Processing large datasets into a vario-scale structure

Martijn Meijers

Faculty of Architecture and the Built Environment,

OTB Research, GIS technology

Feb 12, 2014 – Gemeente Den Haag – Gebruikerscommissie STW

Introduction

- Obtaining a vario-scale dataset: merge and split/collapse areas, simplify boundaries

Introduction

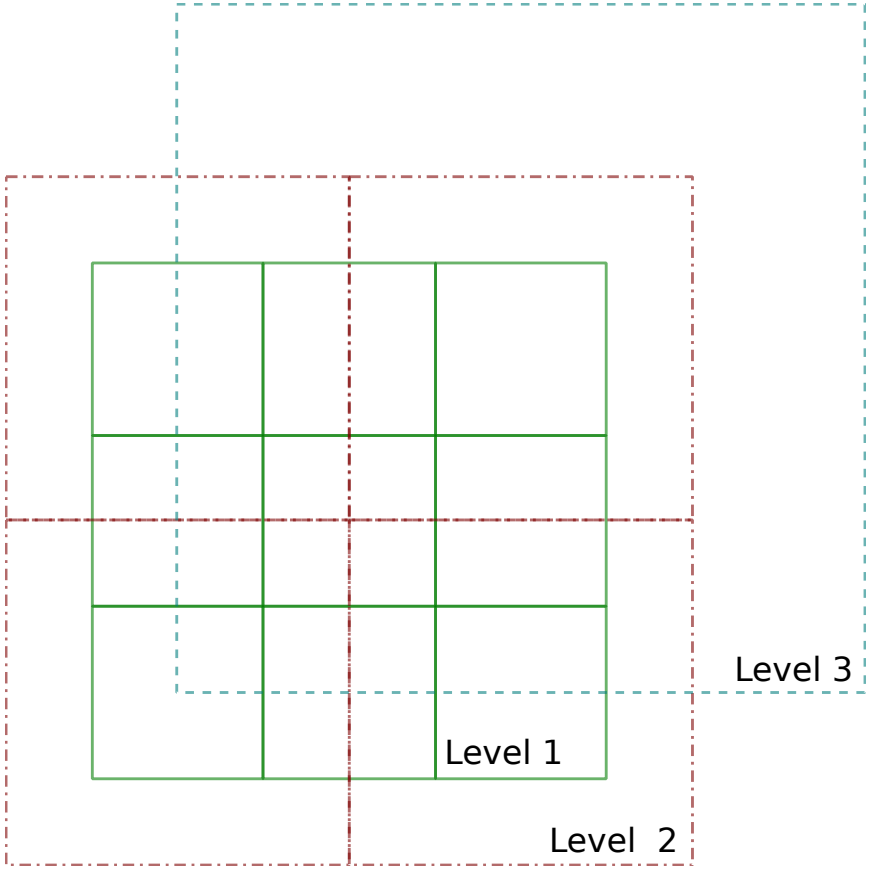
- Obtaining a vario-scale dataset: merge and split/collapse areas, simplify boundaries
- Limited to around 50 000 input areas
- Does not fit in main memory

Introduction

- Obtaining a vario-scale dataset: merge and split/collapse areas, simplify boundaries
- Limited to around 50 000 input areas
- Does not fit in main memory
- Split input in smaller pieces (divide) and union results (conquer)

Split in pieces

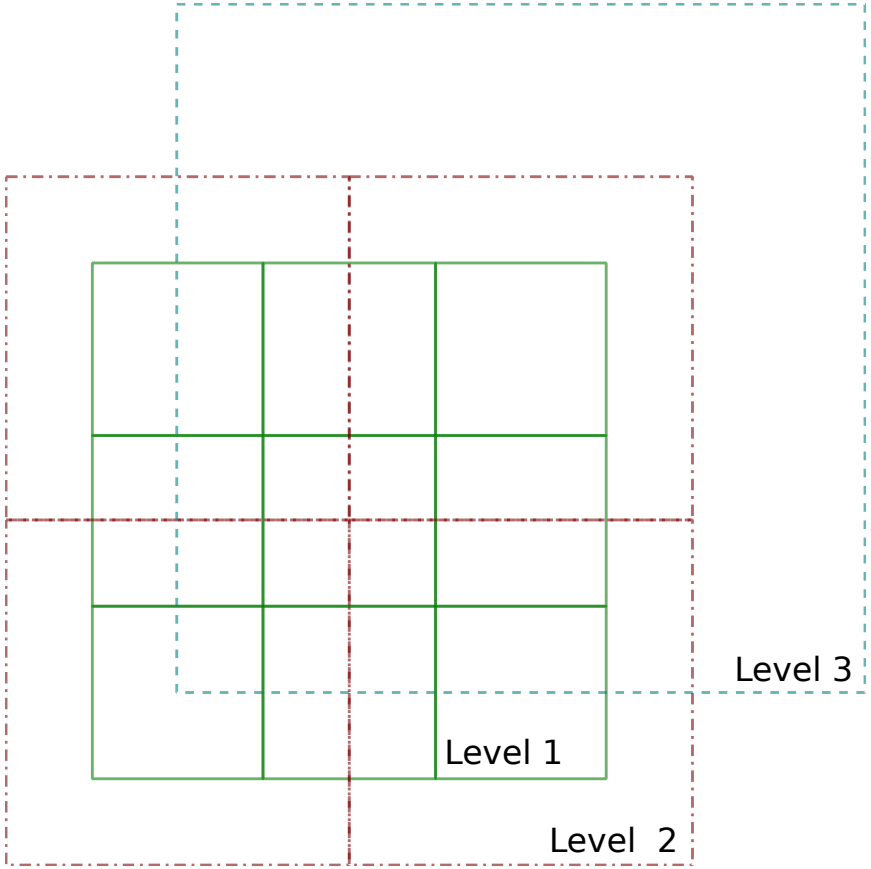
Use partition fieldtree (Frank and Barrera, 1990)



