

Webinaire “Applications des méthodes de machine learning en hydrologie” ...

... urbaine: fouille et fusion de données pour la cartographie des réseaux d'assainissement

Projets THUB, Cart'Eaux, MeDo et thèse CIFRE

Nanée Chahinian, Carole Delenne, Yassine Belghaddar et pleins d'autres gens!

**GDR RHYMA
18 mars 2022**

Petit historique...

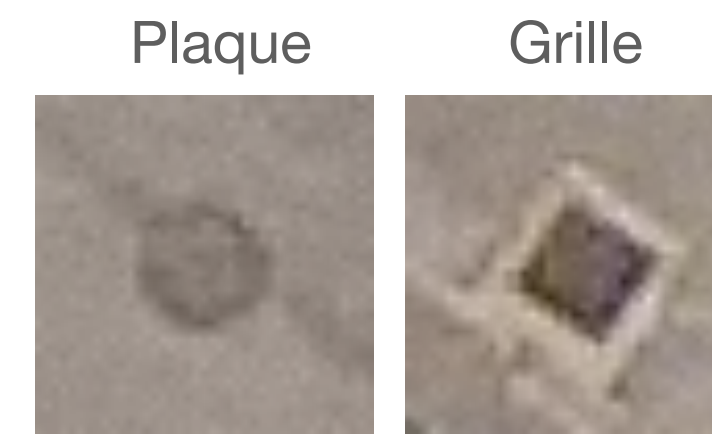
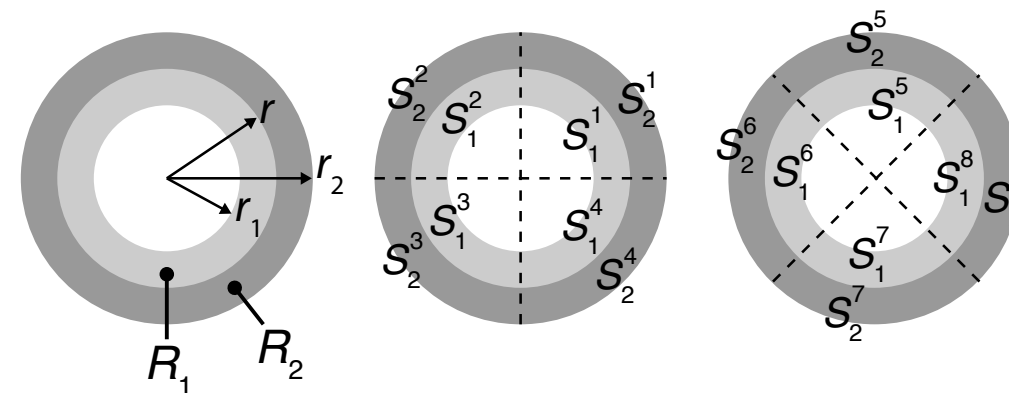
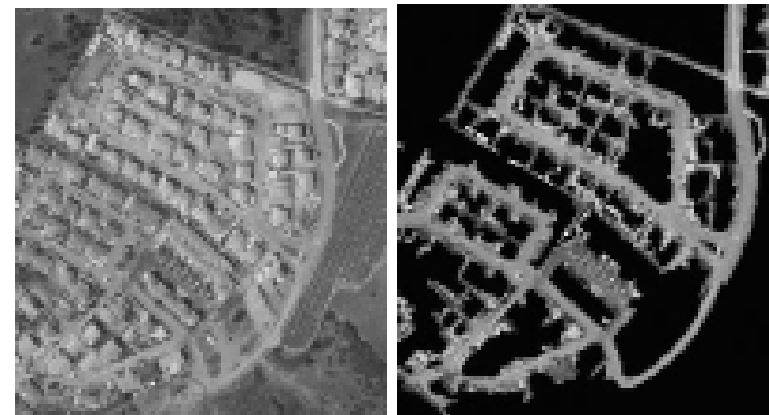
Réseau d'assainissement de Fès: aucune donnée

PNTS « THUB » (2014 N. Chahinian) + projet HSM

- Acquisition d'une image à 5cm de résolution spatiale sur Gigean

Méthodologie

- Pré-traitement: détection des routes, filtre de la végétation et des ombres
- Détection des plaques d'égout: filtre circulaire basé sur la différence de niveaux de gris entre l'objet et l'environnement (coefficient de Bhattacharrya [Niigaki2012])



Rencontre avec l'équipe ICAR (LIRMM) ...

- Comparaison avec une méthode par apprentissage (SVM) développée pour la détection des tombes (Pasquet et al., 2015 JURSE)

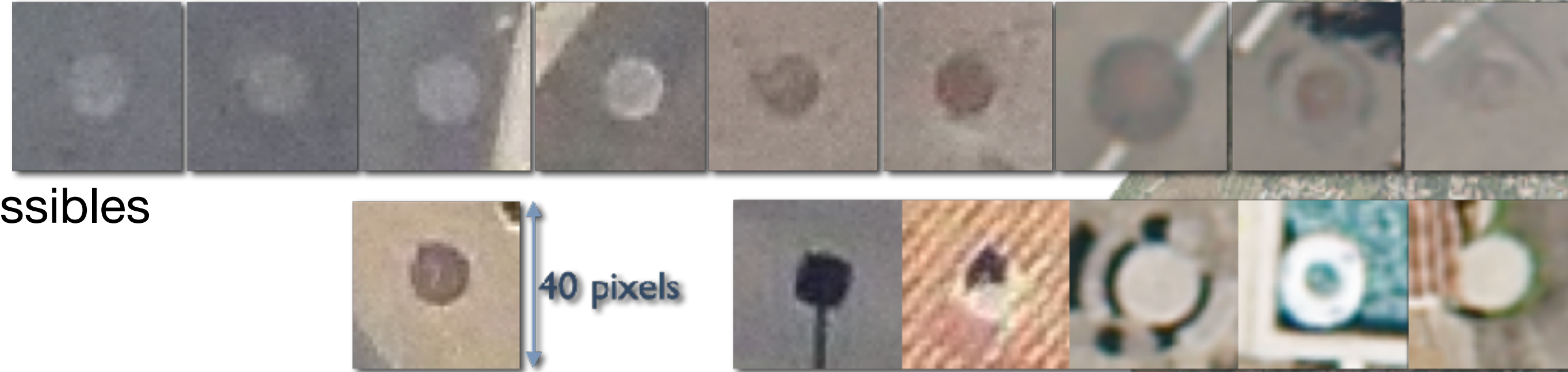
Projet Cart'Eaux

Chercheur d'avenir 2015 (LIRMM, HSM, IRD + LISAH et TETIS)

