

Towards Quality Metrics for OpenStreetMap



Peter Mooney^{1,2} and Padraig Corcoran¹

Introduction

The OpenStreetMap (OSM) project is one of the most popular examples of Volunteered Geographical Information and a powerful example of geo-crowdsourcing. However there is still much skepticism in the GIS community surrounding the value and quality of OSM data. Only a few academic studies (Haklay (2010), Zielstra and Zipf (2010)) have been carried out to measure the quality (positional accuracy) of OSM data. The lack of clearly defined measurements of quality for OSM is a major obstacle towards the use of OSM data for “real” GIS (analysis and research)

OSM Data Gathering

Developed an algorithm (see below) to automatically analyse history of OSM polygons. For each polygon metrics are calculated to help towards understanding how OSM data is collected and edited. Analysis of 1000 heavily edited OSM polygons

Time between Consecutive Versions	No. Occurrences
< 15 minutes	422
15 mins – 30 mins	135
30 mins and 1 hour	2254
1 hour and 2 hours	95
2 hours and 3 hours	331
3 hours and 4 hours	70
4 hours and 5 hours	2129
5 hours and 12 hours	2183
12 hours and 1 day	585
1 day and 1 week	303
1 week and 1 month	240
> 1 month	5669

- 19% of consecutive edits happen within 1 hour
- 37% happen between 1 hour and 1 day
- 44% more than 1 day between edits

