

Un modèle de classification pour l'inférence de la précision spatiale de bâtiments de la base OpenStreetMap avec des indicateurs intrinsèques

Ibrahim Maidaneh Abdi ¹, Arnaud Le Guilcher ², Ana Maria Olteanu Raimond ²

1: University of Djibouti, Djibouti, Djibouti

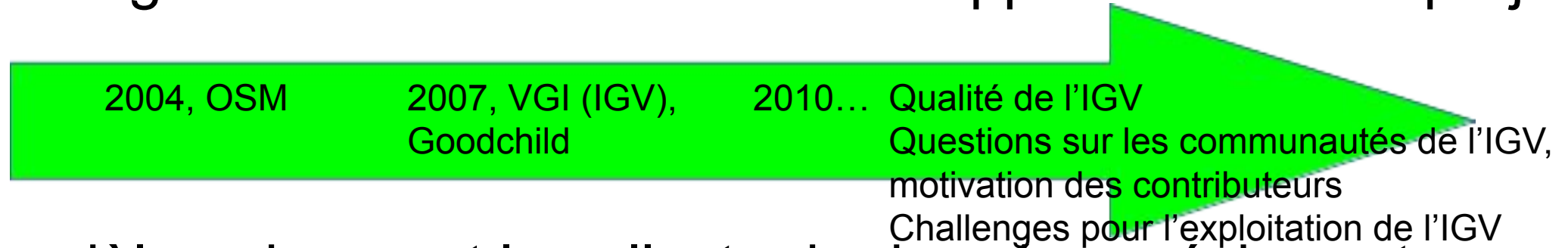
2: LASTIG, Univ Gustave Eiffel, IGN-ENSG

Journée de la Recherche

IGN - ENSG - Université Gustave Eiffel

Le projet OSM et la qualification des données

- Place grandissante de l'IGV et développement d'OSM project



Un modèle qui permet la collecte de données précises et avec un bon niveau de fraîcheur et une information sémantique riche, mais avec une qualité incertaine

- Qualification des données OSM
 - Comparaison avec des bases de référence
 - Qualification intrinsèque, reposant sur l'historique des données, les contributeurs, l'information sémantique, la géométrie

Qualification intrinsèque avec des indicateurs géométriques

- Évaluer la précision géométriques d'objets d'OSM par une méthode d'apprentissage utilisant des indicateurs intrinsèques calculés sur la géométrie (granularité, compacité...)
- Certifier les données d'OSM pour les utiliser dans des contextes opérationnels en l'absence de données de référence

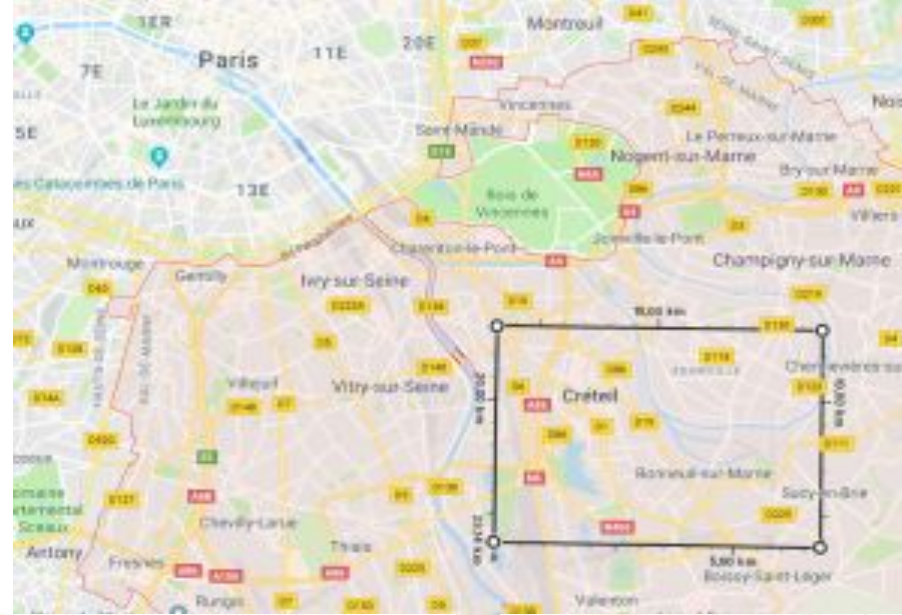
Plan

1. Choix d'une mesure cible pour évaluer la précision géométrique, et définition d'indicateurs intrinsèques
2. Évaluation de la précision de forme par classification
3. Étude de la transférabilité du modèle de classification initial

