AGROCAMPUS OUEST

☐ **CFR Angers**

☑ **CFR Rennes**

Année universitaire : 2015 - 2016 Mention : Biodiversité Ecologie Environnement Spécialité : Ressources Aquatiques et Exploitation Durable, dominante Ressources et Ecosystèmes Aquatiques	**Mémoire de fin d'études** ☐ d'Ingénieur de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage ☐ de Master de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage ☑ d'un autre établissement (étudiant arrivé en M2)

Using otolith microchemistry within Bayesian reallocation models to explore the Allis shad (*Alosa alosa*) metapopulation functioning

Marine RANDON

Soutenu à Rennes le 15/09/2016

Devant le jury composé de :

Président et enseignant référent : Etienne Rivot, Agrocampus Ouest

Maîtres de stage : Hilaire DROUINEAU & Françoise Daverat, Irstea de Bordeaux

Enseignant: Elodie Reveillac, Agrocampus Ouest

Autres membres: Marie NEVOUX, INRA de Rennes

Image de couverture disponible sur https://fr.pinterest.com/pin/41025046579264954/

Fiche de confidentialité et de diffusion du mémoire

Confidentialité

☑ Non ☐ Oui si oui : ☐ 1 an ☐ 5 ans ☐ 10 ans

Pendant toute la durée de confidentialité, aucune diffusion du mémoire n'est possible [1].

Date et signature du maître de stage [2] : 15/09/16

A la fin de la période de confidentialité, sa diffusion est soumise aux règles ci-dessous (droits d'auteur et autorisation de diffusion par l'enseignant à renseigner).

Droits d'auteur

L'auteur[3] **Randon Marine**
autorise la diffusion de son travail (immédiatement ou à la fin de la période de confidentialité)

☑ Oui ☐ Non

Si oui, il autorise

☐ la diffusion papier du mémoire uniquement(4)

☐ la diffusion papier du mémoire et la diffusion électronique du résumé

☑ la diffusion papier et électronique du mémoire (joindre dans ce cas la fiche de conformité du mémoire numérique et le contrat de diffusion)

(Facultatif) ☐ accepte de placer son mémoire sous licence Creative commons CC-By-Nc-Nd (voir Guide du mémoire Chap 1.4 page 6)

Date et signature de l'auteur :

Autorisation de diffusion par le responsable de spécialisation ou son représentant

L'enseignant juge le mémoire de qualité suffisante pour être diffusé (immédiatement ou à la fin de la période de confidentialité)

☒ Oui ☐ Non

Si non, seul le titre du mémoire apparaîtra dans les bases de données.
Si oui, il autorise

☐ la diffusion papier du mémoire uniquement(4)

☐ la diffusion papier du mémoire et la diffusion électronique du résumé

☒ la diffusion papier et électronique du mémoire

Date et signature de l'enseignant :

Didier GASCUEL
Professeur Directeur du Pôle halieutique
AGROCAMPUS OUEST

(1) L'administration, les enseignants et les différents services de documentation d'AGROCAMPUS OUEST s'engagent à respecter cette confidentialité.
(2) Signature et cachet de l'organisme
(3).Auteur = étudiant qui réalise son mémoire de fin d'études
(4) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option)) sera signalée dans les bases de données documentaires sans le résumé

REMERCIEMENTS

Avant de remercier toutes les personnes qui, de près ou de loin, ont participé à rendre ce stage agréable et stimulant, je souhaite exprimer toute ma gratitude à mes futurs encadrants de thèse, Elodie, Olivier et Etienne qui m'ont accordé leur confiance. J'espère que le travail qui suit sera à la hauteur de vos espérances et marquera le départ d'une belle collaboration.

Je souhaite maintenant remercier mes deux encadrants Hilaire et Françoise pour m'avoir permis de réaliser ce stage de fin d'étude à leurs côtés. Un grand merci pour votre soutient, votre disponibilité et vos nombreux conseils. Hilaire, merci pour ta patience et la pertinence de tes remarques qui m'ont permis d'avancer sereinement tout au long du stage. Un immense merci également pour tes conseils qui m'ont aidé à décrocher ma thèse. Françoise, merci pour nos nombreuses et enrichissantes discussions « otolithes » et surtout merci pour ta bonne humeur au quotidien.

Je remercie également Philippe Jatteau pour m'avoir transmis les données d'indices d'abondances et pour avoir pris le temps d'en discuter avec moi.

Merci à Christophe Pecheyran et Gilles Bareille de l'université de Pau et des Pays de l'Adour d'avoir gentiment accepté d'analyser mes échantillons.

Plus généralement, je remercie toute l'unité EABX pour m'avoir accueilli et fourni des conditions de travail agréables. Et bien sûr je n'oublie pas tous les copains stagiaires sans qui ce stage n'aurait vraiment pas été aussi sympathique. Je remercie donc (et j'essaie de n'oublier personne) Alexandre et Virginie de l'Agro, mais aussi Marion, Audrey, Jennifer, Armand, Elise, Julie, Maxime, Johanna, Aurélie, Tom et Coralie. Marion, merci pour t'être intéressée à mon travail et surtout merci pour ton rire communicatif et ta joie de vivre. Audrey, Jennifer, Alex, Virginie, Maxime et Julie, merci pour votre énergie et pour tous ces week-ends qui m'ont donné un air de vacances. Armand, merci pour tes merveilleuses pâtisseries qui nous ont tous enchanté.

Merci à tous mes amis, et surtout merci Alice pour ton amitié et ton soutien depuis toutes ces années.

Enfin je remercie ma famille et plus particulièrement mes parents qui me supportent et m'encouragent dans tous mes projets.

RESUME ETENDU EN FRANÇAIS

Exploration du fonctionnement de la métapopulation de grande alose (*Alosa alosa*) au travers de l'intégration de données de microchimie des otolithes dans des modèles de réassignation Bayésiens.

Contexte

La grande alose (*Alosa alosa*) est un poisson amphihalin anadrome se reproduisant en rivière après une période de maturation en mer de quelques années. Actuellement, cette espèce se répartie des côtes du Portugal au nord de la France. Depuis le milieu du XXème siècle, une contraction de l'aire de répartition s'est accompagnée d'un déclin généralisé des populations. C'est le cas notamment de la population de Garonne, jusqu'alors considérée comme la plus importante d'Europe. A l'heure actuelle, même si la pollution, les obstacles à la migration et la surpêche sont perçues comme des causes aggravantes, l'effondrement de cette population reste à ce jour inexpliquée. Malgré un moratoire sur la pêche mis en place en 2008, la situation de la grande alose en Garonne reste inchangée.

Actuellement, nous disposons de peu de connaissances sur l'écologie de la grande alose et en particulier sur sa tendance à retourner se reproduire dans sa rivière natale, ou au contraire à errer entre rivières. Cette mauvaise connaissance fait qu'à l'heure actuelle, il est difficile de savoir si les aloses de chaque bassin constituent des populations discrètes ou au contraire s'il existe des échanges entre bassins et donc un fonctionnement de type métapopulation. Il est donc crucial pour la gestion de parvenir à évaluer le taux de retour des populations sur leurs rivières natales mais également de qualifier la direction des échanges et de les quantifier. En 2015, Martin *et al.* ont réassigné des adultes de grandes aloses à leurs rivières natales à partir de la microchimie des otolithes pour étudier les capacités de dispersion de l'espèce à l'échelle de son aire de répartition. Il a ainsi été montré que la grande alose est relativement fidèle à son bassin versant d'origine, sans pour autant exclure des échanges entre bassins versants. En revanche, cette étude a soulevé deux principales limites : (1) compte tenu de l'indisponibilité des données d'indices d'abondance, les flux d'échanges entre bassins versants n'ont pas pu être estimés et (2) le modèle de réassignation reposait sur l'hypothèse d'une exhaustivité des sources de référence échantillonnées.

Objectif

La présente étude propose donc de poursuivre ce travail en essayant de palier à ces deux limites. Dans un premier temps nous proposons de construire des modèles alternatifs permettant de s'affranchir de l'hypothèse d'exclusivité, et de choisir parmi ces modèles le plus crédible. Dans un second temps, les sorties du modèle sélectionné sont couplées à des estimations d'indices d'abondance de reproducteurs afin d'étudier les flux entre les bassins versants de naissance

estimés par le modèle, et les bassins versants de reproduction, correspondant au lieu de capture des géniteurs au cours de leur migration de reproduction en rivière. Ce modèle d'échanges entre bassins devrait permettre de confirmer ou d'infirmer l'hypothèse d'une métapopulation. Enfin, cette étude devrait permettre de proposer une priorisation des bassins versants à protéger et à restaurer à l'échelle de l'aire de répartition.

Matériels et méthodes

Les otolithes sont des pièces calcifiées de l'oreille interne du poisson qui grandissent par accumulation de composés présents dans l'eau, depuis la naissance du poisson jusqu'à sa mort. En ciblant une ablation laser sur une zone restreinte de l'otolithe proche du noyau, nous pouvons cibler la phase juvénile, et donc prédire la signature microchimique de la rivière natale à partir de celle de l'otolithe.

Construction et choix du meilleur modèle

Dans un premier temps, une remise à jour du modèle développé par Martin *et al.* (2015) est effectuée. Pour cela, nous couplons des données de microchimie des otolithes d'adultes de grandes aloses (N = 615) capturées pendant leur migration de reproduction sur 15 rivières, avec des données microchimiques de référence. Ces données de référence sont :

- Des données microchimiques d'eau échantillonnée sur 17 rivières considérées comme des rivières de reproduction
- Des données de microchimie des otolithes de juvéniles (N = 44) capturés sur 5 rivières. Cette donnée est une référence puisque les juvéniles capturés dans une rivière y sont nés.

Par inférence Bayésienne, les adultes sont réassignés de manière probabiliste à leur rivière natale. Ce modèle autorisant la réassignation des poissons uniquement dans les rivières de référence, nous proposons de construire un second modèle Bayésien ne fixant pas le nombre de sources natales. Ce modèle à mélange infini de Gaussiennes permet à la fois d'estimer le nombre de sources natales et d'estimer la probabilité d'appartenance des poissons à chacune de ces sources. Etant donné que ce modèle n'intègre pas de données de référence, aucun « lien » entre les sources prédites et des rivières de reproduction ne peut être établi. Nous construisons donc un dernier modèle Bayésien, nommé modèle hybride, qui permet de cumuler les avantages des deux modèles précédents en estimant le nombre de source et associant, quand cela est possible, les poissons aux rivières de référence. Dans le cas contraire, les poissons sont réassignés dans des sources extérieures, donc non connues.

Compte tenu du fait que notre jeu de référence non exhaustif ne permet pas de valider les modèles, nous proposons de comparer leur pertinence. Pour cela nous utilisons plusieurs mesures de cohérence :

- Des critères statistiques : une mesure de déviance et de la convergence des paramètres

- Des indicateurs de cohérence : la probabilité maximale de réallocation, l'entropie de Shannon et le nombre de sources estimées par poisson
- Des critères écologiques : la comparaison des sources prédites par les différents modèles

La comparaison des modèles nous permet ainsi de choisir le modèle le plus pertinent avant de développer un modèle d'échanges de flux entre bassins versants.

Modèle d'échanges entre bassins versants

Ce modèle repose sur une estimation des flux entre les rivières donneuses (i.e. les rivières natales prédites par le modèle sélectionné) et réceptrices (i.e. les rivières de reproduction où ont été capturés les adultes). En multipliant les indices d'abondances de reproducteurs estimés par bassin versant avec les probabilités de naissance estimées par le modèle, nous estimons les flux d'échanges. Il est ainsi possible d'estimer un taux de retour (et donc d'errance) par sous-population et d'établir quelles rivières produisent le plus de reproducteurs. Nous pouvons de cette manière distinguer les rivières « source », qui produisent plus de géniteurs qu'elles n'en reçoivent, et les rivières « puits », qui reçoivent plus de géniteurs qu'elles n'en produisent. Cela permet d'établir une priorisation des bassins versants à protéger et restaurer.

Résultats

Le modèle à mélange infini de Gaussiennes a montré de gros problèmes de convergence et n'a donc pas été conservé pour la suite des analyses. Le modèle hybride a permis de mettre en évidence des poissons aux signatures microchimiques atypiques ne correspondant pas aux sources de références. En revanche, le modèle hybride présentait une convergence et une pertinence écologique plus faible que le premier modèle. De plus, compte tenu du fait que 99% des adultes ont été réassignés dans des sources connues, l'intérêt du modèle hybride reste limité dans notre cas d'étude. Nous avons donc sélectionné le modèle Bayésien fixant le nombre de sources pour étudier le fonctionnement de la métapopulation de grande alose.

Nous avons montré que cette espèce réalise principalement un retour sur rivière natale, et dans une moindre mesure, des échanges significatifs à l'échelle du bassin versant. Un isolement par la distance a été mis en évidence à partir de 250 km entre les rivières natales et de reproduction.

Conclusion

Cette étude a permis de montrer qu'il serait préférable d'établir un programme de suivi de la grande alose sur chaque sous-population dans l'aire de répartition plutôt que d'avoir recours au modèle hybride. Ce dernier pourrait être plus adapté à des études disposant de peu de données de référence. Le modèle d'échanges a permis de mettre en évidence que les sources ainsi que les sous-populations isolées sont à protéger en priorité. Le fonctionnement en métapopulation mis en évidence dans cette étude pousse à la mise en place d'un suivi de sa dynamique spatio-temporelle en mettant en place un programme de suivi des reproducteurs et d'échantillonnage exhaustif sur l'ensemble de l'aire de répartition. Cela permettrait en outre de tester l'hypothèse d'un replacement des grandes aloses vers le nord de l'aire de répartition en lien avec le changement climatique.

CONTENT

LIST OF APPENDIX

LIST OF ILLUSTRATIONS

Figures

Tables

1. Introduction

In the context of global change, understanding fish species distribution and ecology is an important challenge to improve stocks management and ensure populations sustainability (Jorgensen *et al.* 2007; Hamann & Kennedy 2012). By, using more or less distant habitats, migratory species are especially at risk (Wilcove 2008), as such diadromous fishes which complete their life-cycle by doing seasonal or life-stage migrations between the sea and the freshwater habitats (McDowall 2008). During the last century, most temperate diadromous fish species underwent dramatic declines within the North Atlantic area mostly caused by anthropogenic pressures (Limburg & Waldman 2009). Among diadromous fishes, anadromous species spend most of their life cycle at sea before migrating to freshwater habitats to reproduce (Gross et al. 1988; McDowall, 2008). The European Allis shad (*Alosa alosa*), an anadromous Clupeidae, remains as juvenile in freshwater for few weeks before performing a seaward migration as YOY (i.e. Young Of the Year) and spends a several years long growth phase at sea (Lochet 2006). Mature individuals return to freshwater to reproduce aged between 3 and 7 years (Lochet 2006). Reproduction occurs in the higher middle watercourse of rivers in spring and summer (Baglinière *et al.* 2003). Most Allis shads are semelparous (i.e. individuals generally die after reproduction) but a fraction of the population is iteroparous (Lochet 2006).

The distribution area of Allis shad has significantly decreased, from Norway to Morocco in the middle of the 20th century, to an actual range from France to Portugal (Baglinière *et al.* 2003). Allis shad is spread in populations (i.e. groups of individuals which reproduce in the same rivers) through the distribution range. Until the end of the 20th century, the Gironde population was considered as the most important in Europe (Elie & Baglinière 2000; Castelnaud *et al.* 2001). However, at the beginning of the 2000s, the Gironde population collapsed (Rougier *et al.* 2012). Despite a drastic fishing ban measure (i.e. a total moratorium enforced in 2008; COmité de GEstion des Poisson MIgrateurs), the stock has not recovered yet. In parallel, from the mid-twentieth century, a decrease in the number of spawners was observed in the Minho river in Portugal (Mota *et al.* 2015). Water pollution, habitat loss, obstacles to migration (such as dams) and overfishing have been suggested as possible causes of shad populations decline (Jonsson *et al.* 1999; De Groot 2002; Limburg & Waldman 2009). To analyze the collapse and decline of populations, a deeper knowledge of the whole dynamic of Allis shad populations is necessary to understand whether this species form discrete independent populations or have significant exchanges of individuals between sub-populations, resulting in a metapopulation (Kritzer & Sale 2004). Here, a metapopulation stands for a group of sub-populations inter-connected by the dispersal of individuals (Young 1999; Kritzer & Sale 2004).

In this context, whether Allis shad display a homing or a straying behavior is an important question. The homing is defined as the return of adults to their natal stream (McDowall 2008). It occurs at local scale (e.g. natal site) or at larger scale (e.g. river or watershed) (Stewart et al. 2003; Quinn et al. 2012; Hamann and Kennedy 2012). Conversely, the straying behavior corresponds to a spawning migration to non-natal site (Quinn 1993; Keefer & Caudill 2014). The straying and the homing behaviors are two life-history traits in equilibrium in migratory fish populations (Quinn 1984). Homing is probably favored compared to the straying behavior in rivers presenting stables annual characteristics and high-quality habitats because it potentially increases the survival of juveniles through local adaptation (Quinn 1984; Hendry *et al.* 2004). However, the straying behavior could be favored in unstable river, allowing the maintenance of genetic diversity, the ability for adaptation to unpredictable environments and thus the optimization of the individual fitness (Kerr & Secor 2012; Keefer & Caudill 2014). Therefore the estimation of the homing/straying proportion in a population is an important concern in order to enhance the understanding of the adaptive potential of a population across

the distribution area (Keefer & Caudill 2014). It is also important to understand whether populations display a strict homing, which would induce no exchanges between populations, or perform straying, which would favor a metapopulation functioning. The homing behavior is well described for salmon and trout species (Keefer & Caudill 2014). However, very few studies have focused on the homing behavior of Allis shad. In particular, little is known about the proportion of straying individuals in the European catchments. Nevertheless, several methods as genetic (Alexandrino & Boisneau 2000; Alexandrino *et al.* 2006) or morphological analysis (Sabatié *et al.* 2000) have already been tested in order to discriminate Allis shad populations structure without bringing clear results. From the end of the twentieth century, the otolith microchemistry has shown up its efficiency to examine fish migration, natal origin and populations connectivity (Kennedy *et al.* 2002; Daverat *et al.* 2012; Rooker *et al.* 2016).

More specifically, otoliths have been widely used to investigate the homing and straying behaviors. Otoliths of teleost fish are small calcareous concretions located in the inner ear which are involved in the fish balance (Watabe *et al.* 1982). The otolith grows during the entire life of fish by continuous accretion of metabolically inert elements originating from the ambient water (Campana 1999). One of the most important properties of this calcified structure is the continuous growth all along the life without resorption. Consequently, otoliths could be seen as the "black-box" of fishes, and thus these structures are used to estimate the age and the growth rate of fish (Campana 2001), discriminate stocks (Thresher 1999; Campana et al. 2000), determine migration pathways (Walther & Limburg 2012) and reconstruct environmental history (Elsdon & Gillanders 2002). As the otolith incorporates elements from the surrounding water, coupling microchemistry and micro-increments analyses is used to reconstruct the habitat use of the fish from birth (i.e. the core of the otolith) to death (i.e. the edge of the otolith) (Bath *et al.* 2000; Walther & Thorrold 2006). Among elements, the otolith elemental concentrations in Strontium (Sr/Ca) and Barium (Ba/Ca) are mostly influenced by ambient water composition, making them good tracers to reconstruct migration pathways and especially ecological transition between freshwater and marine compartments (Kennedy *et al.* 2000). Bath *et al.* (2000) found a linear relation between otolith and water concentration of Ba and Sr, allowing the estimation of partition coefficients (D). Besides, the isotopic ratio of Strontium ($^{87}Sr/^{86}Sr$) is known to be a powerful fish marker of natal origin, especially for species which carry out their juvenile stage in freshwater (Kennedy *et al.* 2000). This tracer is known to be stable over the years and to reflect the geochemistry of the stream, allowing the discrimination of rivers (Kennedy *et al.* 2000; Walther & Thorrold 2006). Contrary to the Ba/Ca and Sr/Ca ratios, the isotopic ratio of Strontium is not fractionated between the water and the otolith compartments (Kennedy et al. 2000; Blum et al. 2000; Pouilly *et al.* 2014). Thus, the $^{87}Sr/^{86}Sr$ ratio in the otolith is similar to the $^{87}Sr/^{86}Sr$ ratio in the water. Therefore, the combination of the Sr/Ca, Ba/Ca and the $^{87}Sr/^{86}Sr$ ratios appears as a relevant way to discriminate freshwater habitats, and thus investigate the natal origin of fishes.

Because of technical advances, the otolith microchemistry approach currently figures as one of the most relevant methods to investigate Allis shad natal origin (Tomas et al. 2005; Martin et al. 2015). In 2005, Tomas *et al.* examined the structure of Garonne – Dordogne populations using otolith microchemistry analysis, in order to test whether Dordogne River was producing spawners that would reproduce in the Garonne River. Results suggested that Dordogne River acted as a source of spawners in the whole watershed. However, the method used was not precise enough to draw final conclusions. Then in 2014, technical and methodological progresses enable the identification of natal origin of spawners from 15 rivers along the Atlantic coast (Martin *et al.* 2015). In this study, a large data of otoliths and water microchemistry available throughout the distribution range of the species was used (from the Vire River in the North of France to the Mondego River in the South of Portugal). Within a Bayesian model, the

reallocation of natal origin was performed using otolith microchemistry of adults sampled during the spawning migration in rivers across the range. The reallocation was driven by multiple baselines (i.e. reference data sets): the water microchemistry of several reproduction rivers across the range and the otolith microchemistry of juveniles. The use of a Bayesian model is particularly relevant in the case of the analysis of samples from mixed origins because it enables different sources of data to be combined in a model and provides probability distribution of parameters including several sources of uncertainty (Pella & Masuda 2006; Munch & Clarke 2008; Pflugeisen & Calder 2013). Those methods have been performed on otolith microchemistry data in a wide number of studies (Munch & Clarke 2008) and especially in order to cluster individuals to their natal site origin (White *et al.* 2008; Dawson & Belkhir 2009; Neubauer *et al.* 2013). By clustering individuals into groups of common natal origin and comparing these with catch river, Martin *et al.* (2015) found that Allis shad exhibits a high level of natal river fidelity and that straying could occur mostly at a river scale or short distance scale (20 – 100 km), but also at longer distances. Indeed, some individuals were presumed to have traveled ultra-long distance between the natal and the reproduction river. In their results, Martin *et al.* (2015) focused on fishes that were reallocated with a high level of credibility (greater than 80%). They considered that, natal rivers remained too uncertain for other fishes. These uncertainties can be due to many factors. First, fishes could be originated from rivers out of the dataset (i.e. non-sampled rivers). Besides, as water samplings were incomplete (all spawning grounds were not sampled in each river), fishes could also be originated from spawning grounds out of the water baseline. Finally, similarity of signatures between rivers could also generate reallocation uncertainty.