Détection des plaques d'égout sur image aérienne

► Difficultés:

- Petits objets
- Peu de contraste
- Nombreuses confusions possibles



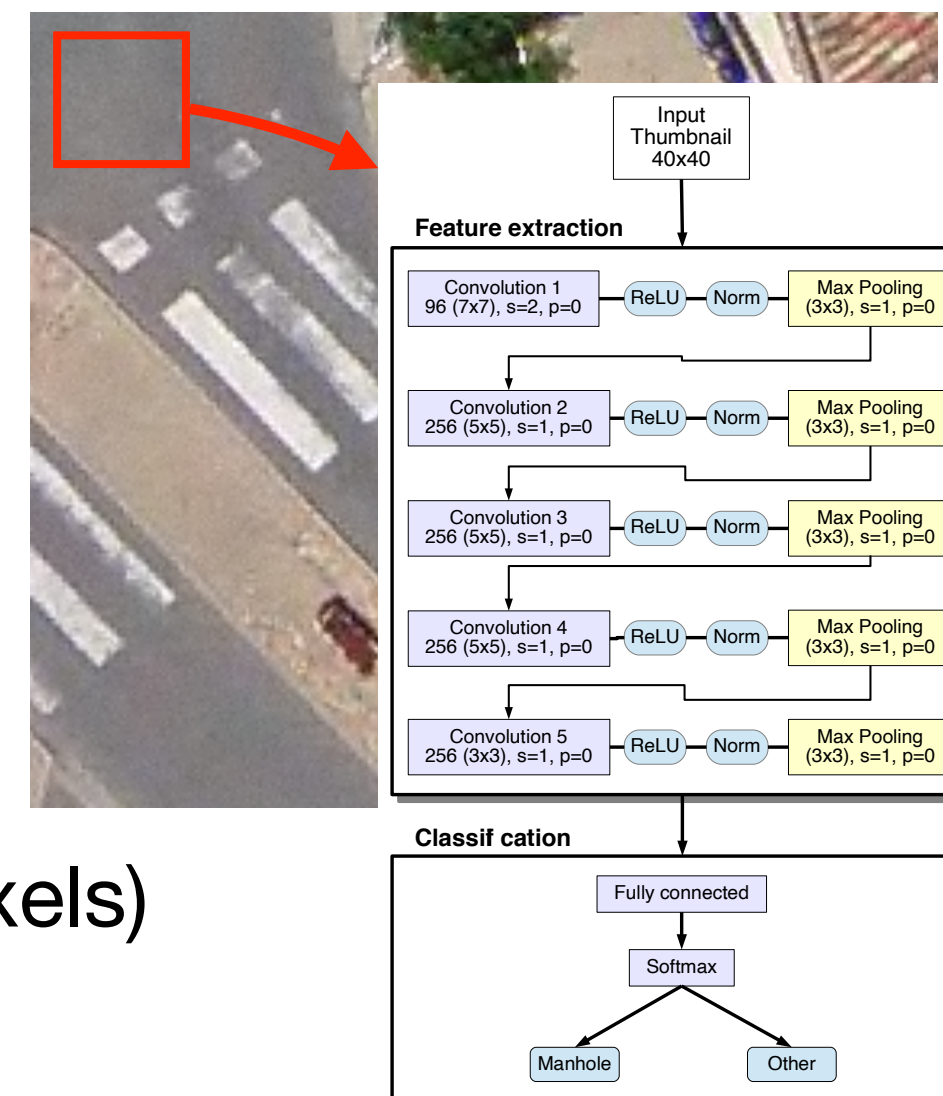
Deep Learning

► Base d'apprentissage:

- Prade le Lez: image en couleurs naturelles à 5cm
- 605 plaques => augmentation de la base (env 20 000 plaques 50 000 autres)

► Réseau AlexNet:

- Convolutional Neural Network CNN
- Paramètres adaptés à la détection des petits objets
- 5 couches
- Fenêtre glissante 40x40 sur toute l'image (pas de 4 pixels)



0 250 500

Projet Cart'Eaux

Validation

Image de Gigean (34): 100 plaques

Comparaison **détection** / **vérité**:

a=%de surface commune

- ▶ TP = True Positive
=> plaque correctement détectée: $a > 50\%$
- ▶ FP = False Positive
=> incorrectement classée comme « plaque »
- ▶ FN = False Negative
=> plaque non détectée



