L'Institut Agro Rennes-Angers
☐ Site d'Angers ☒ Site de Rennes

Année universitaire : 2023-2024 Spécialité : SML-Biologie Parcours : Sciences halieutiques et aquacoles (Ressources et Ecosystèmes Aquatiques)	**Mémoire de fin d'études** ☐ d'ingénieur de l'Institut Agro Rennes-Angers (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement) ☒ de master de l'Institut Agro Rennes-Angers (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement) ☐ de l'Institut Agro Montpellier (étudiant arrivé en M2) ☐ d'un autre établissement (étudiant arrivé en M2)

Developing an operational approach to assess the quality of abundance estimates in the event of missing data: a case study on salmon returning abundances in France

Par : Klervi VERBRUGGHE

Soutenu à Rennes le 13 septembre 2024

Devant le jury composé de :

Président : Pablo BROSSET (Institut Agro, Rennes)

Maître de stage : Clément LEBOT (Institut Agro, Rennes)
Étienne RIVOT (Institut Agro, Rennes)

Enseignant référent : Olivier LE PAPE (Institut Agro, Rennes)

Autres membres du jury (Nom, Qualité)

Marie NEVOUX – Chargée de recherche en écologie et évolution (INRAE, Rennes)

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) :
(ou de l'étudiant-entrepreneur)

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) **Nom Prénom** VERBRUGGHE Klervi

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** : 13/09/24

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.

Date et signature de l'**enseignant** : 13 Sept 2024.

Utilisation de l'Intelligence Artificielle (IA) pour la rédaction du mémoire

L'intelligence artificielle a été utilisée pour améliorer le texte ☒ Oui ☐ Non

Si oui, préciser pour quelle partie de texte et à quelles fins ?

Aide à la traduction (DeepL)

Toute absence de déclaration précise sera considérée comme un manquement aux règles de l'éthique scientifique.

Date et Signature de l'**auteur** : 13/09/24

(1) L'administration, les enseignants et les différents services de documentation de l'Institut Agro Rennes-Angers 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

Acknowledgements (in French)

Avant tout, je tiens à remercier vivement mes encadrants Clément Lebot et Étienne Rivot pour leur accompagnement tout au long de ce stage. Votre pédagogie et votre bienveillance m'ont aidée à réaliser au mieux cette étude. Je vous remercie également pour votre disponibilité et pour le temps accordé à la relecture de ce rapport.

Je remercie également Tristan et Eliot, membres co-fondateurs de la FIACSS, pour ces quelques mois passés dans le même bureau. J'attends toujours de tester Euro Truck Simulator 2 avec vous.

Je tiens également à remercier mes collègues du pôle halieutique, Maximilien, Florian, Vianney, Jérôme, Romain, Hannah, Juliette et Armand, qui ont vivement contribué à l'amélioration de mes compétences ~~en palet et mots croisés~~ dans le vaste domaine qu'est l'halieutique.

Merci à Laurine et Colleen pour cet accueil si chaleureux à l'Agro et pour avoir partagé cette expérience qu'est le congrès de l'AFH.

Merci à mes parents et ma sœur, pour m'avoir soutenue durant toutes ces (longues) années d'étude.

Enfin, merci à Fabien pour ton soutien à toute épreuve et pour m'avoir soufflé tes conseils de grand sage à la croisée des chemins.

List of Figures

of dots informs of the value of the KL index (the value rounded to the hundredth is indicated above every dot) and the color of the dots informs the strength of the multiplicative bias. The KLD is theoretically a non-negative index but can lead to small negative values due to numerical and rounding errors, especially when the difference between the distributions is very small.

List of Appendices

Appendix I: Estimates of the proportion of 1SW in each region obtained in the initial version of the model before the regionalisation of the parameter. The estimates are narrowed around the national grand mean for the 1970-1990 period. Years in the x-axis are return years.

Appendix II: Estimates and data of the parameter of proportion of 1SW in catches (rho) for net fisheries (a) and rod and line fisheries (b) in the regions of interest for the previous and the modified versions of the PastSatAb model. The modified version allows for a better fit to data in the Loire-Allier region.

Appendix III: Estimates of the PastSatAb model under the seven data removal scenarios in comparison to the baseline scenario, for the regional returning abundance estimations (a) and for 1SW proportions estimations (b) for the year 2020. The plots display the median, the 25th and 75th quartiles and the density of the estimates.

Notation

PFA: Pre-fishery abundance

1SW: 1 sea-winter

MSW: multiple sea-winter

ICES: International Council for the Exploitation of the Sea

WGNAS: Working Group on North Atlantic Salmon

NASCO: North Atlantic Salmon Conservation Organisation

SU: Stock Unit

RR model: Run Reconstruction model

DAG: Direct Acyclic Graph

KLD: Kullback-Leibler Divergence

CV: Coefficient of variation

« C’est ainsi que, génération après génération, nous avons oublié la richesse et l’abondance de la vie sauvage qui peuplait la France encore récemment. […] Or on ne le répètera jamais assez, la surabondance est la norme : c’est l’état normal des écosystèmes européens en pleine possession de leurs fonctionnalités, non excessivement perturbées. »

Stéphane Durand, 20 000 ans ou la grande histoire de la nature

Table of contents

I – Introduction

I.1. Ecology, biological traits and life cycle of Salmo salar

Atlantic salmon (*Salmo salar*, Salmonidae) is an anadromous fish and indigenous to the North Atlantic basin (Aas et al., 2011). Its geographic wild range spans from Western Europe, including Portugal to the Barents Sea, and extends to the north-eastern part of America, from New England to the Ungava Bay in Canada (Aas et al., 2011; Mills et al., 2013). Its life cycle is shared between freshwater and marine habitats with the reproduction and juvenile phases occurring in upstream rivers (Aas et al., 2011; Bardonnet & Baglinière, 2011). After their marine sojourn, the return of adults to freshwater after the sea phase usually takes place in the native stream and the species' philopatry is, in part, based on their olfactory memory (Quinn, 1993; Stabell, 1984). This homing behaviour is not entirely strict but is frequently employed to explain the genetic differences observed among salmon populations (King et al., 2001). Spawning occurs between October and January and juveniles stay in freshwater a few years before smoltification happen (Aas et al., 2011; Gueguen & Prouzet, 1994). The length of the freshwater phase varies with latitudes (Webb et al., 2007). During the marine phase, many populations undertake wide-ranging migrations to common feeding grounds in the North Atlantic, located off the coast of West Greenland, in the Labrador Sea and in the Norwegian Sea (Mills et al., 2013; Olmos et al., 2019). At sea, salmon growth is rapid and fish from the French populations generally mature after one to two years at sea, rarely more (Webb et al., 2007). When maturation is reached after a single winter at sea, individuals are referred to as 1 sea-winter (1SW) salmons, and when they spend multiple winters at sea, they are referred to as multiple sea-winter (MSW) salmons (Otero et al., 2012; Rivot et al., 2009). Recent work indicates that sea age at maturation is largely genetically determined (Barson et al., 2015). The migration to the feeding grounds of 1SW extends to the Faroe Islands, while MSW continues to West Greenland waters (Figure 1). The 1SW and MSW sea-age classes differ in aspects such as physical appearance, timing of returning migration and sex ratio. Large MSW salmon often enter into fresh water earlier in the season than smaller 1SW salmon (Niemelä et al., 2006) and sex ratio is not balanced among the sea-age classes.

Because French rivers are located in the southern part of the species' range in Europe (Rivot et al., 2009), the species' life cycle in this area is relatively short compared to that of salmon originating from more northern rivers (Gueguen & Prouzet, 1994) and typically last from two to four years, when it can last nine years in Scandinavia (Prévost et al., 2020). In French populations, most of the MSW fish spent two winters at sea and are referred as 2SW fish (Rivot et al., 2009). The average proportion of females in adult returns is 45% for 1SW and 80% for MSW (Prévost & Porcher, 1996) with MSW females producing around 2 times more eggs than 1SW females (Nevoux et al., 2020). Maturation schedules vary by sex and geographically, some regions being dominated by 1SW fish and others by MSW salmon (Porter et al., 1986; ICES, 2010).

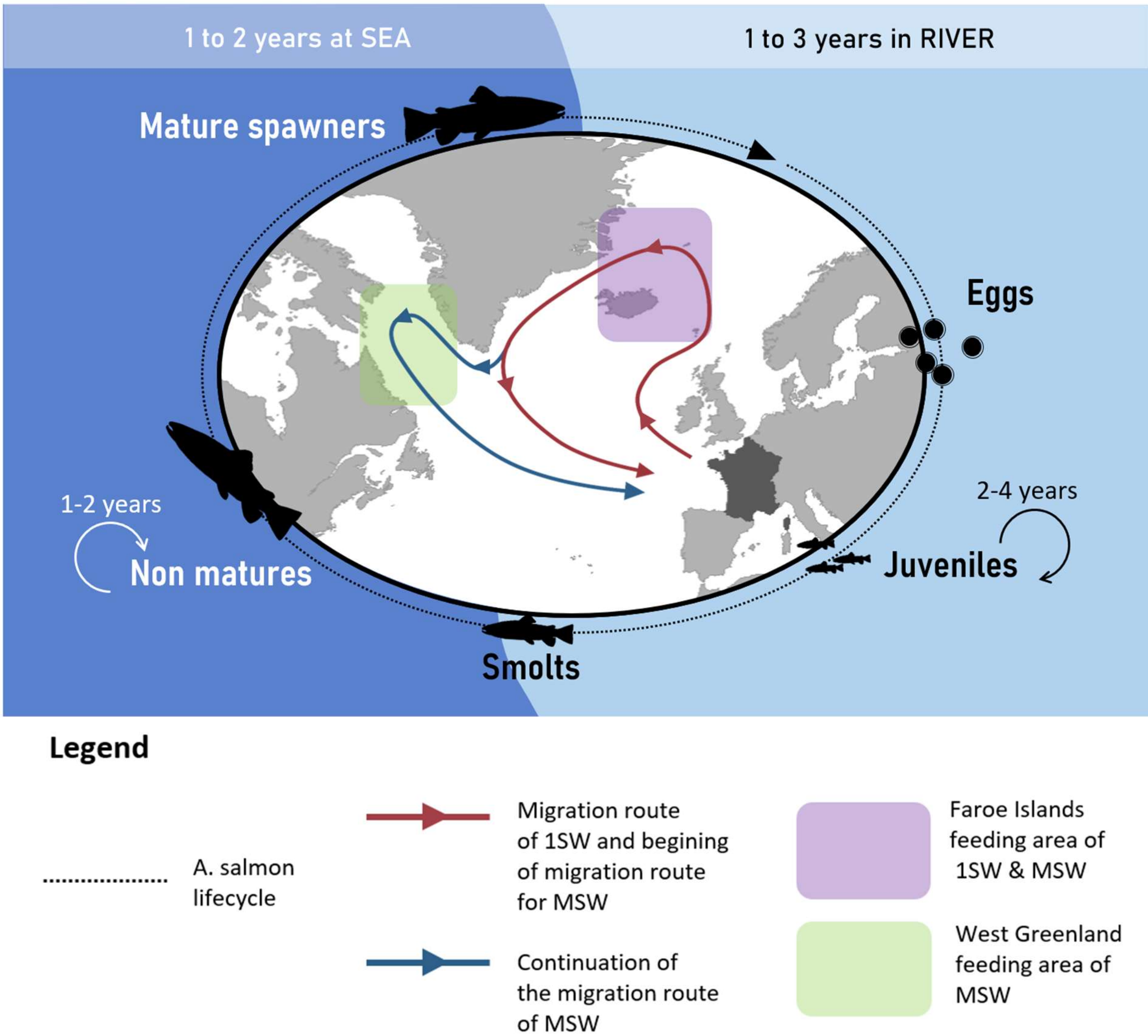


Figure 1: Life cycle, migration routes and feeding areas of 1SW and MSW adapted to French salmon populations (Adapted from Olmos (2019)).

I.2. A species under direct and indirect anthropogenic threats

Atlantic salmon populations have gradually declined in abundance across their distribution range since the 1970s and in most ecological regions, the distribution of salmon areas has become more restricted (Bardonnet & Baglinière, 2011; Chaput, 2012; Jonsson & Jonsson, 2004). The wild Atlantic salmon has a great patrimonial, functional and economic value. They are good indicators of watercourse health, reflecting the overall balance and functionality of aquatic ecosystems. Significant changes in their abundance raise serious concerns about the viability of this species and the condition of their habitat (André et al., 2021; Bal et al., 2017). The conservation status of Atlantic salmon has been downgraded from Least Concern (LC) to Near Threatened (NT) by the IUCN Red List of Threatened Species in 2022 (Darwall, 2022).

Human activities, especially their direct and indirect impacts on freshwater habitat, including dams, pollution, and marine overexploitation, can be considered as major contributors to this decline (Bardonnet & Baglinière, 2011; Dadswell et al., 2022). Alterations to salmon upstream and downstream migration routes can act as barriers to fish migration and reproduction, and therefore represent major obstacles to population survival (Gueguen & Prouzet, 1994).

Salmon populations also have to face a diverse range of fisheries operating in rivers, estuaries, and at sea which may contribute to important mortalities (Lebot et al., in prep). Due to the feeding aggregations behaviour in northern feeding areas during the marine migratory phase, the Atlantic salmon populations are targeted by mixed-stock fisheries (Bradbury et al., 2016). Mixed-stock fisheries exploit individuals originating from multiple source populations (Utter & Ryman, 1993), therefore, Atlantic salmon caught at West Greenland are both North American and European individuals (Reddin & Friedland, 1999). The West Greenland distant water fisheries target multi-sea-winter (MSW) salmon, directly impacting MSW home water stocks (Bradbury et al., 2016; Chaput, 2012; Jensen, 1990). Beginning in the early 1960s, this fishery was initially practiced on a local scale, limited to inland waters and managed on a national level (Reddin & Friedland, 1999; Windsor & Hutchinson, 1994), but it has grown into a more industrial fishery, shifting the scale of exploitation from subsistence to a developing economic activity (Rasmussen & Valdemarsvej, 2001). Catches rose quickly with the development of distant water fisheries and the increasing fishing effort in West Greenland, Norwegian seas and around Faroe Islands (Reddin & Friedland, 1999; Windsor & Hutchinson, 1994) and peaked in the 1970s at about 2 700 tons (Jensen, 1990; Reddin & Friedland, 1999). Nominal catches have declined by 56% over the 2006-2020 period, but this is partly attributed to the reduction of marine fishing intensity during that time period (with banning of the Norwegian drift net fishery and strong reduction of the Faroese long-line fisheries) (ICES, 2021; Jonsson & Jonsson, 2004).

Atlantic salmon are also particularly sensitive to climate change. There is strong evidence that climate change is responsible for altering ecosystem functioning and driving a major shift in the North Atlantic ecosystems (Beaugrand et al., 2008). Atlantic salmon populations at the trailing edge of the species' range are the most vulnerable to environmental change (Chaput, 2012; Valiente et al., 2011), and those living in southern European rivers are likely to be particularly affected by global warming (Valiente et al., 2011). The impact of climate change on marine ecosystem has been suggested as one of the principal causes for the decline in marine survival. The change in marine survival, that began in the Northwest Atlantic in the early 1990s, resulted in a shift in productivity of Atlantic salmon of North America (Chaput et al., 2005; ICES, 2016). This major change in salmon survival has persisted to date and occurs simultaneously with a decline in the proportion of late maturing fish (MSW) in many European and North American stocks (ICES, 2016; Jonsson et al., 2016; Otero et al., 2012) and a delay in the timing of adult migration as a consequence of global warming (Valiente et al., 2011).

I.3. Atlantic salmon stock assessment and limitations of the current method used to estimate return abundances of Atlantic salmon in homewaters

Declining homewaters catches and the rise of distant water fisheries encouraged efforts to rehabilitate and enhance Atlantic salmon stocks, leading in 1984 to the establishment of the North Atlantic Salmon Conservation Organization (NASCO) and a treaty to manage the marine fisheries that can only be managed through international cooperation (Windsor & Hutchinson, 1994). Each year, the International Council for the Exploration of the Sea (ICES) is mandated by NASCO to provide advice on the management of Atlantic salmon stocks. This advice is provided by a group of scientific experts from ICES: the Working Group on North Atlantic Salmon (WGNAS), and for 25 stock units distributed across the species' range, including France which is identified as one of the 25 stock units (ICES, 2024).