Herein, the metapopulation functioning of *Alosa alosa* was investigated using the otolith microchemistry within a Bayesian hierarchical model of reallocation. We first updated the Bayesian model used by Martin *et al.* (2015) with a larger data set. Then, an alternative hierarchical Bayesian model without baselines and without fixing the number of sources was tested, method known as "Infinite Mixture Model" (White *et al.* 2008; Neubauer *et al.* 2013). The clustering was based on the similarity between otolith microchemistry of the adults without reference to any water and juvenile baselines. Finally, an intermediate model was performed by combining the first two approaches. In this hybrid model, fishes were reallocated in the rivers of the baseline or in extra-sources when needed. The models were then compared and used to analyze the metapopulation functioning of Allis shad on the Atlantic-wide scale. Using a step by step construction of Bayesian hierarchical models, the following questions were investigated:

(1) Among the three Bayesian models (1 – simple updates of model from Martin *et al.* (2015), 2 – Infinite Mixture Model, 3 – hybrid approach), which one is, statistically and ecologically, the most appropriate to reallocate adults Allis shad to their natal river?
(2) By coupling the outputs of the best Bayesian model with abundance indicators of spawners, can we improve the understanding of the metapopulation functioning in terms of flux, homing and straying rate, "source" and "sink" rivers and isolation by distance between natal and reproduction rivers?
(3) What are the implications for the conservation of Allis shads through the distribution range?

2. Materials and methods

2.1. Sampling

2.1.1. Water

Seventeen rivers in France and Portugal were sampled in order to analyze water microchemistry (Figure 1). These rivers are considered to be major spawning catchments throughout the range of Allis shad (Aprahamian & Environment Agency 2003).

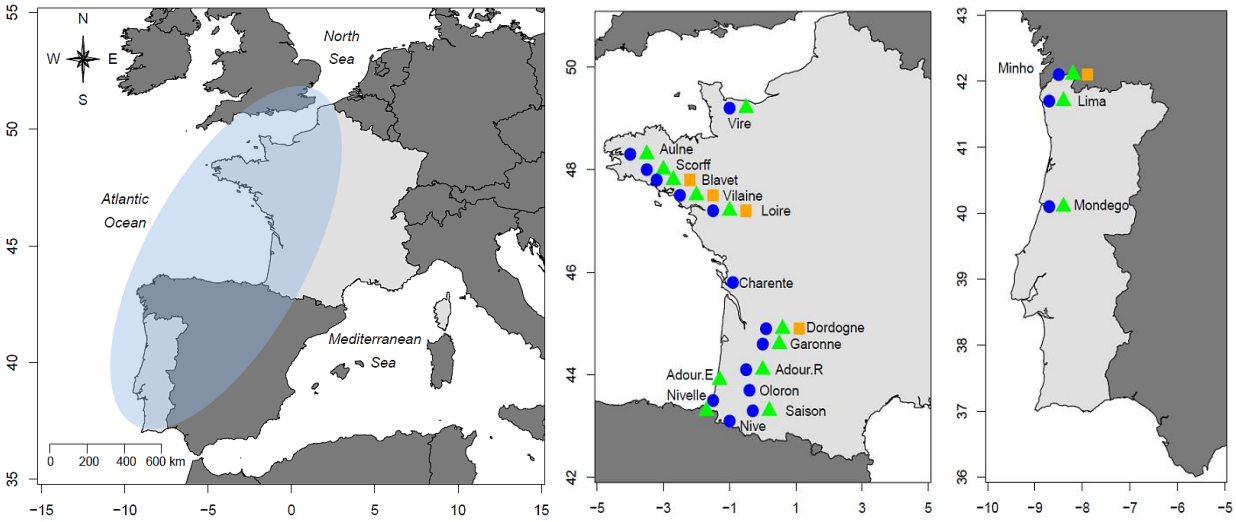

Figure 1 Left panel: Distribution range of the European Allis shad metapopulation. The blue area represents the actual range of the species. France and Portugal are presented in light grey. Middle panel and left panel: respectively the distribution of the water and fish samples in France and Portugal. The water, adult and juvenile samples are respectively indicated by a blue point (●), a green triangle (▲) and an orange square (■). River names figure next to these symbols. The horizontal and vertical axes correspond respectively to the longitude and the latitude

Samplings were performed once a month from late May to September in 2012 and 2013 near the spawning grounds (Table1). At each site, 100 mL of water was collected. In particular, 2 element concentrations, the Barium and the Strontium (Sr/Ca and Ba/Ca), were measured relatively to the Calcium concentration to be comparable, as well as the isotopic ratio of Strontium ($^{87}Sr/^{86}Sr$). Let see Martin *et al.* (2015) for more details about the water sampling.

2.1.2. Juveniles and adults Allis shad

Juveniles Allis shad (n = 44) were collected between June and October 2013 in four French rivers (the Blavet, the Vilaine, the Loire and the Dordogne rivers) and between September and January from 2009 to 2012 in the Minho River (Table 1). Juveniles were collected using seines and bongo nets in the upper-estuarine region of rivers during the downstream migration.

Adults Allis shad (n = 615) were sampled on 15 rivers from upstream spawning sites to tidal freshwater parts of the watercourse between April and June from 2001 to 2014 (Table1). This period matches with the upstream spawning migration of Allis shad. The collected fishes were caught by fishermen, with a trammel net or by sport fishing. Some fishes were found dead after spawning periods. Fishes were measured near the millimeter when possible and then frozen.

Table 1 Number of adults and juveniles Allis shad sampled in each river per year and their mean fork length (mm) ± standard deviation. The number of water sampling in 2012 and 2013 is also specified for each river.

	Adults												Juveniles							Water	
Rivers	2001	2008	2009	2010	2011	2012	2013	2014	***Total***	Mean fork length ± sd			2009	2011	2012	2013	Mean fork length ± sd			2012	2013
Adour E.					2		29		**31**	—	±	—									
Adour R.							6		**6**	—	±	—									5
Aulne							12		**12**	480,8	±	52,1									1
Blavet							7		**7**	—	±	—				18	70	±	10,2		2
Dordogne	38					5	66		**109**	483,4	±	42,7				3	72,7	±	22,0		2
Garonne	43	81				27	37	43	**231**	487,3	±	72,0								1	3
Lima							4		**4**	583,8	±	73,9									1
Loire					4		24		**28**	511,2	±	48,3				4	278	±	28,9	1	4
Minho			24	21	25		17		**87**	595,7	±	47,3	10	4	6		97,9	±	12,1		6
Mondego							15		**15**	494,0	±	30,0									1
Nivelle			16						**16**	484,7	±	39,4									4
Saison							6		**6**	—	±	—									3
Scorff							10		**10**	516,7	±	59,8									1
Vilaine				3	10		6		**19**	525,9	±	28,2				1	350	±	—		4
Vire					7		27		**34**	486,9	±	41,1									1
Charente																					2
Oloron																					5
Nive																				4	

Note that the water baseline contains 3 rivers without fish samples (the Charente, the Oloron and the Nive rivers). The Adour E. corresponds to the estuary and consequently fishes couldn't be reallocated into this site since Allis shad is known to reproduce in the middle watercourse of rivers.

2.2. Samples preparation and microchemistry analysis

Samples were prepared before the start of this work. Protocols are detailed in Martin *et al.* (2015).

Water samples were analyzed to measure elemental concentrations using a solution-based-sensitive Inductively Coupled Plasma Mass Spectrometer (ICP-MS). The isotopic ratio ($^{87}Sr/^{86}Sr$) analysis was performed using a Nu-Plasma Multi-Collector Inductively Coupled Plasma Mass Spectrometer (MC-ICP-MS) following the protocol described by Martin *et al.* (2013).

Since the otolith grow from birth to death, the natal signature of the river experienced by a fish during the juvenile stage corresponds to the portion near the core of the otolith. In order to target this particular stage, Martin *et al.* (2015) performed a C-shaped ablation trajectory. The ablation diameter corresponds to the time during which juveniles experience freshwater. The chemical signature of the core of the otolith reflects the composition of the marine water experienced by the female before the upstream migration (Volk *et al* 2000). In order to exclude this maternal effect on the core signature, the ablation was performed 40µm away from the core. A first semi corona was ablated by a laser to ICP-MS for elemental concentrations analysis and a second semi corona was ablated by a laser to MC-ICP-MS for isotopic ratio analysis (Figure A.1). The width of the two semi coronas was 60 µm, so that the external part of the ablation was placed 100µm from the primordium (Figure A.1). All the elemental concentrations were above the limit of detection (LOD).

Since the two coronas correspond to the juvenile stage, the combined use of $^{87}Sr/^{86}Sr$, Sr/Ca and Ba/Ca defined as a multi-dimensional space allowing the characterization of the natal origin of adults Allis shad.

2.3. Preliminary analysis of water and juvenile baselines

A preliminary analysis of the discriminant capacity of the baselines was performed before examining the outputs of the models. This analysis is necessary to ensure that water and juvenile signatures do not overlap and would allow precise reallocation. The variability of the 3-dimensional water signatures was first tested considering a 'river' effect on the isotopic and the elemental concentrations using a non-parametric Wilcoxon test with Bonferroni adjustment. In parallel, a Canonical Discriminant Analysis (CDA) was performed on the water and the juveniles' otolith microchemistry data sets in order to check the discriminant capacity of the isotopic ratio and the elemental concentrations between rivers and juveniles of the baseline. This analysis was performed using the ade4 package of R software (R Development Core Team, R.3.1.1, 2014). The temporal stability of the juvenile and the water baselines were not tested because of a lack of temporal variability in the sampling dates. Therefore, in this study, the temporal stability of the baselines' signatures was supposed to be checked.

2.4. Construction of Bayesian hierarchical models

In the following subsections, ad and jv correspond respectively to the adult and juvenile stages. In the first model, natal rivers were denoting by N and are included in [1,kb]. This range corresponds to the number of rivers of the water baseline. In the second and the third models, N is included in [1,K], with K the total number of sources which can be superior to kb. Brackets {} denote vectors and braces [] represent matrices. More details about the structure of the models are available in Appendix II.

2.4.1. Bayesian model with fixed number of sources and multiple baselines

The otolith composition could be seen as the result of the integration of the water elements and a partitioning due to three interfaces, the gills, the cellular transport and the crystallization in the otolith (Bath *et al.* 2000). As Bath *et al.* (1999) found a linear relation between water and otolith concentrations in Ba and Sr, a linear regression was performed between the water concentrations in Ba and Sr in the rivers where juveniles were sampled (i.e. the Blavet, the Vilaine, the Loire, the Dordogne and the Minho rivers) and the otolith concentrations of juveniles. The regressions were significant for Sr/Ca ($F=1269$; $df=3$; p-value $=4.865e-05$) and the Ba/Ca ($F=18.06$; $df=3$; p-value $=0.02388$) with a high degree of adjustment between the water and the otolith concentrations (respectively $R^2=0.998$ and $R^2=0.858$ for Sr/Ca and Ba/Ca). Therefore, because the Sr/Ca and Ba/Ca ratios in the otolith are deposited in proportion to their ratios in water, a linear relationship was assumed between water and otolith composition in the Bayesian model. Such a linear regression was not required for the isotopic ratio since it is not submitted to a partitioning (Blum *et al.* 2000; Kennedy *et al.* 2002).

The water and the otolith composition of the adults and juveniles were preliminary centered and scaled for each element. The water isotopic ratio was centered and scale using the mean and variance of adults' otolith isotopic ratio to conserve equality between isotopic ratios in water and otolith after transformation. This transformation was performed to decrease the correlation between regression parameters. The scaling was also useful to provide a single scale of variations among the elements and the isotopic ratio.

The otolith composition of an adult ad was considered to follow a multinormal distribution (MN). The expectation $\{\bar{O}(r)\}$ (i.e. the average composition of the otolith) was defined by a linear relation linking the water composition of a river r with the partitioning coefficients a and b:

$$\left(\{Oto(ad)\} \mid N(ad) = r\right) \sim MN\left(\{a\} \cdot \{Water(r)\} + \{b\}, [\sum]\right) \quad (1)$$

where Oto(ad) and Water(r) correspond respectively to the otolith composition of an adult ad and the water composition of a river r. N(ad) represents the natal river of the adult ad and $[\sum]$ is the variance and co-variance matrix (i.e. the mathematical precision). It was assumed that the partitioning coefficients for the isotopic ratio are b=0 and a=1 because no partitioning occurs between the water and the otolith compartments. For the elemental composition, each partitioning coefficients follows a flat uniform distribution between [0,2] for a and [-3,3] for b. The slope a was supposed to be positive as shown by Bath *et al.* (2000). An uninformative prior was also chosen for $[\sum]$:

$$[\sum] \sim Wishart\left([I], n\right) \quad (2)$$

with [I] the identity matrix (dimension 3x3) and n the degree of freedom (number of elements +1).
For the juveniles, the natal river is already known so their otolith compositions are described by the following relation:

$$\{Oto(jv)\} \sim MN\left(\{a\} \cdot \{Water(N(jv))\} + \{b\}, [\sum]\right) \quad (3)$$

with N(jv) the natal river, and thus the catch river, of the juvenile jv. Finally, a categorical distribution was proposed to reallocate the adults Allis shad to their natal river:

$$N(ad) \sim Categorical\left(\{\theta_{c(ad),y(ad)}\}\right) \quad (4)$$

For each combination of catch river c(ad) and year y(ad), a vector of probabilities of origin was defined for the kb rivers of the water baseline $\theta_{c(ad),y(ad)}(1),\ldots,\theta_{c(ad),y(ad)}(kb)$. In this model, the a priori probability that an adult ad caught in the river r, the year y, born in each river of the baseline was described by a Dirichlet distribution which is an uninformative prior:

$$\{\theta_{c(ad),y(ad)}\} \sim Dirichlet\left(\{\gamma_{1:kb}\}\right) \quad (5)$$

with $\gamma_1 = \ldots = \gamma_{kb} = 1/kb$ and kb = 17 (i.e. the number of rivers in the water baseline).
The Bayesian hierarchical model provides a probabilistic estimate of the natal river of adults. The transfer of information between the juvenile baseline and the otolith microchemistry of adults is performed by means of the variance-covariance matrix $[\sum]$ and the regression parameters a and b. This first Bayesian model supposed that water composition was effectively sampled in each potential source. This constraint introduces bias in the reallocation of fishes, which could be omitted using an Infinite Mixture Model.

2.4.2. Bayesian model without baseline: Infinite Mixture Model

The second model consisted in a Bayesian hierarchical model which estimates the number of sources found in a mixed sample without reference to any baselines data sets. The clustering was based on the similarity between the otolith microchemistry of the adults without reference to the water microchemistry. This method is similar to that developed by Neubauer *et al.* (2013). Considering a mixed sample, a mixture of Gaussian distributions is assumed for the elemental

and isotopic ratios. The Infinite Mixture Model (IMM) provides an estimate of the number of sources, and the proportion of fish in each source (Munch & Clarke 2008).

When possible, in the second and third models, we used conjugate priors for computational ease (Görür & Rasmussen 2010).

In this model, the average otolith composition $\{\bar{O}(r_i)\}$ of an adult is independent from the water composition and was assumed to follow a normal distribution centered on 0 with a large variance (precision = 1 x 10^{-6}). Consequently, the otolith composition became:

$$\left(\{Oto(ad)\} \mid N(ad) = r\right) \sim MN\left(\{\bar{O}(r)\}, [\sum]\right) \quad (6)$$

The mathematical precision $[\sum]$ followed the same distribution as in the first model (equation 2). Besides, the same categorical distribution for the reallocation was used as in the previous model:

$$N(ad) \sim Categorical(\{\theta_{1:K}\}) \quad (7)$$

with K the number of sources. In a purely Infinite Mixture Model, K can theoretically tends to infinity, however in practice, we constraint K in the coda by specifying a K_{max} = 22 (Appendix II). The reallocation in K_{max} sources is theoretically allowed but at the end of the iterative process, all the allowed sources were not filled. The definition of the K probabilities of origin $\{\theta_{1:K}\}$ is based on the "stick breaking process" also called the "Chinese restaurant process" (Sethuraman 1994; Ishwaran & James 2003). Starting with a single stick, K pieces (i.e. K groups of individuals) could be obtained by a breaking process. Each piece presents a particular length which reflects the probability of belonging to this piece. As the lengths of the pieces decrease when the stick breaking process progresses, the probability of belonging to a new group decreases too. Those probabilities are denoted by $\{q_j, j = 1\ldots K\}$. Due to the stick breaking process, the weights of additional sources decrease when the process progresses. The K number of sources could potentially be infinite but the production of a new group depends on the equilibrium between the production cost of this new group and the benefits (i.e. reduction of variance in clusters). The process is based on the maximization of the extra-group variance contrary to the intra-group variance. A simple example with only three groups is presented in Figure 2.

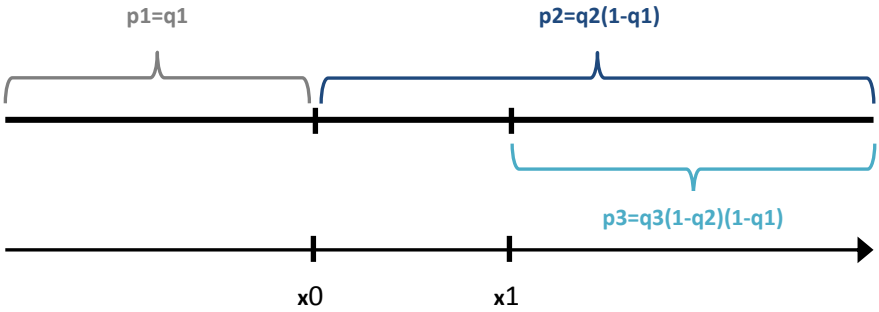

Figure 2 Schematic presentation of the stick breaking process. The large black line represents the stick and the array corresponds to the evolution of the breaking process. This example considers three groups defined by 2 breaking points x0 (the first breaking point) and x1 (the second breaking point). The probabilities associated with each group are presented above or below the brackets.

As presented in Figure 2, the probability of belonging to a new group depends on the size of the previous group. The probability p1=q1 corresponds to the first break, the probability p2=q2(1-q1) is the proportion of the remainder stick from the first break etc.… For each source $\{r_j, j = 1\ldots K\text{-}1\}$, the probability q follows a beta distribution:

$$\mathbf{q(r_j) \sim Beta\left(1,\alpha\right)} \quad \textbf{(8)}$$

with α the concentration parameter defined by $\alpha = 1/\alpha_0$ with α_0 described by a Gamma distribution. The probabilities of origin are defined by construction:

$$\mathbf{\{\theta 1:K\} = \{pi = \prod_{k=1}^{i-1}(1 - q(k)), i = 1 \ldots K\}} \quad \textbf{(9)}$$

Those probabilities $\{\theta_1\}\ldots\ \{\theta_K\}$ follow a Dirichlet distribution, making this model a Dirichlet Process Model (DPM) which belongs to the family of Infinite Mixture Models. The main interest of this model is its capacity to estimate the number of sources, contrary to the first model which assumes reallocations only in rivers of the water baseline. However, in absence of baseline, the second model is not able to associate a source with a river. Here, a source is just a group and is not a precise ecological entity. The use of a hybrid model is a way to combine the advantages of the two first models and to overcome their respective drawbacks.

2.4.3. Bayesian hybrid model: Infinite Mixture Model with multiple baselines

The last model consisted in a combination of the first two models. The baselines (water and juvenile) and the stick-breaking-process were used to allow reallocations in rivers of the water baseline or in extra-sources. When the chemical signatures of individuals do not match those of the baselines, the model produces a new group out of the baselines. Indeed, the inclusion of individuals with "atypical" otolith signatures in a group of the baseline would induce a high intra-group variance and thus, the definition of a new source is preferred for this particular fish. The main interest of the third model lies in its ability to reallocate fishes in extra-sources while keeping the information from baselines.

In this sub-section, the rivers of the baseline are denoted by $\{r_i, i = 1\ldots kb\}$ and the extra-sources are denoted by $\{r_i, i = (kb+1)\ldots K\}$.

Considering a hybrid model, the likelihood was defined by:

$$\mathbf{\left(\{Oto(ad)\} \mid N(ad) = r\right) \sim MN\left(\bar{O}(r_i), [\sum]\right)} \quad \textbf{(10)}$$

with $\{\bar{O}(ri)\} = \{a\} . \{Water(ri)\} + \{b\}$ for rivers of the baseline ($i = 1\ldots kb$) and $\{\bar{O}(r_i)\}$ following a normal distribution for extra-sources ($i = kb +1\ldots K$). In this model, the partition coefficient a was assumed to follow the same uniform distribution as in the first model. However, the coefficient b was different from the first model because of convergence difficulties. This parameter was here described by a normal distribution centered on 0 with a large variance (precision = 1e-6). The mathematical precision [∑] followed the same distribution as in the first model (equation 2).

The otolith composition of juveniles was described by the same multinormal distribution as in the first model (equation 3). Besides, the same categorical distribution for the reallocation was used as in the previous model (equation 7).

Assuming an Infinite Mixture Model, $\{\theta_{1:kb}\}$ was defined by flat Dirichlet priors for the river of the baseline (equation 5). In this model, the a priori probability that an adult caught in the river r, born in an extra-source was described by the stick-breaking process:

$$\{\theta\ kb+1{:}K\} = \{pi = \prod_{j=kb+1}^{i-1} (1-q(j)), i = (kb+1) \dots K\} \quad (11)$$

with q following the same distribution as in equation 8. Contrary to the first model, the year effect was not introduced in the probability of origin $\{\theta_{1:K}\}$ because of convergence constraints.

The structure of the Bayesian hybrid model is outlined in Figure 3. The combination of the model with baselines and fixed number of sources (model 1) with the Infinite Mixture Model (model 2) provides an integrative model of reallocation (model 3).

