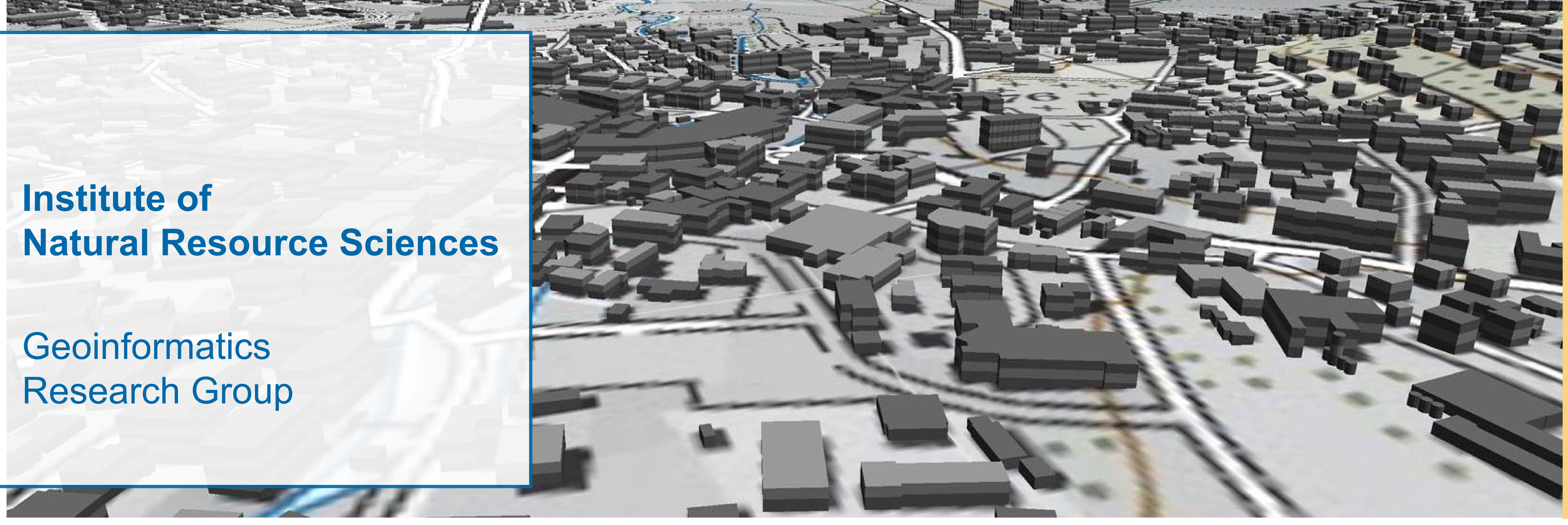
Zurich University
of Applied Sciences
zh
aw
Institute of
Natural Resource Sciences
Geoinformatics
Research Group



modularCity

# Multi-criteria aggregation for sensitive parcel-based census data

**S. Indermühle[1], P. Laube[1], M. Geilhausen[1], T. Zwicker[2 3]**

[1] Zurich University of Applied Sciences ZHAW
Grüental, 8820 Wädenswil, Switzerland
{indu; patrick.laube; martin.geilhausen}@zhaw.ch

[2] tsquare gmbh
Arbergstr. 1, 8405 Winterthur, Switzerland
tzwicker@tsquare.ch

[3] City council Männedorf
Bahnhofstrasse 10, 8708 Männedorf,
Switzerland

## Problem Statement

Modern urban planning considers various socio-economic issues and hence requires fine-grained, potentially sensitive spatial data, raising the *disclosure dilemma*: Whilst it is imperative to aggregate fine-grained sensitive records large enough to avoid disclosure, aggregation blurs the information, potentially hindering insights about spatially explicit relationships. MASC is an aggregation algorithm for sensitive data, balancing five aggregation criteria.

```
foreach reference area p in list do
    difference = INFINITY;
    partner = null;
    foreach neighbor q of p do
        if difference(p, q) ¡ difference then
            difference = difference(p, q);
            partner = q;
        end
    end
    list.remove(p);
    list.remove(partner);
    p.mergeWith(partner);
    if !p.isAnonymous() then
        list.add(p);
    end
end
```