Split in pieces

Use partition fieldtree (Frank and Barrera, 1990)

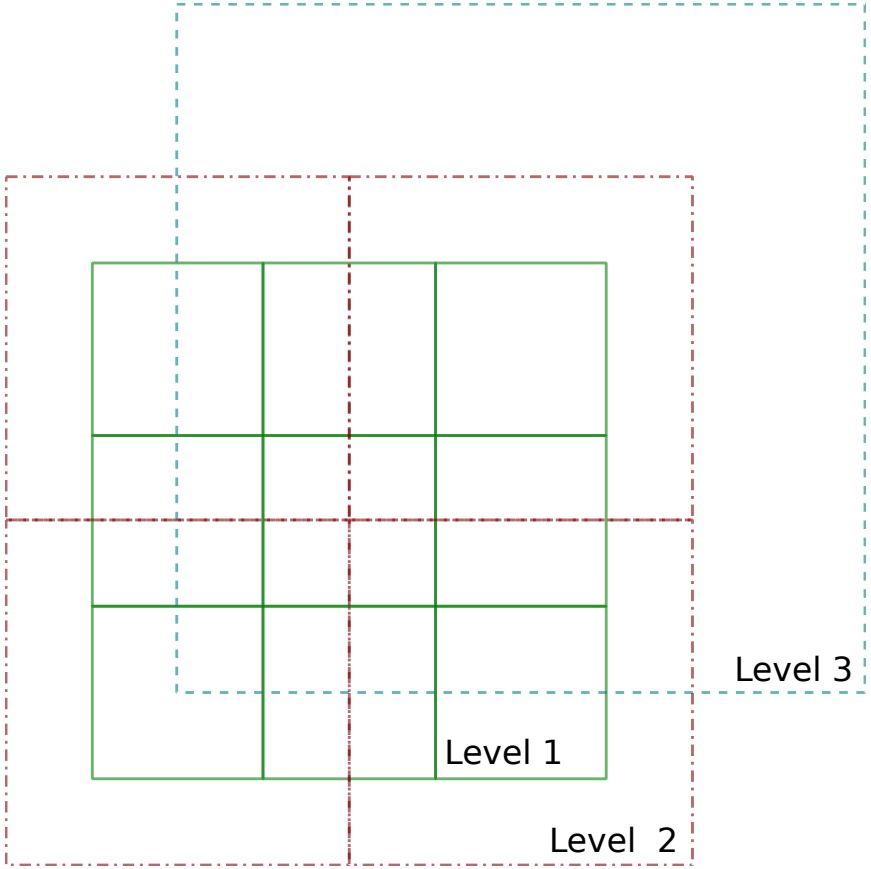
- Tree of fields (a field is a box)



Split in pieces

Use partition fieldtree (Frank and Barrera, 1990)

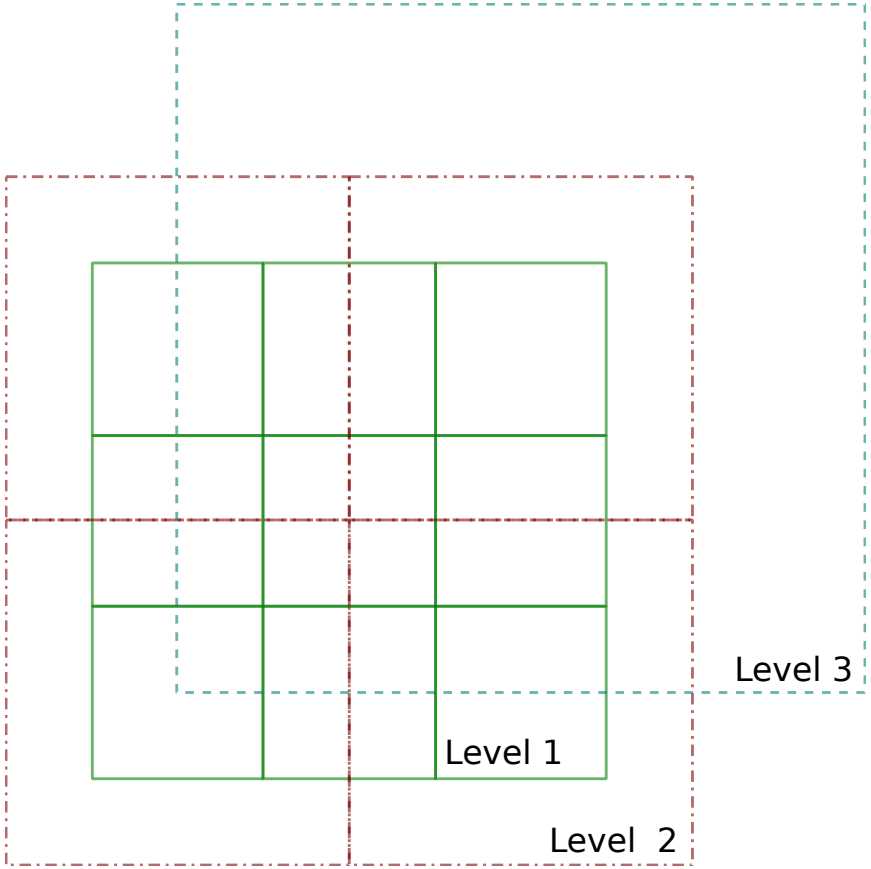
- Tree of fields (a field is a box)
- Start with a field around dataset



