Downscaling météorologique : Une approche statistique pour améliorer la résolution spatiale

Culture Sciences de l'Ingénieur

Matys SAVIDAN

Publié en :
Octobre 2025

école normale supérieure paris–saclay

Matys Savidan est élève de l'ENS Paris-Saclay, au Département d'Enseignement et de Recherche Sciences de l'Ingénierie Electrique et Numérique (DER SIEN). Cet article est issu d'un projet de recherche (Travaux de Recherche Encadré) mené durant sa première année de Master. Ce travail a été réalisé sous la supervision du Professeur Mohammed Nabil El Korso du laboratoire L2S (CentraleSupélec, Université Paris-Saclay).

L'objectif est de présenter, de manière claire et accessible, la méthodologie et les résultats d'un projet portant sur le *statistical downscaling* en météorologie, une technique visant à améliorer la résolution spatiale des données météorologiques à l'aide d'outils mathématiques et d'apprentissage automatique. Tout au long de cet article, l'ensemble des champs météorologiques sont extraits du Copernicus Climate Data Store via son API publique.

1 Pourquoi le downscaling ?

Les modèles météorologiques, telles que les simulations climatiques, fournissent généralement des champs atmosphériques à des résolutions spatiales grossières, de l'ordre de 50 à 100 kilomètres. Bien que suffisante pour décrire les dynamiques à grande échelle, cette résolution ne permet pas de rendre compte des variations locales essentielles à de nombreuses applications : agriculture, hydrologie, prévision météorologique urbaine, ou encore analyse des événements extrêmes.

Le *downscaling* (littéralement réduction d'échelle) désigne le processus consistant à affiner ces champs grossiers afin d'obtenir une version à plus haute résolution, cohérente à la fois sur le plan statistique et spatial. Deux grandes familles d'approches sont couramment utilisées :

- **Downscaling dynamique**, qui repose sur des simulations imbriquées à haute résolution utilisant des modèles climatiques régionaux. Cette méthode est physiquement détaillée mais exigeante en calcul.
- **Downscaling statistique**, qui vise à établir une correspondance entre les données à basse résolution et les champs à haute résolution à partir d'observations historiques. Ces méthodes sont légères en calcul et adaptables, mais nécessitent des données d'entraînement robustes et une cohérence statistique.

Dans ce travail, nous nous concentrons sur les approches statistiques. Nous commençons par comparer les techniques classiques d'interpolation et en évaluons les limites. Nous introduisons ensuite un mécanisme de correction fondé sur la théorie du transport optimal, conçu pour préserver la structure spatiale ainsi que la conservation de masse. Cette correction sert ensuite de base à l'entraînement de modèles capables de l'inférer directement à partir de descripteurs statistiques des champs interpolés.

2 Méthodologie et stratégie

L'étude débute par l'évaluation de méthodes classiques d'interpolation, incluant les interpolations bilinéaire, bicubique, les fonctions à base radiale (Radial Basic Functions RBF) et la pondération par l'inverse de la distance (Inverse Distance Weighting IDW), appliquées à des données haute résolution, préalablement dégradées pour simuler des champs à basse résolution. Ces méthodes servent de référence et sont évaluées à l'aide de métriques standards telles que l'erreur absolue moyenne (Mean Absolute Error MAE) et la comparaison des énergies spectrales. Bien que simples à mettre en œuvre, ces méthodes ne parviennent pas à restituer les structures spatiales fines ni les extrêmes localisées.

Pour surmonter ces limites, le cœur de ce travail se concentre sur le transport optimal régularisé (divergence de Sinkhorn), utilisé comme étape de correction après interpolation. Chaque champ à basse résolution est d'abord

interpolé, puis corrigé par transport optimal. Comme les champs interpolés et cibles présentent des valeurs positives, il n'est donc pas nécessaire d'imposer de contrainte supplémentaire de positivité lors du processus de normalisation. Le coût de transport est calculé sur la base de distances euclidiennes normalisées, et le plan de correction est optimisé par régularisation entropique afin de garantir stabilité numérique et régularité.

