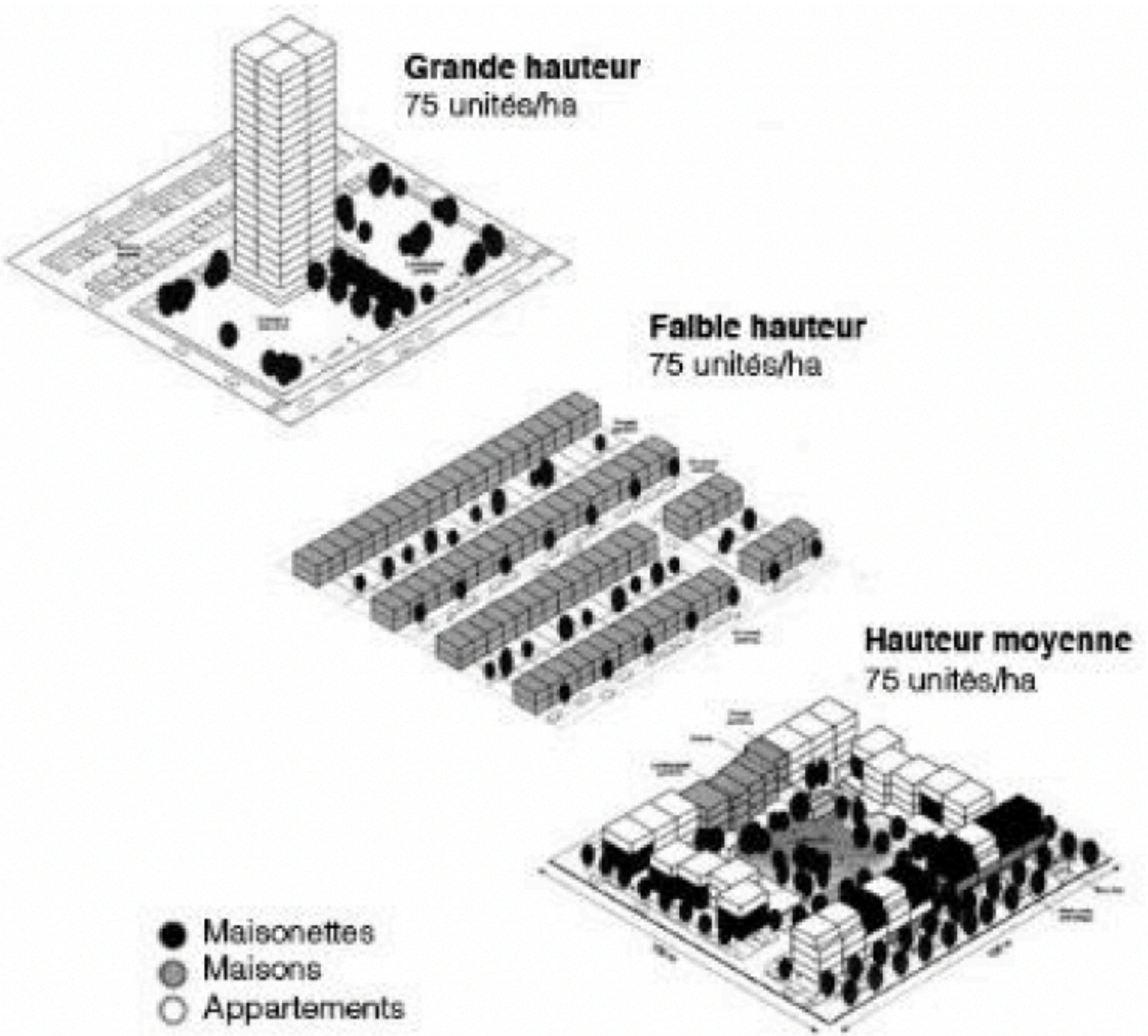
Optimising geospatial vector data matching algorithms for change detection with synthetic data

Paul Guardiola, Juste Raimbault, Ana-Maria Olteanu-Raimond, Julien Perret

CCS 2024

1 CONTEXT



SUB DENSE

Understanding polyrationalities
of space, actors and policies
on suburban densification

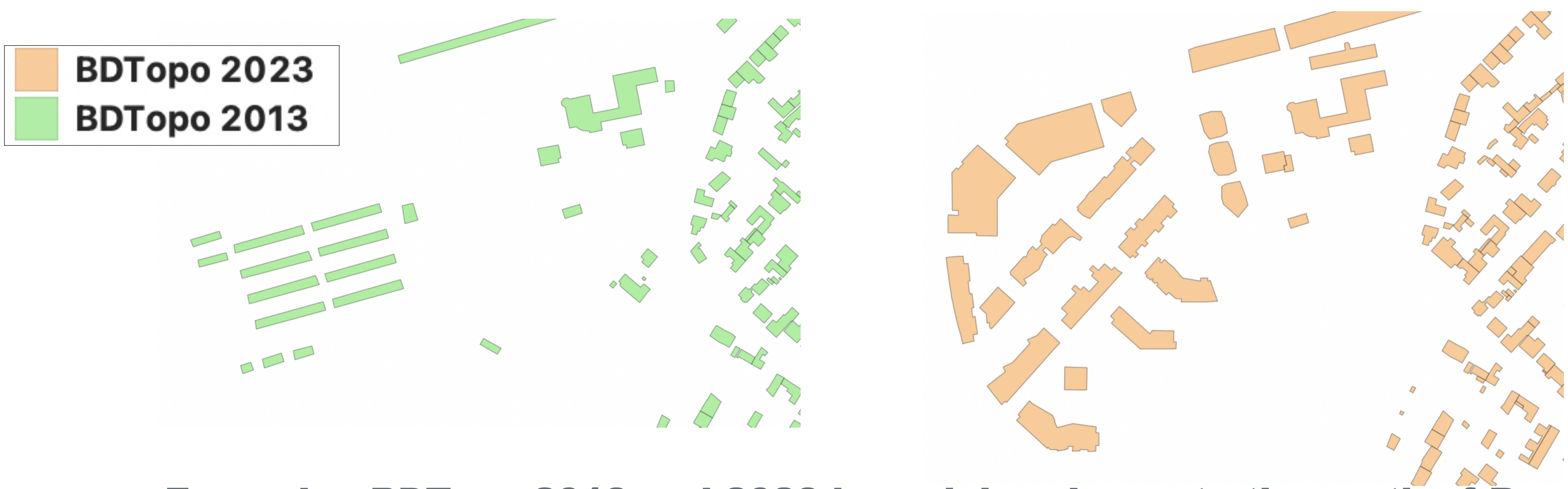
1 - Vivre en Ville, adapt. de Urban Task Force, Towards an urban renaissance, 1999.

Quantification of urban dynamics :