## The MASC algorithm

For each parcel, a difference value is calculated for all neighbors, based on five parameters (see below). Weighting the parameters (slider) allows scenario-based planning. The neighbor with the lowest difference, or the highest similarity, is selected as merging partner. Parcels are merged until they reach an anonymity threshold.

*resident density* | *building regulations* | *distance of centers* | *distance between polygons* | *anonymity penalty*

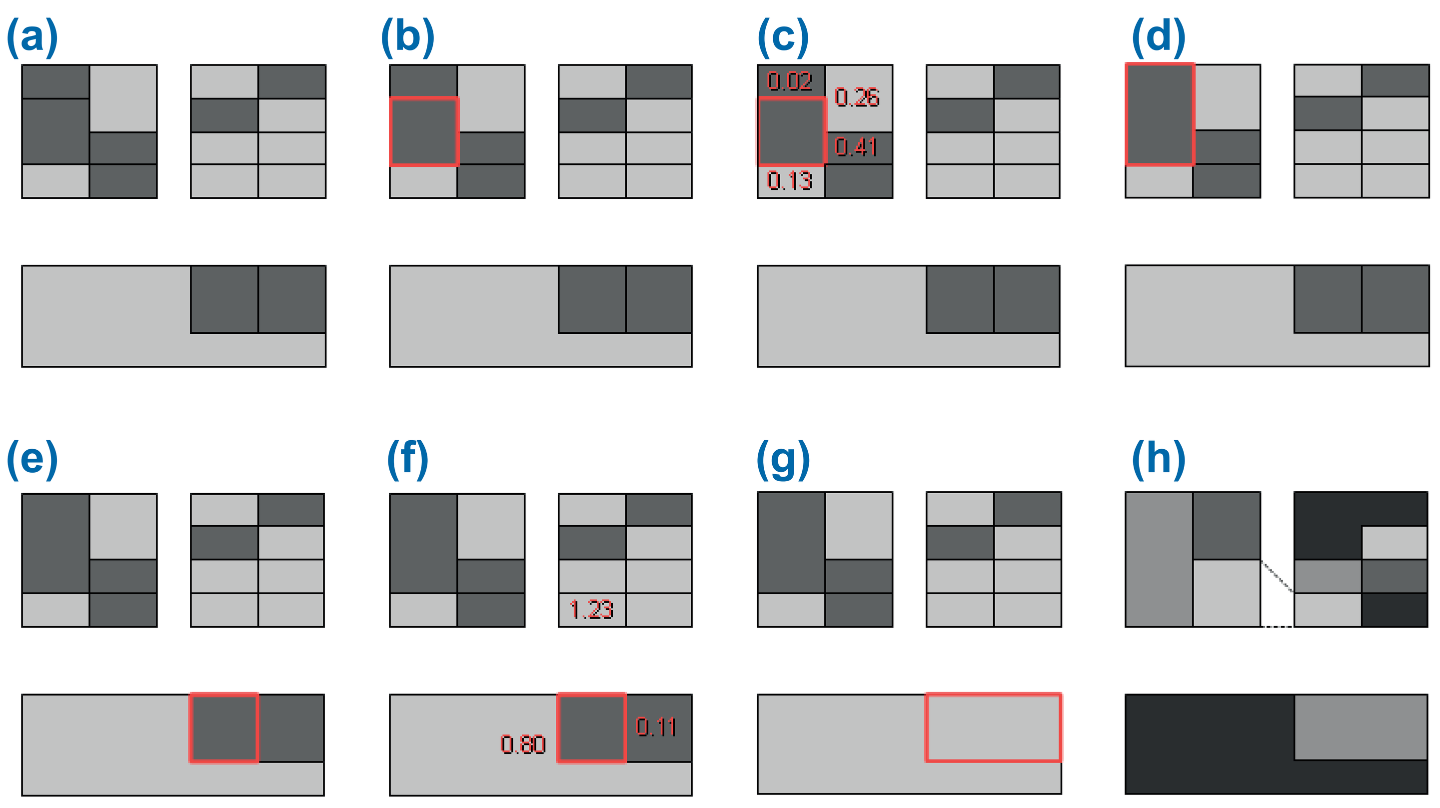



**Figure 1**

MASC aggregation: (a) all reference areas that are not anonymous (in dark grey) are identified; (b) a first reference area is selected; (c) the difference value with each neighbor is calculated; (d) the reference area is merged with the best match. The resulting reference area is not yet anonymous; (e) the next reference area is selected; (f) the difference value of each neighbor is calculated; (g) after merging with the best match the reference area is anonymous; (h) the 10 anonymous regions resulting from the algorithm.

## Case Study

The algorithm is validated using two case studies, Langenthal and Männedorf. The most effective parameter for crystallizing the important socio-economic structure has shown to be resident density (Fig. 2b). Even without any additional information e.g. building type), a high weight of resident density results in zones of the same residential type. By contrast, when accentuating the building regulation parameter (Fig. 2c), the zones are aggregated within respective building zones.

