



SYNTHÈSE DU PROCESSUS D'HARMONISATION DES DONNÉES DE MARQUAGE SUR LES TORTUES OLIVÂTRES EN GUYANE



Prestation à l'attention de l'association *Trans Océan Tortues Marines* réalisée
du 1^{er} mai au 30 juin 2026



AUSTRALMAT
CONSULTING
DATA • ENVIRONNEMENT • ANALYSES

Matthieu PUJOL – BP 110452, 98709 Mahina – N° TAHITI G50777
Tél. : +33 (0) 6 43 70 46 62 | Mél. : tieu.pujol@gmail.com

Remerciements

Cette prestation s'inscrit dans le cadre de la mise en œuvre des actions nationales fédératrices en faveur de l'étude et de la conservation des tortues marines portées par l'association **Trans Océans Tortues Marines** dans le monde entier. Elle a été rendue possible par l'obtention d'un financement spécifiquement dédié à ce travail pour soutenir la mise à jour et la valorisation des données de marquage disponibles en Guyane. À ce titre, je souhaite adresser toute ma reconnaissance pour la confiance accordée à AUSTRALMAT CONSULTING dans la réalisation de cette mission. Tout le socle de cette démarche repose sur la qualité et la régularité des campagnes de terrain coordonnées sur la plage de Rémire-Montjoly par l'association **Kwata** pour collecter les données et le soutien actif structurel et financier de tout le reste des acteurs et bailleurs faisant partie intégrante du Réseau Tortues Marines Guyane. Je remercie l'ensemble de ces organismes contributeurs pour leur travail quotidien.

Déclaration d'éthique

Je certifie que ce travail porte exclusivement sur des données historiques collectées antérieurement au cours de programmes scientifiques autorisés et réglementés. Aucune intervention sur les animaux ni collecte complémentaire de données n'a été réalisée dans cette analyse. Les traitements effectués ont été conduits dans le respect des principes de rigueur scientifique, de traçabilité et de valorisation à des fins de recherche et de conservation.

Disponibilité des données

L'ensemble des données présentées dans cette synthèse constitue le fruit du travail de terrain de l'association **Kwata** et un des livrables relatifs à la prestation initiée par l'association **Trans Océans Tortues Marines**. Toute mise à disposition et utilisation *in fine* se fait sur demande officielle avec l'accord des structures propriétaires des données. Selon les règles de confidentialité, aucune transmission ne sera effectuée par AUSTRALMAT CONSULTING.

Citation

Pujol, M. (2026). Synthèse du processus d'harmonisation des données de marquage sur les tortues marines en Guyane.



La définition de l'état des populations de tortues marines ne repose pas sur un dénombrement des individus d'une population, mais sur une analyse complexe des flux démographiques. Elle intègre des paramètres clés de productivité, comme notamment l'abondance saisonnière de femelles nidifiantes, le nombre de pontes réalisées et la fidélité au site de ponte. Par conséquent, comprendre les tendances démographiques nécessite des données issues de programmes de suivi à long terme, fondés sur le procédé de capture-marquage-recapture (CMR) et la prise en compte des menaces à pluri-échelles. En Guyane, territoire reconnu pour héberger des sites de nidification majeurs pour la tortue verte (*Chelonia mydas*), la tortue olivâtre (*Lepidochelys olivacea*) et la tortue luth (*Dermochelys coriacea*), un récent constat témoigne d'une baisse généralisée de la fréquentation des plages par les femelles adultes. Cette diminution va de pair avec l'accroissement des multiples pressions à terre et en mer, notamment celles liées à la pêche illicite, non déclarée et non réglementée (INN) qui a doublé au cours de la dernière décennie. Face à ces enjeux, la constitution d'un Plan National d'Actions en faveur des Tortues Marines en Guyane (PNATMG) a progressivement permis de structurer les efforts de conservation, en réunissant les différents acteurs institutionnels, scientifiques et associatifs impliqués dans la mise en œuvre de mesures de gestion adaptées. Dans cette dynamique, l'association **Kwata** a conduit, entre 2004 et 2018, des campagnes de marquage de tortues olivâtres et de tortues luths sur la plage de Rémire-Montjoly, localisée dans l'Est guyanais. Ces données, acquises sur plus de seize années d'efforts de terrain, représentent aujourd'hui une ressource scientifique de premier plan. Afin de les exploiter et de les valoriser, l'association **Trans Océans Tortues Marines** a mandaté AUSTRALMAT CONSULTING pour transformer cette matrice brute en une base de données harmonisées et standardisées. Le présent document constitue ainsi une synthèse descriptive et récapitulative des différentes étapes de traitements statistiques menant à identifier chaque femelle marquée et capitaliser les valeurs aberrantes répertoriées. Ainsi, il n'a pas pour ambition d'analyser scientifiquement les données, mais de produire un travail de nettoyage en amont pour les rendre immédiatement implémentables dans des modèles d'analyses démographiques qui seront développées par le CNRS dans le projet INVOCO (Influence des Variables Océanographiques et des activités humaines sur le Comportement, la distribution et la démographie des TORTUES marines dans l'océan Atlantique), grâce au soutien financier de la DGTM Guyane et de l'Office français de la biodiversité (OFB).





TABLE DES MATIÈRES

1. INTRODUCTION	5
1.1. Contexte de l'étude	5
1.2. Objectifs de la prestation	6
2. MATÉRIEL & MÉTHODES	7
2.1. Caractéristiques de la base de données brutes	7
2.1.1. Structure initiale	7
2.1.2. Description des variables d'intérêt.	8
2.2. Logiciels et packages utilisés dans le traitement des données	8
2.3. Description des étapes méthodologiques	9
2.4. Pré-traitement des données	10
2.4.1. Importation et structuration de la base brute	10
2.4.2. Harmonisation des identifiants	10
2.4.2.1. Les PIT	10
2.4.2.2. Les bagues	11
3. RÉSULTATS	12
3.1. Création de la base de données CMR	12
3.1.1. Principe général	12
3.1.2. Consolidation des identifiants individuels	12
3.1.3. Reconstruction des historiques de capture	12
3.1.4. Qualification des événements de capture	13
3.1.5. Structure finale	13
3.2. Création de la base de données des femelles individuelles	14
3.2.1. Principe général	14
3.2.2. Regroupement des identifiants individuels	14
3.2.3. Descripteurs individuels	15
3.2.4. Structure finale	15
3.3. Création de la base de données des valeurs aberrantes	16
3.3.1. Principe général	16
3.3.2. Détection des doublons techniques	17
3.3.3. Réobservations à pas de temps très court	17
3.3.4. Conflit d'identification	18
3.3.5. Événements biologiquement atypiques	18
3.3.6. Structure finale	18
3.4. Quelques statistiques descriptives	19
3.5. Recommandations pour les futures campagnes de terrain	23
4. CONCLUSION	24



1. INTRODUCTION

1.1. Contexte de l'étude

Les données de capture-marquage-recapture (CMR) constituent aujourd'hui l'un des outils méthodologiques de référence pour l'étude de la dynamique des populations de tortues marines. Appliquées aux femelles nidifiantes, elles permettent notamment de reconstruire des historiques individuels, d'identifier les trajectoires de réobservation, d'estimer les fréquences de retour sur les sites de ponte et, à terme,