$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

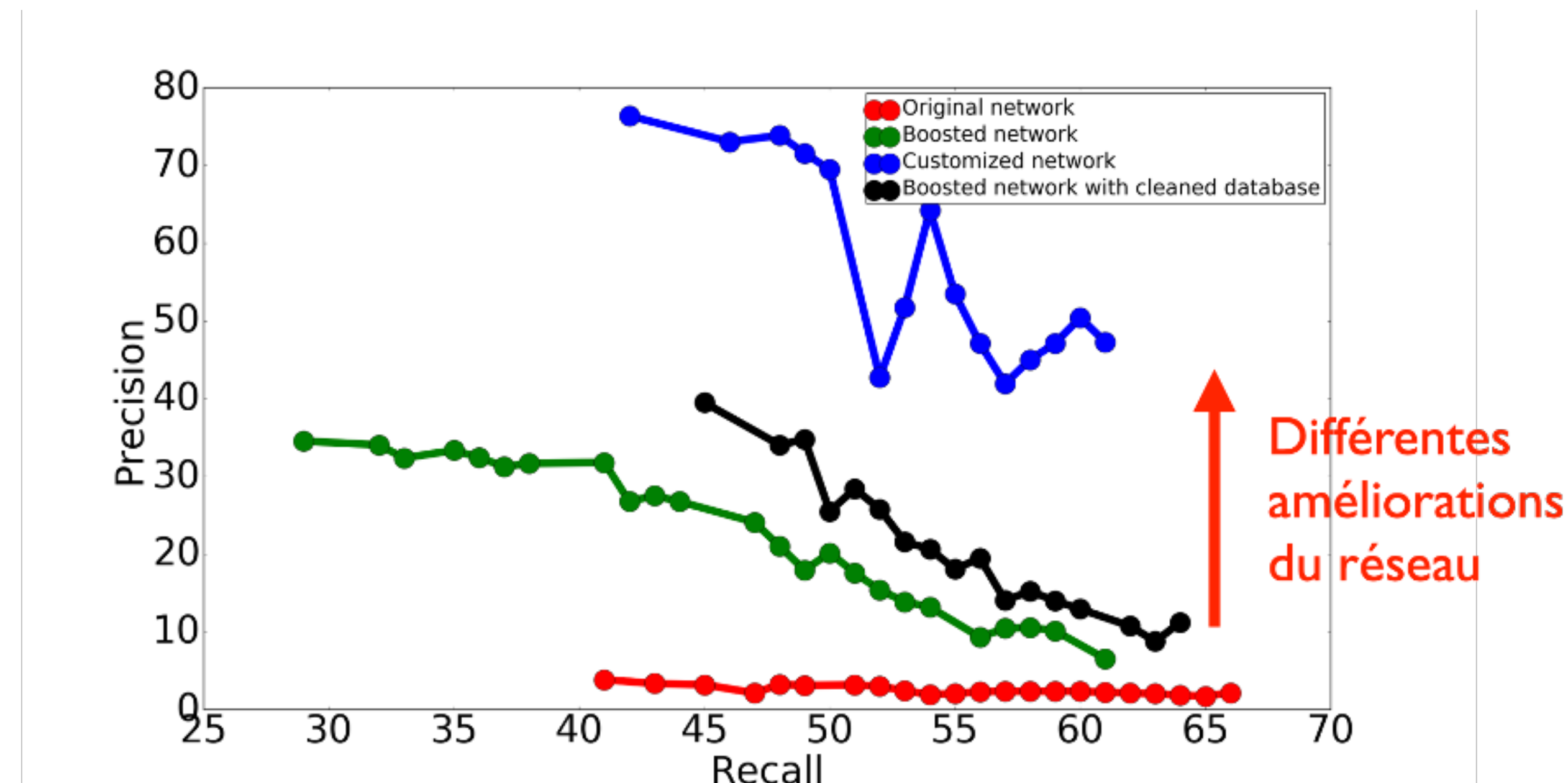
Total détecté

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Total vrai

Courbes ROC

- ▶ Précision fonction du rappel pour différents seuils d'acceptation



Encore beaucoup de FP

Références:

J. Pasquet et al. IEEE J.STARS, 2016 <hal-01275684>.

B. Commandre et al. ISPRS workshop Hanover <hal-01556762>.

Projet Cart'Eaux

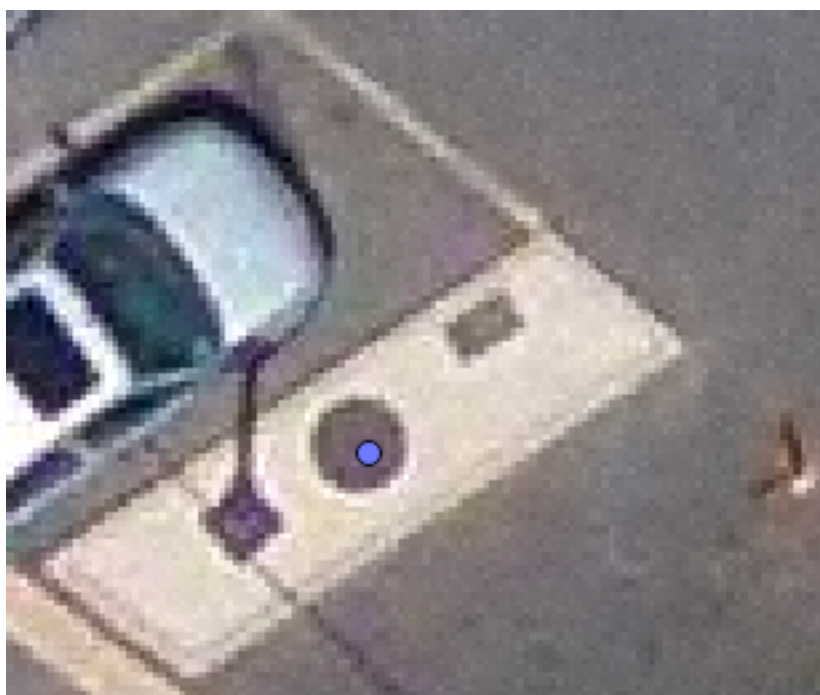
Perspectives

Meilleurs taux de détection possible: seuil de 0.26

- ▶ Rappel = 60% — Précision = 47%

Beaucoup de faux positifs:

- ▶ Certaines plaques ne sont pas dans la base de données (=> autre réseau? Anciennes ou nouvelles plaques?)
- ▶ Tâches sur le sol
=> améliorer le contraste?
=> regrouper les détections trop proches?
- ▶ Beaucoup de Faux en dehors des routes
 - Filtre sur les routes => Rappel = 54% — Précision = 70%
 - Vérification de la présence des plaques sur StreetView?



