

POTENTIEL DES SOURCES DE DONNÉES COLLABORATIVES POUR L'INTÉGRATION DES POINTS DE REPÈRE ET DES ITINÉRAIRES POUR LE SAUVETAGE EN ZONE DE MONTAGNE

par Marie-Dominique Van Damme, Anna-Maria Olteanu-Raimond,
Yann Méneroux

Université Paris-Est, LASTIG, IGN, ENSG
94160 Saint-Mandé, France
prénom.nom@ign.fr

De plus en plus de sites web proposent de partager et de télécharger, entre amateurs de sport de plein air, des traces GPX d'itinéraires ainsi que des points de repère. La connaissance des points de repère le long des itinéraires ou des itinéraires qui passent près de points de repère peut aider les secouristes dans la phase de localisation d'une victime en montagne. Afin de constituer des données métier (i.e. secours en montagne), il est nécessaire d'étudier la pertinence des différentes sources de données et de définir leurs caractéristiques. Dans cet article, nous explorons le potentiel des données issues du web afin de déterminer leur bien-fondé dans le contexte du sauvetage en montagne. Notre objectif est donc d'étudier le contenu des différentes sources de données afin de mieux comprendre comment les points de repère et les itinéraires sont saisis, de démontrer la complémentarité des sources de données en comparaison avec des données d'autorité et d'étudier la possibilité de définir des liens entre les itinéraires d'une part et les points de repère d'autre part. La méthode proposée utilise les techniques d'intégration telles que le recalage de points GPS sur un réseau routier, la reconstruction d'itinéraires et l'appariement de données géographiques. Un premier résultat concerne le grand nombre de données non appariées, ce qui prouve une grande richesse et la complémentarité des sources. De plus, les résultats de l'appariement de données font émerger de nouvelles règles sémantiques concernant les points de repère et les géométries des routes.

Introduction

Plus de 15000 demandes de secours en montagne sont recensées annuellement en France. Malgré les avancées technologiques des systèmes de positionnement et le nombre croissant de géo-applications, la localisation des victimes pendant la phase d'alerte reste un problème pour les équipes de secours. Ne connaissant pas sa position exacte, la victime décrit son environnement spatial en utilisant des points de repère (« Je vois le sommet du Mont Blanc », « Je suis près d'un refuge »), ou en décrivant son itinéraire (« Je suis sur le GR20¹ », « Je marche sur un chemin qui longe un lac »). Les itinéraires et les points de repère peuvent être désignés soit par des noms propres (parking du Col des Montets, sommet du Mont-Blanc, GR20) soit par des noms communs (refuge, route qui longe le lac). Ce type de description qualitative (i.e.

localiser un objet en combinant des relations spatiales et des caractéristiques de référence) est connu dans le domaine de la géomatique sous le nom de géoréférencement indirect (Hill et Zheng 1999). Ceci est cohérent avec le concept de géographie naïve inventé par Egenhofer et Mark (1995), où l'objectif est de modéliser comment les individus perçoivent l'espace. Ce raisonnement géographique naturel peut être rendu plus performant avec une approche d'intégration de données hétérogènes (Duckham, *et al.* 2006).

Pour passer de la localisation relative à la localisation absolue, i.e. en coordonnées géographiques, il faut avoir accès aux données spatiales de type points de repère et itinéraires. De plus, savoir que les points de repère sont à proximité d'un itinéraire ou que les itinéraires passent à proximité de certains points de repère peut aider le secouriste à

¹ GR est une abréviation européenne bien connue qui définit un chemin de longue distance pour la randonnée. Le GR20 est un célèbre GR en Corse de 179 km de long.

valider ou à invalider une hypothèse de localisation de la victime. Ainsi, un nouveau besoin identifié est de définir les relations qui existent entre les points de repère et les itinéraires. Interroger différentes sources de données s'inscrit dans la problématique du web sémantique (Ferrara, Nikolov, et Scharffe, 2013). Traditionnellement, les points de repère font partie des données faisant autorité, issues des institutions telles que les agences nationales de cartographie ou les gazetiers (par ex. Wikipedia, GeoNames), la plupart du temps produites en mode collaboratif ou provenant de sources de données externes, y compris celles faisant autorité (e.g. des organismes public locaux). Au cours de la dernière décennie, de plus en plus de sites web collaboratifs offrent la possibilité de partager et de télécharger des points d'intérêt et des itinéraires. Ces sites fournissent des informations techniques (topoguide, difficulté, etc.), des points de repère et des itinéraires au format GPX qui servent à différentes fins : utilisation de la trace dans une application mobile, suivi des performances de l'utilisateur, etc. Cartographiées et saisies par l'utilisateur, ces données collaboratives ont différents degrés de fiabilité et des structures hétérogènes (e.g. catégories de points de repère, cotations de difficultés différentes, irrégularités dans la description des itinéraires). En raison de la variété des données, de nombreuses requêtes doivent être exécutées dans des sources de données hétérogènes sur le Web pour obtenir des informations relatives aux itinéraires ou aux points de repère. Les données collectées doivent ensuite être intégrées pour constituer une réponse aux secouristes. Ce processus fait partie de la collecte et de l'intégration de données hétérogènes.

Ce travail de recherche s'inscrit dans le cadre du projet CHOUCAS (choucas.ign.fr) qui vise à aider les équipes de secours en montagne à mieux estimer une zone de localisation probable de la victime. L'architecture choisie pour ce projet repose sur une architecture orientée services pour faciliter la fraîcheur des données (Halilali, Gouardères, Devin, *et al.*, 2018). Un médiateur permet, à l'aide d'un catalogue de services, de diriger les requêtes sur différentes sources de données indexées par le système au moyen d'attributs fonctionnels et non fonctionnels. Il est donc nécessaire d'identifier les sources de données pertinentes de type points de repère et itinéraires et de décrire leurs caractéristiques afin de faciliter la définition des liens les unissant pour construire l'indexation de recherche.

Au cours de cette analyse, nous avons exploré le potentiel des données collaboratives candidates comme sources d'information dans le contexte du sauvetage en montagne. Aussi, notre objectif est-il

d'étudier le contenu des différentes sources d'information afin d'avoir une meilleure connaissance de la description des points de repère et des itinéraires, de démontrer la complémentarité des données collaboratives comparées aux données faisant autorité, et d'étudier la faisabilité de définir des liens entre les itinéraires et les points de repère.

Nous évoquons ensuite, en section 2, un aperçu de la littérature existante. La section 3 décrit les sources d'information que nous avons jugées pertinentes en zone de montagne. Les sections 4 et 5 exposent notre approche pour évaluer le potentiel des données collaboratives dans le contexte du sauvetage en montagne et les résultats obtenus. Enfin, la conclusion et les perspectives sont présentées en section 6.

État de l'art

Les points de repère et les itinéraires sont des données essentielles dans différents domaines tels que la navigation (Denis, 1997; Zhou *et al.*, 2017), le géocodage de texte (Moncla, *et al.*, 2014) ou la recherche d'informations géographiques (Schockaert, Cock, et Kerre, 2008). L'analyse des points de repère et la description des itinéraires font également l'objet de nombreuses recherches visant à analyser comment les gens décrivent leur environnement spatial, les lieux ou leurs itinéraires (Goodchild et Li, 2012 ; Schwering *et al.*, 2013).

D'un point de vue cognitif, selon Lynch (1960), les points de repère sont des entités du monde réel permettant à un humain de mieux s'orienter dans un espace qui ne lui est pas toujours familier. Dans le domaine de la description de l'espace, les points de repère sont considérés comme un référentiel (Zhou *et al.*, 2017), tandis que dans le domaine du géocodage textuel, les points de repère représentant les noms de lieux sont considérés comme des points de passage pour géocoder les noms propres (Moncla *et al.*, 2014).