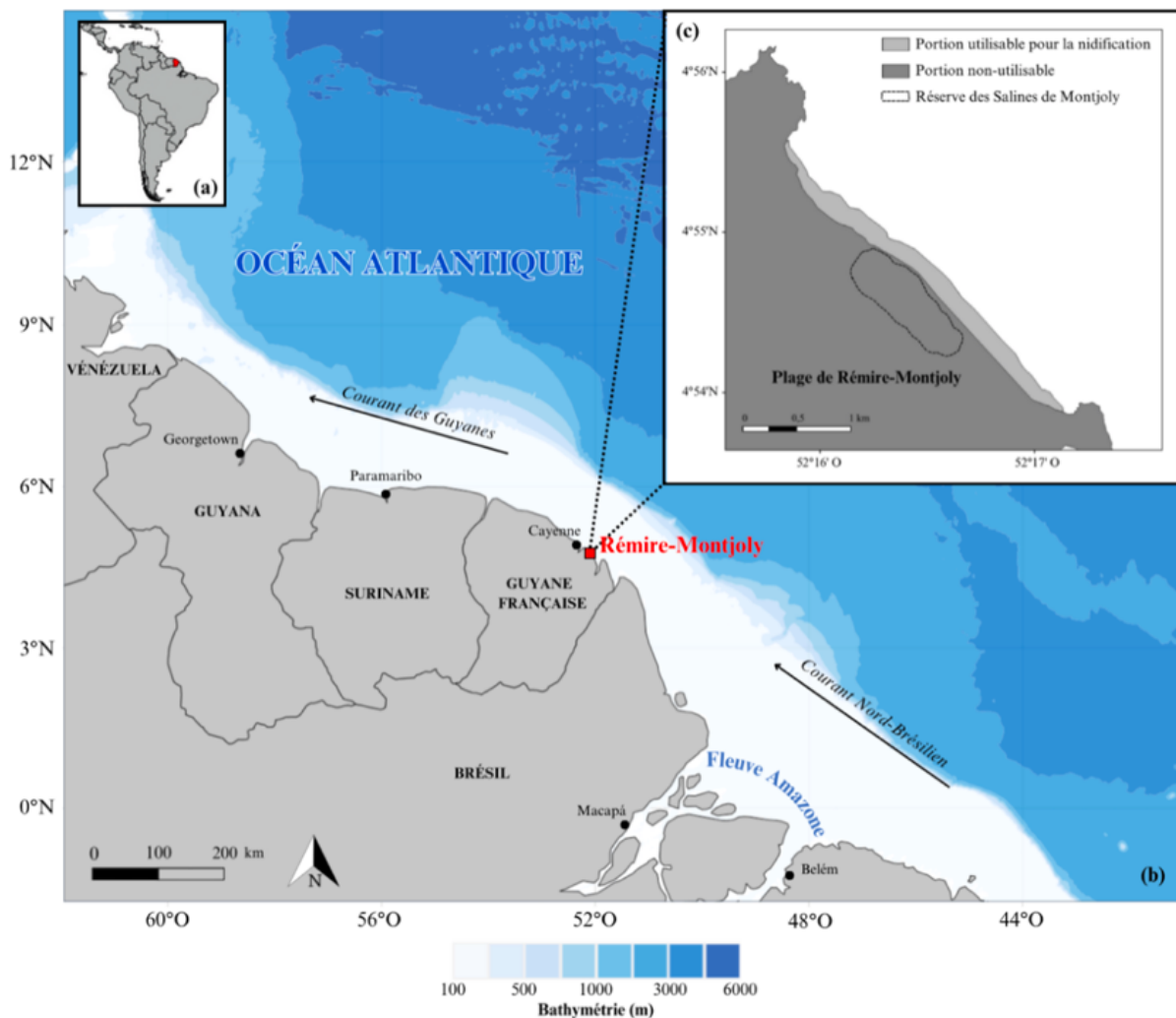


FIGURE 1 | Carte du site d'étude (a) en Amérique du Sud, mettant en avant (b) la localisation de la plage de Rémire-Montjoly en Guyane et (c) un aperçu des portions de la plage favorable à la nidification des tortues marines.

d'alimenter des modèles démographiques visant à mieux comprendre la structure et l'évolution des populations reproductrices.

En Guyane, la population de femelles tortues olivâtres (*Lepidochelys olivacea*) venant nidifier sur les plages de l'Est présente une dynamique des pontes encore méconnue et difficile à anticiper au regard du nombre erratique de femelles réobservées chaque année et la capacité de cette espèce à adopter deux stratégies de nidification différentes : pondre indépendamment les unes des autres « nidification solitaire », ou pondre de manière simultanée « nidification de masse (aussi appelée 'arribadas') ».

Afin de mieux connaître l'écologie comportementale de cette espèce en période de reproduction et obtenir un ensemble historique de données CMR, des campagnes successives de marquage ont été réalisées entre 2004 et 2018 par l'association Kwata sur la plage de Rémire-Montjoly (4,53°N – 52,16°O), localisée sur la presqu'île de Cayenne (FIGURE 1). Ces données reposent sur différents systèmes d'identification individuelle, incluant principalement des marquages électroniques sous-cutanés de type PIT (*Passive Integrated Transponder*), ainsi que des bagues externes utilisées lors des opérations de terrain.

En outre, la constitution progressive de cette base de données, l'utilisation de plusieurs méthodes de marquage et la récupération de différents formats d'enregistrement ont conduit à la production d'un jeu de données hétérogène nécessitant une étape préalable de structuration avant toute exploitation statistique. La matrice disponible correspond ainsi à un tableau historique brut, capitalisant l'ensemble des observations de femelles acquises au cours des suivis réalisés sur les seize années de marquage.

L'exploitation directe de cette base présentait plusieurs limites, parmi lesquelles figuraient notamment l'absence d'harmonisation des identifiants, la coexistence de plusieurs systèmes de marquage pour un même individu, la présence de doublons ou d'incohérences, ainsi que l'impossibilité de reconstruire directement les historiques individuels nécessaires aux futures analyses CMR. Dans ce contexte, une phase de pré-traitement et d'homogénéisation des données apparaissait indispensable afin de transformer la base historique initiale en une structure de données traçable et exploitable à des fins d'analyses démographiques.

1.2. Objectifs de la prestation

Le manque de connaissance actuel sur la dynamique des événements de pontes des tortues olivâtres en Guyane conjuguée à l'exacerbation des menaces en mer et à terre au cours des périodes de reproduction rend indispensable l'analyse des données historiques collectées depuis le début des suivis des femelles. C'est là tout l'enjeu de cette prestation technique : transformer la base historique brute en un ensemble de bases de données structurées permettant l'exploitation future des informations de capture, marquage et recapture chez les tortues olivâtres suivies en Guyane.



À l'issue des traitements réalisés, trois bases de données complémentaires devaient être produites :

- (1) une base de données nettoyée et standardisée prête pour les futures analyses CMR : `BDD_LO_clean_events` ;
- (2) une base de données synthétique construite à l'échelle individuelle permettant la reconstruction des historiques des femelles observées : `BDD_LO_individuals` ;
- (3) une base de données recensant les valeurs aberrantes, incohérences et anomalies identifiées au cours du traitement : `BDD_LO_anomalies`.

Le présent rapport décrit les différentes étapes méthodologiques mises en œuvre pour passer de la base historique initiale à ces trois livrables, depuis les opérations de prétraitement et de qualification des observations jusqu'à la construction des jeux de données finaux.