Deux stratégies sont ensuite explorées pour généraliser cette correction : (1) la constitution d'une base de données de plans de transport pré-calculés, associés à des configurations météorologiques spécifiques, et (2) l'utilisation de descripteurs statistiques globaux (moyenne, variance, asymétrie, kurtosis, énergies fréquentielles) pour identifier des cas similaires ou prédire directement les corrections. Ces descripteurs servent à l'entraînement de réseaux de neurones légers — perceptrons multicouches (MLP) et réseaux récurrents (GRU) — capables d'inférer des séquences de correction à partir de vecteurs de caractéristiques.

3 Méthodes classiques d'interpolation et leurs limites

Comme premier point de comparaison, nous appliquons des techniques d'interpolation standards pour reconstruire des champs de température haute résolution à partir de leurs homologues basse résolution. Ces méthodes sont déterministes, locales et largement utilisées en raison de leur simplicité et de leur efficacité numérique.

Méthodes implémentées

Nous considérons quatre approches courantes :

- **Interpolation bilinéaire** : calcule une moyenne pondérée à partir des quatre points de grille voisins. Elle est rapide et facile à implémenter, mais tend à lisser les transitions et à effacer les gradients marqués.
- **Interpolation bicubique** : utilise un voisinage de 4×4 points et un ajustement polynomial. Elle produit des champs plus réguliers que l'interpolation bilinéaire, mais peut introduire des artefacts à proximité des discontinuités.
- **Pondération par l'inverse de la distance (IDW)** : repose sur l'hypothèse que les points proches ont plus d'influence que les points éloignés. Elle est simple, indépendante de la grille, mais se comporte mal dans les zones où la densité des points est inhomogène ou dans des contextes de relief complexe.
- **Interpolation par fonctions de base radiales (RBF)** : méthode sans maillage, fondée sur des noyaux lisses (souvent des splines fines). Elle donne des résultats visuellement satisfaisants, mais au prix d'un coût de calcul plus élevé.

Évaluation

Chaque méthode est testée sur des données horaires de température sous-échantillonnées, issues du *Copernicus Climate Data Store*. Les résultats sont comparés aux champs haute résolution originaux, principalement à l'aide de l'erreur absolue moyenne (MAE).

Limites

Si ces méthodes préservent les structures à grande échelle, elles échouent à restituer les phénomènes fins et les gradients de température localisés, en particulier dans les régions fortement influencées par la topographie. Ces constats mettent en évidence la nécessité d'un mécanisme de correction plus cohérent, capable de prendre en compte la distribution spatiale des valeurs et pas uniquement leur continuité locale.