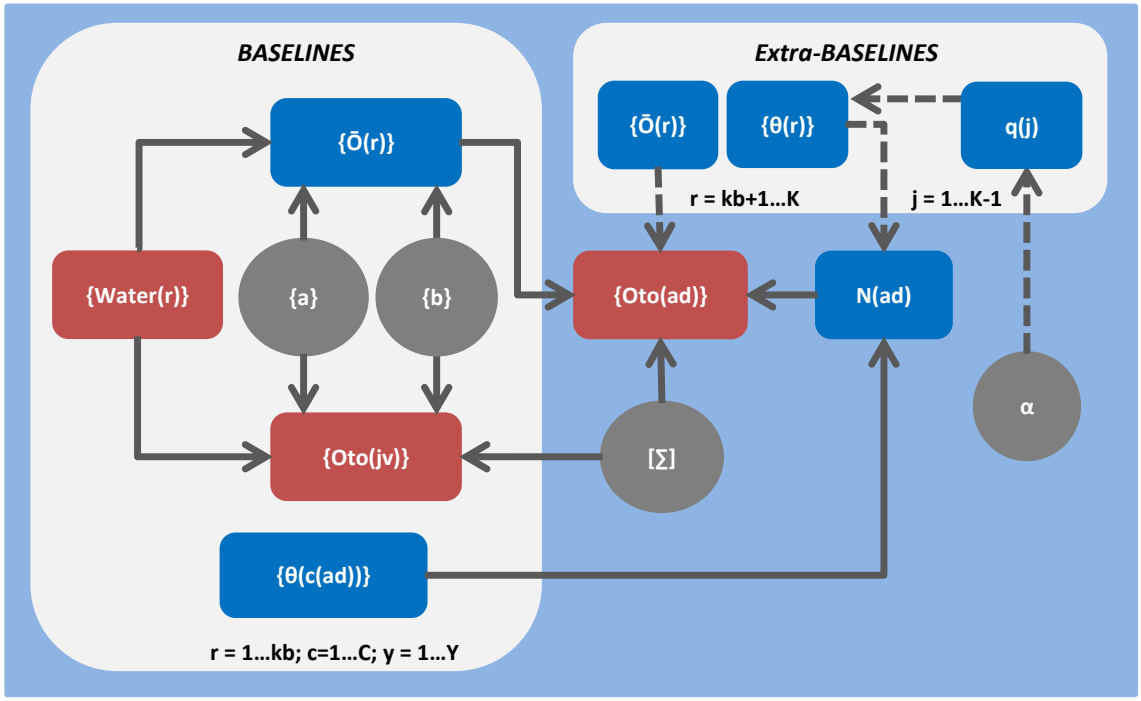

Figure 3 Directed Acyclic Graph of the hybrid Bayesian model considering multiple baselines (water and juveniles) without fixed number of natal rivers. The indexes ad, jv, r, c and y represent respectively the adult stage, the juvenile stage, the river of the water baseline, the catch river and the catch year. Here, kb = 17 (i.e. the number of rivers of the water baseline), C = 15 (i.e. the number of catch river for the adults) and Y = 8 (i.e. the number of catch year for the adults). K corresponds to the total number of sources estimated by the model. The left grey panel represents the rivers of the baseline and the right grey panel corresponds to the extra-baseline. Red boxes point out the data sets. Grey circles are the hyper-parameters and all the other boxes are simple parameters. The dotted lines are relative to the extra-baseline.

2.4.4. Synthesis of the parameters

The parameters of each model are summarized in Table 2.

Table 2 Summary of the parameters of the three Bayesian models. Model 1: model with multiple baselines and a fixed number of sources, Model 2: Infinite Mixture Model, Model 3: hybrid model.

Type of parameters	Parameters	Definition	Priors distribution		
			Model 1	Model 2	Model 3
Hyper-parameters	{a}	Partition coefficients	Uniform(0,2)		Uniform(0,2)
	{b}		Uniform(-3,3)		Normal(0,1e-06)
	[∑]	Variance-covariance matrix	Equation 2	Equation 2	Equation 2
	α	Concentration parameter		InvGamma(1,1)	InvGamma(1,1)
Parameters dependent on hyper-parameters	$\{\bar{O}(r_i)\}$	Average otolith composition	{a}.{Water(ri)}+{b}		{a}.{Water(ri)}+{b} if i = 1…kb; Normal(0,1e-06) if i = kb+1…K
	{Oto(ad)}	Otolith composition of adults	Equation 1	Equation 6	Equation 10
	{Oto(jv)}	Otolith composition of juveniles	Equation 3		Equation 3
	N(ad)	Categorical variable of reallocation of adults	Equation 4	Equation 7	Equation 7
	$\{\theta_{c(ad),y(ad)}\}$	Probability of origin depending on the river and year of sampling	Equation 5		
	$\{\theta_{1:K}\}$	Probability of being originated from one of the K sources estimated		Equation 9	
	$\{\theta_{1:kb}\}$	Probability of being originated from one of the river of the baseline			Equation 5
	$\{\theta_{kb+1:K}\}$	Probability of being originated from an extra-source			Equation 11
	q(r)	Probability of belonging from a group defined by the stick-breaking process		Equation 8	Equation 8

2.4.5. Bayesian posterior distribution using MCMC sampling

Computations were performed with R software (R Development Core Team, R.3.1.1, 2014). The Monte Carlo Markov Chain (MCMC) method was used to draw simulations from Bayesian posterior distributions with the rjags package providing an interface from R to JAGS (Just Another Gibbs Sampling; Plummer 2003) library. Three MCMC chains were run in parallel for each model. For the first model, 20 000 iterations were run after a burn-in period of 10 000 iterations. On account of the stick breaking process used in the second and the third models, the number of iterations was increased to 200 000 with a burn-in period of 50 000 in order to target the chains' convergence. The monitoring was performed on a, b, [∑], α, $\bar{O}(r)$, N(ad) and $\{\theta\}$.

2.4.6. Convergence diagnosis

The convergence was tested for all posterior samplings using the Gelman and Rubin convergence diagnosis (Gelman & Rubin 1992) with the Coda library. The convergence of a parameter is checked if the potential reduction factor is below the threshold of 1.05 (Brooks

and Gelman, 1998). The convergence of categorical variables of reallocation N(ad) were checked by a visual examination of the chains mixing (see Appendix II for example of posterior checking).

2.5. Models comparison

Before selecting one of those three Bayesian models, a comparison of statistical performances and ecological reliability have to be performed. Thus, the three models were compared using both statistical criteria and indicators of reallocation reliability.

2.5.1. Statistical Criteria

The DIC (Deviance Information Criterion) was calculated for each model and compared for models presenting the same data structure (i.e. the first and the third models). This criterion is a Bayesian measure of fit, penalized by the model complexity (Spiegelhalter *et al.* 2002). The DIC is thus defined by the sum of an estimate of fit, plus twice the number of parameters. It is assumed that the model i is better than the model j in term of overall fit if the DIC of the model i is smaller. In addition, convergence diagnosis and posterior checking were used to compare models 'performances.

2.5.2. Indicators of reallocation reliability

The reliability of reallocation was evaluated and compared between models using several indicators. First, the comparison of the maximum probability of reallocation of a fish in a particular source was performed between models. Each iteration generates a reallocation of all fishes in sources. At the end of the iterative process, each fish has been reallocated in one or more sources. The frequency of reallocation of a fish i in a source k was defined as the probability of reallocation of i in k. For example, in the first model, each fish was reallocated 20 000 times in one or several sources. Considering kb sources, each fish presents kb probabilities of reallocation (i.e. frequencies of reallocation). Here, the higher frequency of reallocation for a fish in a source corresponds to the Maximum Probability of Reallocation (MPR). The final reallocation of a fish in a particular source corresponds to the source associated with the MPR. A high MPR is associated with a high reliability in the reallocation process.

The number of sources in which each fish was reallocated during the successive iterations was also considered as a reliability indicator allowing models' comparison. A low number of sources estimated during the MCMC sampling reflects a high reliability in the reallocation.

Besides, the Shannon entropy was calculated for each fish using the Entropy package. Shannon entropy is currently used to assess biodiversity. We found it was very relevant in our context because it summarizes the diversity in reallocations. The Shannon entropy of an adult ad is defined by:

$$\mathrm{H(ad)} = -\sum_{i=1}^{n} Pi * \log(Pi) \qquad \textbf{(12)}$$

with Pi = Ni/N the proportion of reallocation in the source i for the fish ad in regard to the total number of sources $N = N_1 + \ldots + N_n$ estimated for this fish during the iterative process. Thus, a fish presenting only 1 estimated source during the iterative process would present H = 0. This result indicates high reallocation reliability. In opposite, if a fish is reallocated in a large number of sources (i.e. a high diversity of estimated sources), the reallocation reliability decreases and H increases. The Shannon entropy increases when unusual sources are estimated during the iterative process.

Therefore, the MPR, the number of estimated sources and the Shannon entropy of individuals were used to compare the models in terms of reallocation reliability.

2.5.3. Comparison of reallocation and sources between models

The sources and number of fishes reallocated in each sources were compared in order to test the stability of the reallocation results between models. Besides, a focus was done on "inconsistent" fishes (i.e. fishes reallocated in different sources between models) and "stable" fishes (i.e. fishes reallocated in the same sources between models). A comparison of the number of "inconsistent"/"stable" fishes per source and their MPR were compared between models. Another way to check the reallocation reliability is to explore the stability of reallocation between models, for straying fishes performing long distance migration between natal and spawning rivers. Furthermore, visual convergence checking of the categorical variable N and otolith composition of fishes reallocated in extra-sources in the third model were examined.

2.5.4. Confusion of reallocation

Using a matrix constituted by 615 fishes in row and K columns corresponding to the probabilities of reallocation of each fish in the K sources, the confusion of reallocation was examined. More precisely, the confusion of reallocation between sources was investigated using Spearman correlation tests applied on the columns of the previous matrix. These correlations were implemented with the Hmisc package, providing one correlogram per model. This analysis is a way to highlight which sources provide similarity in their probabilities of reallocation, indicating a possible confusion between those rivers. It thus allows the identification of rivers which are potentially poorly discriminated

2.6. Flux between donor and recipient rivers

After comparing the three models, the best model was chosen to develop a model of exchange between rivers. Estimated abundances of adults Allis shad were multiplied with the probabilities of origin $\{\theta\}$ corresponding to the outputs of the selected model to estimate flux between donor and recipient rivers. This approach allows the quantification of flux direction and intensity. A donor river produces spawners (homing and straying fishes) and a recipient river received spawners (homing and straying fishes). The homing occurs when the donor is also the recipient river. A closed river only performs homing.

Abundance estimates were available in several rivers in France from Non-Governmental Organizations (transmitted and updated by P. Jatteau – Irstea Bordeaux). These abundance estimates are presented in Table A.1. Some watersheds are missing because reports do not discriminate shads between *Alosa alosa* and *Alosa fallax* (in the Charente and the Vilaine rivers) or because of a lack of reliable monitoring. The abundances of adults Allis shad in the Minho River in 2009, 2010 and 2011 were retrieved from Mota *et al.* (2015). For the Garonne and Dordogne Rivers, the abundance estimates were derived from a "bull" counting following the method of Carry & Borie (2013). For the other rivers, the abundance estimates were obtained by a video counting system on fishways usually located downstream the spawning grounds. For the Loire River, some fishways were located upstream of the spawning grounds which involved an under-estimation of abundances (www.logrami.fr). Considering that adult otolith samples were concentrated in 2013 (Table 1), the estimated abundances were multiplied with the probabilities of origin estimated for fishes caught in 2013. In 2013, abundances estimates were monitored only in 7 rivers (Table A.1) and thus, an extrapolation of the missing data was needed for the 8 other rivers. Based on the common idea that the population size depends on the habitat size (carrying capacity), the relation between spawner abundances and the surface of the

watershed was investigated to perform an extrapolation of population size in missing data. A nonlinear relation was found (Figure 4).

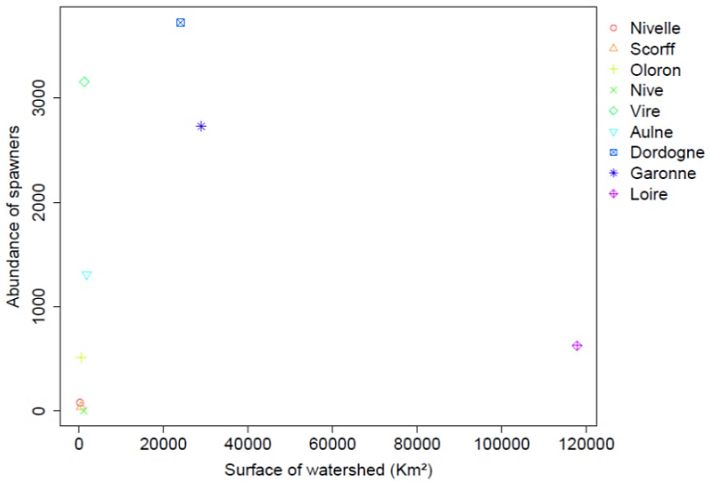

Figure 4 Relation between the surface of watersheds (km²) and the abundance of adults Allis shad during the upstream migration in 2013. Only a few watersheds are presented because of missing data in 2013.

Whereas Loire is the largest watershed of the study, its shad abundance estimate was similar to the shad abundance of small watersheds (e.g. the Oloron watershed), but this figure is underestimated as mentioned above. In opposite, whereas the Vire watershed is 23 times smaller than the Garonne watershed, Garonne and Vire estimated abundances are similar. Therefore, the extrapolation could not be performed using a linear relation between the surface of the watersheds and the estimated abundances. Thus, an Ascending Hierarchical Classification (AHC) of the watersheds was preferred. The analysis was implemented with the stats package. The AHC was performed using the surface of watersheds as classification variable. The Euclidian distance was used as dissimilarity index. The Ward's minimum variance method was implemented to aggregate watersheds. This method minimizes the gain of inertia intra-class each iteration. Finally, the partition was determined by checking the loss of inertia due to an increase in the number of groups (Figure 5.A).

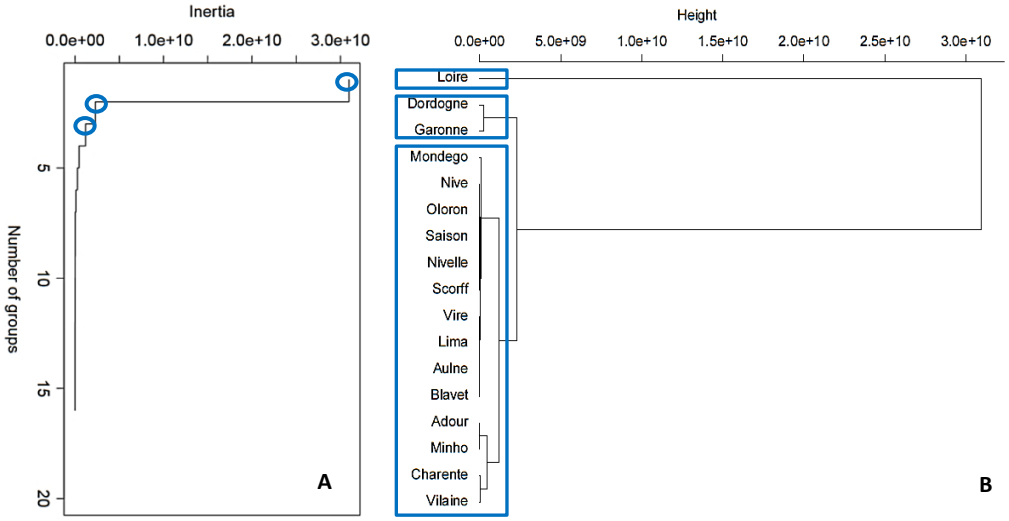

Figure 5 Results of an Ascending Hierarchical Classification implemented on the surface area of the watersheds (km²). A: Evolution of the intra-class inertia with the number of groups. B: Tree produced by the clustering process. Three groups were defined considering that the loss of inertia would be too large with more groups.

Based on the similarity of surface between watersheds, the AHC produced three groups (Figure 5.B). The Loire watershed is the larger (117 800 km²) of the data set and was thus cluster in a unique group. The Garonne and Dordogne watersheds were clustered in another group considering an intermediate surface (respectively 28 900 and 24 000 km²). Finally, small watersheds (N = 14) were classified in the ultimate group presenting small surface compared to the Loire, the Garonne and the Dordogne watersheds (between 238 km² for the Nivelle watershed and 17 080 km² for the Minho watershed). Assuming an incomplete data set of abundances, priors' distributions were chosen for the three groups previously defined. Uninformative flat priors were preferred. Considering the uncertainty of abundance estimates in the Loire river, a large uniform prior was chosen between [500,4000]. For the Garonne and Dordogne rivers, uniform prior between [2000,4000] was assumed. Finally, because of the variability in the abundances in the other rivers, a uniform prior between [0,4000] was defined.

3. Results

3.1. Analysis of water and juvenile baselines

Mean water composition (isotopic ratio and elemental concentrations) were significantly different among rivers of the baseline (Wilcoxon test, $p < 0.05$). However, because of a low number of samples per river, the Bonferroni adjustment test did not found significant differences between couple of rivers for any elements.

Besides, a strong spatial segregation of water and juvenile samples was found using a Canonical Discriminant Analysis (Figure 6).

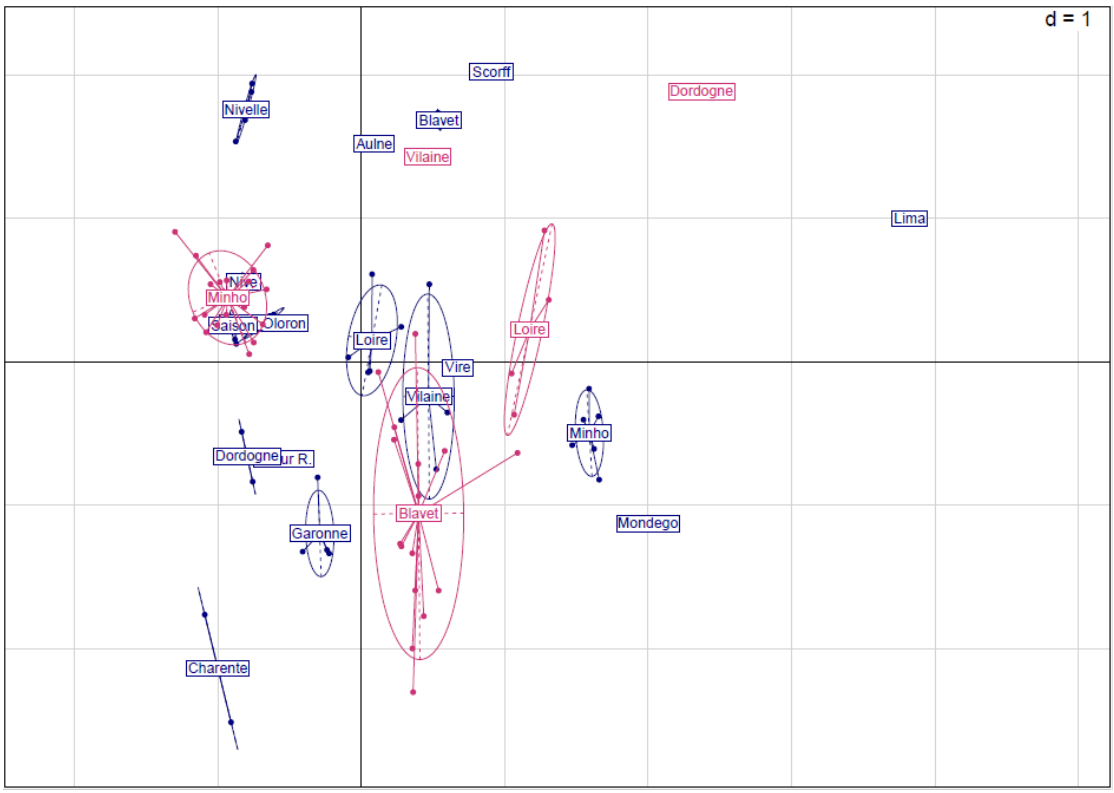

Figure 6 Canonical Discriminant Analysis performed on the water microchemistry of 17 rivers sampled in France and Portugal (in blue) and on the otolith microchemistry of juveniles (in pink). The first canonical variate corresponds to the isotopic ratio and the second is supported by the Sr/Ca ratio and in a less extend by the Ba/Ca ratio. Symbols represent the water and juvenile samples and ellipses are 95% confidence intervals around the mean value.

According to a permutation test, the discrimination between rivers was significant for water (Wilks = 2.7e-5, $p < 0.01$) and juvenile samples (Wilks = 0.0052, $p < 0.01$). Water signatures were mainly segregated by $^{87}Sr/^{86}Sr$ (CV 1) and Sr/Ca (CV 2) (Table 3). The discrimination of juvenile signatures was mostly driven by $^{87}Sr/^{86}Sr$ (CV 1) and both Sr/Ca and Ba/Ca (CV 2) (Table 3). The first two canonical variates (CV) explain respectively 70% and 88% of the signature variability in the water and juvenile data.

Table 3 Canonical weights for the first two canonical variates performed on $^{87}Sr/^{86}Sr$, Sr/Ca and Ba/Ca of the water and juvenile baselines. Higher is the absolute value and better is the contribution of the variable to the construction of the axis. Canonical variate 1 (CV 1) and Canonical variate 2 (CV 2) correspond respectively to the horizontal and vertical axes.

Baselines		CV 1	CV 2
Water	**$^{87}Sr/^{86}Sr$**	-1.12	-0.178
	Sr/Ca	-0.038	0.959
	Ba/Ca	0.155	0.208
Juveniles	**$^{87}Sr/^{86}Sr$**	-0.986	-0.322
	Sr/Ca	0.0850	-0.521
	Ba/Ca	0.00225	-0.582

Apart from the Loire, Vilaine and Vire, rivers are mainly segregated by $^{87}Sr/^{86}Sr$ and Sr/Ca. Juveniles from the Blavet River were discriminated by Sr/Ca and Ba/Ca whereas other juvenile samples were separated by both the isotopic ratio and the elemental concentrations.

The water and juvenile signatures appeared to be relatively stable over the whole 2013 sampling season, in view of the confidence ellipses. Juveniles from the Minho River also presented a relatively stable signature over the sampling dates in 2009, 2011 and 2012.

3.2. Models comparison

3.2.1. Statistical comparison

Since the second model (i.e. Infinite Mixture Model) does not introduce any baselines data sets, the DIC was used to compare the first (i.e. with multiple baselines and fixed number of sources) and the third (i.e. hybrid model) models. The third model presented a better overall fit than the first model (DIC_{model1} = -290.823; DIC_{model3} = -577.368). A river effect was tested in the variance-covariance matrix [$\sum$] in the first and third model in order to take into account the inter-river variability of this parameter. However, the introduction of model complexity was largely penalized by the DIC in both models (DIC_{model1} = 29695; DIC_{model3} = 18735) and thus, this river effect was not kept in the models.

The Gelman and Rubin diagnosis was applied on the main parameters of each model. As the number of parameters is considerable, a global comparison of the convergence checking was performed (Table 4).

Table 4 Percentages of parameters fulfilling convergence condition, tested with the Gelman and Rubin Convergence diagnosis. The convergence is checked when the upper bound of the range of the potential scale reduction factors remains under 1.05. The convergence diagnosis was performed on the monitored parameters of the three models. The number of parameters is specified for each model (n). We found K=21 for the second and third model. Model 1: model with multiple baselines and fixed number of sources, Model 2: Infinite Mixture Model, Model 3: Hybrid model.

Parameters	Model 1		Model 2		Model 3	
	%	n	%	n	%	n
$\{\theta_{1:K}\}$	99	2040	0	K	45	(17+K)*3
$\{\bar{O}(r)\}$	98	51	8.7	K*3	55	51
[$\sum$]	100	9	0	9	12	9
{a}	93	2	–	–	99	2
{b}	99	2	–	–	99	2
α	–	–	0	1	0	1

The first model presents a high proportion of parameters satisfying convergence criteria. The second model appeared to be the worse model according to the Gelman and Rubin's convergence diagnosis. The third model presented a better convergence than the second model for all parameters, however, the convergence was better for the first model than the third model, especially for $\{\theta_{1:K}\}$, $\{\bar{O}(r)\}$ and $[\sum]$.

