Animal Movement Cleaner Application : un outil de pré-traitement des données de biologging

Aurélie Jambon

MINISTERE DE L'AGRICULTURE, DE L'AGROALIMENTAIRE ET DE LA FORET

ECOLE NATIONALE SUPERIEURE des SCIENCES AGRONOMIQUES de BORDEAUX
AQUITAINE
1, cours du Général de Gaulle - CS 40201 – 33175 GRADIGNAN cedex

M E M O I R E de fin d'études
Pour l'obtention du titre
D'Ingénieur de Bordeaux Sciences Agro

Animal Movement Cleaner Application : Un outil de pré-traitement des données de bio-logging

par

Jambon Aurélie

Spécialisation : AgroTIC

Étude réalisée à : Natural Solutions, 68 Rue Sainte 13001 Marseille

- 2019 –

MINISTERE DE L'AGRICULTURE, DE L'AGROALIMENTAIRE ET DE LA FORET

ECOLE NATIONALE SUPERIEURE des SCIENCES AGRONOMIQUES de BORDEAUX
AQUITAINE
1, cours du Général de Gaulle - CS 40201 – 33175 GRADIGNAN cedex

M E M O I R E de fin d'études
Pour l'obtention du titre
D'Ingénieur de Bordeaux Sciences Agro

Animal Movement Cleaner Application : Un outil de pré-traitement des données de bio-logging

Animal Movement Cleaner Application: A pre-treatment tool for bio-logging data

par

Jambon Aurélie

Maitre de stage : Jean-Vitus Albertini

Tuteur de stage : Nicolas Devaux

Remerciements

La rédaction de ce mémoire a été rendue possible grâce à l'accompagnement de plusieurs personnes auxquelles j'aimerais exprimer ma gratitude.

J'aimerais tout d'abord remercier mon maitre de stage, Jean-Vitus Albertini, pour sa patience, sa bienveillance et son expertise. Le développement de cette application web n'aurait pas été possible sans lui.

Je tiens ensuite à remercier l'ensemble de l'équipe de Natural Solutions pour son accueil chaleureux et sa bonne humeur chaque jour. C'est un plaisir de venir travailler dans cette ambiance !

Je voudrais particulièrement remercier Olivier Rovellotti, Frédéric Berton et Christelle Khozian pour m'avoir accordé leur confiance dès le début.

Mes remerciements vont également à l'équipe enseignante, notamment à mon tuteur, Nicolas Devaux, pour avoir été présent aux moments où j'en avais besoin, ainsi qu'à Lionel Bombrun pour avoir accepté de donner son temps et son point de vue à l'occasion du point mi-stage. Je n'oublierais pas le soutien sans faille de Nathalie Toulon, François Thiberville, Léo Pichon et Bruno Tisseyre lors de la recherche de stage.

Enfin je tiens à remercier ma famille et Adrien Pajot pour avoir toujours cru en moi et m'avoir soutenue tout le long de mes études et de ce stage.

Résumé

Le bio-logging consiste à utiliser des balises miniaturisées pour dévoiler les comportements des animaux sauvages. Il permet l'accès à une connaissance jusqu'alors hors de portée. Bien que les données récoltées soient accueillies avec enthousiasme, elles doivent également l'être avec prudence. En effet, les jeux de données brutes peuvent contenir des données aberrantes qui fausseraient les résultats si elles étaient conservées lors de l'analyse.

Ce mémoire s'intéresse au développement d'un outil de pré-traitement de ces jeux de données. Dans un premier temps, les origines des localisations aberrantes en fonction de la technologie de géolocalisation ont été décrites. Ensuite, une analyse des besoins a mis en avant que l'intérêt d'un nouvel outil automatisant le pré-traitement des données aberrantes réside dans sa facilité d'utilisation et dans la place de l'utilisateur en tant que décideur final. Il doit également pouvoir fonctionner pour toutes les espèces, toutes les technologies de géolocalisation et prendre en compte tous les formats de données.

AMCA, Animal Mouvement Cleaner Application, a donc été développée en prenant en compte tous ces paramètres. Il s'agit d'une application web éliminant les données impossibles et celles présentant une vitesse aberrante, en plus de détecter une immobilité. Elle fonctionne actuellement pour la technologie GPS et le format de données de Movebank®. Ses atouts sont sa facilité d'utilisation et la visualisation 3D laissant l'utilisateur libre de modifier les résultats des algorithmes.

Mots clés : *bio-logging, télémétrie, écologie des mouvements, pré-traitement, application web*

Abstract

Bio-logging consists in using miniaturised tags to uncover animals' hidden lives. It gives access to former unreachable data. Resulting data are usually welcomed with eagerness, but caution should also be added. Indeed, Raw data may contain outliers which could skew analysis if kept.

Hence, the present dissertation deals with the development of a pre-treatment tool for those datasets. First of all, outliers' origins have been described depending on the geolocation technology. After that, a need analysis led to the conclusion that a new tool doing automated pre-treatment to remove outliers would only be useful if it is easy to use and if the user has the final word. The tool must also work for any species, any geolocation technology and any dataset structure.

Those parameters have been considered during AMCA, Animal Movement Cleaner Application, development. It is a web application which purpose is to eliminate impossible data and data with out-of-range speed. It is also able to detect immobility. Currently, AMCA is working for GPS technology and Movebank data structure. Easy to use, it also enables users to modify algorithms-kept and eliminated data thanks to a 3D visualization.

Key words: *bio-logging, telemetry, movements ecology, pre-treatments, web application*

Sommaire

Table des illustrations

Liste des abréviations

API : Application programming interface

ARGOS : Advanced research and global observation satellite

BAS : British Antarctic survey

BSPB : Bulgarian Society for the Protection of Birds

CEBC : Centre d'Etudes Biologiques de Chizé

CLS : Collecte localisation satellites

CNES : Centre national d'études spatiales

CSS : Cascading style sheets

GDOP : Geometric dilution of precision

GLS : Global location sensor

GPS : Global positioning system

GSM : Global system for mobile communication

HTML : Hypertext mark-up language

http : Hypertext transfer protocol

JS : JavaScript

LC : Localisation Class

LPO : Ligue de protection des oiseaux

NASA : National aeronautics and space administration

NOAA : National Oceanic and Atmospheric Administration

PTT : Platform transmitter terminal

REST : Representational state transfer

SIG : Système d'information géographique

UHF : Ultra haute fréquence

UICN (IUCN en anglais): Union Internationale pour la Conservation de la Nature

URL : Uniform Resource Locator

VHF : Very high frequency

Glossaire

Back-end : Il s'agit de l'une des deux parties d'une application web. Elle permet de réaliser des calculs sur le serveur en utilisant ses capacités.

Balise : Également appelée tag, logger ou data logger. Elle correspond à l'objet attaché à un animal, embarquant des technologies et capteurs afin d'enregistrer des données au cours de ses déplacements.

Bio-logging : utilisation de balises miniaturisées attachées à des animaux afin d'enregistrer des données sur les mouvements, le comportement, la physiologie et/ou l'environnement d'un individu.

Domaine vital : Le domaine vital correspond à une zone sur laquelle un individu se nourrit et se reproduit.

Framework : Infrastructure logicielle en français. Il donne un cadre lors du développement, met en place la structure d'une application. Son utilisation permet de simplifier et d'uniformiser le travail du développeur.

Front-end : Ce deuxième composant d'une application web est exécutée par le client, ou navigateur, et permet la visualisation de l'application. Elle est écrite à l'aide des trois langages que sont html, CSS et JavaScript.

Télémétrie : Principe consistant à mesurer une distance par des procédés acoustiques, optiques ou radioélectriques.

Introduction

Les mouvements des animaux, comme les migrations, sont connus et observés depuis des siècles, notamment celles des oiseaux. On savait donc que certaines espèces disparaissaient en hiver, sans savoir où elles se rendaient. En 1822, une cigogne blanche (*Ciconia ciconia*) est découverte en Allemagne avec une lance provenant d'une tribu d'Afrique centrale dans le cou. Ce fut la première preuve que cette espèce descendait hiverner en Afrique (Bairlein, 2008). En 1850, les premières recherches sont organisées via des postes d'observation en notant les dates et les heures de passage.

D'après Dan Rubenstein (Princeton University, 2018), "To follow the dynamics of population, you have to follow the fate of individuals. In order to study individuals, you have to recognize every individual". La même idée a amené à la mise en place des premiers bagages en 1899 (Bairlein, 2008). Cette technique basée sur le principe de capture-marquage-recapture permet une grande avancée dans la connaissance des routes migratoires, notamment grâce à l'établissement d'une collaboration internationale. Cette connaissance des mouvements comble une curiosité, mais est aujourd'hui également vitale dans le processus de conservation des espèces et de leur environnement. En effet, connaître leurs habitats permet par exemple d'identifier des menaces ou de les prévenir.

L'arrivée de la télémétrie et du bio-logging ont permis d'atteindre une autre dimension de la connaissance des mouvements, générant des données bien plus précises et continues. Le bio-logging correspond à l'utilisation de balises miniaturisées attachées à des animaux afin d'enregistrer des données sur les mouvements, le comportement, la physiologie et/ou l'environnement d'un individu (Rutz and Hays, 2009). Parmi les données enregistrées, celles de position sont obtenues via télémétrie, principe consistant à mesurer une distance par des procédés acoustiques, optiques ou radioélectriques (Larousse, 04/09/2019).

L'apport de l'ingénierie rend possible la poursuite de la miniaturisation des balises, l'amélioration de la précision des données, de la capacité de stockage et de l'efficacité des réseaux de transmission, afin d'aboutir à des outils de plus en plus performants à des prix réduits (Bograd *et al.*, 2010; Evans, Lea and Patterson, 2012). L'objectif de ces progrès techniques est d'augmenter le panel d'espèces pouvant être étudiées (annexe 1) et de minimiser la subjectivité liée à ces données. En effet, un certain recul est nécessaire lors de l'analyse.

Tout d'abord, attacher une balise à un animal peut **altérer des comportements naturels** en raison de perturbations visuelles, physiques ou aérodynamiques. Le but est donc de diminuer cet impact au maximum en commençant par réduire la taille de l'objet.
Ensuite, la qualité des données peut représenter un deuxième biais en raison d'une **fiabilité plus ou moins importante des technologies de localisation**. Cet aspect peut-être résolu en améliorant directement les technologies embarquées, mais aussi en mettant au point des protocoles de pré-traitement des données (Ropert-Coudert and Wilson, 2004). Ces pré-traitements doivent cependant être **adaptés à chaque espèce** et être réalisés avec **vigilance** dans le cadre d'études sur des animaux, sujets peu prévisibles.
De plus, un suivi continu génère d'une **quantité importante de données** (Li *et al.*, 2015; Thums *et al.*, 2018) pouvant être de divers types et manquant souvent de **standardisation**.

Le développement d'outils s'avère alors utile et jalonné de contraintes afin de faciliter ces pré-traitements pouvant être rendus longs et fastidieux sur de tels jeux de données. **Quel serait alors l'outil optimal pour accélérer et simplifier cette phase de pré-traitement des données de mouvement des espèces, aboutissant à des données utilisables, afin de rendre plus efficace l'accès à la connaissance et donc les projets de conservation ?**

Afin de répondre à cette problématique, nous nous attarderons d'abord sur l'influence de l'espèce et des objectifs d'une étude sur le choix de la balise à utiliser, décision ne pouvant être prise sans une connaissance des technologies de géolocalisation. Ensuite une présentation de l'entreprise et de l'organisation du stage sera réalisée avant de développer chaque étape et leur rôle dans la production de l'outil final. Nous finirons par revenir sur l'outil développé en énonçant ses limites et perspectives.

I. Une balise adaptée, un point clé pour une étude cohérente en bio-logging

Il n'est pas toujours aisé de suivre un animal à la trace pour comprendre ses mouvements et leurs raisons. En fonction des espèces, le suivi peut être plus ou moins évident. En effet, certaines sont sédentaires, fidèles à leur territoire - plus ou moins restreint -, et le sillonneront toute leur vie, comme le chevreuil *Capreolus capreolus* chez les mammifères ou le geai des chênes *Garrulus glandarius* chez les oiseaux (Conservation nature, 03/09/2019). D'autres ont des déplacements variant en fonction des saisons, et nous pouvons alors assister à des mouvements migratoires, comme chez le renne *Rangifer tarandus* (Panzacchi, Van Moorter and Strand, 2013) ou la grue cendrée *Grus grus* (*Migraction*, 03/09/2019). Ainsi, il est nécessaire d'établir des méthodes de suivi adaptées à chacune des espèces à l'aide d'outils technologiques de pointe comme les balises.

Une balise, ou data logger, se compose de différents éléments que sont la source d'alimentation, le mode de communication, la technologie de géolocalisation et les différents capteurs pouvant y être associés. Le choix de ces éléments va influencer les propriétés de l'objet comme son poids, son prix ou encore sa longévité (figure 1). En fonction de la morphologie de l'espèce et du budget dont dispose l'étude, ces propriétés peuvent devenir des contraintes et amener à faire des compromis concernant les données, leur précision, leur récupération, ou encore la durée de l'étude. De cette manière, une connaissance des technologies existantes, de leurs principes, avantages et inconvénients s'impose avant de faire le choix de la balise à utiliser dans un contexte défini.

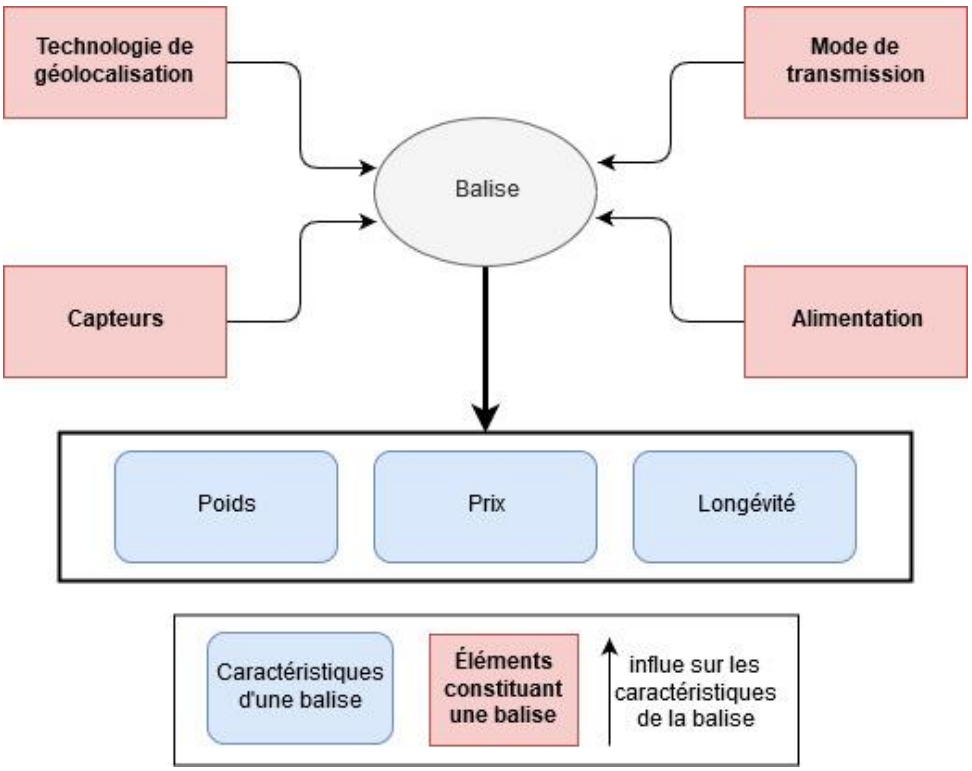


Figure 1. Schéma des éléments et propriétés d'une balise.

A. Les données récupérées par la balise

Parmi l'ensemble des données récupérables par une balise, une attention particulière sera portée aux données de géolocalisation et à leurs caractéristiques afin de comprendre les enjeux et contraintes à appréhender lors du développement d'un outil de pré-traitement.

1. Le positionnement diffère selon les technologies de géolocalisation

Une synthèse des technologies de géolocalisation les plus utilisées a été réalisée afin de comprendre leurs principes et l'origine des erreurs affectant leurs précisions. Ces facteurs font partie de ceux déterminant si une technologie peut être utilisée au cours d'une étude.

a. La radio télémétrie et l'étude des mouvements locaux

La radio télémétrie VHF (Very High Frequency) est apparue au début des années 1960 et a été la première technologie utilisée pour suivre les animaux. Le principe repose sur la communication par ondes radios entre un émetteur placé sur l'animal et un ou plusieurs récepteurs. Plusieurs méthodes peuvent être mises en œuvre afin de localiser ces émetteurs de façon plus ou moins précise.

Quelle que soit la méthode employée, la réception du signal se fait via une antenne radio à régler sur la fréquence recherchée. Cette antenne peut être statique, disposée à un endroit stratégique et accompagnée d'un enregistreur, permettant alors d'avoir des informations sur la présence d'un individu dans un périmètre dépendant de la portée du signal. Néanmoins, lorsque l'étude nécessite des informations plus précises sur les déplacements de l'animal, des moyens plus coûteux doivent être mis en place. Une méthode de suivi à pieds utilise le concept de triangulation, dont l'application la plus rigoureuse utilise trois observateurs munis d'antennes, obtenant ainsi la position moyenne de l'animal au moment de la prise de mesure (figure 2). Un signal très fort dans toutes les directions signifie que l'observateur a atteint l'individu balisé, la localisation est alors certaine. L'utilisation d'un véhicule ou d'un avion permet de couvrir des zones plus larges, même si ces alternatives sont généralement synonymes de perte de précision. En fonction des moyens disponibles dans le cadre de l'étude, un compromis ou une combinaison des techniques peuvent être employés. (Whitworth and FAO, 2007, chap. 7; Malgouyres *et al.*, 2017)

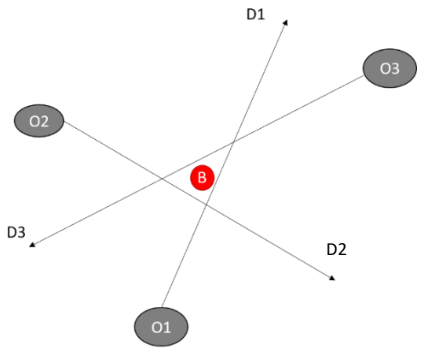


Figure 2. Méthode de géolocalisation par triangulation. OX = Observateur X, DX = Direction du signal de l'observateur X, B = Barycentre du triangle correspondant à la position moyenne de la balise à l'instant t de la mesure.

La description de cette technologie et des méthodes utilisées laisse cependant présager la nécessité d'une proximité immédiate entre l'animal et l'observateur, demandant des ressources importantes et réduisant son application à l'analyse des mouvements locaux.

b. La géolocalisation par la lumière ou GLS (Global location sensor)

Cette technologie développée par le British Antarctic Survey (BAS) repose sur un principe utilisé depuis des siècles dans la navigation maritime qui consiste à déduire une localisation en fonction de la longueur du jour, qui varie selon la latitude, et de l'heure du zénith, variant selon la longitude. De cette manière, les balises (geolocators) attachées aux individus suivis enregistrent l'intensité lumineuse et l'angle du soleil, informations ensuite utilisées pour calculer la durée du jour entre le lever et le coucher du soleil. Ces calculs sont réalisés à l'aide de logiciels (par exemple BAS Track développé par le British Antarctic Survey) une fois les données récupérées à la recapture de l'animal et aboutissent à une estimation des positions et du trajet emprunté par ce dernier entre les deux captures.

Comme on peut l'imaginer, les sources d'erreur sont diverses et notamment liées à des facteurs atténuant le signal lumineux (saison, latitude, couverture nuageuse, pollution lumineuse, ombrage, orientation du capteur, comportement de l'animal), mais aussi à des facteurs biaisant la localisation (vitesse de déplacement, dérive du chronomètre). De plus, d'après le travail réalisé par Jean-Baptiste Thiebot dans sa thèse (THIEBOT Jean-Baptiste, 2011, p. 50), les données récoltées sur les périodes d'équinoxe sont également à écarter car peu qualitatives.

Ainsi, cette technologie engendre des données peu précises, avec une erreur globale estimée à ± 150km, plus importante en latitude (132 ± 75km) qu'en longitude (50 ± 34km). Cette information oriente fortement l'utilisation des geolocators pour l'étude des oiseaux migrateurs au détriment des mouvements locaux, même si ces derniers pourraient être étudiés en considérant seulement la donnée de longitude. (Bächler *et al.*, 2010; Fudickar, Wikelski and Partecke, 2011).

c. ARGOS (Advanced research and global observation satellite)

Le système Argos a été développé par le Centre national d'études spatiales (CNES), la United States national oceanic and atmospheric administration (NOAA), et la National aeronautics and space administration (NASA) dans les années 1970. CLS (Collecte localisation satellites) a été créé en tant que filiale du CNES en 1986 afin d'exploiter, entretenir et commercialiser le système. La maintenance est aujourd'hui assurée en coopération avec plusieurs agences spatiales internationales (CLS, 12/07/2019).