Résultats

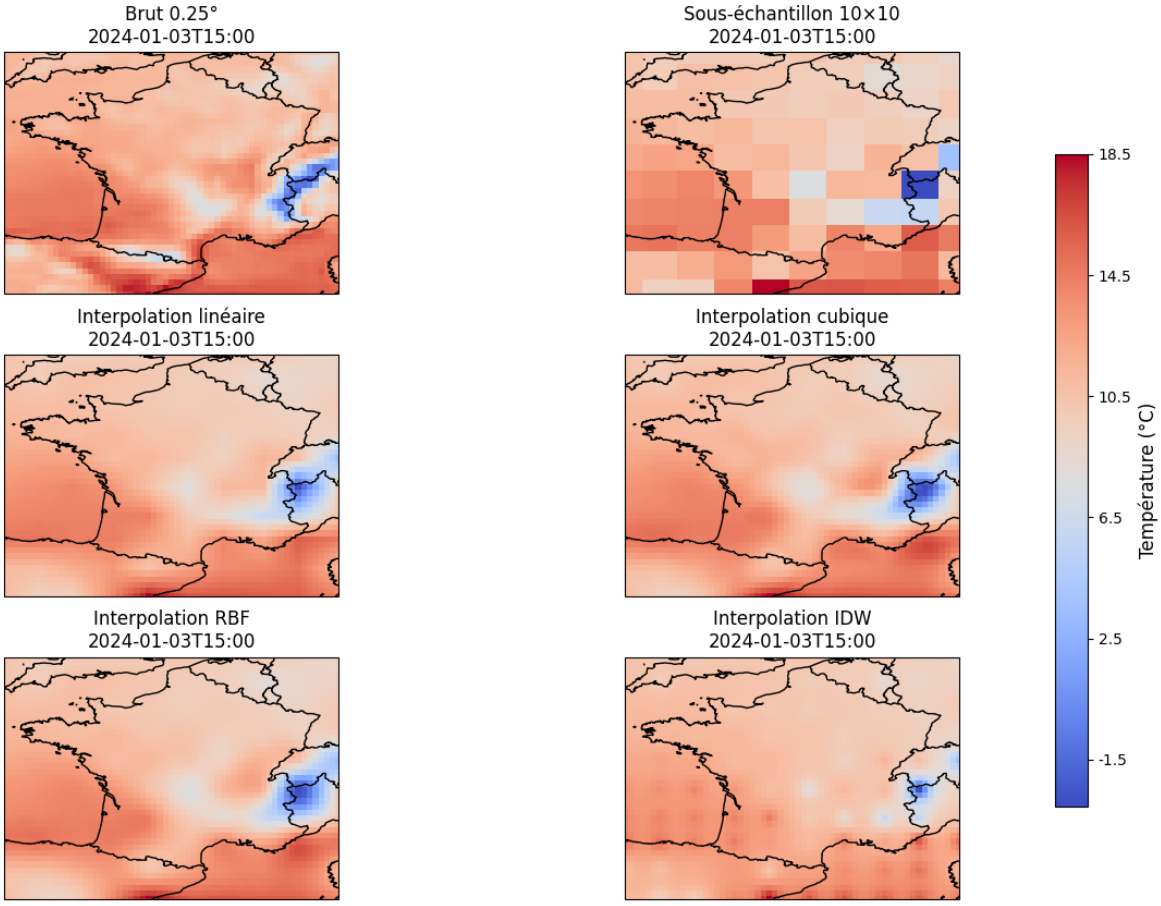

Figure 1 : Interpolation bidimensionnelle classique

Les résultats, résumés dans la Table 1, montrent que bien que les interpolations cubique et RBF obtiennent des performances légèrement meilleures, aucune des méthodes classiques ne parvient à reproduire la variabilité fine présente dans les champs originaux.

Méthode	**MAE (°C)**
Linéaire	0.872
Cubique	0.871
RBF	0.845
IDW	1.320

Table 1 : Erreur absolue moyenne (MAE) sur 168 pas de temps pour différentes méthodes d'interpolation.

4 Transport optimal : un cadre rigoureux

Les méthodes d'interpolation classiques opèrent localement, en s'appuyant sur des hypothèses de régularité pour estimer les valeurs manquantes ce qui induit par la force des choses un gros biais. Ces hypothèses sont souvent violées dans les champs météorologiques, notamment à proximité des reliefs ou des fronts atmosphériques, où les gradients spatiaux sont abrupts. Dans ces contextes, l'utilisation d'un mécanisme de correction global et sensible à la structure devient nécessaire.

Le transport optimal (OT) fournit un cadre mathématique permettant de comparer et de transformer des champs scalaires en calculant la manière la plus efficace de redistribuer une distribution vers une autre. Formulé à l'origine par Monge au XVIII^{e} siècle, puis généralisé par Kantorovich, l'OT a connu un regain d'intérêt grâce aux méthodes numériques modernes et à ses applications en traitement du signal, en apprentissage automatique et en physique.

Éclairage mathématique : qu'est-ce que le transport optimal ?

Problème. Étant données deux distributions a et b de même masse sur un domaine discret, le transport optimal cherche la manière la plus efficace de déplacer la masse de a vers b, en minimisant un coût global.

