



**HAL**  
open science

## Essais de définition d'états de référence pour les estuaires et les lagunes

Hilaire Drouineau, C. Delpech, Philippe Boët, Mario Lepage

► **To cite this version:**

Hilaire Drouineau, C. Delpech, Philippe Boët, Mario Lepage. Essais de définition d'états de référence pour les estuaires et les lagunes. [Rapport de recherche] irstea. 2011, pp.24. hal-02596682

**HAL Id: hal-02596682**

**<https://hal.inrae.fr/hal-02596682>**

Submitted on 15 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Partenariat 2010 – Domaine 2.4. - Action 12**



# Essais de définition d'états de référence pour les estuaires et les lagunes

*Rapport final*

***Hilaire Drouineau (Cemagref Bordeaux)***

***Christine Delpech (Cemagref Bordeaux)***

***Philippe Boët (Cemagref Bordeaux)***

***Mario Lepage (Cemagref Bordeaux)***

## Contexte de programmation et de réalisation

---

En France il n'existe pas de site de référence pour les différents types d'eaux de transition, même historiquement. Cela nécessite donc une construction théorique par simulation. Une première tentative de modélisation par l'approche des modèles CART a été réalisée en 2008 pour prédire des présence/absence d'espèces ou de guildes dans certaines conditions décrivant les habitats physiques et leurs contextes de pressions anthropiques. Les premiers résultats suggèrent qu'il est possible d'aller plus loin.

## Les auteurs

---

**Hilaire Drouineau**

Ingénieur de recherche  
[hilaire.drouineau@cemagref.fr](mailto:hilaire.drouineau@cemagref.fr)  
Cemagref Bordeaux

**Christine Delpech**

Ingénieur d'études  
[christine.delpech@cemagref.fr](mailto:christine.delpech@cemagref.fr)  
Cemagref Bordeaux

**Philippe Boët**

Directeur de recherche  
[philippe.boet@cemagref.fr](mailto:philippe.boet@cemagref.fr)  
Cemagref Bordeaux

**Mario Lepage**

Ingénieur de recherche  
[mario.lepage@cemagref.fr](mailto:mario.lepage@cemagref.fr)  
Cemagref Bordeaux

## Les correspondants

---

Onema : Romuald, Berrebi, [direction, romuald.berrebi@onema.fr](mailto:romuald.berrebi@onema.fr)  
Référence du document :

Cemagref : Mario, Lepage, [Cemagref, mario.lepage@cemagref.fr](mailto:mario.lepage@cemagref.fr)  
Référence du document :

<b>Droits d'usage :</b>	<i>accès libre</i>
<b>Couverture géographique :</b>	<i>France</i>
<b>Niveau géographique</b> [un seul choix] :	<i>National</i>
<b>Niveau de lecture</b> [plusieurs choix possibles] :	<i>Professionnels, experts</i>
<b>Nature de la ressource :</b>	<i>Document</i>

***Essais de définition d'états de référence pour les estuaires et les lagunes***  
***Rapport final***  
***Hilaire Drouineau, Christine Delpech, Philippe Boët, Mario Lepage***

## Sommaire

<u>2</u>	
<b><u>Résumé.....</u></b>	<b><u>5</u></b>
<b><u>Abstract.....</u></b>	<b><u>6</u></b>
<b><u>Synthèse pour l'action opérationnelle.....</u></b>	<b><u>7</u></b>
<b><u>Corps du document.....</u></b>	<b><u>10</u></b>
<u>2</u>	

**Essais de définition d'états de référence pour les estuaires et les lagunes**  
*Hilaire Drouineau, Christine Delpech, Philippe Boët, Mario Lepage*

## Résumé

### RÉSUMÉ

La Directive Cadre sur l'Eau impose de quantifier l'état écologique des masses d'eau par rapport à un état de référence, qui correspondrait à un état vierge. En France, aucune masse d'eau ne peut-être considérée dans un état vierge, et aucune donnée historique ne permet de caractériser cet état, c'est pourquoi il est nécessaire de le modéliser. Pour cela, l'utilisation d'arbres de classification et de régression a été envisagée et a fourni des premiers résultats encourageants les années précédentes. Cette année, une méthode à partir d'arbres de régressions multiples qui avait donné de bons résultats sur les lagunes, a été appliquée aux estuaires. Malheureusement, cette méthode n'a pas permis de fournir de prédiction pertinente pour les estuaires, du fait d'une trop forte hétérogénéité de ces masses d'eau. Des arbres de régression boostés ont également été proposés pour prédire la présence/absence de guildes de poissons. Ces arbres, contrairement aux arbres de régression et de classification traditionnels, sont conçus explicitement pour la prédiction et pourraient donc constituer des outils pertinents pour la prédiction d'un état de référence. Une première analyse sur les poissons d'eau douce donne des résultats satisfaisants, en revanche, la méthode souffre de faiblesses en extrapolation qui risque de limiter les possibilités d'utilisation. Une troisième approche basée sur la construction d'une typologie des lagunes méditerranéennes à partir de la structure de l'assemblage ichtyofaunistique a été proposée. Cette approche a permis de mettre en évidence des disparités fortes entre lagunes. Une corrélation avec les caractéristiques physiques des masses d'eau et avec les seuils de tolérance physico-chimiques définis pour les espèces utilisées dans la classification devrait permettre de mieux comprendre les facteurs influençant la typologie des lagunes, et éventuellement d'améliorer la définition des états de référence.