- **Buildings stability assessment at different dates**

Oblique image in 2014

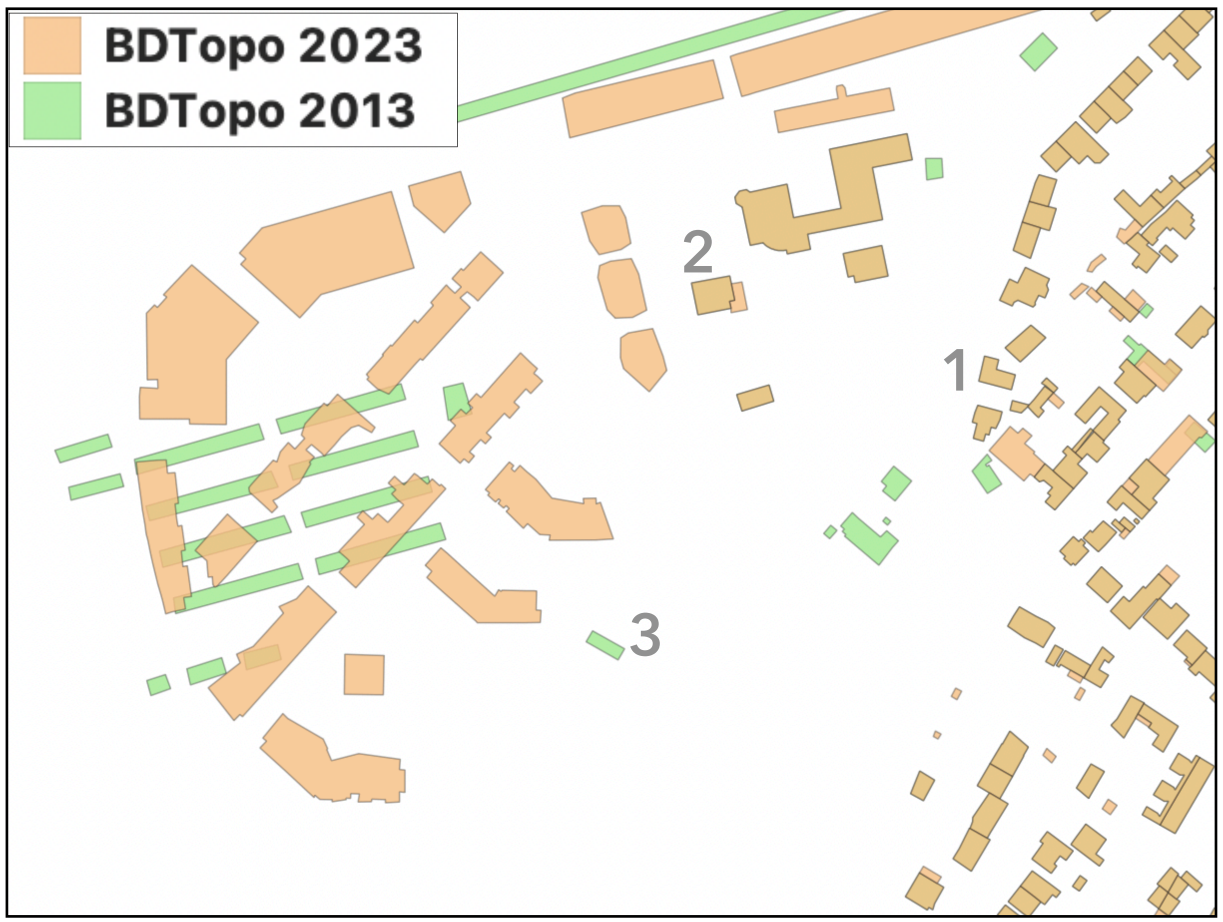
Oblique image in 2023

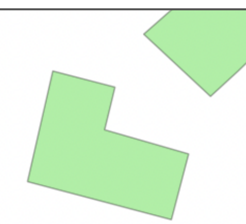
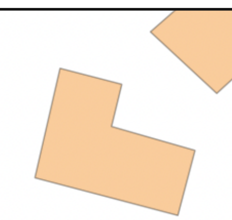
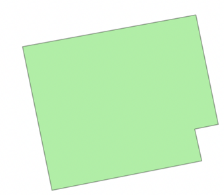
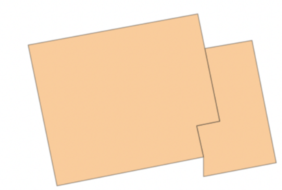


Example : BDTopo 2013 and 2023 in peripheral area to the north of Rennes

Quantification of urban dynamics :

- Establish the types of evolution possible for territorial dynamics

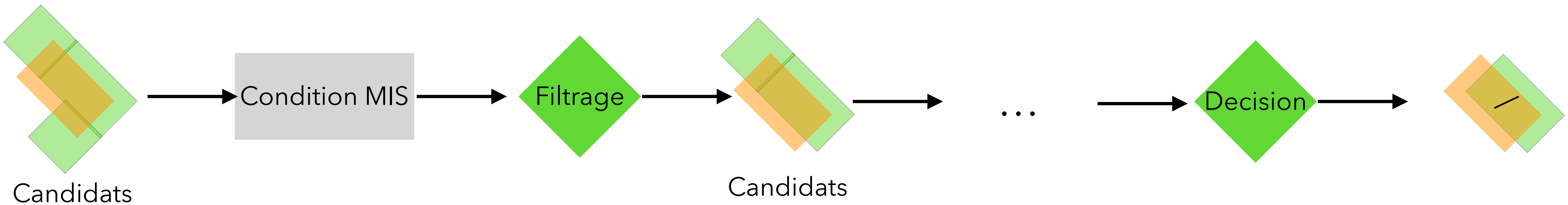


	En 2013 :	En 2023 :	Evolution
1.			Stable
2.			Appeared
3.			Destroyed

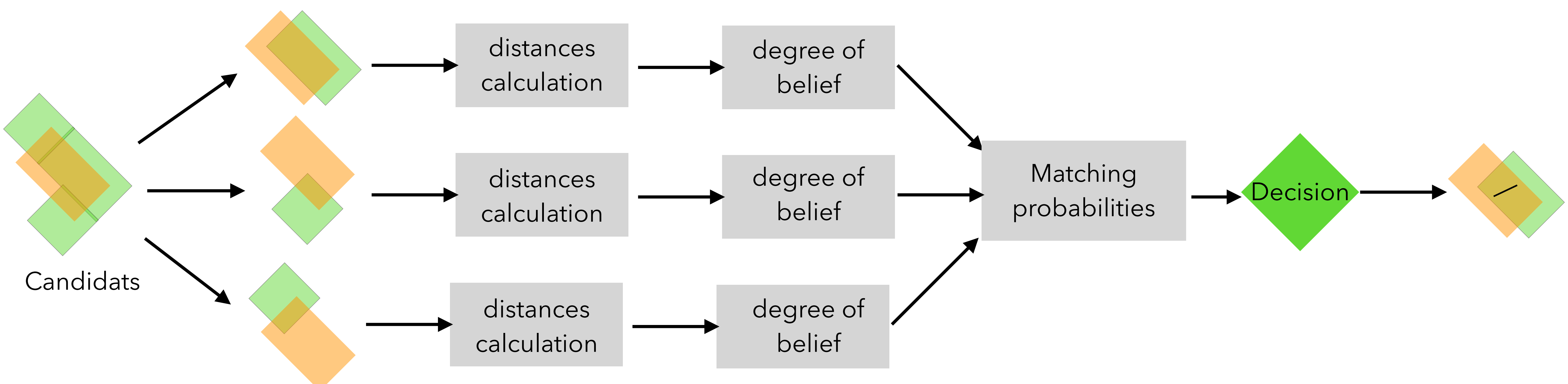
Example : BDTopo 2013 and 2023 in peripheral area to the north of Rennes with types of evolution for several buildings

Mouhamadou NdimJuste RaimbaultBénédicte BucherAna-Maria Olteanu-RaimondJulien PerretA collaborative dashboard to study periurban densification*European Colloquium on Theoretical and Quantitative Geography 2023*, Universities of Coimbra and Porto, Sep 2023, Braga, Portugal