Le fonctionnement de ce système repose sur quatre éléments (CLS, 12/07/2019) (annexe 2):

Une **balise** (souvent nommée PTT = Platform transmitter terminal) est placée sur l'individu à localiser et émet régulièrement un message contenant le numéro d'identification unique de la balise et les données à transmettre. La localisation étant calculée à partir de *l'effet Doppler*, la fréquence d'émission du message (401.650 MHz ± 30 kHz) doit être stable.

Les **satellites Argos**, décrivant une orbite polaire à 850 km d'altitude, reçoivent les messages en provenance des émetteurs et en déduisent la fréquence et la date de réception. Si la fréquence reçue est supérieure à celle réellement émise (connue), alors le satellite se rapproche. Sinon c'est qu'il s'éloigne (figure 3). Une fréquence de réception correspond en réalité à deux positions possibles sur le Terre, symétriques par rapport à la trace au sol du satellite (points à la verticale du satellite).

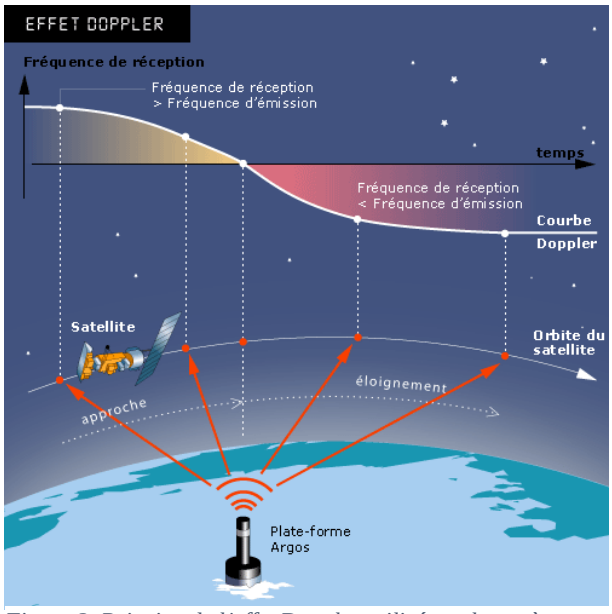


Figure 3. Principe de l'effet Doppler utilisé par le système Argos. Source : (CLS, 12/07/2019)

Ces informations brutes sont ensuite transmises aux **centres de traitement** via des **antennes** afin d'effectuer les calculs pour aboutir à une localisation en choisissant parmi les deux obtenues. Source : (CLS, 12/07/2019)

Les localisations Argos étant données par des satellites en orbites polaires, l'estimation de l'erreur se représente alors à l'aide d'une ellipse décrivant une erreur plus importante en longitude (annexe 3). CLS fournit les informations suivantes liées à l'ellipse : le rayon moyen de l'erreur, le rayon de l'ellipse suivant l'axe le plus long, le rayon de l'ellipse suivant l'axe le plus court, l'orientation de l'ellipse et le GDOP (Geometric dilution of precision).

La précision des localisations Argos peut être très variable. CLS a établi des classes de qualité (LC, Location Class) des données en fonction de l'erreur estimée et du nombre de messages associés à cette localisation (tableau 1).

Tableau 1. Précision des données Argos d'après CLS. Source : (Argos système, 2016, p. 14)

Classe	Erreur estimée	Nombre de transmissions
3	< 250 m	≥ 4
2	250-500 m	≥ 4
1	500 – 1500 m	≥ 4
0	>1500 m	≥ 4
A	Pas d'estimation	3
B	Pas d'estimation	2
Z	Localisation invalide	

L'observation des données récupérées sur le terrain, dans le contexte d'études des animaux, indique une majorité de données de classes de faible qualité (0, A, B, Z). De plus, ces erreurs diffèrent de celles annoncées par CLS. Les données de classes standards (entre 1 et 3) possèdent des erreurs plus importantes qu'annoncé et les données de faible qualité s'avèrent de précisions très variables. Les études confirment la présence d'une erreur horizontale plus importante (même en conditions idéales), ce qui va dans le sens de l'utilisation d'une ellipse d'erreur.

Ainsi, les véritables erreurs seraient plus proches des valeurs indiquées dans le tableau 2 (C. Douglas *et al.*, 2012; Laurentiu Rozylowicz *et al.*, 2018).

Tableau 2.Précision des données Argos d'après (C. Douglas et al.*, 2012)*

Classe	Erreur estimée (68%) (en m)
3	400
2	1000
1	2500
0	>10 400
A	8 100
B	30 500
Z	30 300

Dans le cadre de la technologie Argos, les erreurs de géolocalisation sont majoritairement dues à la vitesse, mais aussi à la visibilité de l'émetteur (topographie, végétation), à la position de l'antenne au moment de la transmission (Laurentiu Rozylowicz *et al.*, 2018), ou à de rapides changements de température.

d. GPS

Le principe de géolocalisation par satellite, GNSS (Global Navigation Satellite System), permet de localiser un objet à la surface du globe via des constellations de satellite comme celle déployée par les Etats-Unis : le GPS (Global positioning system). Cette dernière comporte 30 satellites en orbite à 20000km d'altitude.

Le principe est le suivant : la balise GPS fonctionne comme un récepteur de signaux émis par des satellites. Elle utilise alors deux informations du signal reçu : sa date de réception et sa date d'émission. Sachant que le signal voyage à la vitesse de la lumière, elle peut alors facilement en déduire la distance la séparant du satellite. Ainsi, trois satellites sont nécessaires pour localiser la balise : il s'agit du principe de la trilatération (figure 4) (CNES, 2017).

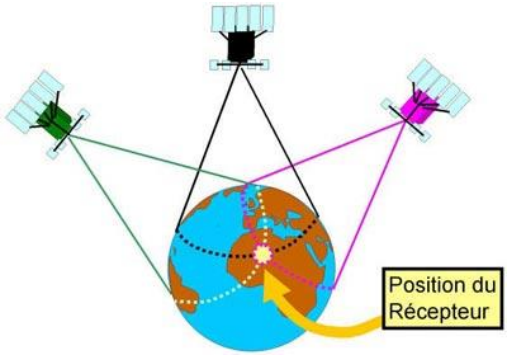
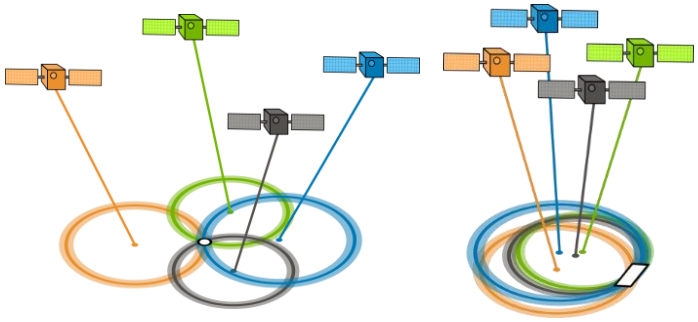


Figure 4. Trouver un point par trilatération.
Source : (Escadrone, 2018)

Il existe plusieurs facteurs pouvant être à l'origine d'erreurs de position par le système GPS. Ces derniers sont décrits dans le tableau ci-dessous (tableau 3) et constituent le mode naturel de la localisation GPS, dont la précision est de l'ordre d'une centaine de mètres.

Tableau 3. Tableau décrivant les erreurs liées à la géolocalisation GPS

Origine de l'erreur	Description
Organisation des satellites **Dilution Of Precision (DOP)** **–** **Geometric DOP, Vertical DOP, Horizontal DOP**	Décrit l'erreur provoquée par la position relative des satellites GPS. Plus le récepteur GPS reçoit des signaux d'un grand nombre de satellites bien répartis dans le ciel (le moins regroupés possible), meilleur (plus faible) sera le GDOP et donc la précision de la position (figure 5).(GIS Geography, 23/07/2019)  *Figure 5. Géométries des satellites participant à la localisation d'une balise. Celle de gauche est éclatée et résulte en une localisation précise, contrairement à la deuxième qui est groupée. Source : (GIS Geography, 23/07/2019)*
Réfraction atmosphérique	Le passage dans la troposphère et dans l'ionosphère peut entrainer un changement de vitesse du signal GPS et donc entrainer une diminution de précision.
L'effet multi-trajets	Le signal GPS peut être réfléchi par des structures – des bâtiments ou des montagnes par exemple - présentes à proximité de la balise. Le récepteur reçoit alors le même signal plusieurs fois à des intervalles de temps différents.
Précision des horloges atomiques	Il peut y avoir un décalage entre les horloges atomiques des satellites et l'horloge de la balise. La localisation étant basée sur une différence de temps, une synchronisation des horloges est nécessaire. Une erreur de trois milliardièmes de seconde résulte en une erreur de 1 mètre dans la localisation. (CNES, 2017)

Contrairement à ce qu'on pourrait penser, la nébulosité ne joue en rien sur la qualité des données GPS.

Plusieurs techniques permettent d'améliorer cette précision en jouant sur un ou plusieurs des facteurs (figure 6) (Escadrone, 2018).

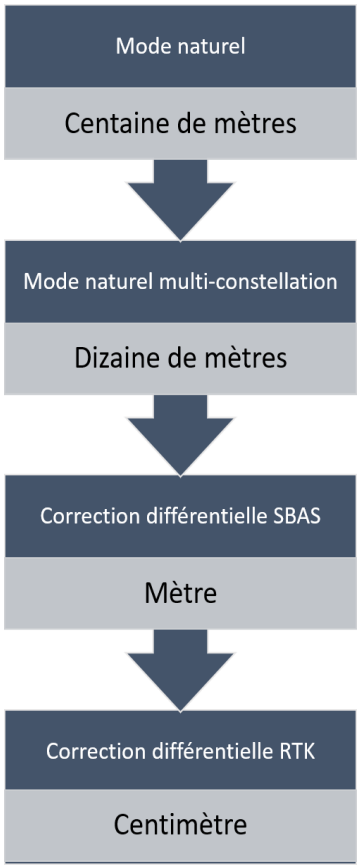


Figure 6.Différentes corrections permettant d'améliorer la précision GPS

Tout d'abord, même si le système GPS utilise le principe de trilatération, un quatrième satellite est utilisé afin de réaliser une synchronisation des horloges atomiques au dixième de milliardième de seconde.

La constellation de satellites GPS n'est pas la seule en orbite, ainsi, certains récepteurs sont capables d'utiliser plusieurs constellations (Galileo, GLONASS…) et plusieurs fréquences pour calculer leur position. Le résultat est alors une moyenne des positions obtenues par chaque constellation et/ou fréquence.

Il existe ensuite des méthodes de correction dite différentielle.

Le Système d'augmentation spatiale satellitaire (SBAS) utilise la position d'un point connu pour calculer l'erreur GPS et la transmettre aux récepteurs GPS compatibles via des satellites géostationnaires (EGNOS pour l'Europe) afin de la retirer de la mesure observée. Il s'agit de la correction utilisée en bio-logging.

La correction différentielle RTK (Real time kinematic) utilise également une base de position connue mais cette fois dont la localisation est très proche de celle du récepteur étudié, ayant donc une valeur d'erreur similaire. Cette correction n'est pas utilisée pour la localisation des animaux.

La localisation représente une des données pouvant être enregistrée par un data logger. D'autres informations complémentaires peuvent venir enrichir le jeu de données final.

2. Des données variées pour des études aux sujets variés

Les études utilisant le concept du bio-logging ne s'intéressent pas seulement aux trajets effectués par les animaux mais également à d'autres données décrivant leur état physiologique ou l'environnement qui les entoure. Le tableau ci-dessous offre un panorama des données existantes en bio-logging (tableau 4).

Tableau 4. Types de données pouvant être récoltées via une balise de bio-logging. 1. Location classes. Seulement pour les balises Argos. Source : (Lotek, 07/09/2019; Microwave Telemetry, Inc., *31/05/2019)*

Type de données	**Informations sur la balise**	**Données de physiologie**	**Données environnementales**	**Donnée de mouvements**	**Qualité des données de position**
Données	Etat de la batterie	Température interne, mort	Altitude, pression, température, luminosité, champ magnétique, salinité, acoustique	Latitude, longitude, vitesse, activité, accélération (en x, y et z), direction	LC[1], HDOP, VDOP, nombre de satellites

Cette diversité de données disponibles dépend des capteurs associés à la balise, offre variant en fonction du fournisseur. Aujourd'hui, un grand nombre d'entreprises propose des produits destinés au suivi des animaux. Certaines, non spécialisées et issues du milieu de l'internet des objets (IOT) ou de l'électronique, ont diversifié leur offre pour s'ouvrir à ce nouveau marché. D'autres sont spécialisées dans le suivi des espèces, voire de certains taxa. En fonction des capteurs développés, chaque fournisseur fournit sa propre structure de données, autant dans le fond que dans la forme. Ce manque de standardisation est un enjeu important dans le cadre du développement d'outils de pré-traitements et d'analyse.

Le premier fournisseur de balises de bio-logging, WildLife Computers, a été créé aux Etats-Unis en 1986. Depuis, leur nombre a augmenté, mais la majorité reste basée aux Etats-Unis et au Japon. Même si quelques entreprises européennes commencent à faire leur place, un grand nombre de balises sont encore réalisées sur-mesure dans les institutions de recherche. (Ropert-Coudert *et al.*, 2009; Le Galliard *et al.*, 2012)

Une fois les données disponibles et leurs caractéristiques connues, le choix idéal des informations espérées peut être contraint par d'autres éléments de la balise… Il convient alors d'étudier le système et les relations entre les composants pour atteindre l'outil optimal.

B. Des caractéristiques dépendantes des choix relatifs aux données

Comme nous avons pu le voir précédemment, les technologies de géolocalisation influent notamment sur la précision des données. Cependant, le choix de cette caractéristique implique des contraintes sur d'autres facteurs à prendre en compte et pouvant amener à une reconsidération de la décision. Il s'agit du coût, de la consommation d'énergie, de la taille de la balise et du milieu d'utilisation (figure 7).

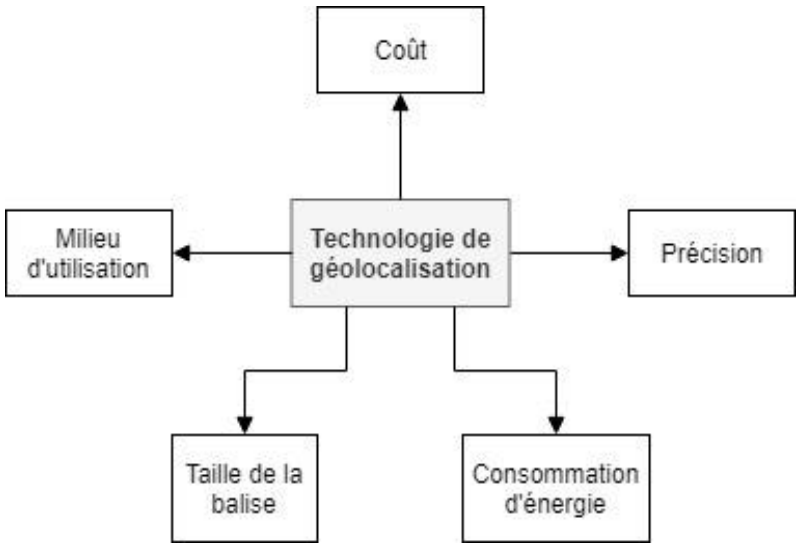


Figure 7. Schéma des différents paramètres impliqués dans le choix de la technologie de géolocalisation.

Ci-dessous, un tableau comparatif (tableau 5) permet d'avoir une vue claire sur les caractéristiques de chaque technologie afin de faire un choix adapté aux besoins.

Tableau 5. Caractéristiques d'une balise en fonction de la technologie de géolocalisation. 1. GPS associé à des technologies de transmission de données. 2. GPS enregistreur nécessitant une recapture. 3. Calculs réalisés en considérant que la balise représente 3% du poids de l'animal, 5% étant le maximum légal (Meyburg and Meyburg, 2009) . Source : (Pathtrack, 08/08/2019; Thomas Robertson, D Holland and Minot, 2012; MTIs Coverage of the World's Bird Species 2018, *2018)*

Technologie / Critère	Radio télémétrie VHF	Géolocalisation par la lumière	ARGOS	GPS
Milieu d'utilisation	Tous, sauf aquatique	Ouvert	Tous sauf aquatique et perte de précision en milieux fermés	Tous sauf aquatique et perte de précision en milieux fermés
Précision	Dépend de la méthode	±150km	0,25 – 15km	cm – 100m
Masse balise	0,26g	0,3g	2g	22g[1] – 0,95g[2]
Masse espèces pouvant être équipées[3]	> 9g	> 10g	>180g	> 700g[1] – >32g[2]
Alimentation	Solaire/batterie			
Récupération des données	Radio	Recapture	Satellite	Radio, recapture, satellite, réseau mobile
Prix	Croissant →			

Ainsi, tout animal dont la masse est inférieure à 32g ne peut pas être suivi en temps réel à une précision permettant d'étudier les déplacements locaux. De plus, l'utilisation d'une balise est limitée aux espèces de masses supérieures à 9g.

Quand certaines technologies nécessitent une recapture (GLS) ou une présence à proximité (radio télémétrie VHF), d'autres sont ou peuvent être associées à des moyens de transfert de données à distance et en temps réel pour un meilleur suivi des individus. Un avantage est également à considérer du point de vue de l'analyse des données : elle peut être faite par étape, sur un jeu de données fragmenté, pouvant aboutir à des résultats avant la fin de l'étude. Néanmoins, cette étape est la plus consommatrice en énergie, et ce augmentant avec la régularité de transmission. La géolocalisation GPS est la seule à avoir plusieurs options parmi lesquelles choisir.

Le tableau ci-après (tableau 6) donne un aperçu des technologies existantes et de leurs caractéristiques.

Tableau 6. Etat des lieux des technologies de communication permettant la transmission de données sur les balises GPS. 1. Par exemple Ornitela inclus 180€ d'abonnement dans le prix de la balise, permettant de transférer un nombre illimité de données lors de 1 à 4 connexions au réseau par jour, et ce pendant au moins 6mois, en fonction de la zone géographique.(Ornitela, 10/08/2019)

	Recapture	**VHF ou UHF**	**GSM (Global system for mobile communications)**	**Argos (Argos système, 2016, p. 13)**
Principe de la transmission	Transmission USB ou filaire entre la balise et un ordinateur	La balise enregistre les données et les transmet lorsqu'elle est à proximité d'un récepteur	Utilise les signaux numériques des réseaux de communication mobile	Transmission radio-satellite
Portée		Faible	Mondiale, en fonction de la couverture réseau	Mondiale
Prix	Ressource humaine	Ressources humaines ou antenne automatique	Abonnement mobile[1]	Dépend de la fréquence de téléchargement
Consommation		-	++	+
Influence sur la taille de la balise		+	++	+

Nous avons donc pu entrevoir la complexité du choix d'une balise de bio-logging. La figure 8 permet d'en faire un bilan, montrant que ce choix dépend d'abord de l'espèce et des objectifs de l'étude. La difficulté vient notamment des liens entre les éléments et caractéristiques, impliquant l'influence de certaines décisions sur d'autres, ce qui aboutit régulièrement à une nécessité de compromis. L'importance est donc d'établir des priorités sur les caractéristiques recherchées.