To produce these recommendations, the WGNAS has developed different assessment models, called "run reconstruction" (RR) models used to estimate times series of national returning abundance of adult salmon to homewater in each of the 25 stock units (including France) by sea-age class from the early 1970s to present (Chaput, 2012; Olmos et al., 2019). The data used for this estimation are collected through an annual data call made by ICES for all stock units. The data requested by the data call pertain to fisheries data, such as catch data and harvest rates, and biological information, as female fecundity and the proportion of females among spawners per sea age class. These data are then used as inputs by the run reconstruction models. In a second step, population dynamics models developed by ICES use those abundance estimates to assess the conservation status of populations within different stock units. This process informs advice on sustainable catch levels under various management scenarios for mixed-stock fisheries, leading to recommendations on quota allocations for mixed-stock fisheries (Lebot et al., in prep).

The run-reconstruction models are a keystone of the assessment workflow. However, these models have limitations that may affect the reliability of the abundance estimates produced and, to some extent, the advice provided by WGNAS (Lebot et al., in prep). For most of the jurisdictions (including France), the model is fed with time series of catches realized in homewaters and estimates of harvest rates (either estimated from data or determined by experts) aggregated at the scale of the whole stock unit, without consideration of the across watershed spatial variability within the stock unit. Essentially, estimates of the number of fish that returns in rivers are derived from the homewater catches raised by the estimates of harvest rates. The raising process may also use estimates of catch declaration rates for jurisdictions where undeclared catches are suspected to occur. Some quantification of uncertainty is included in the process by considering harvest rates and declaration rates are distributed across a certain range (Lebot et al., in prep).

In France, the main limitations of this run-reconstruction procedure are the lack of transparency in the production of harvest rate estimates and the limited information available to justify variations over time. In addition, the spatial variability is not considered, although the temporal and spatial dynamics of returns in French homewaters and the proportion of sea-age classes vary across the different regions in France. Moreover, depending on the regions, the fisheries take place in different locations (estuary and river) and different fishing techniques are used depending on the habitat (net fisheries in estuaries and rod-and-line fisheries in rivers). These two fisheries also target different sea-age classes in different ways. By using catch and harvest rate data aggregated at the stock unit (SU) scale, the RR model cannot take into account those variabilities. In addition, the river is the preferred management scale for Atlantic salmon (ICES, 2023), and some detailed information as direct counts or absolute abundance estimates exist at the scale of index rivers equipped with migration survey devices (Lebot et al., in prep). However, the run reconstruction model does not consider information on return abundances, even though long-term monitoring stations and ad-hoc models (eg. Lebot et al. (2022)) provide total or partial information on the abundance of adult salmon returning to spawn at a river or regional scale.

I.4. PastSatAb project objectives and limitations

To overcome those limitations, the PastSatAb project has been initiated in France, supported by the "*pôle OFB-INRAE-Institut Agro Rennes-UPPA pour la gestion des migrateurs amphihalins dans leur environnement*" (MIAME) The aim of the project is to develop a new methodology to estimate time series of national returning abundance, taking into account the regional scale and the dynamics of the different fisheries. The ultimate goal of this project is to provide more reliable estimates of returns as inputs for the stock assessment model develop by the WGNAS.

A model (referred as the "PastSatAb model" in the subsequent) has already been developed and provide new time series estimates of 1SW and MSW return abundances from 1970 to 2020 at the scale of five regions that are aggregated to obtain national estimates: the Nord, the Bretagne, the Loire-Allier, The Garonne-Dordogne and the Adour (Lebot et al., in prep). This model takes full advantages of a heterogeneous set of data from abundance data to fisheries data, as well as regional expertise on population and exploitation dynamics of Atlantic salmon. Namely, the model makes use of four main sources of data for the different regions: the regional and river estimation of absolute abundance, the observation of river abundance made by counting devices (referred to as direct counts) and the catches. It is developed under a Hierarchical Bayesian Modelling framework which is particularly suited for the joint treatment of such heterogeneous dataset, especially in ecology (Royle & Dorazio, 2008; Cressie et al. 2009; Rivot and Parent 2012) as it allows to extrapolate the information provided by data-rich spatiotemporal units to data-poor units while integrating the inherent uncertainty of this extrapolation (Prévost et al., 2003).

The objective of this work is to address two major limitations and improvement tracks of the PastSatAb model. First, the way the model handles the spatial (among regions) and temporal (among years) variations of the proportion of 1SW/MSW fish in the returns should be improved. Indeed, in the first version of the model, the proportion of 1SW/MSW fish per cohort is supposed to vary over time at the national level, without consideration of the variability among regions. This strong assumption may bias the estimates of 1SW and MSW national returning abundances especially in the early years of the time series where there is less data. We can acknowledge that the proportions of 1SW and MSW could vary greatly between the different regions considered by regionalizing this proportion, thereby improving the modelling structure of this variable by regionalizing this proportion.

The second limitation of the PastSatAb model is that the model relies on various types of data that are produced and collected by a variety of stakeholders at different time of the year. As a result, some data of the year under assessment may not be available by the time of the WGNAS meets, which could alter the PastSatAb estimates for this year. Hence, before the model could be routinely used for estimating the return abundance time series, the relative importance of each of the different data sources must be assessed to evaluate the capacity to produce accurate estimates of abundance even in a context of missing data and to prioritize the data sources to be collated in priority.

To address those limitations, we first developed a new version of the model acknowledging that the proportions of 1SW could vary greatly between the different regions considered. This new version of the model was then used to assess the quality of the estimates that could be produced even in the presence of missing data. This assessment only focuses on year 2020, the last year for which all data are available.

First, we assess the performance of our new model to estimate abundance in a virtual situation where only the catch data are available, a likely situation at the time the WGNAS meets. This allows us to compare the estimate derived from the run-reconstruction (that is only based on catches raised by harvest rates) with those derived from the PastSatAb model with input data limited to catch data. Those estimates are compared to the "reference estimates" (considered as the "truth") obtained from the PastSatAb model run with all data available.

In the second part, we assess the value of information brought by each of the data sources integrated in the PastSatAb framework, seeking to identify which of the data source is critical to accurately estimate abundance. The assessment was expanded to different scenarios of availability for the four data sources used by the PastSatAb model, namely, the regional and river estimation of absolute abundance, the observation of river abundance made by counting devices (referred to as direct counts) and the catches. The PastSatAb model was run under seven data removal scenarios, four theoretical scenarios where one data is removed at a time in order to assess the informative value of each data source—how each data source contribute to the accuracy of the national returning abundances—and three operational scenarios that data sources are removed sequentially from the least available data to the most available data.

II – Materials and Methods

II.1. Characteristics of French salmon stocks

In France, around 50 rivers along the Atlantic coast host or have hosted populations of Atlantic salmon (Lebot et al., in prep). These populations are managed at the scale of five large hydrographic basins (called regions thereafter), each one with its own management committee. These five regions are the Nord, the Bretagne, the Loire-Allier, the Garonne-Dordogne and the Adour (Figure 2). Note that the Nord region is formed by the Artois-Picardie and Seine-Normandie management committees, given the similarities in population dynamics and exploitation of the species in these two regions.

Depending on the region, salmons have different dynamics. We can observe a latitudinal gradient of the size, lifespan and proportion of 1SW. In the Nord and Bretagne regions, populations are smaller (because rivers are usually of smaller size), have a shorter lifespan and adult returns are dominated by the 1SW component. For the regions Loire-Allier, Garonne-Dordogne and Adour, populations are larger, and are mainly composed of MSW individuals.

The Bretagne and Adour regions contribute most to the abundance of the French stock, with historically around 60% of the stock found in Bretagne and around 30% from Adour (Lebot et al., in prep). The Garonne-Dordogne region's contribution to the stock has historically been close to zero.

French salmon populations are exploited by two different fisheries: the professional commercial fishery, essentially operating in estuaries using nets, and a recreational rod-and-line fishery in the rivers. These fisheries are strongly regulated. A fishing license is required since 1987. Estuary fishing is carried out exclusively by professional fishers. It is practiced in the Bretagne and Adour regions. It was practised in the Loire-Allier region before fishing was prohibited throughout the river basin in 1994, following the closure of the Garonne-Dordogne region to all fishing in 1987.

Bretagne is the region where the recreational rod-and-line fishery is the most widely practiced and where catches are highest. Total Allowable Catches (TACs) were introduced at the end of the 1990s in the Bretagne and Nord regions. The Adour is the only region where fishing is authorized without a TAC.

Figure 2: Map of the five French regions considered within the framework of the PastSatAb project (Lebot et al. in prep)

II.2. Outlines of the PastSatAb model

Presentation of the pre-existing PastSatAb model

The PastSatAb model has been developed within an integrated, hierarchical Bayesian framework. Figure 3 illustrates the structure of the model and the data associated in a simplified direct acyclic graph. The return abundance of Atlantic salmon of both age classes (1SW and MSW) is modeled across the five regions studied and for the 1970-2020 period. The salmon abundances are modeled from returns to the shore after the end of their marine migration to spawning escapement in river using latent variables and parameters with associated priors that are informed by multiple data sources and incorporating observation errors. The model's scale of study enables information on return abundance to be integrated on a fine spatiotemporal scale. The model produces Bayesian estimates (posterior distributions) of 1SW and MSW abundances at regional level, taking into account the spatiotemporal heterogeneity of their dynamics. The estimates produced at regional level are aggregated to obtain estimates at the national level.

Notations

In the variables notations (see Figure 3), "*N*" refers to the abundance, "*Y*" to the catches, "*h*" to the harvest rate, "*q*" to the proportion of 1SW in the abundance and "η" to the regional redistribution factor. The number in the variable indicates the spatial area it refers to, with "1" indicating the sea area just before the estuary, *N1tot* representing the total abundance of salmon at sea during their return migration to the French coast. The number "2" refers to estuarine zones before net fisheries occur, "3" corresponds to river entry, after the estuary net fisheries and before rod-and-line fishery. "4" designates upstream river zones with *N4* representing the escape abundance. When an abundance variable has the suffix "*tot*" (*N1tot, N2tot*), it refers to sea-age-aggregated abundance. When there is no "*tot*" (*N2, N3, N4*), the variable is broken down into sea-age classes.

Spatial structure

The model processes and data are structured according to three embedded spatial scales: the national scale, the regional scale and the river scale (Figure 3). Five regions have been defined and correspond to the perimeter of the regional management committees hosting at least one salmon population: the Nord, Bretagne, Loire-Allier, Garonne-Dordogne and Adour regions (Figure 2). In each region, the model tracks changes in salmon abundances in three different areas used by salmons on their return migration, namely the maritime zones preceding estuary entrances, estuary zones and rivers. Abundance and the fishing mortality are distributed sequentially within these three areas. The abundance of salmon coming from the sea before the return to the spawning grounds is called the national returning abundance, and the individuals that survive fishing pressure in the estuary and river are referred to as the escapement abundance. Note that the river scale is used only in two regions, the Nord and the Adour, for which abundance information are only available at that scale.

Sources of mortality

Two different sources of mortality are considered: fishing mortality and all other sources of mortality (natural mortality, predation, bycatch, environmental causes). The model takes into account two types of fisheries that induced targeted mortalities on adult salmon: the net fisheries and the rod-and-line fisheries. Net catches are mainly made by professional fishers in estuaries, and rod-and-line catches are carried out by professional or recreational fishermen in rivers. Distinguishing these two fisheries allows for considering the variability of fishing pressure depending on the type of fishing and the region. For example, net fishing has been practiced for many years in the Adour region, but is considered negligible in the Nord region due to very low catch reports. In this study, all sources of mortality other than fishing mortality are grouped together under a single term. However in the absence of reliable sources of information, we considered that there was no other mortality than fishing mortality (See discussion in Lebot & al. (in prep)) and this parameter is then not represented in Figure 3.

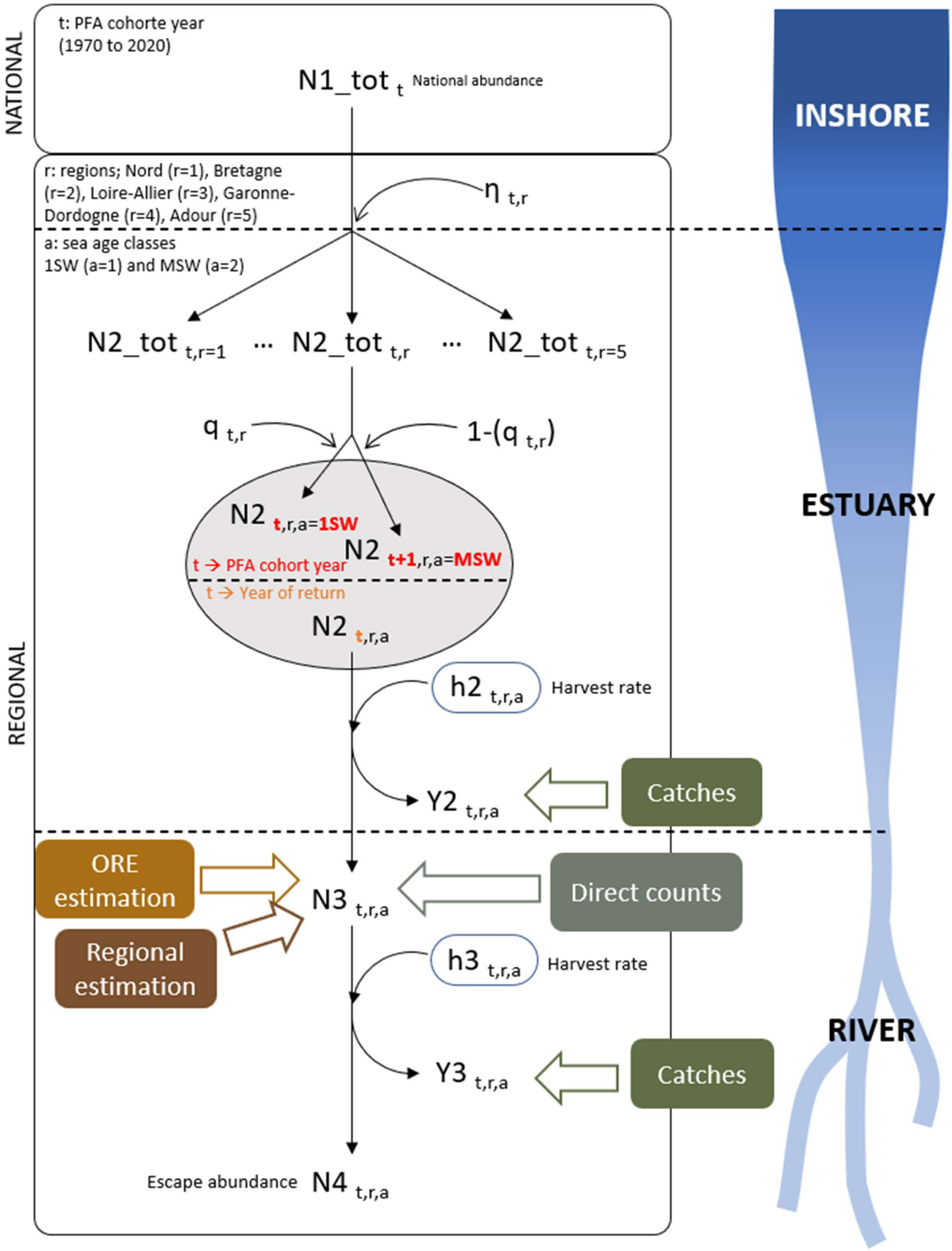


Figure 3: Simplified Direct Acyclic Graph (DAG) of the hierarchical structure of the modified PastSatAb model. Boxes in green and orange indicate the data sources integrated in the model. Those data are connected by an arrow to the latent variable in the model.

Data integration

Several data sources are integrated in the model. In France, the data and expertise collected are highly heterogeneous in nature (catches, total and partial abundance estimates), accessibility over time and space, and reliability. For all data sources, estimates are modelled with observational errors using the plug-in pseudolikelihood method described in Lebot et al. (in prep) to cope with the reliability of each data sources over space and time.

Hypotheses were set to compensate the lack of information on some components of the model. Informative priors or parameters values were set from expert knowledge. Ancillary data were also used to set some parameters. For instance, the efficiency of counting devices was set to 80% for all rivers and all sea-age classes. In addition, the allocation of regional abundance within rivers was set on average proportional to the surface area of habitat available for juvenile development. The average contribution of a river to the regional abundance was set as the ratio of the river's surface area to the total surface area of the region, meaning that a larger river will contribute more than a smaller river (for more detail see Lebot & al. (in prep)).

II.3. Presentation of the data

The data are highly heterogeneous, both spatially and temporally. They are available for the two sea-age classes, at different spatial scale and they are produced and collected by multiple actors, mainly by the "Office Français de la Biodiversité (OFB)", the "Direction Générale des Affaires Maritimes, de la Pêche et de l'Aquaculture (DGAMPA)", the "Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE)" and the "Fédération Nationale de la Pêche en France (FNPF)". As data for 2021 to 2023 were not fully available at the time of the study, we only used data from 1970 to 2020.