2. MATÉRIEL & MÉTHODES

2.1. Caractéristiques de la base de données brutes

2.1.1. Structure initiale

Le jeu de données utilisé `BDD_LO_brute` constitue l'ensemble des observations acquises lors des campagnes de marquage réalisées entre 2004 et 2018 et totalise **21 282 évènements**. Chaque ligne correspond à une observation individuelle effectuée lors d'une opération de marquage d'une femelle observée sur la plage de Rémire-Montjoly.

La structure initiale du tableau comprenait 7 colonnes :

- `PIT.Gauche` et `PIT.Droit` ; deux variables associées à des PIT latéraux
- `PIT.TD` et `PIT.ED` ; deux variables associées à des PIT anatomiques
- `Bague` ; une variable associée à une bague externe
- `Date` ; une variable temporelle
- `Taxa` ; une variable taxonomique

L'exploration préliminaire a montré que les variables `PIT.Gauche` et `PIT.Droit` ne contenaient aucune information, ces colonnes ont ainsi été exclues des traitements ultérieurs et du protocole d'homogénéisation.



2.1.2. Description des variables d'intérêts

Les variables `PIT.ED` et `PIT.TD` correspondent aux numéros des PIT utilisés au cours de l'historique du suivi. Lors des premières années de terrain, les PIT étaient majoritairement implantés au niveau de l'épaule droite des tortues (`PIT.ED`). Cependant, l'expérience acquise au cours des campagnes successives a montré que ces PIT présentaient une tendance à migrer progressivement sous la carapace, réduisant leur détectabilité lors des opérations de recapture. Afin de réduire ce biais, le protocole a évolué vers l'implantation des transpondeurs dans le triceps droit (`PIT.TD`), devenu progressivement le système de référence utilisé pour l'identification individuelle des femelles.

La variable `Bague` correspond aux bagues métalliques externes utilisées lors des premières opérations de marquage. Bien que ces marqueurs puissent être perdus au cours du temps, ils constituent une source d'information essentielle pour les premières années du suivi et permettent, dans de nombreux cas, d'établir des correspondances entre différents systèmes d'identification.

La variable `Date` renseigne la date d'observation de chaque individu et constitue l'information temporelle de référence pour la reconstruction des historiques individuels. Enfin, la variable `Taxa` permet de confirmer l'appartenance taxonomique des observations enregistrées, l'ensemble de la base étant ici associé à la tortue olivâtre (*Lepidochelys olivacea*).

2.2. Logiciels et packages utilisés dans le traitement des données

L'ensemble des traitements statistiques a été réalisé sous le logiciel R version 4.5.2 (R Core Team) au sein de l'environnement RStudio. Les différentes étapes du protocole ont été implémentées sous forme de scripts reproductibles permettant l'importation, le nettoyage, la restructuration et l'export des données.

L'importation des données et la gestion des fichiers Excel ont été réalisées à l'aide des packages *readxl* (**Wickham & Bryan**) et *writexl* (**Ooms**). Les opérations de manipulation, de filtrage, d'agrégation et de création des nouvelles variables ont principalement reposé sur les packages *dplyr* (**Wickham et al.**) et *tidyr* (**Wickham et al.**), tandis que *janitor* (**Firke**) a été utilisé pour l'harmonisation de la structure des tables et des noms de variables.

La standardisation des chaînes de caractères ont été effectués à l'aide du package *stringr* (**Wickam**), permettant notamment l'uniformisation des formats de PIT et de bagues. Les variables temporelles dérivées de la date d'observation ont quant à elles été générées à l'aide du package *lubridate* (**Grolemund & Wickham**), utilisé pour la gestion des informations calendaires et le calcul des intervalles temporels.



Enfin, la consolidation des identifiants multiples et la reconstruction des historiques individuels ont reposé sur des traitements séquentiels réalisés sous *dplyr*, complétés ponctuellement par le package *igraph* (Csárdi et al.) pour l'établissement des correspondances entre différents marqueurs associés à un même individu.

2.3. Description des étapes méthodologiques

La FIGURE 2 présente le pipeline de traitement développé pour transformer la base historique brute de capture-marquage-recapture (CMR) des tortues olivâtres en trois jeux de données consolidés destinés aux futures analyses.

Le protocole débute par une phase de pré-traitement visant à harmoniser les identifiants et qualifier les observations disponibles. Les différents systèmes de marquage sont ensuite comparés afin d'attribuer un identifiant unique à chaque femelle et de reconstituer les correspondances entre les dispositifs associés à un même individu.

À partir de ces identifiants consolidés, les historiques individuels de capture et de recapture sont reconstruits chronologiquement et complétés par la création de variables dérivées. Ce processus aboutit à la production de trois bases complémentaires : une base d'événements prête pour les analyses CMR (*BDD_LO_clean_events*), une base synthétique construite à l'échelle individuelle (*BDD_LO_individuals*) et une base documentant l'ensemble des anomalies détectées au cours du traitement (*BDD_LO_anomalies*).

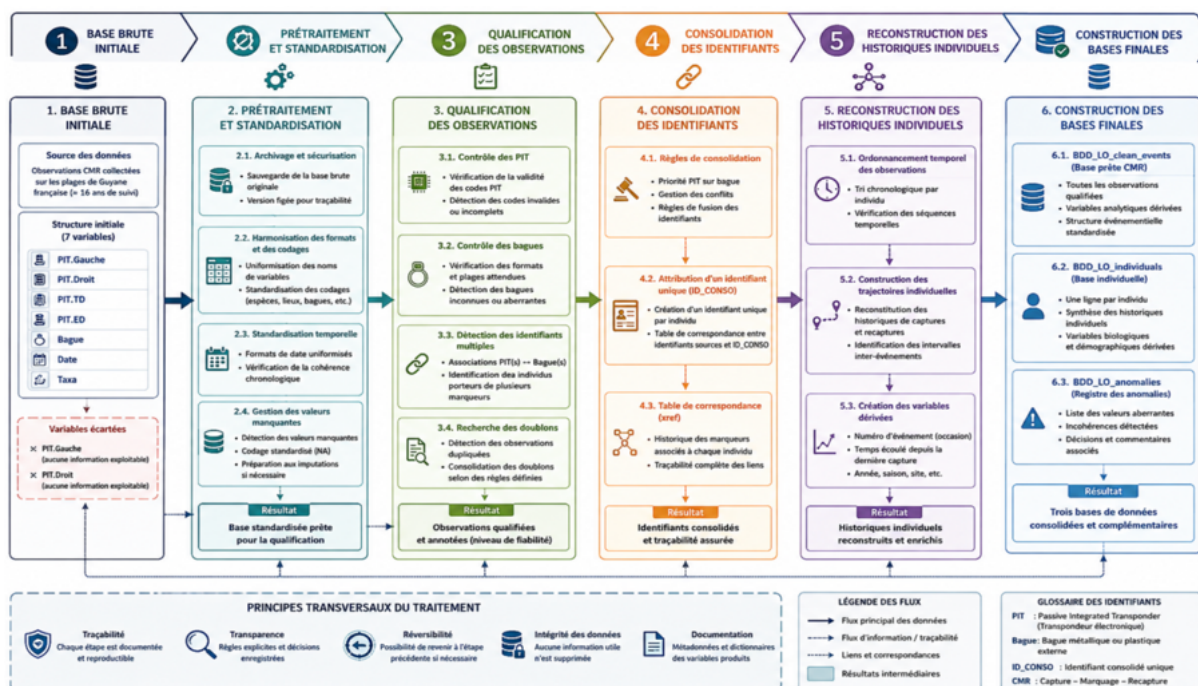


FIGURE 2 | Schéma conceptuel du pipeline de traitement de la base de données brute à la production des trois bases de données consolidées.

