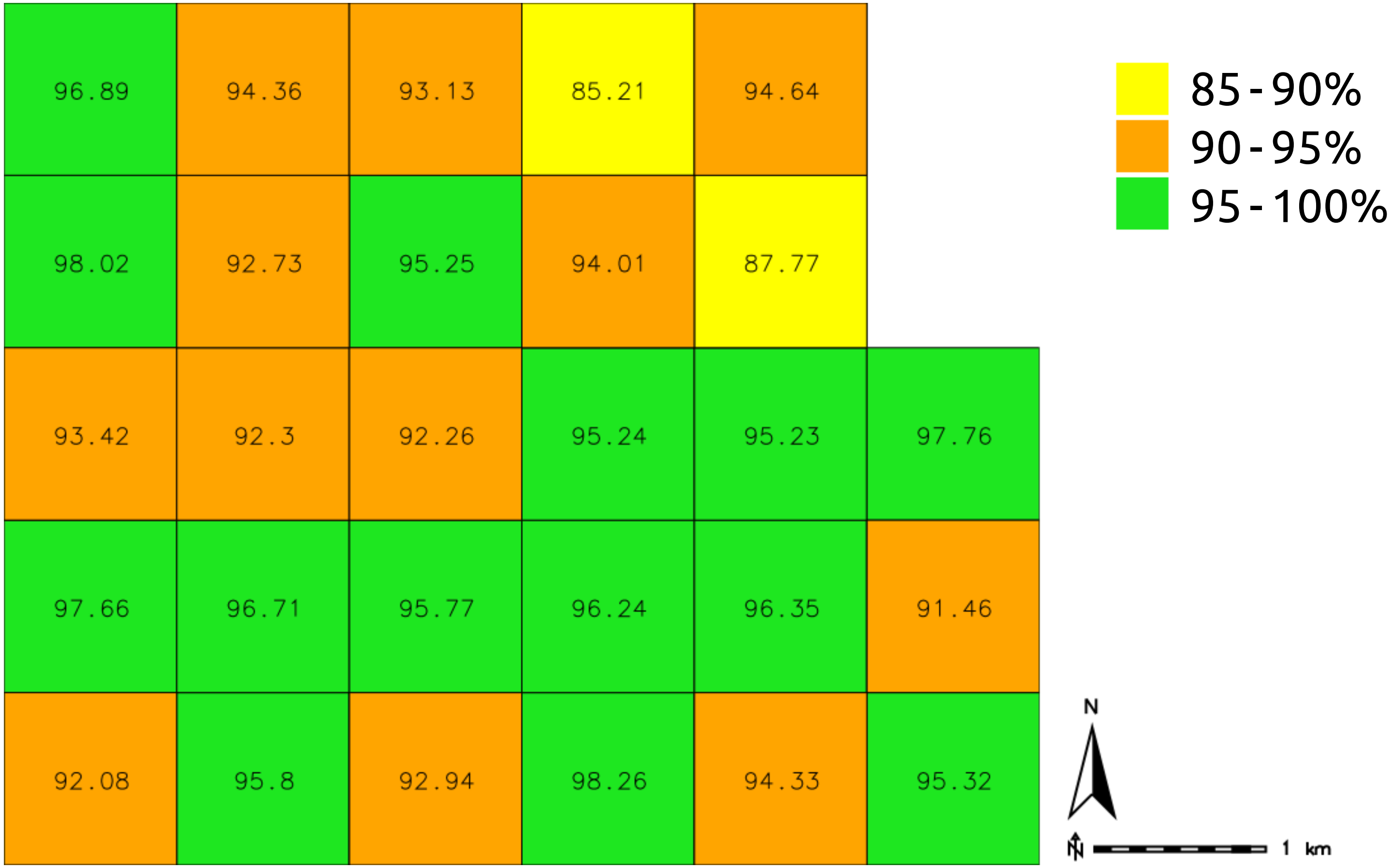
Step 3: grid-based evaluation of OSM accuracy

- For each grid cell, find the OSM maximum deviation from IGN:
  - generalization threshold = 0.5 m, buffer = 11 m, OSM length % = 95