II.3.1. Catches

The model integrates data on the number of salmons caught after returning to estuaries following their marine phase. Catches realized during the marine migration, such as those in the West Greenland fisheries, or bycatch at sea are not taken into account. In France, catch reporting has been mandatory since 1987 (October16th 1996 Arrêté). The data reflect only reported catches. For each fishery, data integrated include the total number of fish caught, regardless of the sea-age classes and the two sea-age class among the number of fish caught all sea-age, expressed as proportion of 1SW. Data are not available for all years and regions, and assumptions have been made to some parameters of the model, such as the harvest rates to account for missing data (See section II.2.4. of Lebot & al. (in prep), for more information on the assumptions of the PastSatAb model).

II.3.1.1. Total number of fish caught

In France, catches reached 2 444 individuals in 2020. Over 50% of these catches were taken by net fisheries in estuaries, and almost entirely in the Adour region. At the national level, the MSW sea-age class is the most heavily targeted, accounting for 59.7% of total catches.

Net fisheries in estuaries

For Bretagne, data are only available for the 4 years 1970, 1971, 2019 and 2020 (Bagliniere & Fontenelle, 1972). For the Loire-Allier and Adour regions, since fisheries are or have been more developed, catch data are more available and reliable. Estuarine catches from the Loire-Allier have been available since 1970 and until 1993, the last year before the closure of salmon fisheries in this region (Lebot et al., in prep). For the Adour, catch data are considered reliable only from 1987 onwards. There is no reliable information for the Nord region and the Garonne-Dordogne because this fishery is not developed in these regions.

Rod-and-line fisheries

Before 1987 and the mandatory reporting, rod-and-line catches were estimated from various sources of information. For the 1970-1986 period, data are mainly taken from Thibault (1996) (in Gueguen & Prouzet, 1994) (except for Garonne Dordogne in which fishing was prohibited until 1985). For Nord and Bretagne regions, the dataset did not cover all rivers and the missing catches (those in the uncovered rivers) were reconstructed from the average relative contribution of these rivers in the total catches of the region. For the period 1987-2020, catches by year and by region are directly drawn from the CNICS database.

II.3.1.2. Proportion of 1SW in the catches

Two main methods are used to determine the sea-age class of adult salmon. The first consists in estimating the age of fish from its size using size-age relationships. The second method consists in estimating age based on scale reading (Baglinière et al., 2020). Today, these methods are regularly used to determine the sea-age class of some individuals caught by different fisheries, but this has not always been the case.

Net fisheries in estuaries

Adour is the region with the most information of 1SW proportion on the net catches. Among the 2 275 fish caught that have been aged during the period 1970-2020, 1,053 were 1SW fish. Hence, in this region almost the same proportion of 1SW (46%) and MSW (54%) is caught by the net fishery (Lissardy & Morandeau, 2006; Marty & Bousquet, 2001). For the Loire, information on sea-age is available for only five fishing seasons from 1989 to 1993. The proportion of 1SW in the sampled catch is very low (3%) in this region. Data on 1SW proportion are not collected in the Nord, the Bretagne and the Garonne-Dordogne regions.

Rod-and-line fisheries

Before 1987, there was no extensive monitoring of the sea-age of the fish caught in the rod-and-line fishery. Only around 10% of the total catch between 1970 and 1986 were aged, among which only 5% were 1SW (Bagliniere et al., 2004; Marty & Bousquet, 2001). Since 1987, rod-and-line fishers have been required not only to report their catches but also to submit scales for aging the fish. This requirement has significantly improved the accuracy of 1SW proportion estimates from that time onward. Now, the vast majority of individual caught (>90%) have been aged, providing a more comprehensive understanding of the 1SW proportion (Lebot et al., in prep).

In the subsequent, total number of fish caught and proportion of 1SW in the catches will be referred to as “catches”.

II.3.2. Abundance data: estimates and direct counts

The abundance data mentioned in the following sections refer to the abundance of adults returning to the rivers, i.e., individuals that have survived net fishery but could be caught by rod-and-line fishery. They are available either at a regional scale or at a river scale.

II.3.2.1. Regional estimations

For the three regions Loire-Allier, Bretagne and Garonne-Dordogne, models have been developed to estimate the absolute abundance of adult returns and for each sea-age class. These models (referred to as ad-hoc models) incorporate the most pertinent data and expert knowledge. They generate Bayesian posterior distributions of abundance for each year and each sea-age class separately, that are integrated into the PastSatAb model using the plug-in pseudo-likelihood method. Data from these models are referred to as "regional estimations" in the subsequent document.

In Loire-Allier, a model has been developed for the Allier tributary (Dauphin & Prévost, 2013). This model estimates the number of adults per sea-age class returning to the Allier to reproduce since 1977. More than 85% of the abundance in this region comes from this river, as it offers the largest area of suitable and accessible breeding habitat. Counts from other rivers were added to the Allier estimates to obtain the total abundance of the region (Lebot et al., in prep).

In Bretagne, a model to estimate the absolute abundance of 1SW and MSW adult returns between 1987 and 2020 was developed in the RENOSAUM project which aimed to renew the regulation system for the exploitation of Atlantic salmon by rod-and-line fisheries in Bretagne (Lebot et al., in prep). The regional estimate is obtained by summing the abundances of the 18 main salmon rivers of the region (Prévost & Lebot, 2022).

The Garonne-Dordogne abundance survey is not similar to the models developed in the other two regions. Management in this region is a special case, as fishing has been banned since 1985, and spawning grounds were only reopened in 1988 thanks to the construction of fish passes at the Golfech and Tuilières dams (Lebot et al. in prep.). Counting devices were installed in 1993 for Golfech and 1994 for Tuilières. Consequently, for this region, it was assumed that there were no fish in the rivers prior to the start of counts in 1994, and the regional abundances entering the rivers were set to 0. From 1994 onwards, lognormal pseudo-likelihood was used to model the observation process of the regional abundance in the Garonne-Dordogne with the adult return abundance set to the sum of counts on these two passes and a very low coefficient of variation (fixed to 3%) (Lebot et al., in prep).

II.3.2.2. River estimations and direct counts

We use estimations of rivers abundance for Nord and Adour regions where estimates are not available at regional level.

Absolute abundance with river estimation models

Each of the two regions has an index river—the Bresle for the Nord region and the Nivelle for the Adour region—where capture-mark-recapture data are collected with a trap device (Lebot et al., in prep). These data are incorporated in ad-hoc models (Servanty & Prévost, 2016) to produce the absolute abundance estimates utilised in this study. They are available for the period 1984-2020 for these two rivers and are referred to as "river estimation" in the subsequent of the document.

Partial counting

The model also incorporates partial count data from counting devices (video counting, trapping devices) available on certain rivers. Counting devices are in operation on four rivers (the Ternoise, the Touques, the Vire and the Orne) in the Nord region, and three rivers (the Nive, the Gave de Pau and the Gave d'Oloron) in the Adour region. The data collected from these devices were treated as partial observations of the absolute abundance in these rivers, acknowledging that some individuals might be missing in the total count. This could happen if they remained downstream of the device or were not recorded during their passage, which can occur during events of floods or failure of the counting device (Lebot et al., in prep). These partial adult return abundance data are referred to as "direct count".

II.4. Improvements of the PastSatAb model: regionalization of the 1SW proportion

Below we detail the modification brought to the model to improve the regionalisation of the 1SW proportion parameter.

In the pre-existing model, the process of ageing the total abundance *N1tot[t]* occurs before the regionalization process to fit with the life cycle of salmon populations. The proportion of 1SW/2SW is modelled at the national level through a parameter *q[t]* that varies across years through an exchangeable hierarchical structure (random effect) (1):

$$(1)\ \mathrm{logit}(q[t]) \sim \mathrm{Normal}(\mu^{q}, \sigma^{q})$$

Hence, the abundance of salmon per sea-age class and region (*N2[t,r,a]*) is obtained by multiplying *N1tot[t]* with the proportion of 1SW *q[t]* and the regional distribution factor *η[t,r,a]*. The drawback of this modelling structure is twofold: first, it has the effect of shrinking the estimates of q[t] around the national grand mean μ^{q}. Second, it constrains the spatial variability of the estimates of the proportion of 1SW/2SW among the different regions.

The shrinkage effect is particularly pronounced before the 1990s when information is limited. Thus, for those years, the regional estimates of 1SW proportions are shrinked to the national grand mean whereas they exhibit significant divergence afterward when the information becomes more precise (Appendix I). This effect is likely to result in an overestimation in the 1SW proportion for the Loire-Allier region and an underestimation for the Bretagne region (Appendix I). To avoid the shrinkage effect and produce more reliable estimates of the proportion of 1SW over the entire time series studied, we regionalize the proportion of 1SW. To allow this parameter and its standard deviation to vary with the regions, the model was modified as follows. It was assumed that the total abundance is first distributed across the five regions using a regional distribution factor $\eta_{t,r}$ no longer distributed according to the sea-age of the salmon ($\eta_{t,r}$ instead of $\eta_{t,r,a}$). Then, each regional abundance (*N2tot[t,r]*) was divided into the two sea-age classes with a regionalized proportion of 1SW *q[t,r]*. An exchangeable hierarchical structure within each region has been set to *q[t,r]* with a mean (μ^{q} *[r]*) and a standard deviation (σ^{q}*[r]*) defined per region (2).

$$(2)\ \mathrm{logit}(q[t,r]) \sim \mathrm{Normal}(\mu^{q}[r], \sigma^{q}[r])$$

The simplified direct acyclic graph (Figure 3) summarizes the revised overall structure of the model. Note that the parameter *q[t,r]* is defined per pre-fishery abundance cohort year. The time index refers to the smolt cohort, which is more relevant in terms of ecology (a certain proportion of post-smolt will mature after one or two years spent at sea). *q[t,r]* is different than the proportion of 1SW/MSW in the return year t, which might be more significant in term of stock assessment. In the results, the proportion of 1SW/MSW will be presented by year of return.

We compared the Watanabe-Akaike information criterion (WAIC) values of the two model versions. The WAIC is a fully Bayesian approach, and a generalized version of the Akaike information criterion (AIC), that estimates the out-of-sample expectation. It calculates the log pointwise posterior predictive density and then adds a correction for the effective number of parameters to account for overfitting (Gelman et al., 2013). This criterion measures the quality of a statistical model by penalizing models according to the number of parameters in order to satisfy the parsimony criterion, based on a compromise between goodness of fit and model complexity. The model with the lowest WAIC value is considered the best-performing.

II.5. Assessing our ability to estimate abundances in the context of missing data

First, we assess the performance of our new model to estimate abundance in a virtual situation where only the catch data are available. To this end, we compare the reference estimates (considered as the "truth") obtained from the PastSatAb model run with all data available with estimates derived from the PastSatAb model with input data limited to catch data and those obtained from the run-reconstruction that is only based on catches raised by harvest rates.

In the second part, we seek to identify which of the data source is the most critical to accurately estimate abundance from the PastSatAb model. To this end, we run the PastSatAb model under several nested data depletion scenarios to assess the value of information brought by each of the data sources.

II.5.1. Assessing abundance estimates from different models using only catch data

The choice of the method for estimating national returning abundances for the year preceding the assessment of the WGNAS results from a compromise between data availability and the cost of obtaining it. Because the catch data are the most easily accessible data sources, it is relevant to assess how informative this data source can be for estimating national returning abundances.

In order to assess our abilities to produce reliable estimates of national returning abundances using only catch data, we compared the estimates produced by three different models to the one considered as our reference. Reference estimates of abundance were obtained by the PastSatAb model integrating all the data sources for all years.

Those reference estimates were first compared to abundance estimates from the PastSatAb model when only the catches were integrated in 2020 (all other data sources integrate for all other years). All the other data sources, i.e., regional and river estimations as well as direct counts are considered as missing data for the year 2020.

Reference estimates were also compared to abundance estimated from the run reconstruction model currently used by the WGNAS for the production of national returning abundance time series. In this model, estimates of the national returning abundances per sea-age class are produced by dividing homewater catches by harvest rate (3):

$$(3)\ N[t,a] = \frac{Y[t,a]}{h[t,a]}$$

where *N[t,a]* is the national abundance of sea-age class a returning the year t, *Y[t,a]* is the total catch (integrating reported and unreported catches) of sea-age class a made the year *t* by all fisheries combined, and *h[t,a]* is the average global harvest rate exerted by all fisheries combined relative to the abundance *N[t,a]*. This model uses official time series of catch data provided by ICES and an expertise on the average global harvest rate. The main limitation of this approach is the lack of transparency in the production of these harvest rate estimates. Very little information is available on how their values have been set and on the justification of the variations over the time series. As a result, since 1995, the same probability distributions have been set to 1SW and MSW harvest rates. They are drawn in two uniform distributions between the intervals [5%; 20%] and [20%; 40%] respectively (ICES, 2024). To compensate for potential unreported catches, the run reconstruction model applies a correction to the reported catch data from 2002 onwards using unreported catch rates. Just like the harvest rates, no information is available on how their distributions have been set. Since 2002, the 1SW and MSW unreported catch rates are drawn in uniform distributions between intervals [20%;40%] and [15%;30%] respectively. Note that this introduces a difference in the assumptions between the run-reconstruction model and the PastSatAb model which uses only reported catch time series considering that information on unreported catches is not sufficient to be integrated to the model.

We assessed estimates derived from a third model that used the same equation (3) than the run-reconstruction model but implemented with the data and output of the PastSatAb model. We refer to this model as the "combined model" because it represents a balance between the data and outputs from the PastSatAb model and the run-reconstruction equation. In the combined model, the probability distribution of the 2020 catches, *Y[2020,a],* is directly taken in a Lognormal distribution centred on the average raw catch data of 2020 and the variance set as the observation variance error used in the PastSatAb model. Given that no information is available on 1SW and MSW harvest rates in 2020, we assume that they should be the same on average than the last five years 2015-2019 average of 1SW and MSW harvest rates as estimated by the PastSatAb model with the full dataset. Hence, we supposed that 1SW and MSW harvest rates were drawn in beta distributions with means equal to the 2015-2019 average of the posterior distributions of 1SW and MSW harvest rates. The coefficient of variation of these distributions have been set to the average annual variability of the harvest rates over the whole 1970-2020 period (taking the variance over the last five years would have led to underestimate the temporal variability). In this way, we take advantage of the information provided by the variety of data available in past years, while producing an estimate with a simple formula for the year for which we are producing an estimate.

The results are represented by the national returning abundance, which groups together the two age classes and the five regions, and by the proportion of 1SW in the national returning abundance.

II.5.2. Model performance in the event of missing data

We assess the capacity of the PastSatAb model to produce accurate abundance estimations even in the presence of missing data. To do so, we evaluated which of the data source is the most critical to accurately estimate abundance from the PastSatAb model. We run several data removal scenarios, removing only 2020 data, the last year of the datasets, and compare the abundance and the 1SW proportion estimations to the reference estimations obtained with all the data (scenario 0).

First, we set up “theoretical” scenarios in which we remove each data source one by one, in order to quantify the informative value of each of the four data sources. In these scenarios, named scenarios 1,2, 3 and 4 (Figure 4), we removed the catches, the direct counts, the river abundance estimates and the regional abundance estimates respectively.

We also created “operational” scenarios based on the relative accessibility of the different data sources. Catches of year t-1 are the most readily available data at the time the working group is held, say in March year t. Indeed, their declaration is mandatory in France and requested by Europe every year via a data call. They are therefore the most accessible data in the model. The direct count of year t-1, which is not mandatory but is collected each year for the various rivers under consideration, is ranked second in terms of accessibility in March year t. Indeed, collating these data “only” require contacting the departmental leisure fishing federations and migratory associations, which collect and produce it, in order to gain access to it. Absolute abundance estimates in the French index rivers (river estimates) involve running estimation models based on catch-mark-recapture data. A meeting of experts is held each year in late winter/early spring year t to update these estimates based on data though year t-1. However, there is no guarantee that this working group will be held before the WGNAS meeting. Hence, river abundance estimates are considered less accessible than direct count. Finally, the regional estimates are produced by models that take advantage of the ORE model estimates. The production of these estimates is therefore dependent on the production of river abundance estimates, and the regional estimates are considered as the least accessible data used by the PastSatAb model.