2.4. Pré-traitement des données

2.4.1. Importation et structuration de la base brute

La base historique a été importée dans R sans modification préalable afin de conserver une version de référence correspondant strictement aux données brutes fournies en entrée du protocole : `bdd_raw`.

Les noms de variables ont été harmonisés avec `clean_names()` afin d'obtenir une nomenclature standardisée, en supprimant les majuscules et remplaçant les points. Cette opération n'affecte pas le contenu des données, mais uniquement la dénomination des champs de la base.

NOM INITIAL	NOM OBTENU
PIT.TD	pit_td
PIT.ED	pit_ed
PIT.GAUCHE	pit_gauche
PIT.DROIT	pit_droit

Enfin, un identifiant technique unique, `record_id`, a été attribué à chaque observation de femelle sur la plage. Cet identifiant permet d'assurer la traçabilité complète des opérations de traitement et de relier chaque enregistrement aux données brutes d'origine tout au long du protocole de consolidation.

La variable `Date` a été convertie au format 'Date' reconnu par R afin de permettre son exploitation dans les traitements chronologiques ultérieurs.

2.4.2. Harmonisation des identifiants

2.4.2.1. Les PIT

Cette étape vise à standardiser l'ensemble des variables utilisées pour l'identification des individus avant toute tentative de reconstruction des historiques de capture. En effet, il est fréquent qu'un même identifiant soit enregistré sous plusieurs formats légèrement différents (*e.g.* 'fg12345', 'Fg12345', 'FG12345', ' FG12345').

Les identifiants ont été convertis au format texte, les espaces parasites ont été supprimés et l'ensemble des caractères a été uniformisé en majuscules. Les chaînes vides ont ensuite été converties en valeurs manquantes (NA) afin de faciliter les contrôles ultérieurs. Cette étape visait à éviter qu'un même individu soit artificiellement représenté par plusieurs identifiants distincts en raison de différences de saisie ou de formatage.



2.4.2.2. Les bagues

La variable Bague a fait l'objet d'un traitement spécifique en complément de l'harmonisation générale des identifiants. L'exploration de la base de données a en effet révélé la présence d'informations concaténées avec des formats hétérogènes. Ces cas correspondaient principalement à des femelles portant simultanément deux bagues externes, généralement une bague sur chaque nageoire antérieure.

Dans la base brute, ces informations étaient parfois enregistrées sous la forme d'une chaîne de caractères uniques combinant les deux numéros de bagues ainsi que des indications de latéralité.

- **Exemple** : l'identifiant 'FG5058 PdroiteG45558' dans la base brute correspond en réalité à deux identifiants distincts associés à une même femelle portant deux bagues 'FG5058' et 'G45558'.

De manière similaire, certaines observations comportaient des indications de positionnement des bagues ou des commentaires ajoutés directement dans le champ d'identification.

- **Exemple 1** : l'identifiant 'gaucheFG1979 droiteFG5429(CNRS)' correspond après standardisation à une même femelle portant deux bagues 'FG1979' et 'FG5429'.
- **Exemple 2** : l'identifiant 'BR32699 (SALVADOR)' correspond à une seule femelle portant la bague identifiée par 'BR32699'.

Les mentions textuelles relatives à la position des bagues ainsi que les commentaires placés entre parenthèses ont été retirés uniquement pour l'extraction des identifiants. Ces informations ne participent pas à l'identification individuelle des femelles et ne sont donc pas nécessaires aux traitements ultérieurs.

Afin de conserver l'ensemble de l'information utile tout en standardisant la structure de la base, deux nouvelles variables ont été créées :

- Bague_1, correspondant au premier identifiant de bague détecté dans la cellule ;
- Bague_2, correspondant à un éventuel second identifiant de bague associé à la même observation.

La colonne Bague d'origine a été conservée dans la base afin de maintenir la traçabilité avec les données brutes. Les variables Bague_1 et Bague_2 constituent quant à elles les versions standardisées utilisées dans les traitements ultérieurs de consolidation des identifiants.

VALEUR INITIALE BAGUE	BAGUE_1	BAGUE_2
FG5058 PDROITEG45558	FG5058	G45558
FG1350 PDROITEG45166	FG1350	G45166
GAUCHEFG1979 DROITEFG5429(CNRS)	FG1979	FG5429
BR32699(SALVADOR)	BR32699	NA



3. RÉSULTATS

3.1. Création de la base de données CMR

3.1.1. Principe général

À l'issue des opérations de prétraitement, les données ont été restructurées afin de constituer une base de données intitulée `BDD_LO_clean_events`. Cette base constitue le principal livrable de la prestation et rassemble l'ensemble des événements de capture, de marquage et de recapture après harmonisation des identifiants.

Contrairement à la base brute, dans laquelle les observations étaient uniquement enregistrées sous forme d'événements indépendants, cette nouvelle structure associe chaque observation à une femelle unique identifiée de manière cohérente à partir de l'ensemble des marqueurs disponibles. Chaque ligne de la base correspond ainsi à une observation individuelle, tandis que l'ensemble des observations appartenant à une même femelle sont reliées par un identifiant commun (`individual_id`).

Cette restructuration constitue le socle des futures analyses de capture-marquage-recapture en permettant de reconstituer les historiques individuels et de suivre chronologiquement les observations de chaque femelle.

3.1.2. Consolidation des identifiants individuels

La première étape de construction de la base a consisté à regrouper les observations correspondant à une même femelle. Cette consolidation repose sur l'ensemble des systèmes d'identification disponibles, comprenant les transpondeurs électroniques (`PIT.TD` et `PIT.ED`) ainsi que les bagues externes (`Bague_1` et/ou `Bague_2`).

Les observations partageant au moins un identifiant commun ont été rattachées au même individu. Cette procédure permet notamment de conserver la continuité des historiques lorsque le système de marquage a évolué au cours du suivi, par exemple lors du remplacement d'une bague perdue ou de l'utilisation successive de plusieurs types de PIT.

Chaque femelle s'est ensuite vu attribuer un identifiant unique indépendant des différents systèmes de marquage utilisés sur le terrain. Cet identifiant constitue désormais la référence utilisée dans l'ensemble des traitements ultérieurs.

3.1.3. Reconstruction des historiques de capture

Une fois les individus identifiés, les observations ont été ordonnées chronologiquement selon leur date d'acquisition. Pour chaque femelle, un rang de capture



(capture_rank_individual) a été attribué afin de numéroté successivement les différentes observations disponibles. Un second rang (capture_rank_year) a également été calculé pour décrire l'ordre des observations réalisées au cours d'une même année de ponte.

Les informations relatives à l'observation précédente ont ensuite été calculées automatiquement. Pour chaque événement, la date de la capture précédente, l'intervalle de temps exprimé en jours, l'année précédente d'observation ainsi que l'intervalle interannuel, si existant, ont été renseignés. Ces variables permettent de décrire la dynamique individuelle des femelles et serviront de base aux analyses futures portant sur les fréquences de retour et les intervalles de remigration.