2 ALGORITHMS

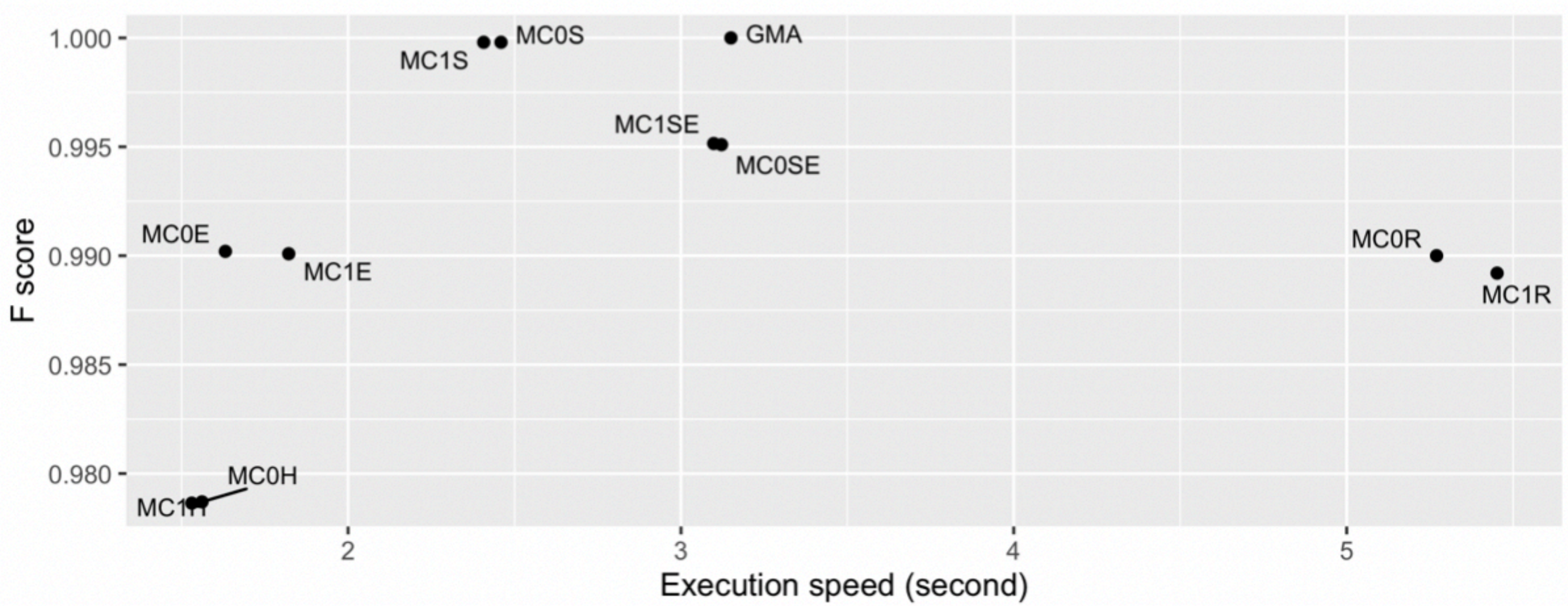


Workflow of Geometric Matching of Areas algorithm (GMA)



Workflow of Multi Criteria Algorithm algorithm (MCA)

- **Paul Guardiola and alt, Benchmarking algorithms for matching geospatial vector data, French Regional Conference Complex System 2024, https://zenodo.org/records/11267401 page 277 - 280**



Optimal GMA parameters: $MIS = 1$, $MIP = 1$, $PIS = 0.8$, $MSD = 0.25$, $MCCF = 0.8$;
MCOSE: Multi-Criteria ($d_b = 0$) with Surface and Euclidien criteria

Pareto comparison of optimal instantiations of the matching algorithms

- **Repeat for different study areas**

3 Synthetic data generator

- **Method : We develop a synthetic data generator, which starts from real-world building data, and perturbs the dataset to generate a virtual future state**

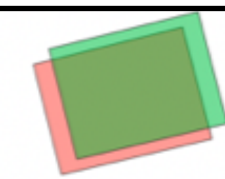
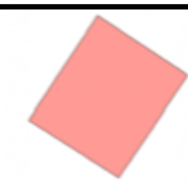
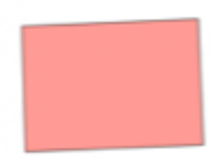
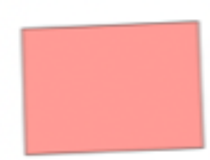
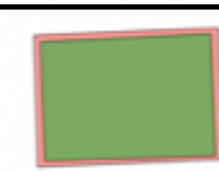
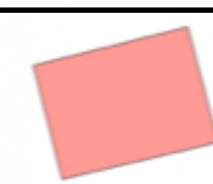
	popRef	popCreate
Decalage		
Noise		
Split		
Scale		
Copy		
Suppress		

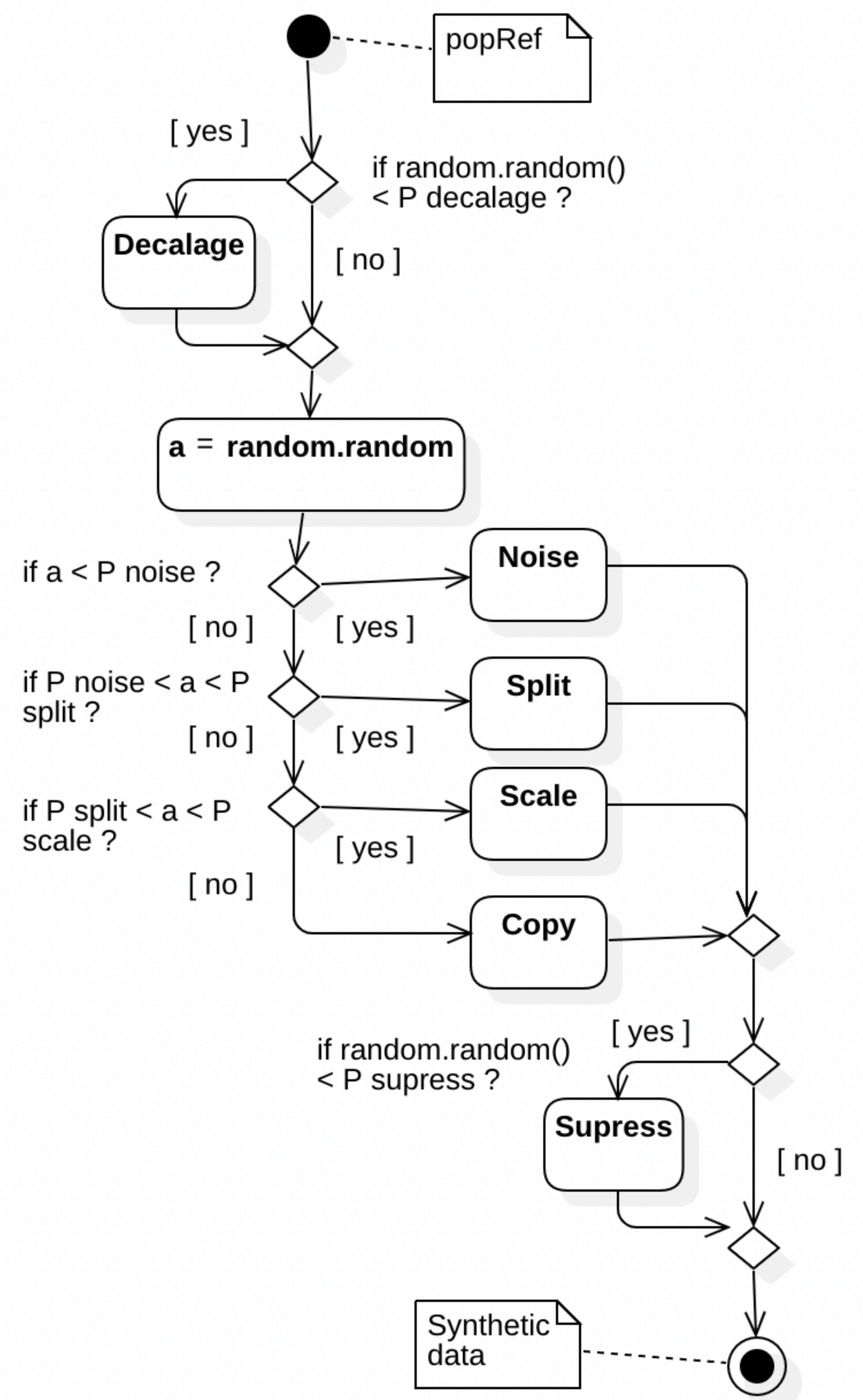
Table of synthetic data generation

Construction of a synthetic data generator

- Use of probabilities to choose to apply a noise or other changement in data specification