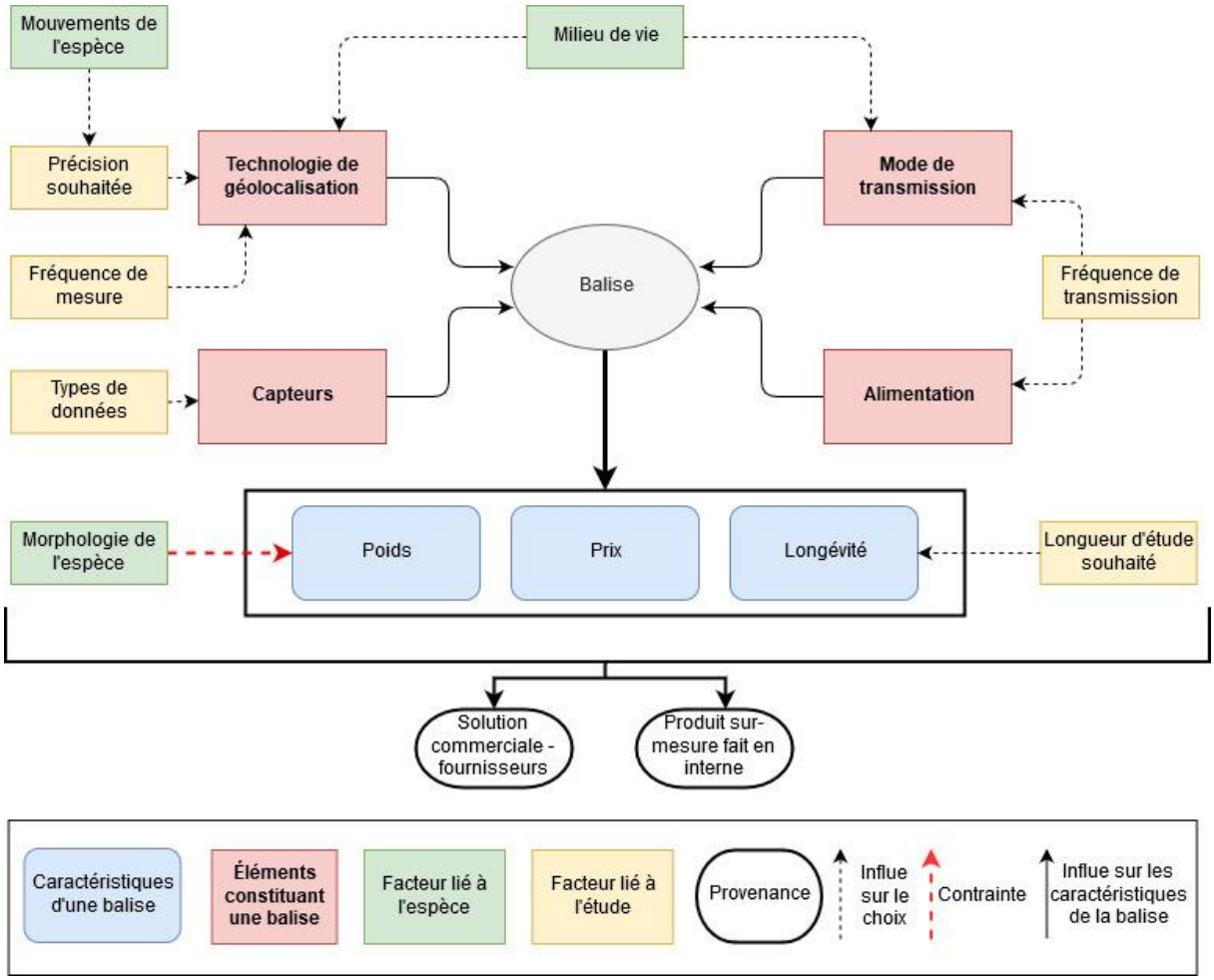


Figure 8. Schéma bilan indiquant les différents facteurs influant sur le choix d'une balise, réalisé en collaboration avec Adrien Pajot (Bordeaux Sciences Agro – stagiaire au CEBC (Centre d'Etudes Biologiques de Chizé))

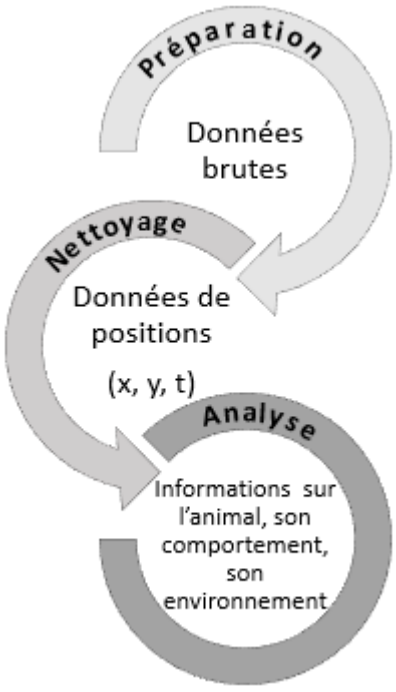


Figure 9. Chaine de traitement des données en écologie des mouvements

Une fois la balise choisie, cette dernière peut collecter les données qui aboutiront à des informations sur l'animal sauvage, sur ses comportements en milieu naturel et son environnement. Cependant, quelques étapes sont nécessaires avant d'atteindre cette connaissance. En fonction de la technologie, les données doivent être préparées afin d'aboutir à des données de positions en fonction du temps, puis nettoyées avant de pouvoir procéder à une phase d'analyse (figure 9). Ces phases de préparation et de nettoyage sont néanmoins dépendantes des **technologies de géolocalisation**, des **fournisseurs** et de **l'espèce étudiée**, rendant toute généralisation compliquée.

Quel serait alors l'outil optimal pour accélérer et simplifier cette phase de pré-traitement des données de mouvement des espèces, aboutissant à des données utilisables, afin de rendre plus efficace l'accès à la connaissance et donc les projets de conservation ?

II. Quel outil développer dans le contexte d'une entreprise telle que Natural Solutions

A. Natural Solutions, une agence digitale spécialisée dans l'innovation pour l'environnement

1. Présentation

Natural Solutions est une entreprise agile spécialisée dans la création et le développement de solutions numériques au service de l'environnement depuis 2008. Son but est de développer des solutions numériques – applications mobiles, web, logiciels… - dans les domaines de l'environnement, la biodiversité et la gestion des territoires. On peut distinguer deux types de projets. Le premier consiste à la réponse à des appels d'offres amenant à la réalisation de produits sur-mesure pour des clients (parcs nationaux, offices du tourisme etc.). Cependant, une partie de l'activité reste dédiée à l'innovation amenant à la création de produits en interne et une activité commerciale de vente. Parmi ces produits, on peut entre-autres citer EcoBalade, une application accompagnée d'un site web permettant notamment aux communes de créer des balades avec un tracé et des espèces clés à voir ; ou encore EcoRelevé, dont l'objectif premier est de gérer des données de terrain, mais restant adaptable à diverses problématiques.

Les réalisations de Natural Solutions se regroupent selon trois axes :

- La saisie de données
- La structuration de données
- La valorisation de données

Ceci implique la présence de compétences pluridisciplinaires pouvant être découpées en pôles :

- Recherche et développement : recherche et conception d'innovations
- Développement informatique : développement web, mobile et base de données à partir de cahiers des charges.
- Commerce : prospection d'appels d'offre et vente des produits Natural Solutions

Ainsi, l'équipe de Natural Solutions est pluridisciplinaire et s'organise comme suit (figure 10):

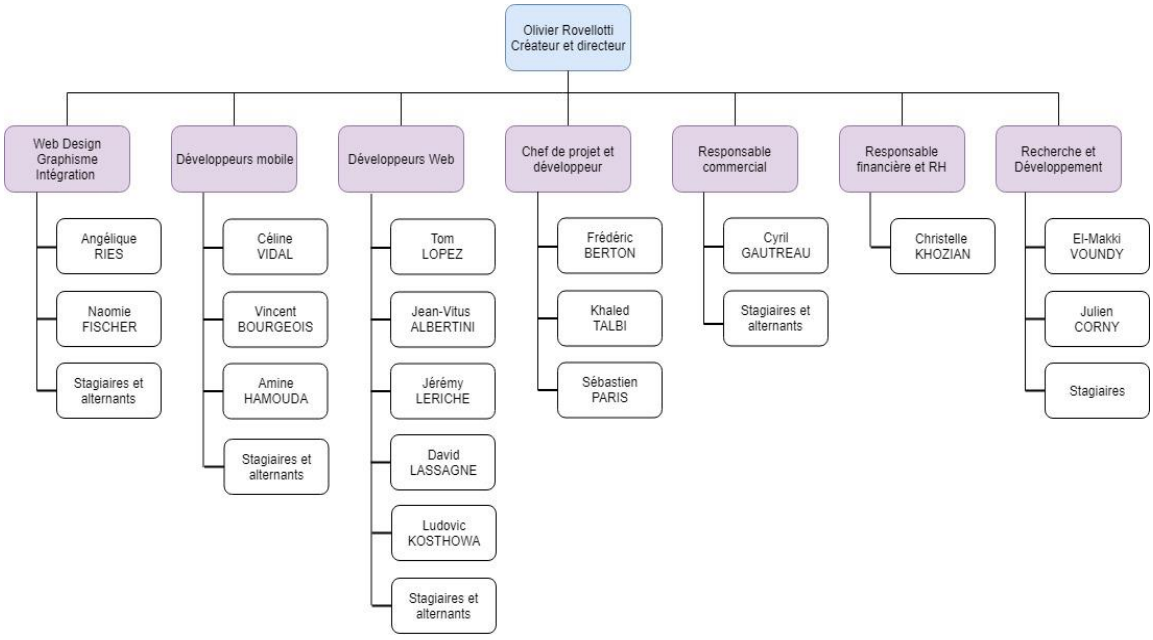


Figure 10.Organigramme de Natural Solutions

2. Contexte du stage et organisation

a. Déroulement du stage

Ma place dans cette organisation a été parmi l'équipe des développeurs web en R&D, afin de développer mon projet de stage qui est considéré comme une création interne.

L'idée est venue avec une demande d'un client majeur, Reneco, d'intégrer un filtre de points aberrants à l'application web EcoRelevé. Reneco est une entreprise spécialisée dans l'élevage conservatoire de l'outarde Houbara (*Chlamydotis undulata)*. Cette entreprise au fonctionnement multisites utilise un certain nombre d'applications développées par Natural Solutions pour accompagner ses différentes activités : élevage, vétérinaire, nourrissage et étude de l'écologie autour de l'espèce.

Si ce client a ce besoin, pourquoi pas les autres ? C'est un sondage réalisé lors de l'analyse des besoins qui répondra à cette interrogation. L'analyse des besoins a constitué une des étapes du stage, la totalité du déroulement de ce dernier étant visible dans le planning de la figure 11.

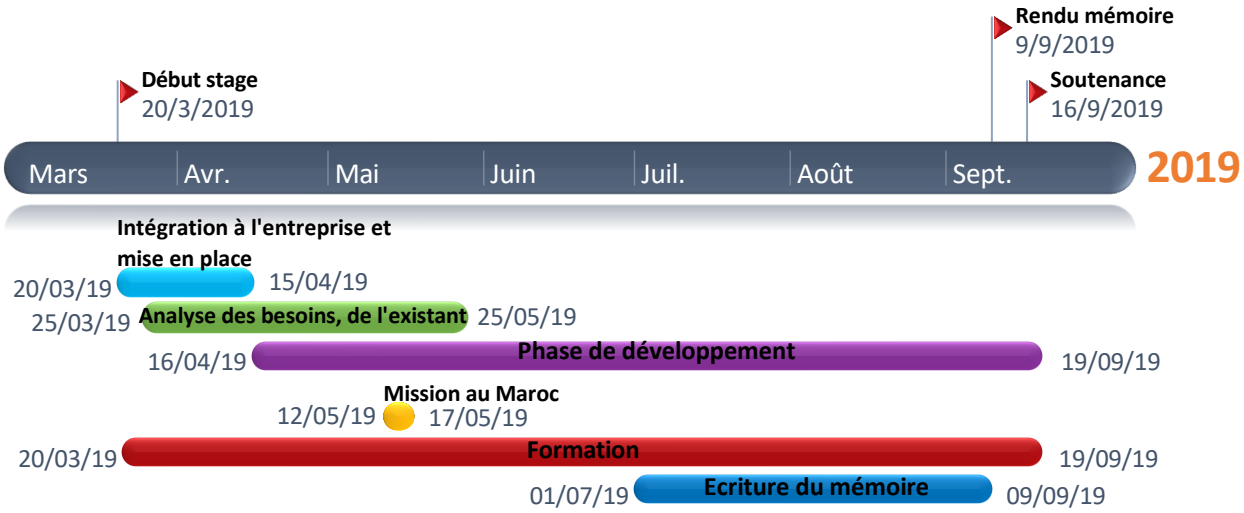


Figure 11. Planning du déroulement du stage

Les étapes d'analyse des besoins et de développement seront abordées dans la suite du mémoire. L'**intégration à l'entreprise** correspond à une phase d'explication du fonctionnement en méthode agile, de présentation des différents outils utilisés en interne et des produits développés.

La **mission au Maroc**, dans un centre de l'entreprise Reneco, a permis de voir sur place le fonctionnement d'un client majeur de Natural Solutions et d'assister à l'utilisation des outils développés pour eux. Cette mission a également été l'occasion de discuter avec le département écologie de leurs besoins en ce qui concerne l'élimination des points aberrants dans les données de suivi des outardes relâchées.

La **formation** indiquée sur toute la longueur du stage fait référence à la formation en développement web nécessaire pour la réalisation de mon projet de stage, mais aussi à une formation en SQL avec Microsoft SQL server sur des bases de données d'écologie.

b. La méthode agile

Natural Solutions fonctionne selon la méthode agile de gestion de projet, qui régit son quotidien. L'unité de temps est le « sprint », ici correspondant à une itération de deux semaines se déroulant comme suit (figure 12) :

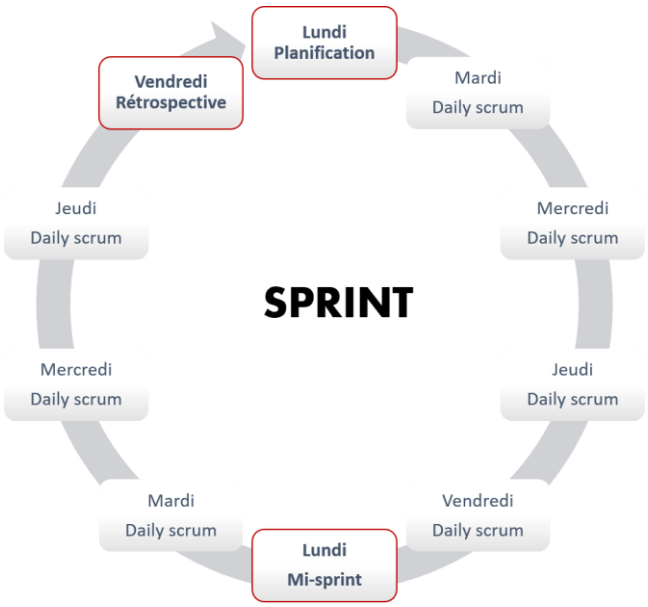


Figure 12. Organisation d'un sprint en méthode agile

La phase de **planification** sert à définir les tâches à réaliser au cours du sprint parmi les stories en attente. Les stories sont gérées et tenues à jour via un outil nommé *Pivotal tracker*. Ensuite, l'avancement du sprint est accompagné par *Scrum Manager*, un outil développé en interne et spécifique aux besoins de Natural Solutions. Ce dernier sert d'appuie lors des réunions de **mi-sprint** et de **rétrospective** et permet de rendre compte de l'avancement des objectifs fixés pendant la planification. Chaque jour, les réunions « Daily scrum » permettent de faire un état des lieux sur le travail qui a été fait la veille, les difficultés rencontrées, et le travail qui est prévu le jour même.

Dans le cadre de mon stage, cette méthode a été un peu détournée de son but initial qui est d'organiser une équipe de projet et de communiquer régulièrement avec le client. Elle m'a surtout permise de structurer mon temps entre missions principale et secondaires, ainsi que de mettre en place un bon séquençage des tâches afin d'avancer par étapes, d'avoir une idée claire des éléments à réaliser, mais aussi de pouvoir changer de stratégie à tout moment.

B. Analyse des besoins

Cette phase d'analyse des besoins et de l'existant part d'une initiative personnelle et ne faisait pas partie des livrables demandés par l'entreprise. Son but est de cadrer le projet afin que le contexte et les limites soient clairs. Elle permet de savoir qui seront les utilisateurs finaux et quels sont leurs besoins dans le but de faire un outil le plus utile possible.

1. Les potentiels utilisateurs identifiés

Les destinataires de l'outil sont tous les utilisateurs de balises de bio-logging. Ces derniers peuvent avoir des profils divers entre écologues et chercheurs, travaillant dans des parcs nationaux, ONG, universités, centres de recherche ou encore organismes publics.

Movebank®, une base de données en ligne créée en 2007 et consacrée au suivi des animaux (elle sera décrite plus en détails dans la partie II.B.3 sur les outils existants), accueillait 5 421 études sur 837 taxa différents en décembre 2018. Une étude comprenant souvent plus d'un individu, la plateforme recense donc aujourd'hui plus de 5 421 individus suivis. Même si elle ne recueille pas la totalité des données de bio-logging, un grand nombre d'acteurs de ce milieu

l'utilisent aujourd'hui, ces chiffres donnant ainsi une bonne indication de l'avancée de ces études.

En effet, comme le représente la figure 13 dans le contexte restreint des espèces plongeuses (oiseaux, cétacés, reptiles) dont les études ont été recensées dans la base de données « Penguiness Book », nous sommes encore bien loin d'un suivi exhaustif des espèces présentes sur notre planète. Ainsi, le nombre d'études ne peut que progresser dans les années à venir.

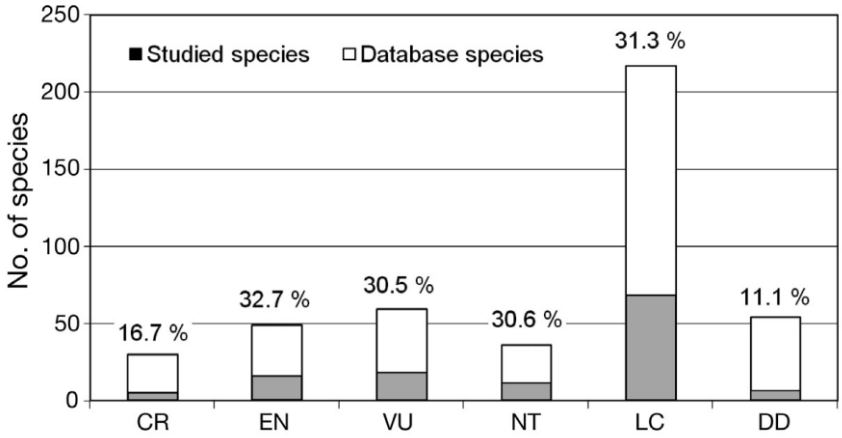


Figure 13. Proportion des espèces plongeuses suivies par bio-logging. Nombre d'espèces considérées dans la base de données Penguiness Book (barres blanches) et le nombre d'espèces correspondant dont l'activité de plongée a été étudiée par bio-logging (barres grises) en fonction du statut de conservation, comme défini dans la liste rouge établie par l'UICN en 2008. CR : en danger critique, EN : en danger, VU : vulnérable, NT : quasi menacée, LC : préoccupation mineure, DD : données insuffisantes. Source : (Ropert-Coudert et al.*, 2009)*

Cependant, le secteur de l'écologie et de la conservation des espèces n'a que peu de budget, amoindrissant les espoirs de vente d'un nouvel outil au milieu d'outils open-source ou réalisés sur-mesure par les organismes et chercheurs. Ainsi, après avoir ciblé les utilisateurs potentiels, il convient d'étudier leurs besoins et les outils existants.

2. Sondage : but et résultats

a. But et mise en place

Afin de bien comprendre la réalité des pratiques, le sondage devait permettre de répondre à quatre grands axes :

1/ Connaitre les caractéristiques des organisations et des études réalisées avec les balises.

2/ Avoir un panorama des outils utilisés pour la gestion des données.

3/ Comprendre les besoins actuels.

4/ Récupérer des contacts pouvant fournir des données à tester.

Il a été réalisé avec Google Form® pour la simplicité de réalisation et de diffusion par mail. La diffusion a notamment commencé par les listes mails du Parc National des Ecrins (grâce à une collaboration existante avec Natural Solutions), du projet Life sur le vautour percnoptère, ainsi qu'une liste interne des chercheurs du CEBC (Centre d'études Biologiques de Chizé). Une autre source de contacts possible est celle des responsables des études stockées sur la plateforme Movebank. Cependant, son usage a été limité afin d'éviter un biais dans les résultats. Ainsi, les contacts sélectionnés pour participer au sondage ont été choisis de manière aléatoire. D'autres

recherches indépendantes visant à trouver des études ou des organismes utilisant des balises ont été réalisées. De cette manière, le Museum National d'Histoires Naturelles, la LPO (Ligue de protection des oiseaux) ou le Parc National des Pyrénées ont été contactés. Pour finir, le sondage a été posté sur Tweeter® et sur le groupe « biodiversity professionals » sur Linkedin®.

b. Résultats

Le sondage a permis de récolter 46 réponses, comprenant 39 organismes différents utilisant des balises. Les réponses concernant les espèces sont de précision variable : certains participants ont répondu pour plusieurs projets/espèces, d'autres par type d'espèces, ou encore seulement pour un projet. Le graph suivant (figure 14) indique la représentation des réponses par type d'espèce.