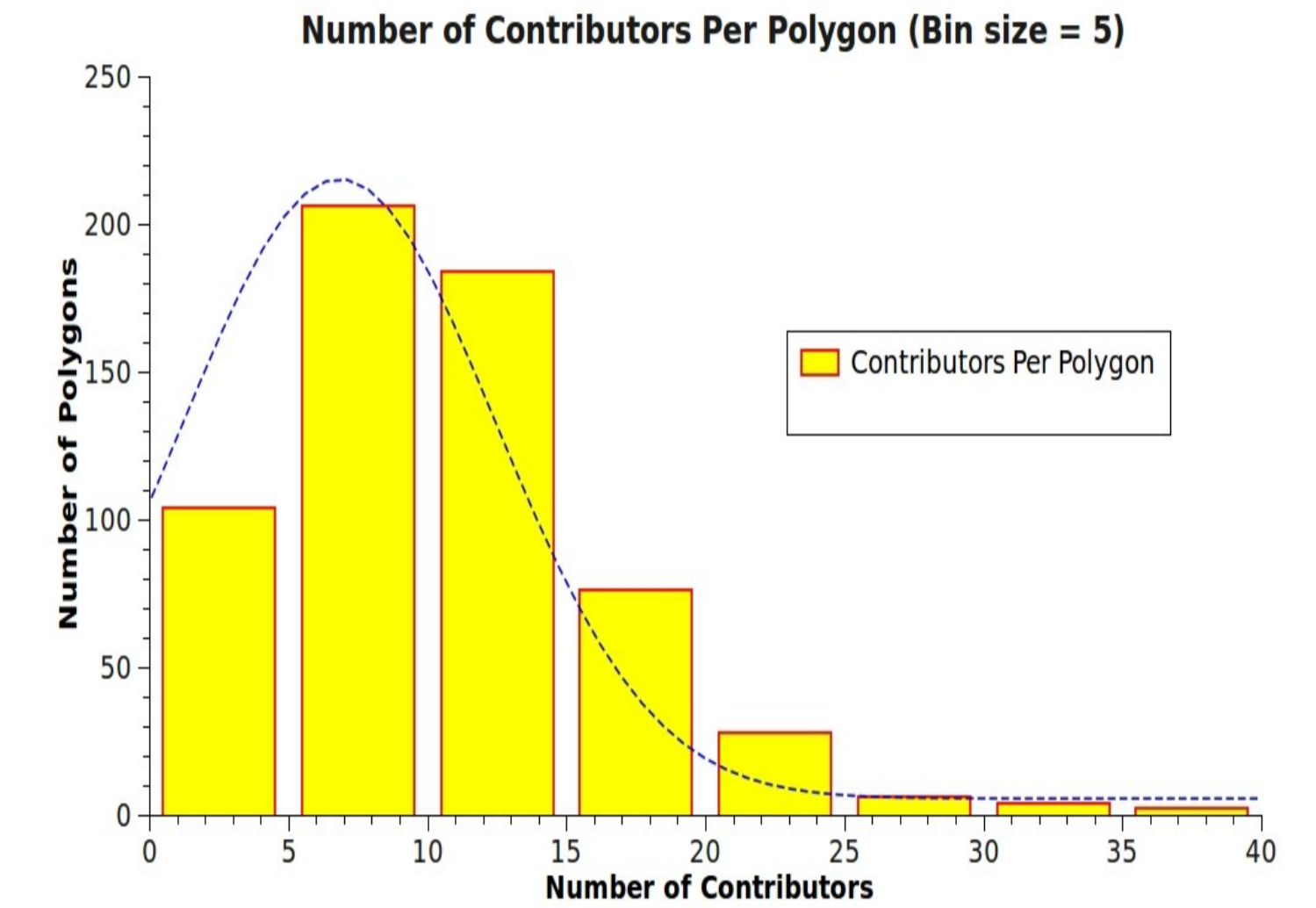
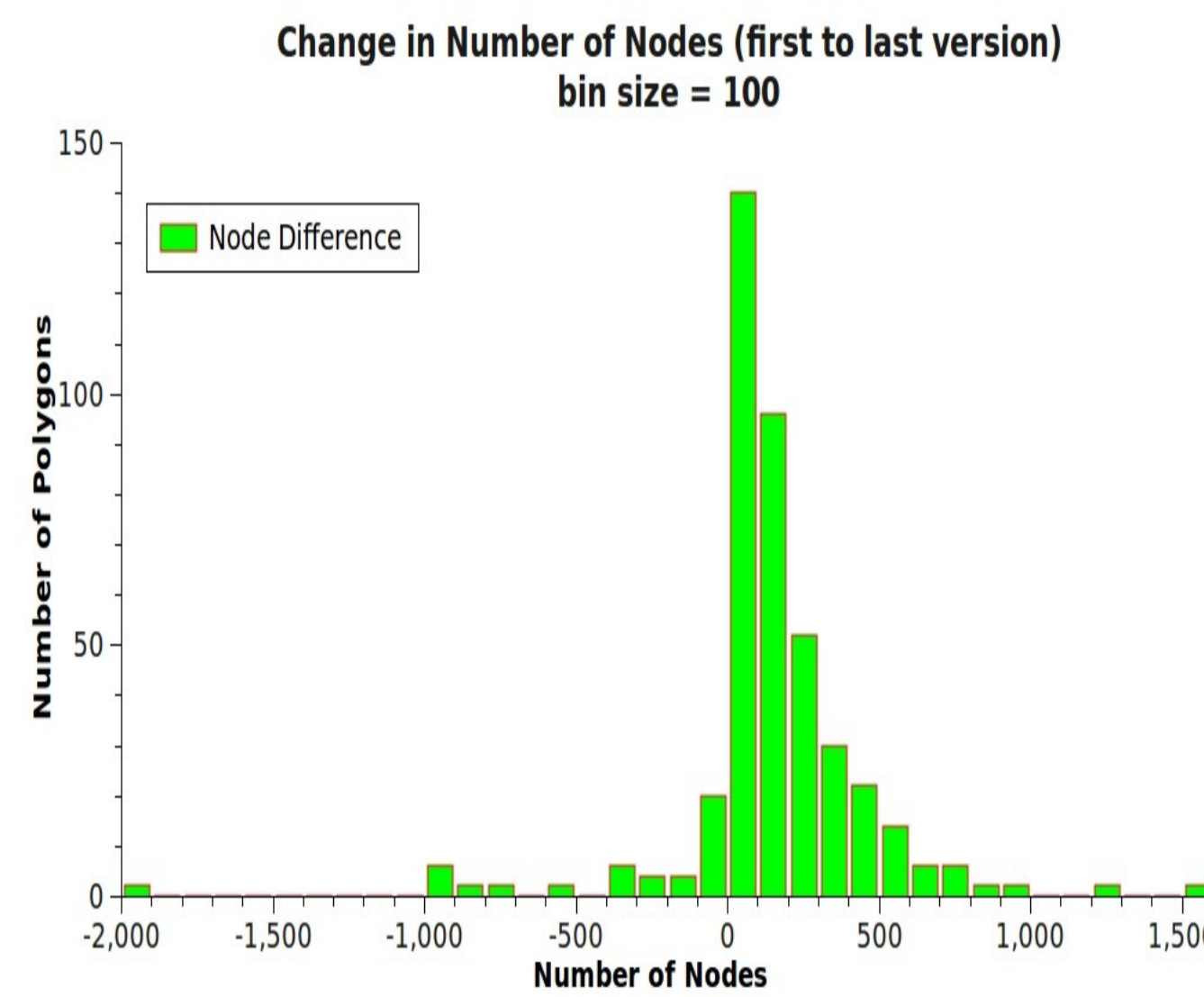