Split in pieces

Use partition fieldtree (Frank and Barrera, 1990)

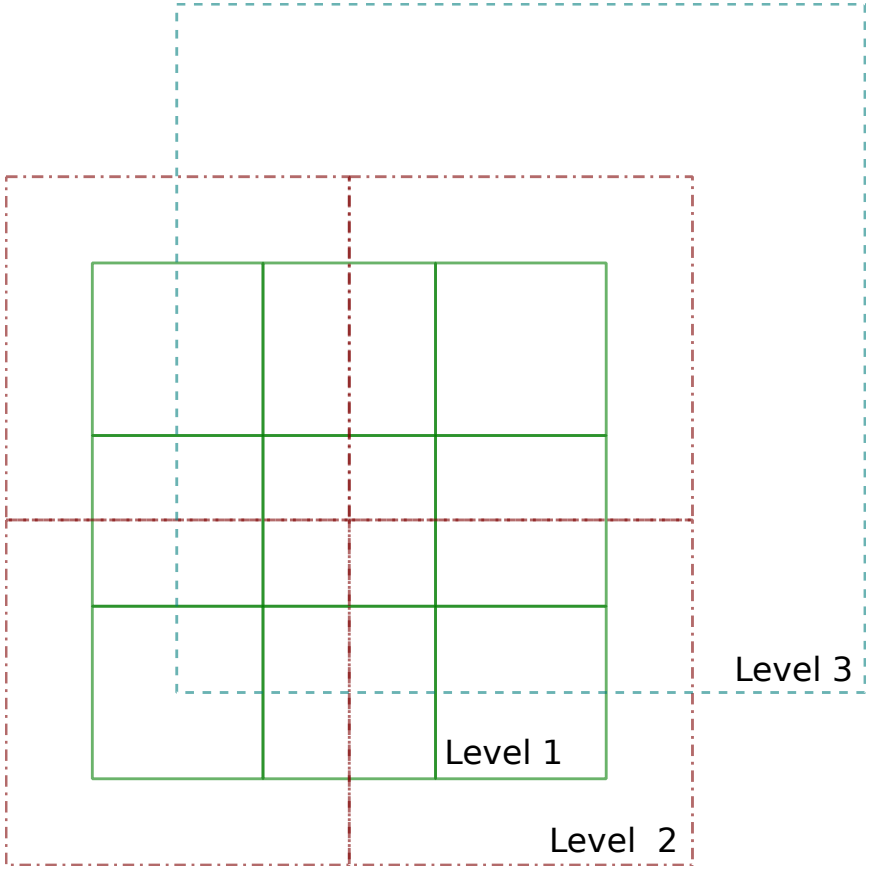
- Tree of fields (a field is a box)
- Start with a field around dataset
- Divide field in 4 and shift ($-\frac{1}{2}$ width, $-\frac{1}{2}$ height), repeat until field has reached a certain size



Split in pieces

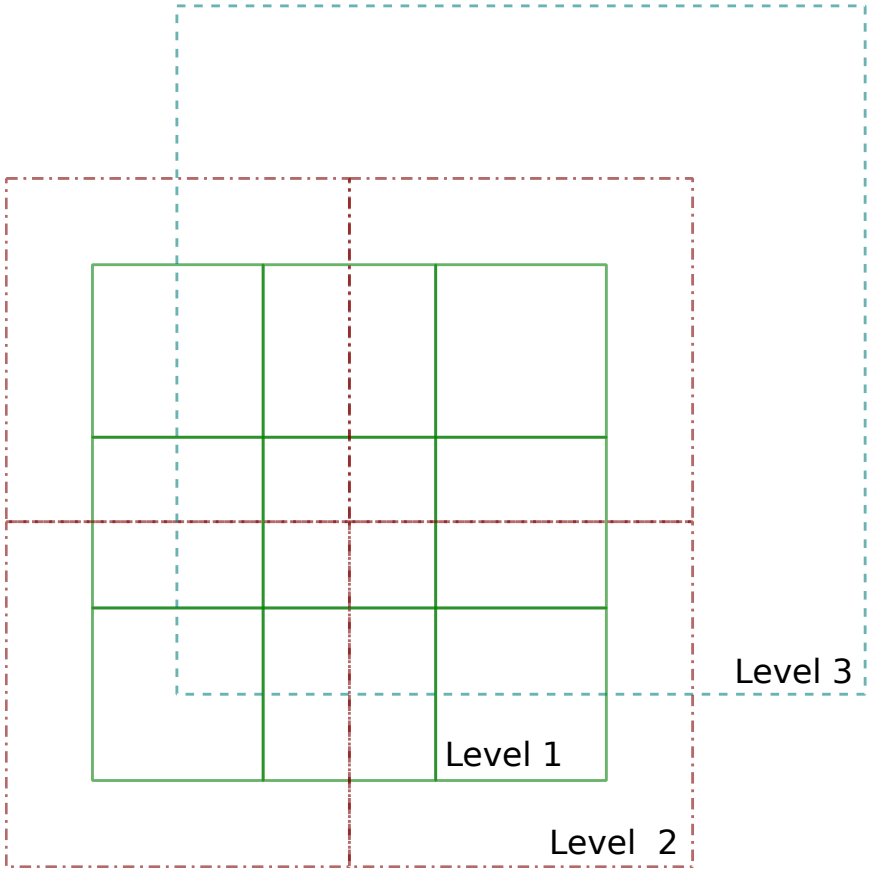
Use partition fieldtree (Frank and Barrera, 1990)

