Crowdsourcing ground-truth datasets for building change detection

Juste Raimbault[1,2,3,4] and Julien Perret[1,5]

01/09/2025

Conference on Complex Systems 2025

Session S1: Urban Complexity 1

[1]LaSTIG, IGN-ENSG-UGE
[2]CASA, UCL
[3]UPS CNRS 3611 ISC-PIF
[4]UMR CNRS 8504 Géographie-cités
[5]LaDéHiS, EHESS

ENSG Géomatique

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES

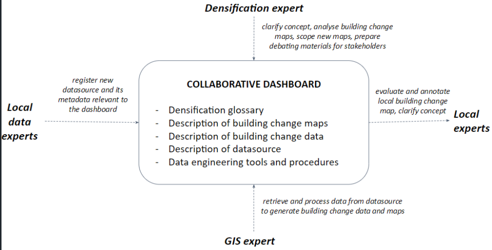
Dynamics of urban densification

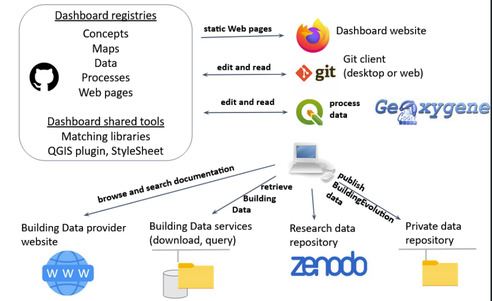
→ Understand urban densification dynamics for sustainable planning [Batty, 2018], with an integrated assessment of trade-offs between Sustainable Development Goals [Raimbault and Pumain, 2022] (linked to densification: increased access and limiteed sprawl vs UHI and less green space [Evers et al., 2024]).

→ In the suburban case, processes at multiple scales with diverse stakeholders; difficult to control through policy [Jehling et al., 2020]; the SUBDENSE European project to understand qualitatively and quantitatively these polyrationalities.

→ Need for a consistent, despite various GIS data qualities across compared countries (FR, UK, DE), quantification of urban change **at the building scale**.

A collaborative dashboard for mediation



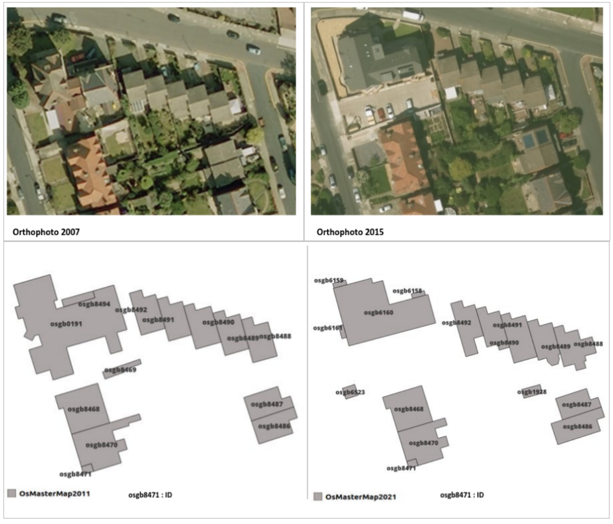


Preliminary work aimed at harmonising operationnal concepts and data sources qualification across the different countries:

→ git-based dashboard to share knowledge and resources between different types of experts [Bucher et al., 2025]

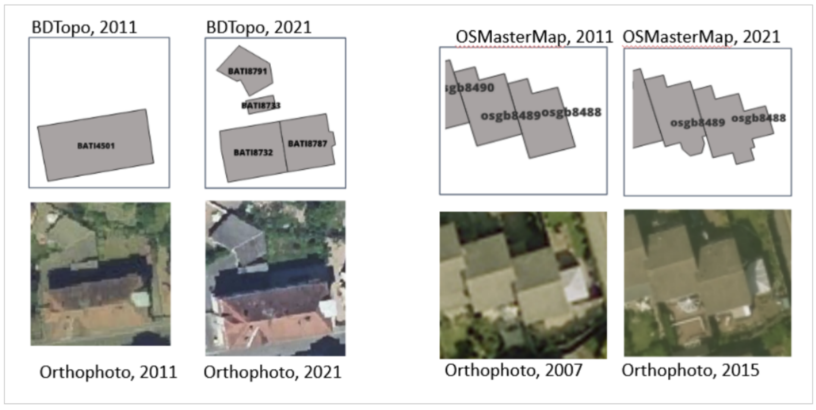
→ reproducible tools and methods, including **building change detection scripts**