The visual convergence check of reallocation in the third model (98%, 600/615 fishes) was better than the second model (3.6%, 22/615 fishes) but worse than the first model (100%). Example of convergence check of reallocation is presented in Figure A.2. We also present posteriors of the parameters a and b from the first and third models in Figure A.3. Then, examples of convergent and non-convergent probabilities of origin are presented in Figure A.4 for the first and third models.

The first model seemed better than the third model in terms of convergence but the third model was better than the first model in terms of overall fit. Thus, based on DIC and convergence analysis, the first and the third models were both selected and were thereafter extensively compared. Comparison of reallocation reliability had to be performed to complete statistical comparisons. However, the second model was already removed from subsequent analysis because of poor convergence.

3.2.2. Indicators of reallocation reliability

The reallocation reliability was investigated using three different metrics, the MPR (i.e. the Maximum Probability of Reallocation), the Shannon entropy and the number of sources per fish. To compare the first and third models, five groups representing the reallocation reliability were defined using those three metrics (Table 5). The five groups were arbitrary defined by cutting the distribution of each metric into five equal parts.

Table 5 Definition of five groups representing the reallocation reliability. Intervals are defined for the MPR, Shannon entropy and number of sources per fish.

Group	Reallocation reliability	MPR	Entropy	Number of sources
A	Very good	≥ 0.80	[0; 0.35[	[1; 2]
B	Good	[0.60; 0.80[	[0.35; 0.70[	[3; 4]
C	Medium	[0.40; 0.60[	[0.70; 1.05[	[5; 6]
D	Poor	[0.20; 0.40[	[1.05; 1.40[	[7; 8]
E	Very poor	[0; 0.20 [	≥ 1.40	≥ 9

The number of fishes per group for each model and metric is reported in Table 6.

Table 6 Percentages of fish per group of reallocation reliability for each metric (MPR, Shannon entropy and Number of sources) and model. Model 1: Bayesian model with fixed number of sources; Model 3: Bayesian hybrid model.

	MPR					Entropy					Number of sources				
Group	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Model 1	79.2	12.2	5.85	2.75	0	69.1	15.8	8.77	5.36	0.97	16.2	40.2	40.7	2.11	0.16
Model 3	83.8	8.30	5.50	2.40	0	78.0	12.4	6.67	2.44	0.49	52.2	21.0	19.5	5.68	1.62

In most cases, fishes were mostly reallocated with a high reliability based on the MPR and the Shannon entropy in the two models. Notice that the third model was slightly better in terms of reallocation reliability using the three metrics. Indeed, fishes were mainly in intermediate groups (B and C) for the first model whereas fishes were mainly in the first group (A) for the third model.

Accordingly to indicators of reallocation reliability, the hybrid model appears as a better candidate to examine the functioning of the metapopulation than the first model. However, such results in not surprising: by increasing the number of potential sources in the third model, it is

logical that we get more homogeneous groups, with consequently higher MPR. Moreover, in the absence of validation data, these results should be taken with caution: these indicators measure whether models are sure of what they "say", but not whether what they "say" is wrong or true. Before selecting one of the two models, we have to compare reallocations between models.

3.2.3. Comparison of reallocation

3.2.3.1. Comparison of sources and homing rate

How many and which sources per model?

In the first model, 15 natal rivers were estimated using the MPR as criterion to reallocate adults. Comparatively, the third model estimated 12 natal rivers according to the MPR, including one extra-source (namely s21). During the iterative process, 4 extra-sources (namely s18, s19, s20, s21) were created but final reallocation (using the MPR) occurred into only one extra-source (s21). The posterior distribution of K for the third model is presented in Figure A.5. The Vire, Scorff, Adour and Saison rivers were not sources in the third model comparatively to the first model. In both models, any fish was reallocated in the Charente and Mondego rivers (i.e. they were not identified as sources).

Changes in homing rate and reallocation reliability between models?

In the first model, the homing behavior (i.e. adults' return to natal site to reproduce) occurred in 10 rivers among 15 natal rivers (Table 7). It represented 46% of adults Allis shad. Homing occurred with a high reliability in 5 rivers (the Aulne, Blavet, Adour, Nivelle and Minho rivers; mean MPR above 0.80). More precisely, strict homing (i.e. 100% of individuals) was found in the Aulne, Blavet and Nivelle rivers with high reliability (mean MPR above 0.80).

Comparatively to the first model, homing represented 39% of adults Allis shad in the third model. More precisely, 7 rivers were concerned by the homing behavior, among which 5 rivers presented a high reallocation reliability (the Aulne, Blavet, Dordogne, Nivelle and Minho rivers; mean MPR above 0.80) (Table 8). The homing behavior observed in the Vire, Vilaine and Adour rivers were not encountered in the third model contrary to the first model.
In the Loire river, the homing rate increased from 46% to 61% between the first and third models, with a gain of reliability (mean MPR increased from 0.40 – 0.60 to 0.60 – 0.80). It was also the case in the Dordogne river where the mean MPR increased from 0.60 – 0.80 to more than 0.80.

Therefore, the total number of fish displaying natal homing decreased from the first to the third model, but the reallocation reliability increased for those fishes.

Focus on Garonne River

In both models, any fishes displayed homing in the Garonne River whereas an important number of fishes were sampled during the upstream migration in 2001, 2008, 2012, 2013 and 2014. However, the number of strayers (i.e. fishes born in other rivers) reallocated in the Garonne River doubled in the third model (from 12 to 28).

The number of sources for fishes sampled in the Garonne decreased from 7 to 2 in the third model. In parallel, the reallocation reliability for those fishes increased in the third model (mean MPR between 0.20 – 0.60 in the first model, mean MPR > 0.60 for 95.7% of fishes in the third model).

Therefore, regardless to the model, no homing was found in the Garonne River.

Table 7 Percentages of adults Allis shad reallocated in each potential natal river with the first model. The number in parenthesis correspond to the effective number of fishes. Reallocations are identified using the mean MPR of fishes caught in collection rivers and reallocated in natal rivers: ■ 0.20 ≤ MPR < 0.40 ■ 0.40 ≤ MPR < 0.60 ■ 0.60 ≤ MPR <0.80 ■ MPR ≥ 0.80

	Natal Rivers																
Collection Rivers	**Vire**	**Aulne**	**Scorff**	**Blavet**	**Vilaine**	**Loire**	**Charente**	**Garonne**	**Dordogne**	**Nive**	**Adour.R.**	**Oloron**	**Saison**	**Nivelle**	**Minho**	**Lima**	**Mondego**
Vire (34)	76,5(26)			14.7(5)	8.82(3)												
Aulne (12)		100(12)															
Scorff (10)				90(9)						10(1)							
Blavet (7)				100(7)													
Vilaine (19)			5.26(1)	10.5(2)	84.2(16)												
Loire (28)					14.3(4)	46.4(13)		39.3(11)									
Dordogne (109)									92.7(101)	4.59(5)				2.75(3)			
Garonne (231)									51.1(118)	2.17(5)	42(97)	1.30(3)	0.43(1)	2.60(6)		0.433(1)	
Adour R. (6)											83.3(5)			16.7(1)			
Adour E. (31)		3.23(1)								54.8(17)	41.9(13)						
Saison (6)											50(3)			50(3)			
Nivelle (16)														100(16)			
Minho (87)					1.15(1)										98.9(86)		
Lima (4)															50(2)	50(2)	
Mondego (15)					6.67(1)			6.67(1)						6.67(1)	80(12)		
Total	26	13	1	23	25	13	0	12	219	28	28	3	1	30	100	3	0

Table 8 Percentages of adults Allis shad reallocated in each potential natal river with the third model. The number in parenthesis correspond to the effective number of fishes. Extra-sources are denoted by s18,19s20 and s21. Reallocations are identified using the mean MPR of fishes caught in collection rivers and reallocated in natal rivers: ■ 0.20 ≤ MPR < 0.40 ■ 0.40 ≤ MPR < 0.60 ■ 0.60 ≤ MPR <0.80 ■ MPR ≥ 0.80

	Natal Rivers																		
Collection Rivers	**Vire**	**Aulne**	**Scorff**	**Blavet**	**Vilaine**	**Loire**	**Charente**	**Garonne**	**Dordogne**	**Nive**	**Adour.R.**	**Oloron**	**Saison**	**Nivelle**	**Minho**	**Lima**	**Mondego**	**S18,19,20**	**S21**
Vire (34)				2,94(1)		97,1(33)													
Aulne (12)		91,6(11)																	8,33(1)
Scorff (10)				80(8)						10(1)									10(1)
Blavet (7)				100(7)															
Vilaine (19)				5,26(1)		89,5(17)													5,26(1)
Loire (28)						60,7(17)		39,3(11)											
Dordogne (109)									94,5(103)					5,50(6)					
Garonne (231)									95,7(221)					3,90(9)		0,433(1)			
Adour R. (6)									83,3(5)					16,7(1)					
Adour E. (31)		3,23(1)						41,9(13)				35,5(11)		19,4(6)					
Saison (6)								50(3)						50(3)					
Nivelle (16)														100(16)					
Minho (87)					1,15(1)										98,9(86)				
Lima (4)															50(2)	50(2)			
Mondego (15)					6,67(1)			6,67(1)							80(12)				6,67(1)
Total	0	12	0	17	2	67	0	28	329	1	0	11	0	41	100	3	0	0	4

3.2.3.2. Stable vs inconsistent fishes

Some fishes were denoted as stable fishes because they were reallocated in the same natal river by the 2 models (66% of fishes). Two kind of stable fishes could be distinguished: the stable and credible fishes and the stable and uncertain fishes. For example, fishes reallocated in the Lima and Minho rivers are stable between models, with mean MPR above 0.60, making them stable and credible fishes. It is also the case of fishes from the Blavet and the Nivelle rivers displaying a homing behavior. For stable and credible fishes, interpretation of reallocations would be the same regardless to the selected model. Conversely, some fishes presented poor reallocation reliability but remained stable between models. For example, 1 fish caught in the Scorff River in Brittany was reallocated in the Nive River in the south of France with a low credibility (MPR respectively equal to 0.33 and 0.30 in the first and third model). Two other fishes were supposed to have strayed between the Vilaine River and the Minho and Mondego rivers in Portugal. Despite those straying fishes were stable between models, the reallocation reliability is questionable because of failure in indicators of reallocation reliability (MPR, Shannon entropy and number of sources). Thus, changes in the model structure had not induced improvement in reallocation reliability for some fishes.

On the other hand, some reallocations varied with the model structure, leading to inconsistencies in river reallocation between the first and third model. Since the third model eliminated the Adour River as natal river, individuals reallocated in this river in the first model were reallocated in other rivers in the third model. Hence, 97 fishes caught in the Garonne river in 2008 and 2012 were reallocated in the Adour river with mean MPR between 0.40 – 0.60 in the first model and were finally reallocated in the Dordogne River in the third model with an increase of mean MPR between 0.60 – 0.80. Contrary to stable fishes, changes in model structure provided changes in credibility of reallocation for inconsistent fishes. More precisely, from the first to the third model, changes in reallocation of inconsistent fishes were associated with an improvement of reallocation reliability.

Therefore, cautions have to be taken before interpreting reallocations for inconsistent fishes and stable but uncertain fishes. This tends to demonstrate that thanks to a higher flexibility/complexity, model 3 enhance the reliability of some fishes, but that results are rather consistent between the two models for other fishes.

3.2.3.3. Focus on extra-sources

Assuming K_{max} = 22 sources in the code (see part 2.4.2), only 4 extra-sources (K = 17 + 4 = 21) were created using the stick breaking process. Respectively 65, 17, 2 and 5 fishes were reallocated at least one time in s18, s19, s20 and s21. The mean MPR was 0.33 for each extra-source. At the end of the iterative process, using the MPR as criterion to reallocate fishes, 4 individuals were finally reallocated in s21. Those fishes were caught in the Vilaine, Scorff, Aulne and Mondego rivers (Table 9).

Table 9 Comparison of characteristics and indicators of reallocation reliability for fishes reallocated in s21 at the end of the iterative process in the third model. M1: model with multiple baselines and fixed number of sources, M3: hybrid model.

Individuals	**ALA138**		**Sco2**		**aul1**		**SM4**	
Catch river	Vilaine		Scorff		Aulne		Mondego	
Catch year	2011		2013		2013		2013	
Model	**M1**	**M3**	**M1**	**M3**	**M1**	**M3**	**M1**	**M3**
Natal river	Scorff	S21	Blavet	S21	Aulne	S21	Nivelle	S21
MPR	0.71	0.33	0.97	0.33	1	0.33	0.99	0.33
Shannon entropy	0.64	0.59	0.17	0.17	0	0.011	0.0030	0.041
Number of sources	3	3	3	3	1	4	2	3

In the first model, the fish named "aul1" caught in the Aulne River was supposed to perform homing with high credibility accordingly to indicators of reallocation reliability. In the third model, this fish was reallocated in s21 with MPR divided by 3. Thus "aul1" was classified into the group D (i.e. poor credibility) according to the MPR criterion. However, this fish was still classified into the group A (i.e. very good credibility) according to the Shannon entropy and into the group B (i.e. good credibility) using the number of sources as criterion. It is also the case for the 3 other fishes reallocated in s21. Thus, the Shannon entropy and the number of sources were relatively stable between models but the MPR decreased systematically.

By analyzing the otolith composition of fishes according to their source of origin, it was clear that individuals reallocated in extra-sources showed atypical otolith composition signature, and thus can be considered as outliers. It is especially the cases of fishes reallocated in s20 and s21 (Figure 7), which present extreme signatures.

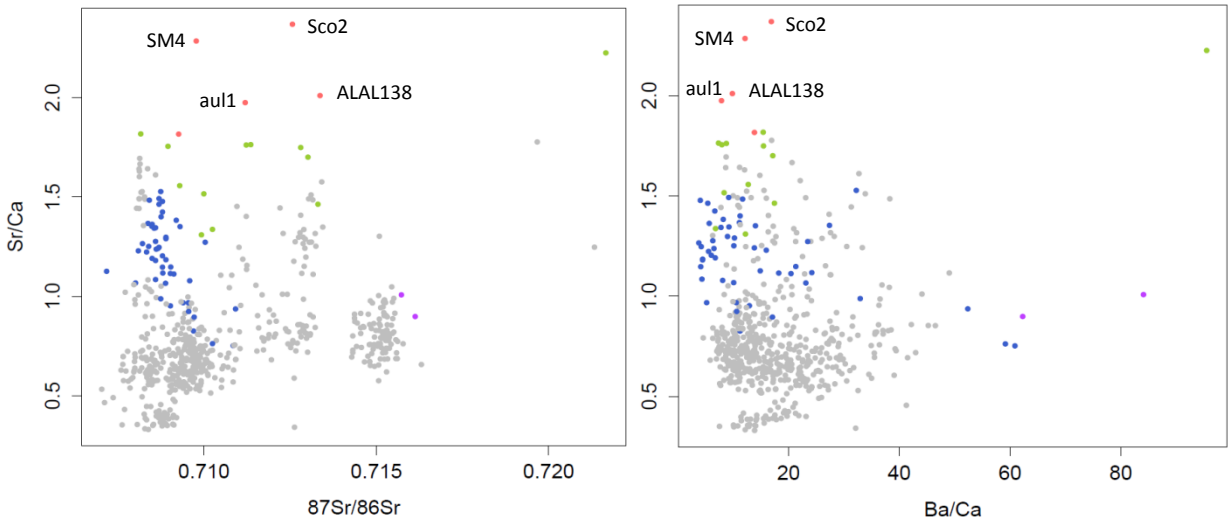

Figure 7 Otolith compositions of adults Allis shad. The vertical axes represent the Sr/Ca ratio. Points are colored according to their reallocation. Grey points ● represent fishes reallocated only in rivers of the baseline during the iterative process. The blue ●, green ●, purple ● and red ● points correspond respectively to fishes reallocated at least one time in sources s18, s19, s20 and s21 during the iterative process. Fishes reallocated in s21 are indicated on each plot.

Fishes from the extra-source s20 showed extreme ratios in Ba/Ca and extreme values of $^{87}Sr/^{86}Sr$. The 4 fishes reallocated in s21 presented extreme ratios in Sr/Ca. Therefore, it seems that reallocation in extra-sources concerned only fishes with extreme otolith signatures.
On the other hand, though uncertain, some fishes remained stable between models because of "classical" signature whereas they presented low MPR (see the example of one fish caught in the Scorff River in Brittany and reallocated in the Nive, as described in 3.2.3.2).

Probabilities of origin in extra-sources ($\{\theta_{kb+1}\}$ … $\{\theta_K\}$) showed poor convergence (potential reduction factors were all above 1.05) and represented 41% of $\{\theta_{1:K}\}$ which failed to converge. Fail in visual convergence of N (i.e. the categorical variable of reallocation) was a supplementary evidence of poor convergence associated with extra-sources. As presented in Figure A.2, the estimate of natal origin was divergent between MCMC chains for fishes reallocated in s21.

Hence, the introduction of extra-sources mainly influences fishes with extreme signatures, and results in poor convergence and limited reliability.

3.2.4. Analysis of inter-river confusion of reallocation

Spearman correlations between probabilities of reallocation (i.e. the frequency of reallocation of a fish in several sources during the iterative process) were investigated for the first and third models. This analysis highlights confusions of reallocation between sources. A negative correlation between sources a and b means that, when the probability of reallocation in a is high, it is low in b, i.e. those sources are well discriminated. Conversely, positive correlations indicate confusion of reallocation, i.e. a low discrimination between sources.

Apart from confusion of reallocation with the Charente River, the Dordogne appeared negatively correlated (i.e. discriminated) with all other rivers in the first model (Figure 8). It was the same case for the Minho River, which presented confusion of reallocation only with the Mondego River. It could be due to a high level of discrimination in water and juvenile signatures between those rivers and other rivers (Figure 6). Despite the Dordogne and Adour rivers showed neighbor signatures (Figure 6), no confusion of reallocation occurred.

In opposite, some rivers presented strong confusions (i.e. high positive correlations). It was especially the case of the Saison, Nive and Oloron rivers which belong to the same watershed (i.e. have the same estuary), and thus presented low discrimination in their water signatures (Figure 6). In the same way, the Scorff and Blavet rivers showed confusion, though more limited. Besides, important confusion occurred between the Loire and Garonne rivers (0.34), likely because of low discrimination in isotopic ratio (Figure 6).

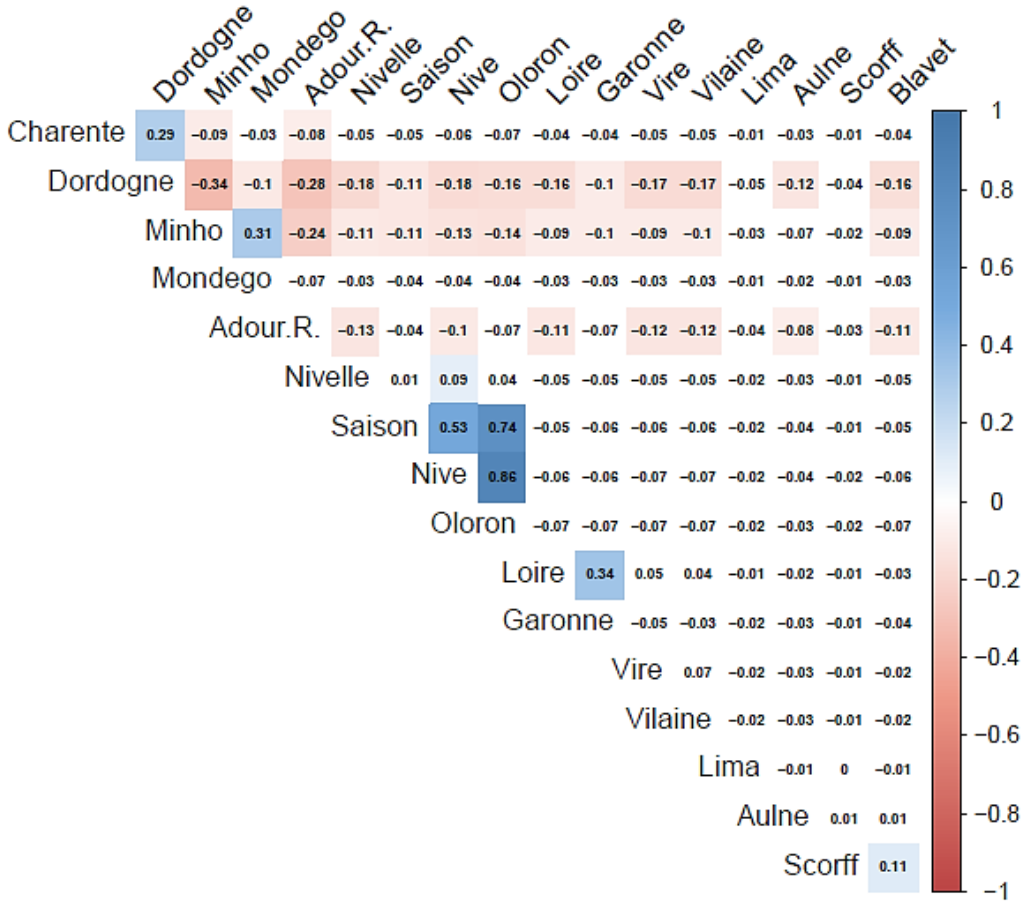

Figure 8 Correlogram of the probabilities of reallocation in rivers of the water baseline for the first model. Positive correlations are represented by blue whereas negative correlations appear in red. Only significant correlations are colored according to Spearman correlation test (the correlation is significant if $p < 0.05$).

The same analysis was realized for the third model (Figure 9). We found that the Dordogne River was independent from all other sources (only negative correlations). Correlation between

the Mondego and Minho rivers decreased but remained present in the third model. Strong confusions between Saison, Nive and Oloron rivers were similar to the first model. However, the Loire and Garonne rivers were not correlated anymore.

Besides, new confusions appeared in the third model, such as confusions between the Garonne and Adour rivers and the Garonne and Charente rivers. We found that the Loire and Vire rivers were both correlated with the Vilaine River, which could be explained by low discriminations in water isotopic ratios (Figure 6).

Furthermore, the Vire River and the extra-source s20 showed high confusion. It was the same case for the Nivelle River and s18 and for the Scorff River and s19. Strong correlations also occurred between extra-sources, such as s19 and s21.