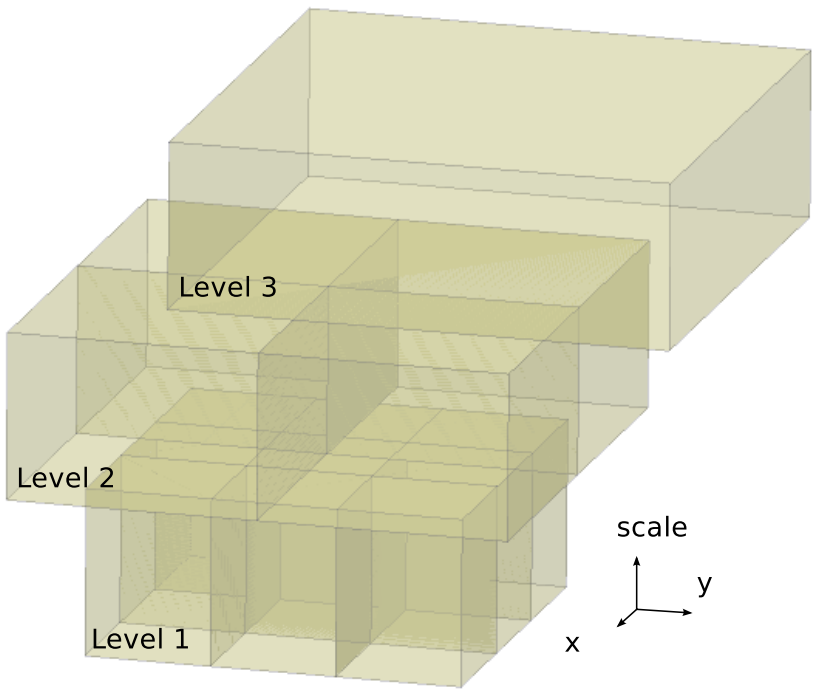
- Tree of fields (a field is a box)
- Start with a field around dataset
- Divide field in 4 and shift ($-\frac{1}{2}$ width, $-\frac{1}{2}$ height), repeat until field has reached a certain size
- Note, every field has (upto) 9 children, and (upto) 4 parents



Split in pieces

Alternative view of fieldtree – Scale = 3rd dimension





Processing large datasets – Method

Steps:

1. Obtain fieldtree structure

Processing large datasets – Method

Steps:

1. Obtain fieldtree structure
2. Per field make partial tGAP structure (a node, edge, face & face hierarchy table per field)

Processing large datasets – Method

Steps:

1. Obtain fieldtree structure
2. Per field make partial tGAP structure (a node, edge, face & face hierarchy table per field)
3. Distribute source data over fields (primitives are stored with smallest field in which they fit completely)

Processing large datasets – Method

Steps:

1. Obtain fieldtree structure
2. Per field make partial tGAP structure (a node, edge, face & face hierarchy table per field)
3. Distribute source data over fields (primitives are stored with smallest field in which they fit completely)
4. Fill partial tGAP structure for every field, i. e. run generalization process (merge areas/simplify boundaries)