Difficulties with data for building change



Example in the UK where changes in OSMasterMap building data includes both real changes and data changes. Source: [Bucher et al., 2025]

Difficulties with data for building change



Example of data specification changes only, for France (BDTopo) and the UK (OSMasterMap). Source: [Bucher et al., 2025]

Quantifying building change

→ Machine Learning on remote sensing raster data had improved performance in recent years, but remains not robust enough in terms of quality and resolution for that particular purpose [Khelifi and Mignotte, 2020]

→ Geospatial vector data matching algorithms as a method for change detection [Xavier et al., 2016], applied to linear data [Costes et al., 2015] and polygonal data [Gregory and Healey, 2007].

→ A new algorithm combining Geometric Matching of Areas [Bel Hadj Ali, 2001] with Multi-criteria Matching [Olteanu-Raimond et al., 2015] presented at CCS last year [Guardiola et al., 2024b].

→ Vector Matching algorithm require extensive parametrisation and optimisation, and provide varying performances in different context.

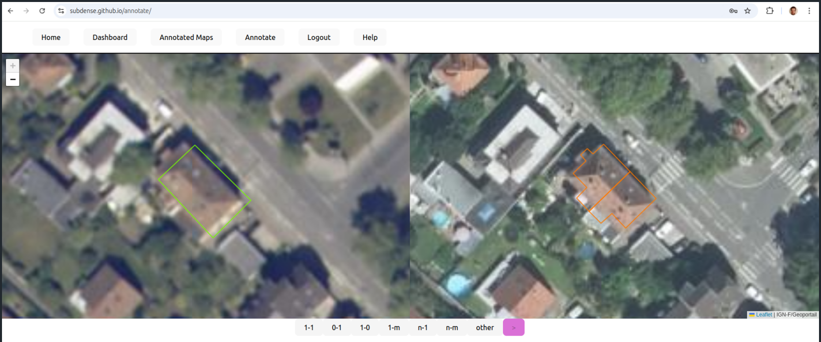
→ Need for large and specific ground-truth datasets (building layers annotated with expected links).

Research objective:

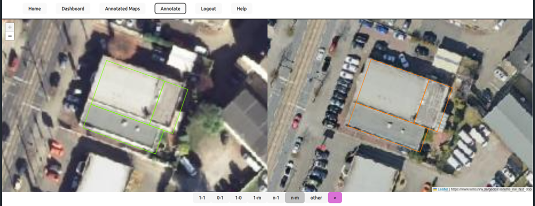
Design and develop a web application aimed at annotating potential matching links between two vector datasets.

Functionalities of the annotation web-application

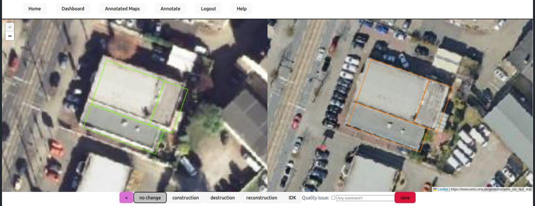
- fully based on git and javascript: no database nor server management, automatically deployed through github pages
- handles multi-user collaboration, with authentication given by writing rights on the git repository
- provides a dashboard to summarise user progress and overall progress
- allows annotating two vector datasets, with aerial photographs being provided in the background for ground-truthing



Two stage annotation process

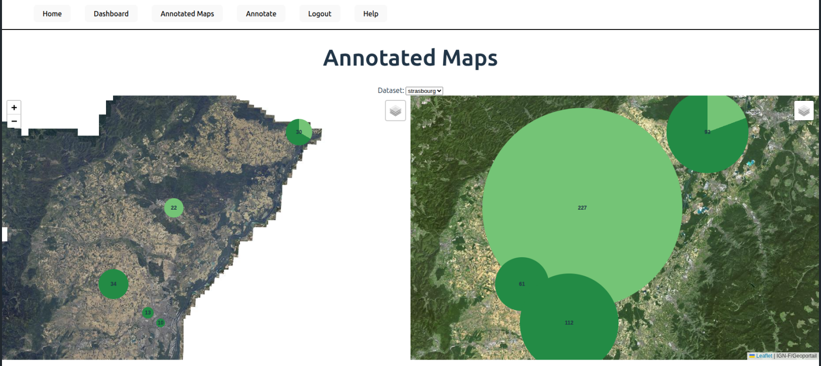


Stage 1: type of matching link (1-1, 0-1, 1-0, 1-m, n-1, n-m).



