Suivi des ressources en eau par Transport Optimal et propagation d'incertitudes

Stage de master

Accroche 100 caractères

Propagation d'incertitudes ; Distance de Wasserstein ; Application au suivi des ressources en eau

Profil recherché

Étudiant·e en dernière année d'école d'ingénieur ou de Master, spécialité mathématiques appliquées, informatique.

— De bonnes connaissances en mathématiques appliquées sont indispensables,

— Des compétences en télédétection, en programmation (python) et la capacité à développer sous Linux seraient appréciées

— Aucune connaissance préalable en hydrologie n'est nécessaire.

Mots clés : mathématiques appliquées, optimisation, programmation, transport optimal, Wasserstein, data science, traitement du signal, traitement d'images, télédétection.

Contexte applicatif

La connaissance de l'évolution des volumes d'eau stockés dans les lacs et reservoirs est crucial pour de nombreux usages (recherche en géosciences, gestion des ressources en eau,. . .) et le manque de mesures de terrain rend les données satellites indispensables.

Pour cela, plusieurs types de données sont fréquemment utilisées :

— Images optiques (Sentinel-2, Landsat) et SAR (Sentinel-1), qui permettent de déterminer l'étendue en eau.

— Données d'altimètres RADAR qui mesurent l'élévation de la surface de l'eau : nadir 1D (Topex/Poseidon, Sentinel-3 et 6) et de fauchée 2D (SWOT, Sentinel-3 NG Topo).

— Des informations topographiques (3D) du sol ; dérivés d'images optiques stereo, de mesures LiDAR, voire de nuages de point SWOT. Ils permettent d'améliorer la connaissance de la relation (loi ZSV) entre la hauteur, l'étendue et le volume d'eau.
— Les données in situ, lorsqu'elles sont disponibles.

Le CNES mène une activité de R&D sur la construction d'indicateurs avancés du remplissage (variation du volume d'eau) des lacs en combinant les données disponibles. Un enjeu central pour le problème est une connaissance fine de la bathymétrie sèche (topographie des zones situées entre les niveaux d'eau les plus bas et les niveau les plus hauts). Les données mentionnées y contribuent directement (mesures 3D optiques stéréo, LiDAR et SAR interférométriques (SWOT) ou indirectement (relation entre la hauteur d'eau et l'étendue d'eau).

Contexte mathématique et problématique

Dans le cadre du stage, l'objectif est d'étudier l'estimation de la bathymétrie et du remplissage des lacs par hybridation de données acquises à des dates et par des capteurs différents au travers des formalismes du transport optimal [3,5].

Soient X, Y deux espaces métriques compacts. On définit $c \in C(X, Y)$ la fonction de coût et $(\mu,\, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$ les marginales. On rappelle le problème de Kantorovich

$$(KP) := \inf \left\{ \int_{X \times Y} c(x,y) \,\mathrm{d}\gamma(x,y) | \gamma \in \Pi(\mu,\, \nu) \right\},$$

qui correspond à un problème d'optmisation sur les plans de transport, et le problème dual

$$(DP) := sup \left\{ \int_X \phi \,\mathrm{d}\mu + \int_Y \psi \,\mathrm{d}\nu | \phi \in \mathcal{C}_b(Y),\, \phi \bigoplus \psi \leq c \right\},$$

pour lesquels des maximiseurs/minimiseurs existent. La formulation de ces problèmes permet de comparer des mesures dans des espaces de probabilités de manière naturelle.

Sur l'espace de probabilités $\mathcal{P}(X)$, on peut définir la p-distance de Wasserstein, par

$$W_p(\mu, \nu) := \left(\min_{\gamma \in \Pi(\mu,\, \nu)} \mathrm{dist}(x,y)^p \,\mathrm{d}\gamma(x,y) \right)^{1/p}.$$

Calculer un barycentre de Wasserstein revient à résoudre le problème de minimisation précédent selon cette distance de Wasserstein.

On construira au cours du stage un barycentre de Wasserstein entre un ensemble de représentations partielles afin de reconstruire la forme et la bathymétrie des lacs, en utilisant ainsi des objets en dimension 2D+T ou 3D+T.

Cette construction pourra se baser sur une régularisation entropique du problème de transport optimal [1, 2, 4].

Le stage pourra s'appuyer sur les résultats obtenus pour la construction de représentations continues implicites à partir de séries temporelles de données topographiques SWOT.

De plus, une attention particulière sera apporté à la prise en compte des incertitudes et à leur propagation. Par "incertitudes" nous entendons trois types d'incertitudes distincts :

1. Incertitude inhérente à la mesure, liée au bruit dans les données
2. Incertitude liée au manque de données
3. Incertitude inhérente au modèle

L'objectif de la modélisation de ces trois types d'incertitudes est de pouvoir les propager jusqu'à l'estimation de la bathymétrie finale. Autrement dit, l'incertitude du produit en sortie doit tenir compte du bruit de mesure, du fait que la donnée peut être incomplète, ou que le modèle est peut être inadapté pour modéliser certains type de phénomènes.

Contacts et déroulement du stage

Le stage s'effectuera et sera financé par le CNES, sur le site de Toulouse, pour une durée de 4 à 6 mois, début à partir de mars 2025. Pour candidater, merci de se rendre sur l'annonce de l'offre via la plateforme de recrutement.
Contacts : Dawa Derksen, Nicolas Gasnier, Jessie Levillain.

Références

[1] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space", SIAM Journal on Mathematical Analysis, 43(2) :904–924, 2011.

[2] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, "Iterative Bregman Projections for Regularized Transportation Problems", SIAM Journal on Scientific Computing 2015 37 :2, A1111-A1138.

[3] F. Santambrogio, Optimal transport for applied mathematicians, Springer, 2015.

[4] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas, "Convolutional wasserstein distances : efficient optimal transportation on geometric domains", ACM Trans. Graph. 34, 4, Article 66, 2015.

[5] C. Villani, Optimal transport : old and new, vol. 338, Springer Science Business Media, 2008.