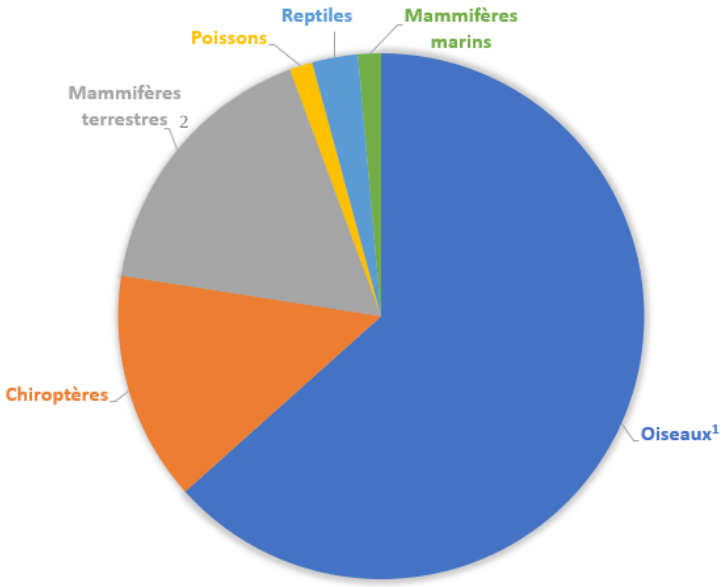


Figure 14. Types d'espèces représentées dans les réponses au sondage. 1. 61% des espèces d'oiseaux représentées sont ici des rapaces, face à 18% d'oiseaux marins et 21% autres. 2. Les ongulés sauvages (bouquetins, chevreuil, chamois) constituent ici une part très importante des mammifères terrestres

En 2014, lors du cinquième symposium réservé au bio-logging, 90% des travaux présentés étaient consacrés aux mammifères marins ou aux oiseaux (GaëlleFehlmann and Andrew J.King, 2016). Ainsi, la proportion des espèces d'oiseaux citées dans le sondage ne semble pas complètement aberrante. Néanmoins, la part des rapaces est certainement surreprésentée, notamment en comparaison avec les oiseaux marins. De plus, la part des mammifères marins est largement sous-représentée. Il est plus difficile de savoir quelles proportions attendre pour les autres taxa, même si la proportion des chiroptères semble très importante.

Pourquoi certaines espèces sont-elles surreprésentées dans les études de bio-logging ?

Plusieurs raisons peuvent expliquer cela. Tout d'abord, le facteur limitant de la taille face aux technologies GPS et Argos impliquent une surreprésentation des espèces de grande taille, favorisant ainsi les grands mammifères marins, les rapaces, et les grands mammifères terrestres. Un deuxième facteur influe sur la représentation d'une espèce dans ces études : la facilité de capture. En effet, plus une espèce est facile à capturer, moins de moyens seront à déployer pour la pose, voire la récupération de la balise en fonction de la technologie utilisée. Ainsi, les espèces vivant en colonie (Manchots), nichant chaque année au même endroit, ayant de petits territoires, etc. ont plus de chances d'être étudiées via bio-logging (Ropert-Coudert *et al.*, 2009).

Parmi ceux ayant répondu utiliser des balises, 83% ont des flottes de moins de 50 balises (figure 15). Ceci reflète les faibles budgets des projets de conservation et de recherche.

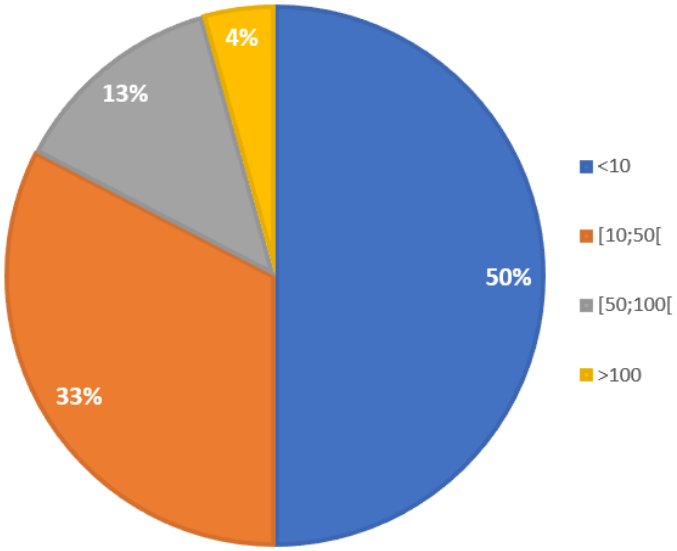


Figure 15. Résultats du sondage concernant l'importance des flottes de balises déployées.

Les technologies de géolocalisation utilisées sont variables (figure 16), même si le suivi par GPS semble représenter une majorité. Ceci paraît logique face à la dominante d'espèces de grandes tailles (rapaces et gros mammifères terrestres), qui par ailleurs sont les espèces les plus étudiées par bio-logging pour cette caractéristique physique. La part importante de VHF est notamment expliquée par la présence de nombreuses études sur les chauves-souris.

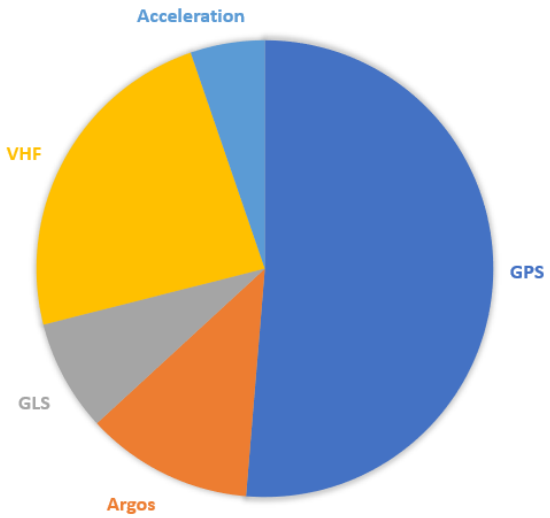


Figure 16. Répartition des technologies de géolocalisation utilisées dans les réponses au sondage.

Ces technologies sont utilisées dans des études analysant souvent plusieurs aspects liés à l'écologie de l'espèce (figure 17). Dans un cadre de conservation, les questions les plus étudiées restent le phénomène de migration, l'étude des menaces pesant sur une espèce, la stratégie de recherche de nourriture et l'estimation du domaine vital.

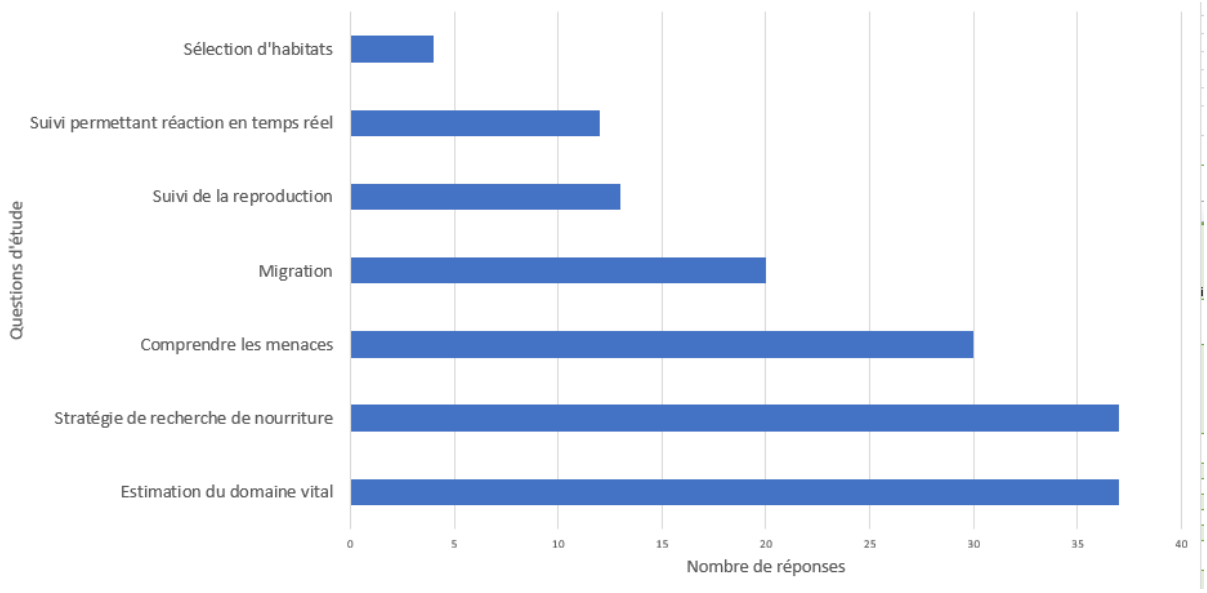


Figure 17. Réponses du sondage concernant les objectifs des études utilisant du bio-logging. Seules les réponses les plus récurrentes (>2 réponses) ont été sélectionnées.

Une grande diversité d'outils sont actuellement utilisés pour stocker/traiter et analyser les données (figure 18). Parmi eux, R® et ses packages, les logiciels de SIG (Système d'information géographique) et Movebank® sont les plus utilisés. La présence de Excel® montre la diversité des niveaux de spécialisation des organismes. Tous les outils cités dans les réponses à cette question ne sont pas destinés aux mêmes utilisations, on y trouve principalement des outils de stockage, et d'analyse.

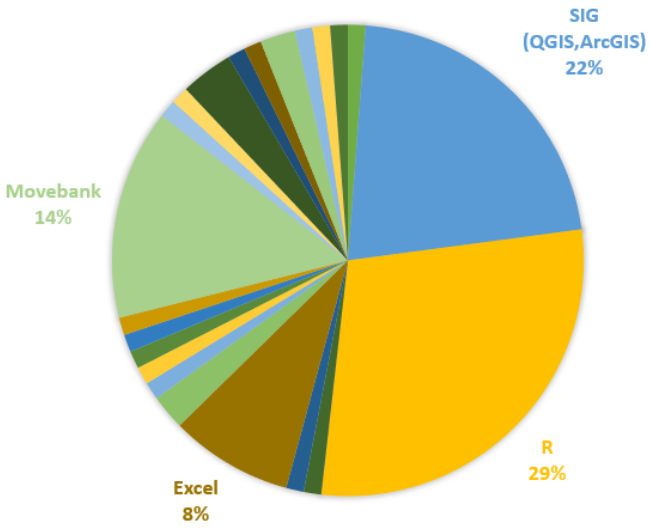


Figure 18. Réponses du sondage concernant les outils utilisés pour gérer les données.

Une question ouverte sur les difficultés rencontrées à l'heure actuelle a reçu des réponses variables, notamment en fonction de l'espèce étudiée. Les problèmes revenant le plus régulièrement sont ceux du temps de traitement (et donc d'argent consacré à payer les ressources), et de son ampleur (nombre de pré-traitements à associer à la quantité importante de données).

La question « Seriez-vous intéressé(e) par un outil effectuant des analyses automatiques de vos données ? (Sur la base d'algorithmes pour le repérage des données aberrantes et de l'apprentissage pour l'analyse de certains comportements) » a accueilli 93% de « oui ». Dont 45% l'utiliserait seulement dans certains cas, ces derniers étant répertoriés dans le tableau 7. Les personnes aillant répondu « non » sont toutes issues d'organismes de recherche réputés et donc aillant accès à des compétences et des outils pour répondre à ces problématiques en interne.

Tableau 7. Réponses des cas dans lesquels serait utilisé l'outil.

Cas des intérêts	Nombre de réponses
Données inutilisables/aberrantes	6
Domaines vitaux	4
Déplacements/ repérage de corridors	3
Traitement données GLS	2
Corrélations entre individus	1
Analyses des interactions oiseaux de mer/bateaux	1
Annoter les comportements sur les données d'accélérométrie	1
Caractérisation des habitats	1

Les potentiels utilisateurs attendent de cet outil (dans l'ordre du nombre de réponses, il s'agissait d'une question ouverte) : qu'il soit facile d'utilisation, qu'il permette un gain de temps et qu'il y ait une bonne visualisation.

A ce stade, nous avons donc les informations suivantes concernant l'outil à développer : la demande de Natural Solutions voulant élaborer un outil pour éliminer les points aberrants et les multiples besoins des utilisateurs. Cependant, nous avons pu voir que certains outils étaient déjà utilisés. A quoi servent-ils ? Une étude des outils existants est donc nécessaire afin de ne pas réaliser de doublon.

3. Les outils existants

Les résultats du sondage ont montré que les utilisateurs se servent déjà de nombreux outils, que ce soit pour stocker ou traiter les données. Le projet du stage est de réaliser un outil de pré-traitement spécifique pour les données de bio-logging géolocalisées. Ainsi, nous laisserons de côté les outils généraux tels que Access, Excel ou PostGRE pour nous focaliser sur les outils spécialisés.

Commençons par Movebank, plateforme déjà citée de nombreuses fois dans ce mémoire. Movebank est une base de données en ligne consacrée au suivi des animaux et créée en 2007 par le Max Planck Institute for animal behavior. Il s'agit d'un outil populaire, gratuit, et avec de nombreuses fonctionnalités, dont les grandes lignes sont décrites dans le tableau 8.

Tableau 8. Détail des fonctionnalités macro de Movebank

Fonctionnalité	Explication
« Create a study »	La création d'une étude est le **point de départ** de l'utilisation de la plateforme. Cette étape permet d'indiquer quelques **informations générales** sur l'étude, comme la personne responsable et celle à contacter, le lieu d'étude et ses objectifs. C'est également à ce niveau que la **confidentialité des données** peut être déterminée (accès exclusif ou public).
« Upload data » / « Live data feeds »	Une fois l'étude créée, l'utilisateur peut **ajouter des données**. Ceci peut être réalisé **à la main** (« Upload data »), ou **automatiquement en direct** par l'intermédiaire de Movebank en fonction des fournisseurs de balises partenaires (« Live data feeds »). Il existe également un système de notification par e-mail contenant des statistiques sur les dernières données téléchargées.
« Deployment manager »	Permet de gérer toutes les informations concernant **l'attribution des balises** : quelle balise est attachée à quel individu, caractéristiques de l'animal, caractéristiques de la balise, date de déploiement… Cette étape est notamment importante si une balise sert plusieurs fois pour différents individus, ou pour restreindre les données stockées aux données de l'animal en liberté par exemple (sans inclure les phases de test ou de voyage de la balise par exemple).
« Event editor »	Permet de voir le jeu de données à la fois sous-forme de tableau et de carte afin de **modifier les attributs associés aux données**. Cette étape peut être réalisée **à la main**, par exemple pour définir un comportement, un état de migration, ou une donnée aberrante. Des **filtres** permettent également de signaler les doublons et données aberrantes en fonction de divers paramètres. Les modifications à la main restent cependant prioritaires.
« Env data »	Cette fonctionnalité permet de **lier les données de mouvements à une multitude de paramètres environnementaux** tels que les vents, l'utilisation des terres, la couverture végétale, la quantité de neige, la topographie… La NASA, l'ESA (European Space Agency), ou encore la NOAA font partie des fournisseurs de ces données.
« Software »	Movebank met à disposition une liste d'outils permettant d'analyser les données.

Movebank est donc un outil complet et gratuit, permettant de faire toute la chaine de la récupération des données à leur analyse. Cependant, une exploitation optimale de la plateforme demande du temps d'exploration et de compréhension des fonctionnalités.

Parmi les outils répertoriés dans « Software », 11 (52%) sont des packages R. D'après (Joo *et al.*, 2019), il existe 59 packages R dédiés aux différentes étapes de traitement des données de suivis que sont la préparation, le nettoyage et l'analyse des données de position (cf figure 9 en

début de mémoire). Le tableau ci-dessous (tableau 9) permet d'entrevoir les possibilités conférées par ces outils.

Tableau 9. Etat de l'art des packages R développés pour les données de suivi. Source : (Joo et al.*, 2019)*

Etape	Traitement	Type de données	Packages
Préparation des données	Permet de transformer les données récupérées par la balise en données de position (x, y, t)	GLS – un enjeu important réside dans la diminution de l'erreur liée à cette technologie, expliquant le développement de plusieurs packages.	GeoLight, probGLS FlightR, trackit, TripEstimation/SGAT, TwilightFree
		Radio	telemetr
		Accélérométrie et magnétométrie	animalTrack, TrackReconstruction
Nettoyage des données	Filtrer les données impossibles	PTT	Argosfilter, SDLfilter
		GPS	SDLfilter
	Retirer les doublons/ filtre sur la vitesse	Tous	T-LoCoH, TrajDataMining, trip
	Compression des données	Tous	adehabitatLT, trajectories, trajr, amt, TrajDataMining, rsMove
	Nouvelles variables de 2nd ou 3ème ordre (ex : angles, distances)	Tous	adehabitatLT, amt, momentuHMM, move, moveHMM, rhr, segclust2d, trajectories, trajr, trip
		Radio	feedr
Informations tirées des données	Animations	Tous	anipaths, moveVis
	Statistiques de déplacements	Tous	amt, movementAnalysis, trajr
	Reconstitution d'un trajet	GLS	HMMoce, kftrack, ukfsst/kfsst
		PTT	argosTrack, bsam
		Tous	Crawl, ctmcmove, ctmm
		GPS	BayesianAnimalTracker, TrackReconstruction
	Identification de schémas de comportements	Tous	EMbC, m2b, adehabitatLT, segclust2d, bcpa, marcher, migrateR, lsmnsd, momentuHMM, moveHMM
		PTT	Bsam
	Utilisation de l'espace	Tous	adehabitatHR, amt, move, rhr, BBMM, ctmm, mkde, movementAnalysis, T-LoCoH, adehabitatHS,

			hab, ctmcmove, moveNT, recurse, rsMove
		Radio	feedr
	Simulation de trajectoire	Tous	Crawl, ctmm, momentuHMM, moveHMM, smam, adehabitatLT, moveNT, SiMRiv, trajr
		PTT	argosTrack, bsam
	Interactions entre individus	Tous	wildlifeDI, movementAnalysis, TrajDataMining

R offre un panel étendu de possibilités pour toutes les étapes de la chaine de traitements des données de suivi des animaux. Cependant, comme on peut le remarquer dans le tableau ci-dessus (tableau 10), il existe souvent plus d'un package pour un type de traitement. Ainsi, l'utilisateur peut très vite se sentir submergé d'informations sans savoir faire un choix sur lequel utiliser. De plus, il est nécessaire de maitriser le logiciel et le(s) package(s), ce qui peut demander du temps et/ou de l'argent s'il faut embaucher quelqu'un de compétent.

Après avoir étudié les packages R répertoriés sur Movebank, 30% des outils restant sont des extensions pour ArcGIS/ArcView. En effet, les SIG permettent également de réaliser des traitements et analyses sur des données de mouvements. Quelques plugins existent sur les logiciels ESRI®, dont la licence est payante, et sur QGIS. Une fois les plugins trouvés, leur utilisation demande une certaine expérience de la pratique des SIG.

Il existe encore d'autres programmes en langages plus anecdotiques comme Java, MATLAB, SAS, SciLab, ou prenant la forme d'applications mobile ou web. De nombreux logiciels sont également développés en interne dans les organisations, voire dans des équipes de recherche, et donc difficiles à trouver, non accessibles ou trop spécifiques aux cas pour lesquels ils ont été développés (Annexe 4).

Nous connaissons maintenant les besoins des utilisateurs et les outils existants. Ainsi, nous avons toutes les cartes en main pour décrire l'outil idéal imaginé.

A. Analyse Fonctionnelle de l'outil idéal

Si on rassemble les idées récoltées dans le sondage ainsi que les défauts des outils existants, l'outil à développer doit permettre une prise en main simple et rapide, accessible à toute personne souhaitant prétraiter des données.

Le tableau ci-dessous (tableau 10) décrit les différentes fonctionnalités imaginées pour l'outil à développer.

Tableau 10. Tableau récapitulatif des fonctionnalités de l'outil à développer.

Type de fonctionnalité	Fonctionnalité	Description	Priorité
Import	Import de données par fichier csv	Les données doivent pouvoir être importées de manière aisée.	1
	Import de données via Movebank	Beaucoup des utilisateurs potentiels utilisent déjà Movebank, voire reçoivent leurs données directement sur cette plateforme. Ainsi, ce type d'import leur évitera une manipulation de téléchargement du jeu de données.	3
Pré-traitements	Elimination des doublons	Les doublons sont ici définis comme étant au moins deux données ayant le même timestamp.	1
	Elimination des points impossibles	Les points impossibles comprennent les cas suivants : données manquantes ou nulles, coordonnées impossibles (|latitude| > 90° ou |longitude| > 180°, ou positionnées dans un habitat impossible pour l'espèce. Ex : bouquetin dans l'océan), timestamp dans le futur.	1
	Elimination des points aberrants	Si la vitesse entre 2 points est supérieure à la vitesse maximale pouvant être atteinte par l'espèce étudiée.	1
	Elimination des données d'immobilité	Détection d'une immobilité et élimination des points correspondants.	2
	Paramétrage avant import	Configuration des paramètres correspondant au jeu de données importé afin de pouvoir réaliser les pré-traitements appropriés.	1
	Paramétrage dynamique	Permet de régler les paramètres une fois l'importation effectuée et avec une visualisation de leurs effets, afin d'avoir un pré-traitement modulable.	2
Informations sur l'animal	Annotation de comportements via apprentissage	Utilisation d'algorithmes d'apprentissage afin de détecter des transitions de comportements, voire un comportement précis.	3
	Visualisation 3D	Permet de visualiser les données sur une carte en prenant en compte le relief.	1
Export	Export	L'utilisateur peut télécharger son jeu de données nettoyé et annoté au format csv.	1

Ainsi, la première utilité de l'outil serait de nettoyer un jeu de données de mouvements.