Stage 2: type of real word evolution (no change, construction, destruction, reconstruction, don't know), signal a quality issue and comment.

Dashboard and maps to monitor progress



- Open source and deployed automatically on github pages `https://github.com/subdense/annotate`
- Developped in scala, compiled into html/js with scalajs
- Datasets (geojson) on their own git repository (access rights to handle user access, login in the app with github login/token)
- git commands through the isomorphic-git library
- Clone all data at login, one commit pushed for each annotation

Deployment

→ pilot deployment of the application on samples (100 points, 500m radius to collect buildings), within functional urban areas of Strasbourg (FR) and Dortmund (DE), between 2011 and 2021; test of the app and first annotation by Subdense researchers (expertise on densification and with country data).

Example of a construction in Strasbourg: 0-1 true link

Example: reconstruction

Example: comment

Example: issue with type of link

Example: missing data

Work in progress: algorithm optimisation

→ Multi-modelling matching algorithm by [Guardiola et al., 2024b] integrate into the OpenMOLE platform.

→ Optimise algorithm parameters for F-score on expected type of link for each annotation cluster, with a NSGA2 genetic algorithm.

→ Preliminary results: algorithm can be fine-tuned to different urban typologies (center, housing projects, suburbs); multi-modelling algorithm has better performance than each separately.

Perspectives

Contributions:

- Web app to annotate matching links
- Proof-of-concept of a “serverless” deployment with user input through git

Work in progress:

- integrate feedback from first tests
- generate consistent samplings (commuting FUAs) on all 6 studied areas
- annotation campaign to produce a “large scale” ground-truth dataset
- matching algorithm optimisation using the ground-truth in OpenMOLE
- optimisation on change detection

Perspectives:

- More precise annotation of links (complicated cases: n-m links e.g.)
- More generic and configurable application: annotate raster data, other polygonal data (land plots), lines vector data, etc.
- Use for data quality issues

References i

- Batty, M. (2018).
  Inventing future cities.
  MIT press.
- Bel Hadj Ali, A. (2001).
  Qualité géométrique des entités géographiques surfaciques. Application à l'appariement et définition d'une typologie des écarts geométriques.
  Theses, université Gustave Effeil ; Anciennement Université de Marne La vallée.

Bucher, B., Brasebin, M., Buard, E., Grosso, E., Mustière, S., and Perret, J. (2012).
Geoxygene: Built on top of the expertise of the french nma to host and share advanced gi science research results.
In *Geospatial Free and Open Source Software in the 21st Century: Proceedings of the first Open Source Geospatial Research Symposium, OGRS 2009*, pages 21–33. Springer.

Bucher, B., Ndim, M., Olteanu-Raimond, A.-M., Raimbault, J., Perret, J., Dembski, S., and Jehling, M. (2024).
Conceptualising a co-operative building evolution dashboard on city regions over the past decades for densification studies.
In *27th AGILE Conference on Geographic Information Science," Geographic Information Science for a Sustainable Future"*, volume 5.

ENSG Géomatique

Bucher, B., Raimbault, J., Ndim, M., Raimond, A.-M., Perret, J., Dembski, S., and Jehling, M. (2025).
A building changes model to support comparative studies and open discussions on densification.
Under review for IJGI.

Costes, B., Perret, J., Bucher, B., and Gribaudi, M. (2015).
An aggregated graph to qualify historical spatial networks using temporal patterns detection.
In *The 18th AGILE International Conference on Geographic Information Science.*

Evers, D., Katurić, I., and van der Wouden, R. (2024).
***Urbanization in Europe: Past Developments and Pathways to a Sustainable Future*.**
Springer Nature.