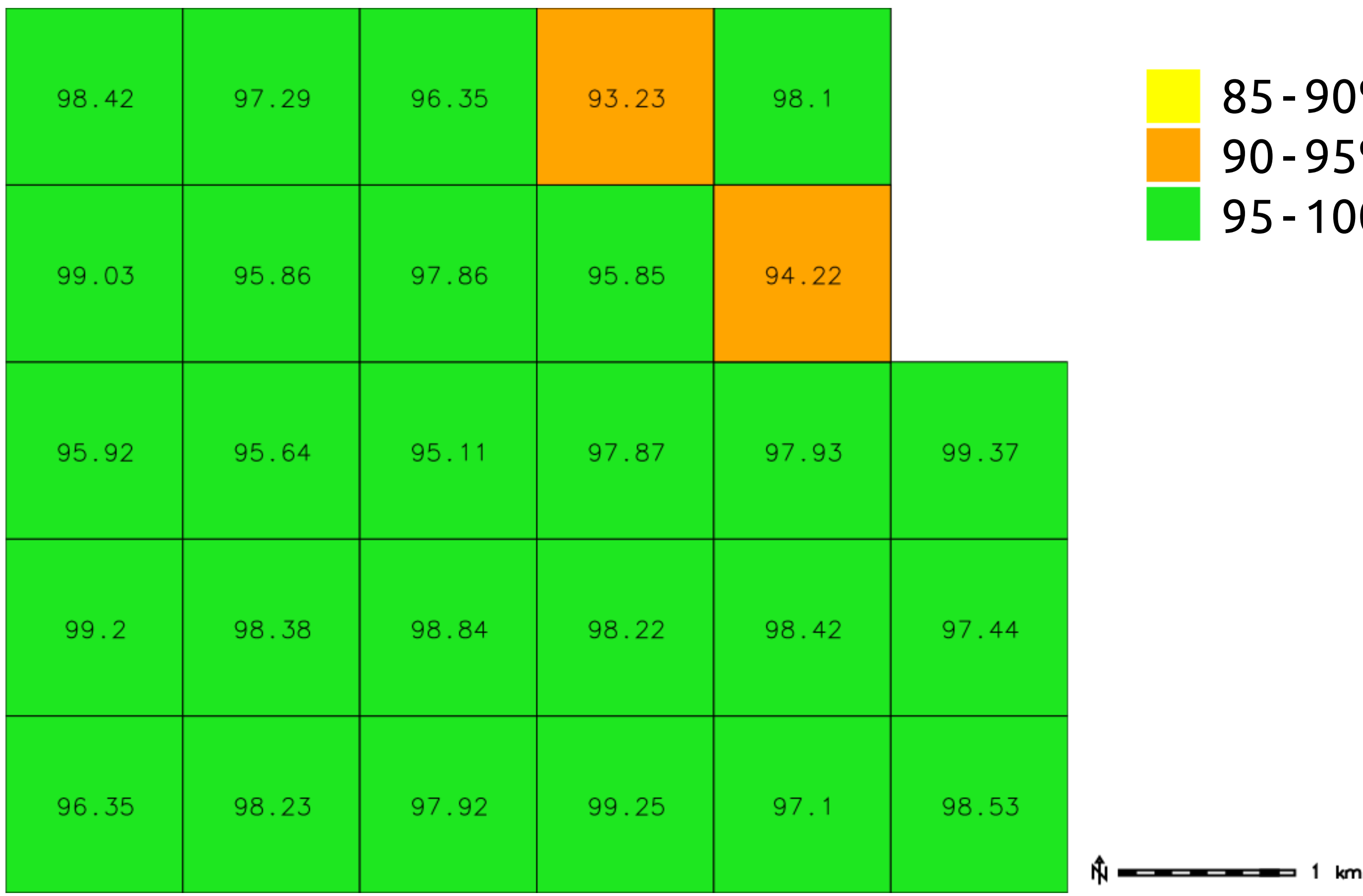
Step 3: grid-based evaluation of OSM accuracy

- For each grid cell, evaluate OSM accuracy against one or more target values of OSM deviation from IGN:
  - length percentage of OSM roads included in the target buffer
  - Area 2: target buffer = 6 m



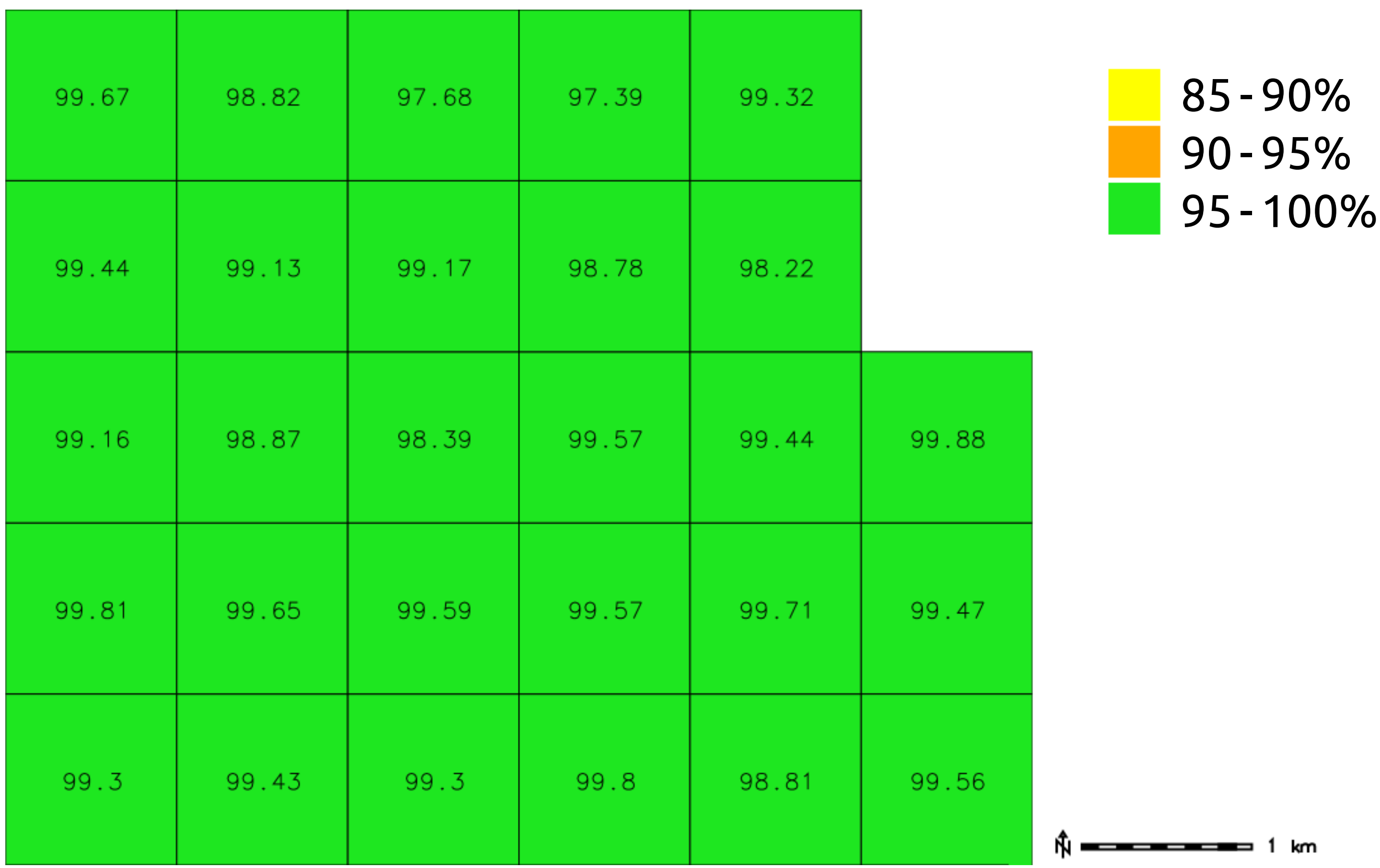
Step 3: grid-based evaluation of OSM accuracy