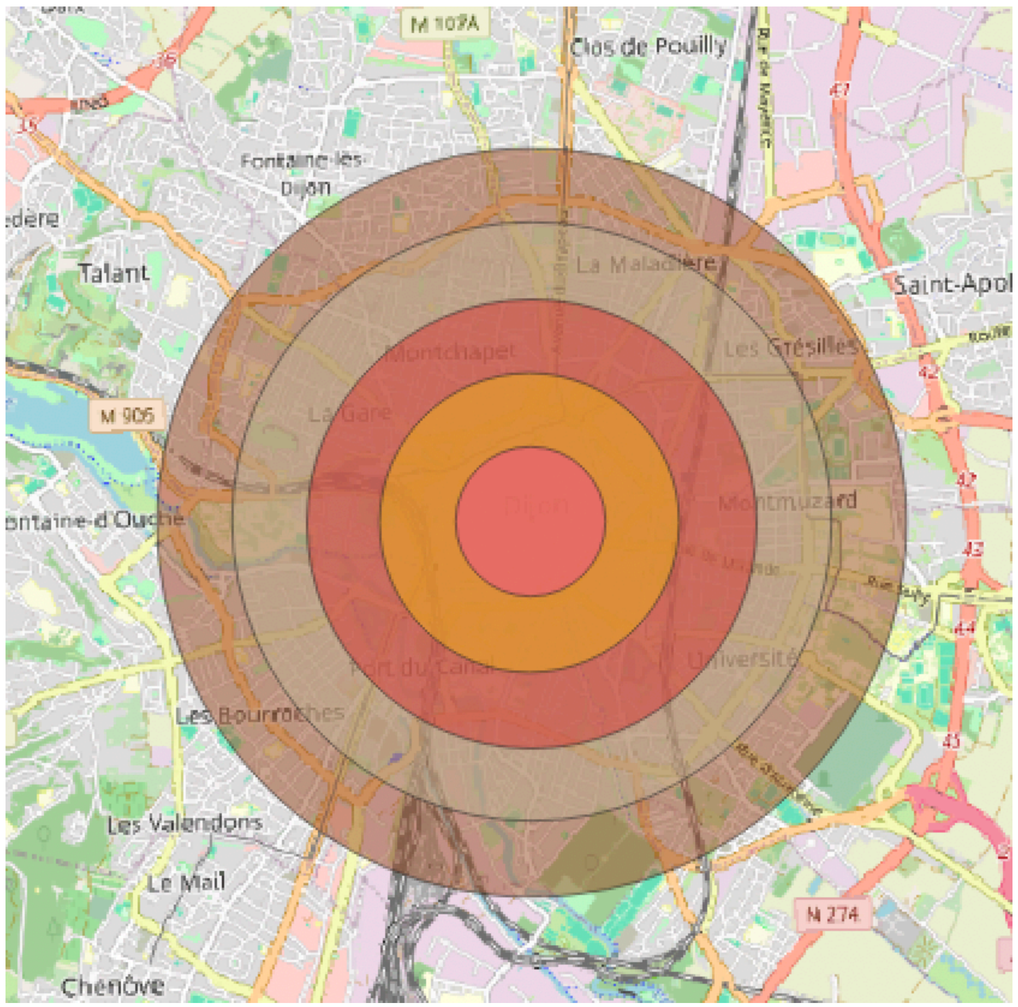
- Decalage and supress intervene during the beginning and at the end of the process to be used on all the data

Exemple : Worfkflow synthetic data generator

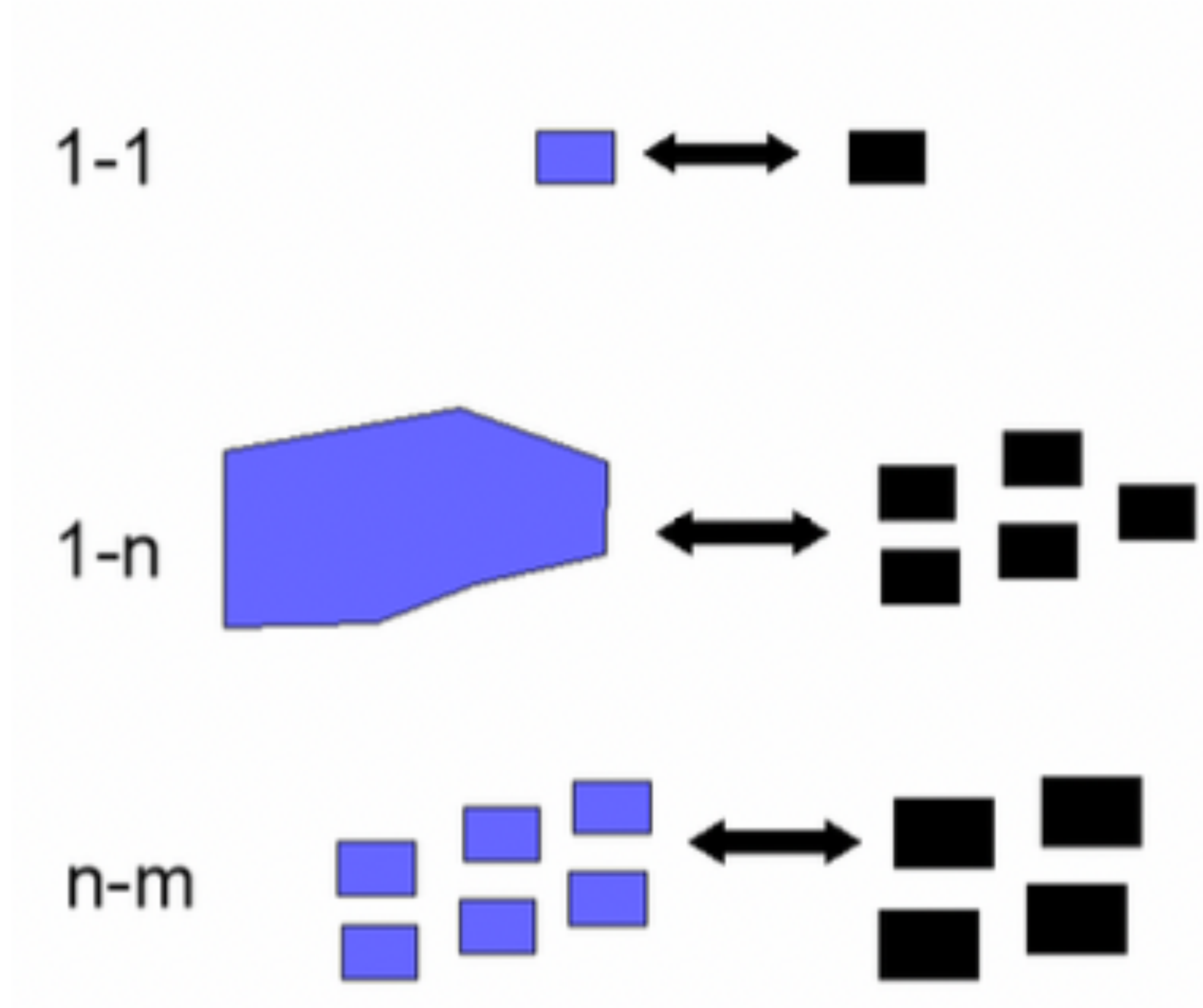


CONTEXT ALGORITHM G.TRUTH OPTIMISATION

- **The problem is to fix the different probabilities**
  - **Recreate urban dynamics**