Monge et Kantorovich. La formulation initiale (Monge, 1781) imposait l'existence d'une application de transport T telle que $b = T_{\#}a$, minimisant le coût total. Kantorovich a proposé en 1942 une relaxation en introduisant des plans de transport $P \in \mathbb{R}_{+}^{n\times n}$, où P_{ij} représente la masse envoyée de x_i vers x_j, sous les contraintes :

$$U(a,b) = \{\, P \in \mathbb{R}_{+}^{n\times n} \;\mid\; P\mathbf{1} = a,\; P^{\top}\mathbf{1} = b \,\}.$$

Fonction de coût. Un choix classique est la distance euclidienne au carré : $C_{ij} = \|x_i - x_j\|^2$. Le plan optimal P^* minimise le coût total :

$$\min_{P\in U(a,b)} \langle C, P\rangle_F = \min_{P\in U(a,b)} \sum_{i,j} C_{ij}P_{ij}.$$

Formulation régularisée. Pour résoudre efficacement ce problème, on ajoute une régularisation entropique :

$$\min_{P\in U(a,b)} \langle C, P\rangle_F + \varepsilon \sum_{i,j} P_{ij}(\log P_{ij} - 1).$$

Cette régularisation lisse la solution et permet d'utiliser des algorithmes rapides et itératifs, comme la méthode de Sinkhorn.