### MOTS CLÉS (THÉMATIQUE ET GÉOGRAPHIQUE)

Directive Cadre sur l'Eau, Eaux de transition, Etat de référence, Arbres de classification et de régression, Arbres de régression boostés.

***Essais de définition d'états de référence pour les estuaires et les lagunes***  
*Hilaire Drouineau, Christine Delpech, Philippe Boët, Mario Lepage*

## **Abstract**

### **ABSTRACT**

In the context of The Water Framework Directive, the ecological status of water bodies must be assessed with respect to a reference state, which corresponds to a pristine state. In the absence of transitional water body in France that can be considered in a virgin state, and of historical data enabling to characterise a pristine state, statistical methods are required to predict and describe a reference state. In this context classification and regression trees have been tested since 2008 to predict reference state, providing promising results. This year, a method based on multiple regression trees which was developed in 2009 to predict reference state in French lagoons, has been applied to estuaries. However, the method does not appear to be relevant for estuaries, because of a greater diversity of physical conditions. Boosted Regression Trees (BRT), which are especially dedicated to prediction, were applied to predict the presence probability of some fish guilds which are poorly modelled with traditional statistical methods. The method has given consistent results for freshwater fishes and the method can probably be applied to other guilds and metrics. However, BRT suffers from limited extrapolation capacity which may be problematic when predicting a reference state. Finally, a lagoons typology based on fish assemblage structure has been carried out. It has pointed out a great variety in fish assemblage structure, which should be correlated with environment and/or anthropogenic factors, and fish tolerance with respect to water quality in the future. It should help to understand which factors are the most important on lagoons assemblage and consequently, should be incorporated in pressure-impact models, and finally it will help to characterise a reference state.

### **KEYWORDS (THEMATIC AND GEOGRAPHICAL AREA)**

Water Framework Directive, Transitional water bodies, Reference state, Classification and regression trees, Boosted regression trees.

**Essais de définition d'états de référence pour les estuaires et les lagunes**  
*Hilaire Drouineau, Christine Delpech, Philippe Boët, Mario Lepage*

## Synthèse pour l'action opérationnelle

### Contexte

La mise en œuvre de la Directive Cadre sur l'Eau (DCE - Directive 2000/60/EC) nécessite d'évaluer l'état des communautés piscicoles des masses d'eau de transition, par le développement de méthodologies d'échantillonnage et d'interprétation des résultats, permettant le classement de cette composante dans l'une des 5 classes de qualité. Cette évaluation se fait par rapport à un état de référence correspondant à un état vierge, ne subissant aucune pression anthropique. En France il n'existe pas de site de référence pour les différents types d'eaux de transition, même historiquement. Une construction théorique par simulation est donc nécessaire. Une première tentative de modélisation par l'approche des modèles CART a été réalisée en 2008 (Boët, 2009) et en 2009 (Delpech *et al.*, 2010b ; Drouineau *et al.*, 2010) pour prédire des présence/absence d'espèces ou de guildes dans certaines conditions décrivant les habitats physiques et leurs contextes de pressions anthropiques. Les premiers résultats suggéraient qu'il était possible d'aller plus loin.

Dans ce rapport, nous présentons la poursuite des travaux menés en 2010 concernant l'utilisation des techniques de classification fondées sur les arbres de régression (CART, *Classification and Regression Trees*) pour aider à la définition des états de référence. Trois actions ont ainsi été engagées.

### Arbres de régressions multiples

En 2009, une approche par arbre de régressions multiples avait été mise en œuvre pour prédire un état de référence pour les lagunes méditerranéennes. Un essai d'application de cette approche relativement simple a été tenté pour les estuaires. Malheureusement, les estuaires présentent une hétérogénéité beaucoup plus importante que les lagunes et finalement aucun arbre de régression pertinent n'a pu être construit. En effet, il s'est avéré impossible de prédire les pressions sans prendre en compte les variables liées aux conditions d'échantillonnage et de l'environnement, et ces pressions devenaient complètement masquées par les variables physiques quand ces dernières étaient intégrées dans l'arbre.

Cette approche qui semblait pertinente dans le cas des lagunes, n'a donc pas été concluante pour les estuaires.

### Arbres de régressions boostés

Les arbres de régression boostés (BRT) sont des méthodes davantage destinées à la prédiction plutôt qu'à l'explication et peuvent donc être particulièrement pertinents pour prédire un état de référence dans un milieu donné. La construction d'un arbre de régression boosté repose en fait sur la construction itérative d'un nombre variable d'arbres de régression simple. Afin de tester la pertinence de cet outil, un BRT a été construit pour essayer de prédire la probabilité de présence/absence de poissons d'eau douce. Le choix a porté sur cette guildes car du fait de la forte proportion de zéro dans le jeu de données, cette guildes s'est en effet

avérée très difficile à modéliser par des GLM. Pourtant, l'utilisation de cette métrique serait intéressante afin de pouvoir mieux prendre en compte les assemblages en zone oligohaline dans l'indicateur poisson. Des graphes de dépendance partielle permettent ainsi de montrer la variation de la probabilité de présence de ces espèces d'eau douce en fonction de diverses variables explicatives. Sans surprise, la zone haline est la variable qui semble avoir le plus d'effet : la probabilité de présence est maximale en zone oligohaline. Puis la saison a un effet non négligeable sur la présence de ces espèces : le printemps semble plus favorable que l'automne. Enfin, on observe un effet des concentrations en polluants, plus important que la saison, car des niveaux faibles de contamination sont globalement plus favorables. Le BRT a aussi permis de tracer des cartes de prédictions pour chaque combinaison de saison et zone haline, à différents niveaux de contamination.

