

Régression sur motifs spatiaux pour les données météorologiques spatialisées

Vihotogbé Houssou & Julie Carreau

28 janvier 2025



Plan

- 1 Introduction
- 2 Données et méthodes
- 3 Résultats préliminaires
- 4 Conclusion

Introduction (1)

- Ressources en eau et changements climatiques (CC)
- Etudes d'impacts \implies Modèles hydrologiques
- Besoin de données spatialisées : disponible sur une grille régulière
 - en période historique
 - interpolation spatiale : krigeage (+ avec dérive externe), inverse distance
 - réanalyses : télédétection, modèle de prévision + observations
 - en période future
 - projections des modèles de climats
- Limites des méthodes d'interpolation
 - plusieurs sources de données auxiliaires
 - effets isotropiques
 - incohérence de la structure spatiale entre le passé et le future

Introduction (2)

- **Notre méthode** : régression sur motifs spatiaux (RMS)
 - utilisation systématique pour interpoler les observations
 - une seule source d'informations auxiliaires
 - garantie la cohérence de l'information spatiale entre le passé et le futur
- **Hypothèse** : représentativité des motifs spatiaux
- Application et comparaison avec des approches classiques d'interpolation spatiale : **cadres contrôlés**
 - taille de la région
 - variabilité climatique (spatiale)
 - densité de stations

Données

- Données du projet ClimEx
 - Ensemble de 50 simulations de CanESM2 : un seul membre
 - Sélection de plusieurs sous-régions emboîtées dans différentes zones
 - croisement densité x taille x variabilité
- Variables d'intérêt : pas de temps journalier
 - précipitations (pr), températures minimum (tasmin) et maximum (tasmax)
- Périodes d'analyses
 - Période de référence : 1980 - 2009
 - Période de test 1 : 2000 - 2009
 - Période de test 2 : 2010 - 2019
- Choix des deux zones
 - Analyse de la variabilité spatiale
 - Choix de sorte à avoir une différence significative de variabilité spatiale
 - Voir plus dans la section "analyses préliminaires"

Régions d'étude (1)

- Québec : 3 sous-régions emboîtées dans chaque zone
 - Sud (zone A) et Nord (zone B) du Québec

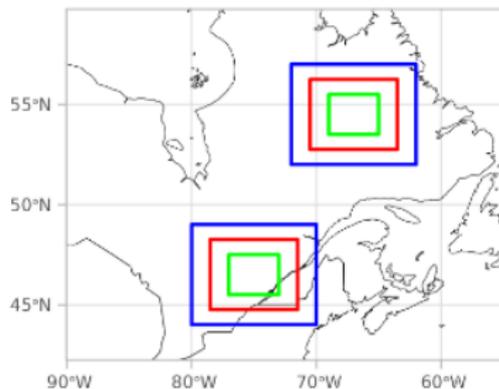
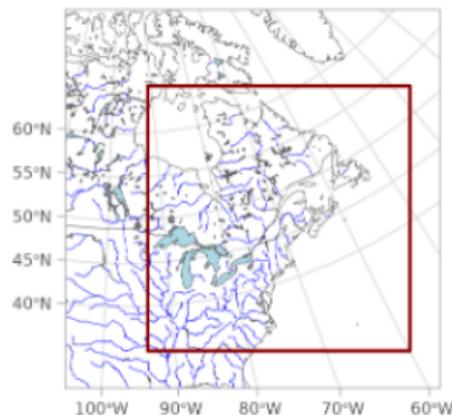


Figure – Région de climex et sous-régions d'étude

Régions d'étude (2)

- Tailles des régions emboîtées

Size of regions	Region A			Region B		
	A1	A2	A3	B1	B2	B3
Number of grid cells	3025	1332	342	2970	1332	342

Table – Tailles des régions

Analyses préliminaires (1)

- Différentes niveau de variabilité climatique (spatiale) entre A et B
 - grandes régions de A1 et B1 : régions en bleu
 - *tests statistiques* : moyennes journalières dans A > B
 - ACP dans A1 et B1 : variabilité spatiale
 - élevée dans les 2 régions, mais plus en A qu'en B pour la période de référence

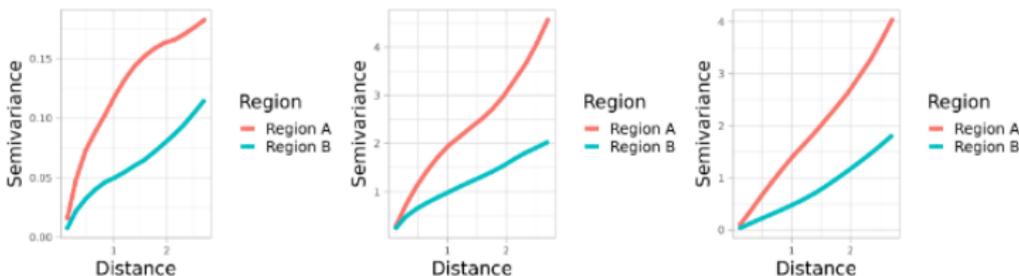


Figure – Semivariances : pr, tasmin et tasmax de gauche à droite

Analyses préliminaires (2)