Gregory, I. N. and Healey, R. G. (2007).
Historical gis: structuring, mapping and analysing geographies of the past.
Progress in human geography, 31(5):638–653.

- Guardiola, P., Raimbault, J., Olteanu-Raimond, A.-M., and Perret, J. (2024a).
  Benchmarking algorithms for matching geospatial vector data.
  In *Proceedings of the French Regional Conference on Complex Systems*, pages 277–280.
- Guardiola, P., Raimbault, J., Olteanu-Raimond, A.-M., and Perret, J. (2024b).
  Optimising geospatial vector data matching algorithms for change detection with synthetic data.
  In *Conference on Complex Systems 2024*.

- Harvey, F., Vauglin, F., and Bel Hadj Ali, A. (1998). **Geometric matching of areas, comparison measures and association links.** In *Proceedings of the 8th International Symposium On Spatial Data Handling*, pages 557–568.
- Jehling, M., Schorcht, M., and Hartmann, T. (2020). **Densification in suburban germany: approaching policy and space through concepts of justice.** *Town Planning Review*, 91(3):217–237.

- Khelifi, L. and Mignotte, M. (2020).
  Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis.
  Ieee Access, 8:126385–126400.
- Mustière, S. (2002).
  Description des processus d'appariement mise en oeuvre au cogit, rapport 2002.0072, chap.6.
  Technical report, IGN.
- Olteanu-Raimond, A.-M., Mustiere, S., and Ruas, A. (2015).
  Knowledge formalization for vector data matching using belief theory.
  Journal of Spatial Information Science, (10):21–46.

- Raimbault, J. and Pumain, D. (2022).
  Trade-offs between sustainable development goals in systems of cities.
  Journal of Urban Management, 11(2):237–245.

- Xavier, E. M., Ariza-López, F. J., and Urena-Camara, M. A. (2016).
  A survey of measures and methods for matching geospatial vector datasets.
  ACM Computing Surveys (CSUR), 49(2):1–34.

Reserve slides

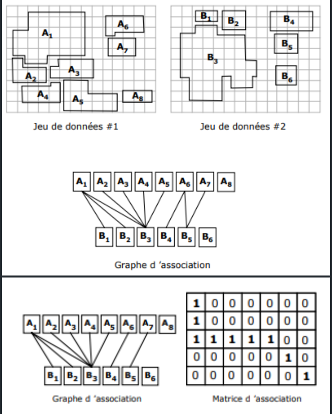
The SUBDENSE project

The SubDense European project studies the dynamics of suburban densification by:

Understanding polyrationalities of space, actors and policies on suburban densification

- exploring how diverse strategies of land policy interact with landowners' and local stakeholders' interest and agency to shape suburban densification and their impact on suburbia across different planning systems (France, Germany, UK);
- combining quantitative approaches (geodata analysis and geosimulation) with qualitative approaches (social and policy science and planning).

Geometric Matching of Areas (GMoA) algorithm

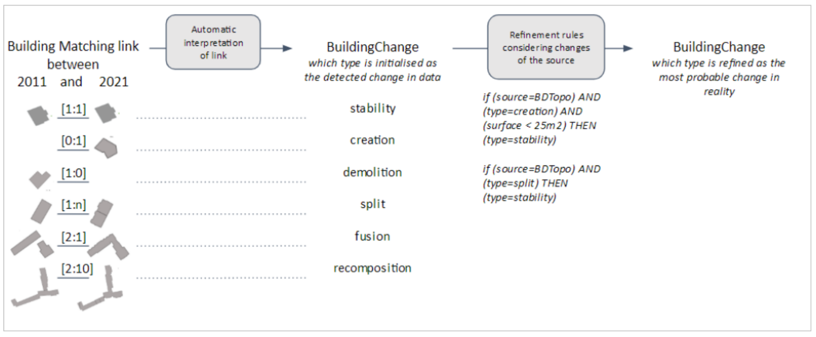


Algorithm to produce m-n links based only on geometries [Harvey et al., 1998] [Bel Hadj Ali, 2001]:

1. Construct all possible association links as surfaces with a non-empty intersection (top left Fig.)
2. Filter links with an intersection surface below a threshold parameter
3. Filter links with an intersection surface too small relatively to matched surfaces (rate parameter)
4. Construct all m-n links as connected clusters in the remaining bipartite graph (bottom left Fig.)