En conclusion, les BRT sont des outils qui peuvent être intéressants pour le développement de l'indicateur poisson DCE, notamment pour approcher un état de référence car ils permettent de répondre en partie à certaines limites des modèles linéaires simples (LM) et généralisés (GLM). Cela étant, ils présentent certaines limites (capacités explicatives et capacités d'extrapolation relativement limitées) qui risquent d'être problématiques quand les variables de pressions auront une influence beaucoup plus faibles que les variables d'environnement.

### **Classification des lagunes françaises à partir de l'assemblage ichtyofaunistique**

Afin d'explorer la possibilité de réaliser une typologie des lagunes méditerranéennes, une analyse des proportions des trois familles les plus abondamment représentées dans les lagunes françaises (atherinidae, gobiidae, mugilidae) a été réalisée. L'objectif était aussi de réaliser une première analyse visant à expliquer les différences obtenues sur les modèles pressions-impacts quand les trois lagunes corses échantillonnées en 2009 sont ajoutées au jeu de données.

Une analyse en composantes principales (ACP) sur la matrice des corrélations de ces proportions par opération de pêche, suivie d'une classification hiérarchique ascendante (critère de Ward, distance de Bray-Curtis), ont ensuite été réalisées pour essayer de faire une typologie des profils d'opérations de pêche. L'ACP a permis de montrer une forte opposition des trois familles, les opérations de pêches sont en général caractérisées par une forte proportion d'une seule famille.

À partir de la classification hiérarchique ascendante, il a été possible de distinguer 4 types d'opérations de pêche en fonction de la famille(s) dominante(s). Le type A (Diana, Urbino, Palo) est marqué par une très forte dominance des athérines. Le groupe B (Biguglia, Grand Bagnas, Salses-Leucate, Thau) est caractérisé par une dominance des athérines mais il présente un peu plus de diversité que le groupe A. Cela peut être dû à une plus grande diversité d'habitats liée à la taille, à une meilleure connexion à la mer ou à une plus forte influence des eaux douces. Le type C (Vaccarès, Or) est caractérisé par de fortes proportions des trois familles, avec cependant de plus fortes proportions de gobiidae. Enfin, le dernier type (Bages-Sigean, Berre, Prévost, Méjean, La Palme) n'est caractérisé par aucune dominance particulière.

Ce travail de typologie est intéressant car il démontre qu'une typologie fondée sur les assemblages de poissons est possible pour les lagunes méditerranéennes. Néanmoins, en l'état, cette dernière ne permet pas de savoir si les catégories obtenues sont liées à des caractéristiques naturelles et/ou anthropiques. Il est cependant déjà intéressant de souligner que des lagunes considérées comme relativement peu impactées par l'indice global de

pression (Diana, La Palme, Vaccarès, Grand Bagnas) se retrouvent dans des types différents. Cette classification montre également que les trois lagunes corses, échantillonnées en 2009, constituent un type relativement à part.

## Conclusion

Trois approches à partir d'analyses multivariées ont été entreprises cette année pour tenter de caractériser les états de référence des masses d'eau de transition. Le développement de ce travail n'a cependant pas pu être mené comme programmé initialement, en raison de l'impossibilité à trouver un stagiaire pour travailler sur cette thématique comme prévu.

Parmi ces trois approches, les arbres de régressions multiples, qui avaient donné de bons résultats l'an passé pour les lagunes, n'ont pas permis de modélisation de l'état de référence en estuaires.

En revanche, les arbres de régression « boostés » apparaissent comme une voie intéressante à poursuivre. Ils souffrent toutefois de certaines lacunes en extrapolation qui risquent d'être gênantes dans certaines situations pour prédire un état de référence.

Enfin la troisième approche par classification hiérarchique a permis d'établir une typologie des lagunes en fonction de familles d'espèces, présentées comme potentiellement pertinentes pour qualifier l'état des masses d'eau de transition (Taverny et al., 2009). Cette typologie devra être approfondie à l'avenir. Il en ressort malgré tout une certaine spécificité des lagunes corses, spécificité qui ne ressort pas dans les indices de pressions, ni dans les variables liées à l'échantillonnage et à l'environnement, prises en compte actuellement dans les modèles pression-impacts.

## Pour en savoir plus :

Drouineau, H., Delpech, C., Boët, P., Lepage, M., 2011. Essais de définition d'états de référence pour les estuaires et les lagunes. Convention de partenariat ONEMA-Cemagref 2010 : 21p.

Drouineau, H., Delpech, C., Lepage, M., 2010. Réponses des assemblages ichtyofaunistiques à l'eutrophisation et proposition de méthode pour la définition d'un état de référence. Convention de partenariat ONEMA-Cemagref 2009 : 12p.

Boët, P., 2009. Conditions de référence pour le compartiment "Poissons" des eaux de transition. Convention de partenariat ONEMA-Cemagref 2008 : 34pp.