Au cours d'un brainstorming avec une partie de l'équipe de Natural Solutions, le nom **Animal Mouvement Cleaner Application,** avec **AMCA** comme acronyme**,** a donc été choisi.

Comme nous avons pu le voir via l'analyse des outils existants, beaucoup sont totalement gratuits, certains nécessitant juste une licence payante (ArcGIS). Ainsi, il parait difficile de réaliser un outil voué à la vente. AMCA sera donc un outil open source et gratuit.

L'outil optimal imaginé utilise un certain nombre de langages et technologies nouvelles et demandant donc un temps de formation et d'adaptation. De plus, la généralisation étant compliquée à mettre en place en raison de la diversité des espèces étudiées et de l'absence de standards, plusieurs cas doivent donc être traités afin de couvrir les besoins d'un maximum d'études. Ceci n'a pas pu être réalisé dans les six mois de stage. Des choix ont donc été effectués pour aboutir à un outil fonctionnel même si incomplet.

III. Le développement d'un « Minimum Viable Product » dans le cadre d'un stage de six mois

A. Des choix

1. Espèces et données utilisées

a. La récupération des données

Il a été choisi d'utiliser deux types d'espèces, au moins un mammifère terrestre et un oiseau, afin de se rendre compte des problématiques liées à chacun. Le choix de l'espèce a été influencé par les données mises à disposition. En effet, au cours du sondage certaines personnes ont laissé leur adresse mail tout en répondant qu'elles seraient en capacité de fournir des données si nécessaire. Ces personnes ont été recontactées et certaines ont envoyé des jeux de données brutes. Cela a été le cas de Jérôme Cavailhes, chargé de mission faune au Parc National de la Vanoise, de Vladimir Dobrev et Volen Arkumarev, chargés de conservation à la BSPB (Bulgarian society for the protection of birds). Les données mises à dispositions correspondent aux types d'espèces les plus représentées lors du sondage, des rapaces et un ongulé sauvage. Elles sont décrites dans la figure 19.

Figure 19. Informations sur les jeux de données mis à disposition par Jérôme Cavailhes (bouquetin des Alpes), Vladimir Dobrev (Vautour percnoptère) et Volen Arkumarev (Vautour fauve). Crédit photo : © P. Saulay- Parc National des Ecrins (bouquetin), © D. Gradinarov (Vautour percnoptère) et camera d'un piège photo de Rewilding Rhodopes (Vautour fauve).

Les jeux de données mis à disposition par la BSPB viennent de cinq fournisseurs différents et ont tous été tirés de Movebank, où ils reçoivent leurs données. Ainsi, un effort de standardisation des jeux de données de la part de la plateforme a pu être constaté, malgré quelques petites différences inter-fournisseurs, liées aux types disparates de données récoltées. Cette différence est également visible à un degré supérieur sur les données du bouquetin qui ne sont pas passées par Movebank.

Les données des vautours percnoptères permettront de travailler sur la détection d'une immobilité grâce aux trois jeux de données d'individus morts. Quant aux vautours fauves, le nombre de nicheurs peut constituer un début de jeu de données pour de faire de l'apprentissage sur la détection de nids. De plus, chaque espèce apporte sa particularité en fonction de son écologie.

b. Le bouquetin des Alpes

Le bouquetin des Alpes (*Capra ibex*), est un mammifère de la famille des bovidés vivant dans les zones montagneuses de l'arc alpin. C'est un animal trapu muni de cornes dont la taille varie en fonction du sexe et de l'âge. En tant qu'espèce grégaire, le bouquetin vit en petits groupes bien distincts : les femelles et jeunes mâles d'un côté, les mâles plus âgés de l'autre. C'est seulement lors de la saison du rut que les deux entités se rejoignent. Cet herbivore se déplaçant peu au quotidien suit néanmoins un rythme de vie saisonnier (figure 20). Il passe l'hiver à basse altitude mais à la recherche de fortes pentes rocheuses où la neige tient moins bien et qu'il peut emprunter grâce à ses sabots d'une grande adhérence. Au printemps, il descend dans la vallée pour se nourrir des premières pousses avant de remonter passer l'été à haute altitude, attiré par la fraicheur et la qualité de la végétation près des crêtes. Ainsi, le bouquetin évolue dans les milieux ouverts allant de la vallée jusqu'à des altitudes pouvant atteindre 3 500m en fonction des massifs (Oncfs, 2019). Sa chasse, interdite depuis 1962, a été à l'origine d'un très fort déclin et d'une grande perte de diversité génétique, nécessitant des programmes de gestion de sa conservation. L'espèce est aujourd'hui considérée comme stable malgré des variations d'une population à une autre. (Parc national de la Vanoise, 01/09/2019).

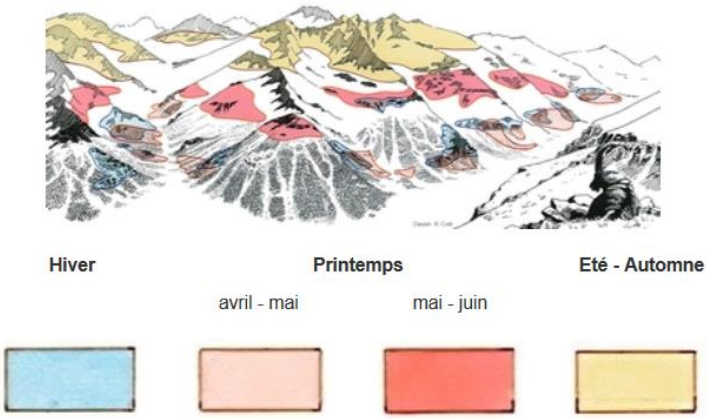


Figure 20. Migration saisonnière du bouquetin. Source : (Oncfs, 2019).

c. Le vautour percnoptère

Le vautour percnoptère (*Neophron percnopterus*), cet oiseau au plumage noir et blanc et à la tête jaune, est le plus petit vautour d'Europe. Son aire de répartition couvre le Sud de l'Europe, le Nord de l'Afrique et le Sud de l'Asie. Comme on peut le voir sur la figure 21, il s'agit d'une espèce migratrice, même s'il existe également des populations sédentaires dans le sud de sa zone de répartition. Les données fournies proviennent d'individus relâchés en Bulgarie et ayant donc connu au moins une migration automnale. Comme beaucoup de rapaces et de grands planeurs, les vautours percnoptères sont des migrateurs diurnes et empruntent le plus souvent des trajets migratoires évitant les grandes étendues d'eau en raison des faibles

courants ascendants s'y trouvant. Chaque printemps, les couples se retrouvent dans le même nid situé dans une cavité rocheuse de falaise abrupte pour y pondre entre 1 et 3 œufs. La taille de leur territoire dépend de la densité de la population, des ressources alimentaires et des gîtes disponibles. Il peut s'étendre jusqu'à 1000km^2 en Provence. Ce petit nécrophage au régime alimentaire varié est aujourd'hui, et ce depuis 2007, sur la liste rouge des espèces en danger d'extinction de l'UICN. Les raisons de son déclin sont multiples et inclues la chasse, les empoisonnements et les collisions avec les infrastructures linéaires. (LPO, 19/08/2019). C'est dans ce contexte que des individus sont balisés et suivis dans le cadre de projets notamment financés par le programme Life de l'Union Européenne (*Egyptian Vulture New LIFE*, 19/08/2019).

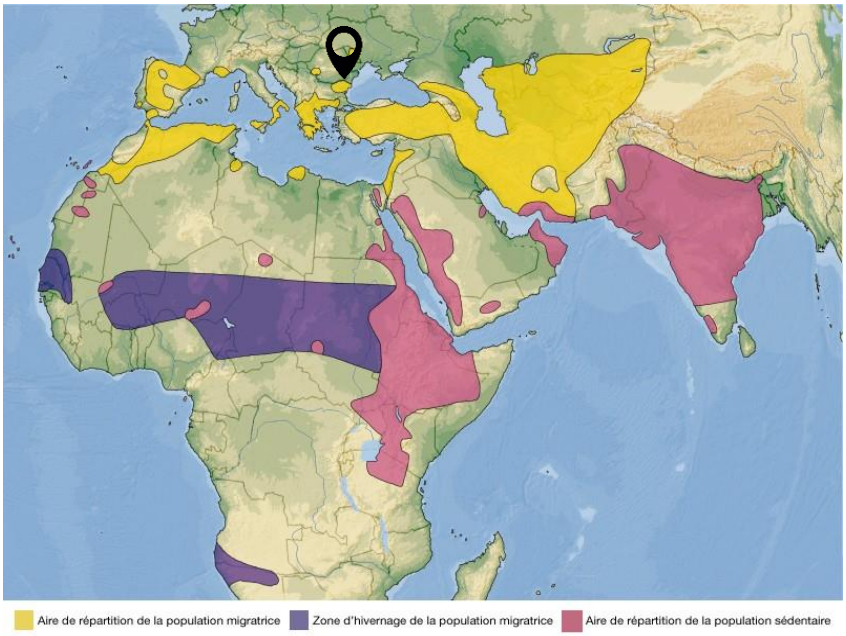


Figure 21. Répartition des populations de vautour percnoptère migratrices et sédentaires. Le repère indique la Bulgarie, lieu d'équipement et de relâché des individus mis à disposition par la BSPB. Source : (LPO, 19/08/2019)

d. Le vautour fauve

Avec plus de 2 mètres d'envergure, le vautour fauve (*Gyps fulvus*) est l'un des plus grands rapaces de France. Son aire de répartition est large, comprenant l'Europe, l'Asie et l'Afrique. Ce grand charognard niche et vit en colonies comptant entre 2 et 100 couples, sur des falaises situées entre 200 et 1600m d'altitude. A l'automne, les jeunes de l'année se dispersent pour une migration de 3-4 ans (en Espagne ou en Afrique pour les vautours français) avant de revenir s'installer dans une colonie, souvent celle d'origine. Les adultes sont sédentaires. Lors de la recherche de nourriture, les groupes s'envolent des falaises grâce aux courants thermiques ascendants et partent prospecter une zone pouvant couvrir plusieurs centaines de milliers d'hectares. Le vautour fauve a subi de forts déclins locaux en raison de manques de nourriture, d'empoisonnements et d'électrocutions, mais n'est aujourd'hui pas considéré en danger notamment grâce à la mise en place de gros dispositifs de nourrissage (LPO, 19/08/2019; Parc national des Pyrénées, 19/08/2019).

Ainsi, les trois espèces mises à disposition donnent accès à trois modes de déplacements différents : migratoire sur de longues distances pour le vautour percnoptère, migratoire sur de courtes distances pour le bouquetin des Alpes, et un sédentaire avec un territoire de prospection étendu.

2. Les études et le choix d'un suivi GPS

Tous les jeux de données fournis sont issus de la technologie de géolocalisation GPS. Une description rapide des projets à l'origine de ces suivis apportera un éclairage sur les raisons derrière ce choix dans chacun des cas, les trois espèces ayant la morphologie requise pour ce type de suivi.

a. ALCOTRA LEMED IBEX, un projet de gestion conservatoire du bouquetin des alpes

Le suivi des bouquetins du Parc National de la Vanoise s'inscrit dans le projet « ALCOTRA LEMED IBEX ». Les objectifs consistent à aboutir à une meilleure connaissance des déplacements de cette espèce, notamment afin d'estimer le domaine vital et d'identifier les menaces potentielles en fonction des interactions avec les activités humaines et le changement climatique. Une veille sanitaire et une étude de la mortalité sont également menées dans le but de mieux comprendre la démographie (Parc national de la Vanoise, 01/09/2019). Ce contexte, en particulier celui de l'étude des interactions avec les activités humaines, nécessite la précision fine des positions fournies par le suivi GPS.

b. Le projet New Life pour le vautour percnoptère

Comme son nom l'indique, le projet dont les données test sont issues est un projet LIFE financé par l'Union européenne. Cette information est un premier argument en faveur d'un suivi GPS : le financement. Il s'agit en effet d'une organisation de grande ampleur impliquant 14 pays. Son but est de sécuriser la route migratoire et de rétablir la population reproductrice des Balkans. Une sécurisation implique donc un repérage des zones à risques, aires pouvant être de taille restreinte…et donc un besoin de données précises. Un dernier avantage est ici la transmission quasiment en temps réel grâce à l'utilisation du réseau GSM. Ceci rend possible un suivi personnalisé et des interventions en cas de problème, comme cela a été le cas en 2018 avec un vautour juvénile passant trop de temps à proximité des habitations (BSPB, 06/09/2019).

c. Life Rewilding Vultures, l'avenir des grands vautours dans les Balkans

Le vautour fauve est l'une des deux espèces cibles du projet Life Rewilding vultures avec le vautour moine (*Aegypius monachus*). Une fois encore, le but est de reconstituer la population sauvage en étudiant notamment la dispersion de l'espèce, les menaces et en prodiguant une alimentation sûre. Pour les mêmes raisons que le vautour percnoptère, la précision GPS est requise dans ce projet également financé par le programme LIFE (Vulture Conservation Foundation, 06/09/2019).

Ces différents exemples sont des applications classiques du suivi des mouvements des espèces pour les conserver. Toutes nécessitent la précision de la technologie GPS.

Nous avons à faire à trois espèces de montagnes, et savons que ce milieu peut être à l'origine d'erreurs de localisation. En plus de provoquer des effets multi-trajets, le relief entraine une visibilité des satellites amoindrie et favorise des géométries groupées aboutissant à des valeurs de GDOP élevées. Néanmoins, la géolocalisation GPS semble ici la plus adéquat.

Ainsi, ces jeux de données permettront d'approfondir le développement de l'outil pour la technologie GPS et de le tester.

B. Analyse technique

L'outil réalisé lors de ce stage et décrit dans ce mémoire est une application web. Ce choix a été fait pour diverses raisons, notamment la simplicité de la structure et de son développement en comparaison avec une application lourde, logiciel installé sur l'ordinateur de l'utilisateur. L'avantage d'une application lourde serait la possibilité d'accéder à des périphériques physiques (camera, lecteur CD…), ce qui n'est pas utile dans notre cas. De plus, l'application web est accessible dès lors qu'il y a un accès internet quand une application lourde peut nécessiter d'utiliser un ordinateur performant pour la faire tourner.

Une application web se constitue en deux parties distinctes appelées « Back-end » et « Front-end » décrites ci-après. Cette structure permet de rendre ce type d'outil performant en utilisant les capacités du serveur pour effectuer les calculs et du navigateur pour la visualisation.

1. Le « back-end »

Le back-end est la partie du code exécutée par le serveur, non-visible par l'utilisateur. Il réalise le travail avant d'envoyer le résultat au client (navigateur). Les langages utilisés pour coder cette partie sont appelés 'langages serveur'. Il en existe plusieurs, alors comment faire le choix ? Le tableau 11 en présente certains avec des critères permettant de les départager :

Tableau 11. Comparaison non exhaustive de langages serveurs dans le contexte du stage. Source : (Les Jeudis, 2018)

Langage	Communauté	Simplicité	Compétences personnelles avant le stage	Compétences internes à Natural Soluitons
php	+++	+++	++	++
python	++++	+++	++	+++
Java	++++	++	-	+
Ruby	++++	+++	-	+

Le langage python se distingue dans toutes les catégories. D'après le sondage de Stack Overflow, il était le langage avec le plus grand nombre de volontés d'apprentissage pour la troisième année consécutive en 2018 et est considéré comme « le langage de programmation majeur ayant la croissance la plus importante » (*Stack Overflow Developer Survey 2018*, 2018). De plus, il y a de nombreuses compétences sur ce langage en interne à Natural Solutions, ce qui permet un encadrement de qualité. Le framework Pyramid®, open-source et codé en python, a été choisi pour sa simplicité de prise en main et sa flexibilité. En effet, sa forme de base contient les fonctionnalités essentielles et permet de développer des applications web très simples. De plus, ces fonctionnalités sont extensibles et rendent également possible la création de projets plus complexes. Plusieurs projets de Natural Solutions utilisent ce framework, assurant un encadrement en interne.

De cette manière, la partie « back-end » (ou back) de l'application web sera configurée avec Pyramid et codée en python. Cette configuration définit notamment le port sur lequel le back pourra écouter. On peut considérer cette information comme l'identifiant de l'application, ici le 6543. Une application web fonctionne selon une architecture 'client-serveur', le client (navigateur) envoyant des requêtes au serveur qui les attend et y répond (Supinfo, 2016). Le code est donc constitué de manière à répondre en fonction du chemin appelé lors de la requête. Ces chemins sont définis dans un fichier nommé '*routes.py'* qui permet de faire le lien avec le code contenu dans des '*views'*. Un exemple permettra de mieux comprendre les interactions entre les différentes parties.

Le code (figure 22) se trouve dans le fichier 'routes' et effectue les actions suivantes :

- La fonction *config.scan* va lire le fichier 'testAJ'.
- La fonction *config.add_route* permet d'appeler une fonction associée à un chemin via une communication interne, ici nommée 'test'. A l'appel du chemin, ici '/', le chemin le plus simple correspondant au nom du serveur associé au port définit en configuration (***localhost :6543/)***, la fonction associée sera exécutée grâce à 'test'.

```
import os
from telemetry2 import views

def includeme(config):
    config.scan('.views.testAJ')
    config.add_route('test', '/')
```

Figure 22. Capture d'écran d'un fichier routes.py de test

Ainsi, en appelant ce chemin on exécute le code ci-contre (figure 23) qui se trouve dans le fichier 'testAJ' nommé précédemment. On retrouve ici 'test' dans l'argument *route_name* du décorateur *@view_config*. Un décorateur de fonction est une fonctionnalité de Python dont le rôle est d'ajouter un comportement à la fonction sur laquelle il est appliqué.(*Appréhendez les décorateurs*, 2019). Ici, il permet de contraindre l'exécution de la fonction my_test_view en lien avec l'appel de la communication nommée 'test'. Par conséquent, l'appel de ***localhost :6543/*** dans un navigateur renverra ici « HELLO WORLD !! ». Cette réponse peut ensuite être récupérée par la partie « front-end » pour l'interpréter et la styliser.

```
from pyramid.view import view_config
from pyramid.response import Response

@view_config(route_name='test')
def my_test_view(request):
    return Response('HELLO WORLD!!')
```

Figure 23. Capture d'écran du fichier testAJ.py

2. Le « front-end »

a. Qu'est-ce que le « front-end » ?

Le « front-end » ou « front » est la partie visible de l'application par l'utilisateur. Elle est codée à l'aide de trois langages, le HTML, le CSS et le JavaScript (JS), puis est traduite par le navigateur (Google Chrome® ou Mozilla Firefox® par exemple) pour donner ce que l'utilisateur voit sur son écran (figure 24). Le langage HTML est un langage à balises permettant de faire le squelette d'une page web. Le CSS lui va apporter la décoration à la page, il permet d'appliquer un style aux éléments HTML et de faire de la mise en page. Ensuite, il incombe au JavaScript d'animer le tout avec de la réactivité, des conditions. Par exemple, s'il y a un champ pour entrer un mot de passe dans une page : le champ permettant d'écrire est une balise HTML, placée à cet endroit et avec cette taille par du CSS. Si un mauvais mot de passe a été entré, une alerte vous le signalera grâce à du code en JavaScript. Le rendu visuel peut varier en fonction du navigateur, chacun interprétant ce code à sa manière (Les Jeudis, 2018).

Figure 24. Fonctionnement de la partie « front-end » d'une application web.

Il existe des framework permettant de simplifier et accélérer le codage. En effet, les framework possèdent des fonctions globales traduisant parfois des dizaines de lignes de code en une seule. Ces outils sont multiples pour des utilisations diverses. Certains, sont pour un seul langage, c'est le cas de VueJS qui est un framework JavaScript, d'autres en utilise plusieurs, comme Boostrap qui utilise les trois langages.

Quels framework et librairies choisir ?

Le tableau ci-dessous présente un aperçu non exhaustif des framework existants et de leur utilité (tableau 12).