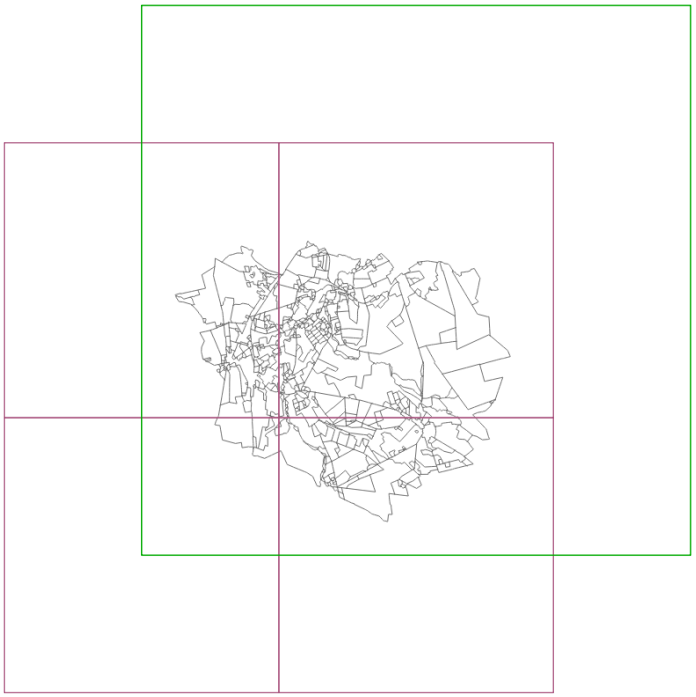
Processing large datasets – Method

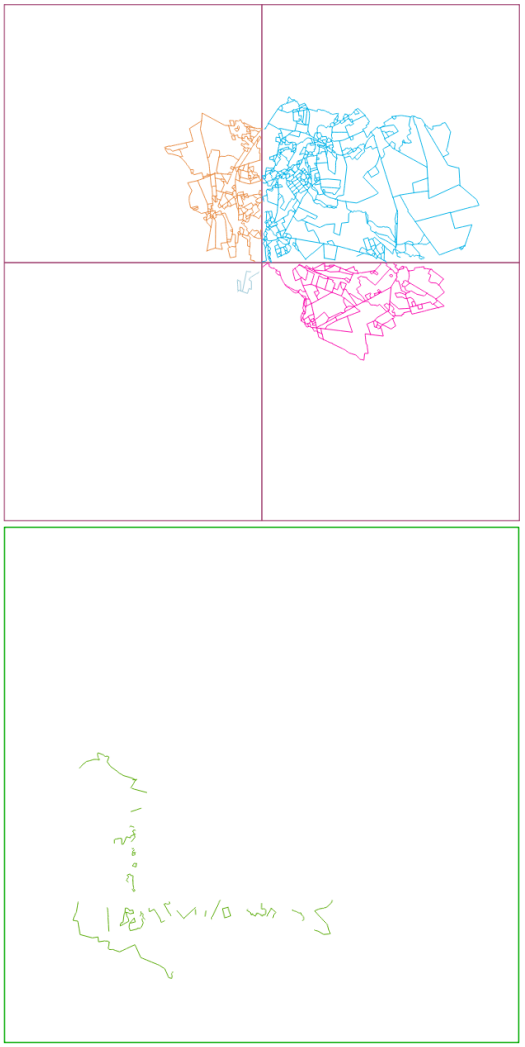
Steps:

1. Obtain fieldtree structure
2. Per field make partial tGAP structure (a node, edge, face & face hierarchy table per field)
3. Distribute source data over fields (primitives are stored with smallest field in which they fit completely)
4. Fill partial tGAP structure for every field, i. e. run generalization process (merge areas/simplify boundaries)
5. When field is processed a field contains less detailed data – input for parent field at next level