- ✔ For each grid cell, evaluate OSM accuracy against one or more target values of OSM deviation from IGN:
  - ➔ length percentage of OSM roads included in the target buffer
  - ➔ Area 2: target buffer = 8 m



Step 3: grid-based evaluation of OSM accuracy

- For each grid cell, evaluate OSM accuracy against one or more target values of OSM deviation from IGN:
  - length percentage of OSM roads included in the target buffer
  - Area 2: target buffer = 10 m



2

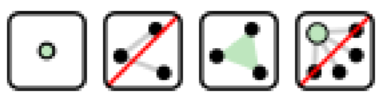
Analysis of OSM contribution patterns

Tagging in OSM

✔ OSM applies a folksonomy approach to tagging with no formal rules or ontologies forced

- ➔ tagging rule-book is the OSM Map Features wiki page
  - ✗ guidance on which tags and combinations of tags to use

Used on these elements



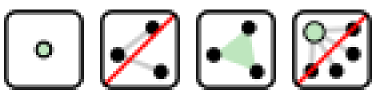
Useful combination

- name=*
- Address
- operator=*
- cuisine=*
- opening_hours=*
- website=*
- phone=*

Tagging in OSM

- OSM applies a folksonomy approach to tagging with no formal rules or ontologies forced
  - tagging rule-book is the OSM Map Features wiki page
    - guidance on which tags and combinations of tags to use
  - taginfo shows that this guidance may not be universally adopted!

Used on these elements



Useful combination

- name=*
- Address
- operator=*
- cuisine=*
- opening_hours=*
- website=*
- phone=*

amenity=restaurant

Overview | Combinations | Map | Wiki | Projects

Combinations

This table shows only the most common combinations of the most common tags.

Page 1 of 13 JSON Displaying 1 to 11 of 135 items

Count →		Other tags
687 829	91.36%	name=*
329 005	43.70%	cuisine=*
246 939	32.80%	addr:street=*
204 893	27.22%	addr:housenumber=*
180 841	24.02%	addr:city=*
168 643	22.40%	addr:postcode=*
140 042	18.60%	building=*
127 409	16.92%	building=yes
113 375	15.06%	phone=*
111 769	14.85%	website=*
93 607	12.43%	source=*

Analysis of OSM tagging practices

- Research questions:
  - do OSM contributors comply to the suggested combinations of tags?
  - does this compliance vary spatially?
- Selection of 10 among the most frequently occurring tags in OSM

Target Tag	TagInfo Ranking	Number of Objects
highway=residential	2	34,688,039
natural=tree	17	7,019,552
highway=footway	18	6,126,861
highway=path	24	4,506,593
highway=tertiary	25	4,328,513
amenity=parking	52	2,061,012
highway=primary	59	1,869,021
highway=bus_stop	66	1,677,724
railway=rail	69	1,584,142
leisure=pitch	93	977,983

Analysis of OSM tagging practices

- Research questions:
  - do OSM contributors comply to the suggested combinations of tags?
  - does this compliance vary spatially?
- Selection of 10 among the most frequently occurring tags in OSM
- Selection of 40 world cities

Methodology

- For each city, for each target tag and for each of the suggested tags to be used in combination:
  - computation of the fraction of objects containing both the target tag and the suggested tag
  - mapping of the fraction to a 5 part Likert Scale
    - 0-20% – POOR
    - 20-40% – FAIR
    - 40-60% – AVERAGE
    - 60-80% – GOOD
    - 80-100% – EXCELLENT
- Example: Christchurch (New Zealand), tag *leisure=pitch*

Report for Tag: `leisure=pitch`			
Total number of objects: 470			
sport	364	77.5%	GOOD
surface	42	9.0%	POOR
Total number of different tags used: 26			