Tableau 12. Aperçu non exhaustif des framework et bibliothèques existantes pour la partie 'front-end'. Source : (Matthieu Bousendorfer, 2013; Chahine, 2018)

Framework	Langage	Utilisation	Simplicité	Communauté et documentation	Encadrement possible à Natural Solutions
ANGULAR	JS	Construction d'interfaces utilisateur	+	++++	++
React	JS	Construction d'interfaces utilisateur	++	+++	+
Vue.js	JS	Construction d'interfaces utilisateur	+++	++++	+
CESIUM	JS	Permettre une visualisation 3D sur un globe gratuitement	++	++	-

	JS	Permettre une visualisation 3D sur un globe gratuitement. Basé sur Cesiumjs, plus simple mais moins complet.	+++	+	-
	html, CSS, JS	Mise en page	++++	+++	+++
	html, CSS, JS	Mise en page	++	++	+

Nous utiliserons ici trois de ces framework :

- VueJS pour la mise en place de la structure de la partie front. Il a été choisi pour sa facilité d'apprentissage et de mise en œuvre.
- CesiumJS pour la visualisation des données sur un globe en 3D.
- Bootstrap pour la mise en page.

Nous développerons par la suite l'utilisation de VueJS et de son rôle dans la structure de l'application.

b. Comment VueJS fonctionne-t-il ?

VueJS fonctionne avec des composants permettant de structurer le code (figure 25).

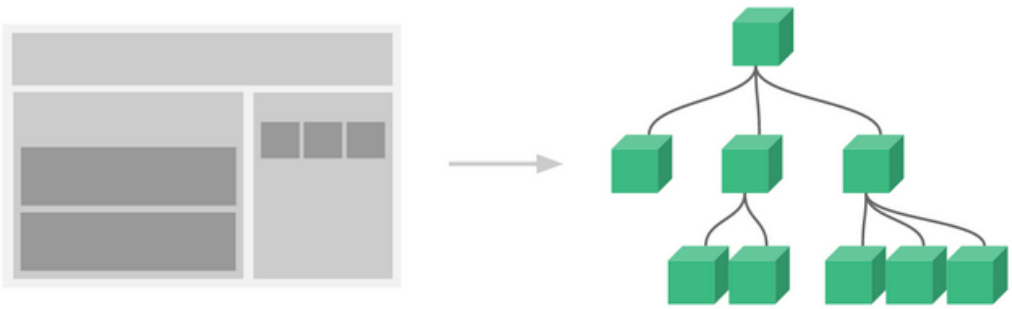


Figure 25. Schéma de la structure en composants de VueJS. Source : (VueJs, 29/08/2019)

Chaque composant est un fichier structuré en trois parties balisées rassemblant la totalité du code du front-end (figure 26) :

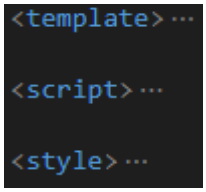


Figure 26. Structure du code avec VueJS

La partie *template* permet l'ajout d'éléments html.

La partie *script* accueil le code JavaScript de l'application.

La partie *style* gère le CSS.

Le lien entre les composants parents et enfants se fait via la partie *template* et *script* du composant parent (figures 27 et 28) :

```
<template>
  <div>
    <h1>Hello !!!!</h1>
  </div>
</template>


<script>
</script>


<style>
```

Figure 27.Composant enfant Hello.vue

La figure 27 est le composant le plus simple pouvant être réalisé : un seul élément html avec du texte. Ce code se trouve dans le fichier Hello.vue qui se trouve dans **./components/chap1**.

Dans le fichier parent App.vue on intègre le composant enfant Hello comme suit :

Via des balises dans *template*

En important le composant dans *script*

En remplissant la partie *components.*

```
<template>
  <div id="app">
    <img src="./assets/logo.png">
    <hello></hello>
  </div>
</template>

<script>
import hello from './components/chap1/Hello.vue'

  export default {
    name: 'app',
    components: {
      hello
    }
  }
</script>
```

Figure 28.Composant parent App.vue

Ce code génère alors le résultat suivant (figure 29) :

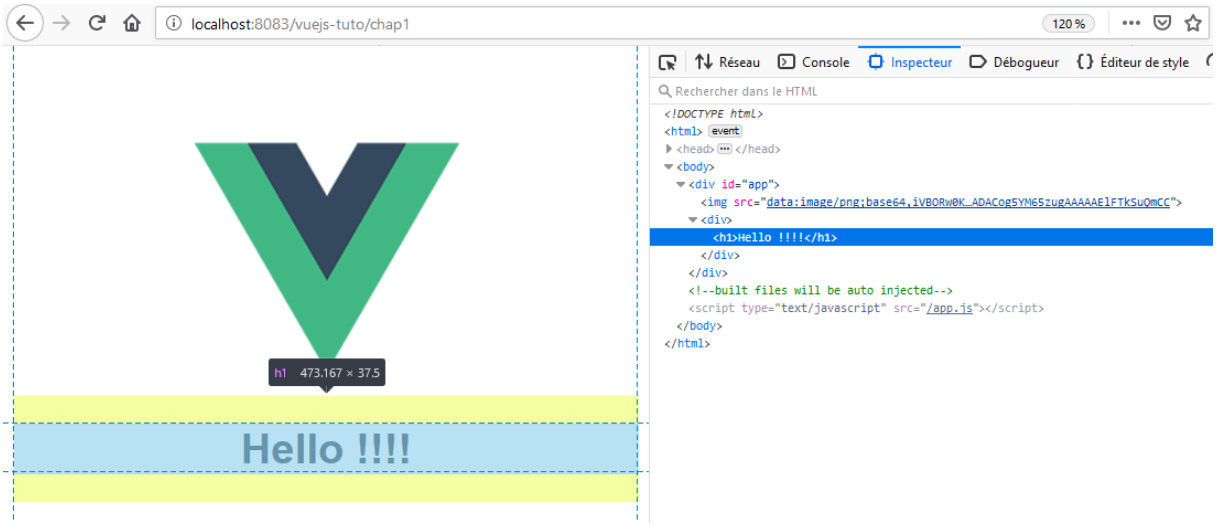


Figure 29. Capture d'écran du résultat de Hello.vue et App.vue

Les composants sont donc développés en remontant l'arborescence jusqu'au composant principal. A l'exécution ils seront chargés dans leur ordre d'apparition dans le composant principal, en finissant l'arborescence d'une branche avant de passer à la suivante

.

De cette manière, VueJS sera utilisé pour faire une application web monopage (AWM, single page application en anglais). Cette catégorie d'application web utilise les potentialités du langage JavaScript et permet de tout avoir sur une page en actualisant seulement les éléments qui ont été modifiés par l'utilisateur via le navigateur.

1. La structure finale de l'application

A ce niveau là nous avons des parties back-end et front-end. **Mais comment ces deux parties communiquent-elles ?**

La mise en place d'une API REST va être la clé de ces interactions. Une API, Application programming interface, est une interface permettant de communiquer des données. Il existe différents types d'API, dont les API REST (REST = Representational state transfer) qui sont basées sur le protocole http (Hypertext transfer protocol). Ce dernier utilise des requêtes pour établir une communication entre le client (front) et le serveur (back). C'est ce qu'on veut !

Un plugin dédié à l'interrogation d'une API REST (donc la demande d'un transfert de données), 'vue resource', a été développé pour VueJs (*Vue-resource*, 2015). Une fois le plugin installé dans le projet, il suffit de l'importer, `import VueResource from 'vue-resource'`, et de signaler à Vue que nous allons l'utiliser, `Vue.use(VueResource)`. Il ne reste alors plus qu'à configurer la communication :

1/ Quelle URL voulons-nous requêter/interroger ? Nous savons que notre partie back-end va générer les données prétraitées, et nous savons qu'elle écoute le port ***6543*** dans notre exemple (cf III.B.1)**.** On définit alors cette URL dans les balises *script* via le code suivant :

```
export default {
  http: {
    root : 'http://localhost:6543'

  }}
```

2/ Pour pouvoir effectuer des pré-traitements, il faut des données. On veut donc envoyer, **publier** une requête contenant les données au back afin qu'il les prétraite, ce qui sous-entend l'utilisation de la méthode *post*. Pour rappel, les scripts effectuant les calculs dans le back sont encapsulés dans des 'view' appelées par des 'route_name' définis grâce à Pyramid. Cette 'route-name' constituera donc l'adresse à laquelle transmettre la requête.

```
this.$http.post('route_name', objet à transmettre)
```

Dans notre cas l'objet sera le plus souvent un fichier csv.

3/ Le back reçoit alors les données, les traite, et les renvoie au front qui va attendre la réponse. Une fois récupérée, cette réponse entrainera des actions : ici l'écriture du message 'SUCCESS !!'

```
.then((response) => {
            console.log('SUCCESS!!')
          })
```

La figure 30 propose un schéma bilan de l'architecture globale de cette application web :

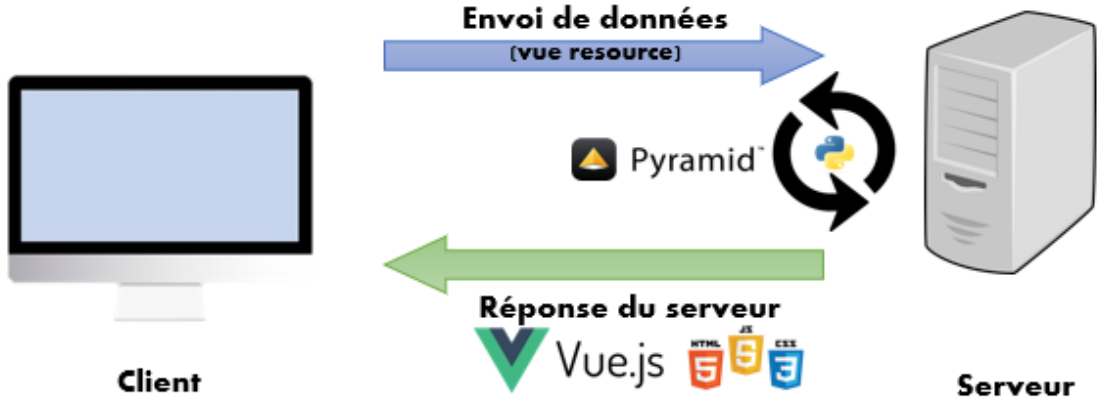


Figure 30. Schéma bilan de l'architecture de l'application web

Une fois la structure mise en place, il n'y a plus qu'à la remplir de fonctionnalités.

C. Les fonctionnalités développées

1. Le choix des fonctionnalités et leur place dans l'application

Au cours des 6 mois de stage, la priorité a été le développement des fonctionnalités d'ordre 1 (cf. II. C.) afin d'avoir une application fonctionnelle pour la faire tester en fin de stage.

La plupart des fonctionnalités décrites dans la figure ci-dessous (figure 31) sont reprises en détails dans la suite du mémoire.

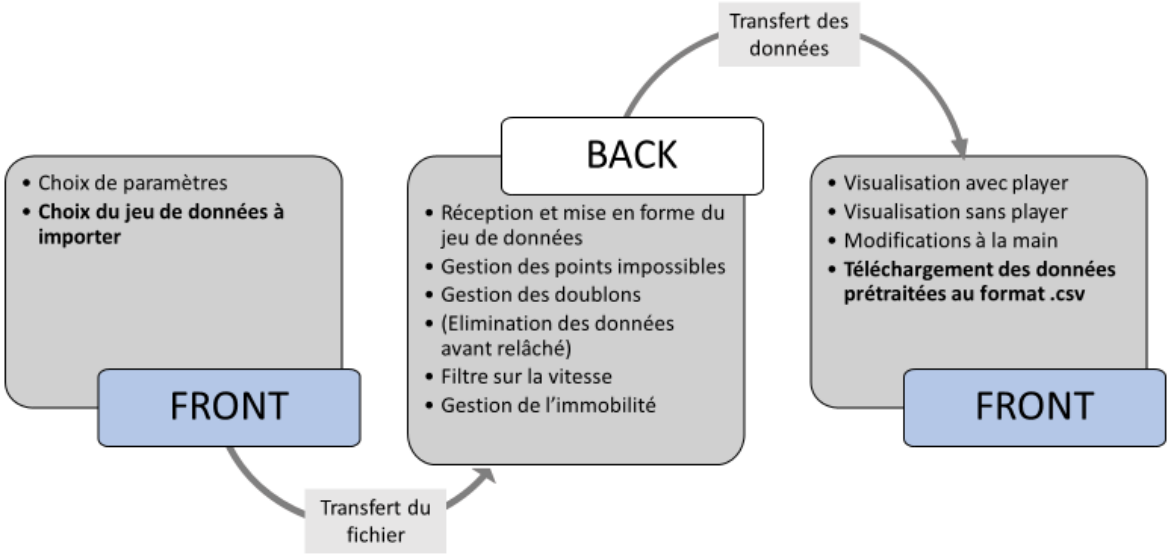


Figure 31. Schéma résumant les fonctionnalités développées et leur place dans le processus de l'application

L'application se présente comme un outil de transit de données. Il n'y a pas de processus d'identification et de compte utilisateur, les données ne sont pas stockées. L'utilisateur entre des données et les récupère en sortie.

2. La phase de pré-traitements

La phase de pré-traitements a été réalisée dans un premier temps pour la technologie GPS grâce aux données de test.

A l'exception de l'algorithme permettant de filtrer en fonction de la vitesse, qui a été inspiré d'une description d'un algorithme utilisé sur Movebank (Movebank, no date), les autres ont été entièrement imaginés au cours du stage.

a. Retirer les points impossibles

La gestion des points impossibles représente ici la problématique des doublons et des coordonnées aberrantes. A l'exception des doublons, ce sont généralement les données les plus faciles à repérées visuellement car très en marge du reste du jeu.

L'algorithme permettant de détecter et éliminer les doublons fonctionne en deux étapes. La première consiste à repérer les données possédant des timestamps (date et heure) identiques. Ensuite, ces données sont comparées entre elles sur le nombre d'informations. Ainsi, la donnée la plus complète est conservée. Si des données ont le même timestamp et le même niveau de détails, la première est conservée. Le code est détaillé dans l'annexe 5.

Les points impossibles sont actuellement détectés en fonction des valeurs possibles de coordonnées (|latitude| < 90° et |longitude| < 180°) et des données complémentaires (valeur avec 'GPS timeout'). Le paramètre du type d'espèce servant à limiter les zones d'occupation possible (continent pour un animal terrestre, mers et océans pour un animal marin…) n'est pas encore utilisé dans cet algorithme.

b. Filtrer sur la vitesse

La vitesse est un facteur permettant un niveau de filtre plus fin, retirant des points pouvant paraitre valides au milieu des autres mais qui ne sont physiquement pas possibles.

Le principe de cet algorithme est le suivant :

La vitesse est calculée entre le point n et n+1. Si cette vitesse est inférieure à la vitesse seuil entrée en paramètre, alors le point est ajouté à la collection de données filtrées. Le calcul suivant sera alors entre n+1 et n+2. Si la vitesse calculée est supérieure au seuil, le point est ajouté à la collection de points éliminés et annoté comme « speed outlier » dans la collection de données brutes. Dans ce cas, le calcul suivant sera entre n+1 et n+3. Le but est de trouver le premier bon point consécutif. La figure 32 ci-contre permet d'avoir une idée plus claire des possibilités. Le code est détaillé dans l'annexe 6.

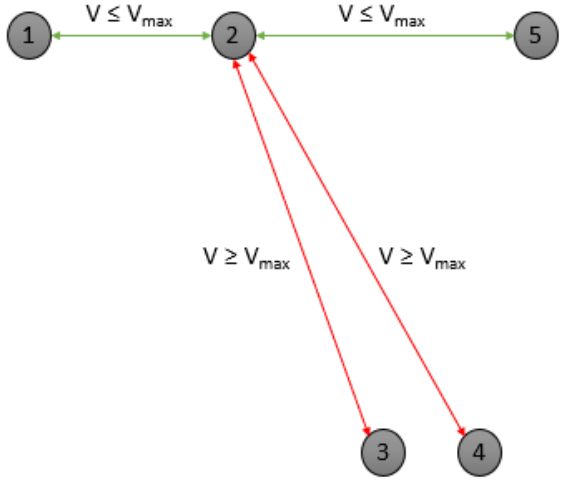


Figure 32. Fonctionnement de l'algorithme de filtre sur la vitesse. V= vitesse calculée, Vmax = vitesse seuil. Ici les points 3 et 4 sont donc éliminés et le point 5 devient le

En utilisant cet algorithme, on considère le premier point comme valide. Il est donc important de s'assurer de la véracité de cette hypothèse grâce à un paramètre d'entrée demandant la date de relâché de l'individu balisé afin de supprimer les points ne correspondant pas aux mouvements de l'animal en liberté. Il est également important de noter que le

paramétrage du seuil de vitesse dépend de la fréquence d'acquisition. En effet, un bouquetin peut atteindre une vitesse de pointe de 70km/h (*Biologie du Bouquetin*, 04/09/2019) mais ne peut la tenir une heure. Ainsi, si les positions GPS sont prises toutes les heures, le paramètre de vitesse maximale doit être inférieur à 70km/h et être de l'ordre de la vitesse moyenne de déplacement de l'espèce. On a pu remarquer que cette problématique était différente pour des espèces planeuses telles que les vautours qui utilisent et sont portées par les courants thermiques. Ainsi, leur vitesse dépend des conditions atmosphériques.

c. Détecter une immobilité

L'information d'immobilité est souvent utile afin de récupérer une balise perdue ou un cadavre. Une fois retrouvée, la balise peut être réutilisée sur un autre individu si elle est toujours en état de marche. Le cadavre est lui analysé afin de comprendre les origines du décès.

La démarche suivie pour détecter une immobilité repose sur le principe suivant : si un individu balisé meurt ou perd sa balise, celle-ci est alors immobile. Cependant, on sait que les technologies de géolocalisation ne sont pas toujours précises. De cette manière, un point immobile peut donner l'impression de se déplacer autour du vrai point d'une distance dépendante de la précision de la technologie. Cependant, ces déplacements resteront inférieurs à l'erreur associée à cette technologie. Ainsi, tous les points seront contenus dans un cercle dont le rayon correspond à l'erreur (on prendra par exemple une erreur surestimée de 50m pour le GPS).

La durée de l'immobilité détectée est un deuxième paramètre à prendre en compte. En effet, le but est de détecter une balise perdue ou un individu mort, or un animal peut rester statique pour diverses raisons comprenant le nourrissage, le repos, la couvaison… De cette manière, une immobilité sera considérée valide si elle dépasse un certain lapse de temps décidé par l'utilisateur, et appelé « seuil » dans la figure 33. Le code est détaillé dans l'annexe7.

Le schéma ci-dessous permet de comprendre le fonctionnement de l'algorithme en utilisant les principes cités précédemment (figure 33).

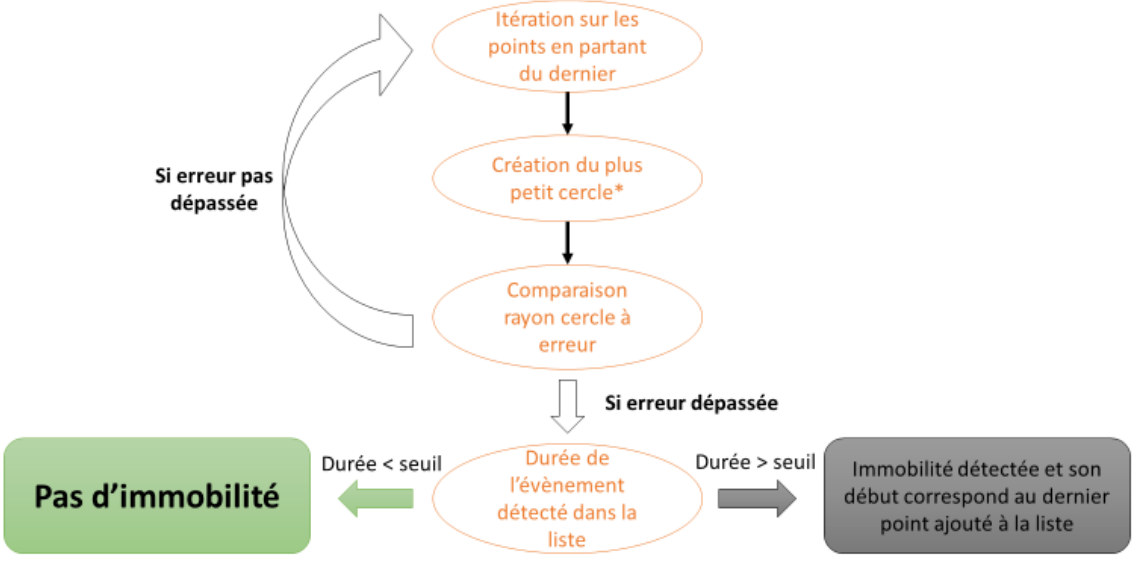


*Figure 33. Fonctionnement de l'algorithme de détection d'une immobilité. * Cette étape est réalisée grâce à la fonction make_circle prenant en paramètre une liste de coordonnées cartésiennes. Cette fonction vient de la librairie pyhton smallest encolsing circle. (Nayuki, 2018)*

De cette manière, le back s'organise en une suite d'algorithmes, représentée par la figure 34, permettant d'effectuer des pré-traitements sur les données de géolocalisation.