Processing large datasets – Example

Split input data over fields



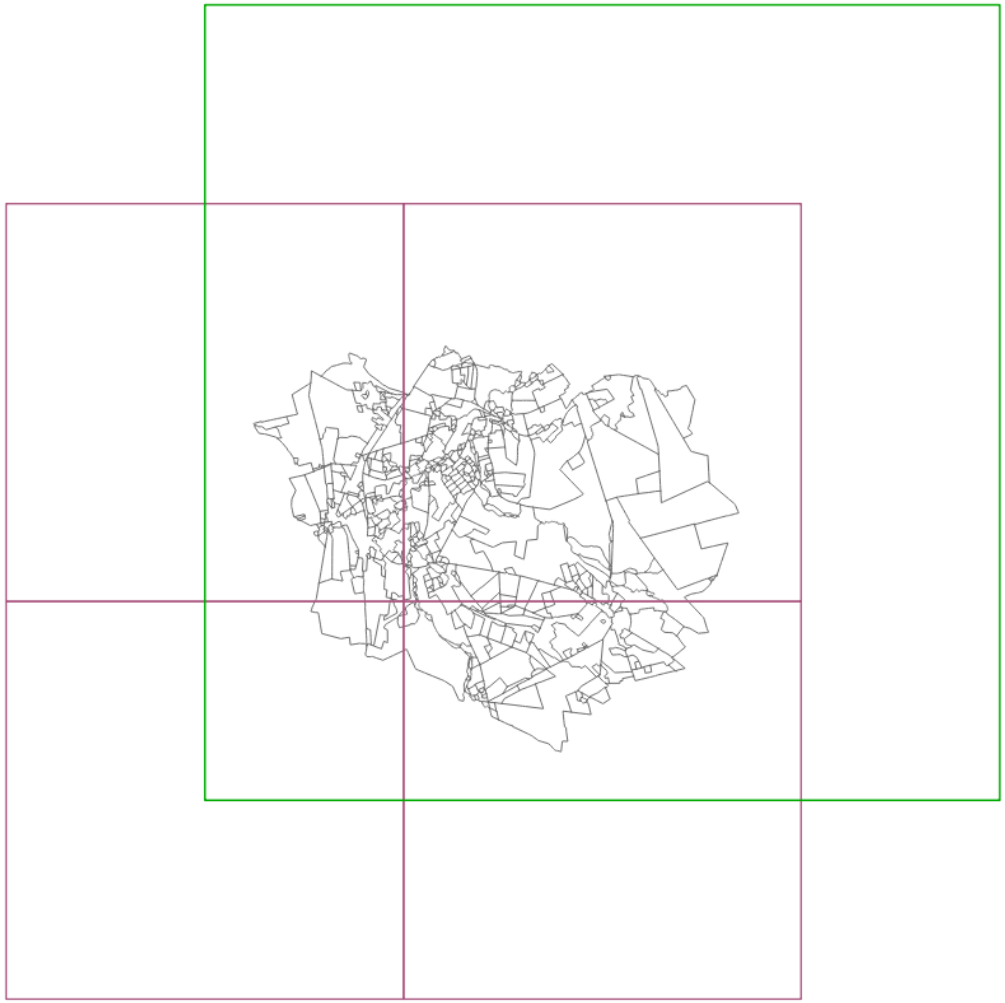


Processing large datasets – Example

Create partial tGAP structure (field by field, level by level)

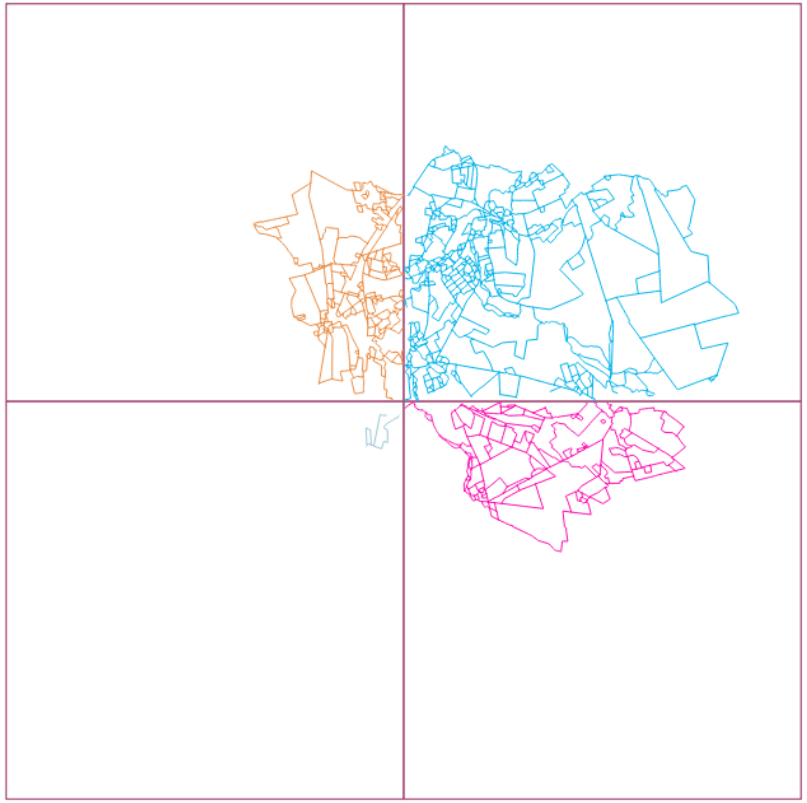
Processing large datasets – Example

Create partial tGAP structure (field by field, level by level)



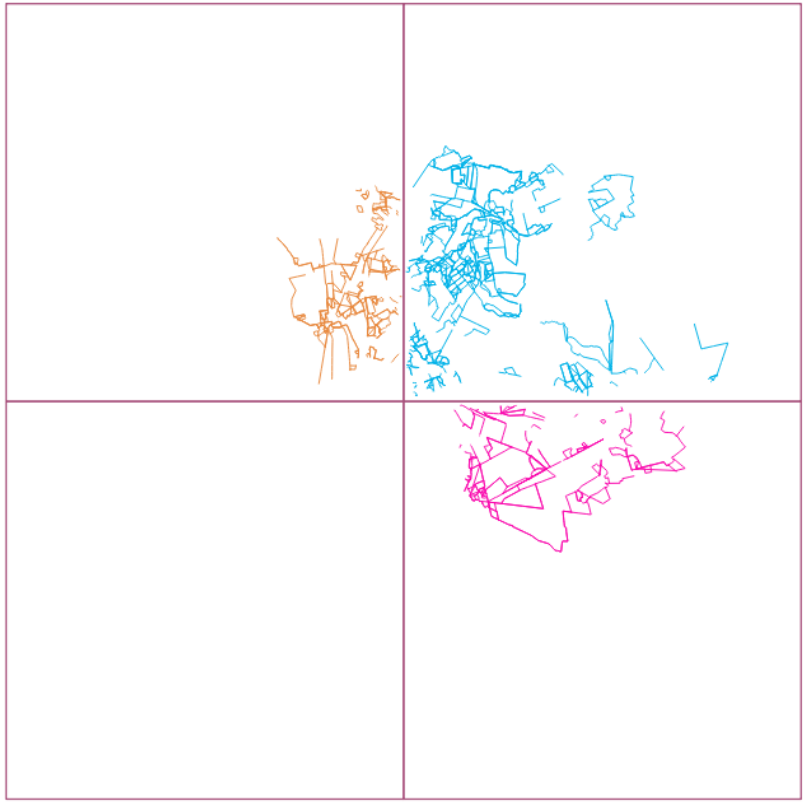
Processing large datasets – Example

Create partial tGAP structure (field by field, level by level)