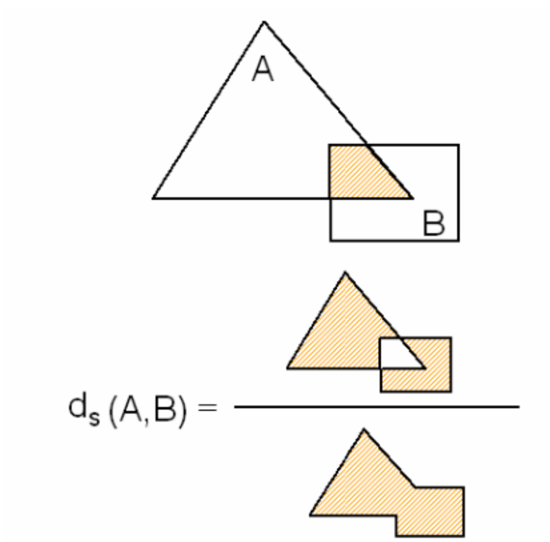
Density of surface distance between popRef and popComp On Dijon on annulus with a radius of 1 km



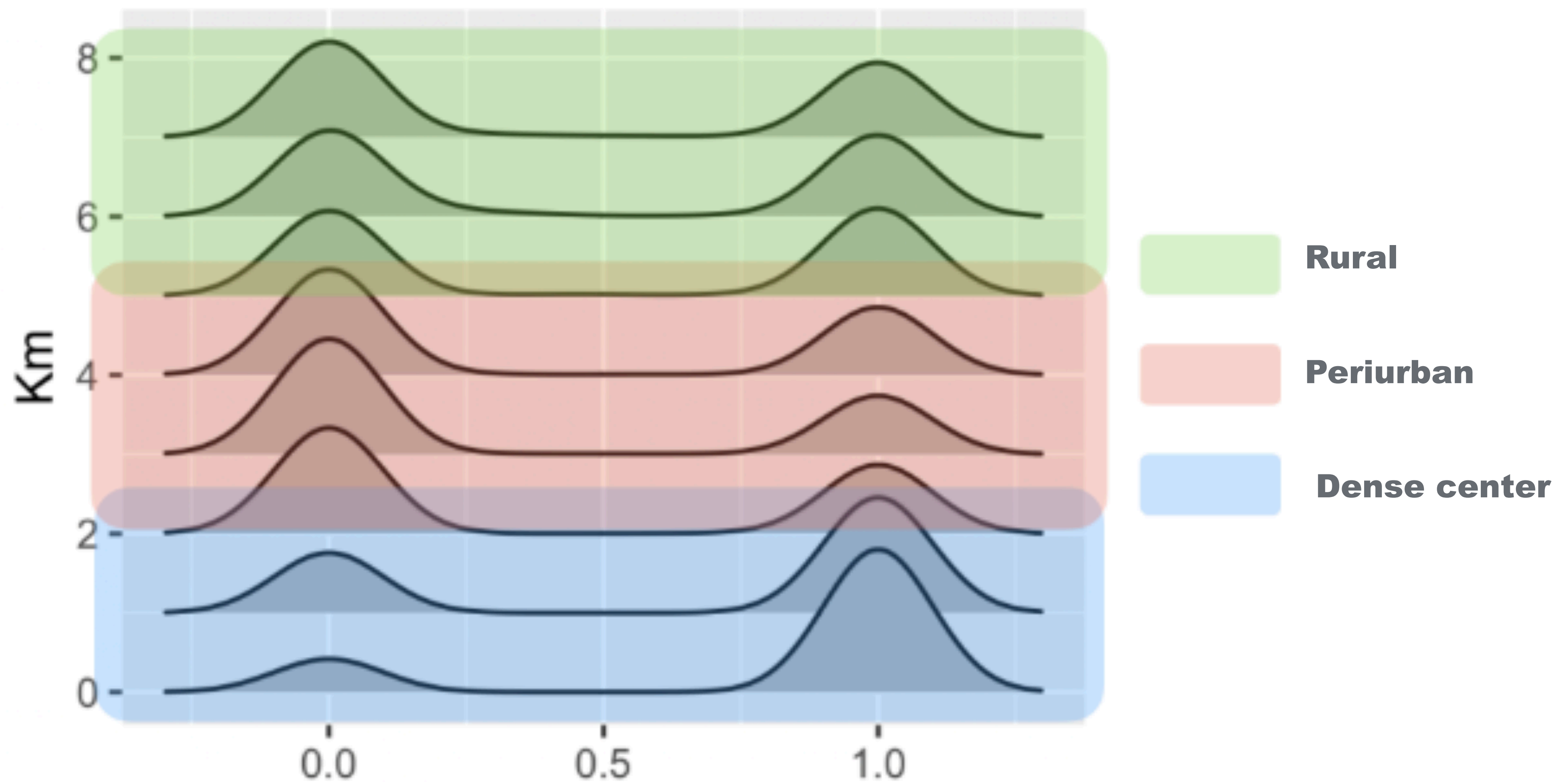
Types of matching links

- **The problem is to fix the different probabilities**
  - **Recreate urban dynamics**

Reminder : Surface distance

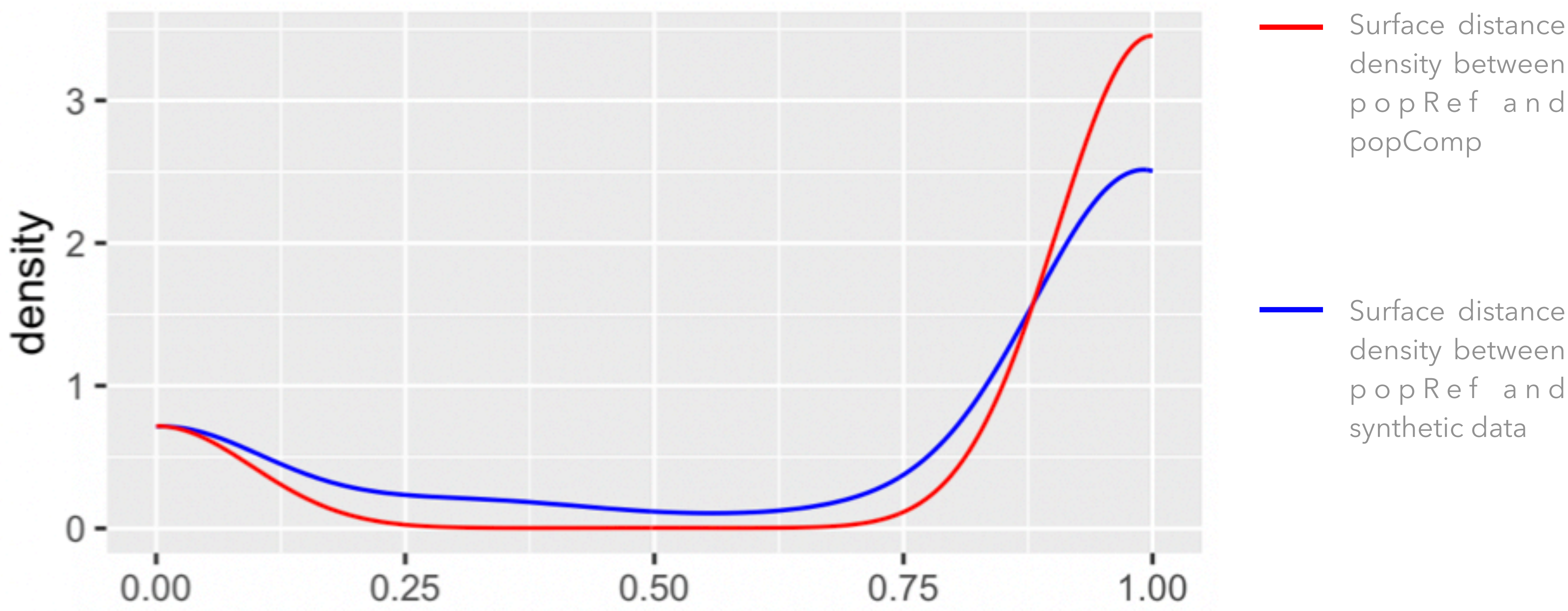


Ana-Maria Olteanu. Fusion de connaissances imparfaites pour l'appariement de données géographiques: proposition d'une approche s'appuyant sur la théorie des fonctions de croyance. Autre [cs.OH]. Université Paris-Est, 2008. Français. NNT : 2008PEST0252 . tel-00469407



