

Régression EMD avec application à la relation santé-climat

Introduction

Contexte

Parmi les conséquences attendues des changements climatiques (CC) se trouve l'augmentation de la mortalité due à la météorologie.

Les maladies cardiovasculaires (MCV) sont notamment très affectées par la météorologie et représentent déjà l'une des principales causes de mortalité au Québec.

Il est donc important de bien quantifier l'effet de la météorologie sur les MCV afin d'anticiper les CC.

On se place donc dans le cadre d'une régression

MCV (nombre de cas au jour i) ~ **Météo** (valeur au jour i)

Problématique

La structure temporelle des données implique deux principaux problèmes :

- Les séries **météo** sont interdépendantes
→ Diminue la précision de la régression
- Les séries **sanitaires** et **météo** ne sont pas stationnaires
→ Crée des corrélations fallacieuses

Régression EMD (R-EMD)

Séries temporelles **sanitaires Y** et **météo X**

Séparation des modes d'oscillation des séries par empirical mode decomposition (EMD) [1] :

$$X(t) = \sum_{k=1}^n c_k(t) + r_n(t)$$

EMD

- c_k : fonction modale intrinsèque (IMF)
→ Oscillation symétrique autour de zéro, donc quasi-stationnaire
→ Si $k \neq l$, c_k et c_l ont des fréquences différentes et sont orthogonales
- $r_n(t)$: reste de la décomposition
→ Fonction ayant zéro ou un seul extremum local ; représente la tendance

2 inconvénients persistent :

- Grand nombre d'IMF c_k
- Colinéarités entre IMF de variables différentes

Utilisation du Lasso [2] : régression linéaire avec contrainte

$$\hat{\beta}^{Lasso} = \arg \min_{\beta} \|Y - X\beta\| \quad s.t. \quad \sum_{j=1}^p |\beta_j| < s$$

Régression

- Diminue la variabilité des estimations
- Opère une sélection des variables

Application

R-EMD1

EMD **X** seulement

1 régression

$$Y \sim \begin{bmatrix} C_1^{X(1)} & \dots & r^{X(1)} \\ \vdots & & \vdots \\ C_1^{X(p)} & \dots & r^{X(p)} \end{bmatrix}$$

R-EMD2

EMD **Y** et **X**

n régressions

$$\begin{bmatrix} \tilde{C}_1^Y \\ \vdots \\ \tilde{r}^Y \end{bmatrix} \sim \begin{bmatrix} \tilde{C}_1^{X(1)} & \dots & \tilde{C}_1^{X(p)} \\ \vdots & & \vdots \\ \tilde{r}^{X(1)} & \dots & \tilde{r}^{X(p)} \end{bmatrix}$$

Données

Santé

- Nombre quotidien de
 - Décès par MCV (1981 - 2011)
 - Hospitalisations par MCV (1996 - 2007)

Météo

- Valeurs quotidiennes (1981 - 2011)
 - Températures (Tmin, Tmoy, Tmax)
 - Précipitations totales (Prec)
 - Hauteur de neige au sol (Snow)
 - Pression moyenne (Pmoy)
 - Humidité relative (Hmin, Hmoy, Hmax)

Région

- Communautés métropolitaines de
 - Montréal (CMM)
 - Québec (CMQ)



Fig. 1 : Carte des régions CMM et CMQ

Résultats

Décomposition

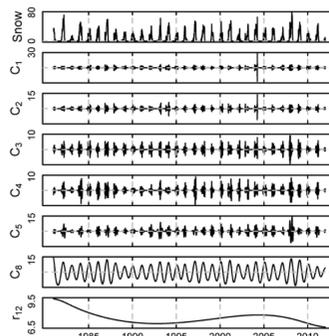


Fig. 2 : Sous-ensemble d'IMF de la variable Snow de la CMM

Exemple : Snow (CMM)

- 12 IMF + Tendance
- $c_1 - c_5$: Hautes fréquences hivernales
- c_8 : Cycle annuel, composante avec la plus forte amplitude

Résumé R-EMD

Périodicité	Décès	Hospitalisations
Court terme 1 mois	<ul style="list-style-type: none"> • Hausses pendant précipitations neigeuses en hiver 	<ul style="list-style-type: none"> • Hausse pour événements chauds et humides l'été
Moyen terme 1 an	<ul style="list-style-type: none"> • Augmentation hivernale due aux basses températures • Hausse hivernale accentuée par l'humidité automnale 	<ul style="list-style-type: none"> • Augmentation hivernale due aux basses températures • Hausse hivernale accentuée par l'humidité automnale • Hausse hivernale allongée par des chutes de neige automnales et printanières
Long terme	<ul style="list-style-type: none"> • Tendance décroissante due à la hausse des températures et l'augmentation de l'humidité 	<ul style="list-style-type: none"> • Tendance croissante due à la hausse des températures et l'augmentation de l'humidité

Évaluation

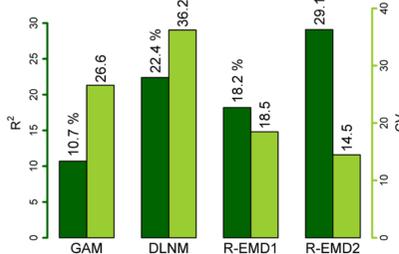


Fig. 3 : Critères d'évaluation des modélisations

Modèles comparés :

- GAM et DLNM [3]
- Critères :
 - R^2 : variance expliquée
 - Validation croisée (CV) : erreur de prédiction

Exemple : Décès CMM

- R-EMD2 meilleur sur R^2 et CV
- DLNM a un meilleur R^2 que R-EMD1 mais moins bon CV

Références

- [1] Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung and H. H. Liu (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **454**(1971): 903-995.
- [2] Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological) **58**(1): 267-288.
- [3] Gasparri, A., B. Armstrong and M. G. Kenward (2010). "Distributed lag non-linear models." Statistics in Medicine **29**(21): 2224-2234.

Conclusion

La régression EMD permet de mettre en évidence les échelles temporelles de la relation entre les MCV et la météo :

- Confirmation du rôle de la température, notamment à très court terme et à l'échelle annuelle ;
- Mise en évidence de l'humidité comme exposition influente ;
- Amélioration de l'ajustement et de la prédiction de la réponse.