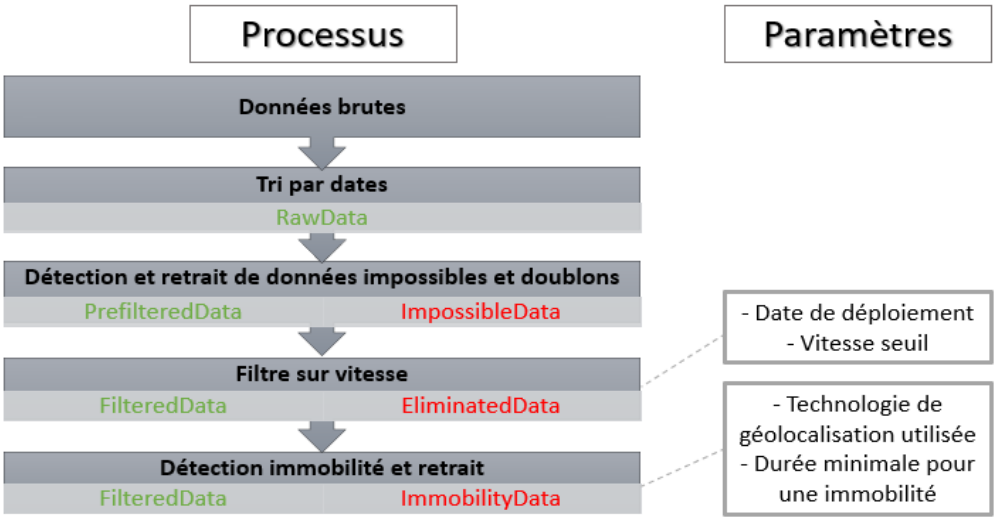


Figure 34. Déroulement du processus de pré-traitements dans la partie 'back-end' de l'application

Une colonne 'status' est ajoutée à la collection des données brutes et est complétée au fur et à mesure du déroulement du processus avec la raison d'élimination du point.

Ensuite, une fois que la partie back-end a fait son travail, chaque collection peut être visualisée sur un globe 3D.

3. La visualisation

Une fois que les pré-traitements automatiques ont été réalisés par le back, le front permet une visualisation des données présentes dans les collections. Cette visualisation est permise par le framework VueJS et la bibliothèque JavaScript CesiumJS. Une capture d'écran de l'application en annexe 8 permet de se rendre compte du rendu du front.

a. Modes de visualisation

Les données peuvent être visualisées avec deux modes différents, l'un affichant tous les points de la/les collection(s), permettant d'avoir une vue d'ensemble du trajet effectué (figure 35), l'autre utilisant une frise chronologique et affichant les points de la collection sélectionnée au cours du temps montrant l'enchainement des positions (figure 36).

Figure 35 Visualisation de tous les points d'une ou plusieurs collection(s).

Figure 36. Visualisation en mode player

b. Gestion de l'altitude

Les tests avec les données de bouquetin ont permis de mettre en évidence la problématique liée à l'altitude. En effet, la capture d'écran ci-dessous (figure 37) montre un point au-dessus du sol, ce qui paraît difficilement réalisable pour un bouquetin. Ceci vient du fait qu'une balise récupère une hauteur ellipsoïdale et non une altitude.

Figure 37. Capture d'écran de données brutes de bouquetin dans

La représentation 3D implique la présence et potentielle disponibilité de données d'un modèle numérique de terrain. De cette manière, l'idée a été de récupérer les données d'altitude associées à des coordonnées précises afin de représenter les mammifères au sol. Cependant, pour les oiseaux les données d'altitude en vol ne peuvent être corrigées. Ainsi, le paramètre définissant le type d'espèce permet également de corriger les altitudes.

c. Une fonctionnalité essentielle : la modification à la main

Une fois ce jeu de données prétraité par les algorithmes du back et du front, l'utilisateur a accès aux différentes collections de données correspondant aux différentes étapes du traitement, permettant d'en voir son évolution. **Mais si un point a été retiré alors que l'utilisateur le considère valide ou le contraire ?**

Lorsqu'on étudie des êtres vivants, tout n'est pas toujours prévisible et donc automatisable, le but de ces études étant de mieux connaitre des espèces. Dans ce contexte, il parait indispensable que l'utilisateur puisse avoir la main sur ses données et que ses modifications soient prises en compte dans le fichier téléchargé en sortie. Par conséquent, il est actuellement possible d'interagir avec les collections « Eliminated Data » et « Filtered Data » pour faire passer des points de l'une à l'autre en les sélectionnant au préalable sur la carte.

La capture d'écran ci-après (figure 38) indique les différentes variations possibles des points des collections « Eliminated Data » et « Filtered Data » :

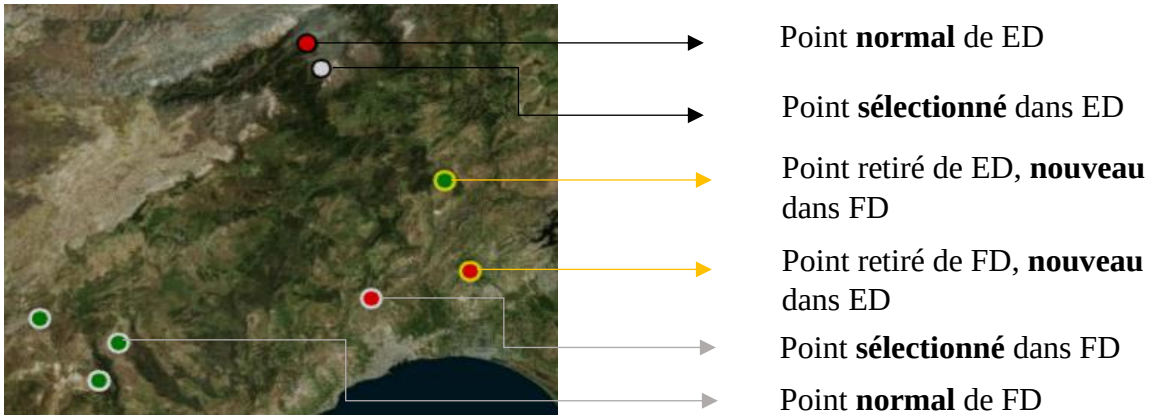


Figure 38. Variantes des points des collections Eliminated Data (ED) et Filtered Data (FD) au cours de la modification à la main

IV. Bilan et perspectives de développement

L'application est aujourd'hui disponible sur github accompagnée d'un guide d'installation (annexe 9).

Suite à une volonté personnelle et afin d'avoir un retour de potentiels utilisateurs, une procédure de test a été envoyée aux trois fournisseurs de données, Jérôme Cavailhes, Vladimir Dobrev et Volen Arkumarev. Pour ce faire, un serveur a été loué et configuré afin de mettre l'application en ligne. Elle est donc disponible en allant sur le lien suivant : vps718375.ovh.net. En plus des instructions pour y accéder, le mail contenait un guide d'utilisation (annexe 10) pour les aider à aborder l'application. Au moment de l'écriture de ce mémoire, aucun résultat de test n'est connu.

A. Bilan et analyse critique

L'outil réalisé est fonctionnel, il permet de prétraiter des données de bio-logging en retirant les données impossibles, en filtrant sur la vitesse et en détectant une immobilité. AMCA est également une application donnant la possibilité de visualiser des données dans un environnement en 3D, laissant à l'utilisateur le choix final des données à conserver. Elle a été conçue pour être paramétrable et facile d'accès sans nécessité de compétences particulières pour l'utilisateur. Il n'a besoin que de la connaissance de son jeu de données et de sa question d'étude pour vérifier la cohérence du résultat.

Cependant, certaines difficultés et imperfections persistent.

Globalement, les principales difficultés ont résidé dans l'utilisation de CesiumJS®. Cette bibliothèque contient un grand nombre de fonctions permettant de multiples interactions avec

le globe. Toutefois, la prise en main a été compliquée, avec une documentation contenant peu d'exemples et une communauté commençant juste à prendre de l'ampleur.

Un des problèmes résiduels concerne la **performance** en fonction de la quantité de données. En effet, lorsque le jeu de données dépasse 8000 données, l'application peut saturer. Le mode « player » peut également avoir des performances réduites ou des dysfonctionnements sur de gros jeux de données. Cette problématique est directement liée à la construction de la carte côté client (navigateur) sachant que Cesiumjs utilise déjà beaucoup de mémoire. Néanmoins, 8000 données représentent une quantité non négligeable, ni déraisonnable pour un déploiement. Un déploiement est ici considéré comme la durée de vie d'une balise sur un individu sans interruption. Comme vu dans la partie I de ce mémoire, la longévité d'une balise est soumise à plusieurs contraintes : le compromis entre la durée de la batterie et la fréquence d'acquisition des données, ainsi que la technique d'attache à l'individu (Hooker *et al.*, 2007). Ainsi, on peut considérer deux grands types de déploiements : longue-durée (maximum 5 ans) /faible-fréquence ou courte-durée/haute fréquence. Ainsi, 8000 données équivalent par exemple à un jeu de données sur environ une année avec une prise de position toutes les heures (24 x 365 = 8760), ce qui correspond à une fréquence dans la tranche moyenne à haute et une durée plutôt moyenne.

Une seconde source de difficulté a été de **rendre l'application applicable à tous les formats de données**. En effet, le milieu du bio-logging ne possède pas encore de standard, chaque fournisseur constituant son format de données de son côté. De cette manière, les données récurrentes peuvent avoir des noms différents (la date peut être appelée 'date' ou 'timestamp' par exemple) et ne pas avoir le même format. Ainsi, la phase de mise en forme de données avant d'y appliquer les algorithmes s'avère compliquée si on veut éviter de demander un paramètre additionnel. A l'heure actuelle, l'application est essentiellement capable de gérer un jeu de données téléchargé depuis Movebank® et un exemple de structure du format pris en charge est téléchargeable depuis l'interface.

Initialement, l'idée était également d'intégrer de l'apprentissage pour détecter des comportements. Cette aspect-là n'a pas pu être développé par manque de temps.

B. Les perspectives

Un grand nombre de fonctionnalités restent à être développées à toutes les niveaux de l'application.

1. Un lien à établir avec Movebank

Movebank s'est révélé être un outil central du milieu du bio-logging. Ainsi, l'intégrer à AMCA ne peut qu'être un atout. Il existe deux pistes à explorer.

La première est une fonctionnalité décrite dans l'analyse fonctionnelle (II.C) et consisterait à mettre en place un import de données via l'API de Movebank, ce qui est rendu possible par la documentation mise à disposition sur github : https://github.com/movebank/movebank-api-doc. L'intérêt de cette option serait double côté utilisateur :

- Ils auraient une manipulation de moins à réaliser dans le processus d'import : le téléchargement du fichier csv sur l'ordinateur.

- Movebank a mis en place un modèle de données standardisé, ce qui permettra sans doute une gestion plus simple de jeux de données provenant de différents fournisseurs, et ce en demandant un nombre réduit de paramètres à configurer pour l'utilisateur.

Un possible partenariat avec Movebank est également envisagé afin de faire apparaitre AMCA dans la liste des logiciels proposés dans la fonctionnalité « softwares ». Ceci donnerait une visibilité importante à l'application pour les utilisateurs, mais également pour les potentiels développeurs qui pourraient être intéressés pour y contribuer.

Ainsi, les utilisateurs de cette plateforme pourraient avoir accès à AMCA à travers un processus continu.

2. L'intégration de l'apprentissage

Les mouvements d'un animal sont la traduction spatio-temporelle de son comportement. Il peut bouger pour chercher de la nourriture, migrer, fuir un prédateur, chercher un partenaire… De cette manière, l'étude des mouvements peut dévoiler des comportements. La détection de comportements par apprentissage (ou machine learning en anglais) serait la troisième valeur ajoutée à AMCA en plus d'être un outil facile d'utilisation donnant accès à une visualisation 3D. Le but de cette technique est de sortir des informations d'un jeu de données. Outre la détection de comportements, nous pouvons également espérer augmenter la précision du trajet réel effectué par l'individu. La notion d'apprentissage regroupe plusieurs méthodes divisées en deux catégories : les algorithmes non-supervisés et supervisés. Le tableau 13 les décrit :

Tableau 13.Panorama non exhaustif des méthodes d'apprentissage pouvant être utilisées sur une étude des mouvements en bio-logging. Source : (Wang, 2019)

Type	Principe	Avantages	Inconvénients	Méthodes
Non-supervisé	Permet de révéler la structure d'un jeu de données non-annoté.	Pas besoin d'avoir des observations associées aux données de mouvements, généralisables à toutes les espèces	Moins précis, ne permet pas de distinguer des comportements de courte durée.	State space models, Hidden Markov Model
Utilisation des algorithmes non-supervisés	**Prédire la localisation suivante** en utilisant les différences entre les coordonnées consécutives et l'angle de rotation. On peut ainsi obtenir un **trajet plus précis** et une idée de **l'erreur de localisation à chaque point**. Ces données d'angles de rotation et de longueur de pas permettent également de **détecter une transition de comportement** (entre migration et recherche de nourriture par exemple).			
Supervisé	Utilise un jeu de données annoté pour entrainer l'algorithme avant de le faire tourner sur un jeu test qu'il doit annoter par lui-même.	Annote les mouvements avec des comportements précis issus des observations	Besoin de données difficiles à récupérer et espèces-dépendantes.	Forêt aléatoire, Machines à vecteurs de support (SVM), classification et arbre de régression (CART), Réseau de neurones artificiels

Utilisation des algorithmes supervisés	Surtout utilisés avec des données d'accéléromètres mesurant l'accélération statique (données de posture) et dynamique (données de locomotion) en continu sur des échantillons prédéfinis (10-20 sec) à une fréquence prédéfinie (toutes les 5-10 min). En associant ces données à des observations de comportements pour entrainer les algorithmes, ces derniers sont ensuite capables **d'affecter un comportement** à chaque échantillon.

Dans un premier temps il paraît plus adapté d'implémenter des méthodes non-supervisées pour aller plus loin dans la correction des trajets et débuter une analyse de comportement. De plus, l'application n'est actuellement pas capable de prendre en charge les jeux de données d'accélérométrie, car trop conséquents.

3. Des fonctionnalités demandées dans les résultats du sondage

Les résultats du sondage sont également une source importante d'idées de fonctionnalités à développer et répondant à des besoins. Elles n'ont pas été réalisées par manque de temps mais ont bien été prises en compte. Parmi ces suggestions on retrouve :

- Une compression des données ou rééchantillonnage avec un pas de temps paramétrable
- L'accès à des statistiques du jeu de données. Par exemple la distance parcourue sur le jeu de données ou la vitesse moyenne.
- Le calcul du domaine vital. Le domaine vital correspond à une zone sur laquelle un individu se nourrit et se reproduit. Il est de coutume de considérer que cette zone contient 95% des positions d'un animal (Kyoto University Field Informatics Research Group, 2012).
- Visualiser des jeux de données de plusieurs individus en même temps

Conclusion :

Que ce soit par simple curiosité, pour les chasser ou pour les conserver, suivre les animaux pour savoir où ils vont et ce qu'ils y font a été une question difficile à résoudre au cours de l'Histoire. Certains sont discrets et difficiles à suivre, d'autres se déplacent trop vite ou trop loin et deviennent alors hors de portée. Avant le bio-logging quelques méthodes existaient mais ne permettaient pas un suivi continu. Cette technique de télémétrie a alors permis de faire un grand pas en avant dans l'étude des espèces, de leurs déplacements et de leurs comportements. De plus, de nouveaux capteurs voient régulièrement le jour, multipliant les types de données récoltables. Ainsi, les animaux sont actuellement autant balisés pour leur étude que pour celle de leur environnement, ils deviennent des sentinelles de leurs milieux (Christophe Guinet, 2007).

La conséquence de ce suivi continu est l'engendrement d'un **grand nombre de données**, qu'il faut ensuite traiter. Le bio-logging a également ses défauts, caractéristiques sur lesquelles travaillent les ingénieurs afin de les amenuiser. De fait, on peut bien imaginer que la présence d'un corps étranger attaché à un animal peut influencer son comportement et ses interactions avec son environnement, ce qui pousse à la **prudence dans les analyses**. Une deuxième source de difficultés réside dans la possibilité **d'erreurs dans la localisation**. Ces erreurs dépendent de la technologie utilisée, la plus précise étant aujourd'hui la localisation GPS. Pour finir, l'absence de standards dans la structure des données représente une difficulté supplémentaire. Ces données sont générées par une grande variété de balises utilisant des technologies et capteurs variables, et conçues par de multiples fabricants. Ce manque de standard peut rendre le travail des données fastidieux et la conception d'outils standardisés difficile. Cette problématique est connue et actuellement traitée par un groupe de travail mis en place par la International Bio-Logging Society (IBLS) (Cagnacci *et al.*, 2017). En attendant le résultat de ce travail, les outils développées doivent pouvoir **prendre en charge toutes les structures de données**.

En d'autres termes, l'outil optimal de pré-traitements de ces données doit pouvoir en gérer un grand nombre, quelle que soit leur origine (technologies et espèces), en retirant au mieux les erreurs de localisation… Tout ça en le rendant simple d'utilisation et permettant de rester vigilant - notamment lorsque le sujet d'étude est un animal, donc peu prévisible. Animal Mouvement Cleaner Application remplit aujourd'hui la plupart de ces objectifs. Elle est pour l'heure adaptée à la technologie GPS et aux jeux de données issus de Movebank. Néanmoins, un élargissement est prévu dans le code conçu pour être paramétrable. Ses valeurs ajoutées sont sa simplicité d'utilisation et la possibilité d'une visualisation en 3D pour un véritable aperçu de l'environnement de l'individu.

Ce stage de fin d'études représente ma première expérience professionnelle dans une entreprise privée. Il a été l'occasion de découvrir ce milieu, son fonctionnement et ses difficultés. J'en retiens avant tout les échanges de savoir, ayant appris le développement web, la gestion de bases de données et la relation client, tout en faisant bénéficier de mes connaissances en écologie et en agronomie. J'ai également pu approfondir mes connaissances en suivi des mouvements des animaux, sujet qui me passionne et que j'avais commencé à découvrir sur le terrain au cours de mon stage de 2ème année d'école d'ingénieur à Bordeaux Sciences Agro. L'opportunité m'a été offerte de poursuivre le développement de AMCA et de la présenter à la conférence Biodiversity Next à Leiden, aux Pays-Bas, fin Octobre 2019.

Bibliographie :

Appréhendez les décorateurs (2019) *OpenClassrooms*. Available at: https://openclassrooms.com/fr/courses/235344-apprenez-a-programmer-en-python/233491-apprehendez-les-decorateurs (Accessed: 3 September 2019).

Argos système (2016) 'Argos users manual.pdf'. Available at: http://www.argos-system.org/wp-content/uploads/2016/08/r363_9_argos_users_manual-v1.6.6.pdf (Accessed: 12 July 2019).

Bächler, E. *et al.* (2010) 'Year-Round Tracking of Small Trans-Saharan Migrants Using Light-Level Geolocators', *PLoS ONE*. Edited by D. M. Evans, 5(3), p. e9566. doi: 10.1371/journal.pone.0009566.

Bairlein, F. (2008) 'The mysteries of bird migration – still much to be learnt', *British Birds*, p. 14.

Biologie du Bouquetin (no date). Available at: https://www.lacsdemontagne.fr/pages%20web/Faune/Bouquetin/bouquetin_description.htm (Accessed: 4 September 2019).

Bograd, S. *et al.* (2010) 'Biologging technologies: new tools for conservation. Introduction', *Endangered Species Research*, 10, pp. 1–7. doi: 10.3354/esr00269.

BSPB (no date) *How not to migrate, tells the Egyptian Vulture Vanya | News |*. Available at: http://bspb.org/en/news/Kak-ne-triabva-da-se-migrira-razkazva-egipetskia-leshoiad-Vanya.html (Accessed: 6 September 2019).

C. Douglas, D. *et al.* (2012) 'Moderating Argos location errors in animal tracking data', *Methods in Ecology and Evolution*, 3, pp. 999–1007. doi: 10.1111/j.2041-210X.2012.00245.x.

Cagnacci, F. *et al.* (2017) *A FUTURE FOR A COMMON BIO-LOGGING LANGUAGE?* Germany: International Bio-Logging Society, p. 21.