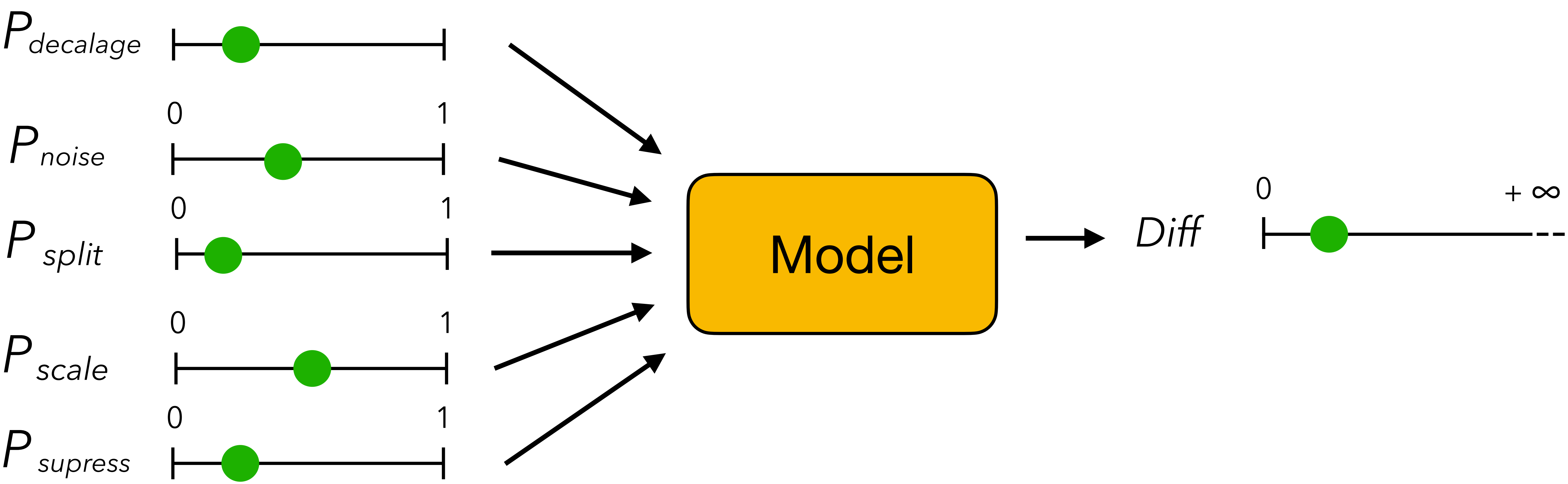
Density of surface distance between popRef and popComp On Dijon on annulus with a radius of 1 km

- **The problem is to fix the different probabilities**
  - **Recreate urban dynamics**



$$Diff = \int (density(test(dS) - density(reference(dS))^2 ddS$$

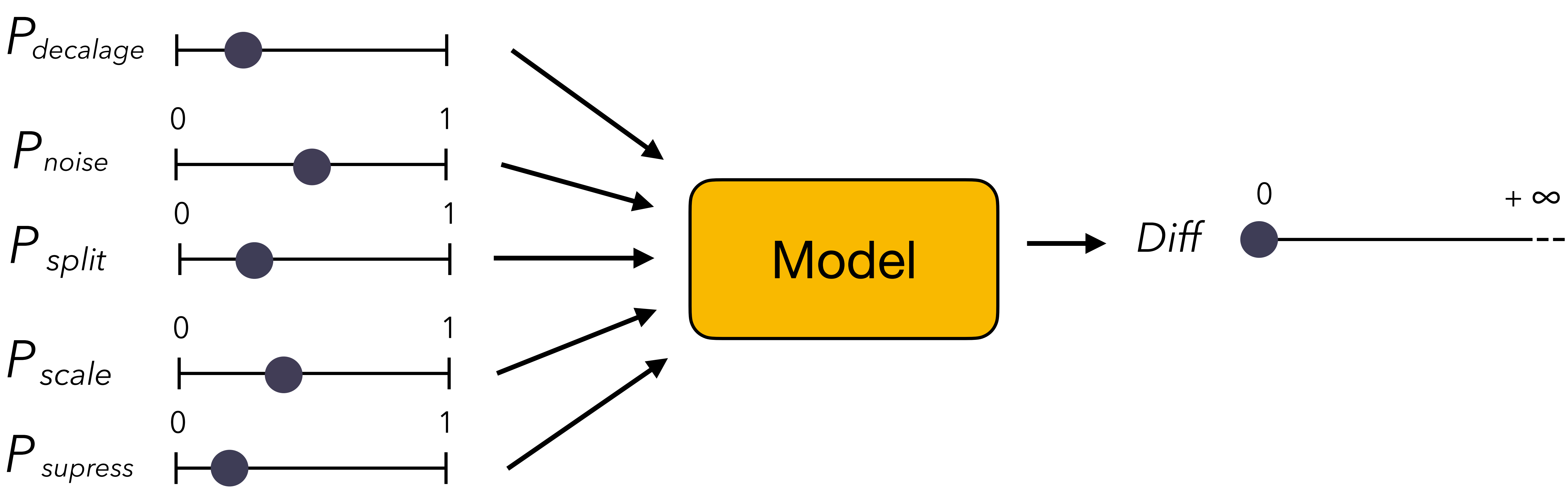
- **The problem is to fix the different probabilities**
  - **Recreate urban dynamics**



