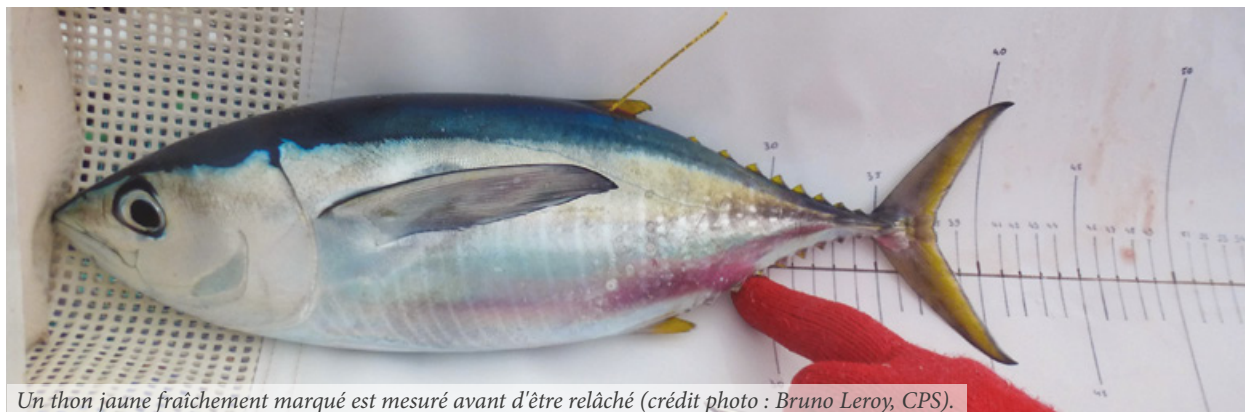


### Les trois V du traitement des données issues de la récupération des marques : validation, validation, validation



Un thon jaune fraîchement marqué est mesuré avant d'être relâché (crédit photo : Bruno Leroy, CPS).

La Communauté du Pacifique (CPS) marque et relâche des thons dans le Pacifique occidental et central depuis 1977. À ce jour, trois programmes de marquage se sont succédé : le Programme d'étude et d'évaluation du stock de bonites (1977-1981), le Programme régional de marquage des thonidés (1989-1992), et le Programme de marquage des thonidés dans le Pacifique (PTTP), lancé en 2006 et toujours en cours. Au total, plus de 700 000 thons ont été marqués et relâchés, et 100 000 marques ont été récupérées et déclarées à la CPS. Les informations tirées de ces campagnes de marquage sont une composante majeure des analyses sur lesquelles reposent la gestion des ressources thonières de la région et les modèles d'évaluation des stocks.

Plusieurs mois peuvent s'écouler entre la recapture du poisson marqué et la découverte de la marque, qui peut intervenir à des étapes très diverses de la filière thon : à bord des bateaux de pêche, pendant le triage ou le transfert du poisson dans les cales ; au cours du chargement à bord des navires de transbordement et de leur déchargement ; dans les chambres froides ; ou dans les conserveries. Plusieurs études ont déjà mis en évidence le lien entre la fiabilité des informations relatives à la récupération des marques et d'autres variables telles que le laps de temps écoulé entre la recapture du poisson marqué et la découverte de la marque. Il est donc primordial que les agents de la CPS effectuent un contrôle rigoureux de la qualité des données, notamment en procédant à la validation croisée des données de récupération des marques au regard des informations tirées d'autres séries de données. À l'issue de cet exercice, chaque marque se voit attribuer une estimation aussi juste que possible de la date et du lieu de récupération, ainsi qu'un indice de fiabilité correspondant au degré d'incertitude associé à ces champs de données essentiels (meilleure estimation de la date de recapture : 12 mars 2015  $\pm$  une semaine, par exemple). La validation croisée consiste à comparer les renseignements relatifs à la récupération des marques tirés de séries de données externes – données de suivi des navires générées par le système de surveillance des navires par satellite (VMS), données extraites des journaux de pêche, par exemple – afin d'estimer la fiabilité des informations reçues des personnes ayant trouvé les marques. Ainsi, en se

référant aux données VMS, on peut confirmer, d'après la position, l'allure et la direction du bateau de pêche censé avoir récupéré la marque, que ce dernier opérait vraisemblablement à proximité de la position signalée de la marque et à la date à laquelle elle a été récupérée. D'autres paramètres peuvent être pris en compte pour vérifier la qualité des données (taux de croissance des thons avant la recapture, distance maximale couverte, bathymétrie, notamment). Le processus de validation croisée représente un investissement considérable en temps et en ressources, en particulier lorsqu'un grand nombre de marques sont récupérées peu de temps après la campagne de marquage, et peut exiger de mobiliser jusqu'à quatre agents à temps partiel pour assurer l'enregistrement et la vérification des données.