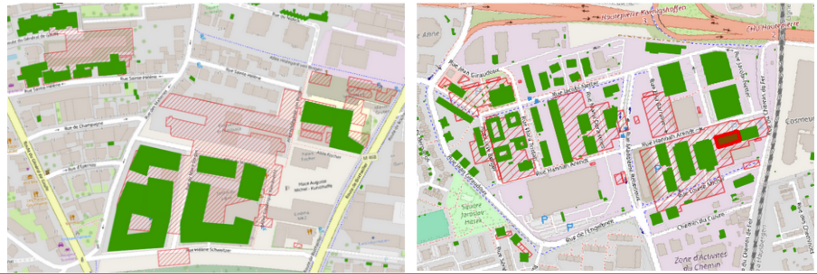
→ implemented in the `geoxygene` java library [Bucher et al., 2012] by [Mustière, 2002]

Application of the GMoA to change detection



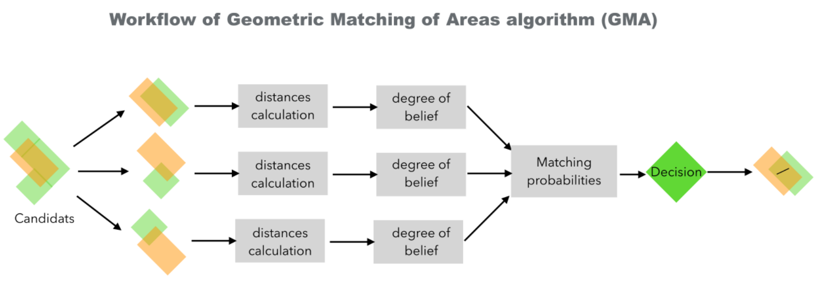
Automatic interpretation of matching links into building change, with specific adjustements for BDTopo [Bucher et al., 2025], by a python script available at `https://github.com/subdense/matching`

Example



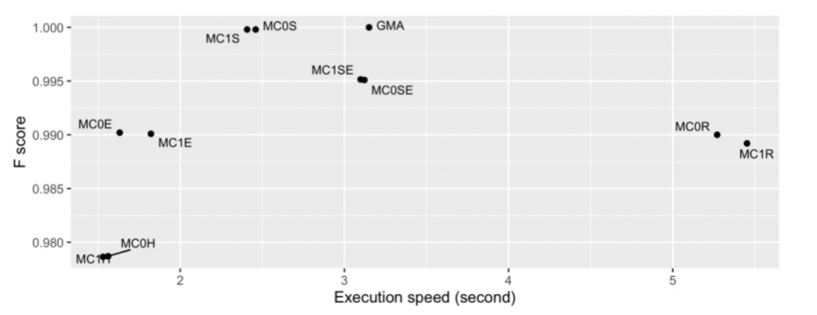
Building evolution for two examples in the urban area of Strasbourg [Bucher et al., 2025]

Multi-criteria matching algorithm (MCA)



Implementation of the Multi-criteria matching algorithm [Olteanu-Raimond et al., 2015] in python by [Guardiola et al., 2024a], with euclidian, surface, radial and Hausdorff distances, and run twice to obtain m-n links.

Benchmarking of algorithms



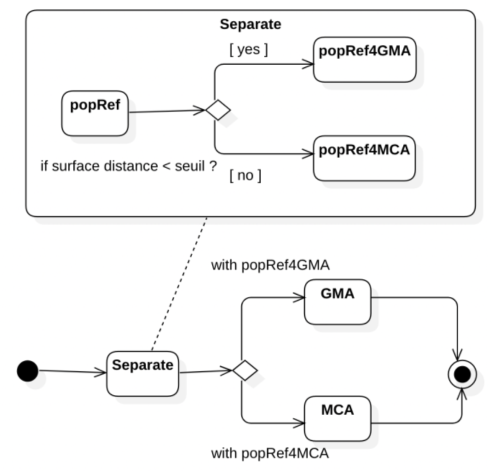
Bi-objective benchmark optimising for performance (F-score on a $\simeq$ 2k buildings ground truth dataset) and runtime, for different parametrisations of the two algorithms, by [Guardiola et al., 2024a]