Results

- *highway=residential*

KEY	Poor	Fair	Average	Good	Excellent
name	8	1	7	5	19
oneway	34	5	1	0	0

- *natural=tree*

KEY	Poor	Fair	Average	Good	Excellent
circumference	38	0	1	0	1
taxon	38	0	0	0	2
leaf_type	34	2	2	1	1
start_date	39	0	0	1	0
height	37	0	1	0	2
denotation	36	1	2	0	1
genus	38	1	1	0	0
species	35	1	2	0	2

- *highway=primary*

KEY	Poor	Fair	Average	Good	Excellent
lanes	10	10	6	6	8
ref	8	10	6	2	14
name	0	2	4	10	24

Results

- *highway=bus_stop*

KEY	Poor	Fair	Average	Good	Excellent
operator	28	4	2	3	3
public_transport	21	7	5	3	4
name	3	4	3	9	21

- *leisure=pitch*

KEY	Poor	Fair	Average	Good	Excellent
sport	0	2	7	16	15
surface	40	0	0	0	0

- Summary:

Tag	Keys	Poor	Fair	Average	Good	Excellent
highway=primary	3	15.00	18.33	13.33	15.00	38.33
highway=tertiary	4	40.00	20.00	13.75	14.38	11.88
highway=bus-stop	3	43.33	12.50	8.33	12.50	23.33
railway=rail	9	46.39	18.61	12.78	11.67	10.56
leisure=pitch	2	50.00	2.50	8.75	20.00	18.75
highway=residential	2	52.50	7.50	10.00	6.25	23.75
amenity=parking	6	90.83	6.67	2.50	0.00	0.00
highway=path	7	91.78	5.71	2.50	0.00	0.00
natural=tree	8	92.19	1.56	2.81	0.62	2.81
highway=footway	6	94.58	4.58	0.83	0.00	0.00

3

Use of OSM to generate Land Use/ Land Cover maps

Land Use/Land Cover (LULC) maps

- LULC maps are crucial products for multiple areas of application:
  - modeling climate and biochemistry of the Earth
  - biodiversity monitoring
  - natural resources management
  - planning/urban studies
  - many others
- LULC maps are created through the classification of satellite imagery and validated using reference data:
  - the creation and updating process is long, costly & time-consuming – insufficient to describe rapidly-changing environments
  - the level of detail and spatial coverage are inadequate for many applications