3.1.4. Qualification des évènements de capture

Enfin, chaque observation a été qualifiée selon sa position dans l'historique individuel. Deux variables logiques ont été générées :

- same_season_reobservation, indiquant si la femelle avait déjà été observée au cours de la même saison de ponte ;
- interannual_reobservation, indiquant si la femelle correspond à une réobservation réalisée après au moins une année d'absence.

Cette distinction permet de différencier les réobservations intervenant au cours d'une même saison de nidification des véritables retours interannuels, ces derniers constituant une information essentielle pour les futures analyses démographiques.

3.1.5. Structure finale

La base finale comporte une ligne par événement de capture et regroupe l'ensemble des informations nécessaires à la reconstruction des historiques individuels. Les principales variables produites sont résumées dans le tableau suivant.

VARIABLE	DESCRIPTION
INDIVIDUAL_ID	Identifiant unique attribué à chaque femelle.
DATE	Date de l'observation.
YEAR	Année d'observation.
TAXA	Espèce observée.
PIT_TD	Identifiant électronique implanté dans le triceps droit (PIT.TD).
PIT_ED	Identifiant électronique implanté dans l'épaule droite (PIT.ED).
BAGUE	Identifiant(s) de bague(s) externe(s).
ALL_IDENTIFIERS	Concaténation de l'ensemble des identi-



CAPTURE_RANK_INDIVIDUAL	fiantes disponibles pour une observation. Rang chronologique de l'observation pour une femelle donnée.
CAPTURE_RANK_YEAR	Rang de capture au sein d'une même année.
PREVIOUS_OBSERVATION_DATE	Date de l'observation précédente.
DAYS_SINCE_PREVIOUS_OBSERVATION	Nombre de jours écoulés depuis la précédente observation.
PREVIOUS_OBSERVATION_YEAR	Année de la précédente observation.
YEARS_SINCE_PREVIOUS_OBSERVATION	Nombre d'années écoulées entre deux observations successives.
SAME_SEASON_REOBSERVATION	Indique une réobservation au cours de la même saison de ponte.
INTERANNUAL_REOBSERVATION	Indique une réobservation entre deux saisons de ponte distinctes.

Cette première base constitue la référence de travail pour l'ensemble des traitements ultérieurs. Elle sera utilisée pour produire la base synthétique des individus (BDD_LO_individuals) ainsi que pour identifier les observations atypiques intégrées au registre des valeurs aberrantes (BDD_LO_anomalies).

3.2. Création de la base de données des femelles individuelles

3.2.1. Principe général

La seconde étape de consolidation a consisté à transformer la base événementielle BDD_LO_clean_events en une base synthétique décrivant les caractéristiques de chaque femelle identifiée au cours du suivi. Contrairement à la base précédente, dans laquelle une même femelle pouvait être représentée par plusieurs lignes correspondant à des événements de capture, cette nouvelle structure regroupe l'ensemble des observations d'un même individu en une seule ligne.

Cette approche permet d'obtenir une vision synthétique des historiques individuels tout en conservant les principales informations nécessaires aux futures analyses démographiques. La base BDD_LO_individuals constitue ainsi une description individuelle de l'ensemble des femelles marquées entre 2004 et 2018.

3.2.2. Regroupement des identifiants individuels

Les observations appartenant à une même femelle ont été regroupées à partir de l'identifiant (individual_id) précédemment attribué lors de la construction de la base événementielle. Cette étape permet de reconstituer l'historique complet de



chaque individu indépendamment du nombre de captures ou de recaptures enregistrées au cours de son suivi.

Pour chaque femelle, la première et la dernière date d'observation ont été identifiées afin de définir la période couverte par son historique. Les années correspondantes ont également été extraites, permettant de caractériser la durée temporelle du suivi disponible pour chaque individu.

Le nombre total d'observations a ensuite été calculé, de même que le nombre d'années distinctes durant lesquelles la femelle a été observée. Ces deux indicateurs permettent de distinguer les individus fréquemment observés au cours d'une même saison de ceux présentant des retours répartis sur plusieurs années de reproduction.

3.2.3. Descripteurs individuels

L'ensemble des identifiants associés à chaque femelle a été regroupé dans une variable unique (`all_identifiers`). Cette variable rassemble tous les systèmes de marquage connus pour un individu, qu'il s'agisse de transpondeurs électroniques ou de bagues externes. Cette consolidation garantit la conservation de l'historique complet des dispositifs d'identification utilisés au cours du suivi.

Parallèlement, les années d'observation ont été synthétisées dans une variable dédiée (`years_observed`), permettant de visualiser rapidement la chronologie des saisons de ponte au cours desquelles une femelle a été observée.

Enfin, deux indicateurs biologiques ont été calculés afin de caractériser le comportement de retour des femelles. Le premier distingue les femelles observées une seule fois de celles ayant été réobservées au moins une fois au cours du suivi. Le second identifie les femelles présentant au moins une réobservation interannuelle, correspondant à un retour sur les plages de ponte au cours d'une année différente. Le nombre minimal de retours interannuels observés a également été calculé pour chaque femelle.

Ces variables constituent les premiers descripteurs individuels directement mobilisables pour les futures analyses démographiques, notamment dans le cadre des estimations de fidélité au site, des probabilités de réobservation et des intervalles de remigration.

3.2.4. Structure finale

La base finale comporte une ligne par femelle identifiée et regroupe l'ensemble des informations synthétiques décrivant son historique de capture-marquage-recapture.



VARIABLE	DESCRIPTION
INDIVIDUAL_ID	Identifiant unique attribué à chaque femelle après consolidation des différents systèmes de marquage.
FIRST_OBSERVATION_DATE	Première date d'observation connue de la femelle dans la base de données.
LAST_OBSERVATION_DATE	Dernière date d'observation connue de la femelle.
FIRST_YEAR	Première année au cours de laquelle la femelle a été observée.
LAST_YEAR	Dernière année d'observation disponible.
TOTAL_OBSERVATIONS	Nombre total d'événements de capture ou de recapture associés à la femelle.
OBSERVED_YEARS_COUNT	Nombre d'années distinctes durant lesquelles la femelle a été observée.
ALL_IDENTIFIERS	Ensemble des identifiants (PIT et bagues) associés à la femelle tout au long de son historique.
YEARS_OBSERVED	Liste chronologique des années de présence de la femelle dans la base de données.
REOBSERVED_STATUS	Statut indiquant si la femelle n'a été observée qu'une seule fois (vue_une_fois) ou au moins deux fois (réobservée).
INTERANNUAL_REMIGRANT_STATUS	Statut indiquant si la femelle présente (remigrante_interannuelle) ou non (non_remigrante_interannuelle) un retour observé entre plusieurs années de reproduction.
INTERANNUAL_RETURNS_COUNT	Nombre minimal de retours interannuels observés, calculé à partir des années distinctes de présence de la femelle.

Cette base constitue le second livrable de la prestation. Complémentaire de BDD_LO_clean_events, elle fournit une synthèse individuelle directement exploitable pour les analyses descriptives de la population.

3.3. Création de la base de données des valeurs aberrantes

3.3.1. Principe général

En complément des deux bases de données réalisées, un registre des anomalies potentielles a été élaboré, BDD_LO_anomalies, afin d'identifier les observations sus-



ceptibles de nécessiter une vérification avant la réalisation d'analyses démographiques. Contrairement aux étapes précédentes, cette procédure ne vise pas à modifier ni à supprimer des observations de la base BDD_LO_clean_events. Son objectif est de signaler les événements présentant une incohérence technique ou biologique potentielle, afin de faciliter leur examen par les gestionnaires de la base de données.