Processing large datasets – Example

Create partial tGAP structure (field by field, level by level)

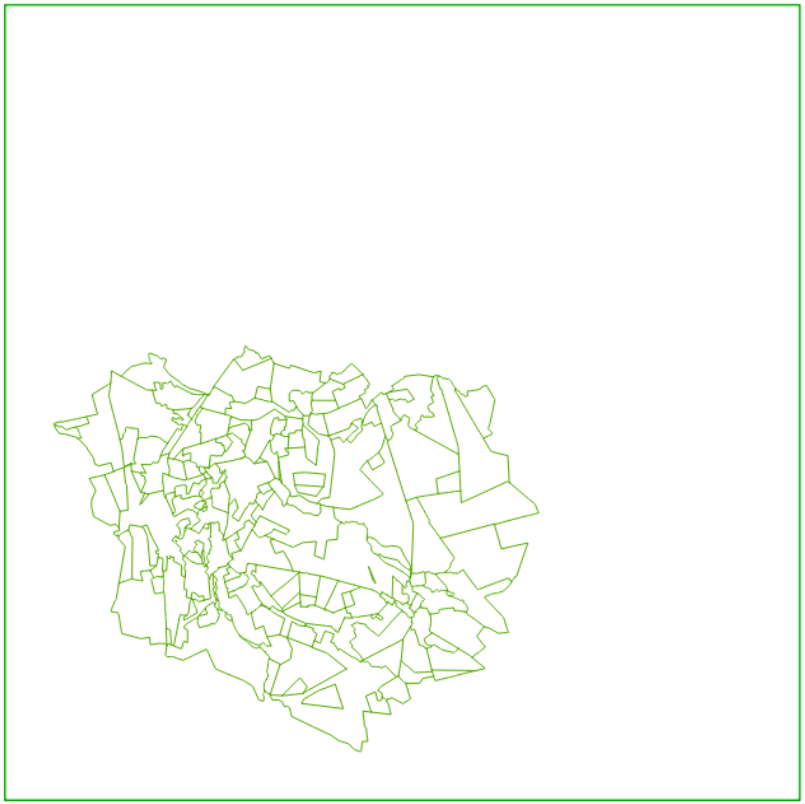


Processing large datasets – Example

Create partial tGAP structure (field by field, level by level)

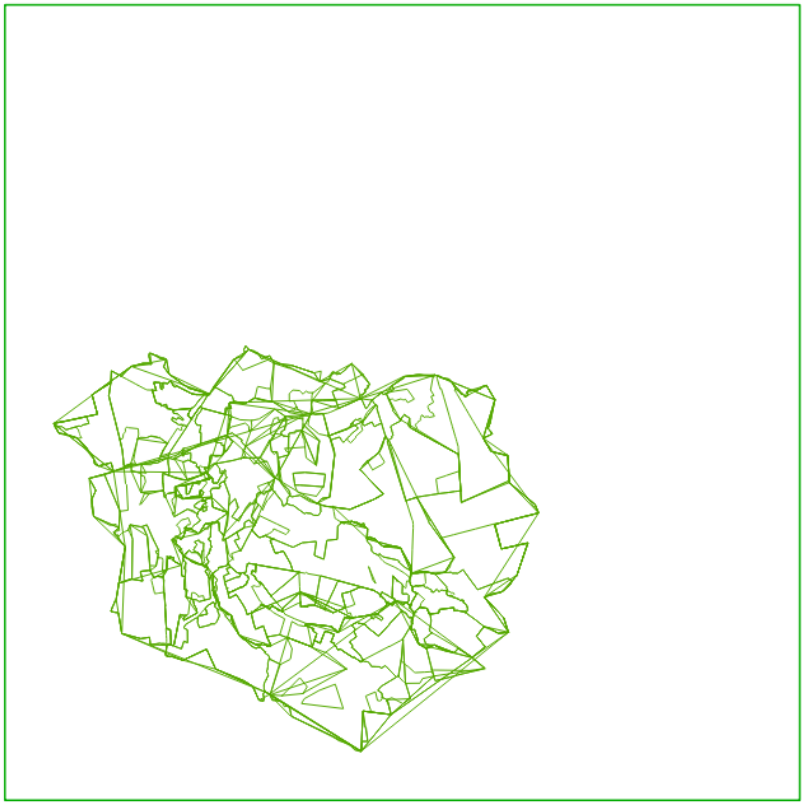
Processing large datasets – Example

Create partial tGAP structure (field by field, level by level)