“Operational” scenarios 4, 5, 6 and 7 mimic plausible scenarios, at the time the WGNAS holds its annual meeting, that cumulate the removal of several data sources simultaneously while following the availability gradient (Figure 4). The least available data source is removed first, then the second least available, etc., until all four data sources are removed (scenario 7). In addition to being a theoretical scenario, as it tests the removal of a single data source, scenario 4 is considered the 1st operational scenario, as it corresponds to the removal of the least available data (the regional estimations of abundance). For the sake of clarity, we will consider it as a theoretical scenario in the rest of this report. Given the recurrent availability and amplitude of this data, scenario 6 is of particular interest in this study. In the scenario 7, no 2020 data is available to estimate abundance in 2020. The model then behaves as pure forecast for 2020. This scenario is unlikely at the time of the WGNAS annual meeting. It does, however, allow us to assess the forecast performance of the model

To compare estimates of returning salmon abundance, we focus on the estimates of the national returning abundance (*N1tot[t]*) as well as the 1SW proportion of in this abundance (q[t]). Because of their distinct dynamics and varied spawning potential, Atlantic salmon management pays particular attention to this proportion. We also investigate regional returning abundance estimates (*N2[t,r,a]*) since the model produces estimates at the regional scale. For the sake of consistency with the current estimates used by the WGNAS, the results will be presented by year of return although these two parameters are modeled by cohort year.

Data availability at the time of the assessement

More available — **Raw data** | **Model outputs** — **Less available**

		Catches	Direct counts	River estimation	Regional estimation
	Scenario 0 (baseline)	x	x	x	x
Theoretical	Scenario 1	-	x	x	x
Theoretical	Scenario 2	x	-	x	x
Theoretical	Scenario 3	x	x	-	x
Theoretical	Scenario 4	x	x	x	-
Operational	Scenario 5	x	x	-	-
Operational	Scenario 6	x	-	-	-
Operational	Scenario 7	-	-	-	-

x : data until 2020
- : data until 2019

Figure 4: Table summarizing the data removed under various data deletion scenarios. In scenario 0, all data sources are available through 2020. Theoretical scenarios (1 to 4) involve the removal of one data source at a time and operational scenarios (5 to 7) involve the removal of multiple data sources simultaneously, following a data availability gradient from the least to the more available data. Each line represents a specific scenario, with data up to 2020 indicated by an "X" (not removed) and data up to 2019 (2020 removed) indicated by a "-".

II.6. Comparing the estimates from each scenario

The estimates of national returning abundance and 1SW proportion from each data removal scenarios are compared to the baseline scenario by a divergence index that measures the dissimilarity of the posterior distributions of interest, combined with the intensity and the direction of the over- or under-estimation bias.

II.6.1. The Kullback-Leibler divergence index

The Kullback-Leibler divergence (KLD) is an index used to quantify the dissimilarity of two probability distributions. The lower the overlap between the two distributions, the higher the KLD index.

Consider two probability density function $f(.)$ and $g(.)$. The Kullback-Leibler divergence between $f(.)$ and $g(.)$ is defined as:

$$KLD\big(f(.), g(.)\big) = \int_{-\infty}^{+\infty} f(x) \cdot log\frac{f(x)}{g(x)} \cdot dx \qquad (4)$$

The KLD has two important properties: its value is always positive and the KLD between a distribution $f(.)$ and itself is zero.

A drawback of the KLD is that it does not consider the sign of the bias between two distributions. For instance, the KLD between a Normal distribution N(0,1) and N(1,1) is identical to the KLD between N(0,1) and N(-1,1), meaning that a bias toward lower values will result in the same KLD as a bias towards higher values.

To address this drawback, we combine the KLD with a second index for the bias direction and intensity (see below).

In our application, the distributions are available through Monte Carlo samples. The KLD were calculated from Monte Carlo samples using the *KL.divergence* function with the *FNN* R package.

Example of the modified Kullback-Leibler divergence:

To illustrate how the index varies given the divergences of probability distributions similar to the ones produced by the model, we calculated the modified KLD between a reference lognormal distribution with a mean of 17 000 and a coefficient of variation (CV) of 13%, and several log-normal distributions with means ranging from 3 000 to 30 000 and CV values ranging from 1% to 42 (Figure 5).

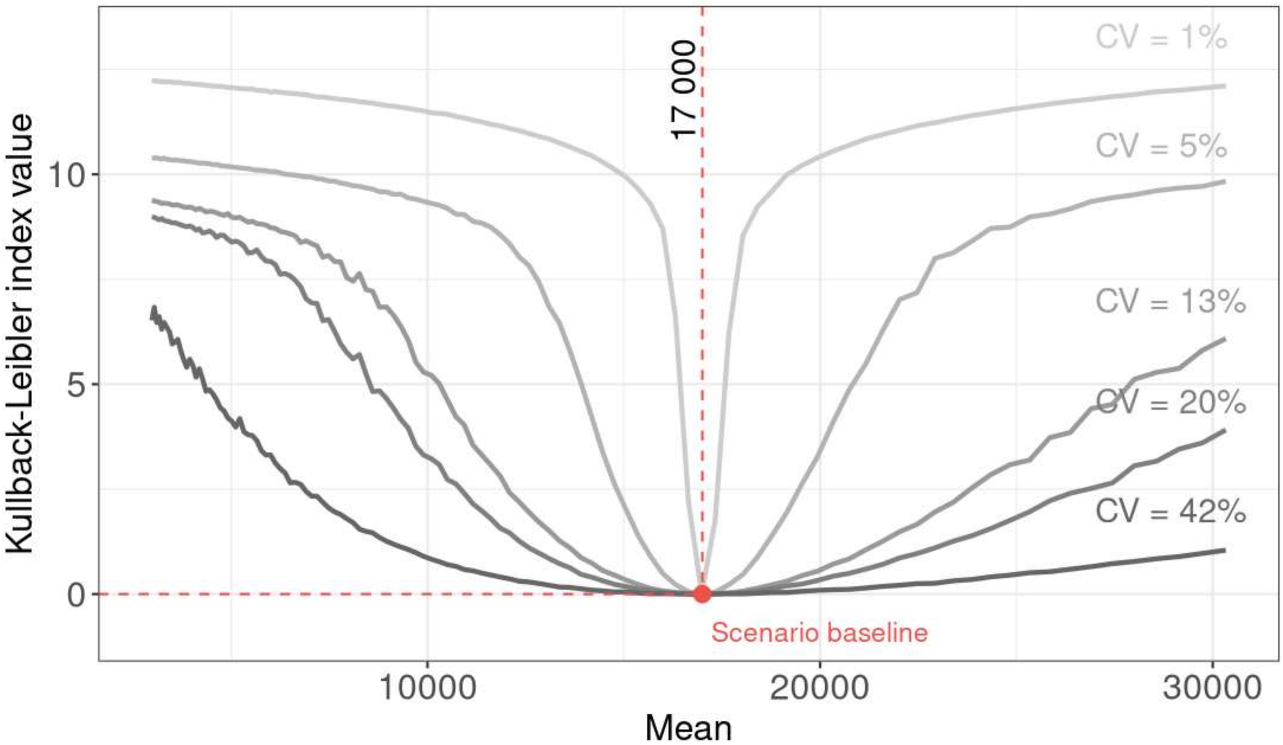


Figure 5: The Kullback-Leibler index value, represented as a function of mean and coefficient of variation, obtained by comparing a reference lognormal distribution with a mean of 17,000 and CV = 13% (a value that most closely represent the CV observed in the data) with different log-normal distributions with a range of negatively and positively biased means and a range of CV from 1% to 42%. The red dot represents the mean of the reference distribution and the KL index value obtained when compared to the alternative distribution.

As shown in the examples (Figure 5), a given value of KLD index can be obtained from a variety of combination of mean and CV.

- For the same mean, the KLD index increases when the CV decreases. Indeed, the lower the coefficient of variation, the more the distribution density will be concentrated around the mean, leading to less overlap between the distributions.
- In the same time, for a given CV, the KLD index increases as the difference between the means grows. The values of KLD index increases following a sigmoidal shape with an inflexion point which is closer to the reference distribution when the CV decrease. It means that the speed at which the KLD index increases in relation to the difference of the means increases as the CV decreases.

II.6.2. Measuring the intensity and direction of the bias

To assess and compare the estimates from various models or scenarios to those produced by the PastSatAb model when all data are available, we aim to understand the direction of dissimilarities measured by the Kullback-Leibler index in the national returning abundances and 1SW proportions. We compare the median estimates from the tested model or scenario with those from the PastSatAb model under scenario 0 by calculating their ratio. This ratio indicates whether the tested model overestimates or underestimates the median—a ratio inferior to 1 indicates underestimation, while a ratio superior to 1 indicates overestimation. The ratio also quantifies the multiplicative bias of the median estimate relative to the reference. The value of the ratio, the estimation bias, is presented in the form of a range of values expressed in percentage, from 10-50% to >200%, associated with blue or red gradient in the case of under- or over-estimates respectively. Indices in the 90%-110% range are shown in grey to indicate a small variation in the median tested compared with the reference median, which cannot be interpreted as a difference in the estimate.

II.7. Bayesian inference

The PastSatAb models were run using Markov chain Monte Carlo (MCMC) sampling to approximate the joint posterior distribution of all unknown quantities. All computations were carried out with the JAGS® software (version 4.3.0.; (Plummer, 2003)) running in rjags, Coda and Dclone packages of R (version 4.3.3, www.r-project.org). Three chains with identical starting points were run in parallel. After an initial phase of sampler adaptation of 5 000 iterations, 3 333 values were drawn for each chain. We set a burn-in phase of 20 000 iterations, and made a subset of 9 999 samples resulting from pooling 3 333 values from each of the 3 chains (1 iteration out of 200 for each chain). These 9 999 samples used to approximate the posterior distributions were examined graphically, using the 5^{th}, 50^{th} and 95^{th} percentiles, in order to verify their stability at the end of the MCMC sampling.

III – Results

III.1. PastSatAb model improvements

The regionalization of the 1SW proportion highly reduced the shrinkage effect observed in the previous model version for the early years of the time series (1970 to 1990) (Figure 6 and Appendix I). Note that from 1990 onwards, the estimates of 1SW proportions by region are the same for both model versions. These results can be explained because after 1990, the proportions of 1SW are adjusted to very precise and reliable abundance data and 1SW proportions in the catches whereas before, the data used to adjust 1SW proportions—only 1SW proportions in the catches—are much less reliable and accurate, causing the estimates of 1SW proportions to be mainly driven by the modelling structure.

Difference of estimates between the two models in the earliest part of the time series (1970 to 1990) are high. In the previous model, the proportion of 1SW in the Loire-Allier region before 1990 was shrinked to the national grand mean and estimated on average around 20%, with an uncertainty of between 5 and 45% (Appendix I). In the 1970s in Bretagne, the proportion of 1SW tended towards 72%, with an uncertainty of between 50 and 85% (Appendix I). With the regionalization of the 1SW proportion parameter, estimates of the proportion of 1SW are more variable among regions. It now takes on a value of approximately 10%, with an uncertainty of between 3 and 25% in the Loire-Allier region (Figure 6) and is now around 77%, with an uncertainty of between 60 and 85% in the Bretagne region (Figure 6). With the modified PastSatAb model, the regional 1SW proportion estimates are more consistent over the entire study period.

Finally, with the modified PastSatAb model we obtained a WAIC value of -1443.7 whereas we obtained a value of -1425.2 with the initial model, which indicate an improvement in the quality of the model. Consistently, the modified model results in a better fit of the estimates of the proportion of 1SW in catches (the *rho* parameter) to the data, particularly for net and rod-and-line fishing in the Loire (Appendix II).

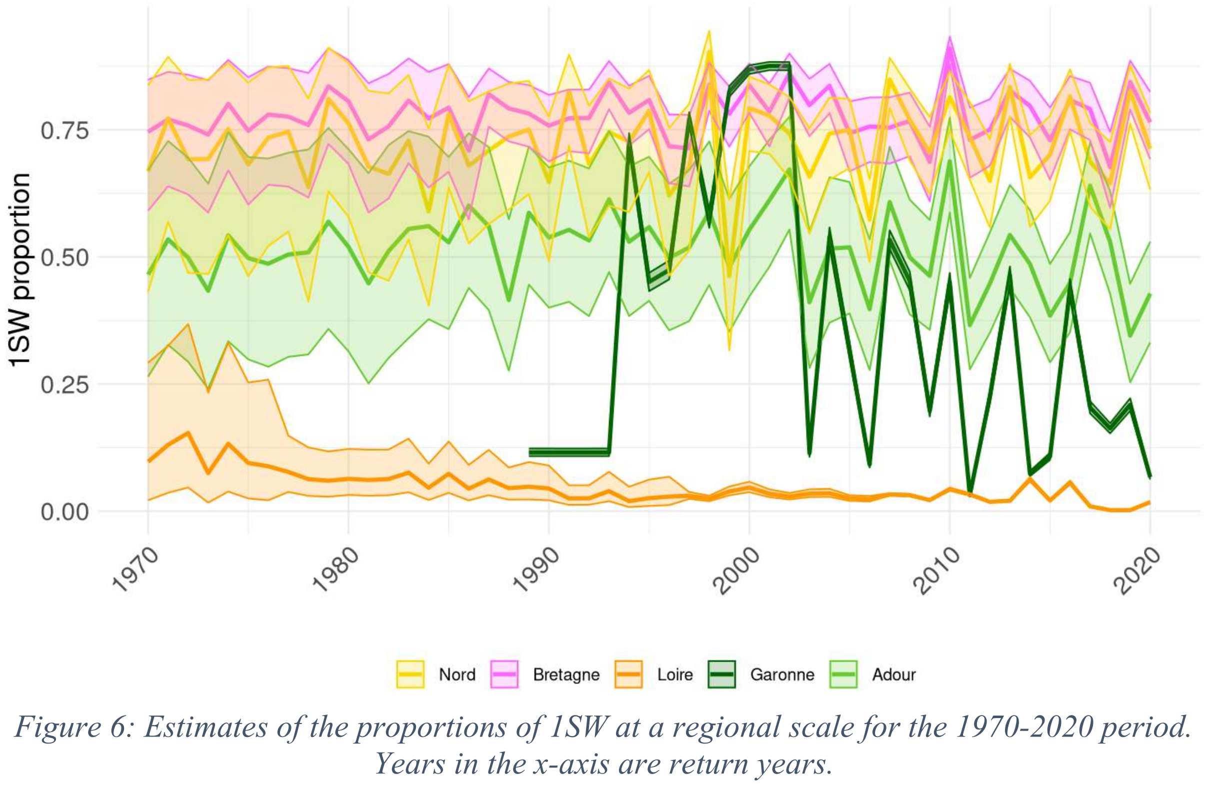


Figure 6: Estimates of the proportions of 1SW at a regional scale for the 1970-2020 period. Years in the x-axis are return years.

III.2. Comparison of the adult return abundance estimation from catch-only models

In the PastSatAb baseline scenario when all data are integrated in the model, results provide a median of national returning abundance around 17 000 individuals (Figure 7.a) and a median proportion of 1SW around 0.6 (Figure 7.b) with a low dispersion in both distributions. Distributions are narrow around the medians, ranging from 13 000 to 22 000 individuals for national returning abundance and between 0.5 and 0.7 for 1SW proportion (Figures 7.a and 7.b).

Running the PastSatAb model with only the catch data in 2020 (scenario 6) provides national returning abundance estimates very similar to those produced when all data are present (scenario 0) (Figure 7.a). The median of national returning abundance distribution is in the 90-110% range of estimation bias and the KLD index has a relatively low value (0.41) (Figure 8). For both the national returning abundance and the 1SW proportion, the distribution obtained under the scenario 6 has a greater uncertainty than those obtained under scenario 0. For the estimation of the 1SW proportion, the median of the scenario 6 is overestimated by 110 to 125%, however the KLD index is below 1, indicating a relatively moderate dissimilarity between the two distributions.

The models based on the classic run-reconstruction approach (catches raised by harvest rates) but with catches drawn in a lognormal distribution produce a broader distribution of national returning abundance than the PastSatAb model (under the baseline scenario), corresponding to a KLD index of 1.12. The median, however, is within the 90-110% range of the estimation bias. The median of the 1SW proportion estimation is identical to that of baseline scenario with a value of KLD index of 1.10, as the distribution is widely dispersed around this median. These highly dispersed distributions are the direct result of the significant variability added to the data used to estimate harvest rates and the lognormal observation errors around catch data.

The 2020 estimated median for the national returning abundance, obtained with the run-reconstruction model, is similar to that of the baseline scenario, with an estimated bias in the 90-110% range. The KLD index for this variable is 0.98, indicating higher dispersion of the estimation produced by the run-reconstruction model. For the 1SW proportion, the bias is also within the 90-110% range, however, Figure 7.b shows a higher estimated median compared to the baseline scenario. Its KLD of 0.71 reflects for the dissimilarity between the estimations produced by these two models. The distributions of national returning abundance and 1SW proportion from the run-reconstruction model are tighter and do not have distribution tails due to the application of a bounded uniform distribution to the harvest rate and the unreported catch rate parameters.