Zones d'étude

- Première zone d'étude : Val-de-Marne

Localisation de la première zone d'étude



Alternance entre villes denses, avec alignements de bâtiments, et zones rurales moins denses



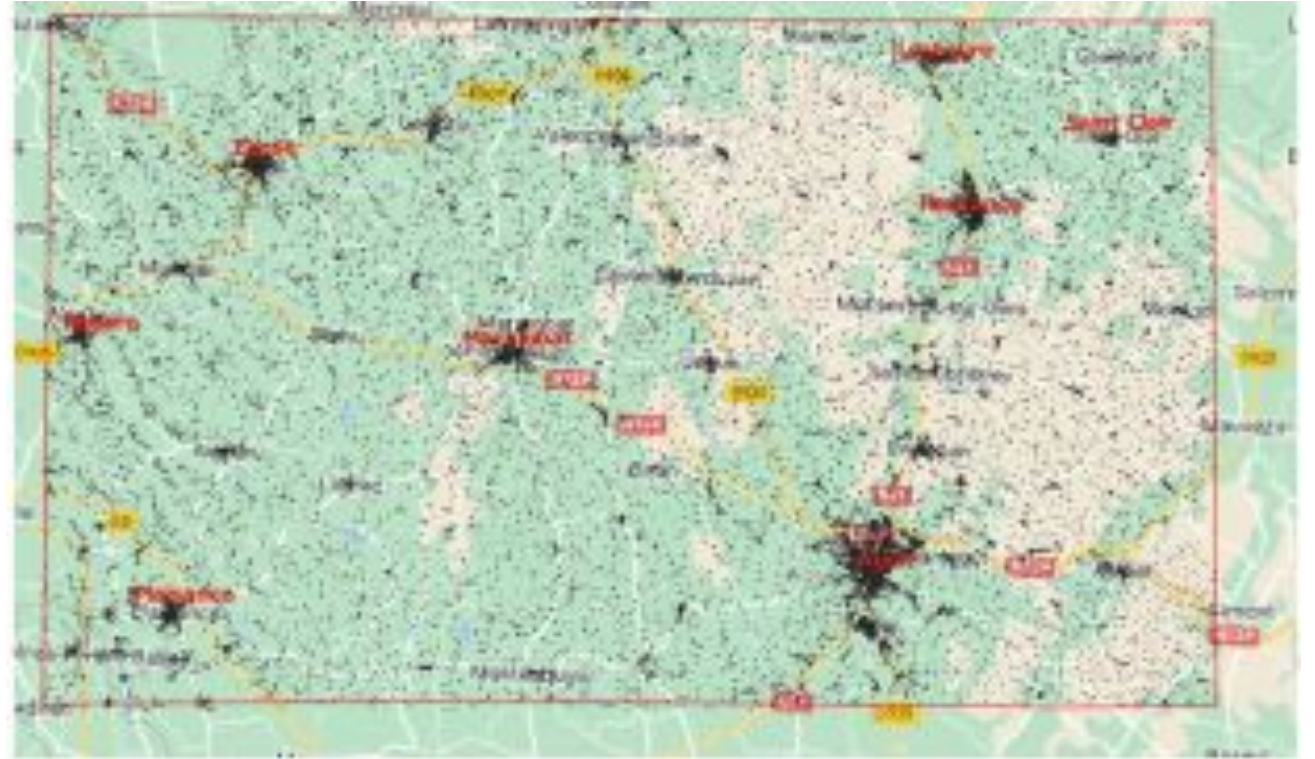
Emprises des bâtiments au nord de la zone



Emprises des bâtiments au Sud-Ouest de la zone

Zones d'étude

- Deuxième zone d'étude : Gers
- Vastes zones peu denses, avec des villes petites et moyennes
- Différente de la première zone, avec de grandes zones peu denses, mais avec plusieurs similitudes (même pays, ni côte ni montagne)