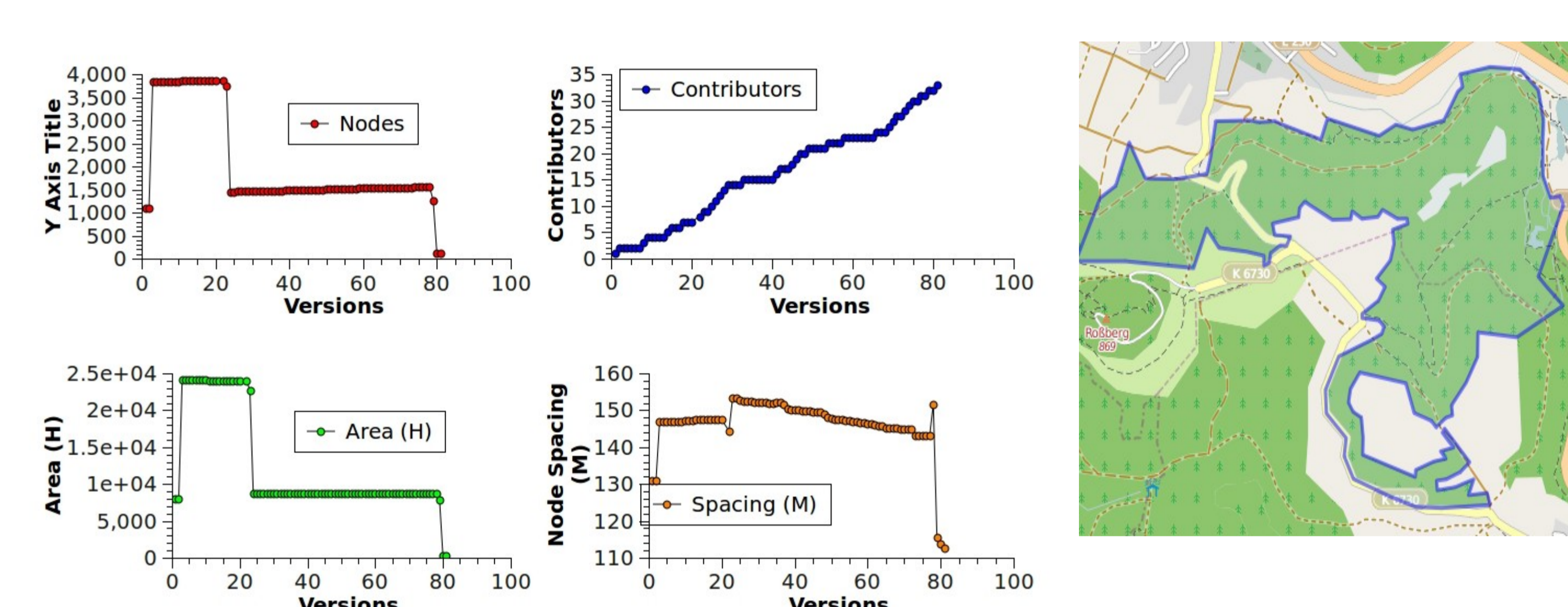
Goals of this research