Ce registre constitue ainsi un outil d'aide à la décision et non un protocole d'exclusion automatique. Chaque observation signalée doit être réévaluée au regard des informations disponibles (fiches de terrain, carnets de suivi, historique de marquage, expertise des observateurs) avant toute décision d'écartement des analyses.

Les critères retenus reposent sur les connaissances actuelles relatives au suivi des tortues olivâtres en Guyane ainsi que sur les principales anomalies susceptibles d'être rencontrées dans une base de données historique de capture-marquage-recapture.

3.3.2. Détection des doublons techniques

Une première série de contrôles a consisté à rechercher les observations strictement identiques au sein de la base de données. Ces situations peuvent résulter d'une duplication accidentelle lors de la saisie ou de la fusion de plusieurs jeux de données historiques.

Les observations concernées ont été intégrées au registre afin d'être vérifiées. Lorsque leur caractère redondant est confirmé, une seule observation pourra être conservée dans les analyses ultérieures.

3.3.3. Réobservations à pas de temps très court

Un second filtre a été appliqué afin d'identifier les femelles réobservées moins de dix jours après une observation précédente. Chez *Lepidochelys olivacea*, un intervalle aussi court est peu compatible avec deux pontes indépendantes au cours d'une même saison de reproduction. Ces situations peuvent notamment correspondre à une tentative de ponte interrompue, une émergence sans ponte, une ponte avortée ou une réobservation de la même tentative de reproduction.

Ces observations ne sont pas nécessairement erronées mais méritent une attention particulière, car leur interprétation peut influencer le nombre d'événements reproducteurs attribués à une même femelle.

Le seuil de dix jours retenus dans cette étude constitue un critère de travail proposé à partir des connaissances biologiques actuellement disponibles. Il pourra être adapté en fonction des objectifs des analyses ou de nouvelles informations issues du suivi de terrain.



3.3.4. Conflit d'identification

Le registre recense également les situations dans lesquelles un même identifiant (PIT ou bague externe) est associé à plusieurs individus consolidés ou, inversement, lorsqu'une même femelle présente des changements inattendus de système d'identification.

Ces cas peuvent résulter d'erreurs de saisie, d'erreurs de lecture des identifiants, d'un remplacement de dispositif de marquage ou plus rarement d'une incohérence lors de la consolidation des historiques individuels.

L'identification de ces observations permet de cibler les vérifications documentaires les plus susceptibles d'améliorer la qualité de la base de données.

3.3.5. Évènements biologiquement atypiques

Enfin, plusieurs indicateurs biologiques ont été intégrés au registre afin de mettre en évidence des historiques individuels présentant des caractéristiques inhabituelles, telles que des intervalles de retour particulièrement courts ou exceptionnellement longs entre deux observations successives.

Ces situations ne doivent pas être considérées comme des erreurs *a priori*. Elles peuvent refléter une variabilité biologique réelle, des conditions environnementales particulières ou simplement les limites inhérentes au suivi de terrain. Leur signalement vise uniquement à attirer l'attention des utilisateurs de la base sur des observations dont la cohérence mérite d'être examinée avant leur utilisation dans des analyses quantitatives.

3.3.6. Structure finale

Le registre des anomalies constitue un troisième livrable indépendant des deux bases de données consolidées. Chaque ligne correspond à une observation ayant satisfait au moins un des critères de détection définis dans le protocole.

VARIABLE	DESCRIPTION
INDIVIDUAL_ID	Identifiant consolidé de la femelle concernée.
DATE	Date de l'observation signalée.
ANOMALY_TYPE	Nature de l'anomalie détectée (doublon, réobservation rapprochée, conflit d'identifiant, etc.).
JUSTIFICATION	Description synthétique du critère ayant conduit au signalement de l'observation.
RECOMMENDED_ACTION	Action proposée (vérification documentaire, contrôle des fiches de terrain, exclusion éventuelle des analyses, etc.).



Le contenu de ce registre est volontairement évolutif. Les critères de détection, les seuils biologiques retenus ainsi que les recommandations associées peuvent être adaptés en fonction de l'évolution des connaissances sur la biologie de l'espèce, de la disponibilité de nouvelles informations de terrain ou des objectifs spécifiques des analyses réalisées. Ainsi, ce registre doit être considéré comme un outil d'aide à l'expertise et au contrôle qualité des données, et non comme une liste définitive d'observations à exclure.

3.4. Quelques statistiques descriptives

La consolidation de la base historique a permis de constituer un jeu de données comprenant **21 282 événements de capture-marquage-recapture**, correspondant à **10 285 femelles identifiées** entre 2004 et 2018. Parmi ces femelles, **5 416 (52,7 %)** n'ont été observées qu'une seule fois au cours de la période d'étude, tandis que **4 869 (47,3 %)** ont été réobservées au moins une fois. Les historiques individuels mettent en évidence **3 673 réobservations intra-saison**, correspondant à plusieurs observations d'une même femelle au cours d'une même saison de ponte. À l'échelle interannuelle, **3 988 femelles (38,8 %)** présentent au moins un retour observé entre plusieurs années de reproduction. Ces historiques représentent **7 324 événements de réobservation interannuelle** enregistrés dans la base consolidée (TABLEAU 1).

TABLEAU 1 | Synthèse des principaux indicateurs descriptifs de la base marquage.

Indicateur	Valeur
Nombre total d'événements	21 282
Nombre de femelles identifiées	10 285
Femelles vues une seule fois	5 416
Femelles réobservées au moins une fois	4 869
Réobservations intra-saison (événements)	3 673
Femelles remigrantes interannuelles	3 988
Réobservations interannuelles	7 324
Moyenne observations/femelle	2,07
Part globale de femelles réobservées (%)	44,6
Maximum observations/femelle	13
Moyenne intervalle interannuel (années)	1,96
Médiane intervalle interannuel (années)	2
Maximum intervalle interannuel (années)	14



La majorité des femelles identifiées n'a été observée qu'une seule fois, tandis qu'un sous-ensemble plus restreint présente plusieurs événements de capture ou de recapture (FIGURE 3).