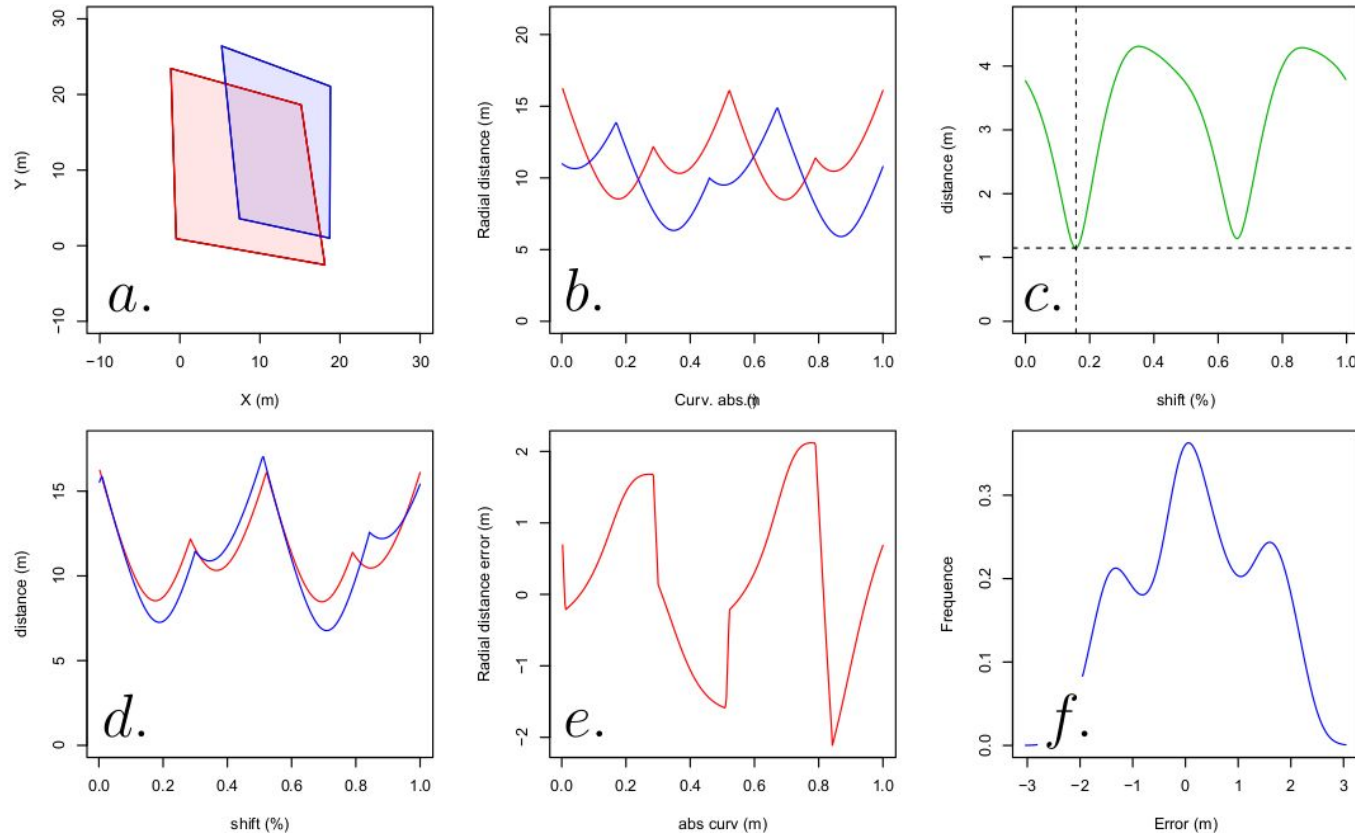


Localisation et emprises des bâtiments dans la seconde zone d'étude

Mesure cible : la distance radiale

- Distance radiale entre un bâtiment OSM et son homologue dans une base de référence(**BD TOPO®**), utilisée comme mesure de précision de forme
- Signature radiale d'un polygone :
$$r(s)=d(O,P(s))$$
 - avec s l'abscisse linéaire normalisée, O le centre de masse du polygone, $P(s)$ le point du polygone d'abscisse linéaire s
 - normalisée de sorte que $\|r\|_{L^2}=1$
- Distance radiale : Différence minimale entre signatures radiales

Mesure cible : la distance radiale



Calcul de la
distance radiale

- A des propriétés de pseudo-distance et sépare bien les formes
- Est une bonne mesure de précision de forme