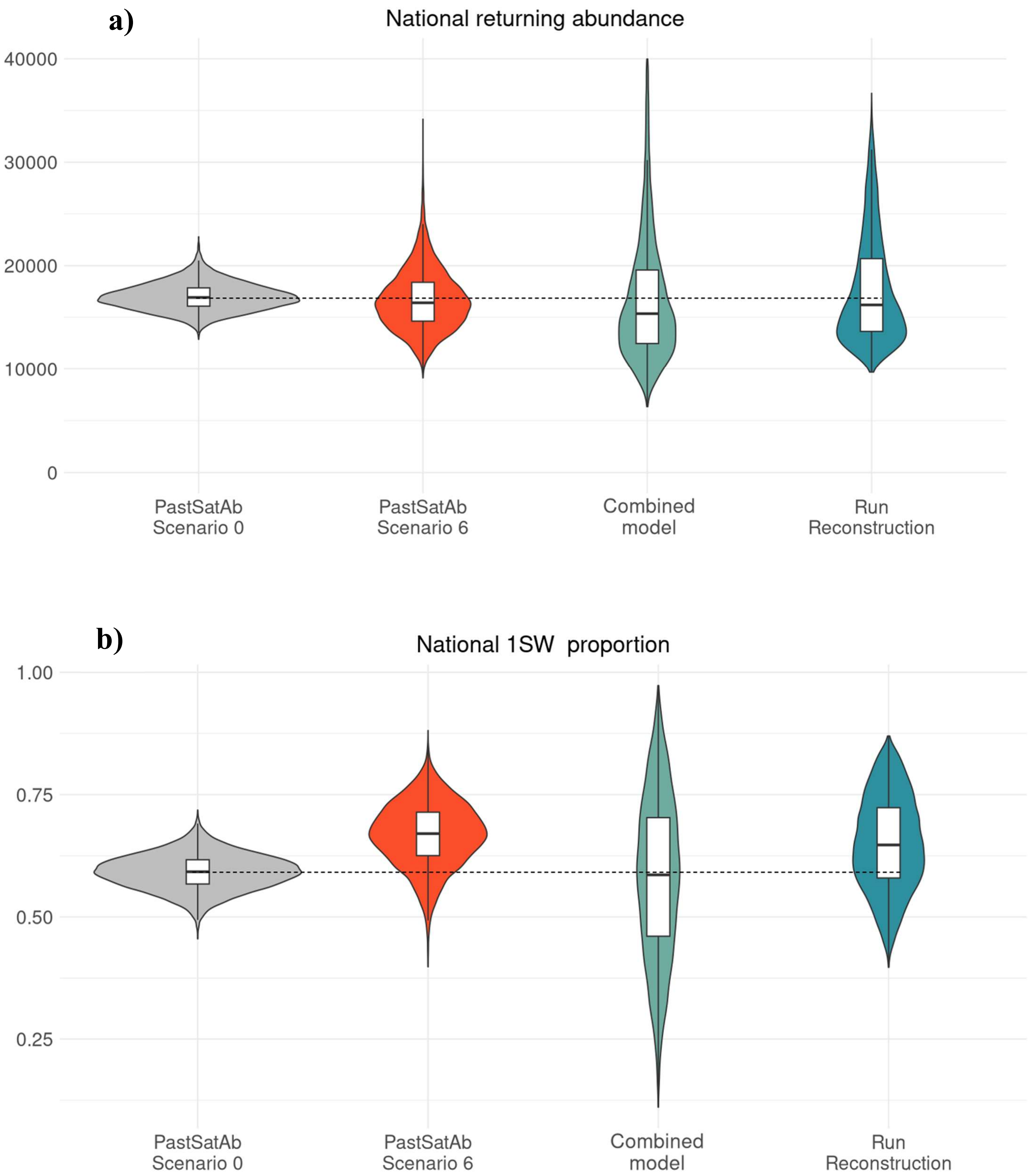


Figure 7: Estimates of the PastSatAb model, scenarios 0 and 6, the "Combined model" and the "run reconstruction model" for the national returning abundance estimations (in number of fish) (a) and for 1SW proportions estimations (b) for the year 2020. The boxplots display the median, the 25th and 75th quartiles and the density of the estimates. The dotted line extends the median value of scenario 0 to facilitate interpretation of the figure.

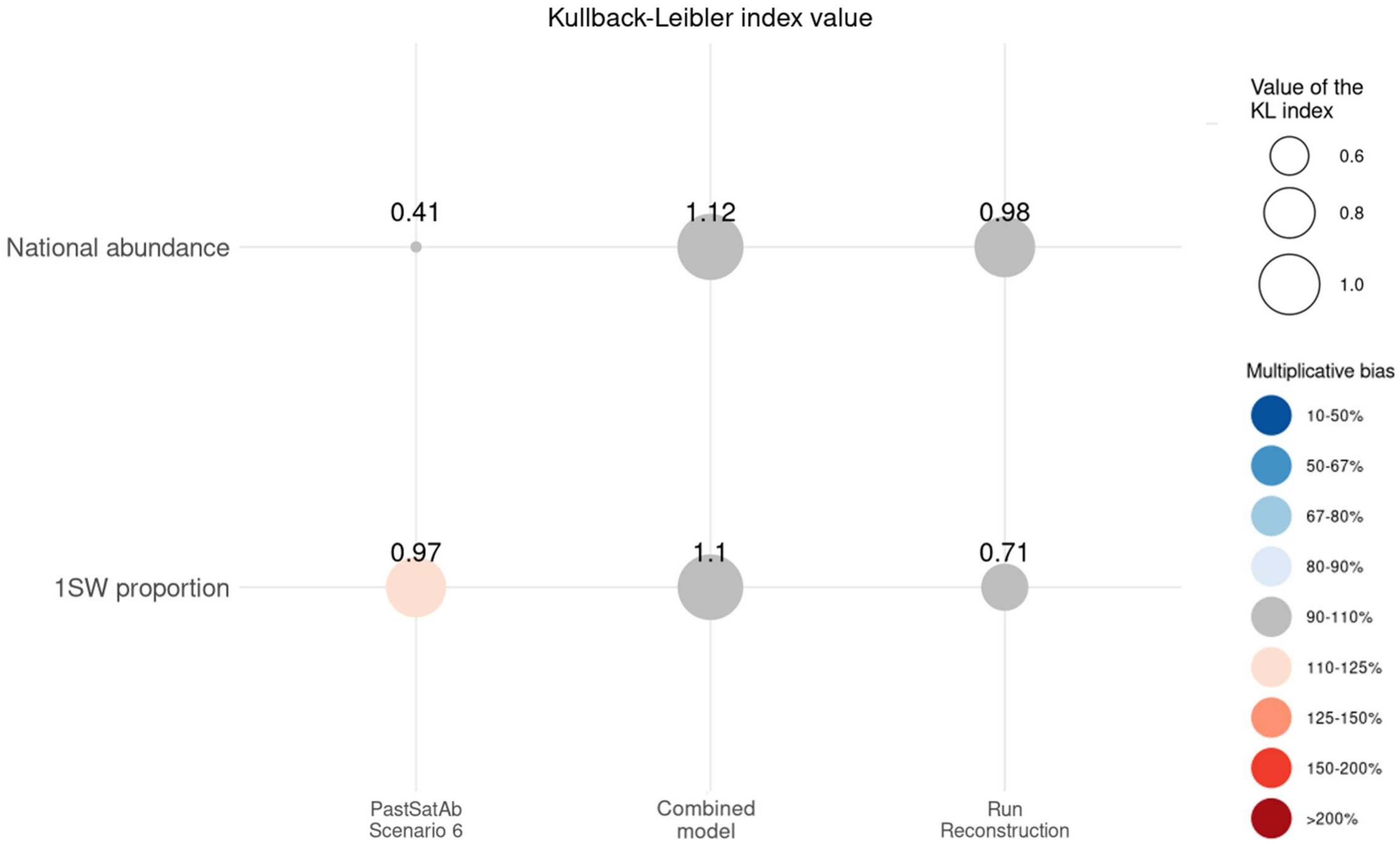


Figure 8: KLD indices values and multiplicative bias range obtained by comparing respectively of the scenario 6, the "Combined model" and the run reconstruction model estimates to the scenario 0 estimates for the national returning abundance estimations and for 1SW proportions estimations. The size of dots informs of the value of the KLD index (the value rounded to the hundredth is also shown above every dot) and the color of the dots informs the strength of the multiplicative bias.

III.3. PastSatAb model estimates with data deletion scenarios

Theoretical scenarios

At the national scale, all data deletion scenarios provide rather accurate estimations when only one of the four data sources is deleted (bias of estimations within the 90%-110% range, Figure 10) except for scenario 2, where the absence of direct counts leads to a fairly large overestimate of national returning abundance (Figures 9.a and 9.b).

Within the "theoretical scenarios", regional abundance estimates and direct counts data contribute the most to the estimations of the PastSatAb model. Deleting data from regional abundance estimates and direct counts abundance (scenarios 2 and 4 respectively) provides very different than estimates of national returning abundance and 1SW proportion estimates of the baseline scenario with all the data (Figure 9), their KLD indices being respectively of 0.68 and 0.47 for the returning abundance and of 0.58 and 0.48 for the 1SW proportion. By contrast, when these data are present (scenarios 1 and 3), the estimates are very close to the reference median and the KLD indices are lower than those from scenarios 2 and 4 (see Figure 10, "National" line). Except for the estimate of national returning abundance in scenario 2, the estimation bias of the medians of these scenarios for both variables remain within the 90%-110% range of the baseline median. The over-estimation of national returning abundance in scenario 2 is in the 110-125% range (Figure 10).

Some estimation biases observed at national level results from bias at regional level. The slight overestimation observed at national level for the national returning abundance under the scenario 2 is caused by strong overestimation at the regional level. At the regional scale, the returning abundance in the Nord region is overestimated by 150-200% and with a KLD index of 4.43 in scenario 2. Similarly, in this region the removal of the river abundance estimations data (scenario 3 and 5) results in an underestimation of the regional returning abundance with similar intensity in both scenarios, with an underestimation of the mean of the order of 50% to 67% and a KLD index value of around 3.8. These results show that, for the Nord region, most of the information comes from direct count data and river abundance estimation.

Operational scenarios

At national level, for estimates of returning abundance, the scenario for which only catch data are available (scenario 6) is the least biased compared to the baseline scenario. Its median is close to the baseline median (and within an estimate bias range of 90-110%, Figure 10) and its KLD index is of 0.41.

Scenario 5 incorporates count data and catch data. For returning abundances, the presence of count data results in an underestimate (cf. scenario 2 Figure 9.a), so scenario 5 produces an underestimated returning abundance compared to the baseline scenario (with a KLD index value of 0.95 and an estimation bias between 80-90%,) but also compared to scenario 6.

The scenario 7 produces the most extreme overestimations and the highest uncertainties for both variables. For 1SW proportion estimations at the national scale, the more data are removed the more the overestimation is important, with KLD indices of 0.27, 0.97 and 2, and bias of estimation within the ranges 90-110%, 110-125% and 125-150% for the scenarios 5, 6 and 7 respectively (Figure 10). The graphical representations of the estimations (Figure 9) illustrate that the informative value of the data is cumulative, as the biases of individual data source accumulate when multiple data sources are removed simultaneously.

When the regional abundance estimations are removed (scenarios 4 to 7), both returning abundance and 1SW proportion of the Loire-Allier and the Garonne-Dordogne regions are biased. In the Loire-Allier region, returning abundances are underestimated, with KL indices values around 3.7 and underestimation factors for median distributions of the order of 10 to 50%; and 1SW proportion is overestimated, with KL indices having values of around 4.8, and the overestimation of medians over 200% from scenario 4 to 7. For the Garonne-Dordogne, returning abundances are underestimated in the same way as for the Loire-Allier, with slightly lower KL indices (of the order of 4.4), and 1SW proportion estimations are relatively different to scenario 0 as the KL indices values are around 3.8 for scenarios 4 to 7. Their underestimation is however mitigated as their estimation bias is in the 67-80% range for the scenario 5, in the 80-90% for the scenarios 4 and 6, and in the 90-110% for the scenario 7 (Figure 10). This result shows that regional abundance estimations are the most informative source of data in these two regions.

For the Bretagne and Adour regions, returning abundances and the proportion of 1SW are relatively similar to those of the baseline scenario in all “theoretical” and “operational” scenarios. In “operational scenarios”, for both for national returning abundances and 1SW proportion variables, the uncertainty around predictions increases as more data are removed from scenario 4 to scenario 7, and the scenario 7 produces the most extreme overestimations, with the highest uncertainties.

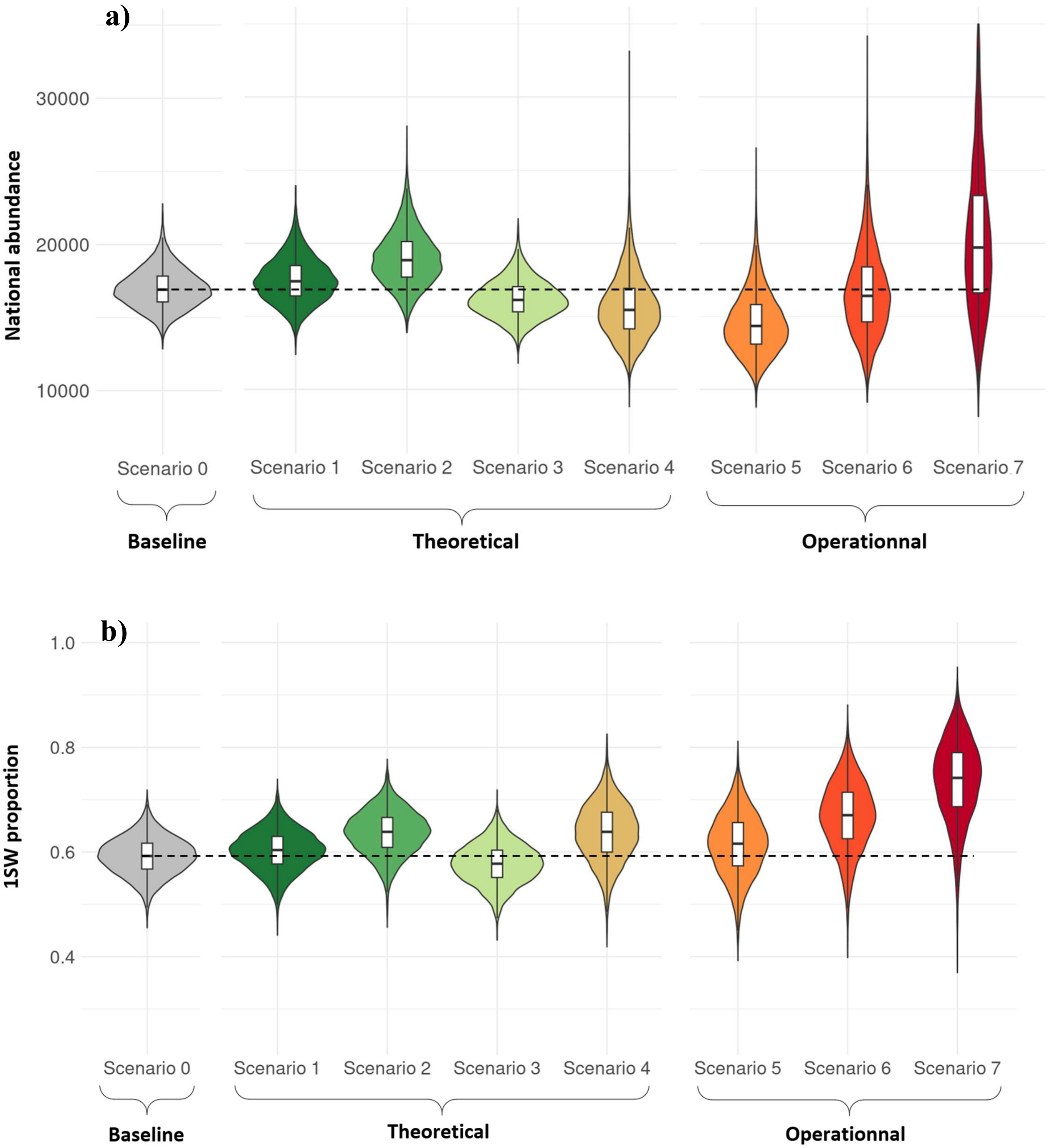


Figure 9: Estimates of the PastSatAb model under the seven data removal scenarios in comparison to the baseline scenario, for the national returning abundance estimations (a) and for 1SW proportions estimations (b) for the year 2020. The boxplots display the median, the 25th and 75th quartiles and the density of the estimates. The dotted line extends the median value of scenario 0 to facilitate interpretation of the figure.

