

# Modélisation de la température de l'eau : comparaison des modèles GAM, Sigmoide et Autorégressif

Fatima Laanaya, André St-Hilaire et Erwan Gloaguen

Institut National de la Recherche Scientifique (INRS), Centre Eau Terre Environnement, Québec (Québec), Canada



## 1. Introduction

- Changement climatique → augmentation de la température de l'air

Température de l'eau

un catalyseur, un dépressur, un activateur, un limiteur, un stimulateur, un contrôleur, un tueur, est l'une des caractéristiques de la qualité de l'eau les plus importants et influents de la vie dans l'eau [1]



## 2. Objectifs

- Évaluer la performance du modèle additif généralisé (GAM) pour prédire la température de l'eau
- Mesurer l'impact du débit

## 3. Données et zone d'étude

Le site de l'étude est situé sur la rivière St-Marguerite Nord-Est à Québec

Variable réponse : moyenne journalière de la température de l'eau

Prédicteurs : moyennes journalières de la température de l'air et de débit

Période d'étude : mai- octobre, 2007 à 2013

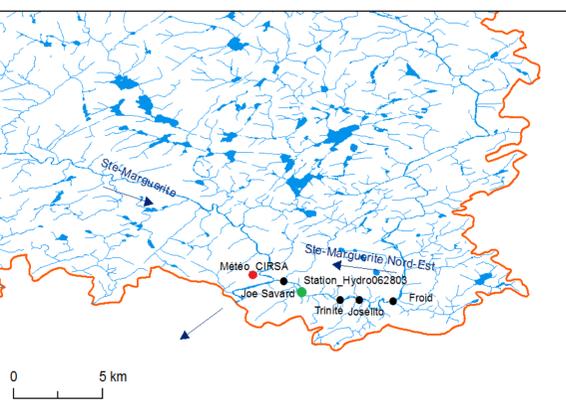


Fig 1 : Zone de l'étude et localisation des stations de mesure utilisées

## 4. Méthodologie

### 4.1. Modèle additif généralisé

- Si on se restreint au modèle linéaire, étant donné une variable réponse

$y$  et  $p$  prédicteurs  $x_1, x_2, \dots, x_p$

- Le modèle linéaire s'écrit :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

- Ce modèle est basé sur l'hypothèse forte de linéarité des paramètres

- Une possibilité pour pallier à cette hypothèse
- conservant l'effet additif de chaque prédicteur

Modèle GAM [2] :

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \varepsilon$$

- L'introduction des fonctions  $f_j$

accorde plus de flexibilité  
les effets non linéaires des prédicteurs

- Les fonctions  $f_j$  sont inconnues et sont présentées par des fonctions de lissage (splines, loess, ...)

- Notre étude utilise GAM + Splines de lissage cubiques (splines pénalisés) qui minimisent :

$$\sum_i [y_i - f_j(x_{ij})]^2 + \sum_j \lambda_j \int f_j''(t_j)^2 dt$$

Somme carrée des erreurs

la pénalité de rugosité

## 4. Méthodologie (suite)

- Les valeurs élevées de  $\lambda$  engendrent des courbes lisses
- les faibles valeurs de  $\lambda$  donnent des courbes plutôt rugueuses
- L'estimation du modèle se fait par «Penalized Iteratively Reweighted Least Squares» [3]

- Les fonctions sont représentées par :

$$f(x) = \sum_{k=1}^q b_k(x) \beta_k$$

Où  $b_k(x)$  sont les bases des splines

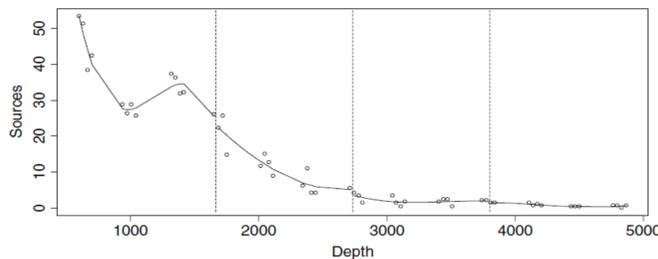


Fig 2 : Illustration de régression spline (Zuur et al., 2009)

### 4.2. Modèle logistique ou sigmoïde

Le modèle sigmoïde ou logistique [4]

$$T_w = \frac{\alpha}{1 + e^{\gamma(\beta - T_a)}}$$

$T_w$  est la température de l'eau

$T_a$  est la température de l'air

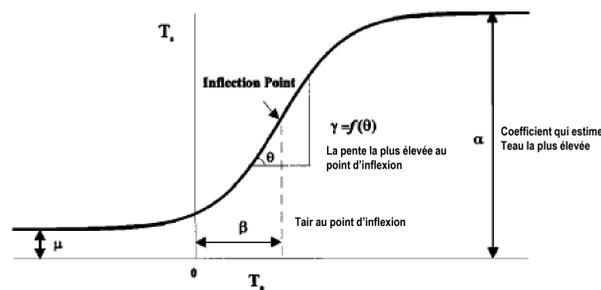


Fig 3 : Schéma représentant les paramètres du modèle sigmoïde [4]