Dans le domaine de la géomatique, selon Teulade-Denantes, Maudet, et Duchêne (2015), les itinéraires représentent des chemins, des parcours ou des sentiers à suivre et sont portés par une infrastructure de réseau physique composée de routes et de sentiers piétonniers. Les itinéraires sont caractérisés par une géométrie et une description textuelle. Dans le domaine de la recherche géographique, la description des itinéraires est étudiée à l'aide de techniques de traitement automatique de la langue pour définir des instructions d'orientation où les points de repère sont identifiés comme des éléments essentiels pour caractériser l'itinéraire (Denis, 1997). Ainsi, les points de repère et les itinéraires ne sont pas des

concepts séparés mais sont liés puisqu'un itinéraire est composé d'un ensemble de repères et de tronçons (Laddada et Saux 2017, Denis, 1997).

Dans la dernière décennie, les descriptions des itinéraires sont aussi disponibles que les traces GPX des itinéraires. De nombreuses recherches utilisent les traces GPX pour estimer des temps de marche en montagne (Pitman, *et al.*, 2012), pour identifier les catégories d'itinéraires de randonneurs et leurs pauses (Kerouanton, *et al.*, 2017), pour construire des réseaux piétonniers (Kasemsuppakorn et Karimi, 2013), pour mettre à jour des données spatiales faisant autorité (Ivanovic, *et al.*, 2016) ou pour enrichir les informations thématiques des réseaux routiers et pour proposer des itinéraires de substitution, dans le cadre de la protection des espaces naturels (Vias et Ocaña Ocaña, 2014).

Un autre aspect important est d'identifier quelles sont les catégories de points de repère et d'itinéraires en zone de montagne. Bien que de nombreuses recherches aient été effectuées pour étudier les points de repère et les itinéraires en milieu urbain (Timpf, 2005), peu de recherches se concentrent sur les points de repère et les itinéraires en milieu naturel, en milieu ouvert et encore moins dans les zones de montagne. Brosset, Claramunt et Saux (2008) ont confirmé le rôle pertinent des points de repère dans la description des itinéraires dans les parcs et les zones forestières et ont constaté que les objets représentés par des points, des lignes et des polygones sont utilisés pour décrire les itinéraires. Concernant la navigation côtière, les points de repère situés en mer ou sur la côte sont utilisés pour définir des itinéraires de bateaux (Laddada et Saux 2017). Des points de repère tels que les rochers, les plateformes, les bouées de danger isolées, les phares sont identifiés à partir de l'analyse des descriptions d'itinéraires pour des navires, faites par des experts. Derungs et Purves (2014) s'intéressent à la description des objets naturels dans un corpus annoté dans le milieu de la montagne (les Alpes suisses). Ils ont identifié 95 catégories d'objets naturels qui ont été comparées avec d'autres études de la littérature, découvrant qu'il y a un chevauchement des catégories pour les 10 plus fréquentes ainsi qu'une grande diversité des catégories en raison de la spécificité du corpus. Dans ce même contexte de la montagne, Egorova, Tenbrink, et Purves (2015) analysent la description textuelle d'itinéraires alpins issus d'une plateforme collaborative afin d'étudier l'importance des points de repère en montagne. Les différentes catégories de points de repère identifiées sont regroupées en quatre groupes : objets du paysage, volet spatial, terrain et constellation.

Contrairement aux itinéraires (descriptions textuelles et fichier GPX) qui sont de nouvelles données traditionnellement non produites par les agences nationales de cartographie, les points de repère peuvent être trouvés à la fois dans les données faisant autorité et dans les données collaboratives. De plus, bien que les points de repère définis par des toponymes puissent provenir de gazetiers, ce n'est pas le cas pour les points de repère exprimés par des toponymes vernaculaires ou des lieux sans nom (*e.g.* un refuge). Ainsi, une autre question de recherche importante est d'étudier la qualité des points de repère provenant de données collaboratives. Évaluer la qualité des données collaboratives en général, et en particulier celle des points de repère, a fait récemment l'objet de différents travaux de recherche. Acheson, De Sabbata et Purves (2017) ont constaté, en comparant les gazetiers GeoNames et Getty Thesaurus of Geographic Names, une hétérogénéité thématique : les lieux habités ont une plus grande couverture que les éléments naturels comme les montagnes, les rivières ou les collines. Ils suggèrent que «les données nationales devraient être utilisées de préférence lorsqu'il s'agit d'éléments naturels». Le même constat d'hétérogénéité thématique est mis en évidence dans les données collaboratives où les collaborateurs préfèrent cartographier certains types d'objets (Bégin, Devillers et Roche, 2013). En ce qui concerne l'hétérogénéité spatiale des données collaboratives, de nombreuses études ont montré qu'il existe une grande hétérogénéité spatiale entre les zones urbaines et les zones moins habitées en raison de la densité de population ou de l'attractivité des lieux (Neis et Zielsstra, 2014). L'exactitude de la localisation des données collaboratives est aussi fréquemment étudiée dans la littérature, et il a été montré que l'exactitude est moins bonne pour les éléments naturels (Girres et Touya, 2010 ; Bégin *et al.*, 2013). D'un point de vue sémantique, Acheson et ses collaborateurs (2016) ont remarqué une hétérogénéité sémantique entre les gazetiers (*e.g.* les montagnes représentent les sommets et les chaînes de montagne dans DBPedia et seulement les sommets dans GeoNames).

Les contributions de ces travaux de recherche sont : 1/ l'utilisation de points de repère, d'itinéraires ou d'une combinaison des deux pour mieux localiser les victimes dans les zones de montagne. Cette approche est légèrement différente de l'approche généralement proposée dans le contexte de la navigation (Denis, 1997), où les points de repère font partie de l'itinéraire. 2/ Générer des connaissances sur la qualité des points de repère et des itinéraires à partir de données collaboratives dans le milieu naturel. 3/ Identifier de nouvelles catégories de points de repère et d'itinéraires dans les zones de montagne.

Description des sources de données

Dans cette étude, nous nous sommes concentrés sur des données provenant de sites web actifs dédiés aux sports de loisirs de montagne, où la communauté et les auteurs des sites font régulièrement des mises à jour, recommandées par les autorités ou par les pratiquants, éventuellement partenaires de l'IGN, publiant suffisamment de données et traitant au moins d'une source thématique spécifique à la montagne. Parmi les sources de données considérées comme les plus pertinentes, nous avons sélectionné les sites web participatifs : RefugeInfo, Camptocamp (C2C) et Visorando. Parmi les sources de données d'autorité, la base de données topographiques BDTPO de l'IGN ainsi que deux sites d'Aires Protégées (AP) sont utilisés pour compléter notre analyse. Presque toutes les données sont fournies via des API. Mentionnons que d'autres portails de données, comme OpenStreetMap (OSM), GeoNames, IGNRando fournissent des points de repère ou des itinéraires mais ces données sont souvent stockées en masse et proviennent de différentes sources de données externes, la source est alors souvent mentionnée (par exemple, dans OSM, certains points de repère sont référencés C2C ou wikidata). Dans cet article, ces sources de données n'ont pas été prises en compte parce que le schéma de données est trop généralisé, que des informations spécifiques sont perdues (comme des informations techniques, les traces GPX, lien entre les itinéraires et les waypoints) ou ne sont pas appropriées (IGNRando utilise le schéma cirkwi dédié aux informations touristiques).

BDTopo® de l'IGN

Parmi les différents thèmes de la base de données topographiques produite par l'IGN, les données utilisées dans cette étude concernent les Points d'Intérêt ou d'Activités (PAI) et les Routes (BDTOPO, 2018). La couverture des données est nationale. Le thème PAI regroupe les points d'intérêt et les activités classées par catégories : administratif, secours, transport, etc. Chaque élément est représenté par sa localisation (x,y), l'origine des données, son type, son nom et sa pertinence. La couche Routes contient à la fois les routes et les chemins ; les données sont modélisées en réseau et sont décrites par type, nom, pertinence, etc.