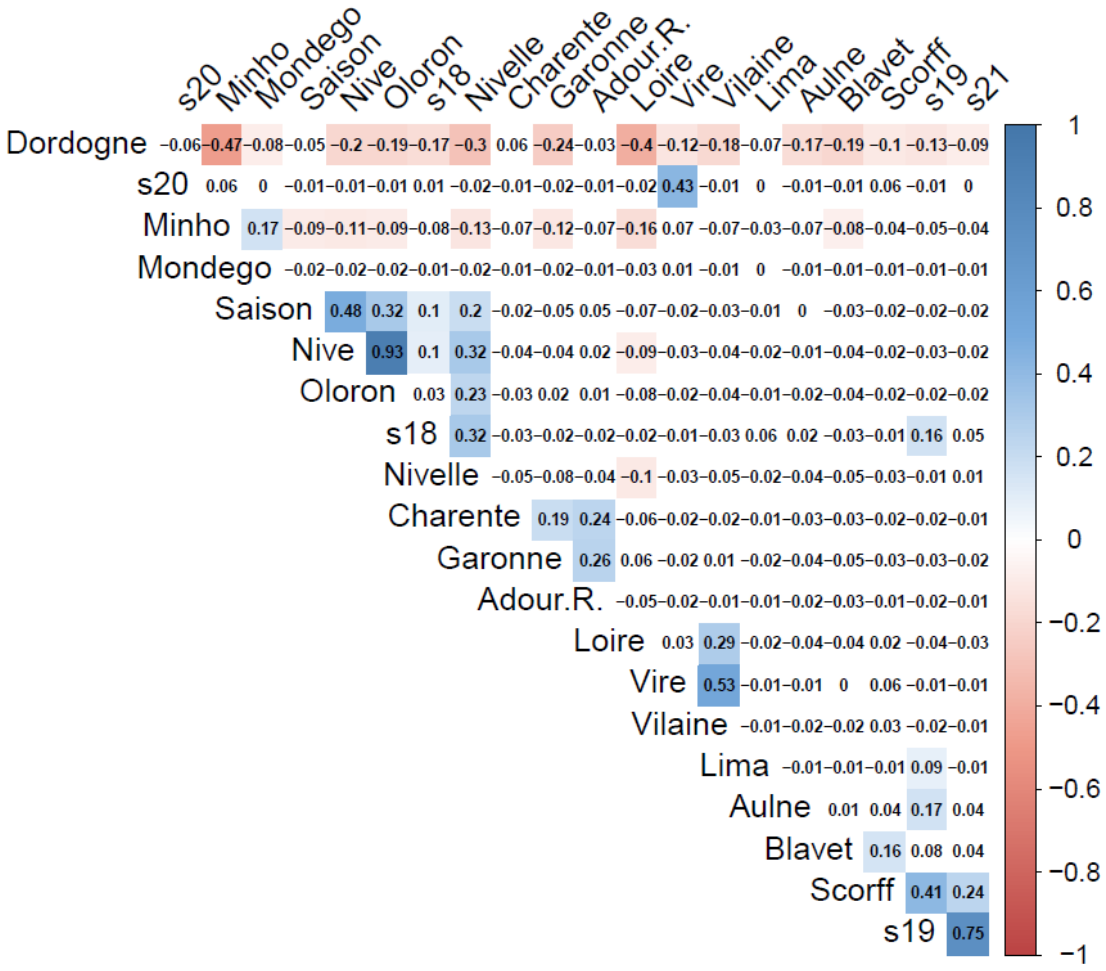

Figure 9 Correlogram of the probabilities of reallocation in rivers of the water baseline and extra-sources for the third model. Positive correlations are represented by blue whereas negative correlations appear in red. Only significant correlations are colored according to Spearman correlation test (the correlation is significant if $p < 0.05$).

Therefore, rivers of the same watershed (such as rivers Saison, Nive and Oloron or rivers Scorff and Blavet) showed strong confusions which were stable between models.

3.2.5. Choice of model to investigate the functioning of the metapopulation

Using the MPR, the Shannon entropy and the number of sources as metrics, we found that the third model reallocated fishes with a better reliability comparatively to the first model. However, the third model generated reallocation with a low reliability in extra-sources. Moreover, the third model showed a worst convergence. Besides, the DIC is relatively

controversial in the case of IMM (White *et al.* 2008) and was thus a poor argument of the hybrid model quality. This point will be developed in discussion.

The third model filled 12 sources (11 rivers of the baseline and 1 extra-source) whereas the first model filled 15 sources of the baseline. More precisely, the Vire Vilaine and Adour were excluded from potential sources from the third model. However, those rivers are known to be important reproduction sites for Allis shad (www.sage-vire.fr; Elie & Baglinière 2000). In parallel, we found strong confusion of reallocation between the Vire, Vilaine and Loire rivers in the third model, which could explain that individuals reallocated in Vire and Vilaine rivers in the first model were finally reassigned in the Loire River in the third model. Therefore, considering that both convergence and ecological failure are strong limits, we decided to keep the first model to investigate the functioning of the metapopulation of Allis shad. Another advantage of this model is that the probability of origin depends on the catch year, which could be useful to examine temporal flux dynamics between sources and sinks.

3.3. Functioning of the metapopulation

3.3.1. Recipient, donor and closed rivers

Because of an incomplete sampling scheme, flux calculation was performed only for 2013 in the following subsections. In 2013, sampled were available for all catch rivers (i.e. recipient rivers) except for the Nivelle River. Thus, this river could not be considered as a recipient river in 2013 and was excluded from flux calculation. However, it was kept as potential donor rivers. Besides, because Allis shad does not spawn in estuary, fishes sampled in the Adour estuary could potentially have chosen the Adour, Oloron or Nive rivers to reproduce. Therefore, the Adour estuary could not be considered as a recipient river and was then excluded from subsequent analysis.

To compare the proportion of produced and received fishes per river, we considered only rivers which are simultaneously recipient and donor rivers. Thus, we estimated flux between 13 recipient rivers (i.e. rivers where spawners were sampled in 2013) and 17 donor rivers (i.e. rivers of the water baseline) by multiplying abundance estimates with probabilities of origin $\{\theta_{c(ad),2013(ad)}\}$ (Figure A.6). We first calculated the proportion of strayers (i.e. incoming fishes) and homing per recipient river. Then we calculated the proportions of strayers produced by each donor river (i.e. outgoing fishes). We finally defined the proportions of total outgoing fishes (i.e. % Total Outgoing) as the ratio between the number of outgoing fishes produced by a donor river (i.e. N_{out}) and the sum of the incoming fishes received by a recipient river (i.e. N_{inc}) and the outgoing fishes produced by a donor river (i.e. N_{out}) (Table 10). It allowed the identification of sources (i.e. rivers which produced more spawners than received) and sinks (i.e. rivers which received more spawners than produced).

Table 10 Total number of fishes estimated for each recipient (N_{totR}) and donor river (N_{totD}) and their relative and effective number of strayers (incoming and outgoing fishes). Incoming and outgoing fishes correspond to straying fishes (i.e. the number of homing fishes are removed from the number of incoming and outgoing fishes). The incoming fishes are received by a recipient river. The outgoing fishes are produced by a donor river. The proportions of total outgoing fishes were calculated to show which river is a source or a sink. Total straying fishes (N_{totout}) corresponds to the sum of incoming (N_{inc}) fishes and outgoing fishes (N_{out}).

Catch sites	Recipient					Donor			Recipient and Donor	
	Homing fishes			Strayers (=incoming fishes)		Strayers (= outgoing fishes)			Total Strayers	
	N_{totR}	N_{home}	%	N_{inc}	%	N_{totD}	N_{out}	%	N_{totout}	% Total Outgoing
Adour R	1041	1013	97.3	**28**	2.70	2396	**1383**	57.7	**1411**	98.0
Aulne	1964	1964	100	**0**	0	1970	**6**	0.30	**6**	100
Blavet	1628	1628	100	**0**	0	4324	**2697**	62.4	**2697**	100
Dordogne	2258	2229	98.7	**29**	1.30	4323	**2095**	48.5	**2124**	98.6
Garonne	2114	2	0.01	**2112**	99.9	644	**642**	99.7	**2754**	23.3
Lima	1235	613	49.6	**622**	50.4	613	**0**	0	**622**	0
Loire	1345	735	54.6	**610**	45.4	735	**0**	0	**610**	0
Minho	1896	1896	100	**0**	0	3821	**1925**	50.4	**1925**	100
Mondego	1442	0	0	**1442**	100	0	**0**	0	**1442**	0
Saison	764	0	0	**764**	100	0	**0**	0	**764**	0
Scorff	2250	0	0	**2250**	100	0	**0**	0	**2250**	0
Vilaine	1664	1231	74	**433**	26.0	1273	**42**	3.30	**475**	8.84
Vire	1911	1899	99.4	**12**	0.60	1899	**0**	0	**12**	0

Besides, contributions of each river to the total production of spawners in the metapopulation were calculated using the number of fishes produced per donor river (i.e. homing and straying fishes) relatively to the total number of fishes produce in the metapopulation (Figure 10).

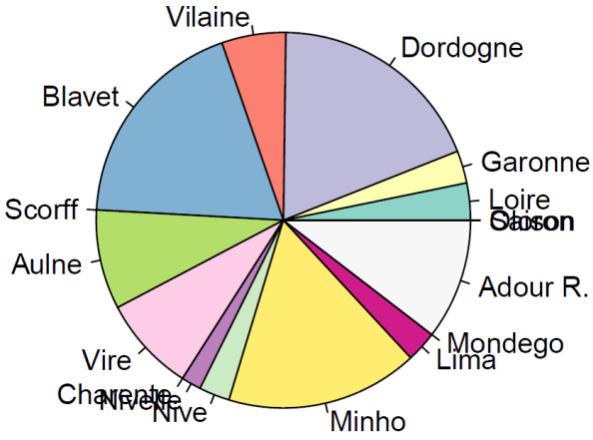

Figure 10 Contribution of each donor river to the total production of spawners in the metapopulation. Productions were estimated using the median of flux distributions, considering that distributions were symmetrical around the median.

Sink rivers

Strict straying (100% of individuals) occurred in the Scorff, Saison and Mondego rivers (Table 10). Those rivers didn't produce any fish (neither homing nor straying). Thus, those rivers were considered to be sink rivers. Obviously, low contributions to the total number of spawners in the metapopulation were found for those rivers (Figure 10).

The Loire and Lima rivers were also sink rivers. Any fish was exported from those rivers (i.e. any straying fish produced) but similar proportion of homing and straying fishes were received. Besides, low contributions to the total number of spawners were found for those rivers (around 3%, Figure 10).

Source rivers

The Minho, Blavet, Adour and Dordogne rivers were the most important sources. Since those rivers received low or zero straying individuals, they produced both homing and straying fishes (Table 10). Strict homing (100% of individuals) was found, for the Blavet and Minho rivers. High homing rates were also found for the Adour and Dordogne rivers (> 97% of individuals). In parallel, we found strong contributions to the total number of spawners for the Dordogne (19%), Blavet (19%), Minho (17%) and Adour (10%) rivers (Figure 10). Those contributions appeared to be the most important among all donor rivers.

Closed populations: Vire and Aulne rivers

The Vire River appeared as a closed system which means that it produced almost exclusively homing fishes (99.4%) and received no fishes from other rivers (Figure 10). Besides, the Vire River showed important contribution (8%) to the total number of spawners in the metapopulation. It was the same case for the Aulne River which received only homing fishes (100%) and produced very few straying individuals (0.30%). Despite this river appeared as a closed system, it largely contributed to the total number of spawners in the metapopulation (9%).

Focus on Garonne River

The Garonne River received high proportion of strayers (99.9%), which means that low homing occurred in this river (Table 10). The probabilities of homing from 2001 to 2014 were calculated from the median of the probabilities of being originated from the Garonne when caught in this river (not presented here). We found null homing probabilities for the Garonne River through the time series. Besides, the Garonne River showed low contribution to the total number of spawners (3%).

Therefore, different types of rivers were encountered: the exclusively donor rivers (i.e. sources) which highly contributed to the production of spawners (the Minho, Dordogne and Blavet rivers), the exclusively recipient rivers (i.e. sinks) which showed low contributions to the total number of spawners (the Scorff, Saison and Mondego rivers) and the closed rivers which produced important proportion of spawners (the Vire and Aulne rivers). The Garonne River appeared as a particular case because no homing occurred whereas straying fishes are received and produced.

3.3.2. Origin of fishes per recipient river

The origin of straying fishes was investigated to understand the dispersal capacity of Allis shad across the distribution range. The origin and proportions of fishes received in each recipient river (except the Nivelle River) are presented in Figure 11.

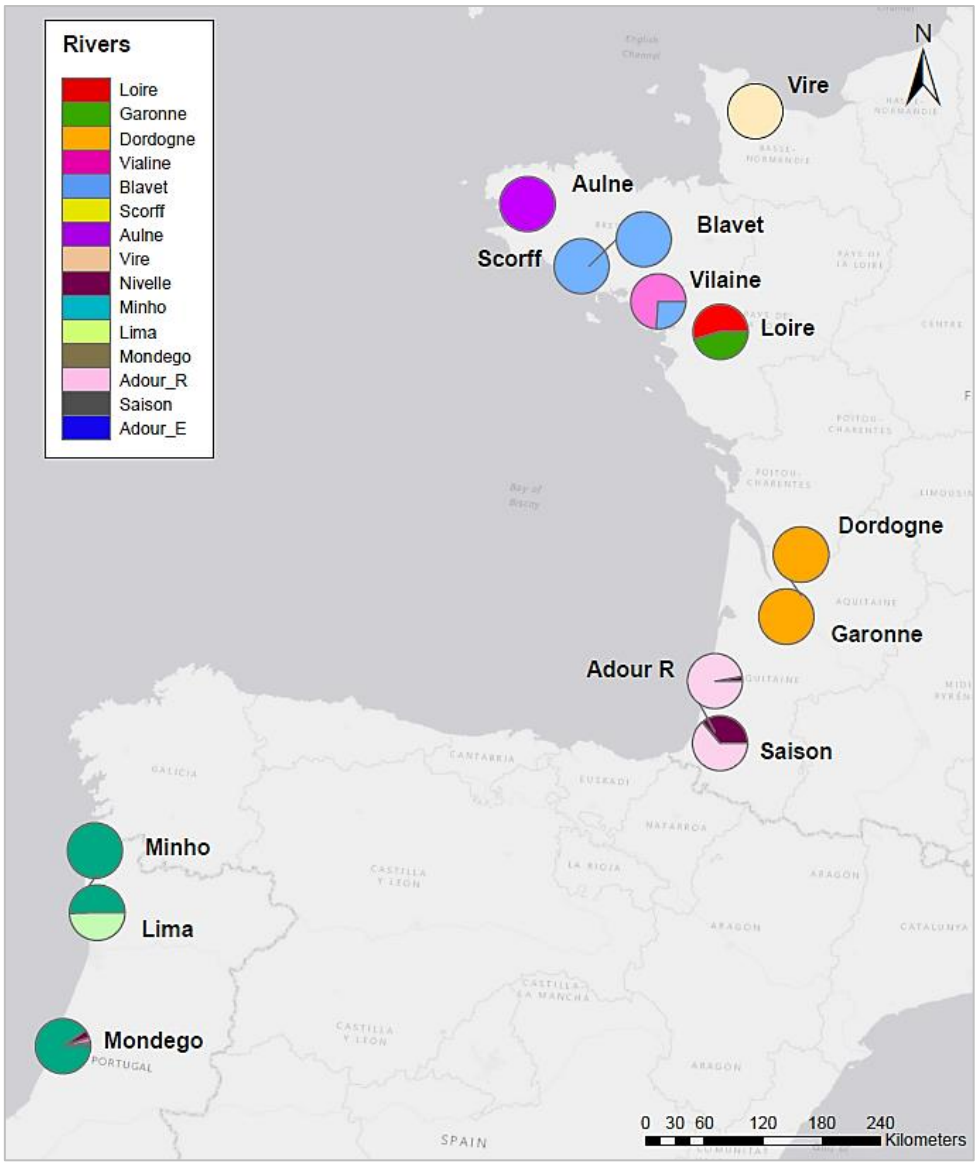

Figure 11 Proportions of fish origin per recipient river (N=13). The pies were performed using the median of the flux distributions, considering that distributions were symmetrica around the median.

Exchanges between donor and recipient rivers occurred mostly at watershed scale (i.e. between rivers having the same estuary). It is the case between the Blavet (i.e. the donor river) and the Scorff (i.e. the recipient river) which has common estuary. The Blavet River was also a source for the Vilaine River which is spatially close. Homing at watershed scale also occurred between the Adour and Saison rivers and in Portugal where the Minho River contributed to the Lima and the Mondego rivers. The Garonne River received exclusively strayers from its neighbor river Dordogne and produced fishes which strayed to the Loire River and in a small proportion to the Mondego River.

3.3.3. Isolation by distance between donor and recipient rivers?

To examine long-distance dispersal capacity of Allis shad, isolation by distance was tested through the relation between flux (i.e. median of flux distributions) and distances between donor and recipient rivers (Figure 12).

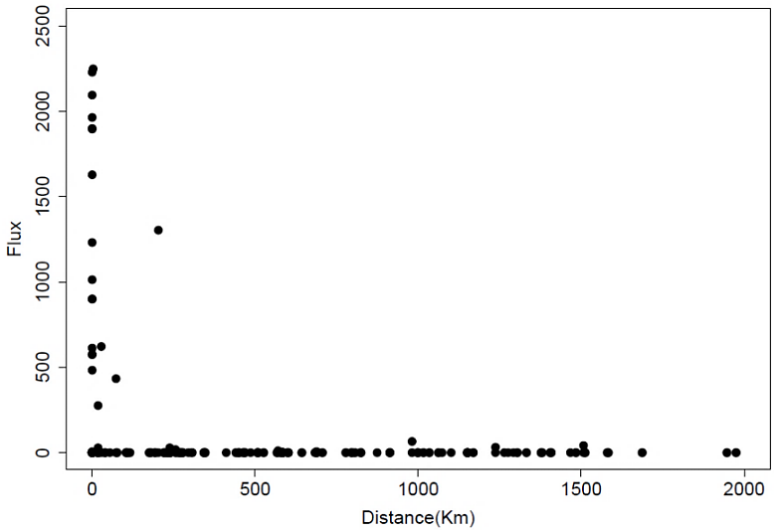

Figure 12 Relation between flux of spawners and distance (km) between donor and recipient rivers. Flux correspond to the median of the flux distributions because they were considered to be symmetrical around the median.

Strong isolation by distance occurred from 250 km between donor and recipient rivers. The intensity of exchanges decreased substantially, even if some exchanges (7%) are displayed at long distances (longer than 250 km) from the donor river to the recipient river. Indeed, a few proportion of fishes performed long-distance stray from the Vilaine (1507 km), Garonne (1237 km) and Nivelle (981 km) rivers to the Mondego river (3% of strayers received in the Mondego river).

3.3.4. Exchanges between the North and the South

Within the Allis shad metapopulation, fishes mostly displayed homing (Figure 13). The straying occurred above all at watershed scale (i.e. between rivers having the same estuary). However, some individuals migrated at long distance and thus, realized southward or northward migrations. Fishes migrated out of their watershed in equivalent proportion towards the South and the North part of the distribution range (Figure 13).

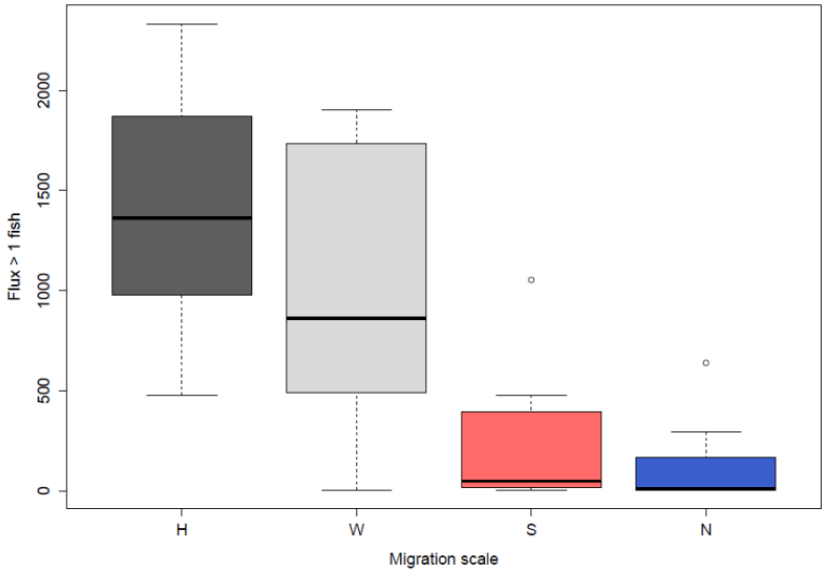

Figure 13 Flux (≥ 1 fish) between donor and recipient rivers considering migration scale. H: homing at river scale, W: homing at watershed scale, S: southward migration, N: northward migration.

Mean flux towards the North and the South were not significantly different (Wilcox test; W = 6797; p-value = 0.4711). Individuals that performed northward migration correspond mostly to fishes which stray from the Garonne to the Loire rivers. In opposite, Allis shad performing southward migrations are mostly strayers from the Blavet to the Vilaine rivers and from the Minho to the Lima rivers.

As we found significant flux between rivers through the distribution range of Allis shad, we conclude that isolated rivers are connected by dispersal.

4. Discussion

In a context of global decline of Allis shad populations through its entire distribution range, the understanding of the whole populations dynamic is an important concern to understand whether this species is distributed in isolated populations or have significant exchange rates between populations (Kritzer & Sale 2004). Herein, we proposed to investigate the metapopulation functioning of Allis shad by inferring natal origin from otolith microchemistry within Bayesian models of reallocation. The information of the natal origin is of interest as it allows the estimate of exchange rates between populations and thus the prioritization of river conservation and restoration efforts.

4.1. Why and how comparing the three models?

We aimed to develop a Bayesian model that combined the advantages from two different kinds of Bayesian models. We first experienced a classical Bayesian model based on the model from Martin *et al.* (2015). In this first approach, in absence of exhaustive otolith training data-set (i.e. otolith signatures with known natal origins available for all potential sources), a water baseline was assumed to be exhaustive (i.e. the chemical signatures of water samples from all potential sources) and used with the limited otolith training data-set to fit a model to reallocate Allis shad. Because catching juveniles Allis shad is difficult, the use of water samples in the reallocation process is an easier alternative which provides a more complete reference atlas of potential sources. We reallocated 79% of adults Allis shad with a high reliability (MPR $\geq$ 0.80), and only 3% of fishes were reallocated with low reliability (MPR $\leq$ 0.40). However, the assumption of exhaustiveness in water baseline and the existence of potential extra-sources may alter the inferences. Thus, as proposed by Martin *et al.* (2015), we chose to explore an alternative model that does not require the exhaustiveness assumption but does not use any baseline. Then, we provided a Bayesian hybrid model that combined the two first approaches.

The Dirichlet Process Model (DPM) belongs to the family of Infinite Mixture Model (IMM), which are particularly relevant to distinguish origin of individuals in a mixed sample and to estimate the number of sources (Munch & Clarke 2008). Rasmussen (1999) pointed out several advantages of IMM, such as they do not limit the number of groups (i.e. sources) to be defined and they determine automatically this number. Based on this approach, Neubauer *et al.* (2013) showed that the DPM was a particularly relevant tool to reallocate some individuals in groups out of the baseline. However, they concluded that the use of DPM without baselines was not efficient, especially because of poor model convergence. They pointed out the need of using informative priors, i.e. an informative baseline. We fitted a DPM and had the same results: our second model (namely Infinite Mixture Model) showed very poor convergence. Thus, we finally decided to exclude this model from or analysis, considering that interpretations of reallocation would be irrelevant and misleading. The use of DPM in our study was just an intermediate step allowing the construction of the hybrid model. We assume that the association of a source with a river was impossible in the case of IMM and would not be a relevant way to

investigate the functioning of the metapopulation, even if the convergence would have been better.