### 4.3. Modèle autorégressif

- Le modèle autorégressif sépare la température de l'eau en deux composantes [5]

$$T_w = T_{A(t)} + R_{w(t)}$$

- une composante saisonnière représentée par une fonction sinusoïdale

$$T_{A(t)} = a + b \sin \left[ \frac{2\pi}{365} (t + t_0) \right]$$

- une composante résiduelle représentée par une chaîne de Markov d'ordre 2

$$R_{w(t)} = \phi_1 R_{w(t-1)} + \phi_2 R_{w(t-2)} + \varepsilon_t$$

## 5. Résultats et discussions

### 5.1. Présentation des données

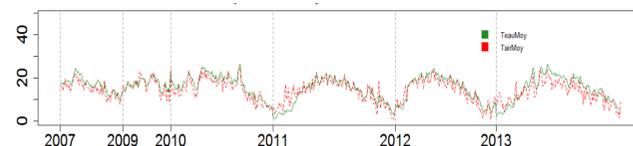


Fig 4 : Présentation de température de l'eau, de l'air et le débit

### 5.2. Modèle additif généralisé

Tab 1 : Significativité des effets

Lissage des variables	Degré de liberté	Test de Fisher	P-value
S(Température de l'air)	3.4	168.3	2.2 e-16
S(Débit)	3.5	75.9	
S(temps)	5.8	72.2	
R <sup>2</sup> ajusté		0.93	
Déviante expliquée		92.8%	

## 5. Résultats et discussions (suite)

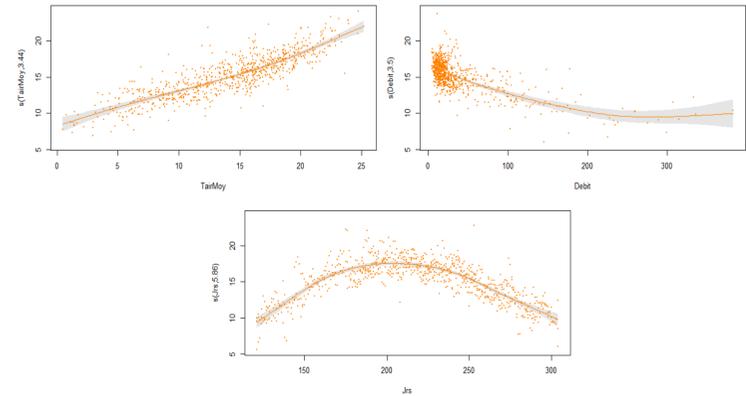


Fig 5 : Présentation des effets de température de l'air, débit et saisonnier

### 5.3. Comparaison des modèles

Tab 2 : Critères de validation par année des trois modèles

Modèle	GAM		Sigmoïde		AR	
	RMSE	Biais	RMSE	Biais	RMSE	Biais
2007	1.28	0.47	1.86	-0.88	2.19	1.20
2009	2.08	1.32	2.35	0.63	1.56	2.24
2010	1.83	-0.28	1.83	-1.08	3.57	-1.34
2011	1.75	0.89	3.55	1.99	2.80	1.39
2012	1.32	-0.01	2.03	0.06	1.86	-0.89
2013	2.18	-1.35	3.40	-0.85	2.87	-0.56
Global	1.79	0.01	2.75	0.02	2.80	0.09

- seules les données de la période chaude sont disponibles en 2009
- L'année 2013 a connue une forte crue, l'air a du mal à chauffer l'eau

### 5.4. Évaluation des prévisions

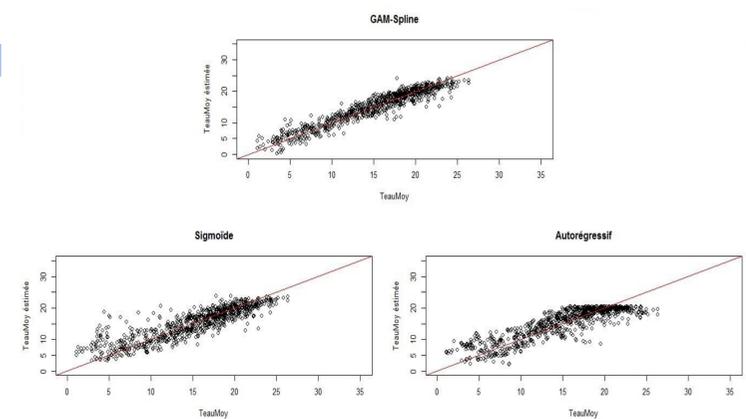


Fig 6 : Évaluation des modèles, valeurs prédites vs observées

- Alignement meilleur pour GAM
- Sigmoïde sous-estime les valeurs faibles
- Autorégressif sous / sur estime les valeurs extrêmes

## 6. Conclusion

GAM est :

- très flexible
- excellent ajustement en présence de relations non-linéaires
- plus robuste aux non stationnarités

## 7. Références

- Federal Water Pollution Control Administration. 1967. *Temperature and Aquatic Life*. Laboratory Investigations No. 6, Technical Advisories and Investigations Branch, Cincinnati, Ohio.
- Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical science*, 297-310.
- Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press
- Mohseni O, Stefan HG & Erickson TR (1998) A nonlinear regression model for weekly stream temperatures *Water Resources Research* 34(10):2685-2692.
- Caissie, D., El-Jabi, N., & St-Hilaire, A. (1998). Stochastic modelling of water temperatures in a small stream using air to water relations. *Canadian Journal of Civil Engineering*, 25(2), 250-260.

## 8. Coordonnées

Fatima Laanaya, Maitrise, Modélisation statistique de la température de l'eau  
Email : fatima.laanaya@ete.inrs.ca