[hilaire.drouineau@cemagref.fr](mailto:hilaire.drouineau@cemagref.fr)  
[christine.delpech@cemagref.fr](mailto:christine.delpech@cemagref.fr)  
[philippe.boet@cemagref.fr](mailto:philippe.boet@cemagref.fr)  
[mario.lepage@cemagref.fr](mailto:mario.lepage@cemagref.fr)

**Essais de définition d'états de référence pour les estuaires et les lagunes**  
*Hilaire Drouineau, Christine Delpéch, Philippe Boët, Mario Lepage*

## Corps du document

### 1. Contexte

La mise en œuvre de la Directive Cadre sur l'Eau (DCE - Directive 2000/60/EC) nécessite d'évaluer l'état des communautés piscicoles des masses d'eau de transition, par le développement de méthodologies d'échantillonnage et d'interprétation des résultats, permettant le classement de cette composante dans l'une des 5 classes de qualité. Cette évaluation se fait par rapport à un état de référence correspondant à un état vierge, ne subissant aucune pression anthropique. En France il n'existe pas de site de référence pour les différents types d'eaux de transition, même historiquement. Une construction théorique par simulation est donc nécessaire. Une première tentative de modélisation par l'approche des modèles CART a été réalisée en 2008 (Boët, 2009) et en 2009 (Delpéch *et al.*, 2010b ; Drouineau *et al.*, 2010) pour prédire des présence/absence d'espèces ou de guildes dans certaines conditions décrivant les habitats physiques et leurs contextes de pressions anthropiques. Les premiers résultats suggéraient qu'il était possible d'aller plus loin.

Dans ce rapport, nous présentons la poursuite des travaux menés en 2010 concernant l'utilisation des techniques de classification fondées sur les arbres de régression (CART, *Classification and Regression Trees*) pour aider à la définition des états de référence. Trois actions ont ainsi été engagées.

La première action prolonge ce qui avait été réalisé en 2009 sur les lagunes (Drouineau *et al.*, 2010) : des arbres de régressions multiples ont ainsi été construits sur les estuaires français pour prédire un état de référence.

Une deuxième analyse teste la pertinence des arbres de régression boostés pour prédire un état de référence, en prenant en exemple une guildes difficilement analysable avec les méthodes usuelles : les poissons d'eau douce.

Enfin, la troisième étude porte sur une classification de l'assemblage ichtyofaunistique des lagunes méditerranéennes en partant d'une liste d'espèces potentiellement indicatrices recensées par Taverny *et al.* (2009).

Il est à noter qu'un stagiaire devait participer à ces analyses en 2010, mais qu'il n'a pas pu être recruté. Une partie des analyses n'a donc pas pu être réalisée.

## 2. Arbre de régressions multiples : possibilité de prédiction d'un état de référence pour les estuaires français

En 2009, une approche par arbre de régressions multiples avait été mise en œuvre pour prédire un état de référence pour les lagunes méditerranéennes (Delpech *et al.*, 2010b ; Drouineau *et al.*, 2010). Cette approche reposait sur deux étapes. Une première étape consistait d'abord à construire un arbre de régressions multiples prédisant des niveaux de pression à partir des observations faites sur l'assemblage ichtyofaunistique. Une proposition de notation en découlait. Dans une seconde étape, l'arbre était ensuite utilisé pour prédire la note qu'obtiendrait une lagune de référence comparée aux notes obtenues par les lagunes échantillonnées.

Un essai d'application de cette approche relativement simple a été tenté pour les estuaires. Malheureusement, les estuaires présentent une hétérogénéité beaucoup plus importante que les lagunes et finalement aucun arbre de régression pertinent n'a pu être construit. En effet, il s'est avéré impossible de prédire les pressions sans prendre en compte les variables liées aux conditions d'échantillonnage et de l'environnement, et ces pressions devenaient complètement masquées par les variables physiques quand ces dernières étaient intégrées dans l'arbre.

Cette approche qui semblait pertinente dans le cas des lagunes, n'a donc pas été concluante pour les estuaires.

## 3. Utilisation d'arbres de régression boostés pour prédire la présence de poissons d'eau douce dans les estuaires

En 2009, une première approche avait été proposée pour prédire la présence/absence de certaines guildes dans les estuaires français en utilisant des arbres de régression (Boët, 2009). Les arbres de régression, tout comme les modèles linéaires (LM ou GLM) sont effectivement des outils qui peuvent permettre de réaliser des prédictions, toutefois ils sont surtout conçus comme des outils explicatifs de données existantes. À l'inverse, certaines méthodes, notamment les arbres de régression boostés (BRT) sont des méthodes davantage destinées à la prédiction plutôt qu'à l'explication (De'ath, 2007 ; Elith *et al.*, 2008). Ces outils pourraient donc être particulièrement pertinents pour prédire un état de référence dans un milieu donné ; de telles prédictions sont nécessaires pour répondre pleinement à la Directive Cadre sur l'Eau.

Tout comme les arbres de régression simples de type (CART), les arbres de régression boostés présentent l'intérêt d'être beaucoup plus souples que les modèles linéaires simples (LM) ou généralisés (GLM) quant aux hypothèses de modélisation. En particulier, ils sont adaptés à des relations linéaires ou non, et permettent de prendre en compte des variables explicatives corrélées entre elles (De'ath, 2007 ; Elith *et al.*, 2008). Cette souplesse de modélisation est particulièrement intéressante dans notre cas, notamment pour des variables

difficiles à modéliser. C'est pourquoi nous avons décidé de tester la méthode des arbres de régression boostés pour prédire les probabilités de présence de poissons d'eau douce en fonction des niveaux de pression anthropique.