Synthetic data generator inputs and output

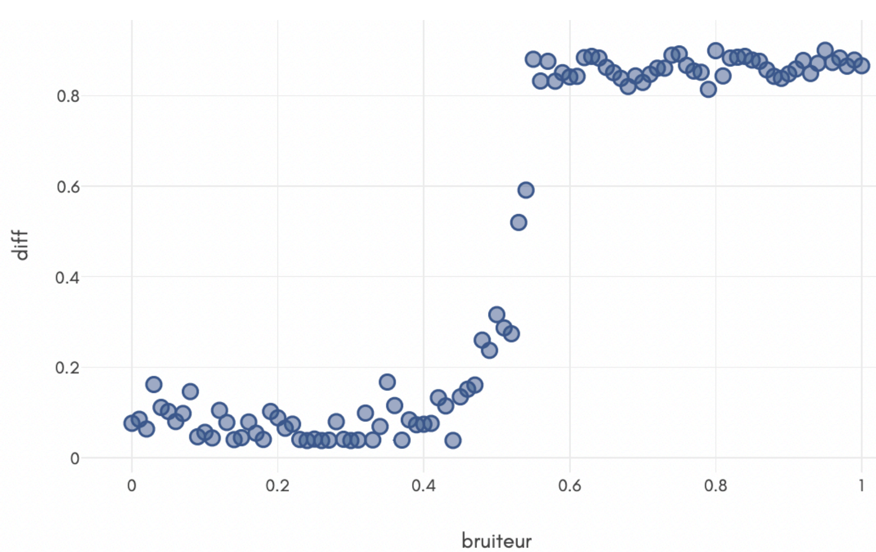
- **Use of OpenMOLE calibration**

Use of OpenMOLE calibration

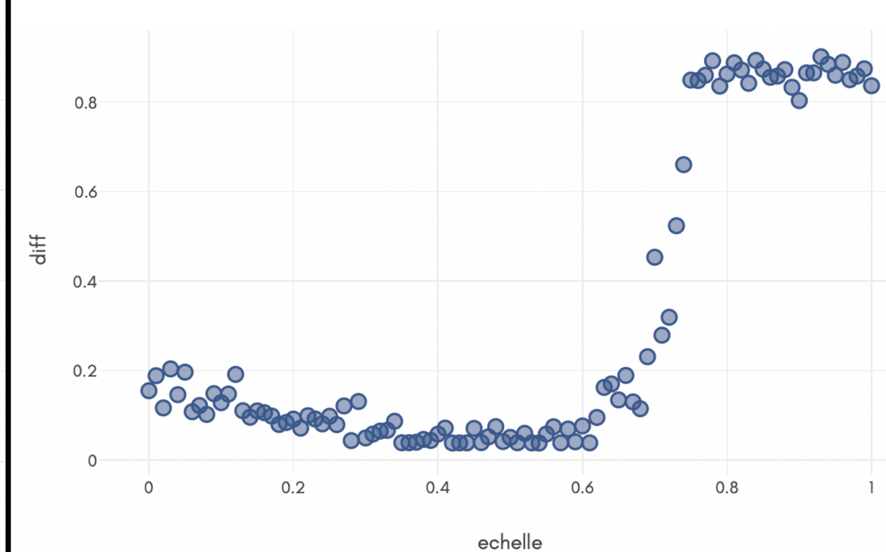


	Calibration 1	Calibration 2	Confidence interval
Input	P noise = 0.26 P scale = 0.46 P split = 0.24 P copy = 0.04	P noise = 0.23 P scale = 0.49 P split = 0.11 P copy = 0.16	I noise = [0.20 , 0.30] I scale = [0.40 , 0.50] I split = [0.05 , 0.25] I copy = [0 , 0.20]