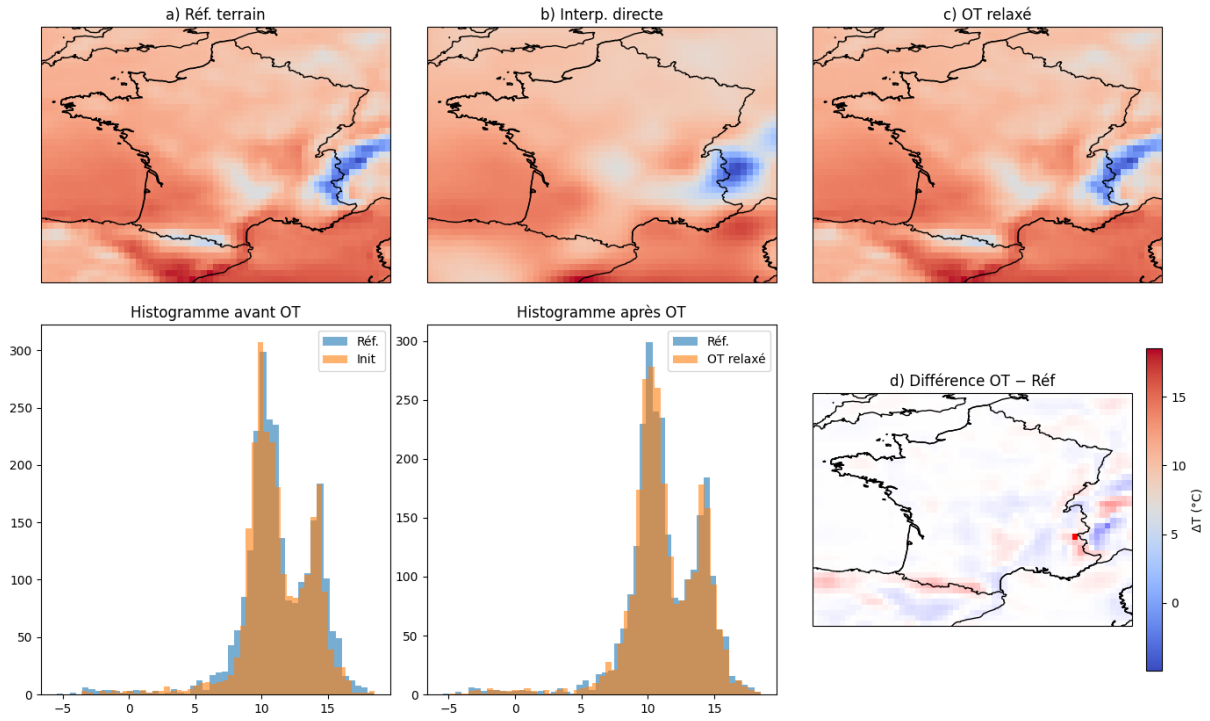

Figure 2 : Résultat obtenu par transport optimal après interpolation cubique

Dans notre cadre, les champs interpolés et les champs cibles sont d'abord normalisés de manière à ce que la somme de leurs valeurs soit égale à un. Cette étape permet de les interpréter comme des distributions de probabilité discrètes sur la grille spatiale, et donc d'appliquer la théorie du transport optimal dans sa forme probabiliste classique. Plutôt que de calculer une correction unique et directe, nous adoptons une approche itérative. À chaque itération, un plan de transport optimal régularisé est calculé entre le champ corrigé courant et la référence haute résolution. Le champ est ensuite mis à jour comme une combinaison convexe de son état précédent et de la version transportée. Ce processus itératif permet un raffinement progressif, garantissant stabilité et convergence vers une configuration spatiale plus proche de la cible.

La mise à jour s'écrit :

$$T^{(k+1)} = (1-\tau)\, T^{(k)} + \tau \cdot \mathrm{Proj}_{G_k}\big(T_{\mathrm{ref}}\big),$$

où τ est un paramètre de relaxation et G_k le plan de transport à l'itération k.

En minimisant pas à pas le coût global de transport, cette approche conserve la structure sous-jacente du champ interpolé tout en l'alignant progressivement sur la référence. La régularisation assure à la fois la régularité des solutions et l'efficacité numérique de la procédure. La Figure 2 illustre l'effet de la correction par transport optimal appliquée à un champ interpolé. Contrairement à l'interpolation cubique seule, qui tend à lisser les contrastes et à effacer les gradients marqués, l'OT redistribue globalement la masse de manière optimale. Cette approche conserve ainsi les structures spatiales fines (reliefs, fronts) et restaure une cohérence locale proche de la référence haute résolution. On voit que l'OT ne se limite pas à un ajustement ponctuel, mais agit comme une correction globale de la géométrie du champ.

5 De la correction par transport à l'apprentissage statistique

La section précédente a introduit le transport optimal comme méthode de correction rigoureuse pour améliorer le réalisme spatial des champs météorologiques interpolés. Dans ce cadre, la correction est calculée en résolvant un problème de transport entre le champ interpolé normalisé et la référence haute résolution.

Cette approche fournit une base théorique et empirique solide, mais elle n'est pas directement exploitable en pratique : elle nécessite l'accès au champ haute résolution que l'on cherche précisément à reconstruire. Autrement dit, il faut déjà connaître la solution pour appliquer la correction, ce qui n'est évidemment pas possible en prévision opérationnelle. En revanche, cette méthode constitue un cadre fiable pour générer des corrections de référence sur des données historiques, et ainsi entraîner un modèle capable de généraliser le processus de correction.

Cette section présente la stratégie adoptée pour passer d'une correction supervisée, fondée sur la référence, à un modèle appris qui prédit les itérations du processus de transport optimal à partir de la structure du champ interpolé.

Apprentissage à partir de corrections pré-calculées

Nous commençons par construire une base d'apprentissage. Pour chaque pas de temps, nous calculons le champ interpolé à partir de sa version basse résolution, puis la correction associée via un transport optimal régularisé (Sinkhorn). Ces plans de transport servent alors de cibles supervisées.

Afin de généraliser, nous faisons l'hypothèse que la correction nécessaire pour un nouveau champ interpolé dépend de sa structure statistique globale. Nous extrayons donc, pour chaque champ, un vecteur de caractéristiques de longueur fixe comprenant notamment :

- Moments d'ordre 1 et 2 : moyenne et écart-type
- Descripteurs d'ordre supérieur : asymétrie (skewness) et aplatissement (kurtosis)
- Indicateurs spectraux : proportions d'énergie dans les basses, moyennes et hautes fréquences (via transformée de Fourier 2D)

Ces descripteurs offrent une synthèse compacte de la forme et de la variabilité du champ.