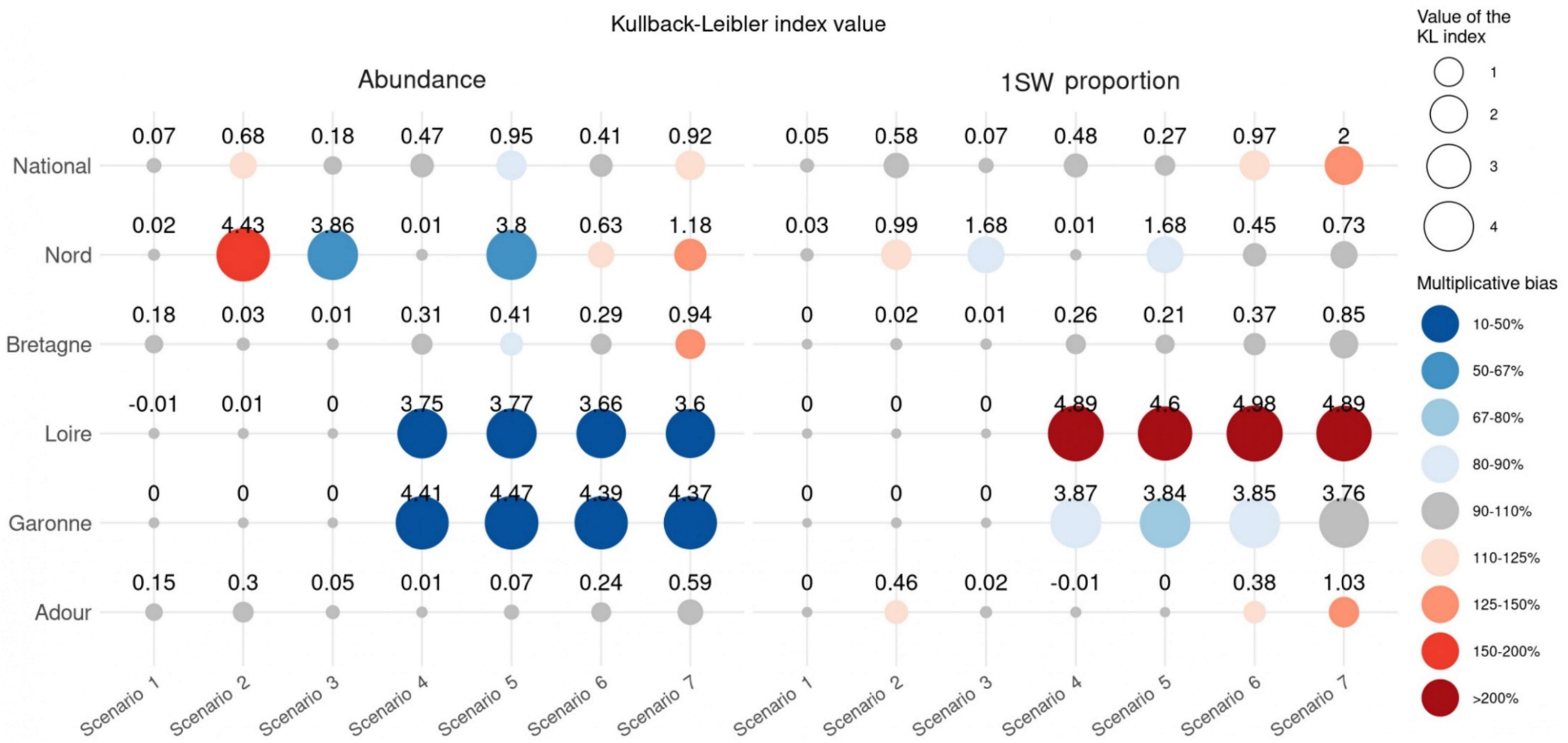


Figure 10: Kullback-Leibler Divergence indices and estimation bias range at a national and regional scale for the returning abundance and the 1SW proportion for the seven scenarios of data deletion in comparison with the scenario when all data are available (scenario 0). The size of dots informs of the value of the KL index (the value rounded to the hundredth is indicated above every dot) and the color of the dots informs of the strength of the multiplicative bias. The KLD is theoretically a non-negative index but can lead to small negative values due to numerical and rounding errors, especially when the difference between the distributions is very small.

IV – Discussion

IV.1. Approach and results overview

In this work, we have developed a framework for assessing the quality of national returning abundance and 1SW proportion estimations of different models and in the event of missing data. To use the PastSatAb model developed by Lebot & al. (in prep) for providing the WGNAS with sea-age national returning abundances estimations for the year under assessment, we focused on i) improving the modelling structure of the model to allow a more realistic estimation of the proportion of 1SW, ii) comparing the PastSatAb estimates with other modelling methods (including the run reconstruction model) and iii) measuring the value of the information of each data source integrated in the PastSatAb model within various scenarios of data removal. We considered four spatio-temporally heterogeneous data sources: catch data (in number of fish), direct counts, estimates of abundance in rivers and estimates of abundance in regions, produced and collected by a various set of management stakeholders. The model was run for each year of the 1970-2020 period for the five French considered regions. The modifications made to the model allow for a better modeling of the proportion of 1SW in the return abundance. The shrinkage effect observed in the previous version of the model has been reduced and we note a better fit of the estimates to the data. Direct counts and regional abundance estimations have the highest informative value out of the four data sources of the PastSatAb model. When the PastSatAb model only relies on the catch data, the national returning abundance estimation provided is close to the estimation provided when all data are accessible, but the 1SW proportion is over-estimated. Removing multiple data sources shows that the informative value is cumulative and that the fewer data sources there are, the greater the uncertainty in the estimates. The combined model presents an accurate estimation of both national returning abundance and the proportion of 1SW, but their distributions are highly dispersed in consequence of the significant variability added to the data used to estimate harvest rates and the pseudolikelihood applied to catch data. This study confirmed that the PastSatAb model is the most appropriate to provide WGNAS with 1SW and MSW national returning abundance estimations when only catch data are available for the year of 2020.

A better representation of sea-age distribution within abundance

The new version of the model allows assigning a different proportion of 1SW to each of the five regions considered in the study. Changing the structure of the model to associate a different proportion of 1SW per region has enabled to better represent the regional proportions of 1SW in the return abundances, and to provide a better fit to the data, particularly in the early years of the time series. However, the fit of the estimates to the data remains improvable (Appendix II) and could lead to bias the regional estimation of 1SW proportion. The results show that the modifications made have improved the model, but the lack of accurate data for the first part of the time series may prevent us from obtaining a totally satisfactory result.

The PastSatAb model, an appropriate method for estimating total return abundance in the event of missing data

The PastSatAb model offers a number of advantages, including the ability to extrapolate information from data-rich spatio-temporal units to data-poor units, while incorporating the uncertainty associated with this extrapolation (Prévost et al., 2003; Brun et al 2011). This enables national abundance estimations to be produced, and detailed at the scale of regions, rivers, and fisheries. For example, the overestimation of returning abundance at the national level under scenario 2 is driven by the significantly overestimated abundance in the Nord region (Figure 10). This level of precision has the advantage of providing a better understanding of the mechanisms underlying national dynamics, as well as enabling more accurate communication with managers. It is relevant to compare the estimates produced by PastSatAb when only catch data are available, here for the year 2020, as catch data are the most easily available. This model takes into account the uncertainty generated by the lack of informative data, as it produces estimates of national returning abundance and proportion of 1SW that have greater uncertainties than those from scenario 0. The overestimation of the proportion of 1SW highlights the limitations of the model's ability to estimate this parameter, despite the improvements.

In France, to date, the run reconstruction model has been used to produce the return abundance estimates provided to the Working Group for North Atlantic Salmon (ICES, 2024). This model has the advantage of using a simple formula to estimate national returning abundance and relies only on catch data that are always available at the time of the annual WGNAS meeting and harvest rates based on expertise, enabling the model to be used routinely. Although this model is different from the PastSatAb in terms of structure and inputs, it is still always worthwhile to compare the estimates they produce because the PastSatAb model was developed in the intention of replacing the run-reconstruction in the medium term.

The combined model uses the average of the PastSatAb harvest rate estimates from the past five years. It allows i) to have more realistic estimates than the expertise currently used in the run reconstruction model, and ii) to moderate potential maximum values or trends if only one year, let say 2019 (the last year with the full dataset) had been used. For example, in 2019, the exploitation rate was 0.20, while the average over the five years prior to 2020 was 0.17 (Lebot et al., in prep). The procedure for increasing the uncertainty around the estimates produced by the combined model allows for this variability to be integrated, as for the estimate of the 1SW proportion (Figure 7.b). This leads to a better estimation of the 1SW proportion by the combined model than by the PastSatAb model when only catch data are available (as its estimation bias is in the 90-110% range whereas the estimation bias of the PastSatAb model is in the 110-125% range).

Knowing that both PastSatAb under scenario 6 and the combined model did not overestimate return abundance estimates is encouraging for the application of these methods when only catch data are available. An over-optimistic vision of the state of a population suggests a risk of increasing TACs and quotas for the species. Given the advantages and disadvantages of both methods, it seems appropriate to consider using the PastSatAb model when only catch data are available. Although its implementation is time-consuming, its excellent estimation of the median for the returning abundance at national level and its ability to provide estimations at more detailed scales makes the PastSatAb model a wiser choice.

Some abundance data are highly structuring for national returning abundance and the proportion of 1SW estimates

Of the four PastSatAb theoretical scenarios, scenarios 2 and 4 produce the most biased estimates of national returning abundance and proportion of 1SW compared with the baseline scenario. These results demonstrate the highly informative nature of direct count data and regional abundance estimates for estimating these variables, and highlight their critical importance in the estimation of the two variables studied. Since the river abundance estimations data concern only two of the five regions and are mainly structural for the Nord region, they cannot be considered as highly informative at national level.

However, the results show that the scenario in which only catch data are available produces the best estimate of national return abundance. Scenarios 4 and 5 do not produce estimates any closer to the baseline scenario, even though they use more data than the scenario 6. Therefore, having more information does not necessarily mean a better estimate because here, the cumulative biases of the data present in scenarios 4 and 5 (important under-estimation due to the presence of count data) do not allow us to obtain a national return abundance estimate as good as or better than that obtained with scenario 6. On the other hand, for the proportion of 1SW, the addition of count data (scenario 5) improves the estimate of the proportion of 1SW compared with scenario 6.

Results at regional level are not necessarily reflected at national level. In the Loire-Allier and in Garonne-Dordogne regions, from the removal of regional model data (from scenario 4 to scenario 7), the national returning abundances are respectively under-estimated and the 1SW proportion is overestimated in the Loire-Allier region, and underestimated for Garonne-Dordogne. The values of their KL indices and multiplicative bias show only slight variations once the regional estimations are removed. These results show that these data provide the majority of information for both national returning abundance estimates and the proportion of 1SW in these regions. These underestimates of abundance are however not reflected at national level as the abundances of these two regions contribute to a very small proportion of the national abundance (see Appendix 3), which also explains the greater uncertainty surrounding the estimates.

The river abundance estimations are an important informative data for the Nord region but not so much for the Adour region. In Adour, the river under assessment of abundance estimation is the Nivelle, which represent a relatively small surface area, and consequently a low abundance in the Adour region, given the assumption of redistribution of regional abundance in rivers. In this region, the direct counts and the large catches data compensate for the absence of this data in scenario 3 (as there are no regional abundance estimates in these regions). Conversely in the Nord region, the Bresle river (the one under assessment) has a larger surface area. Catches and direct counts data are also less informative, leading to an under-estimation of the national returning abundance in this region under the scenario 3.

For the Brittany region, the PastSatAb model produces estimates of abundance and proportion of 1SW that are relatively similar to the baseline estimates in each data deletion scenario. We expected the removal of regional model estimations data to result in a significant bias in the abundance estimates for the region, as these data are informative. The results show that, in this region, the absence of regional estimations is compensated for by the large catch data and count abundances, which make the two variables estimates robust and insensitive to the removal of one or more data sources.

The regions provide different data due to the varying types of monitoring and the different monitoring infrastructures implemented by the management committees. Knowing the contributions of each data source is particularly crucial in our case study to assess the accuracy of PastSatAb estimates at the regional scale in the event of missing data. These results can guide managers in promoting the production and collection of the most informative data by the multiple stakeholders involved.

IV.2. Limitations of the approach

Two main limitations of this study's approach require us to be cautious in interpreting the results obtained. First, our study, carried out for the year 2020, shows that the model is capable of producing estimates of return abundance under different scenarios of missing data. This case study has enabled to develop a methodology to quantify the quality of estimates, as well as to determine the direction and intensity of potential estimation biases generated by the lack of information. In order to be able to characterize potential model trends more robustly, it is necessary to repeat this study in other years (removing only 2019 data, only 2018 data etc.) and to extend it over several years simultaneously (removing both 2020 and 2019 data, etc.). Repeating this study retrospectively from 2020 would enable us to identify trends in the PastSatAb model. This would enable us to identify whether the model's overestimation of the proportion of 1SW when only catch data are available is a trend in the PastSatAb model or a bias relative to the year 2020. The study would also allow us to evaluate the performance of the PastSatAb model under scenario 6 compared with that of the combined model, in order to validate the relevance of using PastSatAb regularly when only capture data are available.

Secondly, given the significant impact of fishing pressure on the river return phase, the assumption that all other sources of mortality are negligible is relevant because i) it simplifies the system under study, allowing for a better understanding of its dynamics, and ii) it allows taking advantage of existing and easily accessible catch data. However, we could question this assumption given the significant impacts of factors such as pollution, habitat loss, disruption of ecological continuity, climate change, and other threats to the survival of Atlantic salmon during the upriver migration phase (Dadswell et al., 2022), and add a mortality factor external to fishing. However, the information needed to quantify the mortality caused by these phenomena is not yet available. Our hypothesis is therefore the most cautious, because an overestimation of the external mortality factor could lead to an overestimation of the estimate of the return abundance, and an over-optimistic view of the state of a population could influence the assessment of its conservation status in the wrong direction.

VI.3. Perspectives of PastSatAb model utilization

Salmon stock assessment plays a key role in the management of its fishery exploitation. The PastSatAb project reflects the desire to produce accurate and unbiased estimates of stock size and its uncertainty. However, given the relative smaller size of the French salmon population, compared to Norway's, which holds 25% of the world's healthy populations (Hindar et al.

2011), improvements to the estimations of the French stock unit alone would likely not have a significant impact on estimates at the North Atlantic scale. Nevertheless, there is a general trend towards developing a more reliable method of estimating stock size to enable more effective management of fisheries. For example, the "England and Wales" jurisdiction employs an extended run-reconstruction model to estimate pre-fishery abundance (PFA). However, like the run-reconstruction approach for estimating PFA in France, this model has certain limitations. Gregory et al. (2023) developed a method for estimating the abundance of the England-Wales stock using rod-and-line fishery data. This model shares similarities with the PastSatAb model, as both are quantitative assessment methods developed using a Bayesian framework to account for uncertainty, and integrating data rich to data poor units. Both models aim to produce accurate estimates that can be used as key inputs on a national and international scale.

The estimations of returning abundances are also implemented in a lifecycle model developed by Olmos et al. (2019). This model allows for the estimation of various parameters, such as the abundance of post-smolts at sea, post-smolt survival rates, and proportions maturing as 1SW at for 13 North Atlantic stock units. They also derived the number of eggs spawned in each stock unit by year from the annual number of returning 1SW and MSW spawners. By generating returning abundance estimates through robust models, life history parameters can be more accurately estimated at the stock unit level. This represents a significant advancement in the evaluation and management of Atlantic salmon, enhancing our understanding of their biology and supporting more effective conservation strategies.

V – Conclusion

This work has developed a methodology to assess the possibility of using the PastSatAb model to provide national returning abundance estimates detailed by sea-age class to the Working Group on North Atlantic Salmon. The use of integrated models, which take advantage of all available data and expertise to provide more realistic estimates, is increasingly favoured. The application of the PastSatAb model, based exclusively on catch data, could thus meet WGNAS requirements by providing the estimates requested. Further analysis over a larger number of years could provide a better understanding of the PastSatAb model's ability to provide data when some of it is missing. However, the results are promising for use in the event of missing data and for these populations highly threatened by anthropogenic factors.

VI – Bibliography

Aas, Ø., Klemetsen, A., Einum, S., & Skurdal, J. (2011). *Atlantic Salmon Ecology*. John Wiley & Sons.

André, G., Guillerme, N., Charles, M., Clermont-Ledez, Y., Legrand, M., Chapon, P.-M., & Beaulaton, L. (2021). *Synthèse sur l'état des populations, des pressions et des modalités de gestion du Saumon Atlantique en France*.

Bagliniere, J.L., and Fontenelle, G. (1972). Les captures de saumon atlantique (Salmo salar L., 1766) dans le monde, essai d'interprétation.