### 3.1. *Mise en œuvre*

#### 3.1.1. *Brève introduction aux arbres de régression boostés*

La construction d'un arbre de régression boosté repose en fait sur la construction itérative d'un nombre variable d'arbres de régression simple (voir le rapport de Boët (2009) pour une présentation de ces arbres) ; les prédictions de chacun de ces arbres sont ensuite combinées pour fournir une prédiction finale. Tout comme les arbres de régression simples, les BRT cherchent à prédire une variable à expliquer, en fonction d'un jeu de variables explicatives quantitatives ou qualitatives. La procédure de construction d'un BRT peut se résumer ainsi :

1. Construction d'un premier arbre de régression simple ajusté sur les données observées et calcul des résidus (différence entre la prédiction de l'arbre et l'observation) ;
2. Répéter m fois :
  - Ajustement d'un nouvel arbre de régression simple sur les résidus précédents ;
  - Calcul des prédictions une fois que le nouvel arbre est ajouté aux m-1 précédents ;
  - Calcul des nouveaux résidus ;
3. Calcul des prédictions finales.

Le fait que les BRT soient particulièrement performants pour la prédiction s'explique par le choix d'ajuster à chaque itération un nouvel arbre sur ce qui n'a pas été bien expliqué par les arbres précédents, c'est-à-dire les résidus. Afin d'éviter le surparamétrage, on choisit en général de tirer à chaque itération un sous-ensemble des observations, évitant que l'arbre ne se « spécialise » trop sur certaines données.

Afin de comprendre quelles variables expliquent la plus grande part de la variabilité de la variable à expliquer, il est possible de tracer des graphiques représentant l'influence relative de chacune des variables.

Le BRT étant constitué d'un nombre élevé d'arbres simples, il n'est en revanche pas possible de le représenter simplement comme un arbre de décision, comme c'était le cas pour les arbres de régression simples ou multiples.

#### 3.1.2. *Modélisation des probabilités de présence/absence de poissons d'eau douce dans les estuaires français*

Afin de tester la pertinence de cet outil, un BRT a été construit pour essayer de prédire la probabilité de présence/absence de poissons d'eau douce. Le choix a porté sur cette guilda car du fait de la forte proportion de zéro dans le jeu de données, cette guilda s'est en effet avérée très difficile à modéliser par des GLM. Pourtant, l'utilisation de cette métrique

serait intéressante afin de pouvoir mieux prendre en compte les assemblages en zone oligohaline dans l'indicateur poisson.

Pour cela, la présence/absence de poissons de cette guilda en estuaire a été modélisée comme une fonction de la zone haline (*sal\_class*), de la superficie de l'estuaire, de l'écorégion, de la saison, et des indices de pression de pollutions métalliques (*métaux\_acp*) et organiques (*organiques\_acp*). C'est-à-dire en reprenant les variables de l'analyse en composantes principales des données RNO de l'IFREMER (Courrat *et al.*, 2009 ; Delpech *et al.*, 2010a ; Girardin *et al.*, 2009).

Pour construire le BRT, nous avons suivi le tutoriel proposé par Elith *et al.* (2008). Différents réglages de complexité d'arbres (1, 2, 3, 5, 7, 10) et de taux d'apprentissage (0.1, 0.05, 0.01, 0.005, 0.001) ont été testés. Ainsi, un taux d'apprentissage de 0.01, qui règle l'évolution des poids des arbres à chaque itération, et une complexité d'arbres de 3, qui règle le nombre de nœuds de chaque arbre, se sont avérés être un bon compromis pour éviter le surparamétrage tout en gardant un nombre d'arbres à construire raisonnable.

Dans une approche parcimonieuse, l'examen de la déviance expliquée quand des facteurs explicatifs étaient retirés a amené à supprimer l'écorégion et la classe de superficie du modèle.

L'arbre ainsi construit a ensuite été utilisé pour prédire la probabilité de présence de poissons d'eau douce pour différentes combinaisons de facteurs environnementaux et d'échantillonnage (classe de salinité et saison) et différents niveaux de contamination organique et métallique.

### 3.2. Résultats

Pour examiner comment varie la probabilité de présence en fonction des variables explicatives, on trace des graphiques dits de dépendance partielle (Figure 1). Sans surprise, ces graphiques montrent que la zone haline est la variable qui semble avoir le plus d'effet (attention, les axes y n'ont pas tous la même échelle) : la probabilité de présence est maximale en zone oligohaline (classe 1). Le printemps semble légèrement plus favorable que l'automne. On voit aussi que des faibles niveaux de contamination sont plutôt plus favorables.

L'analyse de l'influence des 4 variables explicatives (Figure 2) confirme l'influence prédominante de la zone haline, et montre que les contaminations semblent avoir un effet non négligeable, plus important que la saison.

Le BRT a permis de tracer des cartes de prédictions pour chaque combinaison de saison et zone haline, à différents niveaux de contamination (Figure 3). On observe des probabilités nulles en secteur polyhalin, plutôt faibles en mésohalin et assez fortes en oligohalin, que ce soit au printemps ou à l'automne. Par contre, on observe que les probabilités de présence fortes sont essentiellement concentrées à des faibles niveaux de contamination. Ces probabilités pourraient donc éventuellement être utilisées comme prédiction de probabilité de présence dans un état de référence. Cependant, sans que l'on puisse vraiment l'expliquer à cette étape de l'analyse, des forts niveaux de contamination organique couplés à des faibles niveaux de contamination métallique permettent également d'obtenir une probabilité de présence assez forte.

