

Appariement de données anciennes : les cartes de Cassini.

Benoît COSTES

IGN – Laboratoire COGIT, 73 avenue de Paris, Saint-Mandé

benoit.costes@ign.fr

L'objectif du projet ANR GéoPeuple est d'étudier les relations entre l'évolution du paysage français et l'évolution de la répartition de la population sur le territoire.

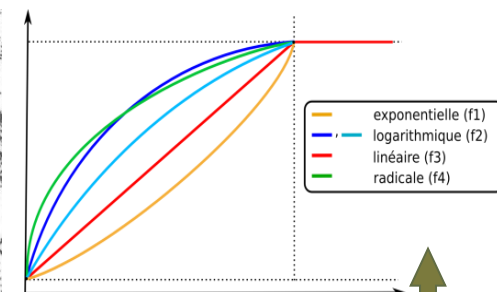
Nous nous focalisons sur l'appariement des données ponctuelles extraites des cartes de Cassini (moulins, églises, hameaux, etc.) avec la BDTopo de l'IGN. Ces données étant imprécises, les techniques d'appariement privilégiant les relations géométriques entre objets ne sont pas adaptées.

Notre approche repose sur un appariement multi-critères (géométrie, sémantique et toponymie) s'inspirant des travaux d'Olteanu tout en s'affranchissant de la théorie de l'évidence pour des raisons de complexité algorithmique.

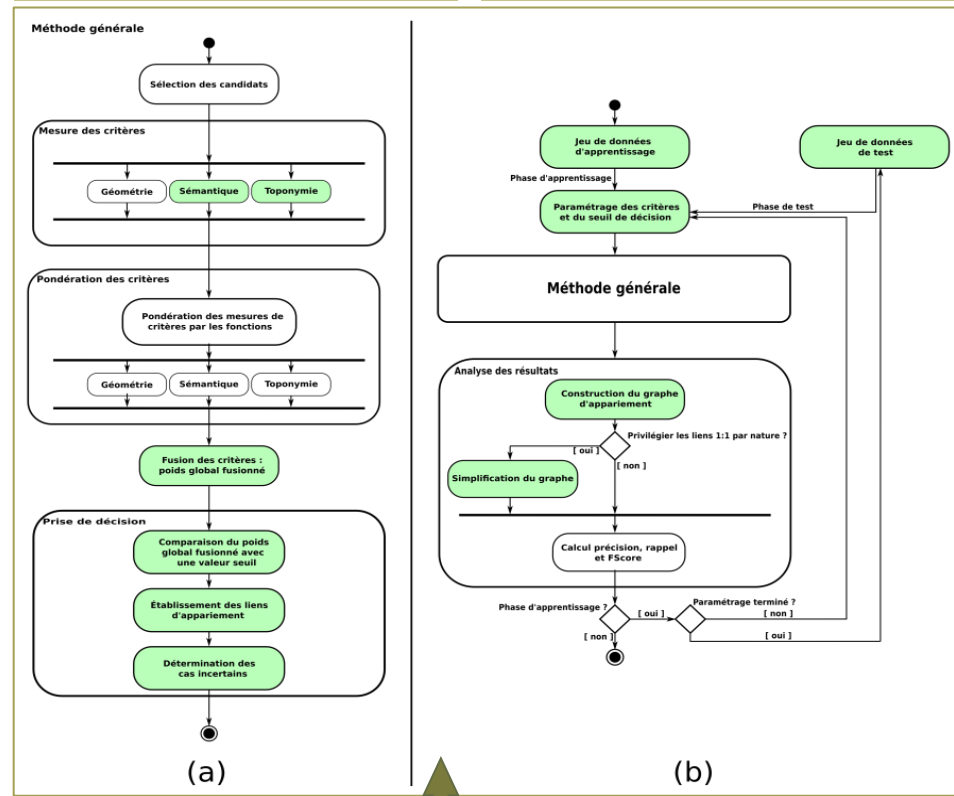
L'étape de pondération des critères permet d'accorder plus ou moins de crédit à un critère en fonction de sa valeur mesurée. Différents types de courbes sont testés. L'analyse des résultats repose sur le calcul d'une mesure combinant de manière égale la précision et le rappel : le F-Score, calculé sur un échantillon de données appariées manuellement.



Données imprécises : modification de la position relative entre deux villages au cours du temps (BDTopo en bleu, Cassini en rouge)



Exemple de quelques fonctions de pondération. Le choix du type de fonction influence sur les résultats et peut varier d'un critère à l'autre.



Méthode s'appuyant sur l'approche d'Olteanu (a) et méthode de paramétrisation automatique (b). Les améliorations apportées sont représentées en vert.

Abstract:

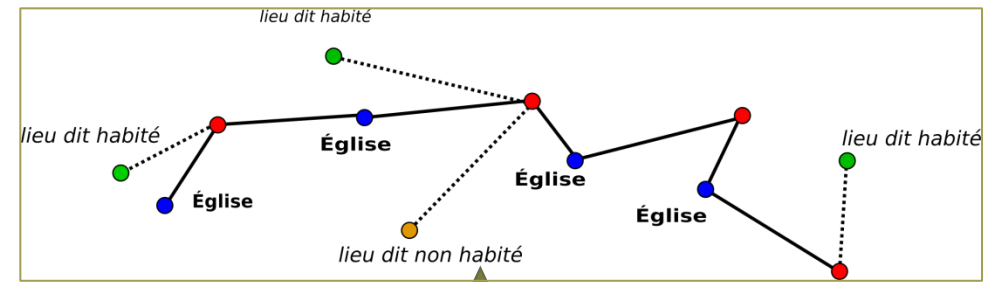
The matching of Cassini maps with current vector topographic database.