Architectures de prédiction

Nous testons deux types de réseaux de neurones pour prédire directement la correction à partir des descripteurs statistiques :

- Un **Perceptron multicouche (MLP)**, qui prédit en une seule passe l'intégralité du vecteur de correction.

- Un réseau de type **Gated Recurrent Unit (GRU)**, qui met à jour le champ de manière itérative, en apprenant à le corriger étape par étape à partir de ses états précédents.

Les deux modèles sont entraînés sur les couples (vecteur de caractéristiques, correction) issus de la base de données pré-calculée. Une fois entraînés, ils peuvent être appliqués directement à de nouveaux champs interpolés, sans nécessiter l'accès au champ de référence haute résolution.

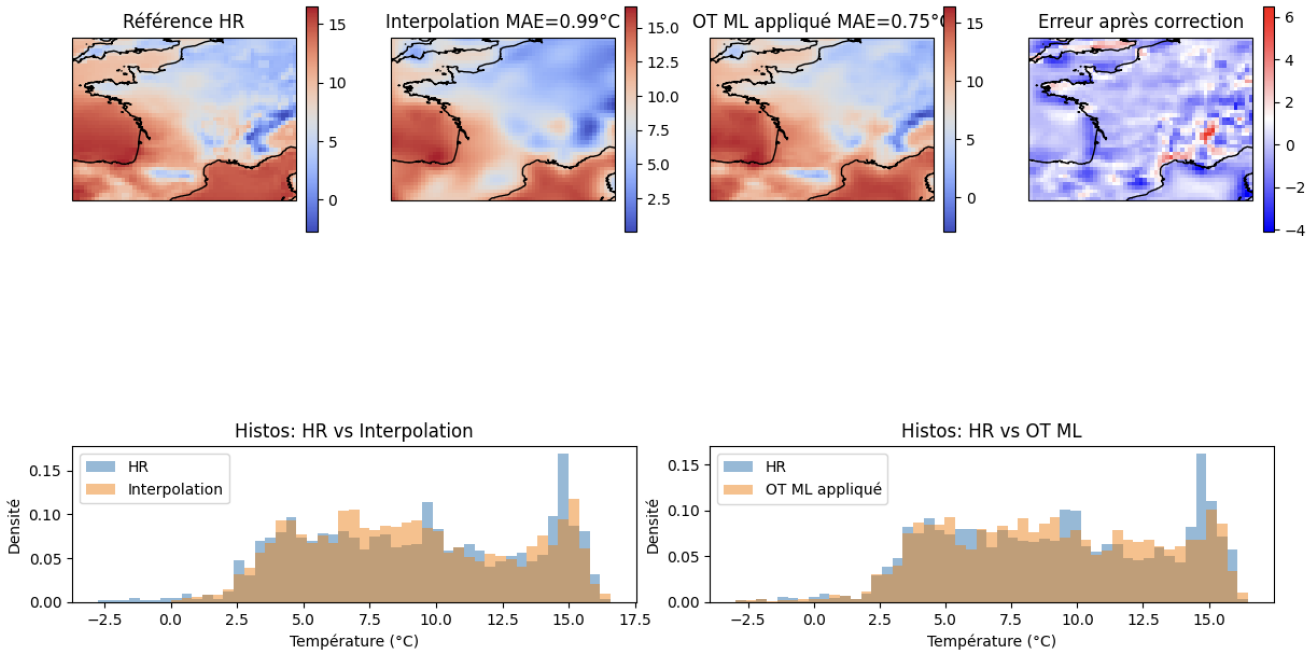

Figure 3 : Correction estimée par OT avec un MLP (performances comparables à celles du GRU)

La Figure 3 montre qu'un modèle statistique (ici un perceptron multicouche) est capable d'apprendre à reproduire les corrections issues du transport optimal, uniquement à partir de descripteurs statistiques globaux. Même sans accès direct au champ haute résolution, le modèle restitue des structures fines et réduit significativement l'erreur par rapport à l'interpolation de départ. Cela illustre que l'apprentissage automatique peut jouer le rôle d'approximation rapide du transport optimal, ouvrant la voie à des applications opérationnelles en temps réel.