$\rightarrow$ various algorithm performances

$\rightarrow$ under-detection of change by GMoA and over-detection by MCA

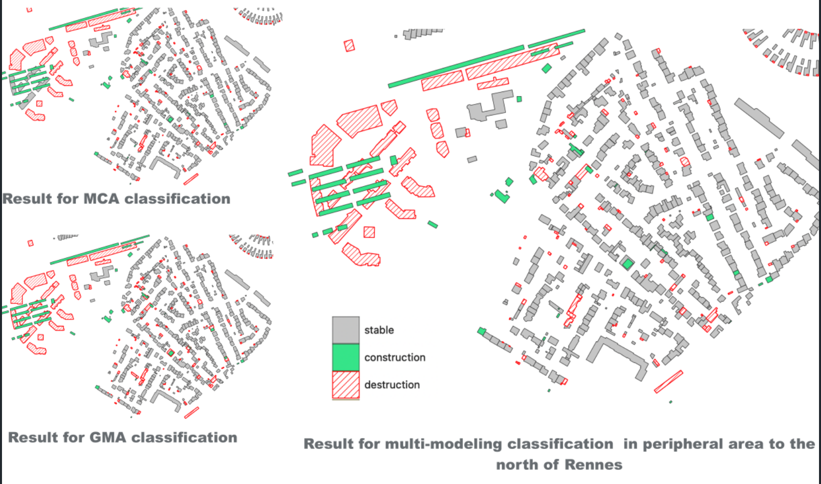
$\rightarrow$ GMA better for m-n links, MCA better for 1-1 links

A new algorithm using multi-modeling



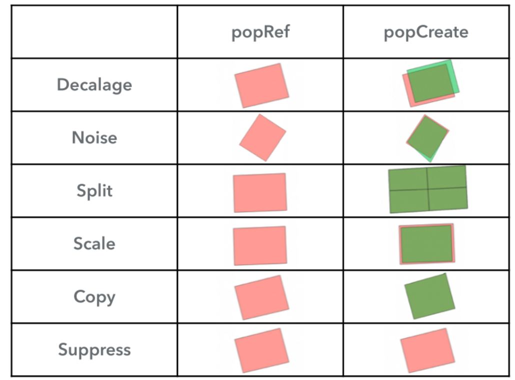
Proposition of a new algorithm combining GMoA and MCA, with the choice made through thresholding surface distance [Guardiola et al., 2024b]

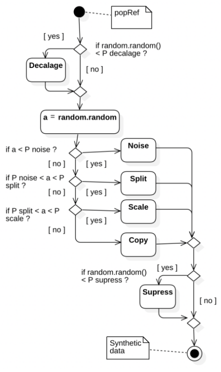
Example



Results obtained on an example in the suburbs of Rennes [Guardiola et al., 2024b]

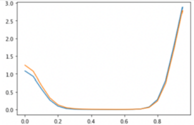
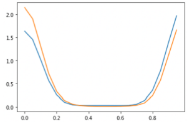
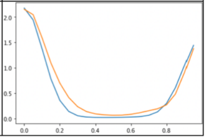
Validating the algorithm with synthetic data





Synthetic data generator introduced by [Guardiola et al., 2024b] to generate datasets on which to validate and optimise the algorithm

First results on synthetic data

Lieu Géographique	Valeur des paramètres optimisés	Courbes des densité des distances surfacique de popRef et popCreee
Centre Urbain	Decallage= 0.05 Brouilleur = 0.45 Echelle = 0.20 Spliteur = 0.25 Copieur = 0.10 Suppression = 0.10	
Banlieue	Decallage= 0.10 Brouilleur = 0.75 Echelle = 0.10 Spliteur = 0.05 Copieur = 0.10 Suppression = 0.10	
Ville périphérique	Decallage= 0.10 Brouilleur = 0.25 Echelle = 0.40 Spliteur = 0.15 Copieur = 0.15 Suppression = 0.10	

Integration into the OpenMOLE software and calibration of the generator for different urban typologies (single fitness: distance between surface distance densities) [Guardiola et al., 2024b]