Refuge.info

Refuge.info est un site web collaboratif proposant différentes données géographiques ponctuelles en zones de montagne. Même si le site Web était focalisé sur les refuges à l'origine, d'autres types de points d'intérêt peuvent y être trouvés aujourd'hui tels que

les points d'eau, sommets, cols, hôtels, camping, parking, bus. Chaque point est caractérisé par sa localisation, son nom, son type et différentes descriptions, telles que comment y accéder, nombre de places, origine des données, etc. Les localisations sont généralement obtenues par numérisation à l'aide de cartes. Concernant la couverture spatiale, les données sont disponibles principalement en France métropolitaine.

Camptocamp.org

Camptocamp (C2C) est un projet collaboratif auquel participent des pratiquants plutôt avertis des sports de montagne. L'objectif du site web est de partager des informations techniques (guides topo, rapports d'incidents et d'accidents, sorties avec les conditions météorologiques et accessibilité) et de contribuer à la sécurité des utilisateurs de l'application. C'est une communauté active qui propose une large couverture, tant du point de vue des activités de montagne que du point de vue géographique (France, Allemagne, Suisse, Espagne). Toutes les données (routes, waypoints, sorties, rapports d'accidents, photos) sont liées les unes par rapport aux autres.

Point de repère. Chaque point est caractérisé par sa localisation, son nom, son type et certaines informations thématiques telles que l'altitude, l'exhaustivité (c'est-à-dire la qualité montrant dans quelle mesure la contribution tient compte des recommandations de partage), la description textuelle et les photos. Différentes catégories existent telles que sommet, col, lac, chute d'eau, localité, canyon.

Itinéraire. Les itinéraires représentent différents types de randonnées en montagne, très peu ont une trace GPX (25% dans notre zone d'étude). La création des géométries est réalisée par les contributeurs avec leurs propres outils de gestion de données et leur propre fond cartographique. Un itinéraire se caractérise par un titre, un type (par exemple une boucle, un aller-retour), une cotation (difficulté globale d'un itinéraire), différentes informations liées à l'altitude (altitude minimale, maximale, dénivelé, pente, etc.), une description textuelle et des photos.

Aires protégées

Parmi les aires protégées (AP), les Parcs naturels régionaux (PNR) et les Parcs nationaux (PN) sont des institutions publiques françaises ayant pour mission de préserver l'environnement naturel et de développer des activités économiques et touristiques. Par ailleurs, les Parcs du territoire français gèrent des informations telles que les points de relief, les points d'intérêt patrimoniaux, les services (informations pratiques comme les points d'eau, les passages difficiles)

et les itinéraires. La plupart d'entre eux utilisent les logiciels de la suite GeoTrek, les données de la BDTopo ainsi que le Modèle Numérique de Terrain (MNT) de l'IGN comme référentiel pour administrer leurs données. Mentionnons que ces données d'autorité (BDTopo et base de données sur les aires protégées) ne sont pas équivalentes en termes de contenu parce qu'elles ont des points de vue différents (topographique vs. environnemental et touristique). Elles sont plutôt complémentaires avec une petite redondance en termes de types de fonctionnalités et d'informations thématiques. L'entrée des données fournies par l'API est l'itinéraire ; les points de repère y sont reliés par référencement linéaire. Les points de repère sont décrits par leur localisation, type, nom, description textuelle et altitude. Le nombre de points de repère est fixé à trois au minimum et ces points doivent être bien répartis sur le parcours. Des types tels que point de vue, point d'eau, abri, col, lac, glacier, sommet, pastoralisme, histoire, sont présents. Les itinéraires sont définis par une géométrie fournie dans un fichier GPX, un titre, une activité sportive, une cotation, des informations liées à l'altitude, une description textuelle, etc.

Visorando

Visorando est un site web offrant la possibilité de télécharger des traces publiées en ligne par des contributeurs. Contrairement à C2C, 100% des géométries des itinéraires ont été produites à partir de données issues de capteurs GPS. En plus des traces, des attributs tels que le titre, le type d'activité (randonnée, vélo, etc.), la description textuelle de l'itinéraire, la durée, la difficulté et le dénivelé sont fournis. Le site Web conseille aux contributeurs d'ajouter plusieurs points de passages aux traces (points de repère ou changement majeur de direction). Les traces sont ensuite ajustées et validées par les contributeurs dans une interface web cartographique utilisant les fonds cartographiques de l'IGN (données faisant autorité).

Méthode

Afin d'identifier et évaluer le potentiel des données issues du web, nous devons définir les relations sémantiques et les relations de correspondance entre les schémas des différentes sources de données. Pour y parvenir, des techniques d'intégration telles que l'appariement de données sont employées. L'hétérogénéité des données provenant de ces différentes sources rend plus complexe la tâche d'appariement.

Notre méthode se déroule en trois étapes. La première étape consiste à extraire les données de dif-

férentes sources d'information. Pour faciliter cette tâche, un plug-In QGis² a été développé. La deuxième étape concerne l'appariement des données. L'objectif est de définir :

1. des itinéraires homologues : sachant qu'un itinéraire est un concept porté par une infrastructure de réseau physique, nous considérons donc que deux itinéraires sont homologues s'ils suivent des chemins proches.

2. et des points de repère homologues : deux points de repères sont homologues s'ils représentent le même objet du monde réel.

La troisième étape vise à évaluer les caractéristiques des objets homologues trouvés en définissant des matrices de confusion ou en évaluant la redondance d'information en calculant des mesures de connectivité dans le réseau d'itinéraires fusionnés. La figure 1 illustre notre approche pour l'appariement des points de repère et des itinéraires.

Les sites web fournissent, à défaut de la géométrie linéaire des itinéraires, au moins la trace GPX d'origine. Pour faciliter l'appariement des itinéraires, il est donc préférable de reconstruire leur géométrie. Deux étapes de pré-traitement sont proposées. Tout d'abord, les points vont être recalés sur un réseau routier de référence en appliquant l'algorithme proposé par Newton et Krumm (2009) fondé sur les modèles de Markov cachés. Cette méthode est plus appropriée dans notre cas, car pour chaque point à recalculer sur le réseau, l'algorithme tient compte des points avant et après, la cohérence des itinéraires étant ainsi bien conservée. Une difficulté concernant cet algorithme est de choisir les bons paramètres : trouver le bon rapport entre les erreurs de positionnement et la tolérance à suivre des routes indirectes. Par exemple, dans la figure 2, le point P1 ne doit pas être mis en correspondance avec le tronçon le plus proche afin de maintenir la cohérence de l'itinéraire et de ne pas créer une impasse.

La deuxième étape du pré-traitement consiste à reconstruire la géométrie linéaire de l'itinéraire. Les points issus de la trace de l'itinéraire sont parfois très éloignés les uns des autres, ce qui peut créer des casures entre les tronçons portant les points recalés. Notre processus intègre l'ajout de ces tronçons manquants, en calculant le chemin le plus court sur le réseau routier entre deux tronçons non connectés. Si le rapport entre la distance à vol d'oiseau et la distance du plus court chemin est inférieur à un seuil fixé, les tronçons manquants sont ajoutés à l'itinéraire (tronçons verts sur la figure 2). A la fin de cette étape, la géométrie d'un itinéraire est composée d'une liste de

2 <https://github.com/ANRChoucas/PluginChoucas>

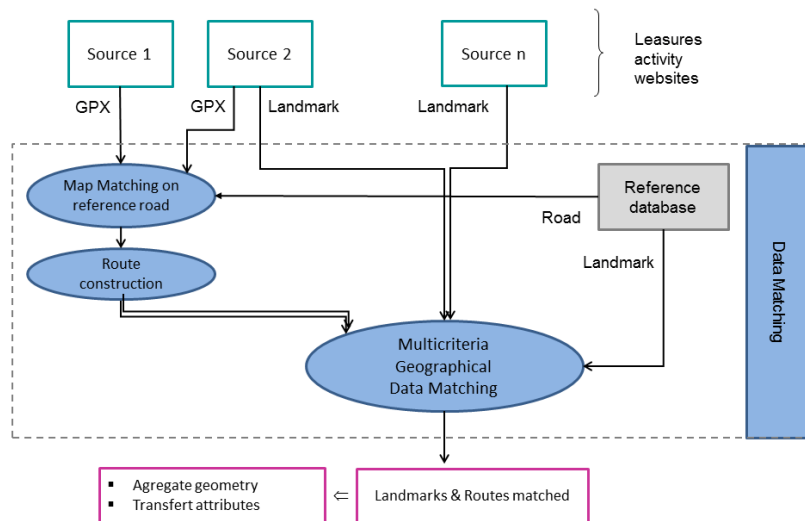


Figure 1 : Méthode proposée pour l'appariement des itinéraires et des points de repère