The integration of information contained in old topographic maps into a historical topographic framework (spatio-temporal database) is a complex task due to the inherent imperfection of data.

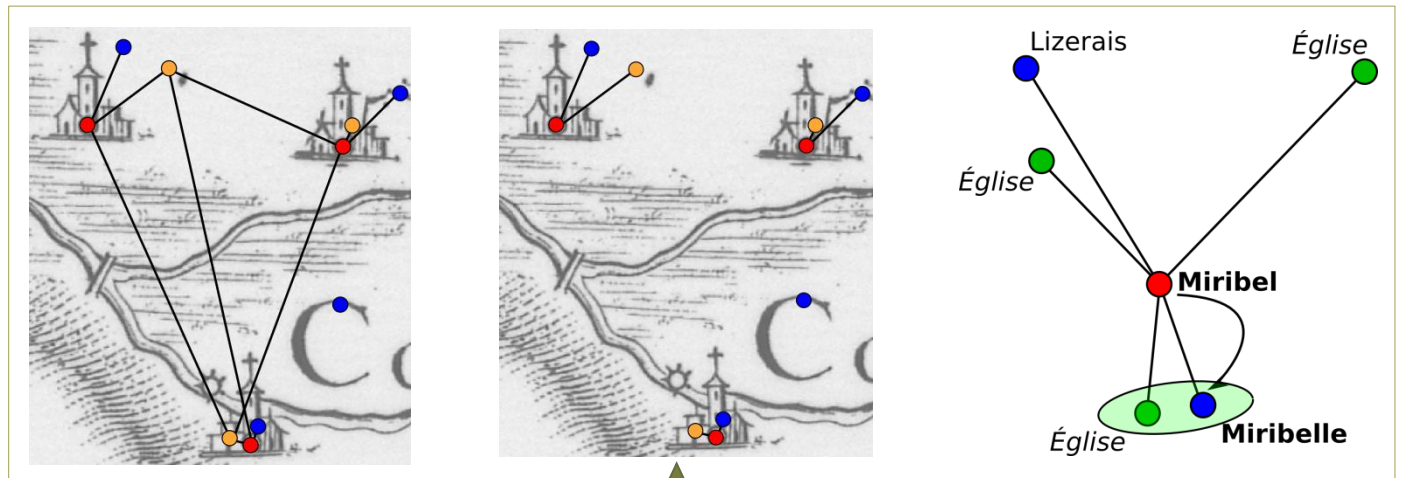
This contribution focuses on the data matching process of vector point data extracted from Cassini's maps with the vector topographic database of the French mapping (BDTOPO, 2011).

Our algorithm follows a multi-criteria analysis approach based on Ana-Maria Olteanu's work about belief functions theory. The automation of the data matching parameterization is done by supervised learning on manually matched data in order to ease results analysis,

Une étape de paramétrisation automatique a été implémentée, s'appuyant sur un apprentissage utilisant l'appariement manuel. La paramétrisation retenue est celle qui maximise la moyenne des Fscores des liens d'appariement et des objets non appariés sur les données d'apprentissage. Nous proposons également une méthode d'étude locale des liens d'appariement par nature d'un objet en construisant le graphe d'appariement de l'objet relatif à une nature N donnée de candidat de proche en proche. Ce graphe permet d'étudier le contexte local d'une entité en visualisant les objets et candidats en conflit à l'issue de l'appariement. Nous proposons de simplifier ces graphes d'appariement conflictuels en choisissant comme meilleur candidat d'un objet Cassini de référence, le plus proche voisin du ponctuel de la BDTOPO portant un toponyme proche de celui de l'objet étudié. Les hameaux, villages et villes de Cassini sont correctement appariés aux lieux dits actuels malgré de fortes variations toponymiques. Nous constatons que la plupart des moulins anciens ont disparus. L'étude des correspondances avec des données intermédiaires comme les minutes d'État-major ou les fonds de carte de 1960 de l'IGN pourra permettre de valider ou d'infirmer les résultats obtenus par notre approche.



Exemple de graphe d'appariement (en trait plein) pour une nature "Église". Les objets Cassini sont en rouge.



Graphe d'appariement pour la nature "Église" avant (1) et après (2) simplification. Les églises Cassini sont en rouge, les églises de la BDTOPO en orange, les lieux-dits de la BDTOPO en bleu. La figure (3) montre la prédilection des candidats proches du meilleur candidat au sens de la distance toponymique.

Zone	Classe	Apprentissage					Validation				
		Préc. LAP	Rapp. LAP	Préc. NAP	Rapp. NAP	F-score moyen	Préc. LAP	Rapp. LAP	Préc. NAP	Rapp. NAP	F-score moyen
Reims	Religieux	100%	97%	94%	100%	98%	100%	97%	100%	100%	99%
	Non religieux	100%	95%	95%	100%	97%	91%	100%	100%	100%	98%
	Moulins	83%	83%	98%	98%	91%	83%	100%	100%	96%	95%
St-Malo	Religieux	99%	97%	75%	100%	92%	100%	98%	100%	99%	99%
	Non religieux	100%	92%	83%	100%	93%	100%	88%	75%	100%	90%
	Moulins	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Grenoble	Religieux	99%	95%	82%	100%	93%	97%	99%	100%	67%	90%
	Non religieux	98%	100%	93%	96%	97%	100%	100%	100%	100%	100%
	Moulins	100%	93%	98%	100%	98%	100%	100%	100%	100%	100%

Résultats d'appariement sur chaque classe de chaque zone étudiée: précisions et rappel des liens d'appariement ("LAP"), précision et rappel des objets non appariés ("NAP"), FScore-moyen, sur les bases d'apprentissage et de validation.