In this context, we set up a Bayesian hybrid model similar to a second model developed by Neubauer *et al.* (2013): they combined a DPM with a baseline of otolith composition of juveniles weakfish, to reassign adults to their natal sites. Herein we combined an incomplete juvenile baseline with a water baseline.

Here, the Bayesian hybrid model reallocated fishes in both identified sources (i.e. rivers of the water baseline) and unidentified sources (i.e. extra-sources) according to the MPR. We found that reassignments in rivers of the baseline concerned 99% of adults Allis shad, with a high reliability (MPR $\geq$ 0.90). This tends to demonstrate that IMM was probably not necessary, at least in our specific case study.

The Bayesian hybrid model suggested the existence of 4 extra-sources and 4 individuals were reallocated in one of those extra-sources, according to the MPR. Those results should be considered with great caution, first because of low reallocation probabilities for those fishes, and more importantly, because of poor convergence for those fishes. Moreover, since extra-sources are not based on any existing baseline, there might be some problems of identifiability (an extra-source can be "named" source 1 by a chain and source 2 by another, inducing switch between labels) (White et al. 2008). However, fishes assigned in extra-source all have extreme signature (particularly in Sr/Ca). The hybrid model was able to detect individuals with signatures inconsistent with the baseline. Before concluding to the presence of spawning sites out of the baseline, alternative hypothesis should be investigated. As explained by White *et al.* (2008), "Does an outlying value represent measurement error, process error (i.e. extreme natural variation) or the presence of an additional source?". The first hypothesis is that extreme signature in Sr/Ca could be due to a measurement error. Indeed, the core of the otolith was excluded from the laser ablation because of maternal effect. However, considering that error in ablation location could occurred in microchemistry analysis, one could not exclude a contamination of the sample by maternal signature, inducing a high ratio in Sr/Ca due to marine habitat previously experienced by the female (Volk et al 2000). On the other hand, high Sr/Ca ratio could be due to the combination of physiological (Sadovy & Severin 1992; Limburg 1995) and environmental factors (Chesney *et al.* 1998; Bath *et al.* 2000). Since the water is known to contribute to more than 80% of the otolith composition (Bath *et al.* 2000), food (Kennedy *et al.* 2000), stress (e.g. inducing otolith checks characterize by high Sr/Ca ratio; Campana & Neilson 1985) or environmental variables, such as the water temperature (Kalish 1991; Townsend *et al.* 1992; Bath *et al.* 2000; Martin *et al.* 2004) could impact the Sr deposition in otoliths. Thus, exclusion of outliers with extreme signatures could occur whereas they could have just experienced different physiological and environmental conditions. Therefore, further investigations would be necessary before concluding to the existence of non-sampled spawning rivers.

Before performing the model of exchange flux, we chose the first model, excluding the Bayesian hybrid model. Many factors have led us to this choice. The most important criterion was the better convergence of parameters of the first model. Indeed, a large proportion of parameters in the third model did not converge according Gelman-Rubin diagnosis. In particular, hyper-parameters such as α or $[\sum]$ did converge neither in the second model (i.e. IMM) nor in the third model (i.e. hybrid). Besides, convergence failures were found for the probabilities of origin (using the Gelman-Rubin test) and for the categorical variable of reallocation (using visual inspection of chains mixing), especially for fishes reallocated in extra-sources. Finally, we pointed out that the use of DIC as criterion to compare models is controversial. Indeed, in their study White *et al.* (2008) experienced several Bayesian IMM to found the optimal number of sources to be estimated. They used the DIC to compare models and select the better model in terms of overall fit. As generally extra-sources have few

individuals reallocated in, the mean Bayesian deviance is minimized when individuals are clustered in extra-source. Therefore, the model selection using the DIC was not a relevant criterion to select the most appropriate model of reallocation. Moreover, some results of the hybrid model were not ecologically plausible. In conclusion, despite theoretical advantages of the hybrid model, all the previous arguments led us to select the first model. Despite we combined statistical and ecological criteria to select the best model, a validation data-set would be required to truly select the best model.

4.2. Recommendations regarding a monitoring program of Allis shad

Contrary to classical methods of classification usually performed to infer natal origin from otolith microchemistry (such as Discriminant Analyses or machine learning methods including Random Forest or Artificial Neural Network; Gillanders & Kingsford 2000; Brown 2006; Hobbs *et al.* 2007), which require large and exhaustive training data-sets, our Bayesian approach used multiple and incomplete baselines to reallocate fishes. However, completing our baselines would provide multiple benefits.

Improving adult sampling and monitoring effort per year and river

First, the adult data-set was not balanced between rivers and years. For example, one can wonder if the strict homing found in the Nivelle River, where only 7 adults were sampled, reflects an exhaustive homing or a simple "breeding aggregation effect". Potential density-dependent reproductive behavior could be involved in the homing process, as found for salmonids (Quinn & Fresh 1984). Allis shad is a Clupeidae, and thus, this species could exhibit a schooling behavior inducing breeding aggregation. In order to ensure that the strict homing rate found in the Nivelle River was a representative pattern of the entire population, more samples are required.

Besides, adult sampling years were not balanced. Indeed, adults were mainly sampled in 2013. In order to study the temporal dynamic of the metapopulation, samples from each year and river are needed. Besides, the monitoring of spawners in all potential spawning rivers, each year, would allow a more precise estimation of flux. In this way, we would be able to enhance homing and straying rates estimates over the years and thus we would investigate temporal and spatial metapopulation functioning. However, cautions are required before analyzing new data-sets. Indeed, herein we reallocated adults Allis shad sampled in 2001 in the Garonne River using water samples from 2012 to 2013. As this species usually performs it spawning migration between 3 and 7 years (Lochet 2006), fishes sampled in 2001 were probably hatched between 1994 and 1998. In our case, we used $^{87}Sr/^{86}Sr$, Sr/Ca and Ba/Ca to reallocate fishes. Despite it was not possible for us to check the temporal stability of the water signature, the use of $^{87}Sr/^{86}Sr$ was relevant because it is known to be stable over the years (Kennedy *et al.* 2000; Walther & Thorrold 2006; Walther & Limburg 2012). Besides, in their study, Martin *et al.* (2015) found that water chemistry was relatively similar to that used by Tomas *et al.* (2005) in 2001 in the Garonne, Dordogne and Adour rivers. Thus, in the present study, we assume that minor temporal changes in water signature did not alter reallocation of adults Allis shad. However, as we were not able to test this presumed temporal stability, both water and adult data-sets are required in future sampling programs. We propose to use elements which are temporally robust instead of introducing temporal effect in the otolith composition, which would introduce complexity in the model.

Extending water baseline coverage and improving rivers discrimination

We propose to improve the spatial coverage of reference data-sets using a larger water sampling across the entire distribution range, in order to improve the reallocation reliability and avoid using a Bayesian hybrid model.

In our study, all potential spawning grounds were not sampled in each river of the water baseline. Since water body could be highly variable between distant spawning sites in a river due to contrast in bedrock geology, individuals from the same river could have experienced different water composition. For example, the Loire River is the largest watershed in France whereas in the present study only 3 spawning sites were included in the water baseline. Thus, geographical coverage could be improved by introducing tributary rivers (such as the Vienne, Creuse and Mayenne rivers in the Loire Basin) in the water baseline where Allis shad may potentially reproduce, as in the Adour watershed where we are able to distinguish 3 tributaries. Reallocation at finer scale (i.e. at spawning site scale) could be accurate in case of heterogeneity in water body signatures in different spawning grounds of the same river. The addition of alternative spawning sites in small rivers such as the Trieux and Elorn rivers in Brittany and the Sèvre Niortaise River in the South of France would also improve the quality of the water baseline.

A possible solution would be to combine models of reallocation with spatial map using hydrologic and geologic data. In their study, Barnett-Johnson *et al.* (2008) evaluated the efficacy of $^{87}Sr/^{86}Sr$ to infer natal origin of salmon using spatial map in geographic information systems. It would be a way to provide exhaustiveness in water baseline through the distribution of Allis shad.

Our study pointed out confusion of reallocation between geographically close rivers (such as the Nive, Saison and Oloron rivers). The addition of ultra-trace elements with $^{87}Sr/^{86}Sr$ and Sr/Ca and Ba/Ca ratios have the potential to increase the discrimination between sources, provided such addition is meaningful and wise regarding discrimination power (Mercier *et al.* 2011). The relevance of elements as geospatial indicators of spawning sites excludes the choice of elements that are under strong physiological control, which could reflect physiological changes more than discrimination between sources (Walther & Thorrold 2008). Temporal variability in the uptake of some trace elements could also introduce noise in or analysis, and thus caution is required before introducing new elements in the analysis.

Ultra-trace elements sometimes under the Limit Of Detection (LOD), are often excluded of the studies because there are not always measurable. Higgins *et al.* (2013) suggested that data below the LOD could contribute meaningfully to the discrimination between rivers, using the presence/absence information to discriminate rivers with overlapping signatures. In their study, De Pontual *et al.* (2000) examined the discriminatory power of several elements. They found that Pb was under the LOD for majority samples but was detected in samples from the Loire estuary. Therefore, we suggest improving discrimination between sources using geospatially meaningful additional ultra-trace elements to be treated in a mix of quantitative and qualitative data.

In our study, we pointed out that the Bayesian hybrid model would have reallocated individuals in rivers of the baseline instead of extra-sources because of “classical” otolith signature. Therefore, increasing the discrimination between reference sources would increase the reallocation reliability of the hybrid model.

Completing juvenile baseline to improve reallocation and validate the models

In our study, the water baseline was the main driver of natal origin inferences. As we did not used juveniles from each river of the water baseline to reallocate adults, the juvenile baseline had limited interest. A complete juvenile sampling scheme is crucial so that the reallocation does not only arise from the linear relation between the otolith and water composition, and we would be able to account for inter-river variability in the otolith composition.

Besides, as we explained previously, the model choice and the interpretation of natal origin would be improved by a validation method. One of the most relevant ways to validate models of reallocation is the use of juveniles as training data. Indeed, as the Allis shads reproduce in

the middle watercourse of the rivers, the natal origin of juveniles sampled in freshwater is already known. Therefore, we propose to sample juvenile in each river of the water baseline to test the reallocation performance of our models. We propose a new way to undertake determination of origin of fish using otolith chemistry, by using exhaustive water sampling as a baseline data set and otolith composition of juveniles as a validation data set.

Therefore, an adequate monitoring program and a model validation would provide a better interpretation and estimation of the homing and straying rates.

4.3. Strict homing, complete straying or limited diffusion: evidence and consequences

In their study, Martin *et al.* (2015) were not able to conclude exchanges rates, and consequently homing and straying rates, because abundance indices were not available. In the present study, we went one step further by coupling the results of reallocations with abundance indices to achieve an estimation of exchange and homing rates.

Variability in homing rate among rivers

Diadromous fishes display two opposite behaviors: straying and homing (Quinn 1984). Straying, or dispersal, is an opportunity to experience richer and sustainable habitats. It also impairs adaptation to specific freshwater habitat (Kritzer & Sale 2004). Conversely, homing facilitates the emergence of locally adapted populations, providing benefits when environmental conditions are constant through time (Keefer & Caudill 2014). The definition of homing is not clear enough for Allis shad, and we actually cannot conclude whether the natal site fidelity is due to specific orientation system (i.e. sun, orientation, polarized light, geomagnetism, temperature, currents and olfaction; Kritzer & Sale 2004) similar to that of salmon and trout species or if natal homing is the result of limited dispersal capacity (i.e. fishes would come back to their natal site because they do not travel enough time to reach other spawning rivers). In salmonids, the olfactory imprinting is the primary mechanism involved in the homing (Dittman & Quinn 1996). It occurs during the juvenile stage at natal site, rearing sites and along migration routes (Keefer & Caudill 2014). This phenomenon is related to olfactory odor detection, hormonal activities and memories (Ueda 2012). The existence of such a guiding mechanism underlying the homing behavior is very important for management. For example, for salmon restocking, the existence of this mechanism imposes that young salmons should be released young enough to be imprinted by the water source. It also impairs some management measures such as assisted migrations which are often proposed to mitigate the effect of climate change.

Herein, we found various degree of homing among rivers. On the one hand, the Vire, Vilaine, Blavet, Minho, Dordogne and Adour rivers exhibited high homing rates. On the other hand, strict straying was found in the Scorff, Saison and Mondego rivers. Between those extremes, homing and straying were well balanced in other rivers, such as the Loire and Lima rivers. Three hypotheses could be proposed to explain the variability in homing behavior among rivers. First, the homing behavior could be preferred in stable high quality habitat because it would provide high juvenile survival (Quinn 1984; Hendry *et al.* 2004). It would be interested to explore whether rivers with high straying rates (e.g. sinks) correspond to highly degraded and fragmented rivers. On the other hand, the choice of the spawning river could be density-dependent. Indeed, as explain previously, the Allis shad is a Clupeidae species with an aggregative behavior which could lead adults to choose their spawning river according to the population size. Consequently, large populations would present high homing rate contrary to small populations where straying would be prevalent. Then, homing could be mainly due to low dispersal capacity.

Therefore, more researches are needed to understand the underlying mechanism inducing homing behavior for Allis shad.

Straying behavior and isolation by distance

To explore the former explanation, i.e. high homing rate due to limited diffusion capacity, we had a look at the estimated distance between natal and spawning rivers. We found that most straying shads display limited distance movements, with a strong homing at basin scale. This is consistent with the results brought by Alexandrino *et al.* (2006) and Jolly *et al.* (2012) which performed genetic analyses and found significant isolation by distance for Allis shad, so that gene flow primarily takes place between neighboring rivers. Combining genetic and microchemistry tracers, Martin *et al.* (2015) showed high site fidelity with a weak genetic differentiation between populations, indicating that moderate straying was sufficient to standardize genetic structuration among populations. In our study, a few proportions (i.e. 3%) of flux between Portugal and France seemed to be sufficient to homogenize Portuguese and French populations. Lassalle *et al.* (2008) suggested that potential marine growing areas are probably located near the watershed of origin. Adults would thus migrate from these areas to neighboring watersheds, inducing a homing at basin scale. Moreover, compared to salmonids, alosines spend only few months in the natal river, limiting the possibility of imprinting (Kritzer & Sale 2004). In this context, it is not possible to exclude that our results, i.e. strong return to natal rivers, is not only due to limited diffusion during the marine growth phase.

Martin *et al.* (2015) suggested combining straying rates with population relative abundance of spawners to highlight northward dispersion tendency caused by climate change. However, our study did not confirm this hypothesis as we did not found a clear northward migration pattern. Indeed, the most important part of northward flux corresponded to straying from the Garonne River to the Loire River. However, we found confusion of reallocation between the Garonne and Loire rivers; therefore we cannot conclude about the exchanges between those rivers.

4.4. Metapopulation functioning: sinks and sources and implication for management

Metapopulation or isolated populations?

The present study proposed to examine metapopulation structure of the European Allis shad through a model of exchange between populations. Up to now, we are unaware of any study dealing with the metapopulation functioning for this species. As explained by (Schtickzelle & Quinn 2007), the question of metapopulation dynamics in anadromous fish is not well studied and remained marginal. Nevertheless, metapopulation dynamic is of interest to management through the prioritization of watershed conservation, restoration and fisheries regulation. Populations form a metapopulation if populations are well separated from each other and connected by dispersal (Schtickzelle & Quinn 2007). Given our results, Allis shad seems to fulfill the criteria. We thus conclude that rivers are sub-populations connected by some degree of exchange, resulting in a metapopulation functioning.

Identification of source, sink and closed populations: implication for management

We found that the Scorff, Saison and Mondego rivers acted only as sinks and do not contribute significantly to the total number of spawners in the metapopulation. This suggests that either the spawning grounds are of poor quality or hardly accessible, or that juvenile survival is low. We also pointed out important sources, such as the Blavet, Dordogne, Minho and Adour rivers. Intermediate rivers, such as the Loire and Lima rivers, presented homing and straying in the same proportion. Finally, we also found two isolated populations, the Vire and Aulne rivers which appeared as important sources for the metapopulation.

Therefore, in terms of prioritization of watershed conservation and restoration, we proposed to focus on the most productive rivers (the Blavet, Dodrogne, Minho and Adour rivers) and the isolated populations which highly contribute to the production of spawners (the Vire and Aulne rivers). Besides, to understand whether strayers have positive (i.e. demographic gain, resilience, genetic diversity) or negative effects (i.e. demographic stochasticity, adult mortality in the new environment, competition, low reproductive success, sex-ratio imbalance) on local populations (Keefer & Caudill 2014), supplementary studies on the recruitment success for example, are needed.

5. Conclusion and prospects

Despite an increasing use of Bayesian models of reallocation in recent years (White *et al.* 2008; Neubauer *et al.* 2013; Martin *et al.* 2015), we are unaware of any previous study that combined the outputs of a Bayesian model of reallocation with abundance estimates of Allis shad spawners to develop a model of exchange between basins. We finally showed a metapopulation functioning for Allis shad with a low magnitude of long dispersal migration. We also proposed a prioritization of watersheds to be protected, using the source and sink rivers concept. More studies are required to understand why populations acted as sinks or sources (e.g. differences in accessibility, habitat quality or reproductive success).

We used a Bayesian model of reallocation with fixed number of sources instead of an Infinite Mixture Model with baseline to develop the model of exchange between watersheds. Indeed, the hybrid Bayesian model was not performing in our case and we proposed to improve the monitoring program for Allis shad instead of using a hybrid model which could be useful when acquisition of data is too difficult.

In the future it would be interesting to use the approach developed in the present study to monitor the entire metapopulation of Allis shad. Herein, we focused on the spatial metapopulation dynamic in 2013 to investigate the exchanges between watersheds. We propose to improve the monitoring program of Allis shad to study the spatial and temporal dynamics of the metapopulation. We suggest developing a standard assessment method of the abundance of spawners on each river across the distribution range. Besides, because the hybrid model is not performing in our case, improving the spatial coverage of the water baseline is required. The model validation is also required and we propose to sample juveniles in each spawning river across the distribution area to build a training data set. In practice, we suggest sampling several adult, water and juvenile samples per spawning grounds in each river, each year, to provide a spatial and temporal monitoring of the Allis shad metapopulation.

In the future, this kind of analysis would allow the investigation of Allis shad northward migration in response to global change. Indeed, Lassalle *et al.* (2009) proposed that the climate change would be favorable for the Northern populations. Herein, however, we did not found a clear pattern of northward migration, but we pointed out two isolated populations, the Vire and Aulne rivers. A few years ago the Vire River population was anecdotic whereas this river appeared to be extremely attractive now. Besides, exchanges with British populations are probable for the Vire population. Thus, one can wonder if the actual Northern populations are the result of high quality conditions or whether immigrants had strayed and settled. Therefore, in the future, the kind of analysis we developed here must be associated with analysis of the reproductive success of Allis shad in order to eventually point out replacement difficulties caused by density-dependent effects (Rougier *et al.* 2012) or others. Considering that the European Allis shad is a relatively warm water species, the climate change would probably not impact the metapopulation functioning in the same magnitude as other anadromous species, but this assumption remained to be study.

6. References

Alexandrino, P. J., & Boisneau, P. (2000) Diversité génétique. *Les aloses Alosa alosa et Alosa fallax spp*. 179-198.

Alexandrino, P., Faria, R., Linhares, D., Castro, F., Le Corre, M., Sabatié, R., Baglinière, J. L. & Weiss, S. (2006) Interspecific differentiation and intraspecific substructure in two closely related clupeids with extensive hybridization, *Alosa alosa* and *Alosa fallax*. *Journal of Fish Biology* **69**: 242–259.

Aprahamian, M. W. & Environment Agency (2003) *Alosa alosa and Alosa fallax spp.: literature review and bibliography*. 374 pp.

Baglinière, J.-L., Sabatié, M. R., Rochard, E., Alexandrino, P. & Aprahamian, M. W. (2003) The allis shad *Alosa alosa*: Biology, ecology, range, and status of populations. *American Fisheries Society Symposium* **35**: 85–102.

Barnett-Johnson Rachel, Thomas E. Pearson, Frank C. Ramos, Churchill B. Grimes & R. Bruce MacFarlane (2008) Tracking natal origins of salmon using isotopes, otoliths, and landscape geology. *Limnology and Oceanography* **53**: 1633–1642.

Bath, G. E., Thorrold, S. R., Jones, C. M., Campana, S. E., McLaren, J. W. & Lam, J. W. (2000) Strontium and barium uptake in aragonitic otoliths of marine fish. *Geochimica et cosmochimica acta* **64**: 1705–1714.

Blum, J. D., Taliaferro, E. H., Weisse, M. T. & Holmes, R. T. (2000) Changes in Sr/Ca, Ba/Ca and $^{87}Sr/^{86}Sr$ ratios between trophic levels in two forest ecosystems in the northeastern USA. *Biogeochemistry* **49**: 87–101.

Brooks, S. P., & Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics* ***7***: 434-455.

Brown, J. A. (2006) Using the chemical composition of otoliths to evaluate the nursery role of estuaries for English sole *Pleuronectes vetulus* populations. *Marine Ecology Progress Series* **306**: 269–281.

Campana, S. (1999) Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Marine Ecology Progress Series* **188**: 263-297.

Campana, S. (2001) Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**: 197–242.

Campana, S. E. & Neilson, J. D. (1985) Microstructure of Fish Otoliths. *Canadian Journal of Fisheries and Aquatic Sciences* **42**: 1014–1032.

Campana, S. E., Chouinard, G. A., Hanson, J. M., Frechet, A. & Brattey, J. (2000) Otolith elemental fingerprints as biological tracers of fish stocks. *Fisheries Research* **46**: 343–357.

Carry, L. & Borie, G. (2013) Suivi de la reproduction de la grande alose sur la Garonne en 2012. Rapport MIGADO 6GRT-11. 26 pp.

Castelnaud, G., Rochard, E. & Le Gat, Y. (2001) Analyse de la tendance de l'abondance de l'alose *Alosa alosa* en gironde à partir de l'estimation d'indicateurs halieutiques sur la période 1977-1998. *Bulletin Français de la Pêche et de la Pisciculture* **362–363**: 989–1015.

Chesney, E. J., McKee, B. M., Blanchard, T. & Chan, L.-H. (1998) Chemistry of otoliths from juvenile menhaden *Brevoortia patronus*: evaluating strontium, strontium: calcium and strontium isotope ratios as environmental indicators. *Marine Ecology Progress Series* **171**: 261–273.

Daverat, F., Morais, P., Dias, E., Babaluk, J., Martin, J., Eon, M., Fablet, R., Pécheyran, C. & Antunes, C. (2012) Plasticity of European flounder life history patterns discloses alternatives to catadromy. *Marine Ecology Progress Series* **465**: 267–280.