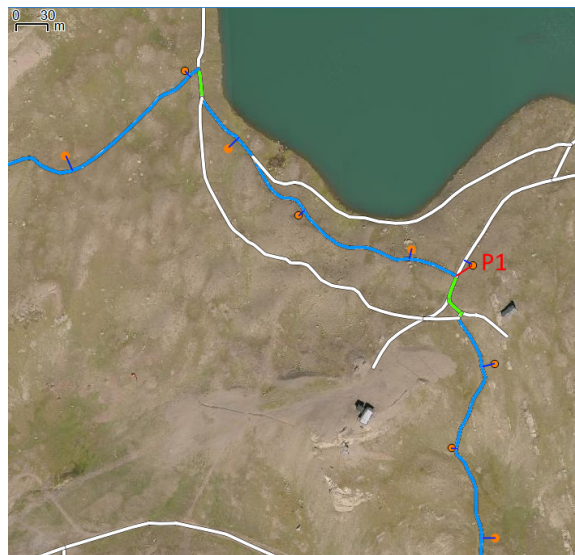


Figure 2 : Points GPS (orange) recalés sur le réseau routier de la BDTopo (bleu). Le point P1 est recalé avec le nœud du réseau pour maintenir la consistance de l'itinéraire.

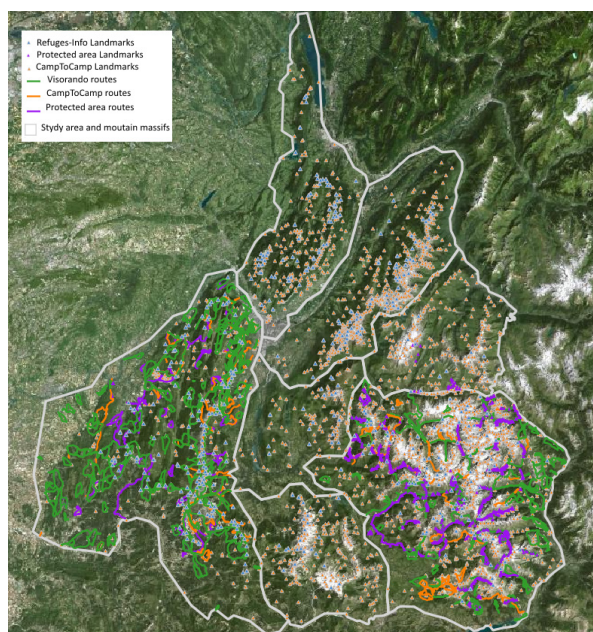


Figure 3 : Zone d'étude : zone de montagne dans les Alpes françaises

tronçons du réseau routier.

Une fois les étapes de pré-traitement terminées, l'appariement de données peut être effectué séparément pour les routes et les points de repère. L'algorithme d'appariement de données appliqué est un appariement de données multi-critères open-source³ s'appuyant sur la théorie des fonctions de croyances (Olteanu-Raimond, Mustière et Ruas, 2015). L'algorithme d'appariement s'utilise sur un jeu de données de référence et sur un jeu de données de comparaison donnant ainsi une direction à l'appariement (pour chaque donnée du jeu de référence, l'algorithme recherche les données homologues et candidates dans le jeu de comparaison). Mentionnons que le jeu de données de référence peut être un jeu de données faisant autorité ou un jeu de données collaboratif.

Connaissant les caractéristiques des jeux de données issus des sites web décrits dans la Section 3, nous avons opté pour deux critères (position et nom) pour l'appariement des itinéraires et trois critères (position, toponyme et sémantique) pour l'appariement des points de repère. Les données de la BDTOPO seront utilisées comme données de référence.

Le critère de position s'appuie sur la distance géométrique entre l'élément de référence et un candidat (à savoir la distance euclidienne pour les points de repère et la moyenne de la distance moyenne de Hausdorff pour chaque tronçon routier des itinéraires). Le critère toponymique compare le nom de l'élément de référence avec le nom du candidat. Parmi les différentes mesures testées pour comparer les chaînes de caractères, la distance de Samal (Samal *et al.*, 2005) et la distance Cosinus sont considérées comme les plus appropriées pour les points de repère et les routes respectivement. Enfin, le critère sémantique compare les catégories des points de repère. Afin de calculer les similitudes entre les catégories, la mesure Wu-Palmer (Wu et Palmer, 1994) est appliquée en utilisant l'ontologie de domaine (Mustière *et al.*, 2011). Tous ces critères sont enfin fusionnés pour prendre la décision finale.

Nous pouvons noter que l'algorithme d'appariement de données ne prend aucune décision (*i.e.* les objets géographiques ne sont pas appariées) si les critères sont contradictoires (par exemple dans le cas de deux objets homologues potentiels, si le nom est identique et si la sémantique est différente avec une distance importante). Ces cas sont étiquetés comme étant des cas incertains et seront gérés manuellement.

Résultats

Description de la zone d'étude

Notre étude s'est portée sur une partie des Alpes françaises, près de Grenoble (9 870 km²), sur un territoire couvrant sept massifs montagneux (figure 3). Concernant les points de repère, dans notre zone d'étude, les 25 906 points sont répartis comme suit : BDTOPO (21 097 points), C2C (3 229 points), Refuge.info (478 points), AP (1 102 points). Les points de passage de Visorando ont été exclus car leurs catégories n'étaient pas bien identifiées et le nom ne correspondait pas directement à un toponyme. Les itinéraires sont localisés, quant à eux, au cœur de deux massifs alpins de la zone d'étude, le Vercors et les Écrins. Les 706 itinéraires se répartissent comme suit : Visorando (398), C2C (127) et AP (181). Dans notre approche, nous n'avons pas tenu compte des relations spatiales et topologiques existant entre les points de repère et les itinéraires. Toutes les sources ne fournissent pas nécessairement les deux types de données. De plus, les types de relations entre point de repère et itinéraire ne sont pas clairement définis, un lien peut représenter une relation de proximité aussi bien qu'une relation d'intervisibilité. Les résultats seront donc analysés séparément. Comme le montre la figure 3, la couverture des points de repère et des itinéraires est bien diffuse. La précision des itinéraires est hétérogène d'une source de données à l'autre, en moyenne, la distance entre deux points consécutifs est de 25 m, 50 m, 5 m pour Visorando, C2C et AP respectivement.

Résultats d'appariement

Les paramètres concernant le recalage des points des traces sur le réseau routier sont de : 80 m pour la recherche de candidats et 4 m pour l'écart type des mesures d'erreur GPS. Dans l'ensemble, 97% des points (298 348 points) sont recalés et la racine carrée de l'erreur quadratique moyenne des déplacements induits par le recalage est égale à 12,94 m. Il est à noter que quelques impasses ont été générées après le processus de recalage des traces. Concernant l'étape de reconstruction de la géométrie linéaire des itinéraires, le rapport maximal entre la distance à vol d'oiseau et la distance du chemin le plus court entre deux tronçons non connectés est estimé empiriquement à 5. La longueur totale des tronçons du réseau routier manquants ajoutés est égale à 337 km, soit 4% de la longueur totale des itinéraires. Ceci montre la pertinence du processus de reconstruction des routes.