*Figure 1. Dépendance partielle de la probabilité de présence en fonction des variables explicatives. Ces graphiques montrent l'effet moyen de chaque variable explicative sur la variable à expliquer.*

*Figure 2. Influence relative des variables explicatives.*

### *3.3. Discussion : avantages et limites de la méthode*

Cette méthode est donc intéressante car elle permet de répondre en partie à certaines limites des modèles linéaires simples (LM) et généralisés (GLM), notamment leur performance en prédiction, et leur souplesse dans les hypothèses. On voit dans l'exemple présenté précédemment que la présence de poissons d'eau douce a ainsi pu être prédite en fonction de différents facteurs naturels et anthropiques, alors que cela est très difficile avec les modèles statistiques usuels. Les BRT peuvent donc être un outil intéressant pour la modélisation d'un état de référence.

À contrario, la méthode présente malgré tout quelques limites qu'il faut mentionner. Tout d'abord, les capacités explicatives des BRT sont relativement limitées, ce qui n'est pas nécessairement gênant si on l'utilise en combinaison avec des modèles de type LM/GLM. Un inconvénient plus gênant réside dans les faibles capacités d'extrapolation des BRT. En effet, le domaine de prédiction se limite aux domaines représentés par le jeu de données. Pour reprendre l'exemple précédent, la probabilité de présence prédite à un niveau de pression nulle est bornée par les probabilités observées dans l'estuaire le moins soumis à pression. Si la superficie de l'estuaire avait été la variable la plus explicative, la probabilité de présence à un niveau de pression nulle dans un grand estuaire aurait été bornée par la probabilité dans le grand estuaire le moins soumis à pression.

En conclusion, les BRT sont des outils qui peuvent être intéressants pour le développement de l'indicateur poisson DCE, notamment pour approcher un état de référence. Cela étant, ils présentent certaines limites dont il faut avoir conscience, lesquelles risquent d'être problématiques quand les variables de pressions auront une influence beaucoup plus faibles que les variables d'environnement.



## 4. Classification des lagunes françaises à partir de l'assemblage ichtyofaunistique

Le but de cette étude était d'explorer la possibilité de réaliser une typologie des lagunes méditerranéennes à partir d'une liste d'espèces potentiellement indicatrices, proposée par Taverny *et al.* (2009). L'objectif était aussi de réaliser une première analyse visant à expliquer les différences obtenues sur les modèles pressions-impacts quand les trois lagunes corses échantillonnées en 2009 (Diana, Urbino, Palo) étaient intégrées au jeu de données (voir le rapport de l'action 16 de la convention de partenariat ONEMA-Cemagref 2010).

À cette fin, les proportions des captures représentées par trois familles abondamment représentées dans les lagunes françaises : atherinidae, gobiidae, mugilidae ont d'abord été calculées pour chaque opération de pêche du jeu de données. Puis, une analyse en composantes principales (ACP) sur la matrice des corrélations de ces proportions par opération de pêche (Figure 4), suivie d'une classification hiérarchique ascendante (critère de Ward, distance de Bray-Curtis), ont ensuite été réalisées pour essayer de faire une typologie des profils d'opérations de pêche (Figure 5).

L'analyse en composantes principales montre une forte opposition des trois variables. On a donc en général des opérations de pêches caractérisées par une forte proportion d'une famille, mais assez peu de plusieurs familles. L'axe 1 semble représenter un gradient des proportions d'athérines, alors que l'axe 2 semble plus opposer les opérations à mugilidae (Or et Méjean) par rapport aux opérations à gobiidae (Vaccarès).

a)



b)

*Figure 4. Analyse en composantes principales des proportions des captures représentées par trois grandes familles de poissons dans la capture totale. a) position des opérations de pêche et du centroïde de chaque masse d'eau dans le plan factoriel défini par les deux premières composantes. b) cercle de corrélations des variables (*prop\_ath* : proportion d'athérines, *prop\_mugilidae* : proportion de mugilidae et *prop\_gobiidae* : proportion de gobiidae) avec les deux premiers axes de l'ACP.*

**Figure 5. Arbre de classification des proportions des trois grandes familles pour chaque opération de pêche.**

À partir de la classification hiérarchique ascendante (Figure 5), il semble que définir 4 types d'opérations de pêche soit un bon compromis entre gain de précision (hauteur de l'arbre) et utilisation pratique (nombre de classes pas trop élevé). Les proportions moyennes de chacune des familles dans les 4 classes ainsi définies ont été calculées pour caractériser les types d'opérations de pêche (Tableau I).

**Tableau I. Caractérisation des trois types d'opérations de pêches.**

Type	prop_at h	prop_mugilidae	prop_gobiidae	Nom attribué au type
<b>1</b>	<b>0.55</b>	<b>0.03</b>	<b>0.08</b>	<b>Opérations à athérines et autres espèces</b>
<b>2</b>	<b>0.91</b>	<b>0.00</b>	<b>0.02</b>	<b>Opérations à athérines uniquement</b>
<b>3</b>	<b>0.06</b>	<b>0.02</b>	<b>0.03</b>	<b>Autres opérations</b>
<b>4</b>	<b>0.15</b>	<b>0.18</b>	<b>0.43</b>	<b>Opérations aux trois familles</b>

Deux types, « opérations à athérines et autres espèces » et « opérations à athérines uniquement », sont caractérisés par de fortes proportions d'athérines, le premier étant quasi uniquement composé d'athérines, alors que dans les opérations de type 2, d'autres espèces sont aussi capturées, mais assez peu de mugilidae et de gobiidae. Le type « opération aux trois familles » se caractérise par une forte proportion de gobiidae, et des quantités non négligeables des 2 autres familles, mugilidae et gobiidae. Enfin, le dernier type, « autres opérations », se caractérise par des captures faibles des trois familles.

