

Transport optimal et rectangularisation de série temporelle

Abanish BASKARADEVAN Félix HUSSON

Encadrant : LACAILLE Jérôme

May 22, 2024



Sommaire

- ▶ Définition du problème de rectangularisation de série temporelle
- ▶ Présentation du transport optimal et du DTW
- ▶ Présentation de la méthode par échantillonnage

Intérêts de la rectangularisation

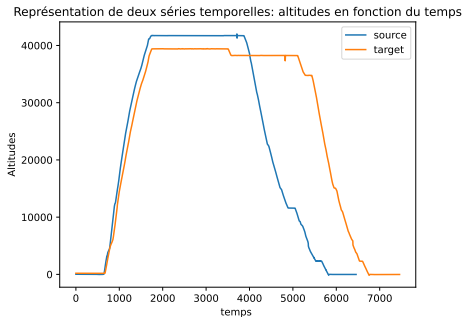


Figure 1: Représentation d'altitudes au cours du temps

Définition de la rectangularisation

Definition 1

Soit $t_1 \rightarrow f(t_1)$ et $t_2 \rightarrow g(t_2)$ deux séries temporelles.

On cherche deux fonctions de dilatation ϕ_1, ϕ_2 telles que $t'_1 = \phi_1(t_1)$ et $t'_2 = \phi_2(t_2)$ pour ainsi écrire $f(t'_1)$ et $g(t'_2)$ qui sont les séries temporelles rectangularisées.

Présentation du transport optimal

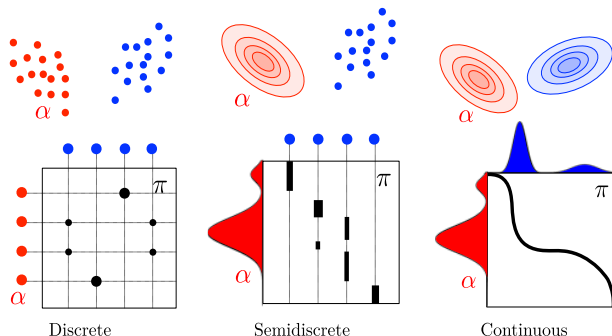


Figure 2: Représentation de fonction de couplage

1

¹[Peyré and Cuturi, 2020]

Description du transport optimal

μ_k : distributions discrètes (k=0 la source et k=1 la cible)

Π : Matrice de transport

C : Matrice de coût

$$\forall k = 0, 1, \quad \mu_k = \sum_{i=1}^{n_k} p_{k,i} \delta_{x_{k,i}}$$

$$\mathcal{P} = \left\{ (\Pi_{i,j})_{i,j} \in (\mathbb{R}^+)^{n_0 \times n_1} : \forall i, \sum_j \Pi_{i,j} = p_{0,i}, \forall j, \sum_i \Pi_{i,j} = p_{1,j} \right\}$$

$$\Pi^* \in \arg \min_{\Pi \in \mathcal{P}} \sum_{i,j} \Pi_{i,j} C_{i,j}$$

$$C_{i,j} = \|\mathbf{x}_{0,i} - \mathbf{x}_{1,j}\|^2$$

$$W_2(\mu_0, \mu_1)^2 = \sum_{i,j} \Pi_{i,j}^* C_{i,j}.$$

Problème temporel du transport optimal

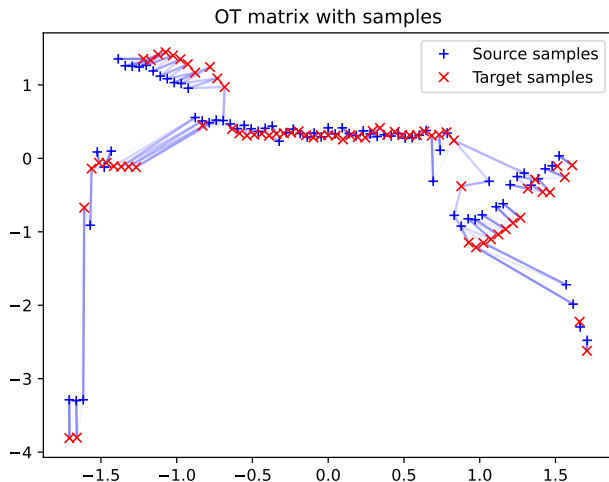


Figure 3: Appariement du transport optimal de deux EGT: cas normalisation en temps

DTW

Le DTW (Dynamic Time Warping) est un algorithme qui permet de comparer deux séries temporelles.