Dawson, K. J. & Belkhir, K. (2009) An agglomerative hierarchical approach to visualization in Bayesian clustering problems. *Heredity* **103**: 32–45.

De Groot, S. J. (2002) A review of the past and present status of anadromous fish species in the Netherlands: is restocking the Rhine feasible?. *Ecological Restoration of Aquatic and Semi-Aquatic Ecosystems in the Netherlands* **478**: 205-218.

De Pontual, H., Lagardere, F., Troadec, H., Batel, A., Desaunay, Y. & Koutsikopoulos, C. (2000) Otoliths imprinting of sole (Solea solea) from the Bay of Biscay: a tool to discriminate individuals from nursery origins? *Oceanologica acta* **23**: 497–513.

Dittman, A. & Quinn, T. (1996) Homing in Pacific salmon: mechanisms and ecological basis. *Journal of Experimental Biology* **199**: 83–91.

Elie, P. & Baglinière, J.-L. (2000) *Les aloses (Alosa alosa et Alosa fallax spp.): Écobiologie et variabilité des populations*. Quae. 277 pp.

Elsdon, T. S. & Gillanders, B. M. (2002) Interactive effects of temperature and salinity on otolith chemistry: challenges for determining environmental histories of fish. *Canadian Journal of Fisheries and Aquatic Sciences* **59**: 1796–1808.

Gelman, A. & Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences. *Statistical science* **7**: 457–472.

Gillanders, B. M. & Kingsford, M. J. (2000) Elemental fingerprints of otoliths of fish may distinguish estuarine'nursery'habitats. *Marine Ecology Progress Series* **201**: 273–286.

Görür, D. & Rasmussen, C. E. (2010) Dirichlet process gaussian mixture models: Choice of the base distribution. *Journal of Computer Science and Technology* **25**: 653–664.

Gross, M. R., Coleman, R. M., & McDowall, R. M. (1988) Aquatic productivity and the evolution of diadromous fish migration. *Science **239***: 1291-1293.

Hamann, E. J. & Kennedy, B. P. (2012) Juvenile dispersal affects straying behaviors of adults in a migratory population. *Ecology* **93**: 733–740.

Hendry, A. P., Morbey, Y. E., Berg, O. K. & Wenburg, J. K. (2004) Adaptive variation in senescence: reproductive lifespan in a wild salmon population. *Proceedings of the Royal Society of London B: Biological Sciences* **271**: 259–266.

Higgins, R. M., Danilowicz, B. S., Brophy, D., Geffen, A. J., McGowan, T. & Gillanders, B. (2013) Influence of the limit of detection on classification using otolith elemental signatures. *Canadian Journal of Fisheries and Aquatic Sciences* **70**: 922–929.

Hobbs, J. A., Bennett, W. A., Burton, J. & Gras, M. (2007) Classification of Larval and Adult Delta Smelt to Nursery Areas by Use of Trace Elemental Fingerprinting. *Transactions of the American Fisheries Society* **136**: 518–527.

Ishwaran, H. & James, L. F. (2003) Generalized weighted chinese restaurant processes for species sampling mixture models. *Statistica Sinica* **13**: 1211–1235.

Jolly, M. T., Aprahamian, M. W., Hawkins, S. J., Henderson, P. A., Hillman, R., O'Maoiléidigh, N., Maitland, P. S., Piper, R. & Genner, M. J. (2012) Population genetic structure of protected allis shad (Alosa alosa) and twaite shad (Alosa fallax). *Marine Biology* **159**: 675–687.

Jonsson, B., Waples, R. S. & Friedland, K. D. (1999) Extinction considerations for diadromous fishes. *ICES Journal of Marine Science: Journal du Conseil* **56**: 405–409.

Jorgensen, C., Enberg, K., Dunlop, E. S., Arlinghaus, R., Boukal, D. S., Brander, K., Ernande, B., Gardmark, A. G., Johnston, F., Matsumura, S., Pardoe, H., Raab, K., Silva, A., Vainikka, A., Dieckmann, U., Heino, M. & Rijnsdorp, A. D. (2007) Ecology: Managing Evolving Fish Stocks. *Science* **318**: 1247–1248.

Kalish, J. M. (1991) Determinants of otolith chemistry: seasonal variation in the composition of blood plasma, endolymph and otoliths of bearded rock cod *Pseudophycis barbatus*. *Mar. Ecol. Prog. Ser* **74**: 137–159.

Keefer, M. L. & Caudill, C. C. (2014) Homing and straying by anadromous salmonids: a review of mechanisms and rates. *Reviews in Fish Biology and Fisheries* **24**: 333–368.

Kennedy, B. P., Blum, J. D., Folt, C. L. & Nislow, K. H. (2000) Using natural strontium isotopic signatures as fish markers: methodology and application. *Canadian Journal of Fisheries and Aquatic Sciences* **57**: 2280–2292.

Kennedy, B. P., Klaue, A., Blum, J. D., Folt, C. L. & Nislow, K. H. (2002) Reconstructing the lives of fish using Sr isotopes in otoliths. *Canadian Journal of Fisheries and Aquatic Sciences* **59**: 925–929.

Kerr, L. A. & Secor, D. H. (2012) Partial Migration Across Populations of White Perch (*Morone americana*): A Flexible Life History Strategy in a Variable Estuarine Environment. *Estuaries and Coasts* **35**: 227–236.

Kritzer, J. P. & Sale, P. F. (2004) Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. *Fish and Fisheries* **5**: 131–140.

Lassalle, G., Béguer, M., Beaulaton, L., & Rochard, E. (2008). Diadromous fish conservation plans need to consider global warming issues: an approach using biogeographical models. *Biological Conservation* **141**: 1105-1118.

Lassalle, G., & Rochard, E. (2009). Impact of twenty-first century climate change on diadromous fish spread over Europe, North Africa and the Middle East. *Global Change Biology* **15**: 1072-1089.

Limburg, K. E. (1995) Otolith strontium traces environmental history of subyearling American shad *Alosa sapidissima*. *Mar. Ecol. Prog. Ser.* **119**: 25–35.

Limburg, K. E. & Waldman, J. R. (2009) Dramatic Declines in North Atlantic Diadromous Fishes. *BioScience* **59**: 955–965.

Lochet, A. (2006) Devalaison des juvéniles et tactiques gagnantes chez la grande Alose *Alosa alosa* et l'alose feinte *Alosa fallax*: Apports de la microchimie et de la microstructure des otolithes. Thèse de Doctorat. Université Bordeaux I. Bordeaux. 220 pp.

Martin, G. B., Thorrold, S. R. & Jones, C. M. (2004) Temperature and salinity effects on strontium incorporation in otoliths of larval spot (*Leiostomus xanthurus*). *Canadian Journal of Fisheries and Aquatic Sciences* **61**: 34–42.

Martin, J., Bareille, G., Berail, S., Pecheyran, C., Daverat, F., Bru, N., Tabouret, H. & Donard, O. (2013) Spatial and temporal variations in otolith chemistry and relationships with water chemistry: a useful tool to distinguish Atlantic salmon *Salmo salar* parr from different natal streams: Otolith chemistry of *Salmo salar*. *Journal of Fish Biology* **82**: 1556–1581.

Martin, J., Rougemont, Q., Drouineau, H., Launey, S., Jatteau, P., Bareille, G., Berail, S., Pécheyran, C., Feunteun, E., Roques, S., Clavé, D., Nachón, D. J., Antunes, C., Mota, M., Réveillac, E., Daverat, F. & Morán, P. (2015) Dispersal capacities of anadromous Allis shad population inferred from a coupled genetic and otolith approach. *Canadian Journal of Fisheries and Aquatic Sciences* **72**: 991–1003.

McDowall, R. M. (2008) Diadromy, history and ecology: a question of scale. *Hydrobiologia* **602**: 5–14.

Mercier, L., Darnaude, A. M., Bruguier, O., Vasconcelos, R. P., Cabral, H. N., Costa, M. J., Lara, M., Jones, D. L. & Mouillot, D. (2011) Selecting statistical models and variable combinations for optimal classification using otolith microchemistry. *Ecological Applications* **21**: 1352–1364.

Mota, M., Bio, A., Bao, M., Pascual, S., Rochard, E. & Antunes, C. (2015) New insights into biology and ecology of the Minho River Allis shad (Alosa alosa L.): contribution to the conservation of one of the last European shad populations. *Reviews in Fish Biology and Fisheries* **25**: 395–412.

Munch, S. B. & Clarke, L. M. (2008) A Bayesian approach to identifying mixtures from otolith chemistry data. *Canadian Journal of Fisheries and Aquatic Sciences* **65**: 2742–2751.

Neubauer, P., Shima, J. S. & Swearer, S. E. (2013) Inferring dispersal and migrations from incomplete geochemical baselines: analysis of population structure using Bayesian infinite mixture models (L Borger, Ed.). *Methods in Ecology and Evolution* **4**: 836–845.

Pella, J. & Masuda, M. (2006) The gibbs and split merge sampler for population mixture analysis from genetic data with incomplete baselines. *Canadian Journal of Fisheries and Aquatic Sciences* **63**: 576–596.

Pflugeisen, B. M. & Calder, C. A. (2013) Bayesian hierarchical mixture models for otolith microchemistry analysis. *Environmental and Ecological Statistics* **20**: 179–190.

Plummer, M. (2003, March). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing* 124-125.

Pouilly, M., Point, D., Sondag, F., Henry, M. & Santos, R. V. (2014) Geographical Origin of Amazonian Freshwater Fishes Fingerprinted by $^{87}Sr/^{86}Sr$ Ratios on Fish Otoliths and Scales. *Environmental Science & Technology* **48**: 8980–8987.

Quinn, T. P. (1984) Homing and straying in Pacific salmon. *Mechanisms of migration in fishes* **14**: 357–362.

Quinn, T. P. (1993) Biological Interactions of Natural and Enhanced Stocks of SalmonA review of homing and straying of wild and hatchery-produced salmon. *Fisheries Research* **18**: 29–44.

Quinn, T. P. & Fresh, K. (1984) Homing and Straying in Chinook Salmon (*Oncorhynchus tshawytscha*) from Cowlitz River Hatchery, Washington. *Canadian Journal of Fisheries and Aquatic Sciences* **41**: 1078–1082.

Quinn, T. P., Rich, H. B., Gosse, D. & Schtickzelle, N. (2012) Population dynamics and asynchrony at fine spatial scales: a case history of sockeye salmon (*Oncorhynchus nerka*) population structure in Alaska, USA. *Canadian Journal of Fisheries and Aquatic Sciences* **69**: 297–306.

Rasmussen, C. E. (1999) The infinite Gaussian mixture model. *NIPS* **1**: 554-560.

Rooker, J. R., David Wells, R. J., Itano, D. G., Thorrold, S. R. & Lee, J. M. (2016) Natal origin and population connectivity of bigeye and yellowfin tuna in the Pacific Ocean. *Fisheries Oceanography* **25**: 277–291.

Rougier, T., Lambert, P., Drouineau, H., Girardin, M., Castelnaud, G., Carry, L., Aprahamian, M., Rivot, E. & Rochard, E. (2012) Collapse of allis shad, *Alosa alosa*, in the Gironde system (southwest France): environmental change, fishing mortality, or Allee effect? *ICES Journal of Marine Science* **69**: 1802–1811.

Sabatié, M. R., Boisneau, P., & Alexandrino, P. (2000) Variabilité morphologique. *Les aloses Alosa alosa et Alosa fallax spp.* 137-178.

Sadovy, Y. & Severin, K. P. (1992) Trace elements in biogenic aragonite: correlation of body growth rate and strontium levels in the otoliths of the white grunt, *Haemulon plumieri* (Pisces: Haemulidae). *Bulletin of Marine Science* **50**: 237–257.

Schtickzelle, N. & Quinn, T. P. (2007) A metapopulation perspective for salmon and other anadromous fish. *Fish and Fisheries* **8**: 297–314.

Sethuraman, J. (1994) A constructive definition of dirichlet priors. *Statistica Sinica* **4**: 639–650.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**: 583-639.

Stewart, I. J., Quinn, T. P. & Bentzen, P. (2003) Evidence for Fine-Scale Natal Homing Among Island Beach Spawning Sockeye Salmon, *Oncorhynchus nerka*. *Environmental Biology of Fishes* **67**: 77–85.

Thresher, R. E. (1999) Elemental composition of otoliths as a stock delineator in fishes. *Fisheries Research* **43**: 165-204.
Tomas, J., Augagneur, S. & Rochard, E. (2005) Discrimination of the natal origin of young-of-the-year Allis shad (*Alosa alosa*) in the Garonne-Dordogne basin (south-west France) using otolith chemistry. *Ecology of Freshwater Fish* **14**: 185–190.
Townsend, D. W., Radtke, R. L., Corwin, S. & Libby, D. A. (1992) Strontium:calcium ratios in juvenile Atlantic herring *Clupea harengus* L. otoliths as a function of water temperature. *Journal of Experimental Marine Biology and Ecology* **160**: 131–140.
Ueda, H. (2012) Physiological mechanisms of imprinting and homing migration in Pacific salmon *Oncorhynchus* spp. *Journal of Fish Biology* **81**: 543–558.
Volk, E. C., Blakley, A., Schroder, S. L., & Kuehner, S. M. (2000) Otolith chemistry reflects migratory characteristics of Pacific salmonids: Using otolith core chemistry to distinguish maternal associations with sea and freshwaters.*Fisheries Research* **46**: 251-266.
Walther, B. D. & Limburg, K. E. (2012) The use of otolith chemistry to characterize diadromous migrations. *Journal of Fish Biology* **81**: 796–825.
Walther, B. D. & Thorrold, S. R. (2006) Water, not food, contributes the majority of strontium and barium deposited in the otoliths of a marine fish. *Marine Ecology Progress Series* **311**: 125–130.
Walther, B. D. & Thorrold, S. R. (2008) Continental-scale variation in otolith geochemistry of juvenile American shad (Alosa sapidissima). *Canadian Journal of Fisheries and Aquatic Sciences* **65**: 2623–2635.
Watabe, N., Tanaka, K., Yamada, J., & Dean, J. M. (1982) Scanning electron microscope observations of the organic matrix in the otolith of the teleost fish *Fundulus heteroclitus* (Linnaeus) and *Tilapia nilotica* (Linnaeus). *Journal of Experimental Marine Biology and Ecology* ***58***: 127-134.
White, J. W., Standish, J. D., Thorrold, S. R. & Warner, R. R. (2008) Markov chain Monte Carlo methods for assigning larvae to natal sites using natural geochemical tags. *Ecological Applications* **18**: 1901–1913.
Wilcove (2008) Animal Migration: An Endangered Phenomenon? Issues in Science and Technology **24**: 71–78.
Young, K. A. (1999) Managing the decline of Pacific salmon: metapopulation theory and artificial recolonization as ecological mitigation. *Canadian Journal of Fisheries and Aquatic Sciences* **56**: 1700–1706.

Sitographie

http://www.logrami.fr, consulté le 23/06/2016
http://www.sage-vire.fr, consulté le 02/08/2016

7. Appendix

Appendix I: otolith ablation

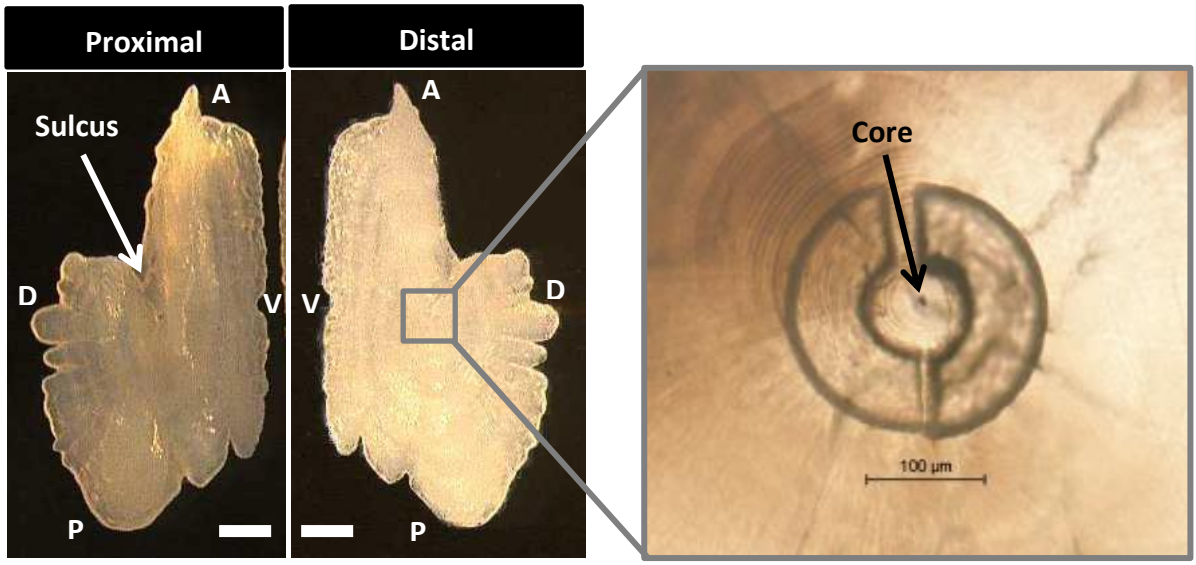

Figure A. 1 On the left: the left sagitta of an adult Allis shad on the proximal and distal side. D: dorsal face. V: ventral face. A: anterior face. P: posterior face. The scale represents 500 µm. On the right: two semi coronas ablated on the otolith of an adult Allis shad near the core. The right and the left semi coronas correspond respectively to the ICP-MS (elemental analysis) and the MC-ICP-MS (isotopic analysis) analyses. Two left photographs: Lochet (2006). Right photograph: Martin *et al.* (2015).

Appendix II: JAGS code

Model 1: Bayesian model with fixed number of sources and multiple baselines

```
model
{

# ------------------------------------------------------------------------------------------------------------
# Priors on the parameters of the regression water/otolith
# ------------------------------------------------------------------------------------------------------------
            for (elem in 2:(nbelement)) {
                        a[elem] ~ dunif(0,2)
                        b[elem] ~ dunif(-3,3)
            }
            a[1]<-1
            b[1]<-0

# ------------------------------------------------------------------------------------------------------------
# Regression between water and otoliths for nbriv sources (the number of sources of the water baseline)
# ------------------------------------------------------------------------------------------------------------
            for (riv in 1:nbriv) {
                        for (elem in 1:(nbelement)) {
          mu_elem[riv, elem] <- Eau[riv, elem]*a[elem]+b[elem]
       }
    }

# ------------------------------------------------------------------------------------------------------------
# Reallocation process
# ------------------------------------------------------------------------------------------------------------
    for (poisson in 1:nbad) {
       NaisAd[poisson] ~ dcat(ProbaNais[CaptAd[poisson],CaptY[poisson],1:nbriv])
       OtoAd[poisson, 1:nbelement] ~ dmnorm(mu_elem[NaisAd[poisson],
          1:nbelement], tau_elem[1:nbelement, 1:nbelement])
    }

    for (ju in 1:nbjuv) {
       OtoJuv[ju, 1:nbelement] ~ dmnorm(mu_elem[CaptJuv[ju],
          1:nbelement], tau_elem[1:nbelement, 1:nbelement])
    }

# ------------------------------------------------------------------------------------------------------------
# Probability of origin considering catch river and catch year effects
# ------------------------------------------------------------------------------------------------------------
            for (riv in 1:nbriv) {
       proba[riv]<-1/nbriv}

            for (riv in 1:ncapt){
                        for(y in 1:nyear){
                        ProbaNais[riv,y,1:nbriv]~ddirch(proba[1:nbriv])
             }
    }

# ------------------------------------------------------------------------------------------------------------
# Prior on variance covariance matrix
# ------------------------------------------------------------------------------------------------------------
            tau_elem[1:nbelement, 1:nbelement] ~ dwish(R[, ], scale)
            scale <- nbelement +1
```

```
# ------------------------------------------------------------------------------------------------------------
# Prediction of otolith composition for each elements and each river of the water baseline
# ------------------------------------------------------------------------------------------------------------
        for (i in 1:nbriv){
        Oto_pred[i,1:nbelement]~dmnorm(mu_elem[i,1:nbelement],tau_elem[1:nbelement, 1:nbelement])
        }

} # end of the model
```

Model 2: Infinite Mixture Model

```
model
{

# ------------------------------------------------------------------------------------------------------------
# Prior on variance covariance matrix
# ------------------------------------------------------------------------------------------------------------
    tau_elem[1:nbelement, 1:nbelement] ~ dwish(R[, ], scale)
    scale <- nbelement +1

# ------------------------------------------------------------------------------------------------------------
# Prior on alpha (the concentration parameter)
# ------------------------------------------------------------------------------------------------------------
            ialpha ~ dgamma(1,1) T(0.1,3.4)
            alpha<-1/ialpha

# ------------------------------------------------------------------------------------------------------------
# Construction of the stick breaking process
# ------------------------------------------------------------------------------------------------------------
for (riv in 1:nbsource){
      for (elem in 1:(nbelement)){
         mu_otho[riv,elem]~dunif(-2,2) # average composition of an otolith within given river
      }
   }

for (source in 1:(nbsource-1)){
            q[source]~dbeta(1,alpha)
            }

            weight.ini[1]<-q[1]
            for (j in 2:(nbsource-1)){
                        weight.ini[j]<-q[j]*(1-q[j-1])*weight.ini[j-1]/q[j-1]
            }
            weight.ini[nbsource]<-1-sum(weight.ini[1:(nbsource-1)])
            sorted.weights<-sort(weight.ini[1:nbsource])

# ------------------------------------------------------------------------------------------------------------
# Probability of origin
# ------------------------------------------------------------------------------------------------------------
            for (j in 1:nbsource){
                        ProbaNais[nbsource-j+1] <- sorted.weights[j]
            }

# ------------------------------------------------------------------------------------------------------------
# Reallocation process
# ------------------------------------------------------------------------------------------------------------
   for (poisson in 1:nbad){
      NaisAd[poisson] ~ dcat(ProbaNais[1:nbsource])
      OtoAd[poisson, 1:nbelement] ~ dmnorm(mu_otho[NaisAd[poisson],
         1:nbelement], tau_elem[1:nbelement, 1:nbelement])
   }

# ------------------------------------------------------------------------------------------------------------
# Summary statistics
# ------------------------------------------------------------------------------------------------------------
            for (i in 1:nbad) {
                        for (j in 1:nbsource) {
                                    obs.by.cluster[i,j] <- equals(NaisAd[i], j)
                        }
```

```
        }
        for (j in 1:nbsource) {
                active[j] <- step(sum(obs.by.cluster[,j])-1)

# ------------------------------------------------------------------------------------------------------------
# Prediction of otolith composition for each elements and each source
# ------------------------------------------------------------------------------------------------------------
        Oto_pred[j,1:nbelement]~dmnorm(mu_otho[j,1:nbelement],tau_elem[1:nbelement, 1:nbelement])
        }

} # end of the model
```