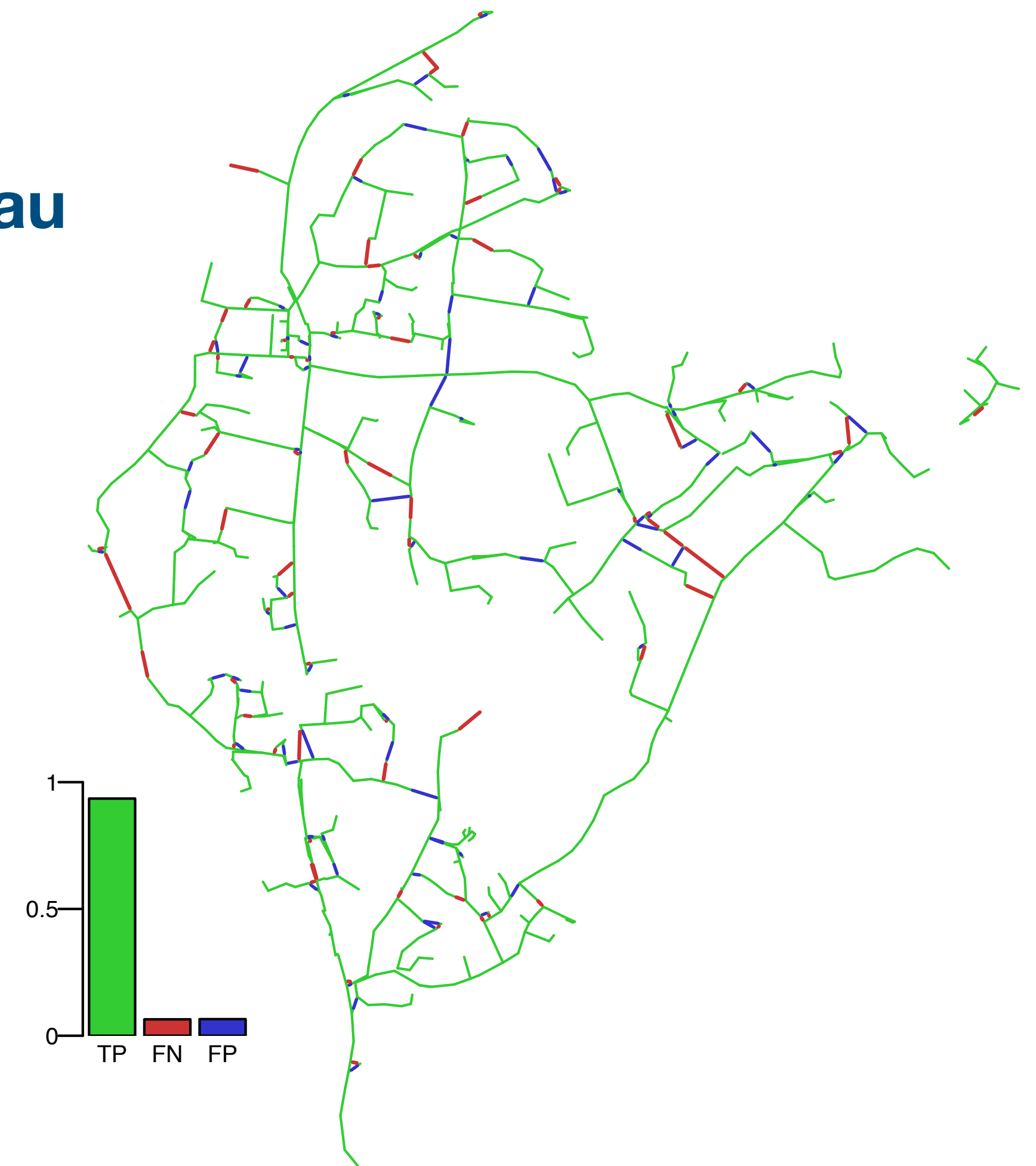
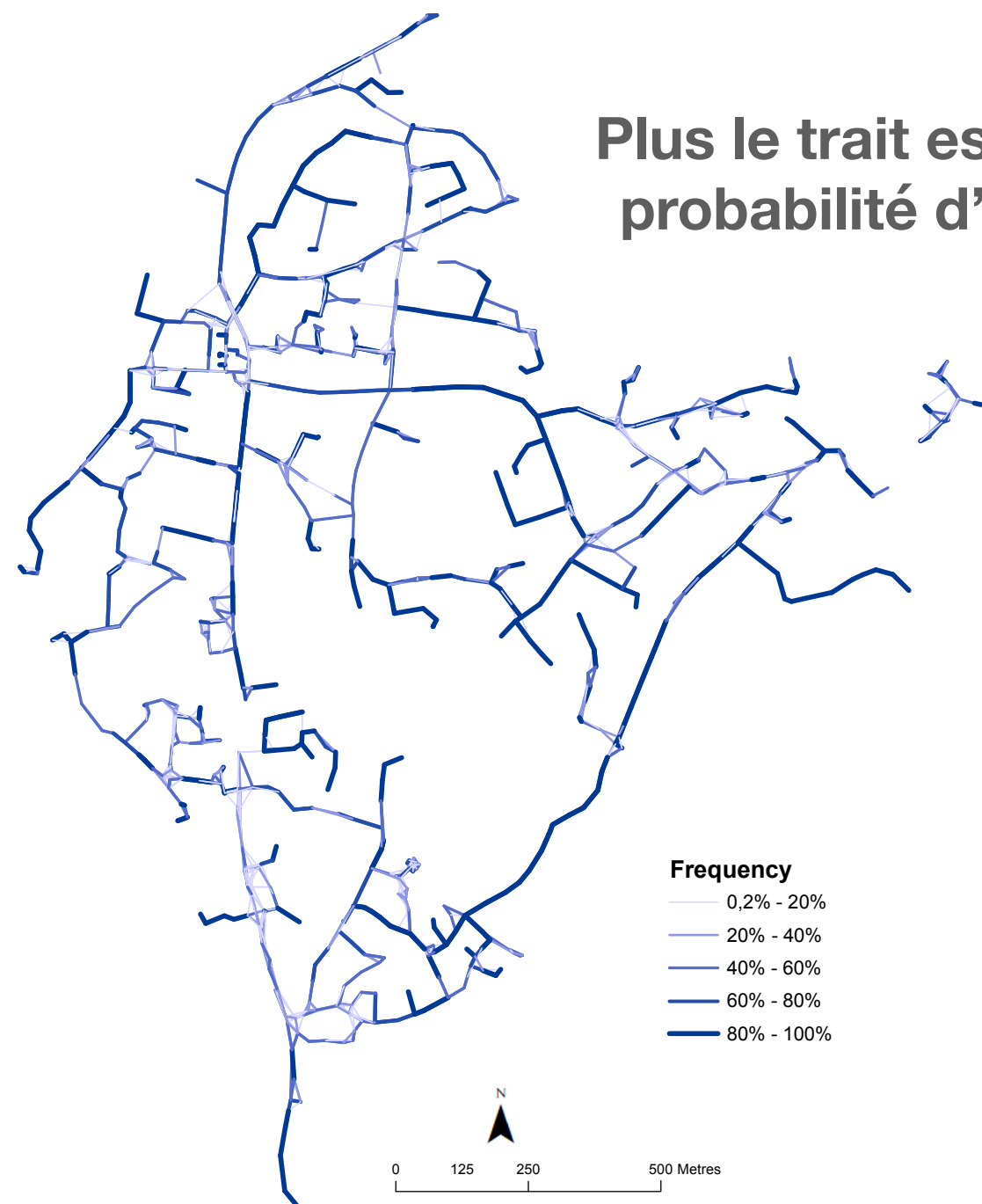
Projet Cart'Eaux