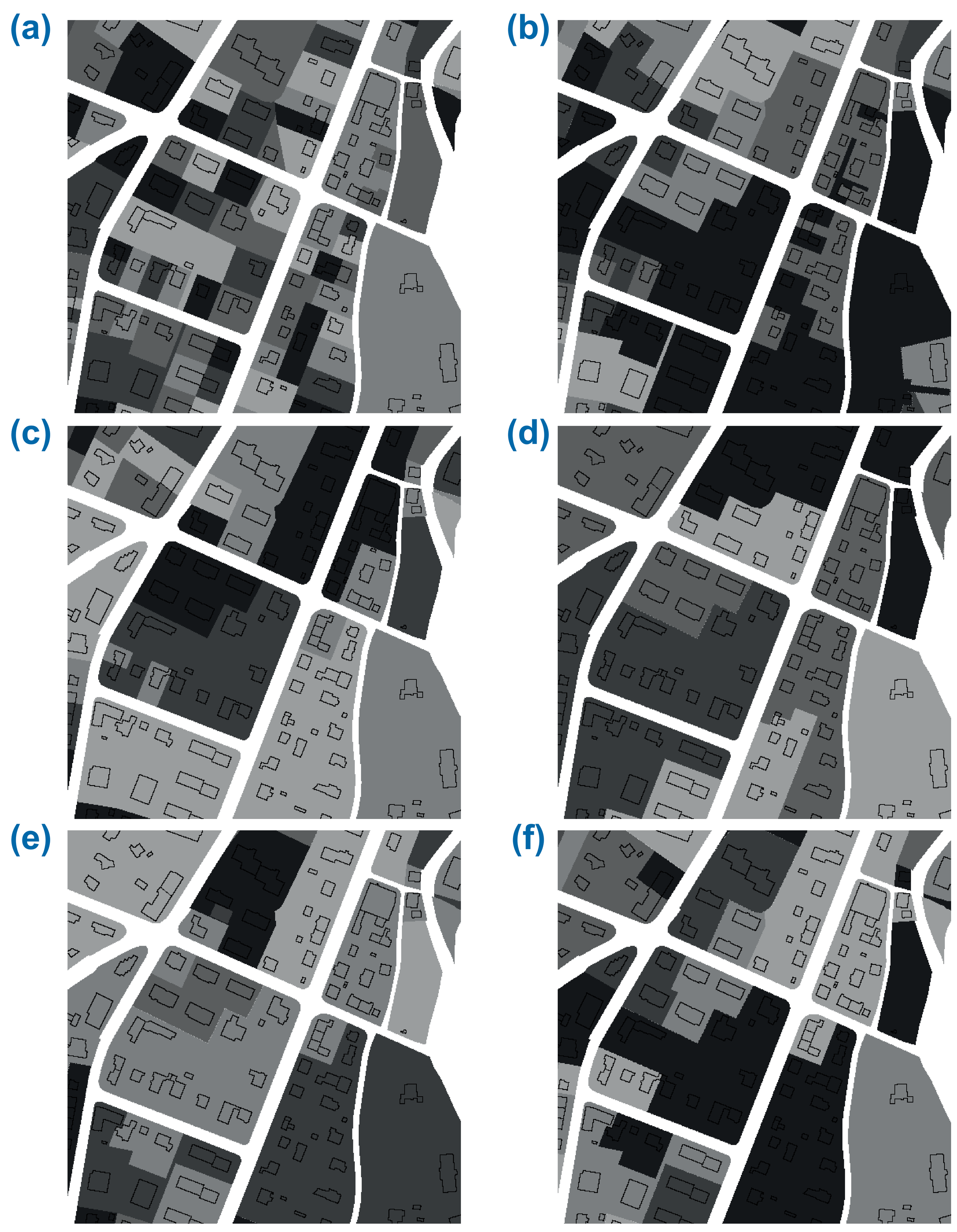



**Figure 2**

Different aggregation scenarios, shown for the case study of Langenthal. *n* was set to 30 residents (a) initial parcels that are used as the input; (b) result based on resident density only; (c) result based on floor area ratio only; (d) results based on geometric distances only; (e) result based on the following exemplary weights: density 1, floor area ratio 0.5, center distances 2.3, polygon distances 2.3, anonymity penalty 2; (f) result based on the following exemplary weights: density 1, floor area ratio 1, center distances 2, polygon distances 0, anonymity penalty 1.

## Validation and Conclusions

The MASC aggregation procedure was evaluated through qualitative plausibility tests with local decision makers. In planning workshops, the stakeholders explicitly noticed that MASC can uncover the fine-grained socio-economic structures whilst safeguarding confidentiality regulations. Manual adjustments were only suggested for a very small number of parcels (around 1% of all parcels).

Given its weighted multi-criteria character, MASC empowers the stakeholders to experiment with different planning scenarios adhering to different priorities (different weights).