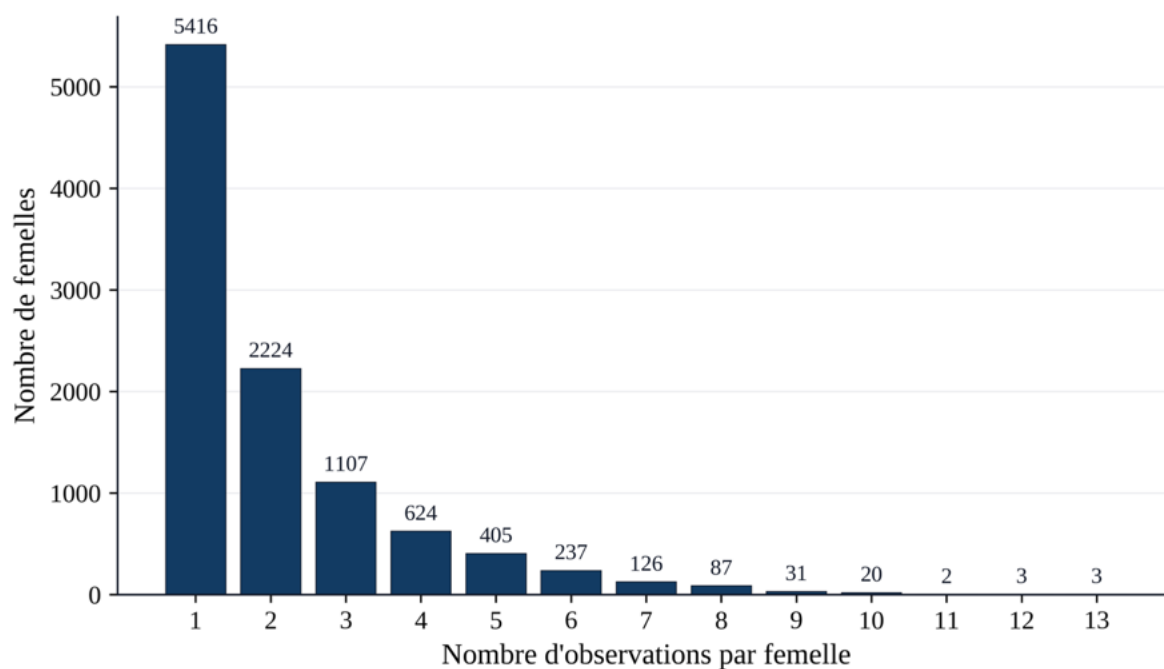


FIGURE 3 | Distribution du nombre d'observations par femelle.

La courbe du nombre de femelles identifiées cumulées illustre l'enrichissement progressif de la base de données au cours de la période 2004-2018. Chaque année contribue à augmenter le nombre total d'individus connus, ce qui reflète l'accumulation des observations issues des campagnes successives de terrain (FIGURE 4). Cela permet également d'apprécier la croissance du référentiel individuel disponible pour les analyses démographiques, tout en rappelant que l'interprétation des tendances doit tenir compte des variations potentielles d'effort d'échantillonnage entre années.

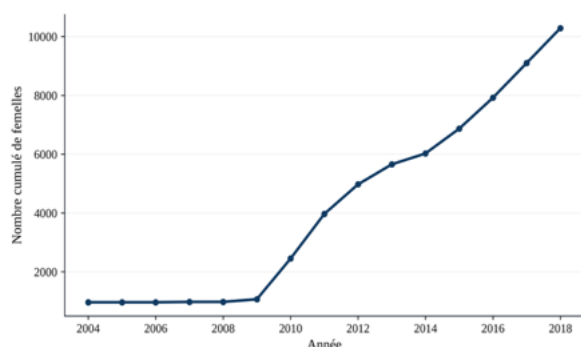


FIGURE 4 | Évolution cumulative du nombre de femelles identifiées.

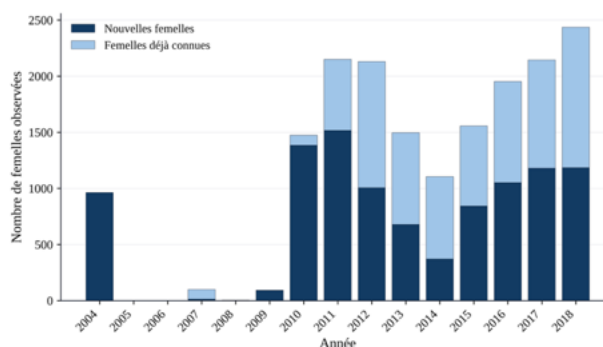


FIGURE 5 | Nombre annuel d'évènements d'identification.

La distribution annuelle des femelles nouvellement identifiées et des femelles déjà connues décrit l'évolution temporelle de la base consolidée entre 2004 et 2018. Les nouvelles femelles renseignent l'entrée progressive d'individus dans le suivi, tandis que les femelles réobservées traduisent la capacité du dispositif de terrain à retrouver des individus déjà marqués lors de saisons précédentes ou au sein d'une même saison de ponte (FIGURE 5).

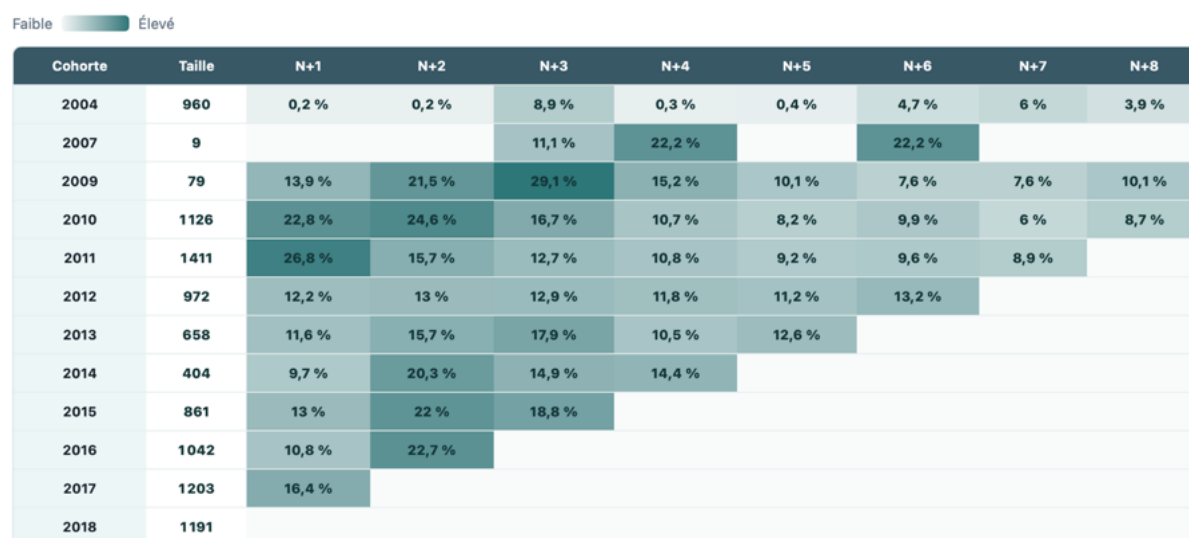


FIGURE 6 | Pourcentage de femelles réobservées en fonction de leur cohorte de première observation et de l'intervalle de retour observé (N+1 à N+8) chez la tortue olivâtre (*Lepidochelys olivacea*) en Guyane entre 2004 et 2018. La taille des cohortes correspond au nombre de femelles identifiées pour la première fois au cours de chaque année. L'intensité de la couleur indique la proportion de femelles réobservées pour chaque intervalle de retour.

La représentation des fréquences de réobservation par cohorte met en évidence une forte variabilité des probabilités apparentes de retour selon les années de première identification des femelles (FIGURE 6). Les cohortes les plus abondantes, notamment celles de 2010 ($n = 1126$), 2011 ($n = 1411$), 2012 ($n = 972$), 2016 ($n = 1042$) et 2017 ($n = 1203$), présentent des effectifs suffisants pour permettre une interprétation robuste des tendances observées. À l'inverse, certaines cohortes de faible taille, telles que celle de 2007 ($n = 9$), doivent être interprétées avec prudence, car quelques individus réobservés peuvent conduire à des pourcentages relativement élevés.