Algorithme de cartographie à partir des noeuds du réseau

Basé sur les fonctions coûts "métier"

- ▶ pente
- ▶ longueur
- ▶ angle entre deux conduites
- ▶ Pénalité si sort de la route ou croise un bâtiment ...

Version stochastique



Mais les caractéristiques des conduites sont toujours inconnues !

Références

- N. Chahinian et al. Computers, Environment and Urban Systems, 2019 <hal-02275903>.
B. Commandre et al., ICUD, 14th IWA/IAHR International Conference on Urban Drainage, 2017.

Thèse CIFRE Y. Belghaddar

Fusion de données et complétion des données manquantes

Mise en place d'un modèle générique pour la fusion de données multi-sources

- images, textes, bases de données...

Fusion de données géographiques

- Utilisation des fonctions de croyances pour l'appariement d'éléments de réseaux issus de plusieurs cartes

Complétion des données manquantes

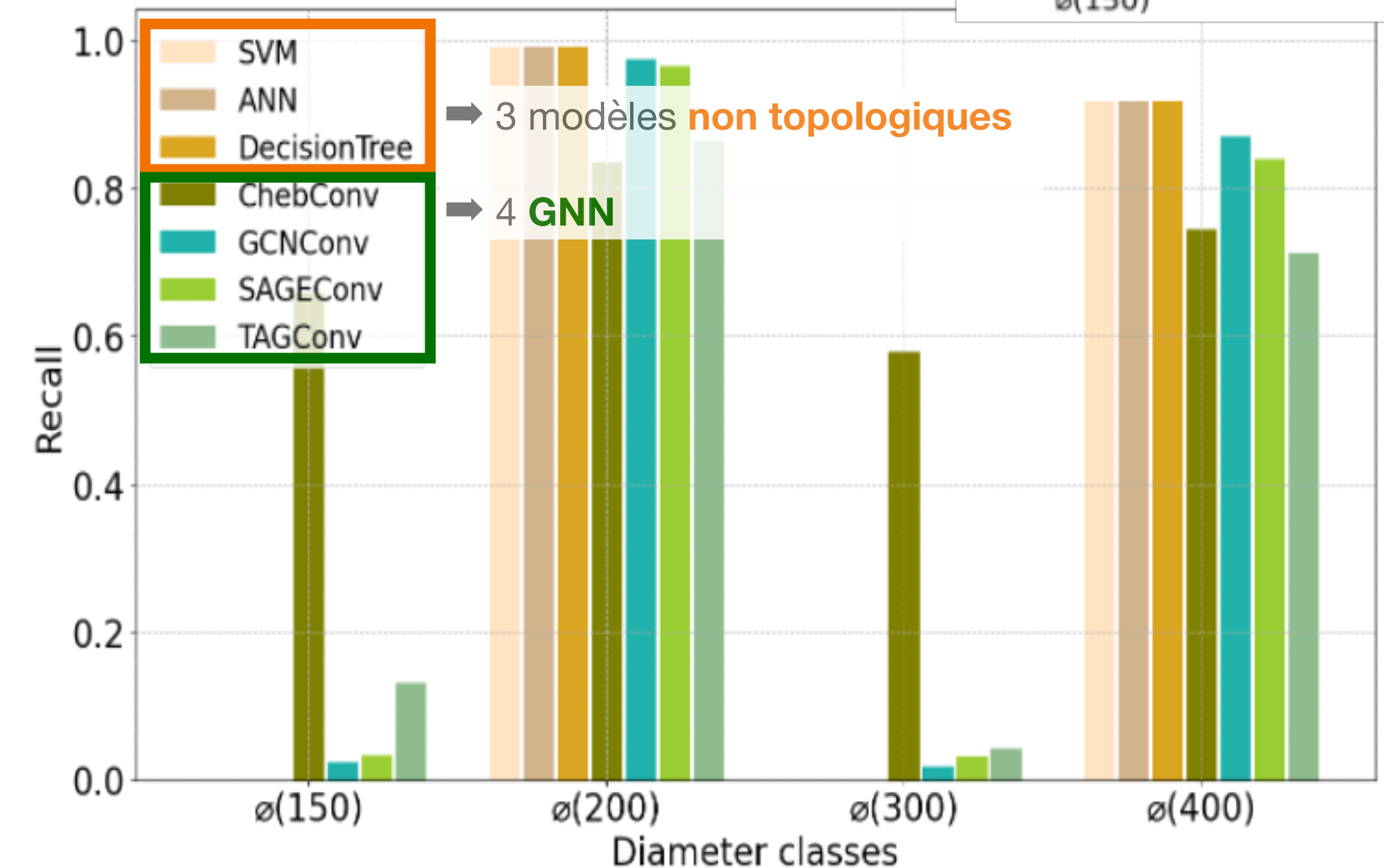
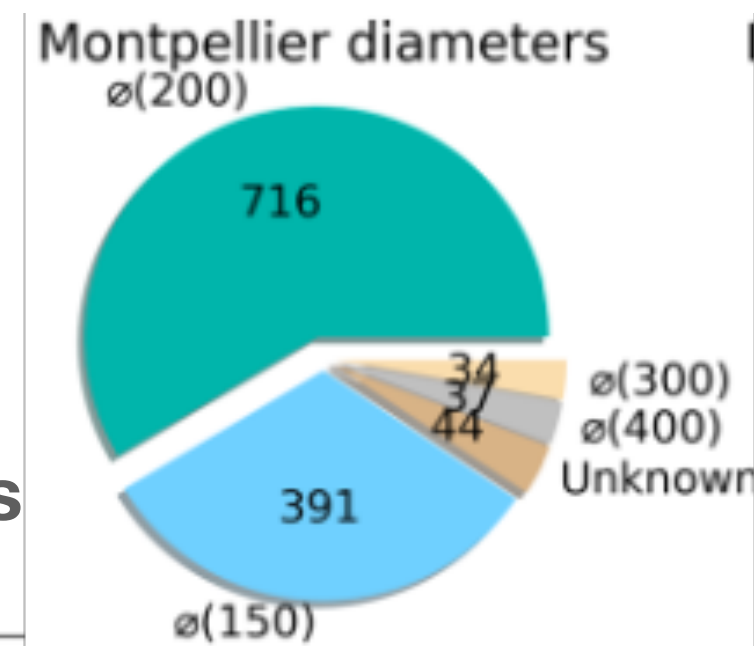
- Utilisation des Graph Neural Networks: prise en compte de la structure des réseaux => meilleure prédiction des classes minoritaires

