

# Tendance nationale de l'Indice Poisson Rivière (IPR) estimée par un Modèle Additif Généralisé (GAM)

P. Irz<sup>a</sup>, M. Levi-Valensin<sup>a</sup>

<sup>a</sup> Ministère de l'Ecologie-Service de l'Observation Et des Statistiques  
Commissariat Général au Développement Durable  
5 route d'Olivet, 45 000 Orléans  
pascal.irz@developpement-durable.gouv.fr  
michael.levi-valensin@developpement-durable.gouv.fr

**Mots clefs** : Biodiversité, Modèle additif, Données répétées, Package mgcv .

L'IPR ou **Indice Poisson Rivière** [1] est un outil d'évaluation de la qualité des cours eau produit par l'Onema (Office national de l'eau et des milieux aquatiques). Il décrit l'écart entre les caractéristiques du peuplement observé lors d'une pêche et les valeurs de référence. À partir de cet indice synthétique et composite, notre objectif est d'**évaluer si l'état des cours d'eau tend à s'améliorer, ou à se dégrader sur le long terme**. La modélisation statistique des mesures (brutes ou transformées) en fonction des stations et de l'année de collecte est d'usage courant dans l'analyse des indices biologiques d'abondance. Elle a pour principaux avantages de permettre la prise en compte de covariables et de fournir des statistiques de diagnostic. Les prédicteurs sont l'identifiant de la station, l'année et une ou plusieurs covariables comme effets fixes ou aléatoires [2] : dans tous les cas, la tendance estimée correspond à l'effet lié à la variable *année*.

On peut envisager un **modèle additif généralisé** (ou GAM) si la relation est non linéaire mais indéterminée *a priori* .

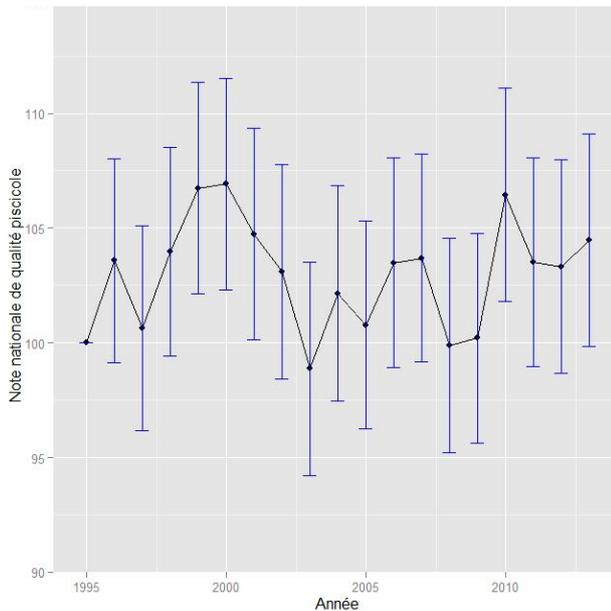
Plusieurs modèles ont été testés avec comme covariables :

- les deux variables *stratégie et moyen de pêche* pour contrôler statistiquement la variabilité des mesures liée aux changements de méthodes de pêche.
- le *code bassin* pour exclure de la tendance d'évolution d'une station son appartenance à un bassin versant.
- une fonction spline des *coordonnées cartographiques (X,Y)* de la station pour tenir compte de l'hétérogénéité spatiale du modèle.
- le numéro du jour de la mesure dans l'année (*jour.année*) pour prendre en compte une éventuelle saisonnalité.
- le *code station* introduit comme effet aléatoire du fait de la variabilité du champ des stations d'une année sur l'autre.

Plusieurs packages permettent de construire des modèles linéaires généralisés, des modèles mixtes (lme4) ou non linéaires mixtes (nlme). Notre choix s'est porté sur le package *mgcv* (Mixed GAM Computation Vehicle) [3] qui autorise la modélisation d'effets aléatoires ainsi que de relations additives, supposées non linéaires avec les options de lissage de la fonction spline (*bs="cc"*, *bs="re"*). La fonction *bam* s'applique à des données volumineuses et permet des traitements plus rapides que la fonction *gam* surtout en présence d'effets aléatoires [4].

Deux options sont envisageables pour inclure l'année :

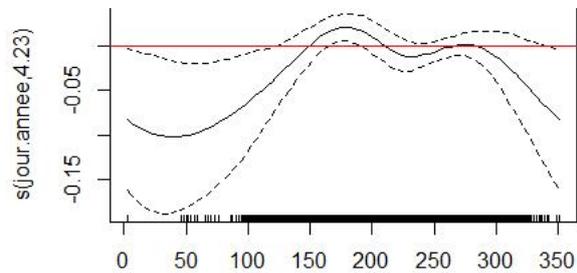
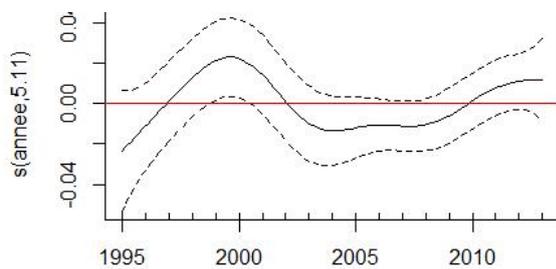
- en variable qualitative et quantifier l'effet annuel comme un écart à une année de référence.



```
mod6 <- bam (logipr ~ as.factor(annee) +
s(code.station, bs="re") + s(bassin, bs="re")
+ s(jour.annee,bs="cc") + s(xlambert93, ylam-
bert93) + intit.method + intit.moyen, fa-
mily=gaussian, data = sub3)
```

Seules les années 2002, 2007 et 2013 présentent une différence significative avec l'année de référence 1995 avec une situation plus favorable. Par rapport à la tendance pluri-annuelle à la hausse, l'année 2003, marquée par la canicule, détonne.

- ou par une fonction spline de la variable quantitative qui représente une courbe lissée de la tendance annuelle (à.g) et des effets journaliers permettant de prendre en compte une éventuelle saisonnalité (à.dr).



```
mod7 <- bam (logipr ~ s(annee) + s(jour.annee, bs = "cc") +s(code.station, bs = "re") +
s(bassin, bs = "re") + s(xlambert93, ylambert93) + intit.method + intit.moyen,data=sub3)
```

On constate une saisonnalité significative avec des notes IPR moins bonnes que la moyenne lorsque les pêches sont réalisées de janvier à avril, et meilleures que la moyenne entre les jours 165 et 190 (de mi-juin à mi-juillet).

## Références

- [1] Oberdorff, T., Pont, D., Hugueny, B. & Porcher, J.-P. (2002). Development and validation of a fish-based index for the assessment of « river health » in France. *Freshwater Biology*, 47(9), 1720-1734.
- [2] Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. (Ed.). (2009). *Mixed effects models and extensions in ecology with R*. New York, NY : Springer.
- [3] Wood, S. (2015). Package **mgcv** version 1.8-7. <http://cran.stat.auckland.ac.nz/web/packages/mgcv>
- [4] Wood, S. N., Goude, Y., Shaw, S. (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*