- Pics de pr, tasmin et tasmx dans la même période
- Climatologies dans A1

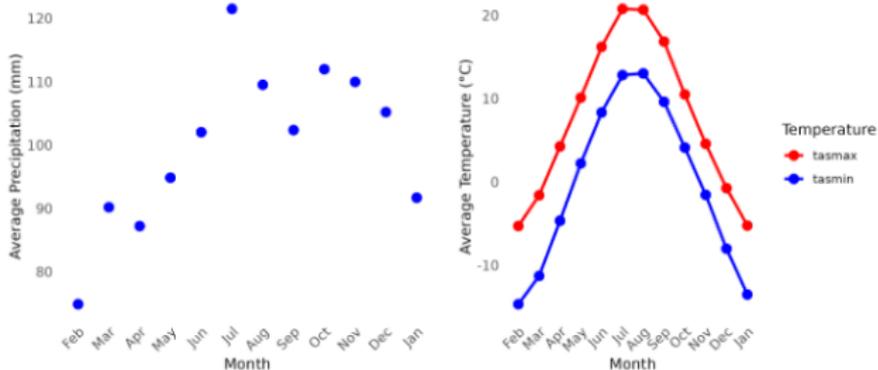


Figure – Climatologies : pr, tasmin et tasmx de gauche à droite

Cadre expérimental (1)

- Les stations synthétiques instrumentés correspondent à un sous-ensemble des mailles
- Le reste des mailles constituent les points où l'estimation (ou interpolation) doit être faite
- Cadre expérimental synthétique
 - contrôle total des données
 - élimination des biais des observations
 - manque de stations dans certaines régions
 - panne des instruments de mesure
 - incohérence dans les données historiques
 - couverture homogène et continue dans l'espace-temps

Cadre expérimental (2)

- Création de configurations synthétiques : croisement des facteurs
 - 2 régions x 3 tailles x 5 densités
 - 30 cas de comparaisons/variable/méthode
 - effets de chaque facteur sur la performance
- Contrôle de l'effet des facteurs
 - Variabilité spatiale : 2 régions spatialement variables
 - Taille des régions : 3 tailles de régions par zone
 - Densité de stations : 5 différents % de mailles sans données

Méthodes d'interpolation spatiale (1)

- Pondération de Distance Inverse (PDI)

$$\hat{Z}(s_0) = \sum_{i=1}^n W_i Z(s_i) \quad (1)$$

avec $W_i = \frac{\frac{1}{d_i^p}}{\sum_{i=1}^n \frac{1}{d_i^p}}$

- Krigeage Ordinaire (KO) : le modèle est : $Z(s_0) = \mu + \epsilon(s_0)$

$$\hat{Z}(s_0) = \sum_{i=1}^n W_i(s_0) Z(s_i) \quad (2)$$

Les poids de krigeage $W_i(s_0)$ sont déterminés en résolvant un système d'équations linéaires impliquant la fonction de semivariance.

Méthodes d'interpolation spatiale (2)

- Krigeage avec Dérive Externe (KDE)

$$Z(s_0) = \mu(s_0) + \epsilon(s_0) \quad (3)$$

avec $\mu(s_0) = a + b X(s_0)$, X est la dérive externe

Régression sur motifs spatiaux (RMS)

- Extraire les motifs spatiaux ("spatial patterns") projections climatiques
 - Les projections climatique couvrant la même région d'intérêt
 - EOF/ACP pour obtenir les motifs spatiaux : vecteurs propres
 - un motif spatial (MS) couvre toute la région d'intérêt
 - Pour chaque champ : combinaison **optimale** de motifs spatiaux
- Calibrer la régression
 - Restreindre de la couverture spatial du MS aux mailles contenant des stations
 - ajuster un modèle de régression à chaque pas de temps considéré

Régression sur motifs spatiaux (suite)

- RMS : illustration
 - soit $X_{n,p}$ les données du RCM sur grilles régulières et $Z_{m,p}$ les observations à interpoler avec $d < p$ stations valides.
 - soit k est le nombre de MS retenus. Pour chaque jour j de $Z_{m,p}$, on a :