Bagliniere, J. L., Denais, L., Rivot, E., Porcher, J. P., Prévost, E., Marchand, F., & Vauclin, V. (2004). *Length and age structure modifications of the Atlantic salmon (Salmo salar) populations of Brittany and Lower Normandy from 1972 to 2002. UMR INRA ENSA, Ecobiologie et Qualité des Hydrosystèmes Continentaux, Conseil supérieur de la pêche, UESP, INRA, Rennes.*

Baglinière, J.-L., Hamelet, V., Gueraud, F., Aymes, J.-C., Goulon, C., Richard, A., Josset, Q., & Marchand, F. (2020). *Guide pour l'interprétation des écailles et l'estimation de l'âge chez la truite commune (Salmo trutta) dans les populations françaises*. Office Français de la Biodiversité. https://hal.inrae.fr/hal-02880916

Bal, G., Montorio, L., Rivot, E., Prévost, E., Baglinière, J. -L., & Nevoux, M. (2017). Evidence for long-term change in length, mass and migration phenology of anadromous spawners in F rench A tlantic salmon *Salmo salar*. *Journal of Fish Biology*, *90*(6), 2375–2393. https://doi.org/10.1111/jfb.13314

Bardonnet, A., & Baglinière, J.-L. (2011). Freshwater habitat of Atlantic Salmon (Salmo salar). *Canadian Journal of Fisheries and Aquatic Sciences*, *57*, 497–506. https://doi.org/10.1139/cjfas-57-2-497

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., … Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, *528*(7582), 405–408. https://doi.org/10.1038/nature16062

Beaugrand, G., Edwards, M., Brander, K., Luczak, C., & Ibanez, F. (2008). Causes and projections of abrupt climate-driven ecosystem shifts in the North Atlantic. *Ecology Letters*, *11*(11), 1157–1168. https://doi.org/10.1111/j.1461-0248.2008.01218.x

Bradbury, I. R., Hamilton, L. C., Sheehan, T. F., Chaput, G., Robertson, M. J., Dempson, J. B., Reddin, D., Morris, V., King, T., & Bernatchez, L. (2016). Genetic mixed-stock analysis disentangles spatial and temporal variation in composition of the West Greenland Atlantic Salmon fishery. *ICES Journal of Marine Science*, *73*(9), 2311–2321. https://doi.org/10.1093/icesjms/fsw072

Brun, M. 2011. Un modèle de la dynamique de renouvellement de la population. In Aide à la décision pour la conservation des populations de saumon atlantique (Salmo salar L.), Université de Pau et des Pays de l'Adour. Pau. p. 252.

Chaput, G. (2012). Overview of the status of Atlantic salmon (Salmo salar) in the North Atlantic and trends in marine mortality. *ICES Journal of Marine Science*, *69*(9), 1538–1548. https://doi.org/10.1093/icesjms/fss013

Chaput, G., Legault, C. M., Reddin, D. G., Caron, F., & Amiro, P. G. (2005). Provision of catch advice taking account of non-stationarity in productivity of Atlantic salmon (Salmo salar L.) in the Northwest Atlantic. *ICES Journal of Marine Science*, *62*(1), 131–143. https://doi.org/10.1016/j.icesjms.2004.10.006

Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V., and Wikle, C.K. (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19(3): 553–570. doi:10.1890/07-0744.1.

Dadswell, M., Spares, A., Reader, J., McLean, M., McDermott, T., Samways, K., & Lilly, J. (2022). The Decline and Impending Collapse of the Atlantic Salmon (*Salmo salar*) Population in the North Atlantic Ocean: A Review of Possible Causes. *Reviews in Fisheries Science & Aquaculture*, *30*(2), 215–258. https://doi.org/10.1080/23308249.2021.1937044

Darwall, W. (2022). IUCN Red List of Threatened Species: Salmo salar. *IUCN Red List of Threatened Species*. https://www.iucnredlist.org/en

Dauphin, G., & Prévost, E. (2013). *Viability analysis of the natural population of Atlantic salmon (Salmo salar L.) in the Allier catchment Analyse de la viabilite d'une population naturelle de saumon atlantique (Salmo salar L.) dans le bassin de l'Allier*.

Gelman, A., Hwang, J., & Vehtari, A. (2013). *Understanding predictive information criteria for Bayesian models* (arXiv:1307.5928). arXiv. http://arxiv.org/abs/1307.5928

Gregory, S. D., Gillson, J. P., Whitlock, K., Barry, J., Gough, P., Hillman, R. J., Mee, D., Peirson, G., Shields, B. A., Talks, L., Toms, S., Walker, A. M., Wilson, B., & Davidson, I. C. (2023). Estimation of returning Atlantic salmon stock from rod exploitation rate for principal salmon rivers in England & Wales. *ICES Journal of Marine Science*, *80*(10), 2504–2519. https://doi.org/10.1093/icesjms/fsad161

Gueguen, J. C., & Prouzet, P. (1994). *Le saumon atlantique: Biologie et gestion de la ressource*. Editions Quae.

Hindar, K., Hutchings, J. A., Diserud, O. H., and Fiske, P. (2011). Stock, recruitment and exploitation. In Atlantic Salmon Ecology, pp. 299–332. Ed. by Ø. Aas, S. Einum, A. Klemetsen, and J. Skurdal. Wiley-Blackwell, Oxford, UK.

ICES. (2010). Report of the Study Group on Biological Characteristics as Predictors of Salmon Abundance. *ICES Document*, 158.

ICES. (2016). *Report of the Working Group on North Atlantic Salmon (WGNAS)* [Report]. ICES Expert Group reports (until 2018). https://doi.org/10.17895/ices.pub.19284557.v1

ICES. (2021). *Working Group on North Atlantic Salmon (WGNAS)* [Report]. ICES Scientific Reports. https://doi.org/10.17895/ices.pub.7923

ICES. (2023). *Working Group on North Atlantic Salmon (WGNAS)* [Report]. ICES Scientific Reports. https://doi.org/10.17895/ices.pub.22743713.v1

ICES. (2024). *Working Group on North Atlantic Salmon (WGNAS)* [Report]. ICES Scientific Reports. https://doi.org/10.17895/ices.pub.25730247.v1

Jensen, J. M. (1990). Atlantic salmon at Greenland. *Fisheries Research*, *10*(1), 29–52. https://doi.org/10.1016/0165-7836(90)90014-M

Jonsson, B., & Jonsson, N. (2004). Factors affecting marine production of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, *61*(12), 2369–2383. https://doi.org/10.1139/f04-215

Jonsson, B., Jonsson, N., & Albretsen, J. (2016). Environmental change influences the life history of salmon *Salmo salar* in the North Atlantic Ocean. *Journal of Fish Biology*, *88*(2), 618–637. https://doi.org/10.1111/jfb.12854

King, T. L., Kalinowski, S. T., Schill, W. B., Spidle, A. P., & Lubinski, B. A. (2001). Population structure of Atlantic salmon (Salmo salar L.): A range-wide perspective from microsatellite DNA variation. *Molecular Ecology*, *10*(4), 807–821. https://doi.org/10.1046/j.1365-294X.2001.01231.x

Lebot, C., al., & al. (in prep). *PastSatAb—Reconstruction des séries historiques d'abondance et de taux d'exploitation du saumon en France—Une approche Bayésienne pour intégrer les données et l'expertise disponible à l'échelle régionale.*

Lebot, C., Arago, M.-A., Beaulaton, L., Germis, G., Nevoux, M., Rivot, E., & Prévost, E. (2022). Taking full advantage of the diverse assemblage of data at hand to produce time series of abundance: A case study on Atlantic salmon populations of Brittany. *Canadian Journal of Fisheries and Aquatic Sciences*, *79*(4), 533–547. https://doi.org/10.1139/cjfas-2020-0368

Lissardy, M., & Morandeau, G. (2006). *Caractérisation et abondance des captures professionnelles des poissons migrateurs dans le bassin de l'Adour et les courants côtiers landais en 2006.*

Marty, A., & Bousquet, B. (2001). SITUATION DES POISSONS MIGRATEURS AMPHIHALINS SUR LE BASSIN DE L'ADOUR. *Bulletin Français de la Pêche et de la Pisciculture*, *357–360*, 345–356. https://doi.org/10.1051/kmae/2001054

Mills, K. E., Pershing, A. J., Sheehan, T. F., & Mountain, D. (2013). Climate and ecosystem linkages explain widespread declines in North American Atlantic salmon populations. *Global Change Biology*, *19*(10), 3046–3061. https://doi.org/10.1111/gcb.12298

Nevoux, M., Rivot, E., Gregory, S., Marchand, F., Besnard, A-L, Meslier, L., Jousseaume, T., Launey, S., King, A. Stevens. J., (2020). Changes in sex ratio and fecundity of salmonids. SAMARCH project, Deliverable D3.3.1, 18 pp.

Niemelä, E., Orell, P., Erkinaro, J., Dempson, J. B., BrØrs, S., Svenning, M. A., & Hassinen, E. (2006). Previously spawned Atlantic salmon ascend a large subarctic river earlier than their maiden counterparts. *Journal of Fish Biology*, *69*(4), 1151–1163. https://doi.org/10.1111/j.1095-8649.2006.01190.x

Olmos, M., Massiot-Granier, F., Prévost, E., Chaput, G., Bradbury, I. R., Nevoux, M., & Rivot, E. (2019). Evidence for spatial coherence in time trends of marine life history traits of Atlantic salmon in the North Atlantic. *Fish and Fisheries*, *20*(2), 322–342. https://doi.org/10.1111/faf.12345

Otero, J., Jensen, A. J., L'Abée-Lund, J. H., Stenseth, N. Chr., Storvik, G. O., & Vøllestad, L. A. (2012). Contemporary ocean warming and freshwater conditions are related to later sea age at maturity in Atlantic salmon spawning in Norwegian rivers. *Ecology and Evolution*, *2*(9), 2192–2203. https://doi.org/10.1002/ece3.337

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Working Papers*.

Porter, T. R., Healey, M. C., O'Connell, M. F., Baum, E. T., Bielak, A. T., and Côté, Y. (1986). Implications of varying the sea age at maturity of Atlantic salmon (Salmo salar) on yield to fisheries. In Salmonid Age at Maturity, pp. 110– 117. Ed. by D. J. Meerburg. Canadian Special Publication of Fisheries and Aquatic Sciences, 89. 118 pp.

Prévost, E., & Lebot, C. (2022). *RENOSAUM: Rénovation de la stratégie de gestion du saumon en Bretagne Rapport final* (p. 146 p.) [Final report]. OFB ; INRAE ; UPPA ; Association Bretagne Grands Migrateurs. https://hal.inrae.fr/hal-04055849

Prévost, E., Nevoux, M., Rivot, E., & Baglinière, J. -L. (2020). Le saumon atlantique. In *Les poissons d'eau douce de France* (2ème édition, pp. 494–497). Biotope Editions, Mèze ; Muséum national d'Histoire naturelle, Paris (collection Inventaires et biodiversité).

Prévost, E., Parent, E., Crozier, W., Davidson, I., Dumas, J., Gudbergsson, G., Hindar, K., McGinnity, P., MacLean, J., & Sættem, L. M. (2003). Setting biological reference points for Atlantic salmon stocks: Transfer of information from data-rich to sparse-data situations by Bayesian hierarchical modelling. *ICES Journal of Marine Science*, *60*(6), 1177–1193. https://doi.org/10.1016/j.icesjms.2003.08.001

Prévost, E., & Porcher, J. P. (1996). *Méthodologie d'élaboration de totaux autorisés de captures (TAC) pour le saumon atlantique (Salmo salar L.) dans le massif armoricain. Propositions et recommandations scientifiques*.

Quinn, T. P. (1993). A review of homing and straying of wild and hatchery-produced salmon. *Fisheries Research*, *18*(1), 29–44. https://doi.org/10.1016/0165-7836(93)90038-9

Rasmussen, R. O., & Valdemarsvej, K. (2001). *The development of fisheries in Greenland*.

Reddin, D. G., & Friedland, K. D. (1999). A history of identification to continent of origin of Atlantic salmon (*Salmo salar* L.) at west Greenland, 1969–1997. *Fisheries Research*, *43*(1), 221–235. https://doi.org/10.1016/S0165-7836(99)00074-0

Rivot, E., Perrier, C., Jouanin, C., Roussel, J.-M., Riera, P., Porcher, J.-P., & Baglinière, J.-L. (2009). Climate Change May Have Affected Growth and Life History in Atlantic Salmon Juveniles over the Past 30 Years. *American Fisheries Society Symposium*, *69*, 855–857.

Rivot, E., and Parent, E. (2012). Bayesian hierarchical models in statistical ecology. In Introduction to hierarchical Bayesian modeling for ecological data. CRC Press, Boca Raton. pp. 3–42.

Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Elsevier.

Servanty, S., & Prévost, É. (2016). *Mise à jour et standardisation des séries chronologiques d'abondance du saumon atlantique sur les cours d'eau de l'ORE DiaPFC et la Bresle*.

Stabell, O. B. (1984). Homing and Olfaction in Salmonids: A Critical Review with Special Reference to the Atlantic Salmon. *Biological Reviews*, *59*(3), 333–388. https://doi.org/10.1111/j.1469-185X.1984.tb00709.x

Utter, F., & Ryman, N. (1993). Genetic Markers and Mixed Stock Fisheries. *Fisheries*, *18*(8), 11–21. https://doi.org/10.1577/1548-8446(1993)018<0011:GMAMSF>2.0.CO;2

Valiente, A. G., Juanes, F., & Garcia-Vazquez, E. (2011). Increasing Regional Temperatures Associated with Delays in Atlantic Salmon Sea-Run Timing at the Southern Edge of the

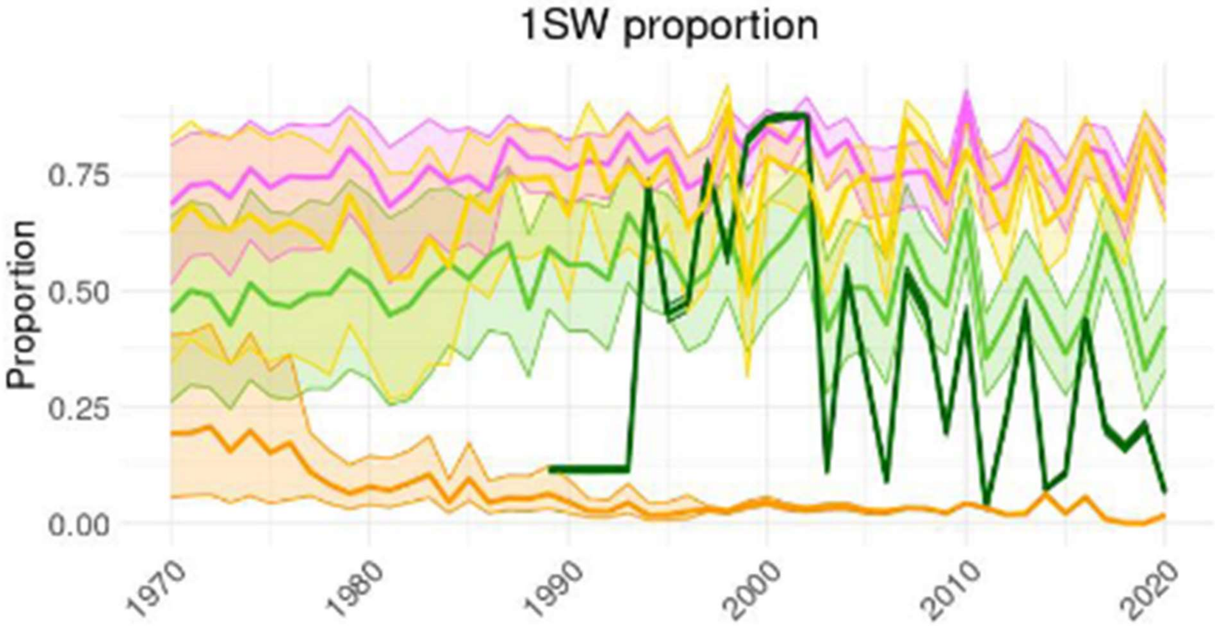
European Distribution. *Transactions of the American Fisheries Society*, *140*(2), 367–373. https://doi.org/10.1080/00028487.2011.557018

Webb, J., Verspoor, E., Aubin-Horth, N., Romakkaniemi, A., & Amiro, P. (2007). The Atlantic Salmon. In *The Atlantic Salmon* (pp. 17–56). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470995846.ch2

Windsor, M. l., & Hutchinson, P. (1994). International management of Atlantic salmon, Salmo salar L., by the North Atlantic Salmon Conservation Organization, 1984–1994. *Fisheries Management and Ecology*, *1*(1), 31–44. https://doi.org/10.1111/j.1365-2400.1970.tb00004.x