Afin d'examiner si ces types d'opérations de pêche correspondaient à des types de lagunes, les proportions des opérations de pêche appartenant à chacun des types ont été calculées pour chaque lagune (Tableau II).

**Tableau II. Proportions d'opérations de pêche de chaque type pour chaque lagune.**

Lagunes	Opérations			
	À athérines et autres	À athérines uniquement	Autres	Aux trois familles
<b>Bages-Sigean</b>	<b>0.33</b>	<b>0.15</b>	<b>0.48</b>	<b>0.04</b>
<b>Berre</b>	<b>0.10</b>	<b>0.17</b>	<b>0.62</b>	<b>0.10</b>
<b>Biguglia</b>	<b>0.38</b>	<b>0.38</b>	<b>0.04</b>	<b>0.21</b>
<b>Vaccarès</b>	<b>0.07</b>	<b>0.00</b>	<b>0.14</b>	<b>0.79</b>
<b>Diana</b>	<b>0.06</b>	<b>0.83</b>	<b>0.06</b>	<b>0.06</b>
<b>Grand Bagnas</b>	<b>0.35</b>	<b>0.45</b>	<b>0.03</b>	<b>0.16</b>
<b>La Palme</b>	<b>0.06</b>	<b>0.06</b>	<b>0.56</b>	<b>0.31</b>
<b>Méjean</b>	<b>0.00</b>	<b>0.00</b>	<b>0.69</b>	<b>0.31</b>
<b>Or</b>	<b>0.13</b>	<b>0.00</b>	<b>0.27</b>	<b>0.60</b>
<b>Palo</b>	<b>0.07</b>	<b>0.93</b>	<b>0.00</b>	<b>0.00</b>
<b>Prévost</b>	<b>0.06</b>	<b>0.06</b>	<b>0.88</b>	<b>0.00</b>
<b>Salses-Leucate</b>	<b>0.19</b>	<b>0.65</b>	<b>0.10</b>	<b>0.06</b>
<b>Thau</b>	<b>0.20</b>	<b>0.61</b>	<b>0.07</b>	<b>0.11</b>
<b>Urbino</b>	<b>0.10</b>	<b>0.90</b>	<b>0.00</b>	<b>0.00</b>

On voit ainsi que certaines lagunes, notamment Diana, Urbino et Palo, sont très marquées par des opérations à athérines, tandis que d'autres comme Vaccarès sont marquées par des opérations à gobiidae. Afin de tester la possibilité de faire une typologie des lagunes à partir de ces opérations de pêche, une classification hiérarchique ascendante (distance de Bray-Curtis, critère de Ward) a été réalisée sur ce tableau II (Figure 6).

L'examen de l'arbre de cette classification (Figure 6) montre que considérer 4 types de lagunes est un bon compromis entre précision (hauteur de l'arbre) et utilisation pratique (ne pas multiplier le nombre de types). Sur cette base, chacun des 4 types de lagunes est donc caractérisé en calculant la proportion moyenne représentée par chacun des 4 types d'opération de pêche (Tableau III).

**Figure 6. Arbre de classification des lagunes en fonction de la proportion représentée par chacun des types d'opérations de pêche.**

Cette classification semble correspondre à un gradient sur l'axe 1 de l'ACP (Figure 4). Les types A et B sont caractérisés par une forte proportion d'opérations à athérines. Ce sont des lagunes très productives, où l'on retrouve d'ailleurs en général de la production ostréicole (Diana, Urbino, Salses-Leucate, Thau). Le type A est encore davantage dominé par les athérines, tandis que le groupe B présente un peu plus de diversité, peut-être en raison d'une plus grande diversité d'habitats, liée à la taille (Salses-Leucate, Thau, Biguglia), à une meilleure connexion à la mer (Thau, Salses-Leucate) ou à une plus forte influence des eaux douces (Biguglia, Grand-Bagnas). Le type C présente une majorité d'opérations aux trois familles, marquées cependant par une forte proportion de gobiidae. On peut noter que les deux lagunes de ce type sont les deux seules lagunes à ne plus présenter de graux naturels. Enfin, le dernier type n'est caractérisé par aucune forte dominance particulière.

Ce travail de typologie est intéressant car il démontre qu'une typologie fondée sur les assemblages de poissons est possible pour les lagunes méditerranéennes. Par contre, en l'état, cette dernière ne permet pas de savoir si les catégories obtenues sont liées à des caractéristiques naturelles et/ou anthropiques. Poursuivre ce travail en essayant d'expliquer ces types par des variables de pressions et/ou d'environnement serait donc particulièrement intéressant. Il serait en particulier très pertinent de croiser cette typologie avec les seuils de qualité physico-chimiques listés par Taverny *et al.* (2009). Il est cependant déjà intéressant de souligner que des lagunes considérées comme relativement peu impactées par l'indice global de pression (Diana, La Palme, Vaccarès, Grand Bagnas) se retrouvent dans des types différents. Il est aussi notable que selon cette classification, les trois lagunes corses, échantillonnées en 2009, constituent un type relativement à part.