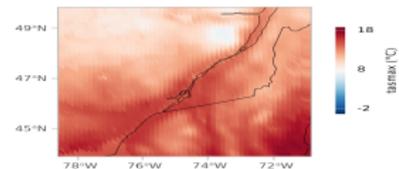
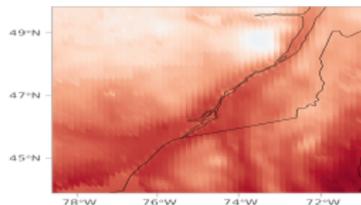
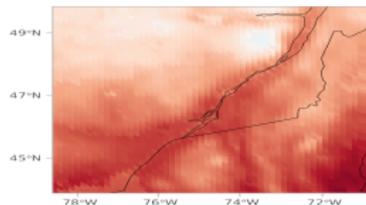
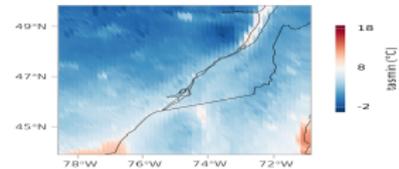
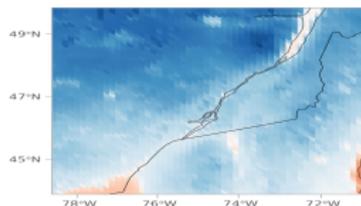
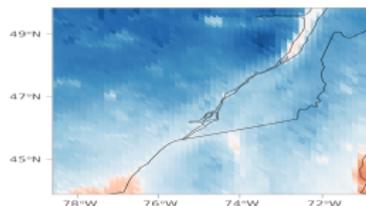
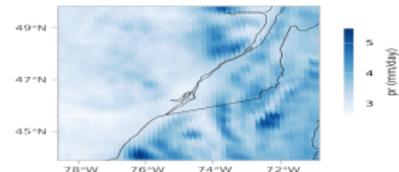
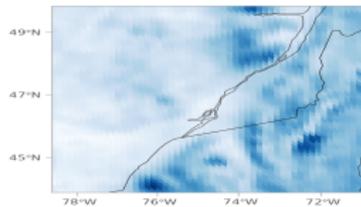
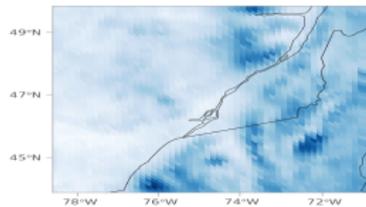
$$\begin{array}{l}
 \text{PCA} \implies \overbrace{X_{n,p}}^{\text{RCM Data}} = \overbrace{S_{n,k}}^{\text{Loading}} * \overbrace{A_{p,k}^T}^{\text{RCM Spatial patterns}} \\
 \Downarrow \\
 \text{GLM} \implies \overbrace{Z_{d,1}^j}^{\text{Valid Observations}} = \overbrace{A_{d,k}}^{\text{Restricted Spatial patterns}} * \overbrace{\beta_{k,1}}^{\text{Observations Temporal patterns}} + \varepsilon_{d,1} \\
 \Downarrow \\
 \text{Interpolation} \\
 \overbrace{Z_{p,1}^j} = \overbrace{A_{p,k}}^{\text{RCM Spatial patterns}} * \overbrace{\beta_{k,1}}^{\text{Observations Temporal patterns}}
 \end{array}$$

Figure – Illustration de RMS

Résultats préliminaires : un cas expérimental

- Premier cas expérimental
 - Région A : grande variabilité spatiale par rapport à B
 - Taille 1 : grande taille de la zone A
 - Densité : grande densité avec seulement 10% de mailles manquantes
 - Période : 2000 - 2009

Résultats préliminaires : Moyennes journalières

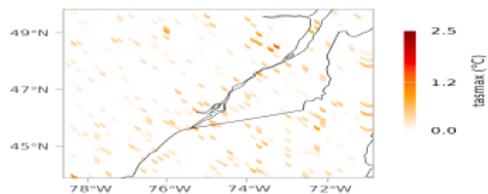
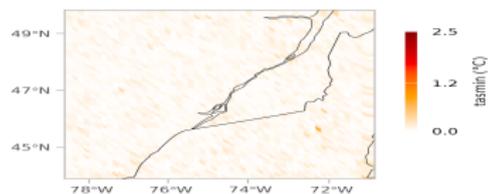
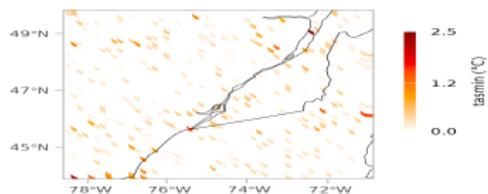
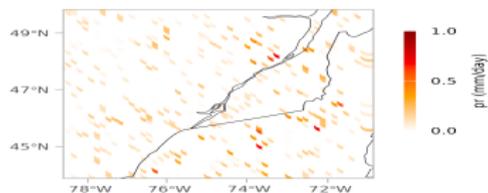


IDW

SPR

OBS

Résultats préliminaires : RMSE



IDW

SPR

Perspectives et conclusion

- Plus de métriques seront considérées par la suite
 - Prédications **vs** observations synthétiques
 - Moyennes spatiales
 - Hétérogénéité spatiale
 - Climatologie
 - Extrêmes
- Plus de méthodes ...
 - Krigeage ordinaire
 - Krigeage avec dérive externe
 - Splines à plaques minces
- Plus de cas expérimentaux : 30 au total /variable/méthode