Pour les paramètres de l'appariement de données multi-critères, le buffer choisi pour la recherche des

³ <https://github.com/IGNF/MultiCriteriaMatching>

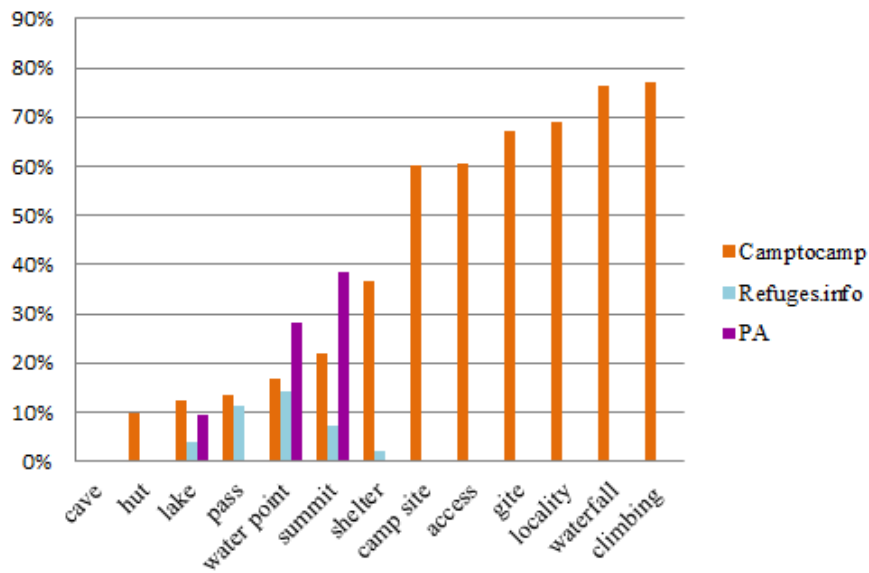


Figure 4 : Distribution des points de repère non appariés par catégorie

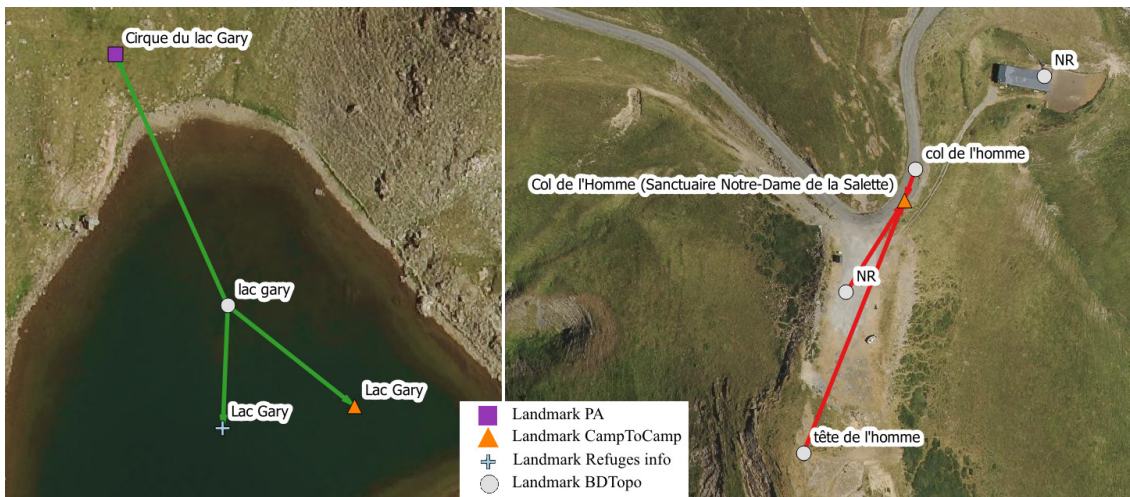


Figure 5 : Résultats d'appariement : lien 1 :1 à gauche ; lien incertain à droite



Figure 6 : Deux itinéraires appariés qui ne partagent pas tous les mêmes tronçons.

candidats est de 500 m et 300 m pour les itinéraires et les points de repère respectivement. Choisir les seuils appropriés est une tâche difficile, il faut trouver le bon compromis entre la qualité des résultats et le nombre de liens. Les seuils de l’algorithme d’appariement ont été fixés empiriquement en analysant les distributions des distances entre chaque couple d’objets des deux jeux de données et pour chaque critère ou en utilisant des seuils déjà suggérés dans la littérature (Olteanu-Raimond *et al.*, 2015). C’est le cas, par exemple, des seuils utilisés pour le nom et le critère sémantique pour les points de repère. Ce qui donne pour les itinéraires, des seuils pour le critère de position et le critère de similarité des noms de randonnées, 200 m et 0,3, respectivement. Pour les points de repère, les seuils ont été estimés respectivement à 100 m, 0,6 et 0,7 pour la position, le nom et la sémantique.

Résultats d’appariement des points de repère

Chaque source de données collaboratives est comparée séparément aux données de la BDTopo, base de données qui fait autorité dans notre étude. Les résultats obtenus sont présentés dans le tableau 1. Un lien 1:0 signifie qu’un point de repère d’une source de données (e.g. C2C) n’a pas de point de repère homologue dans les données faisant autorité (e.g. BDTopo), tandis qu’un lien 1:1 signifie qu’un objet homologue existe dans les données faisant autorité.

	1:0	1:1	Incertains
C2C-BDtopo	1167	1938	214
Refuges.info-BDtopo	80	367	34
AP - BDTopo	863	180	61

Tableau 1 : Résultats des appariements des points de repère.

En ce qui concerne les objets non appariés (colonne 2 du tableau 1), deux types de cas peuvent être observés. Premièrement, les objets ne sont pas appariés en raison de la faible précision des points de repère collaboratifs (*i.e.* le point de repère est éloigné de son homologue dans la BDTopo) ; deuxièmement, les objets ne sont pas appariés parce qu’il n’y a aucun objet homologue dans la BDTopo. Ce dernier cas concerne 70% des points de repère non appariés. Par exemple, un abri en ruine est étiqueté comme historique dans les données des zones protégées et n’existe plus dans les bases de données faisant autorité ou dans les autres sources, car l’objet a été détruit.

La figure 4 montre la distribution des points de repère non appariés pour les principales catégories. Cette ventilation montre bien la complémentarité dans

les différentes sources de données. Par exemple, C2C a plusieurs catégories spécifiques aux sports de montagne comme l’escalade, les cascades, points de repère moins présents dans la base de données d’autorité.

Une validation manuelle des liens 1:1 a été effectuée pour évaluer la précision. Cette étape est nécessaire pour l’analyse de la sémantique dans les différentes sources de données. La précision est estimée à 95% pour C2C et Refuges.info et 71% pour AP. La figure 5 illustre les liens d’appariement 1:1 dans les trois sources de données où le point de repère nommé Lac Gary présente un point de repère homologue.

Enfin, les cas incertains sont des cas complexes où aucune décision n’est prise par l’algorithme d’appariement de données. Ils sont corrigés manuellement. La figure 5, à droite, montre un exemple d’un tel cas. Le point de repère «Col de l’homme (Sanctuaire Notre Dame de la Salette)» dans C2C, dont la catégorie est type d’accès, a trois candidats potentiels dans la BDTopo : « Col de l’homme » (col), « Tête de l’homme » (sommet) et « Parking Col de l’homme » (parking). Dans ce cas, aucune décision n’est prise du fait que la catégorie type d’accès n’est pas un concept dans le domaine de l’ontologie et que les noms des candidats sont plutôt similaires.

Résultats d’appariement des itinéraires

Les résultats obtenus pour l’appariement de données des itinéraires sont présentés dans le tableau 2 et les liens 1:1 ont également une bonne précision (85%), précision obtenue manuellement.