Références

Belghaddar et al., Software and Systems Modeling, soumis

Belghaddar et al., Water, 2021, <hal-03264611>.

Exemple:
Prédiction des diamètres manquants



Rappel par classe (Ncorrects/Nclasse)

30% de la base pour l'apprentissage, 70% pour le test.

Mégadonnées, Données liées et fouille de données pour les réseaux d'assainissement

Objectifs:

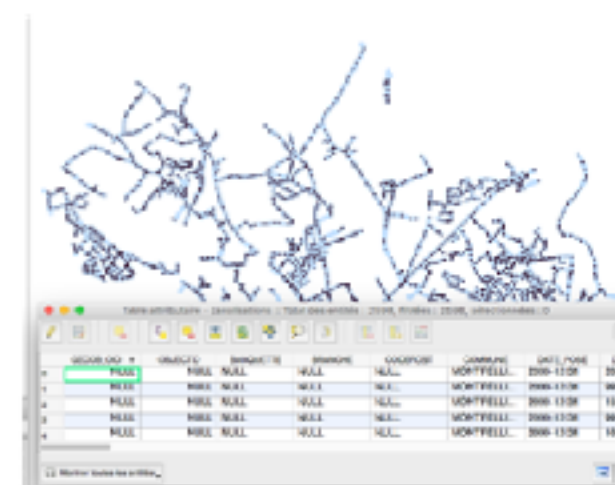
- ▶ Tirer profit des mégadonnées du web pour renseigner la géométrie du réseau et les données «annexes» pouvant servir aux gestionnaires.
- ▶ Créer d'un corpus relatif à l'histoire du réseau
- ▶ Faire l'historique des pannes et des interventions effectuées sur le réseau

Défis

- ▶ Traduire des informations textuelles non structurées en données quantitatives et en connaissance structurée du réseau.
- ▶ Transformer une information approximative en connaissance incertaine.

La **Métropole** va procéder à des travaux d'extension du **réseau d'eaux pluviales** **avenue de la Colline** dans le **quartier Cévennes**, à partir du **jeudi 20 octobre 2009**.
Ces **travaux**, qui débutent pour une **durée de 3 semaines**, consistent à réaliser une **extension** de la **conduite d'évacuation des eaux pluviales** de l'**avenue de la Colline**, notamment pour **raccorder** le **nouveau bâtiment du lycée Jules Ferry**. Cette **conduite** s'arrête **actuellement au-dessus de la rue Favier** et elle sera **prolongée jusqu'à la rue des Eglantiers**.
(source: www...)