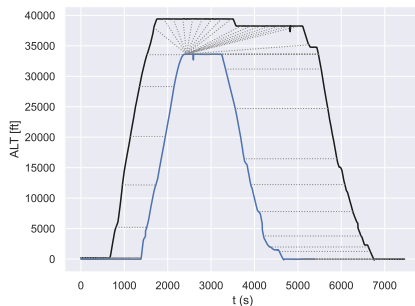


Figure 4: Visualisation de l'algorithme de DTW

Chemin du DTW

La distance usuelle utilisée dans les bibliothèques utilisant DTW est la suivante :

$$D(i, j) = \begin{cases} |x(i) - y(j)|, & i = j = 1 \\ |x(i) - y(j)| + \min(D(i-1, j), D(i-1, j-1), D(i, j-1)), & \text{sinon} \end{cases}$$

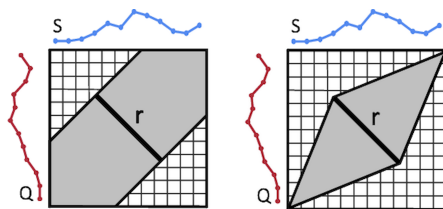


Figure 5: Différentes méthodes de calcul ²

²Dynamic Time Warping: Itakura vs Sakoe-Chiba

Chemin du DTW

Le chemin DTW permet de voir les correspondances entre deux séries temporelles. Nous avons effectué le calcul de ce chemin en utilisant deux méthodes distinctes.

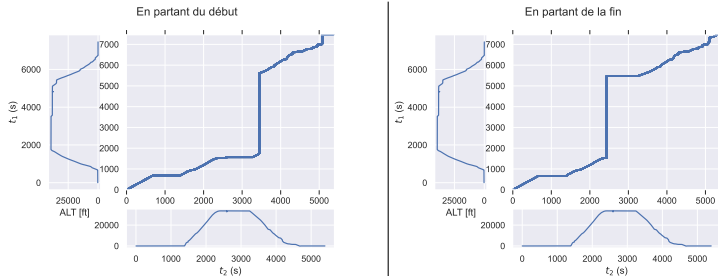


Figure 6: Calcul du chemin selon le départ initial du chemin

Utilisation du chemin

Le calcul du DTW permet de trouver les correspondances entre les deux séries. À l'aide de ce chemin, nous pouvons établir deux nouvelles séries temporelles, toutes deux de même dimension.

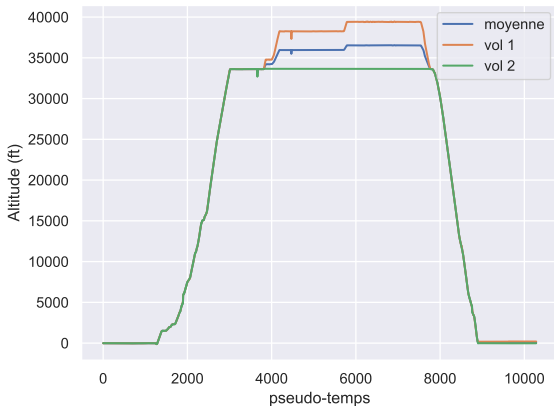


Figure 7: Exemple de calcul statistique d'une moyenne de 2 vols de durées différentes

Échantillonnage

Idée : Échantillonner le chemin

- ▶ Le DTW nous donne un chemin entre deux séries temporelles
- ▶ Échantillonnage du chemin à la dimension de la série 'cible'
- ▶ Transformation de la série 'source' vers la série 'cible' via ce chemin échantillonné

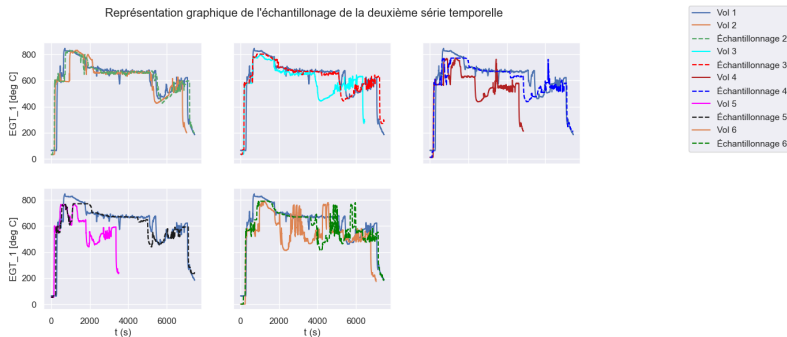


Figure 8: Échantillonnage de plusieurs vols vers un vol 'cible'