	1:0	1:1	Incertains
C2C- Visorando	76	20	21
Geotrack- Visorando	92	59	14
C2C- Geotrack	77	14	6

Tableau 2 : Résultats d’appariement des itinéraires

Afin de construire un réseau d’itinéraires de randonnées, on considère que deux itinéraires sont homologues si leurs géométries sont relativement proches. Différents cas d’itinéraires appariés sont alors observés : 1/ lorsqu’il existe une variante de l’itinéraire et que le détour est court ; 2/ lorsque deux itinéraires sont très proches et ont la même direction (typiquement lorsqu’ils bordent une rivière (ce qui est satisfaisant dans notre contexte puisque la victime décrit l’itinéraire par « je marche le long d’une rivière ») (figure 6) ; 3/ lorsque les points de départ sont à quelques mètres les uns des autres. Dans ce cas, un itinéraire peut faire une boucle et l’autre itinéraire

un aller-retour. Précisons que dans cette étude, les tronçons manquants sur le réseau routier n'ont pas été gérés (e.g. hors sentier, chemin sur glacier, voie d'escalade, etc.). Ces tronçons ne sont pas pris en compte dans l'appariement de données.

Les itinéraires incertains font partie des trois cas mentionnés ci-dessus. Remarquons que le nombre d'itinéraires appariés est faible. En réalité de nombreux itinéraires n'ont pas le même point de départ ou la même destination, ils n'ont que quelques tronçons en commun.

Bilan

En conclusion, l'appariement des points de repère et des itinéraires donne de bons résultats. Le grand nombre d'objets non appariés prouve la richesse des données collaboratives issues du web. Dans les deux cas, les liens incertains représentent plutôt des objets non appariés, ce qui correspond à notre besoin. Concernant les points de repère, nous avons remarqué que l'appariement échoue lorsque les concepts n'existent pas dans l'ontologie utilisée ou lorsque le nom n'est pas rempli dans les données faisant autorité. Dans ce cas, le critère le plus pertinent est la distance. Dans l'appariement des itinéraires, l'effet du critère du nom de la randonnée est très difficile à analyser. Le critère de localisation joue un rôle important dans la décision finale, qu'il s'agisse d'itinéraires appariés ou non.

Couverture et redondance des données

La mesure de connectivité d'un graphe d'itinéraires est un indicateur intéressant pour observer leurs couvertures et leurs présences. En examinant séparément les différents réseaux des itinéraires pour les trois sources de données, on obtient (figure 7) la répartition spatiale des composants (tronçons) connectés (aires) par sources de données (couleurs) et labellisés par le nombre d'itinéraires : 58 routes connectées pour Visorando, 28 pour C2C et 28 pour AP. Cette répartition spatiale signifie que les sources de données couvrent une grande superficie. En calculant ce même indicateur, mais avec les graphes fusionnés deux par deux, le nombre de composants connectés reste stable (59, et 58, pour la fusion de Visorando et C2C, respectivement pour la fusion de Visorando et AP). Avec la fusion des sources, le nombre de tronçons augmente par composant connecté, mais la couverture ne varie pas. Les itinéraires des différentes sources ne se sont pas toujours connectés avec la fusion. Les différentes sources de données sont donc non redondantes. En ce qui concerne la redondance des points de repère entre les différentes sources de données, seuls 37 points de repère homologues sont présents dans toutes les

sources de données. Le tableau 3 montre la redondance des points de repère par rapport aux sources de données deux à deux. La diagonale représente le nombre total de points de repère pour chaque source. On peut remarquer que les sources C2C et BDTopo sont plus comparables en termes de données : 1 790 points de repère de C2C ont un point de repère homologue dans BDTopo.

	C2C	Refuge.info	AP	BDTopo
C2C	3,009	217	81	1,790
Refuge.info		478	41	352
AP			1102	127
BDTopo				21,097

Tableau 3 : Analyse de la redondance des points de repères homologues dans les sources de données

Analyse sémantique des liens de correspondance

Comme mentionné dans le premier paragraphe, un des principaux besoins est de pouvoir faire des requêtes concernant les points de repère, les itinéraires ou une combinaison des deux. Pour cela, il est nécessaire de définir des règles de correspondance sémantique entre les différents concepts des sources de données (e.g. un col dans le monde réel est apparié à un point de passage dans C2C et à un col dans Refuges.info). Pour identifier les règles de correspondance sémantique, nous avons produit des matrices de confusion en utilisant les objets homologues (liens 1:1). Faute de place dans cet article, elles ne sont pas toutes représentées ; seules les principales correspondances sémantiques sont décrites ci-dessous.

Une requête fréquente sur le web pour la recherche d'itinéraires concerne le niveau de difficulté de l'itinéraire. Dans la recherche de localisation d'une victime, les gens prennent en compte le niveau de difficulté en décrivant leur itinéraire comme « nous avons pris un chemin facile pour aller au refuge Mont Blanc », « nous sommes sur la partie difficile du parcours ». Prendre en compte les règles de correspondance trouvées pour les itinéraires homologues s'avère donc intéressant (tableau 4). Malgré le peu d'itinéraires trouvés, nous avons analysé la matrice de confusion entre les itinéraires Visorando et les itinéraires des zones protégées (59 itinéraires trouvés). Chaque source contient cinq niveaux, les noms des niveaux sont différents, mais les paramètres considérés sont similaires (e.g. effort physique, mental et différence d'altitude). Dans l'ensemble, les deux sources sont en désaccord. Par exemple, le niveau moyen est réparti sur les niveaux facile, moyen et

Visorando AP	Facile	Moyen	Difficile	Très difficile	Extrem. difficile
Très facile	2		1		
Facile	4	6			
Moyen	5	15	7		
Difficile	1	7	9	2	
Très difficile					

Tableau 4 : Matrice de confusion des niveaux de difficulté des itinéraires entre Visorando et les zones protégées

difficile dans les deux cas. Plus précisément, le nombre d'itinéraires de niveau facile dans Visorando prend toutes les valeurs des niveaux des itinéraires des zones protégées. Un parcours de niveau difficile est classé dans le niveau facile et vice-versa. Ce travail préliminaire devrait être examiné plus en détail.

Pour les points de repère, une matrice de confusion est calculée pour chaque lien 1:1. Concernant l'appariement des sources de données C2C-BDTopo, on peut remarquer que les sommets dans C2C sont souvent appariés avec des sommets (547 objets), des pics (259 objets), des crêtes (29 objets), des rochers (29 objets). Une autre règle pertinente concerne les points de repère ayant la catégorie type d'accès. Ces points de repère sont appariés à des points de stationnement (40) ou aux cols (33). Cela confirme la définition du concept type d'accès dans C2C (e.g. les points de repère avec la catégorie type d'accès représentent les objets du monde réel liés à une position donnée ou comme point de départ d'une position A à une position B). Deuxièmement, pour Refuges.info, on peut remarquer que les passages sont toujours appariés à des cols. D'autres appariements intéressants sommets/pics (16 points de repère) et les points d'eau/chutes d'eau (5) ont été trouvés. Un point d'eau dans Refuge.info représente une source d'eau mais aussi des objets tels que des ponts. Ces points de repère n'ont pas été appariés. Enfin, l'analyse de la matrice de confusion AP-BDTopo donne des résultats très surprenants. Par exemple, les points d'intérêt ayant une catégorie telle que « géologie », « histoire » sont associés à des canyons, des cols, des grottes et des abris. Cela montre la complémentarité des données des zones protégées par rapport aux autres sources de données. De plus, les points de vue sont souvent associés à des sommets et des cols. En ce qui concerne l'utilisation combinée des points de repère et des itinéraires, la figure 8 illustre comment les points de repère sont reliés aux routes.

La figure 9 représente un objet du monde réel présent dans toutes les sources de données (« Lac de la Muzelle »), et lié (dans les sources de données d'origine) à deux itinéraires homologues.

Conclusion

Dans cet article, nous avons étudié le potentiel des sources de données hétérogènes spatiales participatives et issues du web pour l'intégration de données pour le secours en montagne. Nous nous sommes concentrés sur les itinéraires et les points de repère. Pour atteindre cet objectif, nous avons proposé une méthode fondée sur les techniques d'appariement afin de comparer et d'évaluer ces sources.