Résultats et avantages

Même avec un jeu de données limité, les modèles appris reproduisent des corrections pertinentes et réduisent significativement la MAE par rapport aux interpolations de référence (en moyenne 0,74 °C sur le lot). Le modèle GRU, en particulier, se montre plus performant dans les zones à forte complexité spatiale, car il ajuste dynamiquement le champ au fil des itérations. Cette stratégie fondée sur l'apprentissage permet ainsi d'intégrer les avantages du transport optimal dans un substitut rapide et peu coûteux, ouvrant la voie à des applications en temps réel.

Il convient toutefois de noter que l'ensemble d'entraînement utilisé dans cette étude est resté relativement restreint, limité à seulement trois mois de données météorologiques, dans des conditions saisonnières proches. Cette contrainte soulève naturellement des questions sur la capacité de généralisation des modèles appris, notamment lorsqu'ils seraient appliqués à des situations plus diverses ou extrêmes. Néanmoins, avec davantage de ressources de calcul, la constitution d'un jeu d'entraînement beaucoup plus vaste et diversifié sur le plan saisonnier serait tout à fait réalisable, et permettrait probablement de surmonter ces limites.

6 Conclusion et perspectives

Cette étude a présenté un cadre statistique pour le *downscaling* météorologique, visant à reconstruire des champs de température haute résolution à partir d'entrées grossières. Nous avons commencé par évaluer les méthodes

classiques d'interpolation qui, malgré leur simplicité et leur efficacité numérique, échouent à restituer la variabilité fine et la cohérence spatiale des champs.

Pour dépasser ces limites, nous avons introduit le transport optimal régularisé comme mécanisme de correction. Bien qu'il dépende de la disponibilité de données de référence haute résolution, il constitue un repère théorique robuste pour évaluer et quantifier les ajustements spatiaux nécessaires.

Dans une perspective de généralisation, nous avons construit une base de données de plans de transport précalculés et entraîné des modèles de réseaux de neurones légers (MLP et GRU) capables de prédire les corrections à partir de descripteurs statistiques globaux. Ces modèles apportent des améliorations notables, en particulier dans les régions à forte hétérogénéité spatiale, malgré un apprentissage effectué sur un jeu de données relativement restreint et limité à une seule saison.

Parmi les perspectives, citons l'extension aux champs multivariés, l'utilisation de bases d'apprentissage plus vastes et plus diversifiées, ainsi que la mise en œuvre de schémas d'apprentissage spatio-temporels conjoints. Ce travail de preuve de concept illustre que des approches hybrides, guidées par la théorie, peuvent offrir des solutions évolutives à l'un des problèmes inverses centraux de la météorologie.

Références

[1] Douglas Maraun, Martin Widmann, and et al. Statistical and dynamical downscaling of precipitation : An evaluation and comparison of downscaling methods in Europe. *Hydrology and Earth System Sciences*, 23 :773–783, 2019.

[2] Bastien François, Mathieu Vrac, Alex J. Cannon, Yoann Robin, and Denis Allard. Multivariate bias corrections of climate simulations : which benefits for which losses ? *Earth System Dynamics*, 11(2) :537–562, 2020.

[3] Yoann Robin, Mathieu Vrac, Philippe Naveau, and Pascal Yiou. A dynamical optimal transport correction for multivariate biases in climate simulations. *Hydrology and Earth System Sciences*, 23(12) :773–793, 2019. Exact page range à vérifier dans le PDF original.

[4] Matthew Thorpe. Introduction to optimal transport. Lecture notes, University of Cambridge, 2018. Lent term 2017–2018, current version 8 March 2018.

[5] Thomas CAVALLAZZI. Notes on optimal transport. Technical report, Cours/notes internes, n.d. PDF fourni par l'auteur, à compléter si l'éditeur ou l'année précise est connue.