Indicateurs intrinsèques de qualité

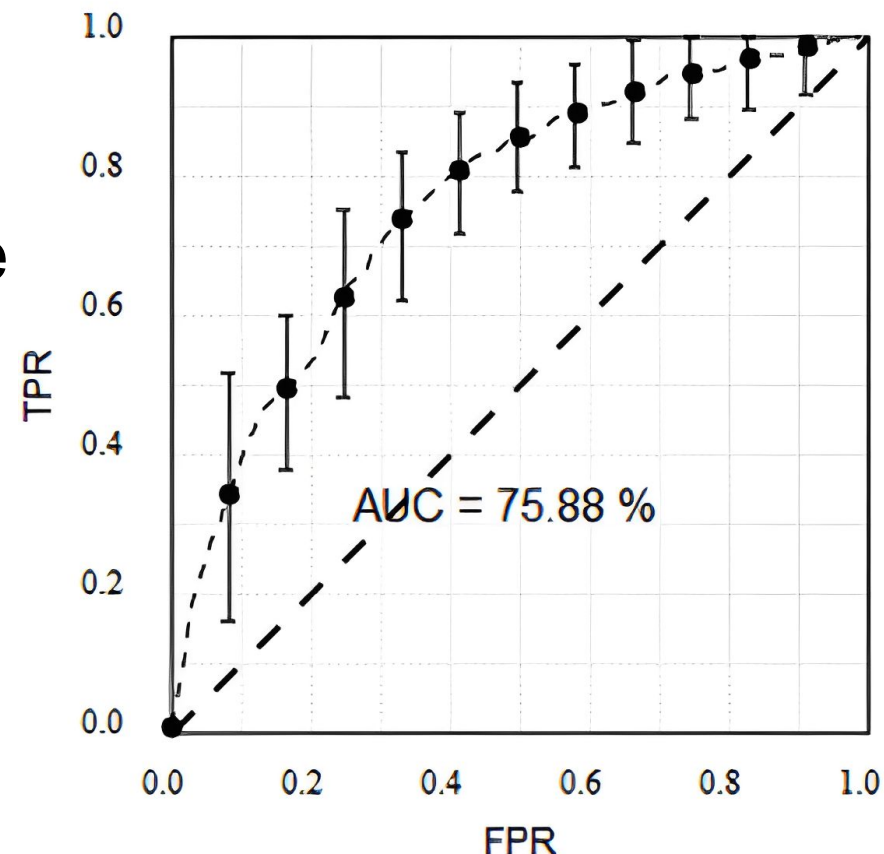
- Indicateurs calculés en utilisant uniquement la géométrie des objets
- Choix d'indicateurs intrinsèques
 - indicateurs connus pour leur pertinence dans la mesure de la qualité (granularité, compacité)
 - large gamme d'indicateurs pour une étude exploratoire
 - exemple : longueur moyenne des côtés, convexité, présence d'*outliers*, restructurabilité à partir d'une version sous-échantillonnée

Méthode d'apprentissage : forêt aléatoire

- Classification binaire : identifier les objets ayant une distance radiale supérieure à un certain seuil (positifs)
- Forêt aléatoire : combinaison d'arbres de décision
- La moyenne des verdicts des arbres donne une probabilité inférée que l'objet soit positif
- Choix d'un seuil de probabilité pour classer les objets comme positifs ou négatifs
- 3-fold cross validation

Initial results

- Courbe ROC
 - représente la performance d'un classifieur pour l'ensemble des seuils de probabilité
 - TPR : capacité à trouver les positifs
 - FPR : proportion des objets négatifs étiquetés incorrectement
 - l'aire sous la courbe (AUC) mesure une performance globale de la méthode
- AUC=75% → opère une discrimination convenable, mais insuffisante pour un usage opérationnel