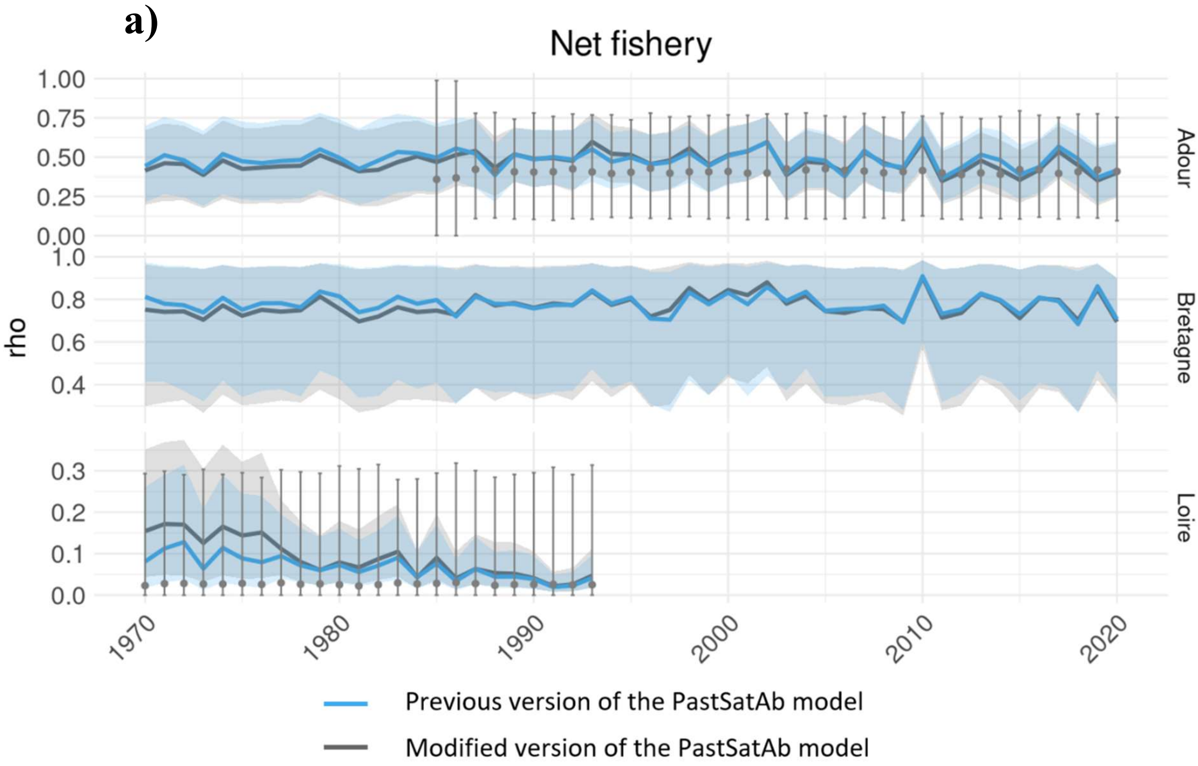
VII – Appendices

Appendix I: Estimates of the proportion of 1SW in each region obtained in the initial version of the model before the regionalisation of the parameter. The estimates are narrowed around the national grand mean for the 1970-1990 period. Years in the x-axis are return years.

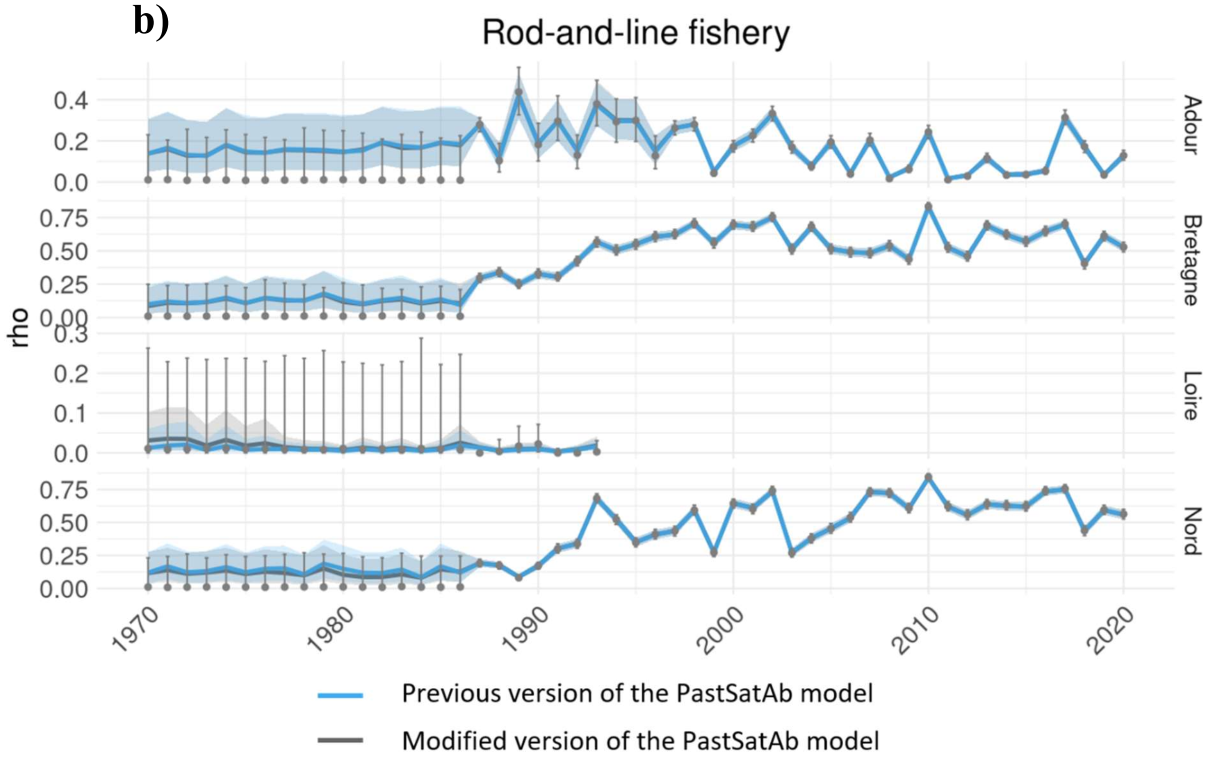


Appendix II: Estimates and data of the parameter of proportion of 1SW in catches (rho) for net fisheries (a) and rod and line fisheries (b) in the regions of interest for the previous and the modified versions of the PastSatAb model. The modified version allows for a better fit to data in the Loire-Allier region.

a)

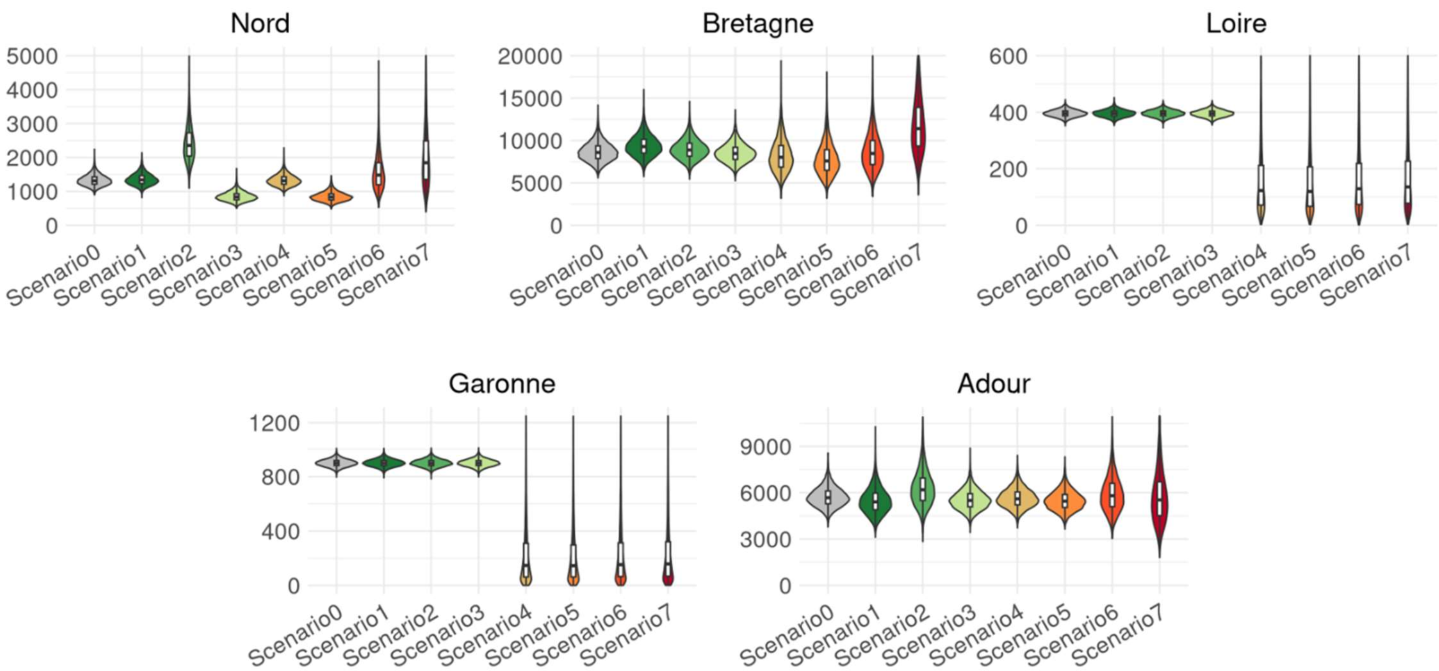


b)

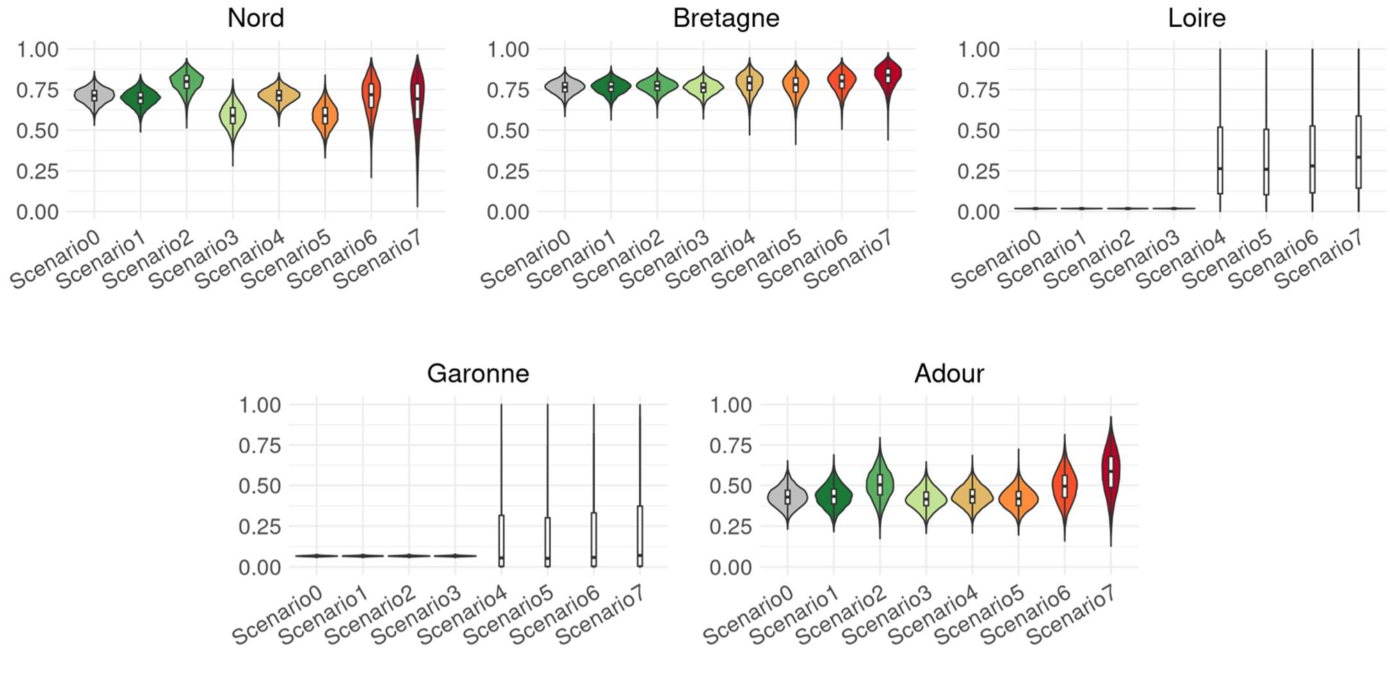


Appendix III: Estimates of the PastSatAb model under the seven data removal scenarios in comparison to the baseline scenario, for the regional returning abundance estimations (a) and for 1SW proportions estimations (b) for the year 2020. The plots display the median, the 25th and 75th quartiles and the density of the estimates.

a) **Regional returning abundance**



b) **1SW proportion**



L'INSTITUT agro Rennes Angers	Diplôme : Master Sciences de la Mer et du Littoral Spécialité : SML-Biologie Parcours : Sciences Halieutiques et Aquacoles (Ressources et Ecosystèmes Aquatiques) Enseignant référent : Pablo BROSSET, Enseignant chercheur, L'institut Agro Rennes

Auteur(s) : Klervi VERBRUGGHE Date de naissance* : 10/04/1999	Organisme d'accueil : Institut Agro Adresse : 65 rue de Saint Brieuc, CS 84215, 35042 Rennes Cedex Maître de stage : Clément LEBOT et Etienne RIVOT
Nb pages : 32 Annexe(s) : 3	
Année de soutenance : 2024	

Titre français : Développement d'une approche opérationnelle pour évaluer la qualité des estimations d'abondance en cas de données manquantes : une étude de cas sur les abondances de retour de saumons en France

Titre anglais : Developing an operational approach to assess the quality of abundance estimates in the event of missing data: a case study on salmon returning abundances in France

Résumé (1600 caractères maximum) :

Chaque année, les juridictions de l'Atlantique Nord doivent fournir au groupe de travail sur le saumon de l'Atlantique Nord (WGNAS) des estimations des abondances nationales de saumons par classe d'âge, permettant au WGNAS de donner un avis sur la gestion de l'espèce. Le modèle PastSatAb a été développé en partie pour fournir ces estimations, sans tenir compte de la variabilité régionale de la proportion de saumons d'un hiver de mer (1SW). Afin d'améliorer le modèle et son application lorsque des données sont manquantes, nous nous sommes concentrés sur 1) améliorer l'estimations de la proportion de 1SW, 2) comparer le modèle PastSatAb avec d'autres modèles d'estimation d'abondance, et 3) évaluer la valeur informative de chaque source de données dans des scénarios de suppression de données, pour l'année 2020. Nous avons analysé quatre sources de données hétérogènes : les capture, les comptages, les estimations d'abondance en rivière et régionales, dans cinq régions françaises pour la période 1970-2020. Les améliorations du modèle permettent une modélisation plus précise de la proportion de 1SW. Les comptages directs et les estimations d'abondance régionales se sont avérés être les données plus informatives. Cette étude confirme que le modèle PastSatAb est le meilleur pour les estimations d'abondance lorsque seules les données de capture sont disponibles. Une analyse sur plusieurs années permettrait d'améliorer ces résultats. Cependant, les résultats sont prometteurs pour une utilisation dans des scénarios de données manquantes et pour des populations menacées par des facteurs anthropiques.

Abstract (1600 caractères maximum) :

Each year, North Atlantic jurisdictions must provide the North Atlantic Salmon Working Group (WGNAS) with estimates of sea-age national returns, enabling WGNAS to advise on species management. The PastSatAb model was developed partly to provide these estimates, assuming the proportion of one-sea-winter (1SW) salmon is determined nationally, without considering regional variability. To improve the model's realism and applicability when data is missing, we focused on 1) enhancing the model's structure for better 1SW proportion estimates, 2) comparing PastSatAb with other estimations models, and 3) assessing the informative value of each data source under various data suppression scenarios for 2020, the most recent year studied. We analyzed four heterogeneous data sources: catch data, direct counts, river and regional abundance estimates, across five French regions for the 1970-2020 period. The model's improvements allow for more accurate 1SW proportion modeling. Direct counts and regional abundance estimates has been proved most informative. This study confirms PastSatAb as the best model for WGNAS sea-age estimations when only catch data is available. Further analysis over more years could deepen understanding of its effectiveness with missing data however, the results are promising for use in scenarios of missing data and for populations threatened by anthropogenic factors.

Mots-clés : Saumon atlantique, modèle hiérarchique, cadre d'étude bayésien, modèle à espace d'état, évaluation de modèle, scénarios de suppression de données, indice de Kullback-Leibler.

Key Words: Atlantic salmon, hierarchical model, bayesian framework, state-space model, model assessment,data deletion scenarios, Kullback-Leibler index.

* *Elément qui permet d'enregistrer les notices auteurs dans le catalogue des bibliothèques universitaires*