Différente catégorie d'appareillement possible

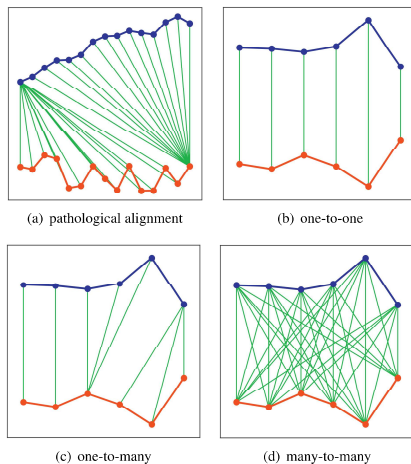


Figure 9: Catégorie d'appareillement possible

Conclusion

- ▶ Transport optimal et clustering
- ▶ Une résolution déterministe du problème?
- ▶ Utilisation d'autre distribution ou autre chemin

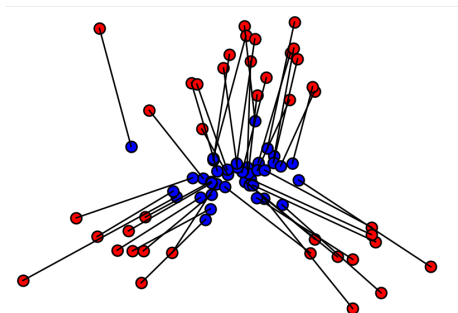


Figure 10: Représentation d'un transport entre la mesure rouge et la bleue

Représentation de la matrice de coût cas 2D

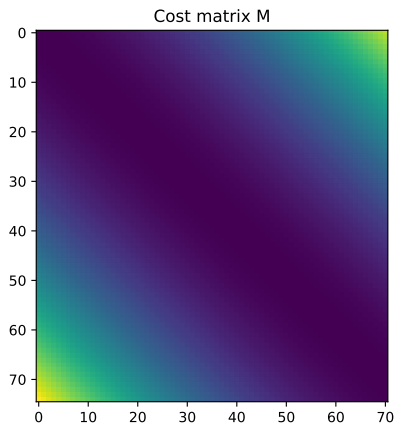


Figure 11: Matrice de cout du transport optimal sur des séries temporelles

Représentation de la matrice de transport de Sinkhorn cas 2D

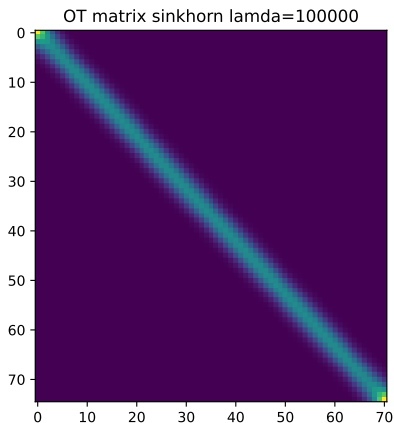


Figure 12: Matrice de transport de sinkhorn

Transport optimal classique sur des séries non normalisé en temps

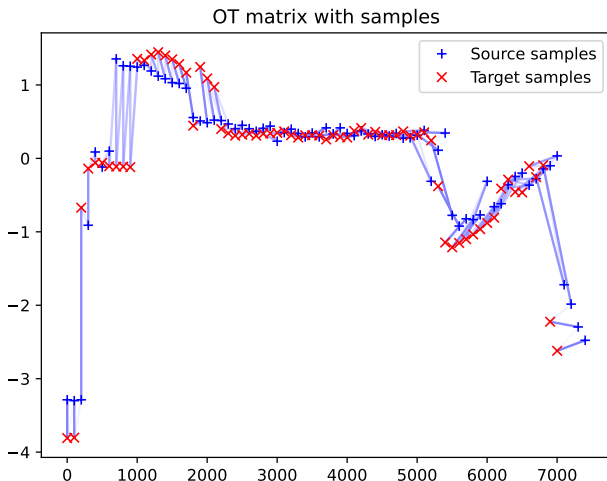


Figure 13: Appariement du transport optimal de deux EGT sans la normalisation en temps

Références I



Jupyter Notebook Viewer.



Al-Naymat, G. et al. (2012).

Sparsedtw: A novel approach to speed up dynamic time warping.

arXiv.



Desai, A. et al. (2021).

Timevae: A variational auto-encoder for multivariate time series generation.

arXiv.



François-Petitjean (2014).

Db.

Consulté le 14 avril 2024.



Morel, M. et al. (2018).

Time-series averaging using constrained dynamic time warping with tolerance.

Pattern Recognition, 74:77–89.



Peyré, G. and Cuturi, M. (2020).

Computational optimal transport.

arXiv.

Références II



Zhang, Z. et al. (2020).

Time adaptive optimal transport: A framework of time series similarity measure.

IEEE Access, 8:149764–74.