## Acknowledgements

This research was supported by Swiss Commission for Technology and Innovation CTI, project 12910.1 PFES-ES, “modularCity – Software zur Unterstützung nachhaltiger Stadtplanung” and the municipality of Langenthal. The authors furthermore thank Dr. Markus Zahnd, City of Langenthal, for his support.

Poster presented at 8th *Int. Conf. on Geographic Information Science* (GIScience), Vienna - September, 23-26, 2014

GIScience 2014

## References

[1] Armstrong M P, Rushtong G and Zimmerman D L, 1999, Geographically masking health data to preserve confidentiality. *Statistics in Medicine*, 18(5):497–525.

[2] Assunção R M, Neves M C, Câmara G and Da Costa Freitas C, 2006, Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees, *International Journal of Geographical Information Science*, 20(7):797–811.

[3] Cockings S, Harfoot A, Martin D and Hornby D, 2013, Getting the foundations right: spatial building blocks for official population statistics. *Environment and Planning A*, 45(6):1403–1420.

[4] Joung C, Martins D and Skinner C, 2009, Geographically intelligent disclosure control for flexible aggregation of census data. *International Journal of Geographical Information Science*, 23(3-4):457–482.

[5] Shortt N, 2009, Regionalization/zoning systems. In: Kitchin R and Thrift N (eds), *International Encyclopedia of Human Geography*, Elsevier, Oxford, 298–301.