Notre approche a été testée sur trois sources de données aussi bien pour les points de repère que pour les routes. Nos résultats ont mis en valeur la richesse et la complémentarité des différentes sources de données et montrent une bonne couverture de la zone d'étude. Suite à cette analyse, de nouvelles connaissances concernant la sémantique des données (classification des points de repère), la géométrie et le niveau de difficulté des routes sont apparues. L'intégration des données par le biais de sources de données identifiées n'est pas une tâche facile. L'étape suivante consiste à formaliser ces nouvelles connaissances de différentes sources au moyen de métadonnées.

Concernant les relations existant entre les itinéraires et les points de repère dans certaines sources de données, nous avons étudié la faisabilité de redéfinir cette relation a posteriori en utilisant les liens de correspondance. Les premiers résultats sont prometteurs, comme le montre la figure 8. Ainsi, grâce aux liens d'appariement, les points de repère d'une route (violet dans la figure) peuvent être transférés vers une autre route (orange dans la figure) et vice versa. La suite des travaux consiste à concevoir un nouveau schéma global afin de définir les relations entre les points de repère et les itinéraires depuis les différentes sources.

D'un point de vue technique, ce travail a suscité le développement d'un plug-In QGIS dont la fonctionnalité est de télécharger les données de cette étude depuis les API d'accès⁴. Ce travail a permis aussi de développer un algorithme de construction de routes à partir d'une liste de tronçons dans le plug-In MapMatcher⁵. De plus, comment optimiser le résultat de la requête pour les cas où le même objet du monde réel est représenté dans de nombreuses sources de données ? Dans cette situation, la question est de savoir quelle représentation de la donnée doit renvoyer le service Web : une représentation multiple ou une représentation floue pour les fonctions appariées semble être une solution intéressante. Plus généralement, comment le contexte spatial peut-il être pris en compte par le service web afin de rendre l'information la plus complète possible ?

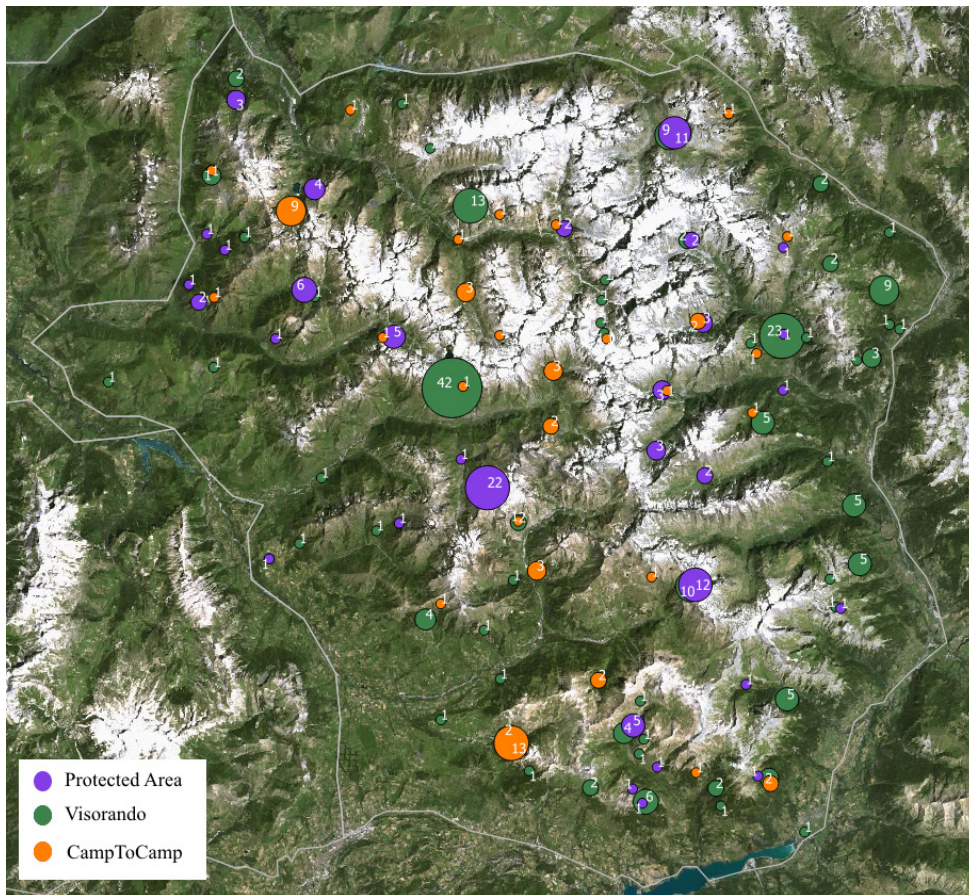


Figure 7 : Pour chaque source, répartition spatiale des tronçons connectés, le label indique le nombre d'itinéraires



Figure 8 : Liens d'appariement des routes et des points de repère

Pour continuer, comment identifier les points de repère dans les régions montagneuses, quand de nombreuses catégories de points de repère ne sont pas représentées dans les gazetiers faisant autorité. Ainsi, une recherche future consiste à identifier toutes les catégories des points de repère qui ne sont pas présentées dans les données faisant autorité et à élargir l'ontologie de domaine (Mustière *et al.*, 2011) pour les y inclure.

Les résultats de l'appariement des routes indiquent certaines faiblesses dans notre définition. Par exemple, deux routes très proches ne peuvent pas passer par un même repère de balisage ou peuvent avoir une dénivellation très différente. En améliorant certains critères d'appariement des données, comme la pondération en fonction de l'état du terrain (différence d'altitude) dans les critères spatiaux ou la pondération des noms de lieux identifiés dans le nom de l'itinéraire, la définition des itinéraires appariés peut être étendue à : *deux itinéraires sont équivalents quand il y a plusieurs points de repères identiques et les caractéristiques du terrain sont similaires entre deux points de repère.*

Dans ce travail, nous nous sommes concentrés sur une zone d'essai dans les Alpes françaises. Nous pensons que notre méthodologie peut être appliquée dans

d'autres pays de montagne. Cela dépend, bien sûr, de la disponibilité des données. En ce qui concerne les données collaboratives, certaines des données utilisées, telles que C2C, ont des caractéristiques hors de France (par exemple, Alpes italiennes et Suisse, Espagne, Allemagne, Népal). Il existe également d'autres initiatives existantes menées par un grand nombre et ayant une portée spatiale mondiale (Wikiloc) ou des répertoires toponymiques mondiaux (GeoNames, DBpedia). Pour les données faisant autorité, il existe des nomenclatures nationales dans la plupart des pays développés. Pour étendre la méthode à d'autres pays montagneux, les méthodes d'appariement pourraient être adaptées aux caractéristiques des données, généralement dans le choix des critères et des types de distances.

Enfin, dans le cadre de recherches futures, nous prévoyons de travailler pour concevoir un algorithme d'appariement multi-sources. En effet, l'algorithme multi-critères compare un ensemble de données à une référence ou au plus grand ensemble de données (Visorando pour les routes). Les cardinalités des liens appariés correspondent à 1:1 ou 1:n (par post-traitement). Pour n sources de données et sans ensemble de données de référence, nous devons exécuter l'algorithme $n(n-1)$ fois, ce qui n'est pas une solution optimale.