- Easily computable metrics for assessment of the quality of OpenStreetMap (OSM) data
- Initial research concentrates on polygons only!
- The metrics should include OSM specific measures: that is unsupervised quality analysis of OSM
- Provide a feedback mechanism for the OSM community to improve overall data quality

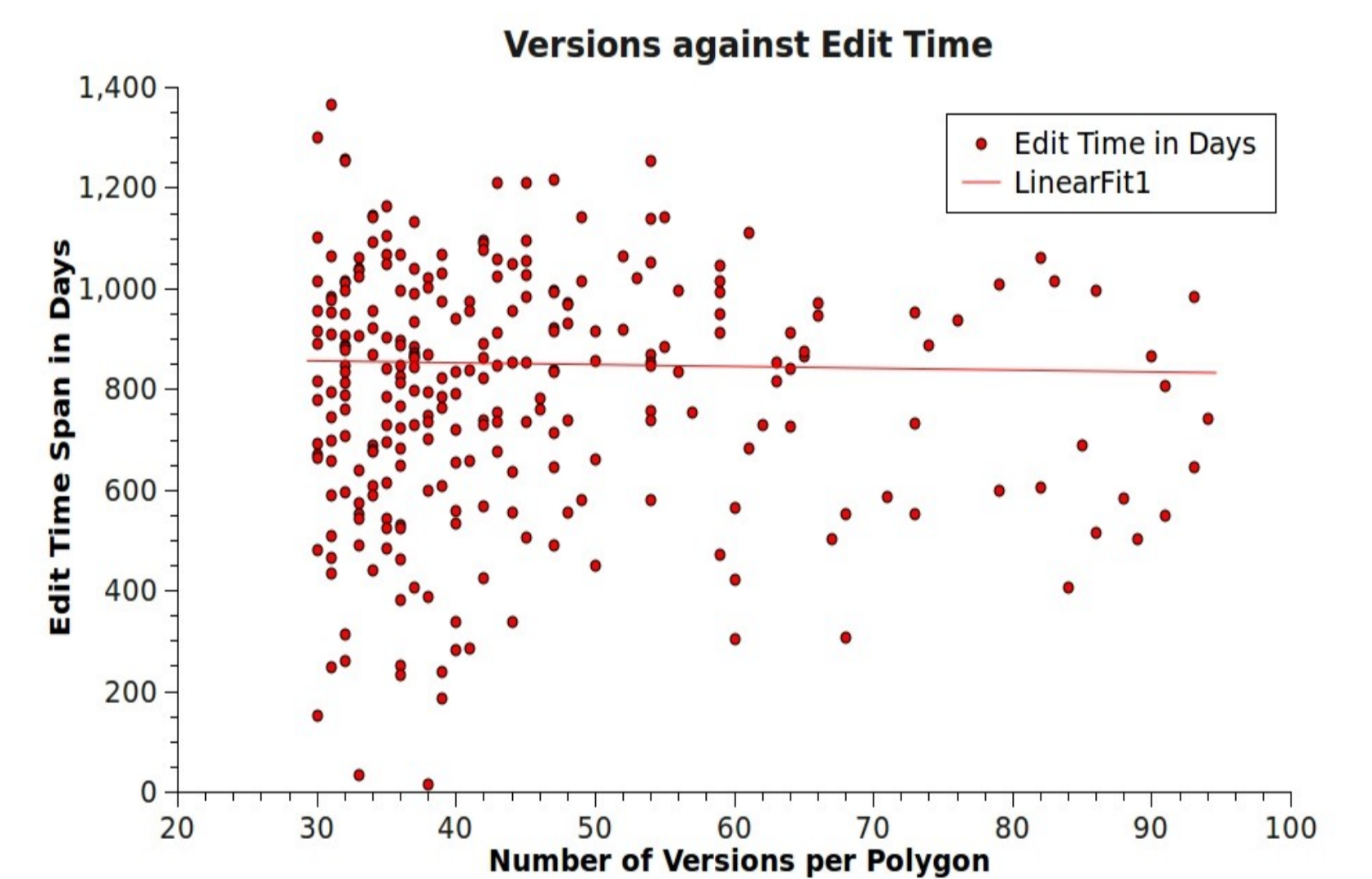
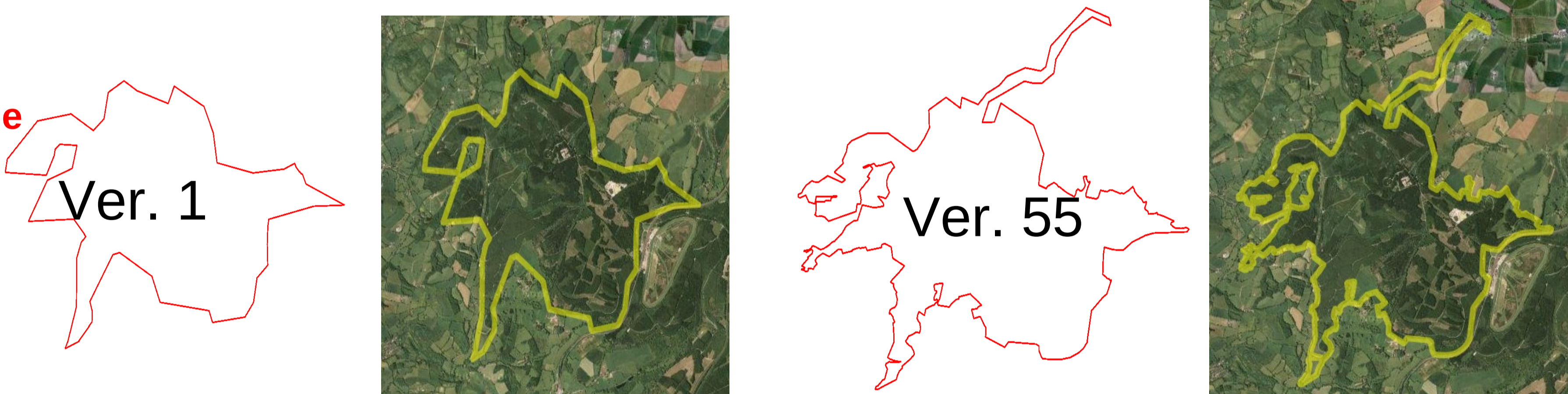
```

Data: OSM-ID of polygon P for analysis
Result: GPX files and Data files for each version v_i of P. Summary analysis of historical changes to polygon P
1 Using wget or curl download history XML file using OpenStreetMap API
2 Parse the history XML file into objects for each version v_i of P
3 for v_i in P do
4   for v_j in v_i do
5     check if n_j exists global node object list
6     if inGlobalList(n_j) then
7       Node n_j has been processed already
8     else
9       download n_j from OSM API
10      success = downloadFromOSM = API(n_j)
11      if success then
12        n_j exists in OSM API
13      else
14        this is a deleted node
15        parse from OpenStreetMap webpages
16        parseWebOSM(n_j)
17      end
18      add to node object list for v_i
19    end
20  end
21 end
22 finished with v_i of P
23 calculate stats/characteristics for v_i
24 insert v_i into PostGIS
25 write DAT,GPX, KML, SHP file for v_i
26 end
27 Write History Summary File for P
    
```