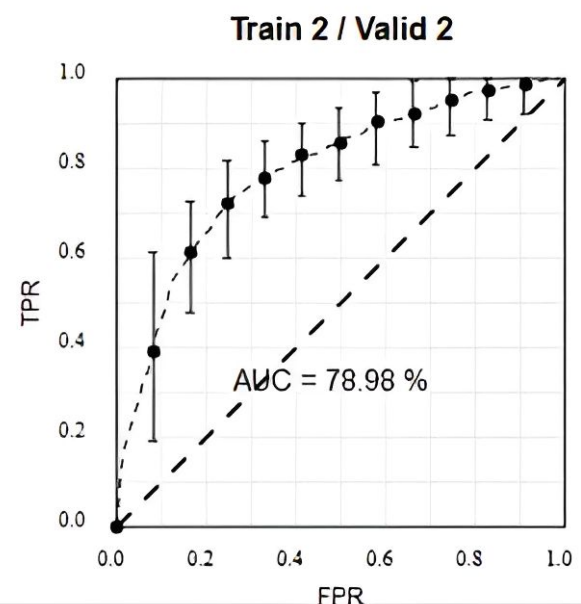
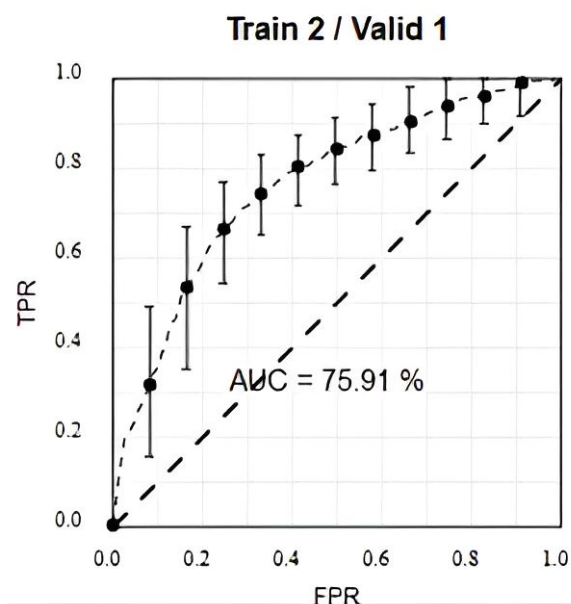
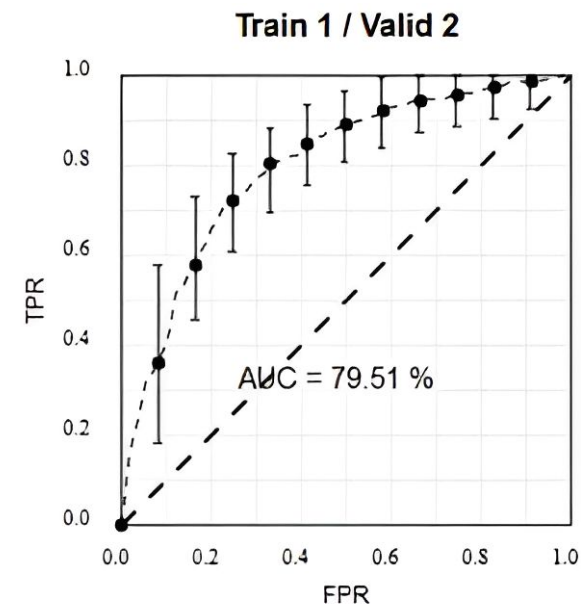
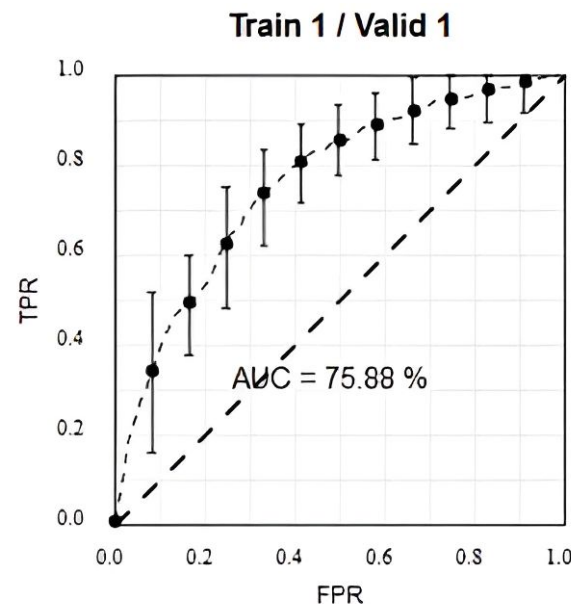


Étude de la transférabilité

- 4 expériences
 - 2 algorithmes entraînés sur la première zone, et évalués sur la première zone (Train 1 / Valid 1) et sur la seconde zone (Train 1 / Valid 2) respectivement
 - 2 algorithmes entraînés sur la seconde zone, et évalués sur la première zone (Train 2 / Valid 1) et sur la seconde zone (Train 2 / Valid 2) respectivement
- Validation croisée pour les expériences utilisant la même zone pour l'entraînement et l'évaluation
- Tous les algorithmes sont entraînés avec le même nombre d'exemples

Résultats

- Performances très similaires
- Possibilité d'utiliser un algorithme entraîné avec des données contrôlées sur une nouvelle zone
- A confirmer avec d'autres types de zone (montagne, côtes, autres pays)



Conclusion

- La méthode de classification effectue une séparation acceptable, mais insuffisante pour un usage opérationnel
- Un algorithme entraîné sur une zone peut être efficace sur une autre zone
- Perspectives d'amélioration:
 - La transférabilité doit être contrôlée sur des zones plus diversifiées
 - Possibilité d'étendre à la précision de position
 - Mobiliser des indicateurs prenant en compte le contexte spatial (bâtiments voisins, autres thèmes...)