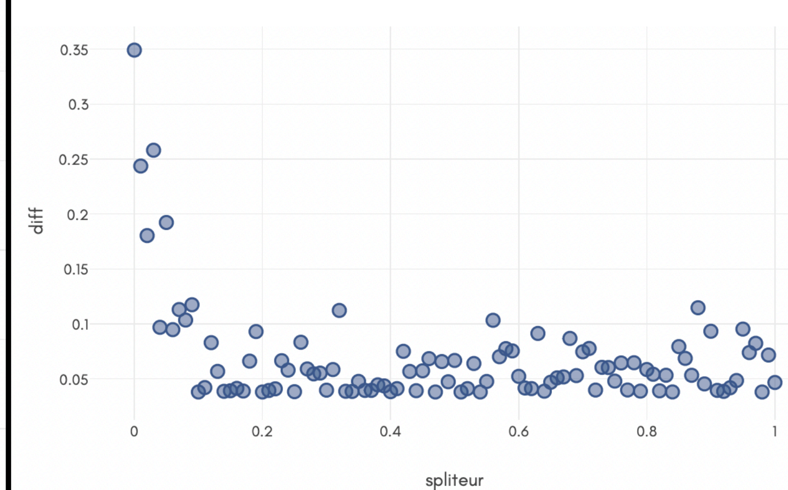
P noise



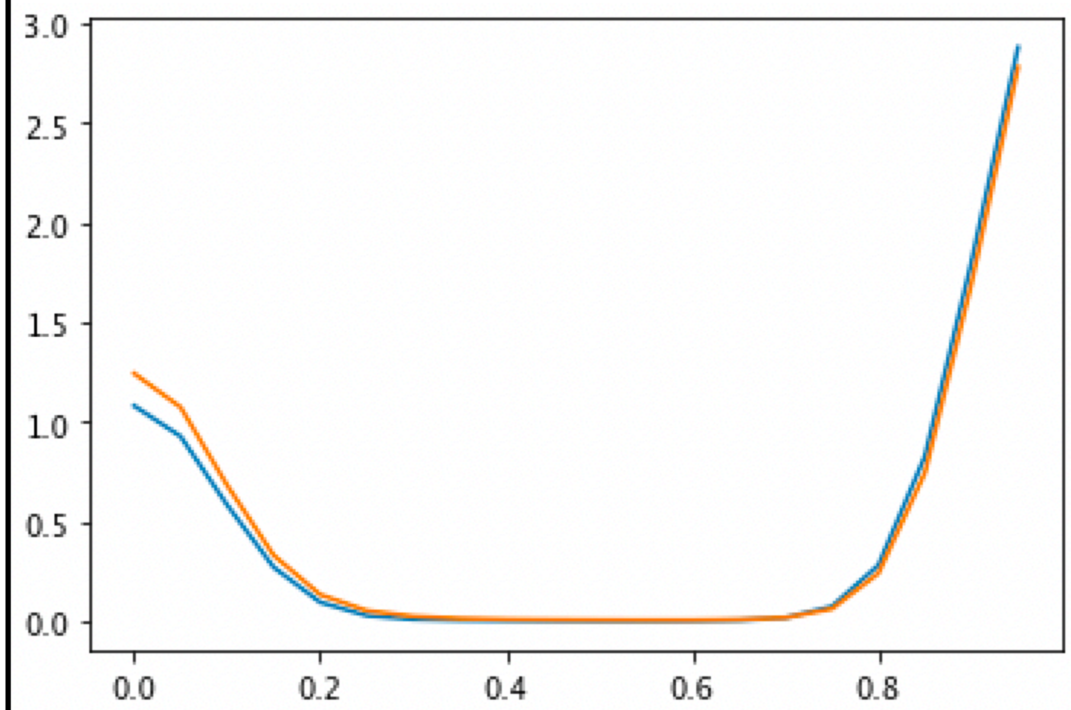
P scale



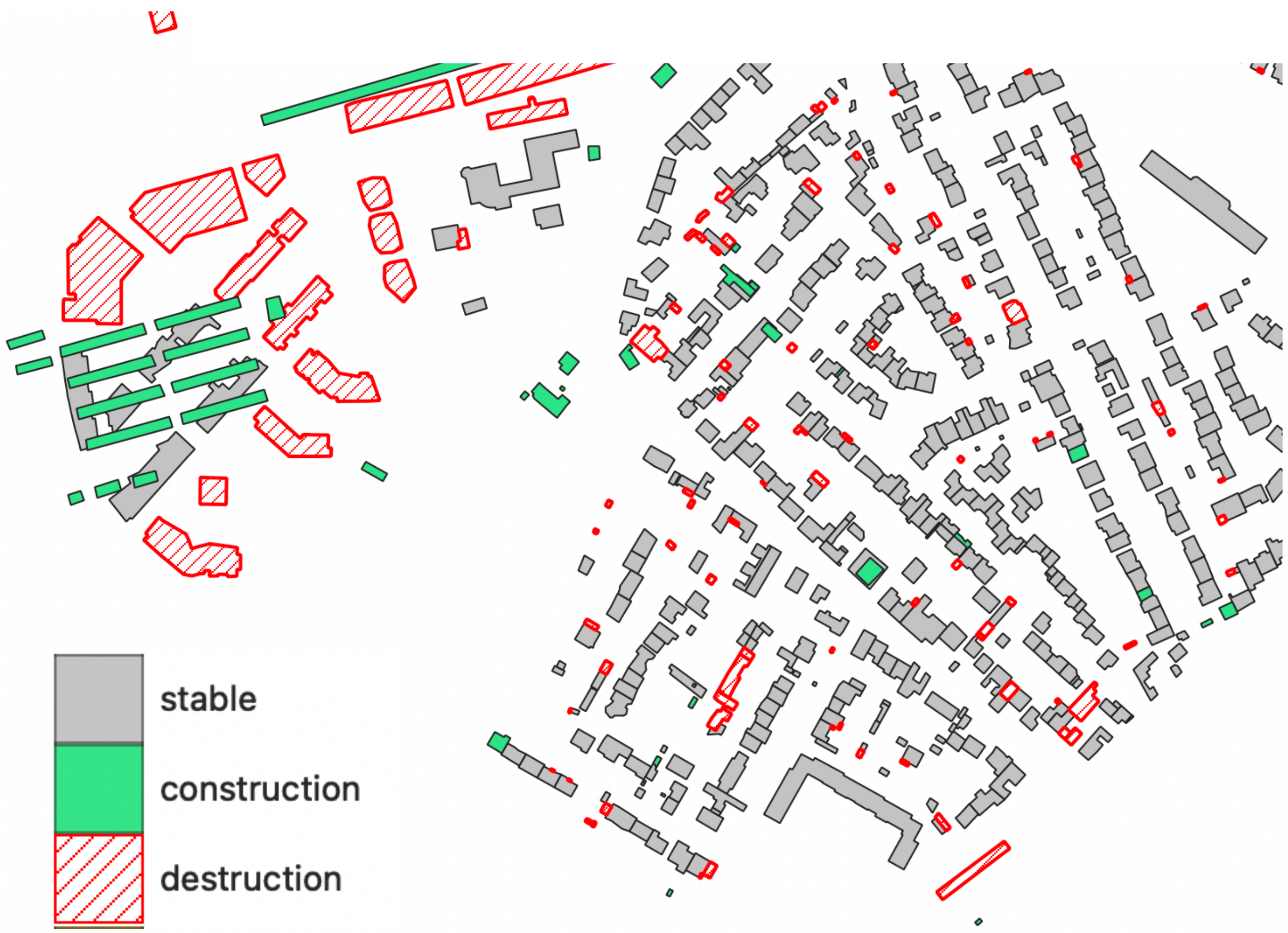
P split



Résultats de la calibration d'OpenMOLE pour différentes types de zone sur Rennes

Urban morphologies	Optimized probabilities	Surface distance density between popRef and synthetic data
Dense center	Decalage= 0.05 Noise = 0.45 Scale = 0.20 Split = 0.25 Copy = 0.10 Supress = 0.10	

Matching algorithm classification with GMA and MCA



Result for MCA classification

Result for GMA classification

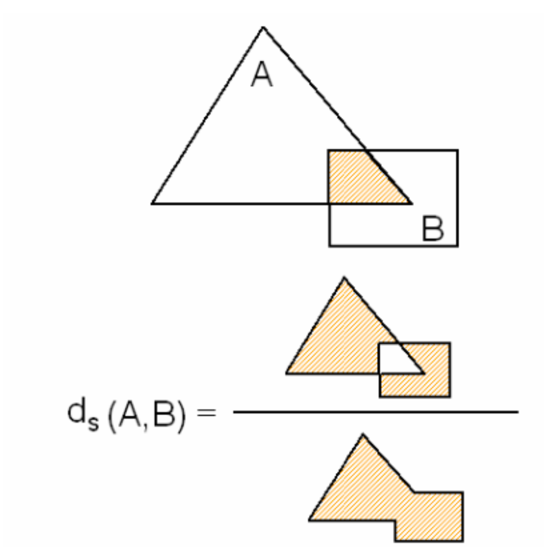
- **We find an under-detection for the GMA and an over-detection for MCA**
- **GMA is a well method for n-m links rather than MCA is better for 1-1 links**

4 Optimisation

Multi-modeling

- **Apply different Matching algorithm according their zone of predilection**

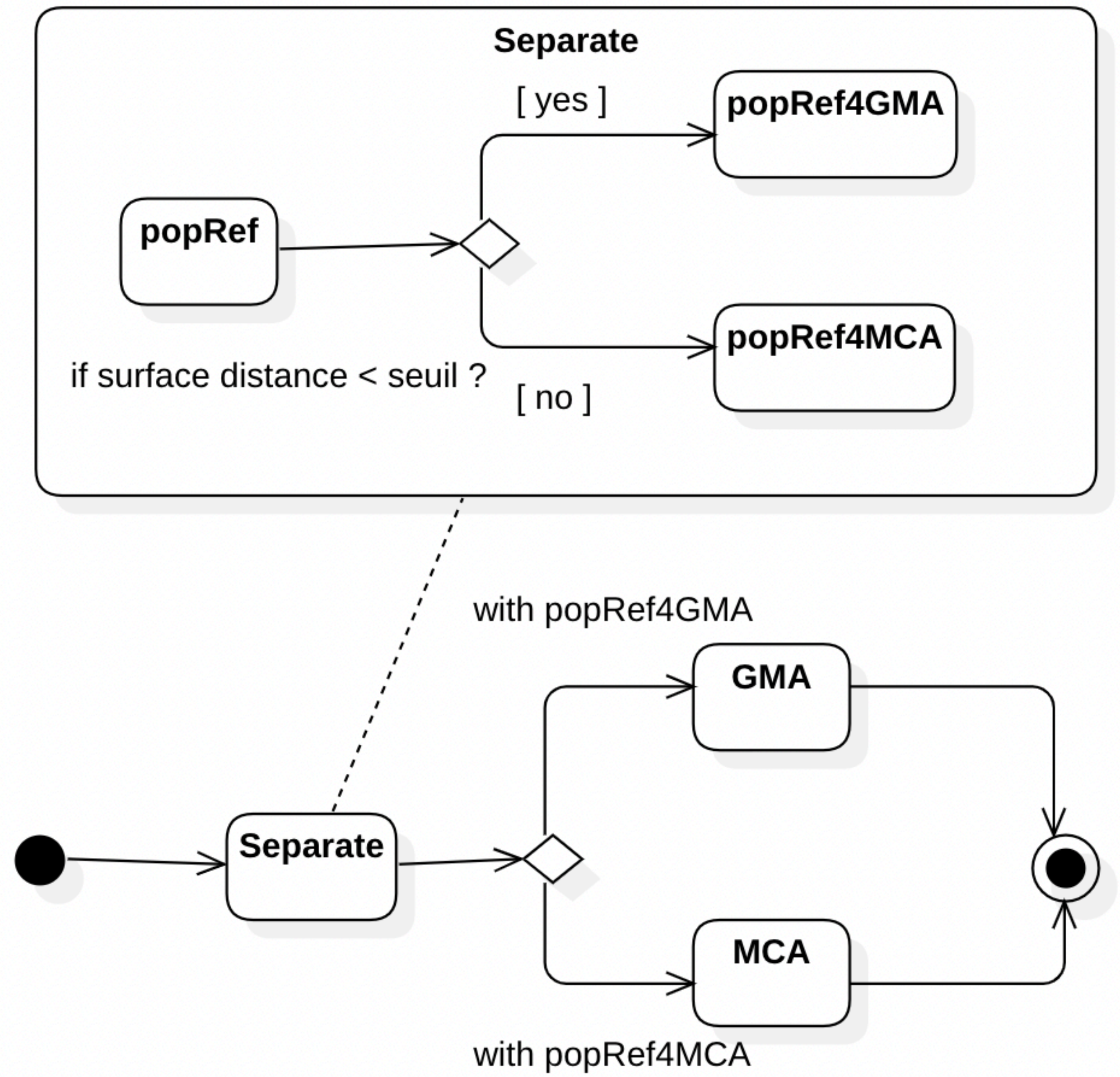
Reminder : Surface distance



Ana-Maria Olteanu. Fusion de connaissances imparfaites pour l'appariement de données géo- graphiques: proposition d'une approche s'appuyant sur la théorie des fonctions de croyance. Autre [cs.OH]. Université Paris-Est, 2008. Français. NNT : 2008PEST0252 . tel-00469407

popRef with surface distance

- **Apply different Matching algorithm according their zone of predilection**



Worflow du modèle pour séparer la population de référence

Result for MCA classification

Result for GMA classification

Result for multi-modeling classification in peripheral area to the north of Rennes

Future works :

- to test this work to other European countries
- to test other matching algorithms
- to compare change detection results with machine learning results on remote sensing images

While our results are quite specific to geospatial data and the case of building change detection, they provide some insights into the broader question of spatial sensitivity analysis for complex systems, since we describe a methodology to generate synthetic data by introducing noise into a real world dataset, a research direction that was remaining to be fully explored in the future.

This presentation proposes :

- a synthetic data generator for Matching algorithm for geospatial vector data
- a method to obtain a matching algorithm optimized

Source code and data are available on an open git repository at :
https://github.com/paulguardiola/projetAppariement.