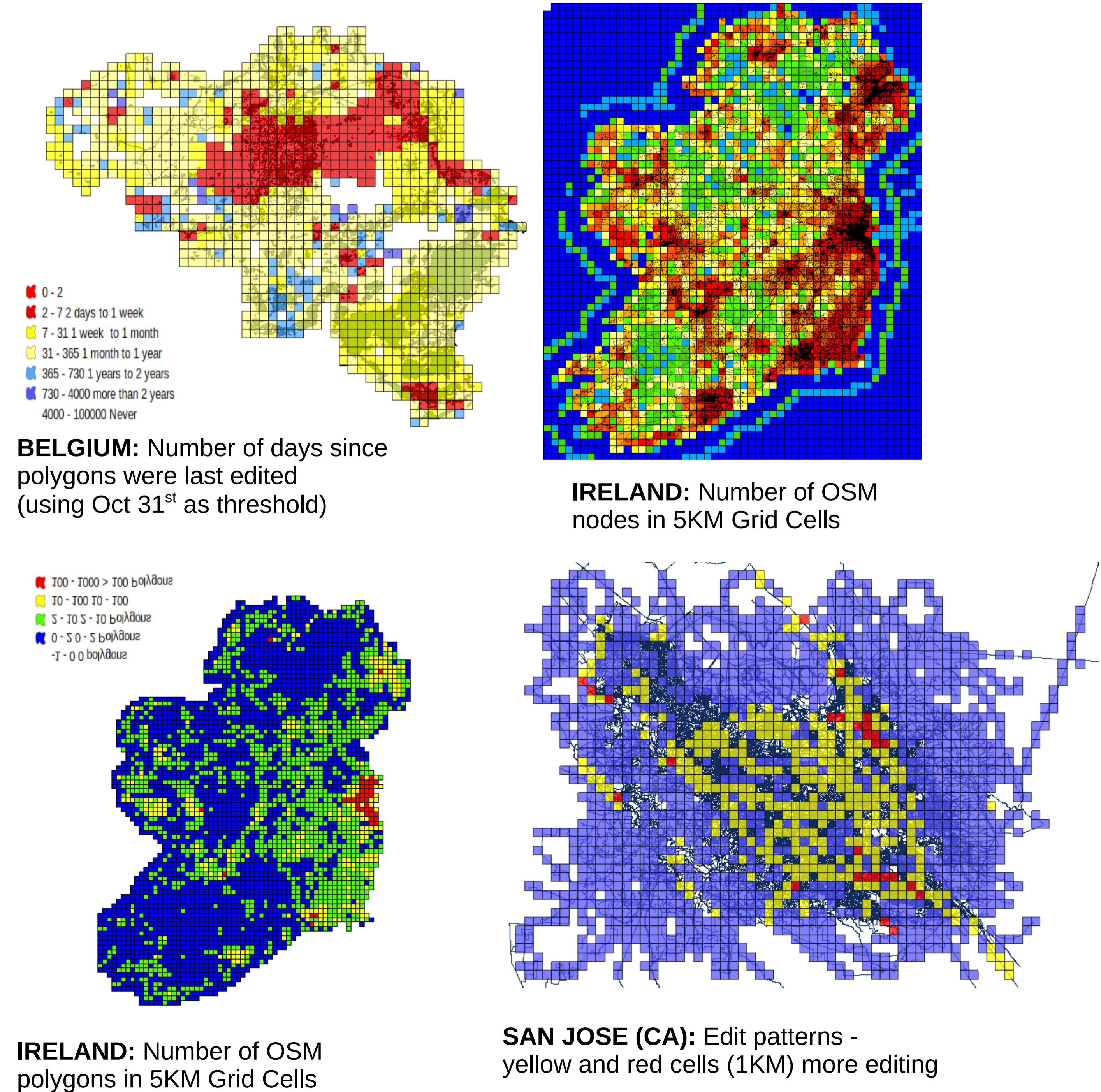
Example of a volatile polygon evolution (ID 26164873)



EXAMPLE: Polygon Change (ID 20142797) 55 Versions



Visualisation of OpenStreetMap spatial data homogeneity



Overall Observations and Conclusions

Given the diversity in quality and representation it is difficult to make definitive statements (quantitative measures) about the quality of OSM natural features (Mooney et al, 2010)

“While professional data is not without it’s faults the coverage of OSM in rural areas is too small to be seriously considered a sophisticated alternative for any application” (Zielstra and Zipf 2010)

“Regional differences in OSM map coverage and accuracy make it difficult to make general statements about the suitability of OSM as a replacement for professionally captured and edited data” (Over et al, 2010)

“All quality considerations are use-case sensitive in VGI” (USGS, 2010)



1. Department of Computer Science, National University of Ireland Maynooth (NUIM), Co. Kildare. Ireland
 Email: peter.mooney@nuim.ie
 Web: <http://www.cs.nuim.ie/~pmooney>

2. Environmental Research Center, Environmental Protection Agency, Richview, Clonskeagh, Dublin 14. Ireland
 +353 (1) 268 0100 p.mooney@epa.ie