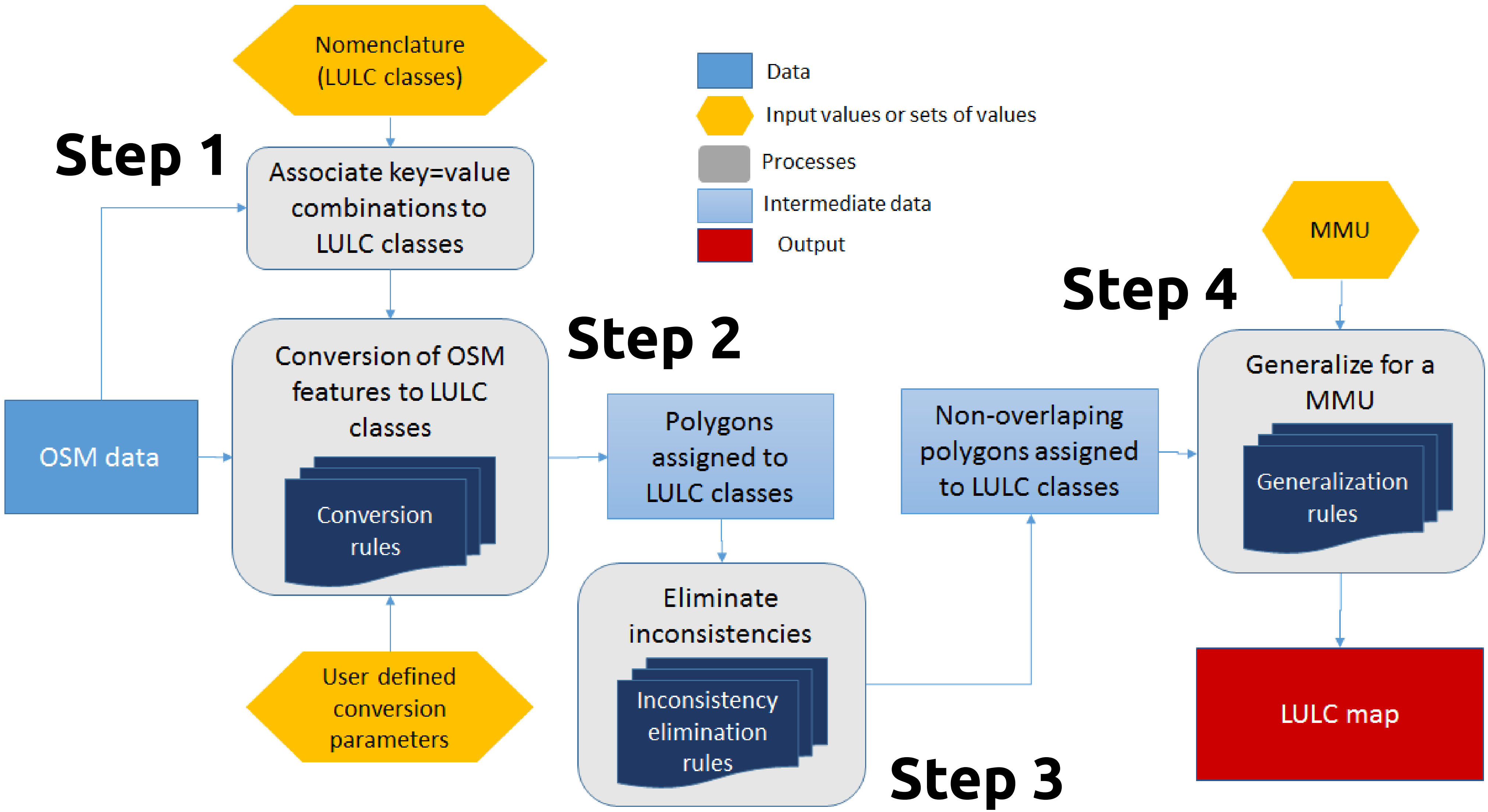
OSM as a source of LULC maps

- Exploiting OSM as a source for LULC maps has a number of advantages:
  - OSM full spatial coverage in the world
  - OSM richness
  - OSM non-stop updating
  - OSM open license
- Exploiting OSM as a source for LULC maps has some disadvantages:
  - OSM uneven spatial coverage
  - OSM positional accuracy & geometrical inconsistencies
  - OSM semantic inconsistencies
- Purpose: creating an automated procedure which converts OSM data in a specific area into a LULC map
  - reference nomenclatures of current EU and global LULC maps (e.g. Urban Atlas, Corine Land Cover, GL30)

Example of nomenclature: Urban Atlas

<table>
<tr><th>Level 1</th><th>Level 2</th><th>Level 3</th></tr>
<tr><td rowspan="4">1 Artificial Surfaces</td><td>1.1 Urban Fabric</td><td>1.1.1 Continuous urban fabric
1.1.2 Discontinuous urban fabric
1.1.3 Isolated Structures</td></tr>
<tr><td>1.2 Industrial, commercial, public, military, private and transport units</td><td>1.2.1 Industrial, commercial, public, military and private units
1.2.2 Road and rail network and associated land
1.2.3 Port areas
1.2.4 Airports</td></tr>
<tr><td>1.3 Mine, dump and construction sites</td><td>1.3.1 Mineral extraction and dump sites
1.3.3 Construction sites
1.3.4 Land without current use</td></tr>
<tr><td>1.4 Artificial non-agricultural vegetated areas</td><td>1.4.1 Green urban areas
1.4.2 Sports and leisure facilities</td></tr>
<tr><td colspan="3">2 Agricultural, semi-natural areas, wetlands</td></tr>
<tr><td colspan="3">3 Forests</td></tr>
<tr><td colspan="3">5 Water</td></tr>
</table>

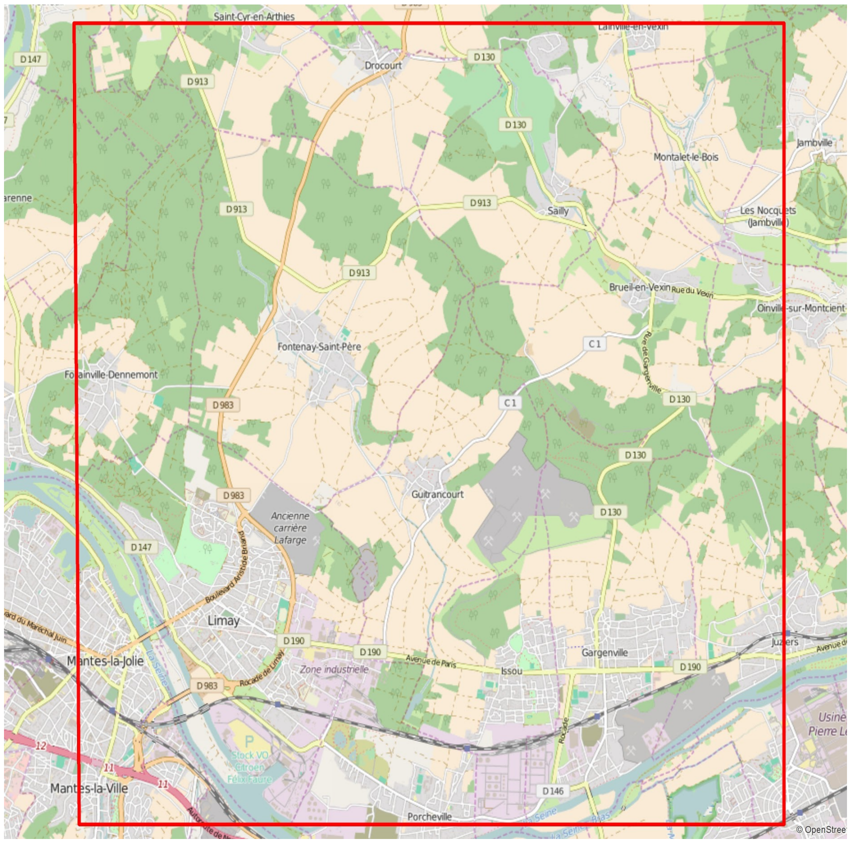
Methodology to convert OSM into LULC maps