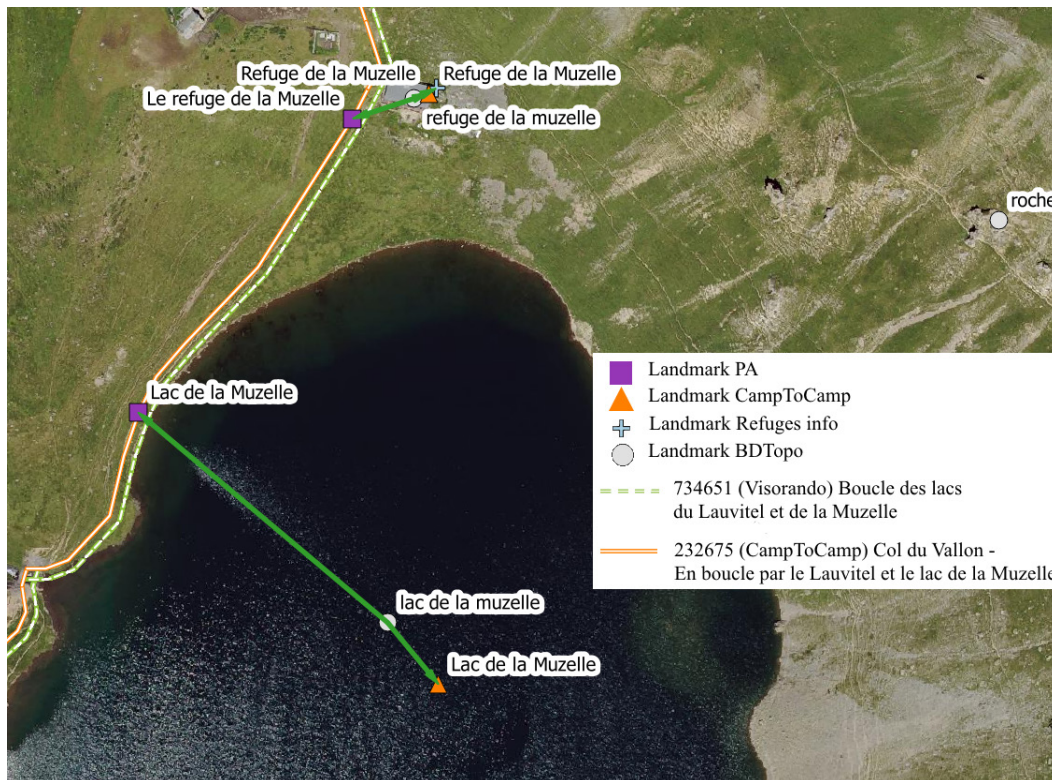


Figure 9 : Mêmes points de repère représentant un lac, lié à 2 itinéraires appariés

4 <https://github.com/ANRChoucas/PluginChoucas>

5 <https://github.com/IGNF/mapmatcher>

Bibliographie

- Acheson, E., De Sabbata, S. and S Purves, R. 2017. "A quantitative analysis of global gazetteers: Patterns of coverage for common feature types". *Computers Environment and Urban Systems* 64, pp. 309–320.
- Bégin, D., Devillers, R., Roche, S 2013. "Assessing volunteered geographic information (VGI) quality based on contributors' mapping behaviours". In *Proceedings of the 8th International Symposium on Spatial Data Quality*, Kowloon, Hong Kong, China; pp. 149–154.
- Brosset, D., Claramunt, C. and Saux, E. 2008. "Wayfinding in natural and urban environments: A comparative study". *Cartographica* 43, pp. 21–30.
- Denis, M. 1997. "The description of routes: A cognitive approach to the production of spatial discourse". *Cahiers de Psychologie Cognitive* 16, pp. 409–458.
- Derungs, C., S. Purves, R. 2013. "From text to land- scape: Locating, identifying and mapping the use of landscape features in a swiss alpine corpus". *International Journal of Geographical Information Science* 28, pp. 1272–1293.
- Duckham, M., Lingham, J., Mason, K , Worbosy, M. 2006, "Qualitative Reasoning About Consistency in Geographic Information", *Information Sciences—Informatics and Computer Science, Intelligent Systems, Applications: An International Journal* vol. 176 (6), pp. 601-627.
- Egenhofer, M. J., Mark, D. M. 1995. "Naive geography". Dans Frank A. U. et Kuhn W. (eds), *Spatial Information Theory A Theoretical Basis for GIS*, Berlin - Heidelberg, Springer, pp. 1–15.
- Ferrara, A., Nikolov, A. , Scharffe, F. 2013. "Data linking for the semantic web". *Semantic Web: Ontology and Knowledge Base Enabled Tools, Services, and Applications*, pp. 316-326
- Girres J.-F., Touya G. 2010. "Quality Assessment of the French OpenStreetMap Dataset", *Transactions in GIS* 14(4), pp. 435–459.
- Goodchild M., Li L. 2011, «Formalizing space and place ». Dans CIST2011 - Fonder les sciences du territoire, Paris, Collège international des sciences du territoire, pp. 177–183.
- Ivanovic, S, Olteanu-Raimond, A-M, Mustière S., Devogele T. 2016, "Detection of outliers in crowdsourced GPS traces", Spatial Accuracy 2016 Symposium, July 2016, Montpellier, France.
- Halilali, M., Gouardères, E. Devin, F., Gaio, M. 2018 « Plateforme logicielle pour l'intégration et la composition de services géospatiaux ». Dans Proceeding of SAGEO'2018 (Spatial Analysis and GEOmatics), November 6-9, 2018, Montpellier, France.
- Hill L., Zheng Q. 1999, "Indirect Geospatial Referencing through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers", in *Proceedings of the 62nd Annual Meeting of the American Society for Information Science*, Washington, DC, USA, pp. 57–69.
- Kasemsuppakorn P, Kasemsuppakorn A. Karimi, H. 2013. "A pedestrian network construction algorithm based on multiple GPS traces". *Transportation Research Part C: Emerging Technologies* 26, pp. 285–300.
- Kerouanton, C., Duparc, A., Jolivet, L., et al. 2017. « Outdoor recreationists, how to find their pauses by using ST-DBSCAN clustering algorithm". Dans treizièmes Rencontres de Théo Quant, Besançon, France.
- Lynch K. 1960. *The Image of the City*. Cambridge Mass., MIT Press.
- Moncla L., Renteria-Agualimpia W., Noguerras-Iso J., Gaio, M. 2014. « Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus". Dans ACM (éd.), *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2014)*, Dallas, Texas, United States.
- Mustière, S., Abadie, N., Aussenac-Gilles, N., et al. 2011. « Analyses linguistiques et techniques d'alignement pour créer et enrichir une ontologie topographique », *Revue Internationale de Géomatique*, 21(2), pp. 155–179.

- Neis P., Zielstra D. 2014. „Recent developments and future trends in volunteered geographic information research: The case of Openstreetmap”. *Future Internet* 6(1), pp. 76–106.
- Newson P., Krumm J. 2009, “Hidden markov map matching through noise and sparseness”. Dans *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09*, ACM, New York, NY, USA, pp. 336–343.
- Olteanu-Raimond A.-M., Mustière S., Ruas, A., 2015. « Knowledge formalization for vector data matching using belief theory”. *Journal of Spatial Information Science*, No 10.
- Pitman A., Zanker M., Gamper J., Andritsos, P. 2012. “Individualized hiking time estimation”. *23rd International Workshop on Database and Expert Systems Applications*.pp.101–105.
- Samal A., Seth S., Cueto, K. 2004. “A feature-based approach to conflation of geospatial sources”. *International Journal of Geographical Information Science*, vol. 18, n°5, p. 459-489
- Schockaert,S., Cock M. D., Kerre, E. E. 2008. “Location approximation for local search services using natural language hints”. *International Journal of Geographical Information Science* 22(3), pp. 315–336.
- Timpf S. 2005. *Cognitive Wayfinding Agents in Public Transportation Networks*, Geographic Information Science Center, University of Zurich, Suisse.
- Teulade-Denantes J., Maudet, A., Duchêne, C. 2015, “Routes visualization: Automated placement of multiple route symbols along a physical network infrastructure”, *Journal of Spatial Information Science*.
- Vias J., Ocaña Ocaña C., 2014. “Multicriteria evaluation by GIS to determine trail hiking suitability in a natural park”, *Boletín de la Asociación de Geoógrafos Españoles* pp. 323–339.
- Wu Z., Palmer, M. , 1994, “Verb Semantics and Lexical Selection”. Dans : *Proceedings of the 32nd Annual Meetings of the Associations for Computational Linguistics*, pp. 133-138.
- Zhou S., Winter S., Vasardani M., Zhou S., 2017. “Place descriptions by landmarks”. *Journal of Spatial Science* 62(1), pp. 47–67.