Model 3: Bayesian hybrid model

```
model
{

# ------------------------------------------------------------------------------------------------------------------
# Priors on the parameters of the regression water/otolith
# ------------------------------------------------------------------------------------------------------------------
   for (elem in 2:(nbelement)) {
                        a[elem] ~ dunif(0,2)
                        b[elem] ~ dnorm(0,1e-6)
            }
            a[1]<-1
            b[1]<-0

# ------------------------------------------------------------------------------------------------------------------
# Regression between water and otoliths for nbriv sources (the number of sources of the water baseline)
# ------------------------------------------------------------------------------------------------------------------
   for (riv in 1:nbriv) {
      proba[riv]<-1/nbriv
      for (elem in 1:(nbelement)) {
         mu_elem[riv, elem] <- Eau[riv, elem]*a[elem]+b[elem]
      }
   }

# ------------------------------------------------------------------------------------------------------------------
# Prior on variance covariance matrix
# ------------------------------------------------------------------------------------------------------------------
    tau_elem[1:nbelement, 1:nbelement] ~ dwish(R[, ], scale)
    scale <- nbelement +1

# ------------------------------------------------------------------------------------------------------------------
# Prior on alpha (the concentration parameter)
# ------------------------------------------------------------------------------------------------------------------
            ialpha ~ dgamma(1,1) T(0.1,3.4)
            alpha<-1/ialpha
# ------------------------------------------------------------------------------------------------------------------
# We build the weights of additional sources through a stick breaking process
# ------------------------------------------------------------------------------------------------------------------
            for (riv in (nbriv+1):(nbsource-1)){
                        q[riv-nbriv]~dbeta(1,alpha)
                        }
            for (riv in (nbriv+1):nbsource){
                        for (elem in 1:(nbelement)) {
                                    mu_elem[riv, elem] ~ dunif(-2,2)
                        }
            }
            weight.ini[1]<-q[1]
            for (j in (nbriv+2):(nbsource-1)){
                        weight.ini[j-nbriv]<-q[j-nbriv]*(1-q[j-nbriv-1])*weight.ini[j-nbriv-1]/q[j-nbriv-1]
            }
            weight.ini[nbsource-nbriv]<-1-sum(weight.ini[1:(nbsource-nbriv-1)])
            sorted.weights<-sort(weight.ini[1:(nbsource-nbriv)])


            for (riv in 1:ncapt){
                        #flat dirichlet priors for the nbriv first rivers (the one for which we have water)
                        Priv[riv,1:nbriv]~ddirch(proba[1:nbriv])
                                    for(j in 1:nbriv){
                                                ProbaNais[riv,j]<-Priv[riv,j]*(1-(1-(1/1+alpha))^nbriv)
```

```
                        #assuming a dirichlet process model, the nbriv first riv would have this total weights
                }

                #for the remaining sources we use the weights from the stick breaking process
                for (j in (nbriv+1):nbsource){
                        ProbaNais[riv,nbsource-j+1+nbriv]<-sorted.weights[j-nbriv]*((1-1/1+alpha)^nbriv)
#we removed the expected weight of the nbriv first rivers
                }
        }

# ------------------------------------------------------------------------------------------------------------
# Reallocation process
# ------------------------------------------------------------------------------------------------------------
for (poisson in 1:nbad) {
    NaisAd[poisson] ~ dcat(ProbaNais[CaptAd[poisson],1:nbsource])
    OtoAd[poisson, 1:nbelement] ~ dmnorm(mu_elem[NaisAd[poisson],1:nbelement], tau_elem[1:nbelement,
1:nbelement])
  }

for (ju in 1:nbjuv) {
    OtoJuv[ju, 1:nbelement] ~ dmnorm(mu_elem[CaptJuv[ju],
      1:nbelement], tau_elem[1:nbelement, 1:nbelement])
  }

# ------------------------------------------------------------------------------------------------------------
# Summary statistics
# ------------------------------------------------------------------------------------------------------------
        for (i in 1:nbad) {
                for (j in 1:nbsource) {
                        obs.by.cluster[i,j] <- equals(NaisAd[i], j)
                }
        }
        for (j in 1:nbsource) {
                active[j] <- step(sum(obs.by.cluster[,j])-1)

# ------------------------------------------------------------------------------------------------------------
# Prediction of otolith composition for each elements and each source
# ------------------------------------------------------------------------------------------------------------
        Oto_pred[j,1:nbelement]~dmnorm(mu_otho[j,1:nbelement],tau_elem[1:nbelement, 1:nbelement])
        }

} # end of the model
```

Appendix III: Posterior checking

Convergence of reallocation was checked for ALAL138 in the first model since the three chains predicted natal origin in the same proportion during the iterative process (Figure A.2.A). Therefore, in the first model, this individual was supposed to be originated from the Blavet (near 30%) and the Scorff (near 70%) rivers. The final reallocation was the Scorff River according to the MPR. Those predictions are presented in Figure A.2.B.
However, the convergence of reallocation was not checked in the third model. Indeed, as presented in Figure A.2.C, the three chains predicted variable sources, leading in inconsistencies in the distribution of reallocation. We explained the reallocation of this individual in extra-source in the third model by extreme signature in Sr/Ca (Figure A.2.D).

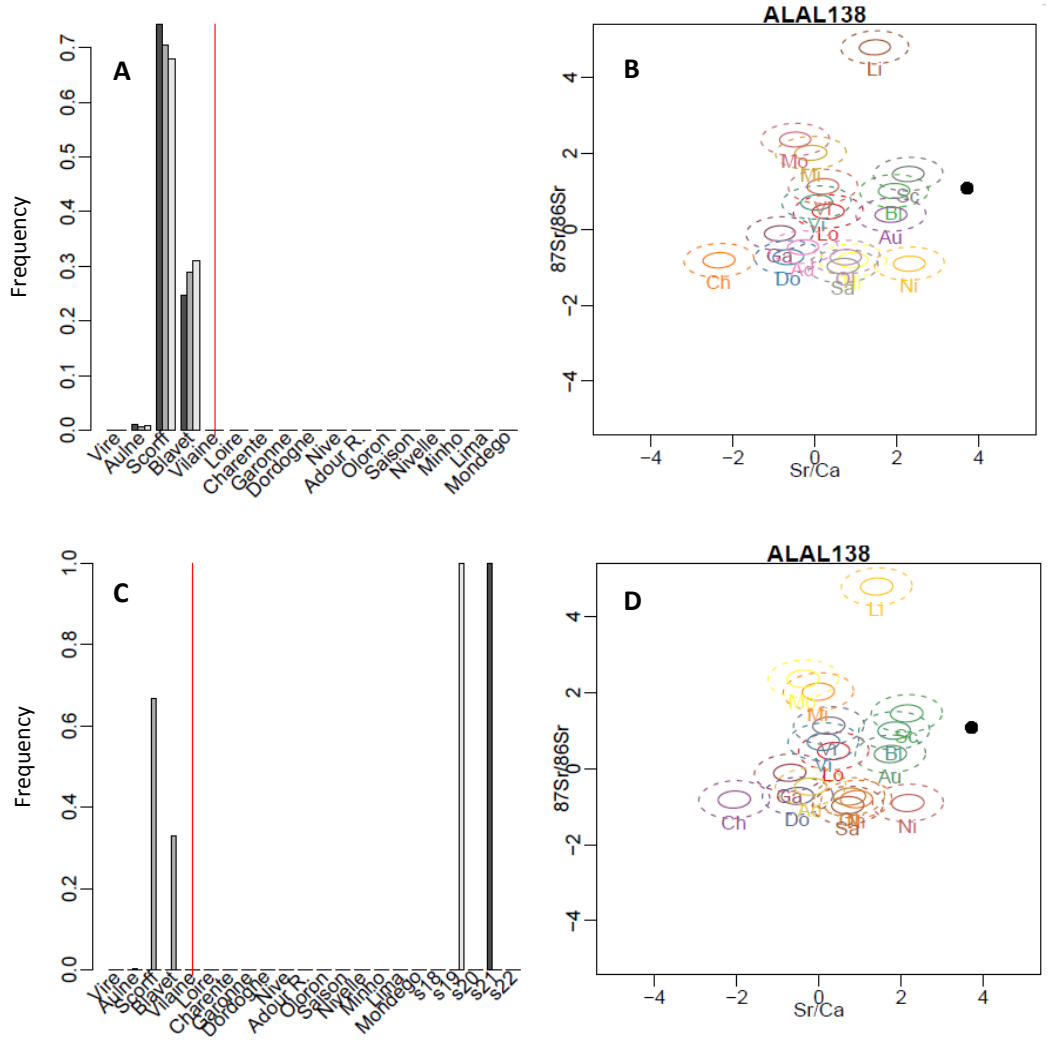

Figure A. 2 Posterior distribution of the categorical variable of reallocation N of the individual 'ALAL138' for the first (A) and third model (C). Red lines indicate to the catch river. Grey bars are the MCMC chains distribution. Based on the water composition, prediction of otolith composition ($^{87}Sr/^{86}Sr$ and Sr/Ca) for each source are presented for the first (B) and third model (D). Ellipses are 95% (in dotted line) and 75% (in solid lines) confidence intervals. The black point represents the otolith composition of 'ALAL138'. Variables are centered and scale. Predictions of extra-sources composition are not visible because ellipses were too large to be represented in the same plot as other sources (D). Ch: Charente; Do: Dordogne; Ga: Garonne; Sa: Saison; Ni: Nivelle; N: Nive; Ol: Oloron; Lo: Loire; Bl : Blavet ; Au: Aulne ; Sc : Scorff; Ad : Adour ; Vir: Vire; Vi: Vilaine; Mi: Minho; Mo: Mondego; Li: Lima.

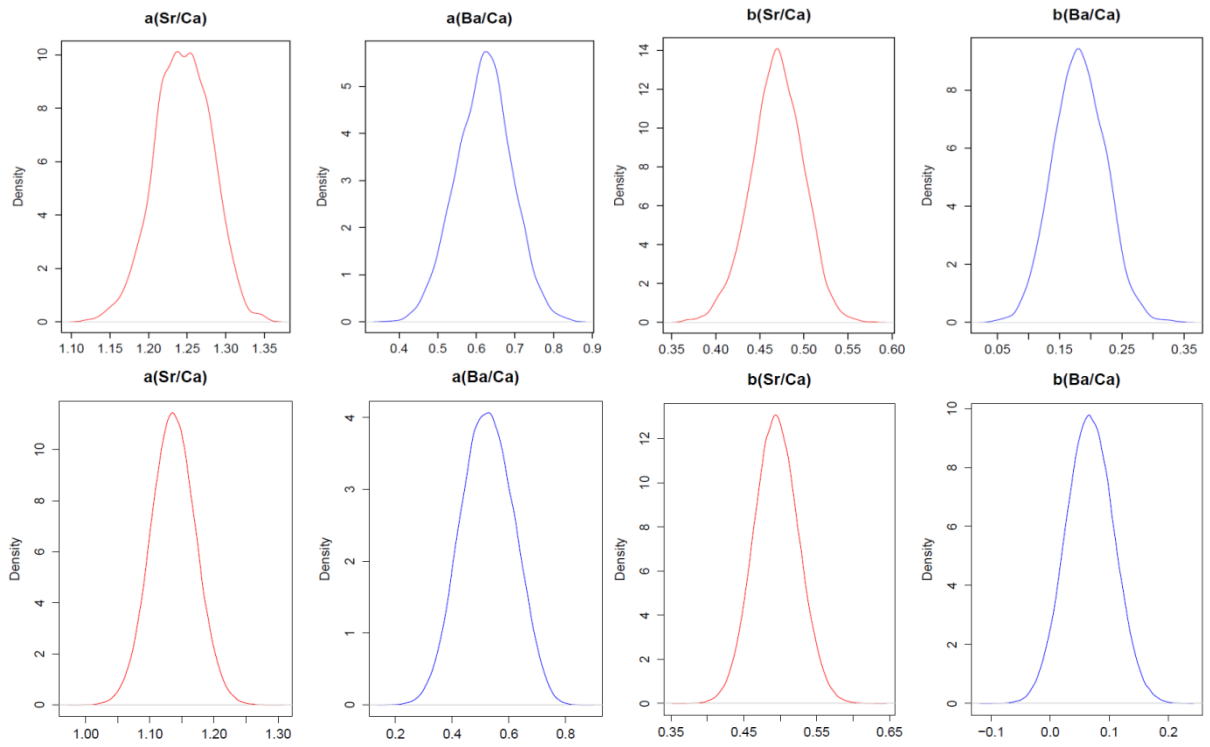

Figure A. 3 Posterior distribution of the partition coefficients a and b for the Sr/Ca and Ba/Ca ratios estimated by the first model (on the top) and the third model (on the bottom).

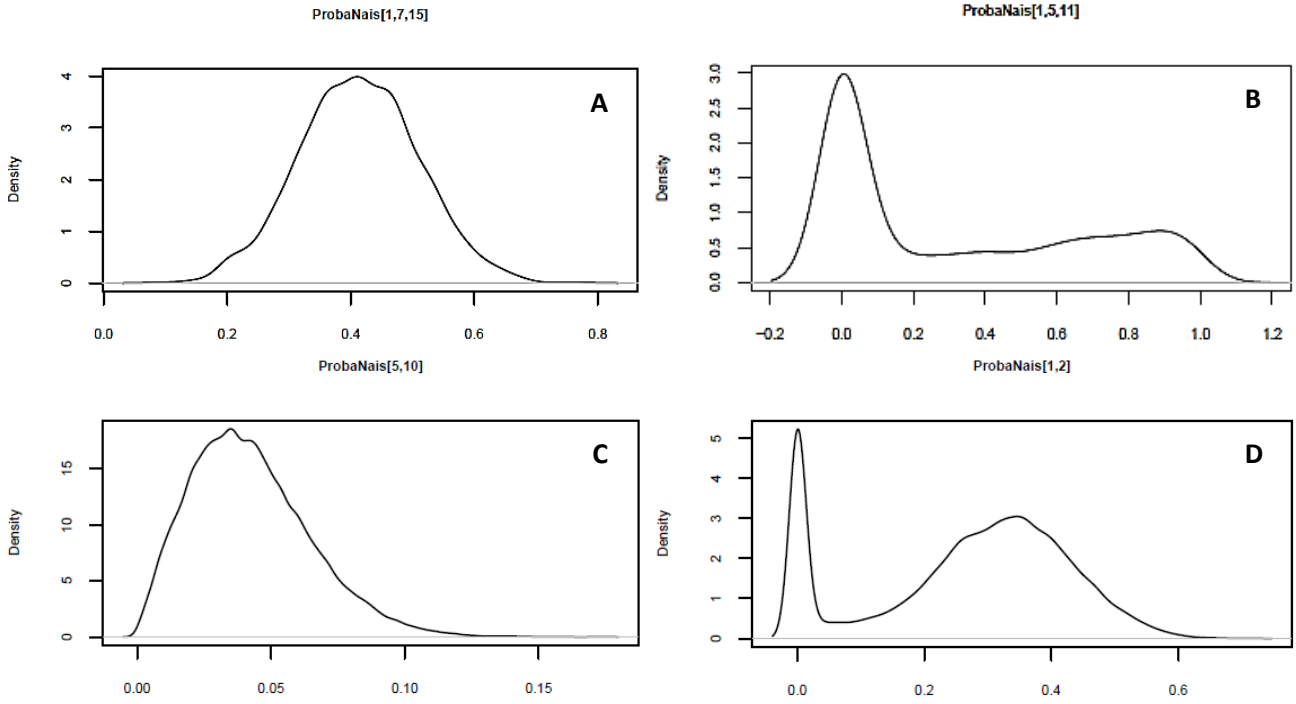

Figure A. 4 Example of posterior distributions of the probabilities of origin. A and B correspond to posteriors from the first model. A and B represent respectively the probability of being originated from the Adour river for fishes caught in the Adour estuary in 2013, and the probability of being originated from the Nive river for fishes caught in the Adour estuary in 2011. C and D correspond to posteriors from the third model. C and D represent respectively the probability of being originated from the Nivelle river for fishes caught in the Dordogne river and the probability of being originated from the Garonne for fishes caught in the Adour estuary.

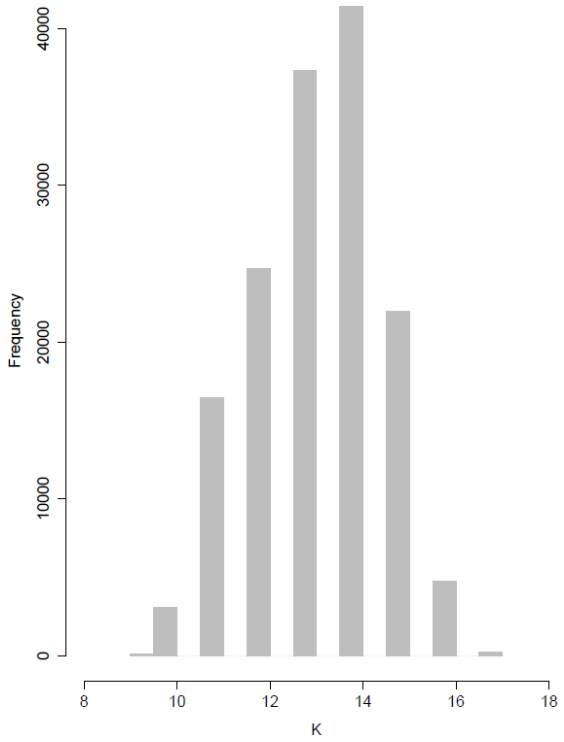

Figure A. 5 Posterior distribution of the number of sources K estimated by the Bayesian hybrid model.

Appendix IV: Abundances of spawners per year and watershed

Table A. 1 Surface of the watersheds (km²) and abundance indices of adults Allis shad per sampling year. Grey cells correspond to data matching with samplings of adult otoliths of Allis shad. Blue cells represent data with samplings of adult otoliths of Allis shad shad but without abundance indices, so that an extrapolation of abundance has to be done.

Watershed	Surface (km²)	Abundance per samplig year							
		2001	2008	2009	2010	2011	2012	2013	2014
Nivelle	238	29	224	91	103	330	332	78	103
Scorff	480	4	23	73	57	89	10	38	35
Oloron	533	2	4	0	12	368	60	513	204
Saison	627	NA	NA	NA	NA	NA	NA	NA	NA
Nive	1101	0	2	0	1	4	1	0	1
Vire	1260	NA	8000	5142	3081	2603	3786	3157	5489
Aulne	1821	NA	2771	2885	1143	918	120	1308	1156
Blavet	2089	NA	NA	NA	NA	NA	NA	NA	NA
Lima	2535	NA	NA	NA	NA	NA	NA	NA	NA
Mondego	6644	NA	NA	NA	NA	NA	NA	NA	NA
Charente	10549	NA	NA	NA	NA	NA	NA	NA	NA
Vilaine	11190	NA	NA	NA	NA	NA	NA	NA	NA
Adour	16880	NA	NA	NA	NA	NA	NA	NA	NA
Minho	17080	NA	NA	261	133	73	NA	NA	NA
Dordogne	24000	207748	55635	64906	55475	12985	2231	3724	14585
Garonne	28900	186735	72075	103314	87092	46409	3817	2729	12909
Loire	117800	3233	7258	2557	1893	540	1986	624	1539

Appendix V: Flux distributions between donor and recipient populations

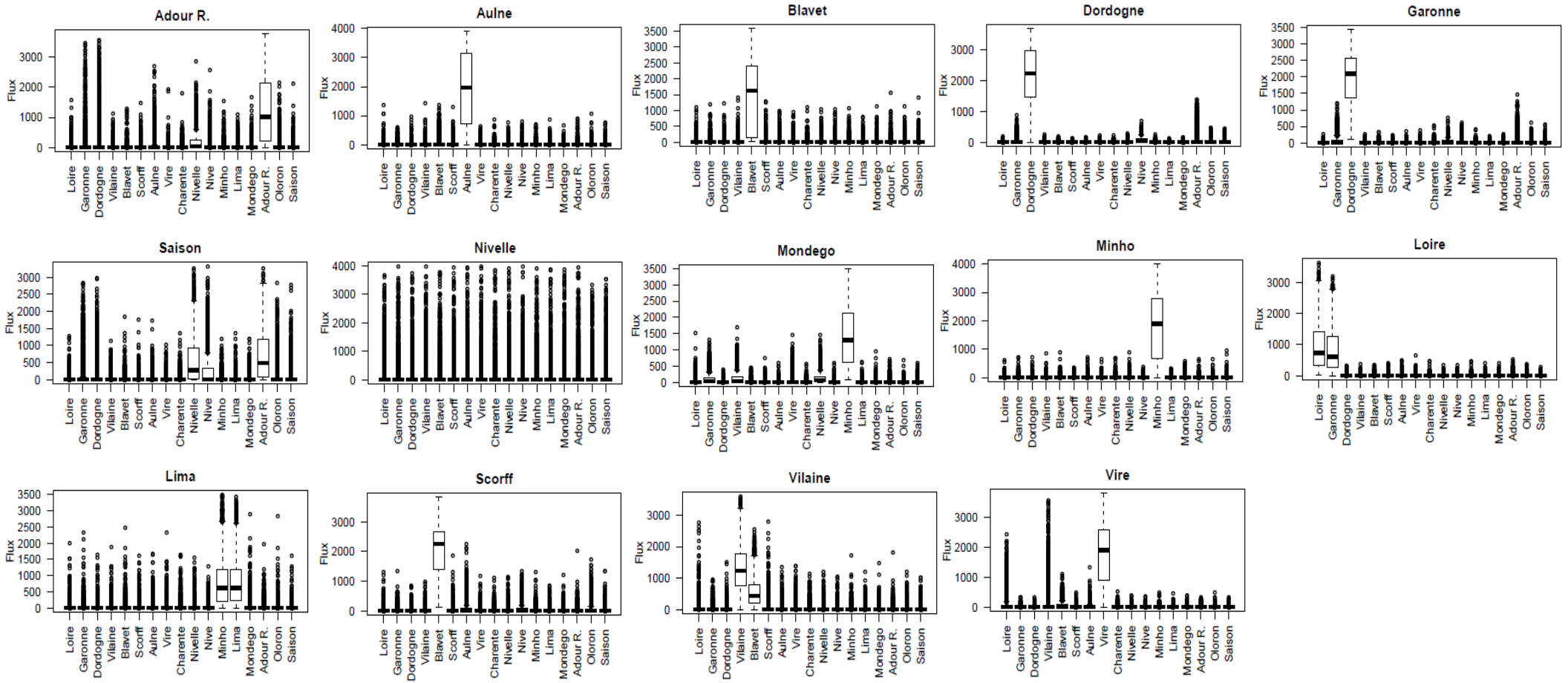

Figure A. 6 Distribution of flux between donor and recipient populations estimated by multiplying the probabilities of origin estimated by the first Bayesian model with the abundance estimates of spawners. Each plot corresponds to a recipient river (i.e. a catch river). The horizontal axes represent the natal river.