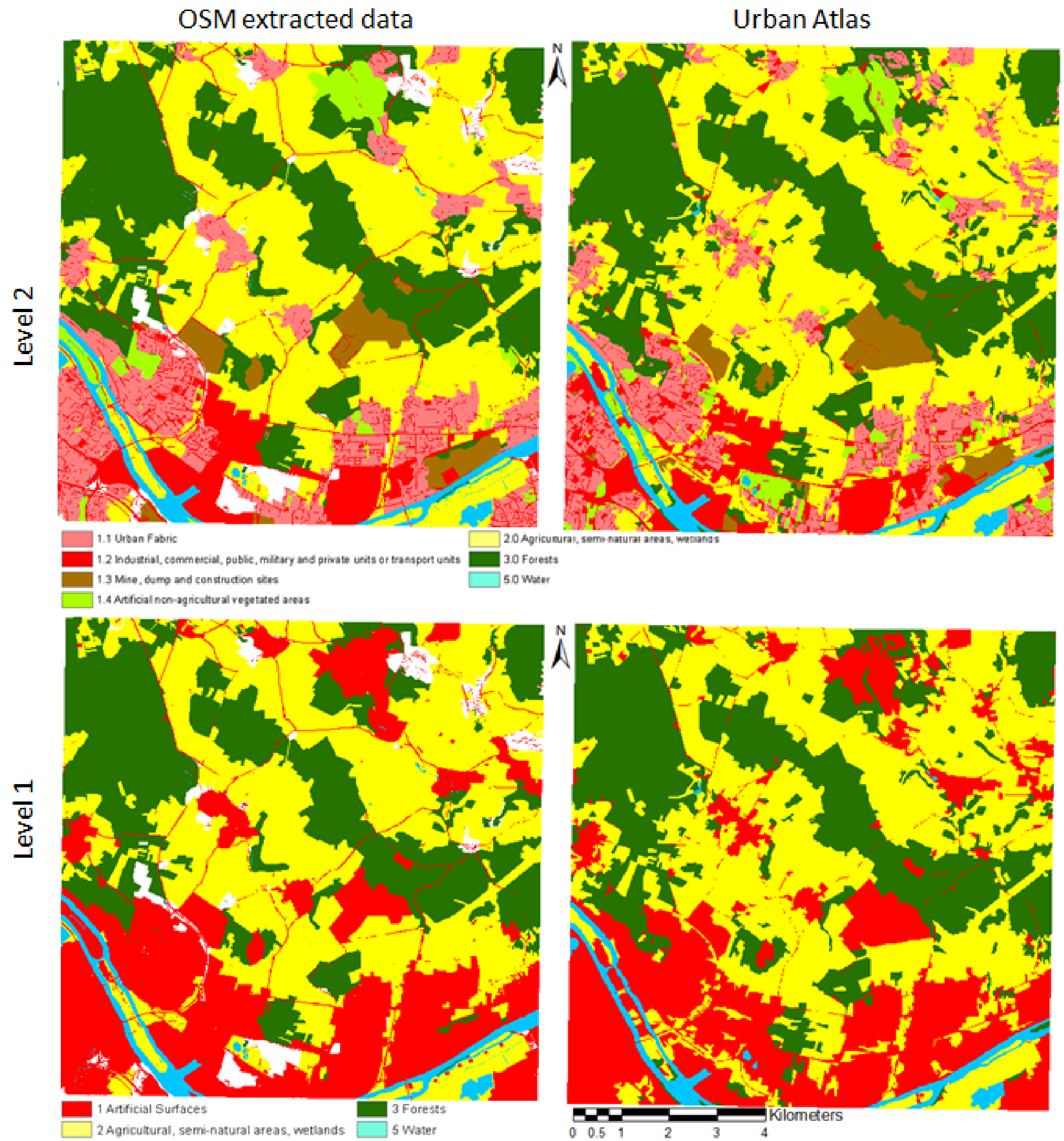
- ✔ Source code: https://github.com/JoaquimPatriarca/senpy-for-gis
- ✔ Web service: http://vgi.mat.uc.pt/vgi/osm/osm2lulc – work in progress!