Compatibilité BD gestionnaire



Quantification des incertitudes



Dynamique des trajectoires du réseau

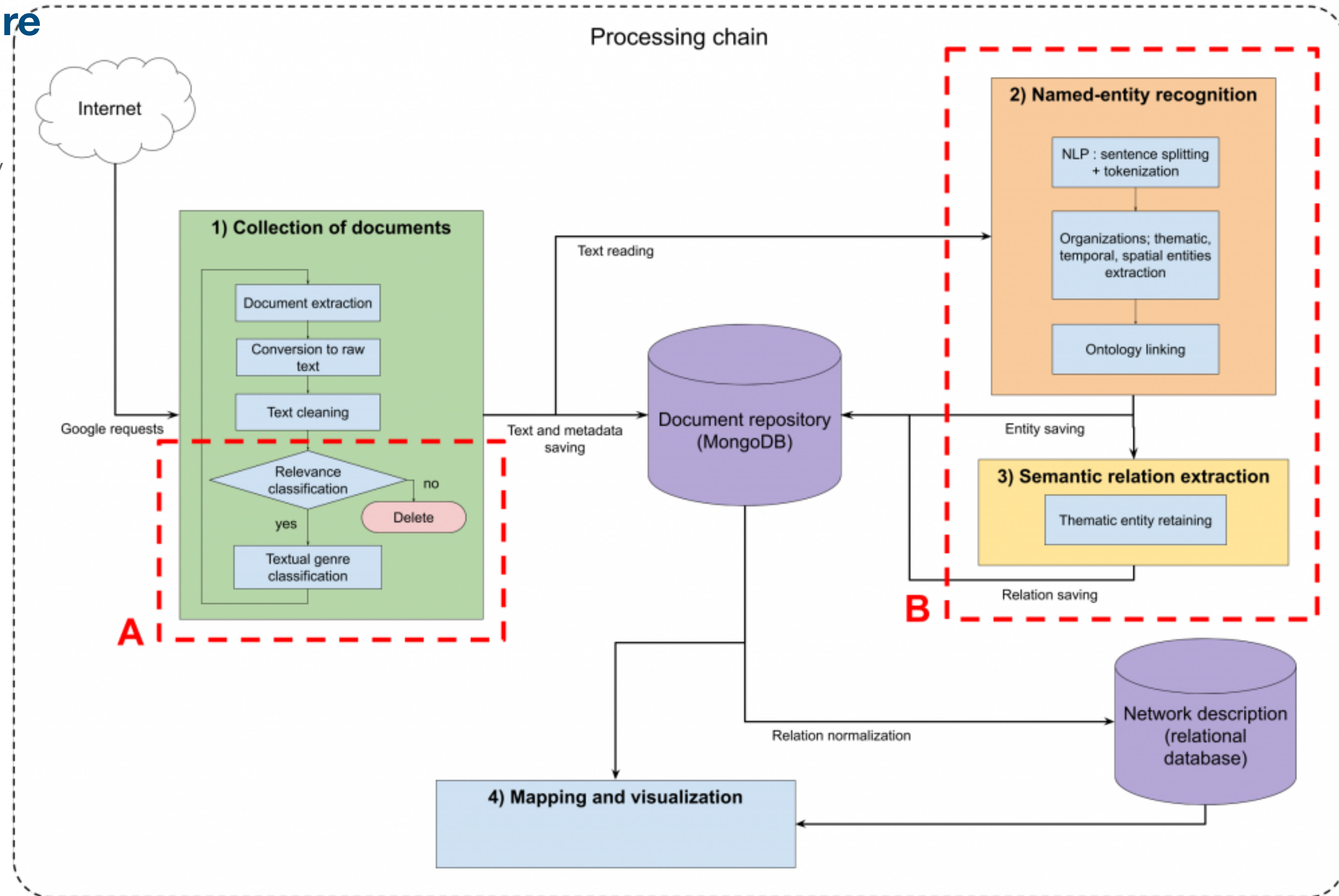


Organisation, Type réseau, Spatial, Temporel, Travaux, Élément réseau

Projet MeDo

Schéma de l'architecture

<http://webmedo.msem.univ-montp2.fr/>



Projet MeDo

Repères méthodologiques et premiers résultats

Méthode:

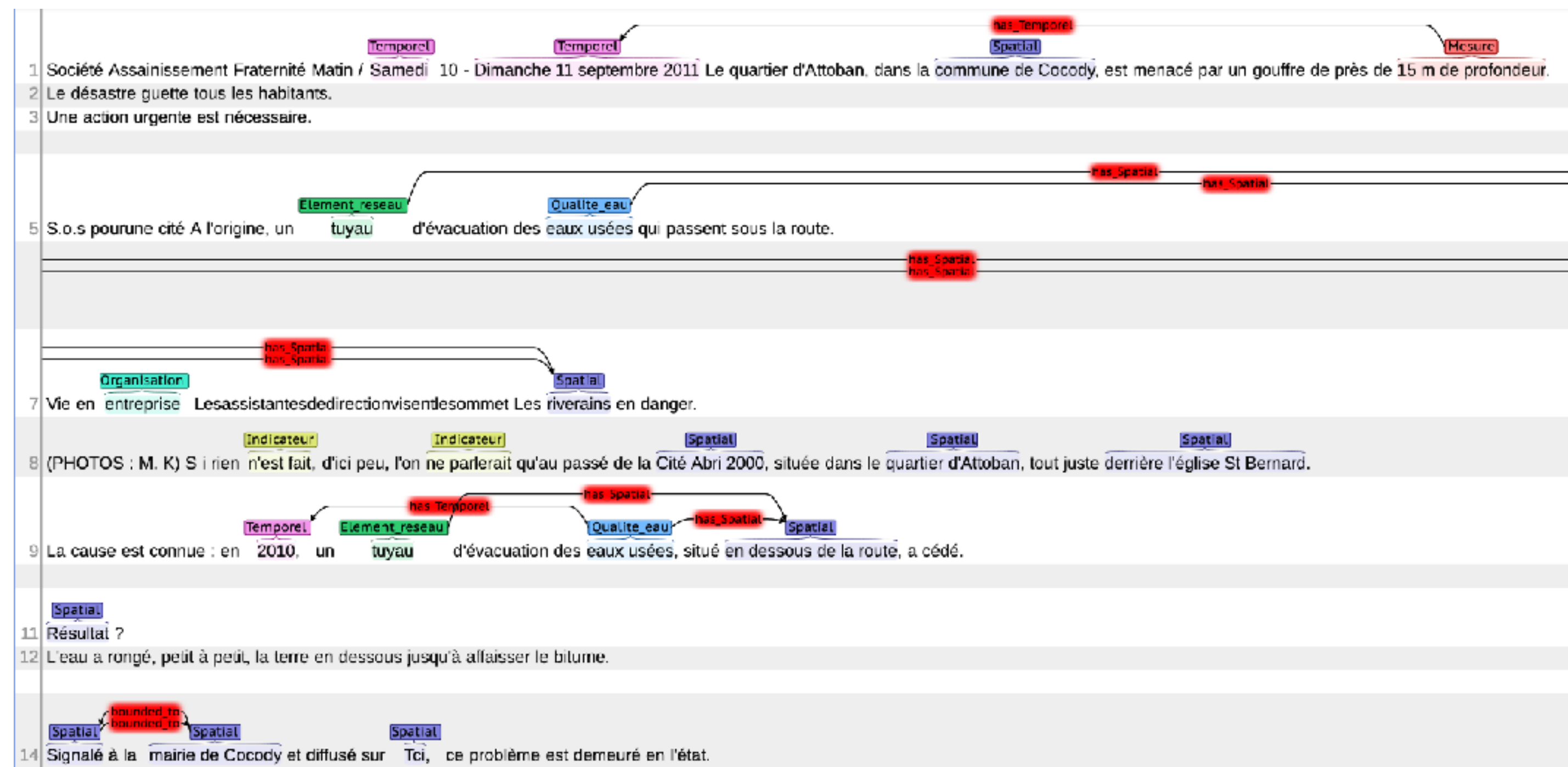
- Classification de genre/pertinence : Scikit-learn, Multinomial Naive Bayes.
- Modèle d'annotation spécifique : 13 entités (élément réseau, travaux, qualité...)
- Reconnaissance d'entités nommées : Core-NLP, SpaCy, Heideltime
- Mise en relation spatiale : BAN, Geonames & Nominatim