Toutefois, une question majeure se pose : quel est le degré de précision des indices de fiabilité associés aux données sur la date et le lieu de la récupération des marques ? Et comment répondre à cette question en l'absence d'informations précises sur la date et le lieu de récupération des marques ? Les tests de faux marquage sont un moyen de répondre à ces questions : les observateurs posent discrètement des marques classiques sur des thons à bord de senneurs, et les données tirées de la récupération des fausses marques font l'objet du même contrôle qualité que les marques récupérées dans le cadre du PTTP. On peut ensuite comparer directement le lieu et la date de récupération de la fausse marque aux meilleures estimations de date et de lieu résultant du processus de validation croisée,

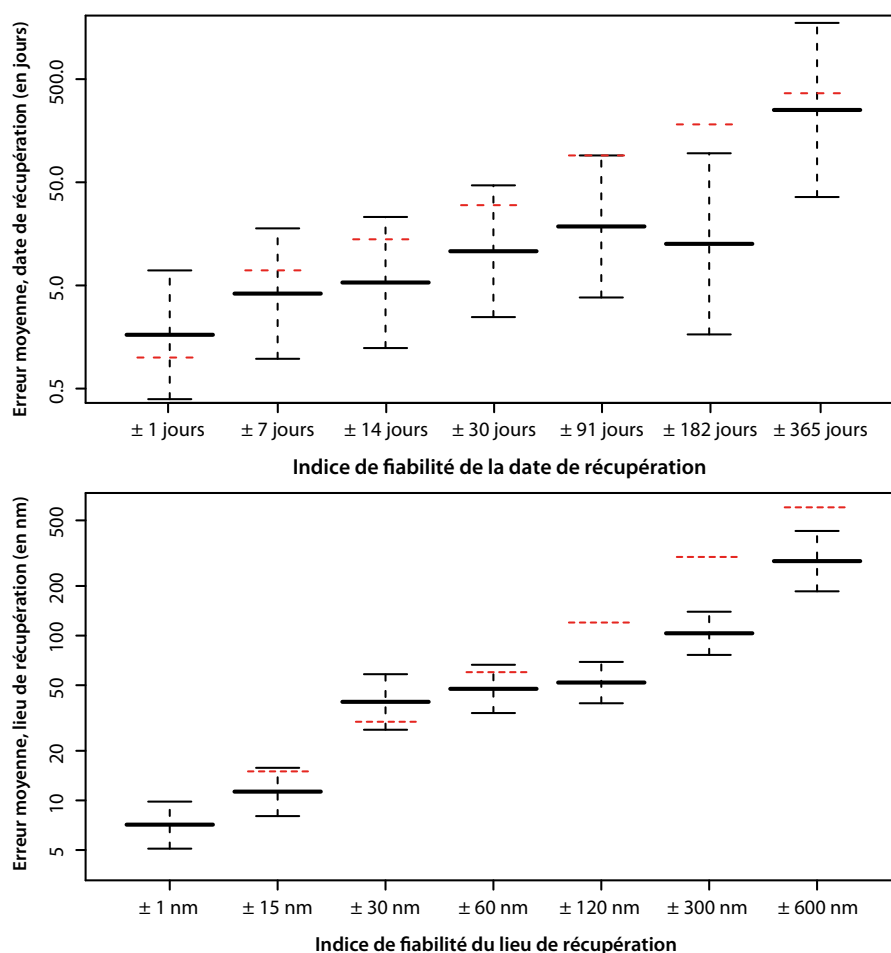


Figure 1. Schéma comparatif des erreurs moyennes estimées dans la date (en haut) et le lieu (en bas) de récupération des marques et des indices de fiabilité attribués. Les lignes noires en gras correspondent aux meilleures estimations et les lignes noires fines au degré d'incertitude qui leur est associé. Les lignes rouges en pointillés représentent le degré d'incertitude tel qu'il ressort du processus de validation croisée.

et comparer les erreurs relevées dans les meilleures estimations à l'indice de fiabilité attribué durant le processus de validation croisée.

Les erreurs relatives à la date et au lieu des recaptures augmentent proportionnellement au degré d'incertitude qui leur a été attribué (figure 1), ce qui démontre que la validation croisée est un moyen efficace de déterminer avec précision la fiabilité relative des informations sur la récupération des marques. Dans le cas des marques récupérées pour lesquelles les informations disponibles sont fortement entourées d'incertitude, les meilleures estimations de la date et du lieu de récupération sont généralement plus fiables que ne le laissent supposer les résultats du processus de validation. Cela étant, les erreurs de date et de lieu concordent généralement avec l'indice de fiabilité établi à l'issue du processus de validation croisée. Ce constat indique que la validation croisée peut aussi permettre de déterminer le degré d'incertitude des informations relatives à la date et au lieu de récupération des marques.

Grâce aux informations tirées des essais de faux marquage, on peut aussi analyser les facteurs influant sur le degré de précision de l'incertitude estimée associée à la date et au lieu de récupération des marques. À titre d'exemple, la majeure partie des vérifications et des comparaisons effectuées dans le cadre du processus de

validation croisée reposent sur l'identification précise du bateau de pêche ayant récupéré la marque. Les éléments dont on dispose montrent clairement, et il n'y a là rien d'étonnant, que les erreurs liées à la date et au lieu de récupération des marques sont plus importantes lorsque les personnes qui trouvent les marques n'identifient pas correctement les bateaux de pêche les ayant récupérées.

L'analyse résumée dans cet article ouvre des pistes encourageantes et montre que la validation croisée fournit des estimations précises de l'incertitude inhérente aux informations sur la récupération des marques. On peut donc intégrer à l'analyse des données de marquage le degré d'incertitude des informations relatives à la récupération des marques. De plus, les données de faux marquage offrent la possibilité d'analyser les facteurs qui ont le plus d'impact sur l'incertitude inhérente aux données de récupération des marques. Tous ces éléments peuvent contribuer utilement à la définition des modifications à apporter au programme de marquage pour renforcer le degré de confiance dans les séries de données issues des opérations de marquage.

## Pour plus d'information :

**Thomas Peatman**

Chargé de recherche halieutique principal  
(modélisation statistique), CPS  
[thomaspc@spc.int](mailto:thomaspc@spc.int)