Case studies

- ✔ Paris area

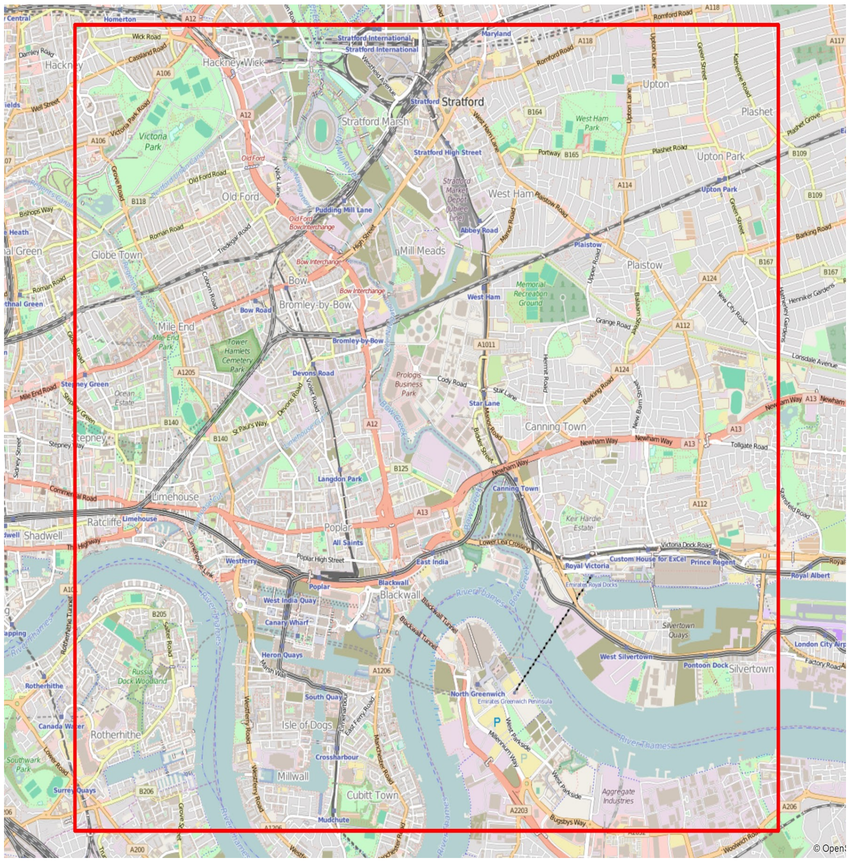


© OpenStreetMap contributors

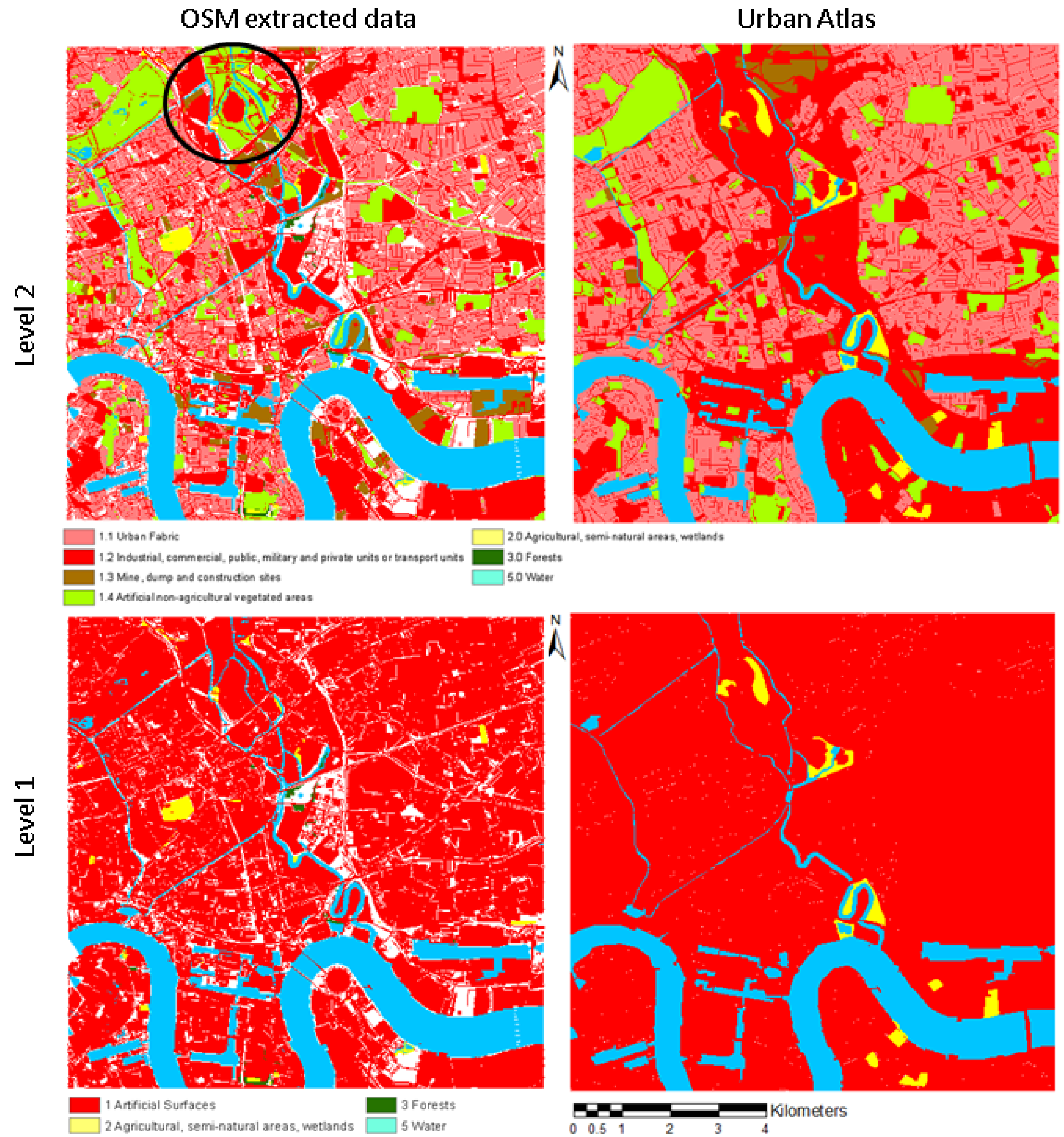


Case studies

- ✔ London area



© OpenStreetMap contributors



Case studies

- Areas [ha] occupied by Level 2 classes associated to the overlapping regions in the Urban Atlas & the OSM-derived maps – Paris study area