**Tableau III. Caractéristiques des 4 types et appartenance des lagunes à chacun de ces types.**

Types	Opérations				Lagunes
	À athérines et autres	À athérines uniquement	Autres	Aux trois familles	
<b>A</b>	<b>0.07</b>	<b>0.89</b>	<b>0.02</b>	<b>0.02</b>	<b>Diana, Urbino, Palo</b>
<b>B</b>	<b>0.28</b>	<b>0.52</b>	<b>0.06</b>	<b>0.14</b>	<b>Biguglia, Grand Bagnas, Salses-Leucate, Thau</b>
<b>C</b>	<b>0.10</b>	<b>0.00</b>	<b>0.20</b>	<b>0.69</b>	<b>Vaccarès, Or</b>
<b>D</b>	<b>0.11</b>	<b>0.09</b>	<b>0.65</b>	<b>0.15</b>	<b>Bages-Sigean, Berre, Prévost, Méjean, La</b>

Améliorer la typologie des lagunes devrait à terme permettre de mieux comprendre les facteurs environnementaux et/ou anthropiques expliquant celle-ci, et donc nécessaires à prendre en compte dans les modèles pression-impacts. Cela devrait également permettre de faciliter la description d'un état de référence, en particulier si la description de certains types s'explique essentiellement par des niveaux de pression plus faibles. Les métriques de l'assemblage de poissons discriminant ce type par rapport aux autres types seraient alors des métriques potentiellement intéressantes pour l'indicateur multi-métriques.

## 5. Conclusion

Trois approches à partir d'analyses multivariées ont été entreprises cette année pour tenter de caractériser les états de référence des masses d'eau de transition. Le développement de ce travail n'a cependant pas pu être mené comme programmé initialement, en raison de l'impossibilité à trouver un stagiaire pour travailler sur cette thématique comme prévu.

Parmi ces trois approches, les arbres de régressions multiples, qui avaient donné de bons résultats l'an passé pour les lagunes, n'ont pas permis de modélisation de l'état de référence en estuaires.

En revanche, les arbres de régression « boostés » apparaissent comme une voie intéressante à poursuivre. Ils souffrent toutefois de certaines lacunes en extrapolation qui risquent d'être gênantes dans certaines situations pour prédire un état de référence.

Enfin la troisième approche par classification hiérarchique a permis d'établir une typologie des lagunes en fonction de familles d'espèces, présentées comme potentiellement pertinentes pour qualifier l'état des masses d'eau de transition (Taverny *et al.*, 2009). Cette typologie devra être approfondie à l'avenir. Il en ressort malgré tout une certaine spécificité des lagunes corses, spécificité qui ne ressort pas dans les indices de pressions, ni dans les variables liées à l'échantillonnage et à l'environnement, prises en compte actuellement dans les modèles pression-impacts.

## 6. Références

- Boët, P. 2009. Conditions de référence pour le compartiment "Poissons" des eaux de transition. Convention de partenariat ONEMA-Cemagref 2008 : 34pp.
- Courrat, A., Lobry, J., Nicolas, D., Laffargue, P., Amara, R., Lepage, M., Girardin, M., Le Pape, O. 2009. Anthropogenic disturbance on nursery function of estuarine areas for marine species. *Estuarine, Coastal and Shelf Science*, 81 : 179-190.
- De'ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology*, 88 : 243-251.
- Delpech, C., Courrat, A., Pasquaud, S., Lobry, J., Le Pape, O., Nicolas, D., Boët, P., Girardin, M., Lepage, M. 2010a. Development of a fish-based index to assess the ecological quality of transitional waters: The case of French estuaries. *Mar Pollut Bull*, 60 : 908-918.
- Delpech, C., Drouineau, H., et Lepage, M. 2010b. Amélioration des performances de la méthode ELFI pour l'évaluation de la qualité des eaux de transition. Convention de partenariat ONEMA-Cemagref 2009 : 95pp.
- Drouineau, H., Delpech, C., et Lepage, M. 2010. Réponses des assemblages ichthyofaunistiques à l'eutrophisation et proposition de méthode pour la définition d'un état de référence. Convention de partenariat ONEMA-Cemagref 2009 : 12p.
- Elith, J., Leathwick, J., et Hastie, T. 2008. A working guide to boosted regression trees. *J Anim Ecol*, 77 : 802-813.
- Girardin, M., Lepage, M., Amara, R., Boët, P., Courrat, A., Delpech, C., Durozoi, B. et al. 2009. Développement d'un indicateur poisson pour les eaux de transition. Rapport scientifique de fin de contrat : 35p.
- Taverny, C., Élie, P., et Boët, P. 2009. La vie piscicole dans les masses d'eau de transition : proposition d'une grille de qualité pour la température, l'oxygène dissous, la salinité et la transparence. Étude CEMAGREF Bordeaux n° 131 : 44p + annexes.

Onema  
Hall C – Le Nadar  
5 square Félix Nadar  
94300 Vincennes  
01 45 14 36 00  
[www.onema.fr](http://www.onema.fr)

*Cemagref*  
*Parc de Tourvoie*  
BP 44,  
92163 Antony cedex  
01 40 96 61 21  
[www.cemagref.fr](http://www.cemagref.fr)