Les résultats montrent que les fréquences de réobservation les plus importantes sont généralement observées entre une et trois années après la première capture. La cohorte 2011 présente par exemple une fréquence maximale de retour à N+1 (26,8



%), tandis que les cohortes 2010, 2015 et 2016 atteignent leurs valeurs les plus élevées à N+2, avec respectivement 24,6 %, 22,0 % et 22,7 % de femelles réobservées. De même, la cohorte 2009 présente une fréquence maximale de retour à N+3 (29,1 %). Cette récurrence d'un maximum de réobservation observé entre deux et trois années après la première identification suggère l'existence d'une périodicité des retours compatible avec les intervalles de remigration décrits chez *Lepidochelys olivacea* dans d'autres régions tropicales.

Au-delà de ces intervalles préférentiels, les fréquences de retour tendent globalement à diminuer avec l'augmentation du temps écoulé depuis la première observation. Néanmoins, plusieurs cohortes anciennes conservent des proportions non négligeables de femelles réobservées après cinq années ou plus. La cohorte 2004, qui constitue l'une des séries les plus longues de la base de données ($n = 960$), présente encore des taux de réobservation de 4,7 %, 6,0 % et 3,9 % respectivement à N+6, N+7 et N+8. De manière similaire, la cohorte 2009 maintient des fréquences de retour supérieures à 7 % jusqu'à huit années après la première observation. Ces résultats témoignent de l'existence d'historiques individuels particulièrement longs et soulignent l'intérêt des suivis de longue durée pour documenter les trajectoires reproductrices des femelles.

L'interprétation de ces résultats doit toutefois prendre en considération les effets liés à la durée du suivi disponible. Les cohortes les plus récentes disposent nécessairement d'un horizon temporel d'observation plus restreint que les cohortes anciennes, ce qui limite la possibilité d'observer des retours à long terme. L'absence de valeurs dans les colonnes correspondant aux intervalles les plus élevés pour les cohortes récentes ne traduit donc pas une absence biologique de remigration, mais résulte principalement d'un phénomène de censure temporelle.

Dans leur ensemble, ces résultats suggèrent que les femelles olivâtres suivies en Guyane présentent une probabilité apparente de retour maximale au cours des deux à trois années suivant leur première observation, suivie d'une diminution progressive des fréquences de réobservation à mesure que l'intervalle temporel augmente. La persistance de retours observés jusqu'à huit années après le marquage met néanmoins en évidence une fidélité individuelle potentiellement importante aux sites de ponte, dont l'étude détaillée nécessitera la mise en œuvre d'analyses démographiques dédiées intégrant explicitement les probabilités de détection et les processus de survie.

Ces statistiques descriptives constituent une première synthèse de la population suivie et fournissent les principaux indicateurs nécessaires aux futures analyses de capture-marquage-recapture. Elles permettent notamment de quantifier l'intensité du suivi individuel, la fréquence des réobservations et la fidélité apparente des femelles aux sites de ponte. En revanche, l'identification d'une éventuelle périodicité des retours interannuels nécessite une analyse spécifique des intervalles de remigration, qui dépasse le cadre de la présente prestation.



3.5. Recommandations pour les futures campagnes de terrain

L'estimation du nombre de PIT nécessaires pour les campagnes futures constitue un élément important de la planification logistique des opérations de terrain. Il convient toutefois de souligner que le nombre de PIT déployés n'influence pas directement la probabilité individuelle de réobservation, laquelle dépend également de facteurs biologiques (survie, fidélité au site, périodicité des remigrations), méthodologiques (effort de suivi, probabilité de détection) et environnementaux. En revanche, l'augmentation de la taille des cohortes permet d'accroître le nombre absolu de femelles susceptibles d'être réobservées et améliore ainsi la robustesse statistique des analyses démographiques futures.

À partir des fréquences de réobservation observées dans les cohortes récentes, plusieurs scénarios opérationnels peuvent être proposés pour la campagne de marquage 2026.

TABEAU 2 | Scénarios d'acquisition de PIT pour la campagne de marquage 2026.

Scénario	Nombre de PIT prévus	Femelles potentiellement réobservées (20 %)	Interprétation
MINIMAL	1 000	≈ 200	Cohorte exploitable mais offrant une marge limitée pour les analyses futures
RECOMMANDE	1 300	≈ 260	Compromis satisfaisant entre contraintes logistiques et objectifs scientifiques
SECURISE	1 500	≈ 300	Marge de sécurité permettant d'absorber les variations interannuelles et les pertes potentielles de matériel
AMBITIEUX	1 750	≈ 350	Scénario optimisant le suivi à long terme et la puissance des futures analyses CMR

Au regard des effectifs historiques observés à Rémire-Montjoly, une dotation comprise entre **1 300 et 1 500 PIT** apparaît comme une cible réaliste pour la saison 2026. Un tel effectif permettrait raisonnablement d'espérer disposer, à moyen terme, d'environ **250 à 300 femelles réobservées à N+1 ou N+2**, sous réserve du maintien d'un effort de suivi comparable à celui réalisé au cours des années précédentes.

Au-delà de la seule campagne 2026, la poursuite de ces opérations de marquage revêt un intérêt majeur pour l'amélioration des connaissances sur la dynamique de la population de tortues olivâtres de Guyane. La constitution de longues séries tempo-



relles constitue en effet une condition indispensable pour estimer de manière robuste les paramètres démographiques tels que la survie apparente, les probabilités de réobservation, les intervalles de remigration, la fidélité aux sites de ponte ainsi que l'abondance des femelles reproductrices.

Dans cette perspective, la pérennisation des suivis CMR apparaît comme un investissement essentiel pour renforcer la qualité des jeux de données disponibles et soutenir le développement futur d'approches démographiques avancées permettant une meilleure compréhension de l'état de conservation de cette population.

4. CONCLUSION

Cette prestation a permis de consolider, harmoniser et structurer l'ensemble des données historiques de capture-marquage-recapture acquises sur les femelles tortues olivâtres (*Lepidochelys olivacea*) suivies sur la plage de Rémire-Montjoly entre 2004 et 2018. À partir d'une base de données brute présentant des hétérogénéités de saisie, de formatage et d'identification, un protocole reproductible de traitement statistique sous R a été développé afin de garantir la cohérence des historiques individuels et de faciliter leur exploitation future.

Les traitements réalisés ont conduit à la production de trois livrables complémentaires : une base événementielle standardisée (BDD_LO_clean_events) décrivant l'ensemble des observations individuelles, une base synthétique à l'échelle des femelles (BDD_LO_individuals) regroupant les principaux descripteurs biologiques et temporels associés à chaque individu, ainsi qu'un registre des anomalies potentielles (BDD_LO_anomalies) destiné à documenter les observations nécessitant une expertise complémentaire avant leur éventuelle exclusion des analyses.

L'exploration descriptive de ces données met en évidence l'existence de plusieurs milliers de femelles identifiées au cours de la période d'étude, une proportion importante de réobservations intra- et interannuelles ainsi que des historiques individuels s'étendant sur plusieurs années, témoignant du potentiel de ce jeu de données pour l'étude de la fidélité aux sites de ponte, des intervalles de remigration et de la dynamique démographique de la population.

Au-delà de la structuration des données historiques, ce travail constitue désormais une base méthodologique directement exploitable pour le développement futur d'analyses démographiques avancées fondées sur des approches de capture-marquage-recapture telles que celles qui seront développées par le CNRS dans le cadre du projet INVOCTO. Enfin, ce travail permet de démontrer l'utilité d'une telle campagne de marquage, d'autant plus quand les données sur la pêche illégale en Guyane montrent un doublement de l'effort de pêche au minimum au cours de la dernière décennie selon les récentes études menées par le CRPMEM.