PARIS		Classes assigned to the overlapping regions in the OSM-derived map							Match/ Row Sum (%)
		1.1	1.2	1.3	1.4	2	3	5	
Classes assigned to the overlapping regions in UA	1.1	967	106	1	11	50	24	1	83
	1.2	186	640	37	20	50	13	3	67
	1.3	19	24	227	0	45	7	0	71
	1.4	56	26	0	161	57	6	5	52
	2	108	148	33	43	3545	124	10	88
	3	21	28	11	44	138	2425	5	91
	5	3	4	1	1	6	5	221	92
Match/Column Sum (%)		71	66	73	57	91	93	90	85

Case studies

- Areas [ha] occupied by Level 2 classes associated to the overlapping regions in the Urban Atlas & the OSM-derived maps – London study area

LONDON		Classes assigned to the overlapping regions in the OSM-derived map							Match/ Row Sum (%)
		1.1	1.2	1.3	1.4	2	3	5	
Classes assigned to the overlapping regions in UA	1.1	2346	796	16	86	8	2	21	72
	1.2	525	2323	214	174	32	8	86	69
	1.3	25	51	18	26	5	3	7	14
	1.4	19	111	5	644	17	5	18	79
	2	5	18	41	23	3	3	9	3
	3	0	0	0	0	0	0	0	0
	5	12	22	8	5	0	0	1107	96
Match/Column Sum (%)		80	70	6	67	4	0	89	73

References

- ✔ Antunes F., Fonte C.C., Brovelli M.A., Minghini M., Molinari M.E. & Mooney P. (2015) Assessing OSM Road Positional Quality with Authoritative Data. *Proceedings of the VIII Conferência Nacional de Cartografia e Geodesia*, Lisbon (Portugal), October 29-30, 2015.
- ✔ Brovelli M.A., Minghini M., Molinari M.E. & Mooney P. (2017) Towards an automated comparison of OpenStreetMap with authoritative road datasets. *Transactions in GIS* 21(2), 191-206.
- ✔ Brovelli M.A., Minghini M. & Molinari M.E. (2016) An automated GRASS-based procedure to assess the geometrical accuracy of the OpenStreetMap Paris road network. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B7, 919-925.
- ✔ Brovelli M.A., Minghini M., Molinari M.E. & Mooney P. (2015) A FOSS4G-based procedure to compare OpenStreetMap and authoritative road network datasets. *Geomatics Workbooks* 12, 235-238.
- ✔ Davidovic N., Mooney P., Stoimenov L. & Minghini M. (2016) Tagging in Volunteered Geographic Information: An Analysis of Tagging Practices for Cities and Urban Regions in OpenStreetMap. *ISPRS International Journal of Geo-Information* 5(12), 232.
- ✔ Fonte C.C., Minghini M., Patriarca J., Antoniou V., See L. & Skopeliti A. (2017) Generating Up-to-Date and Detailed Land Use and Land Cover Maps Using OpenStreetMap and GlobeLand30. *ISPRS International Journal of Geo-Information* 6(4), 125.
- ✔ Fonte C.C., Patriarca J., Minghini M., Antoniou V., See L. & Brovelli M.A. (2017) Using OpenStreetMap to Create Land Use and Land Cover Maps: Development of an Application. In: *Volunteered Geographic Information and the Future of Geospatial Data*. IGI Global, 113-137.
- ✔ Fonte C.C., Minghini M., Antoniou V., See L., Patriarca J., Brovelli M.A. & Milcinski G. (2016) An automated methodology for converting OSM data into a land use/cover map. *Proceedings of the 6th International Conference on Cartography & GIS* 1, 462-473, Albena (Bulgaria), June 13-17, 2016.

Acknowledgements

- Thanks to all the amazing people I have worked with on these exciting topics!
  - Vyron Antoniou
  - Francisco Antunes
  - Maria Antonia Brovelli
  - Cidália Costa Fonte
  - Nikola Davidovic
  - Monia Elisa Molinari
  - Peter Mooney
  - Joaquim Patriarca
  - Linda See
  - Andriani Skopeliti
  - Leonid Stoimenov
- Thanks to all OSM contributors for making this possible :)

Contacts

Thank you!

Marco Minghini

Politecnico di Milano
Dept. Of Civil and Environmental Engineering | GEOlab

marco.minghini@polimi.it | @MarcoMinghini

This presentation can be downloaded from here:

Special Issue on JGS on VGI

- ✔ The Journal of Geographical Systems (JGS) is an interdisciplinary journal aiming to encourage and promote high quality scholarship on important theoretical, methodological & empirical issues with a central spatial or regional dimension
  - → Impact Factor: 1.314 (2016), Journal Citation Reports®
- ✔ Special Issue "Volunteered Geographic Information: Looking Towards the Next 10 Years":
  - → the first 10 years of VGI have seen an explosion of activity, particularly in the form of projects such as OpenStreetMap – but what will the next 10 years hold?
  - → Guest Editors:
    - ✗ Linda See, IIASA, Austria
    - ✗ Cidália Costa Fonte, University of Coimbra, Portugal
    - ✗ Vyron Antoniou, Hellenic Military Geographical Service, Greece
    - ✗ Marco Minghini, Politecnico di Milano, Italy