Session No.

The timescales of the relationship between a response and exposures by EMD-regression

Pierre Masselot* (1), Fateh Chebana (1), Diane Bélanger (1,2), André St-Hilaire (1), Belkacem Abdous (2,3), Pierre Gosselin (1,2,4)

(1) Institut national de la recherche scientifique (2) Centre hospitalier universitaire de Québec
(3) Université Laval (4) Institut national de santé publique du Québec
** pierre-lucas.masselot@ete.inrs.ca*

Introduction

Context

In future years, it is expected that weather related mortality and morbidity should increase because of climate change.

In particular, cardiovascular diseases (CVD) are expected to be impacted. They already represent the second cause of mortality in the province of Quebec.

An important public health challenge is thus to well understand the weather impact on CVDs, usually through a regression model

$$Y = \sum_j \beta_j X_j + \varepsilon$$

Y: **CVD** (# cases), X_j: **weather variable**

Problematic

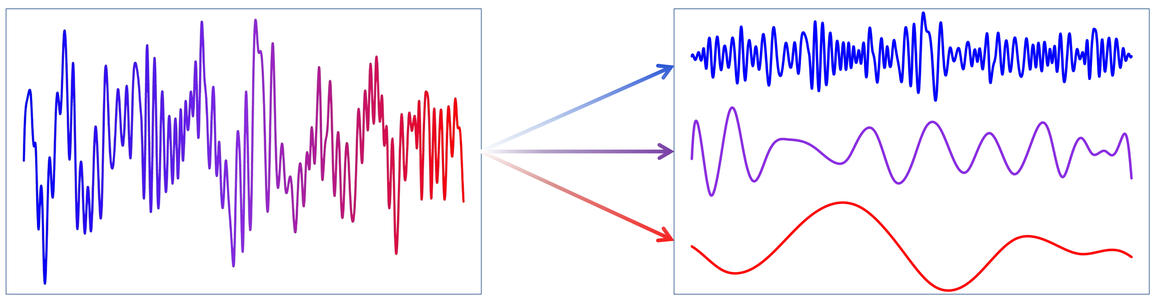
The time structure of the data prevents the regression analysis to be applied in a straightforward way. Two main issues occur:

- Weather variable are correlated because of their common seasonalities ;
- Series are not stationary.
  → Decreases estimations accuracy ;
  → Increases the chances of fallacious correlations.

Proposed solution: decompose the time series into basic oscillating component through empirical mode decomposition (EMD) and use them in regression analysis.
→ Outlines the important time scales of the relationship
→ Respects the hypotheses of regression

Methods: EMD regression (EMD-R)

EMD

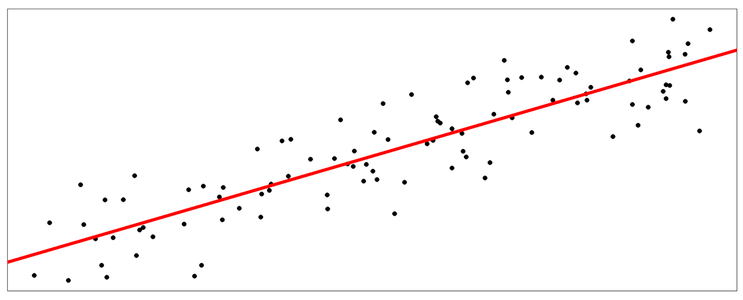


Decompose a series into basic oscillating components:

$$X(t) = \sum c_k(t) + r(t)$$

- $c_k(t)$: intrinsic mode function (IMF) which is a symmetric oscillating component around the zero line ;
- $r(t)$: monotone component assimilated to the trend.

Regression model



EMD results in a large number of components, so we make use of the *Lasso* for regression:

$$\hat{\beta} = argmin\|Y - X\beta\| \quad \text{w.r.t} \quad \sum|\beta_j| < s$$

- Shrinks several coefficients to zero to operate variable selection ;
- Manages the remaining correlation.

Design

EMD-R1

1 regression model

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

EMD-R2

n regression models

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

Two different EMD-R model:

- *EMD-R1*: $Y \sim X$ IMFs
  → outlines the main scales of influence of the weather ;
- *EMD-R2*: Y IMF $\sim X$ IMFs
  → provides the detail of the relationship at each scale

Results

Computation of a *Sensitivity* value:
→ $\hat{\beta}$ scaled by the amplitude of the associated IM

EMD-R1 results:

- Only humidity affects mortality at lower scales
- Strong effect of temperatures at 1-year scale
- Strong effect of both humidity and temperature at the trend level

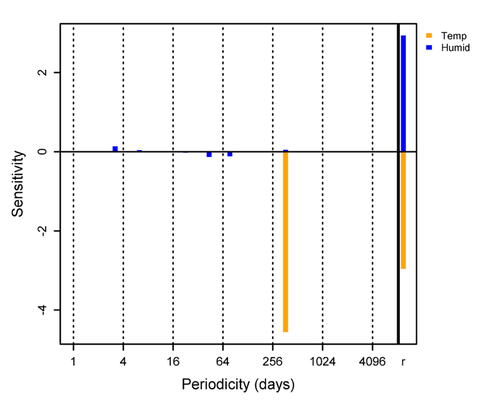


Fig.2: EMD-R1 sensitivity values

EMD-R2 results:

- Influence of humid and hot events in summer at the 5-day scale
- Influence of dry periods at the monthly scale
- Still strong effect of temperatures at the yearly scale

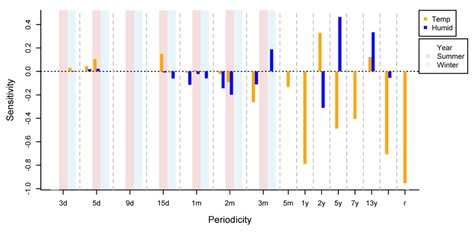


Fig.3: EMD-R2 sensitivity values

Comparison with classical models:

- EMD-R has better performances
- EMD-R2 is better for explaining but EMD-R1 is better for predicting

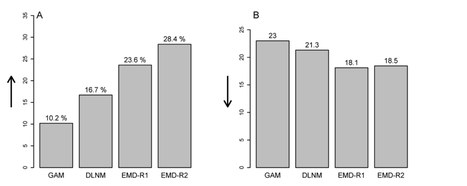


Fig.3: R^2 (left) and GCV (right) of the models

Data

Y: daily CVD death number

X: daily mean temperature and humidity

Period: 1981 – 2011 included

Region: Greater Montreal area (Canada)

Fig. 1: Montreal location

Conclusion

Methodological side:
→ EMD-R allows reporting an association in terms of time scale.
→ Outlines the main scales at which a relationship occurs
→ Enhance the explaining and predicting power of models

Epidemiological side:
→ Strong annual effect of temperatures found
→ At short scales, influence of humidity

References

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.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological) **58**(1): 267-288.

Yang, A. C., Fuh, J.-L., Huang, N. E., Shia, B.-C., Peng, C.-K., & Wang, S.-J. (2011). Temporal Associations between Weather and Headache: Analysis by Empirical Mode Decomposition. *PLoS ONE, 6*(1), e14612. doi: 10.1371/journal.pone.0014612

Institut national de santé publique Québec

Fondsvert

INRS UNIVERSITÉ DE RECHERCHE