Corpus test :

- 1,557 documents (Montpellier, France)

Premiers résultats

- Bonne performance en classification (F-score >0.90)
- Utile pour détecter certaines entités e.g. travaux, éléments réseaux (F-score > 0.63).
- Pas beaucoup de données quantitatives (géométrie, débit)
- Validation sur d'autres villes : Abidjan ?



Projet MeDo

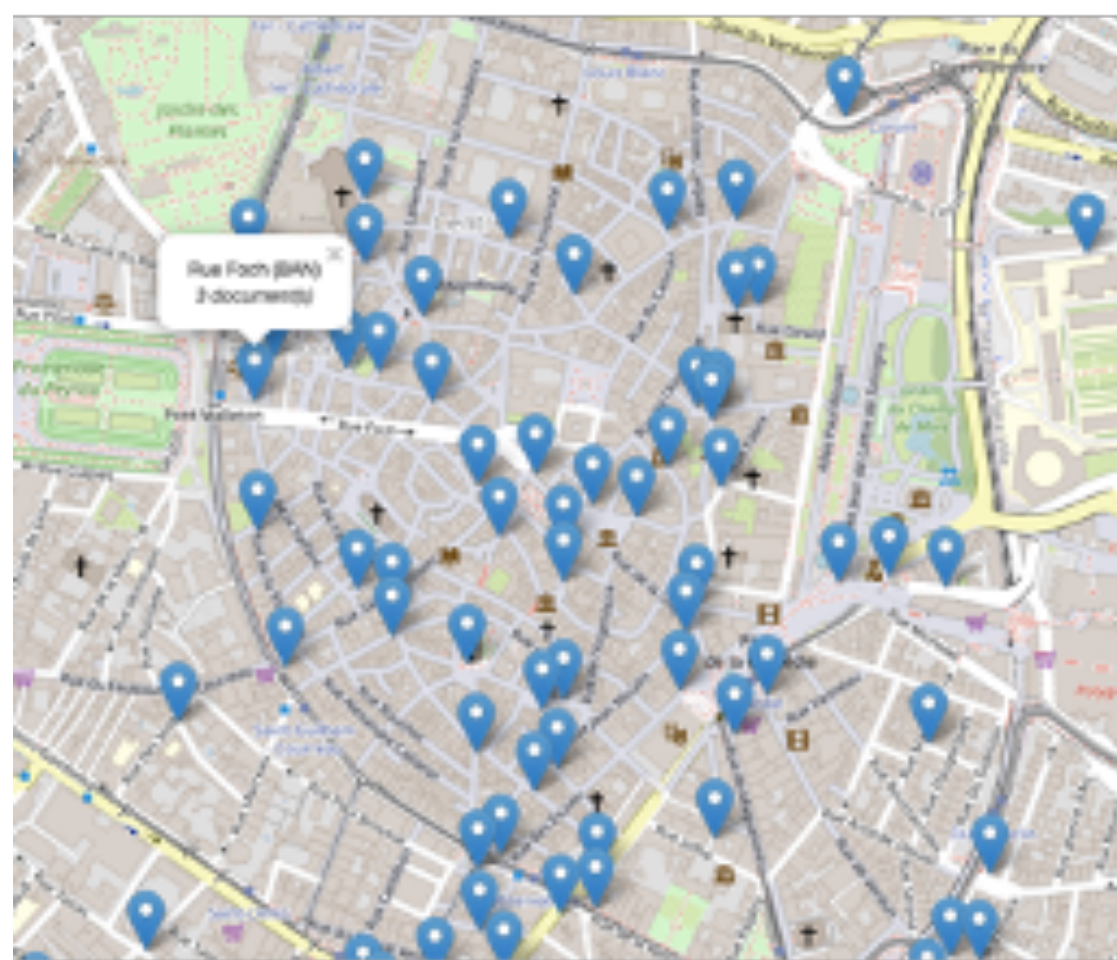
Transformer des informations en données

Objectif:

Intégrer les informations dans une base de données géographique

voir démo sur

http://webmedo.msem.univ-montp2.fr/?page_id=12



La **Métropole** va procéder à des **travaux d'extension** du **réseau d'eaux pluviales** **avenue de la Colline** dans le **quartier Cévennes**, à partir du **jeudi 20 octobre 2009**.

Ces **travaux**, qui débutent pour une **durée de 3 semaines**, consistent à réaliser une **extension** de la **conduite d'évacuation des eaux pluviales** de l'**avenue de la Colline**, notamment pour **raccorder** le **nouveau bâtiment du lycée Jules Ferry**. Cette **conduite** s'arrête **actuellement au-dessus de la rue Favier** et elle sera **prolongée jusqu'à la rue des Eglantiers**.

(source: www...)