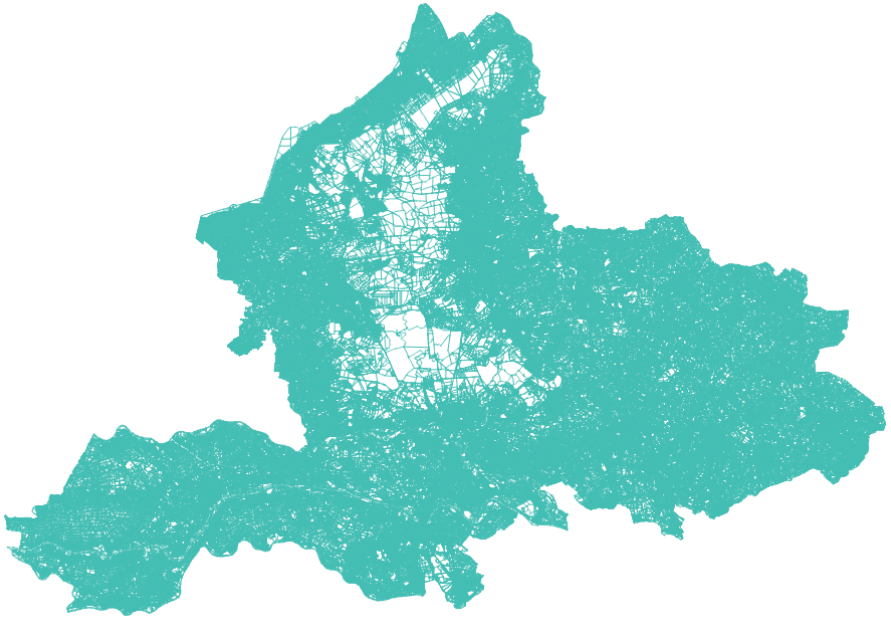
Processing large datasets – Example

Create partial tGAP structure (field by field, level by level)



Results

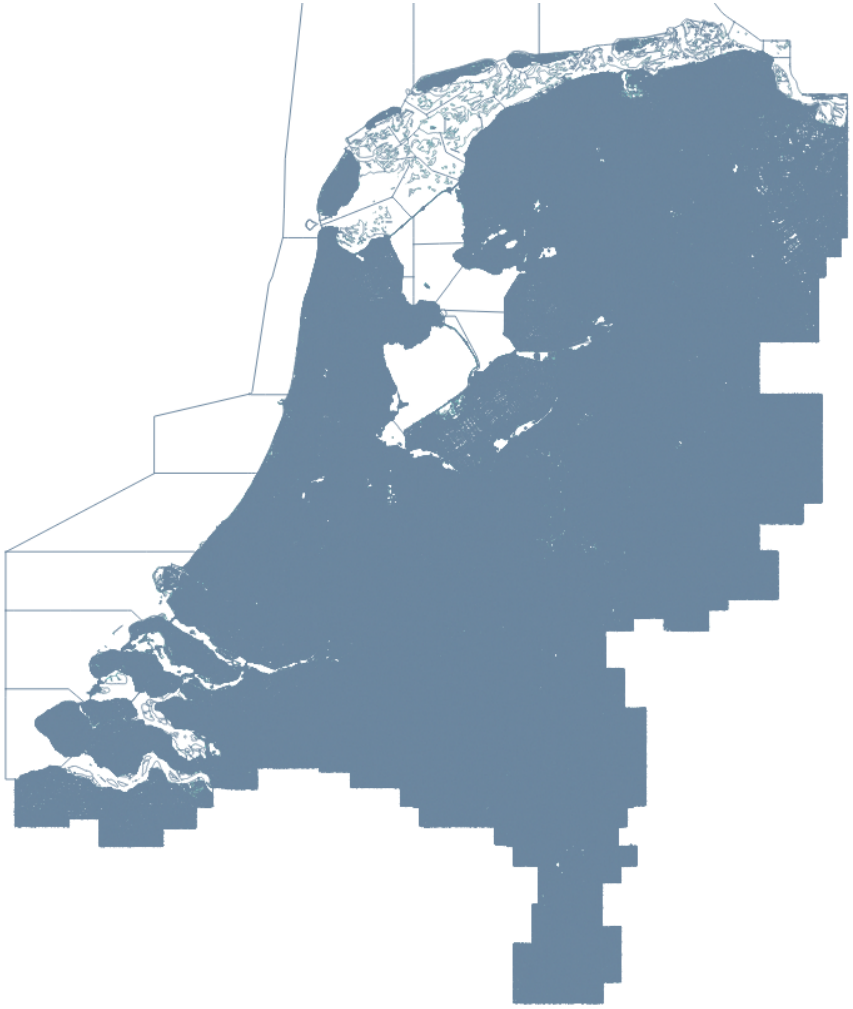
- Initial implementation finished (January 2014)
- Small bugs: at the borders of the fields and propagating to next level (setting scale range for primitives)
- Able to process dataset with 50 000 cadastral parcel polygons, trying with 300 000 & 800 000



Future work

Short term

- Fix issues
- Make work in parallel (multiple fields at same time)
- Make split/collapse operation work again
- Test with Top10NL (roads, water, terrain — 5.2 million area features)



Future work

Longer term

- Investigate use of fieldtree for:
  - Data retrieval with large dataset
  - Processing updates
- Integrate improvements of Radan/Lina (better quality maps/smooth transitions)

Thank you for your attention

- Delft University of Technology
  Faculty of Architecture and the Built Environment
  OTB Research
  GIS Technology

- dr. ir. Martijn Meijers
  b.m.meijers@tudelft.nl
  tel. (+31) 15 27 85642

References

Frank, A. and Barrera, R. (1990). The Fieldtree: A data structure for geographic information systems. In Buchmann, A., Günther, O., Smith, T., and Wang, Y.-F., editors, *Design and Implementation of Large Spatial Databases*, pages 29–44. Springer Berlin / Heidelberg.