Chahine, H. (2018) 'ReactJS vs Angular vs VueJS : Que choisir en 2018 ? ⋆ Ambient Formations', *Ambient Formations*, 28 June. Available at: https://www.ambient-it.net/reactjs-vs-angular-vs-vuejs-que-choisir-en-2018/ (Accessed: 7 September 2019).

Christophe Guinet (2007) *interventions_christopheguinet.pdf*. CNRS, p. 3. Available at: http://www2.cnrs.fr/sites/communique/fichier/interventions_christopheguinet.pdf (Accessed: 1 September 2019).

CLS (no date a) *3.2 Principe de la localisation Argos*. Available at: http://www.argos-system.com/manuel/3-location/32_principle.htm (Accessed: 12 July 2019).

CLS (no date b) *Argos - Système mondial de suivi et d'étude par satellite dédié à l'environnement, Argos*. Available at: http://www.argos-system.org/fr/ (Accessed: 12 July 2019).

CLS (no date c) *Comment fonctionne argos, Argos*. Available at: http://www.argos-system.org/fr/argos/comment-fonctionne-argos/ (Accessed: 12 July 2019).

CNES (2017) *As-tu pris ton Galileo ?, jeunes*. Available at: https://jeunes.cnes.fr/fr/tu-pris-ton-galileo (Accessed: 6 September 2019).

Conservation nature (no date) *Arrêté fixant la liste des espèces de gibier dont la chasse est autorisée.* Available at: http://www.conservation-nature.fr/article3.php?id=103 (Accessed: 3 September 2019).

Egyptian Vulture New LIFE (no date). Available at: http://www.lifeneophron.eu/ (Accessed: 19 August 2019).

Escadrone (2018) *Guide du GNSS : RTK et PPK sur drones et résultats*, *Escadrone*. Available at: https://escadrone.com/guide-gnss-rtk-ppk/ (Accessed: 8 September 2019).

Evans, K., Lea, M.-A. and Patterson, T. (2012) 'Recent advances in bio-logging science: Technologies and methods for understanding animal behaviour and physiology and their environments Introduction', *Deep Sea Research Part II Topical Studies in Oceanography*, 88–89. doi: 10.1016/j.dsr2.2012.10.005.

Fudickar, A., Wikelski, M. and Partecke, J. (2011) 'Tracking migratory songbirds: Accuracy of light-level loggers (geolocators) in forest habitats', *Methods in Ecology and Evolution*, 3, pp. 47–52. doi: 10.1111/j.2041-210X.2011.00136.x.

GaëlleFehlmann and Andrew J.King (2016) 'Bio-logging', 26(18), pp. R830–R831. doi: https://doi.org/10.1016/j.cub.2016.05.033.

GIS Geography (2017) 'GPS Accuracy: HDOP, PDOP, GDOP, Multipath & the Atmosphere', *GIS Geography*, 13 March. Available at: https://gisgeography.com/gps-accuracy-hdop-pdop-gdop-multipath/ (Accessed: 23 July 2019).

Hooker, S. K. *et al.* (2007) 'Bio-logging science: Logging and relaying physical and biological data using animal-attached tags', *Deep Sea Research Part II: Topical Studies in Oceanography*, 54(3–4), pp. 177–182. doi: 10.1016/j.dsr2.2007.01.001.

Joo, R. *et al.* (2019) 'Navigating through the R packages for movement', *arXiv:1901.05935 [q-bio, stat]*. Available at: http://arxiv.org/abs/1901.05935 (Accessed: 16 August 2019).

Kyoto University Field Informatics Research Group (2012) 'Biologging', in *Introduction to field informatics*. Springer. Ishida, Toru, p. 174. Available at: http://www.ai.soc.i.kyoto-u.ac.jp/field_en/english_textbook/Biologging.pdf (Accessed: 1 September 2019).

Larousse, É. (no date) *Définitions : télémétrie*. Available at: https://www.larousse.fr/dictionnaires/francais/t%C3%A9l%C3%A9m%C3%A9trie/77097 (Accessed: 4 September 2019).

Laurentiu Rozylowicz *et al.* (2018) 'Empirical analysis and modelling of Argos Doppler location errors in Romania | bioRxiv'. doi: https://doi.org/10.1101/397364.

Les Jeudis (2018) *Développement front-end et back-end : Quelles différences ?*, *Les Jeudis - Blog d'actualité IT*. Available at: https://blog.lesjeudis.com/developpement-front-end-et-back-end-quelles-differences (Accessed: 28 August 2019).

Li, J. *et al.* (2015) 'Social Information Improves Location Prediction', in, p. 8.

Lotek (no date) *Products – Lotek Wireless*. Available at: https://www.lotek.com/products/ (Accessed: 7 September 2019).

LPO (no date a) *Présentation de l'espèce - Vautour percnoptère - LPO Rapaces*. Available at: http://rapaces.lpo.fr/vautour-percnoptere/presentation (Accessed: 19 August 2019).

LPO (no date b) *Vautour fauve - observatoire-rapaces.lpo.fr*. Available at: http://observatoire-rapaces.lpo.fr/index.php?m_id=20066 (Accessed: 19 August 2019).

Malgouyres, F. *et al.* (2017) *Etude du fonctionnement de la population du petit rhinolophe de la forêt de duesme (21) dans un objectif de gestion conservatoire*. ONF Réseau Mammifères, p. 84.

Matthieu Bousendorfer (2013) 'Foundation vs BootStrap, point de vue d'un webdesigner / intégrateur', *Captain Poulpe*, 30 October. Available at: https://blog.edenpulse.com/foundation-vs-bootstrap-point-de-vue-dun-webdesigner-integrateur/ (Accessed: 7 September 2019).

Meyburg, B.-U. and Meyburg, C. (2009) 'Satellite tracking of Birds'. Available at: http://www.raptor-research.de/pdfs/a_sp100p/a_sp139_en.pdf (Accessed: 8 July 2019).

Microwave Telemetry, Inc. (no date). Available at: https://www.microwavetelemetry.com/ (Accessed: 31 May 2019).

Migraction (no date). Available at: https://www.migraction.net/index.php?m_id=1517&bs=96 (Accessed: 3 September 2019).

Movebank (no date) *General purpose data filters | Movebank*. Available at: https://www.movebank.org/node/27252 (Accessed: 27 August 2019).

MTIs Coverage of the World's Bird Species 2018 (2018). Available at: www.microwavetelemetry.com/evolution_of_the_ptt (Accessed: 31 May 2019).

Oncfs (2019) *Oncfs - Le Bouquetin des Alpes*. Available at: http://www.oncfs.gouv.fr/Connaitre-les-especes-ru73/Le-Bouquetin-des-Alpes-ar1527 (Accessed: 18 August 2019).

Ornitela (no date) *Ornitela - Ornithology and Telemetry Applications, ornitela*. Available at: https://www.ornitela.com (Accessed: 10 August 2019).

Panzacchi, M., Van Moorter, B. and Strand, O. (2013) 'A road in the middle of one of the last wild reindeer migration routes in Norway: crossing behaviour and threats to conservation', *Rangifer*, pp. 15–26. doi: 10.7557/2.33.2.2521.

Parc national de la Vanoise (no date) *Bouquetins du Parc national de la Vanoise*. Available at: http://bouquetins.vanoise-parcnational.fr/ (Accessed: 1 September 2019).

Parc national des Pyrénées (no date) *Vautour fauve | Parc national des Pyrénées*. Available at: http://www.pyrenees-parcnational.fr/fr/des-connaissances/le-patrimoine-naturel/faune/vautour-fauve (Accessed: 19 August 2019).

Pathtrack (no date) *nanoFix GEO - Mini*. Available at: https://www.pathtrack.co.uk/products/nanofix-geo-mini.html (Accessed: 8 August 2019).

Ropert-Coudert, Y. *et al.* (2009) 'Diving into the world of biologging', *Endangered Species Research*, 10, pp. 21–27. doi: 10.3354/esr00188.

Ropert-Coudert, Y. and Wilson, R. P. (2004) 'Subjectivity in bio-logging science: do logged data mislead?', (58), pp. 23–33.

Rutz, C. and Hays, G. (2009) 'New frontiers in biologging science', *Biology letters*, 5, pp. 289–92. doi: 10.1098/rsbl.2009.0089.

Stack Overflow Developer Survey 2018 (2018) *Stack Overflow*. Available at: https://insights.stackoverflow.com/survey/2018/?utm_source=so-owned&utm_medium=social&utm_campaign=dev-survey-2018&utm_content=social-share (Accessed: 27 August 2019).

Supinfo (2016) *ARCHITECTURE CLIENT / SERVEUR | SUPINFO, École Supérieure d'Informatique*. Available at: https://www.supinfo.com/articles/single/2519-architecture-client-serveur (Accessed: 3 September 2019).

THIEBOT Jean-Baptiste (2011) *Déplacements et sélection d'habitat chez les animaux non contraints par la reproduction : une étude de l'écologie en mer des Manchots durant les phases d'immaturité et inter-nuptiale*. UNIVERSITÉ PIERRE ET MARIE CURIE. Available at: https://tel.archives-ouvertes.fr/tel-00660333/document (Accessed: 1 August 2019).

Thomas Robertson, B., D Holland, J. and Minot, E. (2012) 'Wildlife tracking technology options and cost considerations', *Wildlife Research*, 38, pp. 653–663. doi: 10.1071/WR10211.

Thums, M. *et al.* (2018) 'How Big Data Fast Tracked Human Mobility Research and the Lessons for Animal Movement Ecology', *Frontiers in Marine Science*, 5, p. 21. doi: 10.3389/fmars.2018.00021.

VueJs (no date) *Introduction — Vue.js*. Available at: https://fr.vuejs.org/v2/guide/index.html#Composer-avec-des-composants (Accessed: 29 August 2019).

Vue-resource (2015). Pagekit. Available at: https://github.com/pagekit/vue-resource (Accessed: 7 September 2019).

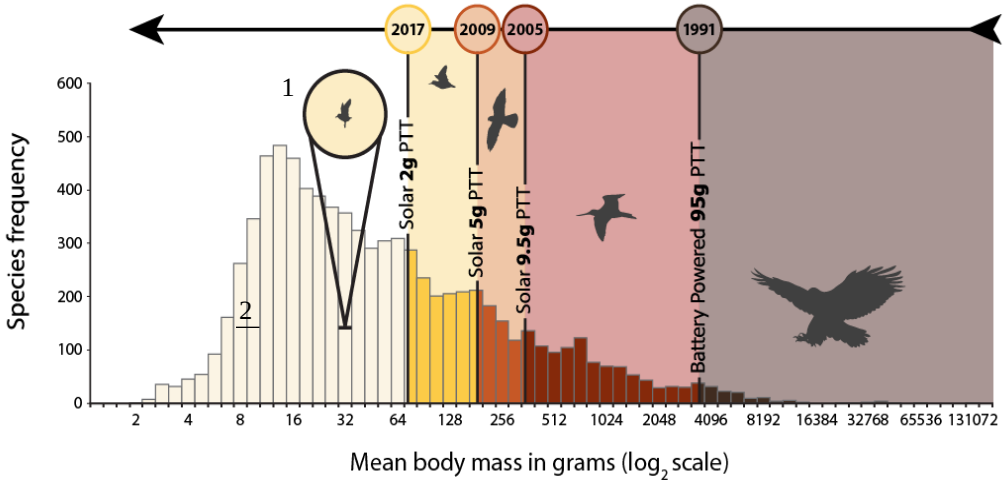
Vulture Conservation Foundation (no date) *LIFE Rewilding Vultures - Conservation of Black and Griffon vultures in the cross-border Rhodopes mountains, Vulture Conservation Foundation*. Available at: http://www.4vultures.org/life-projects/re-vultures/ (Accessed: 6 September 2019).

Wang, G. (2019) 'Machine learning for inferring animal behavior from location and movement data', *Ecological Informatics*, 49, pp. 69–76. doi: 10.1016/j.ecoinf.2018.12.002.

Whitworth, D. and FAO (eds) (2007) *Wild birds and avian influenza: an introduction to applied field research and disease sampling techniques*. Rome: Food and Agriculture Organization of the United Nations (FAO animal production and health manual, 5).

ANNEXES

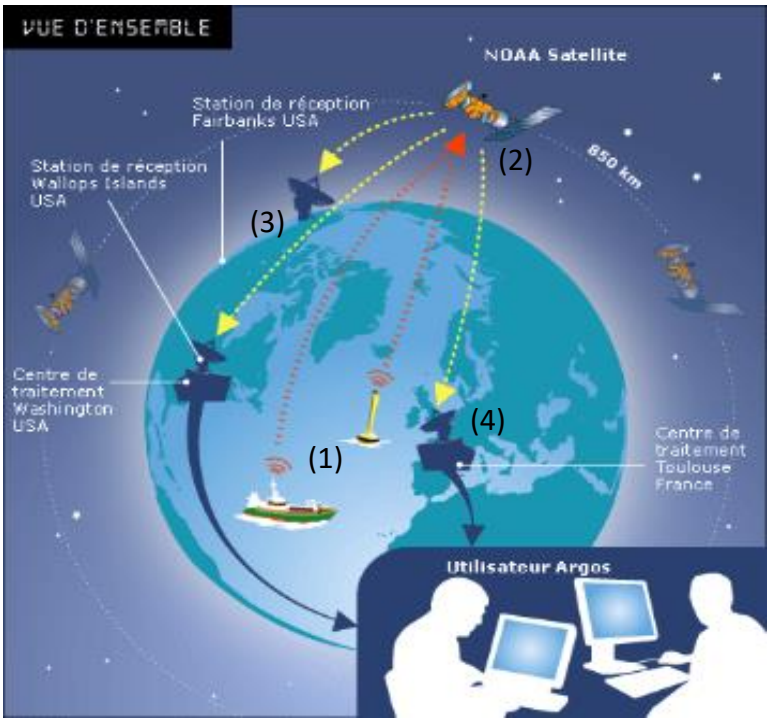
Annexe 1 : Distribution et évolution de la proportion des espèces d'oiseaux pouvant être suivies avec des balises Argos (35%). 1. Indique la limite de masse pour les balises GPS. 2. Indique la limite pour tout suivi en bio-logging. Source : (MTIs Coverage of the World's Bird Species 2018, 2018)



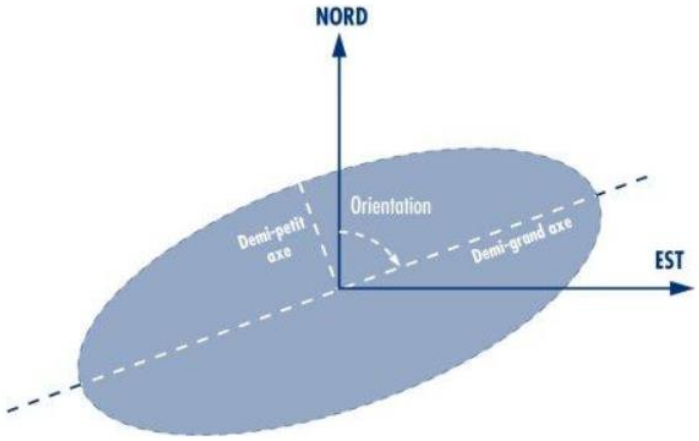
Annexe 2 : Schéma bilan du fonctionnement du système Argos.

1.Balises déployées, 2. Satellite, 3. Antenne, 4. Centre opérationnel.

Source : (CLS, 12/07/2019)



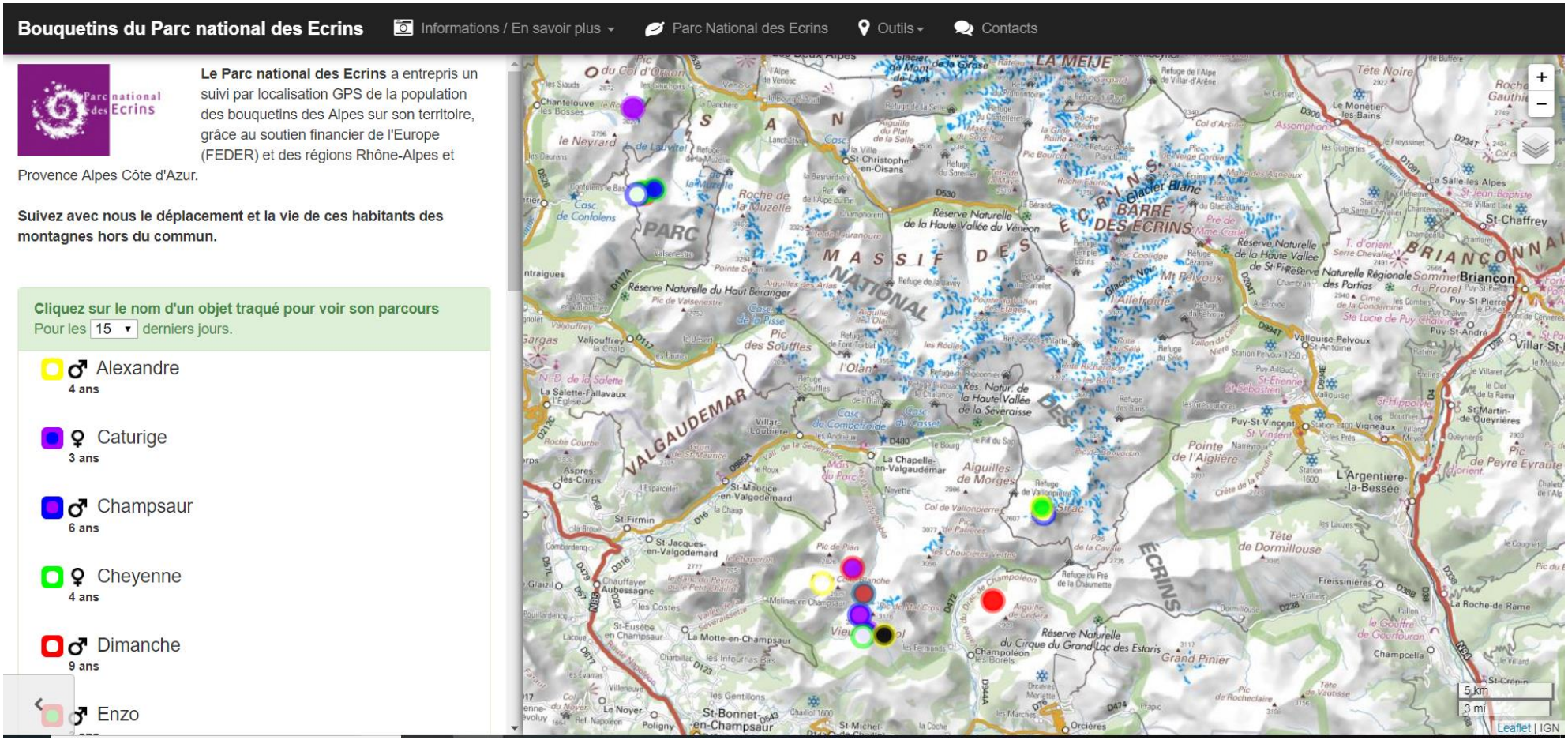
Annexe 3 : Ellipse d'erreur fournie par CLS (Argos système, 2016)



Annexe 4 : FollowDem, un outil développé par le Parc National des Ecrins.

L'application web est utilisée par plusieurs parcs nationaux de montagnes (Parc National des Ecrins, Parc National de la Vanoise) pour le suivi des bouquetins. Elle permet suivi visuel et une communication pédagogique sur l'espèce.

Pour celle du Parc des Ecrins : http://bouquetins.ecrins-parcnational.fr.



Annexe 5 : Code python de l'algorithme de gestion des doublons

```
def findDuplicates(candidateDf):
      # Dataframe rassemblant tous les doublons sur la date
      allDuplicatedDf = candidateDf[candidateDf.duplicated(['date'],keep=False)]
      # Récupération des dates pour lesquelles il y a des doublons
      listDateGroup = allDuplicatedDf['date'].unique().tolist()
      duplicatedRowsToDelete = None
      # Récupération des doublons d'une date donnée
      for date in listDateGroup:
            currentDf = allDuplicatedDf.loc[allDuplicatedDf['date']==date]
            # Dénombrement des colonnes vides pour chaque doublon
            currentDf['total'] = currentDf.isnull().sum(axis=1)
            currentDfOrdered = currentDf.sort_values(by='total',ascending=True)
            # Ajout à la collection des doublons sauf pour le premier qui est donc
            conservé car étant plus complet
            duplicatedRowsToDelete=pd.concat([currentDfOrdered[1:],duplicatedRowsToD
            elete])
      if duplicatedRowsToDelete is not None:
            duplicatedRowsToDelete = duplicatedRowsToDelete.drop(['total'] , axis=1)
      return duplicatedRowsToDelete
```

Annexe 6 : Code python de l'algorithme de filtre sur la vitesse

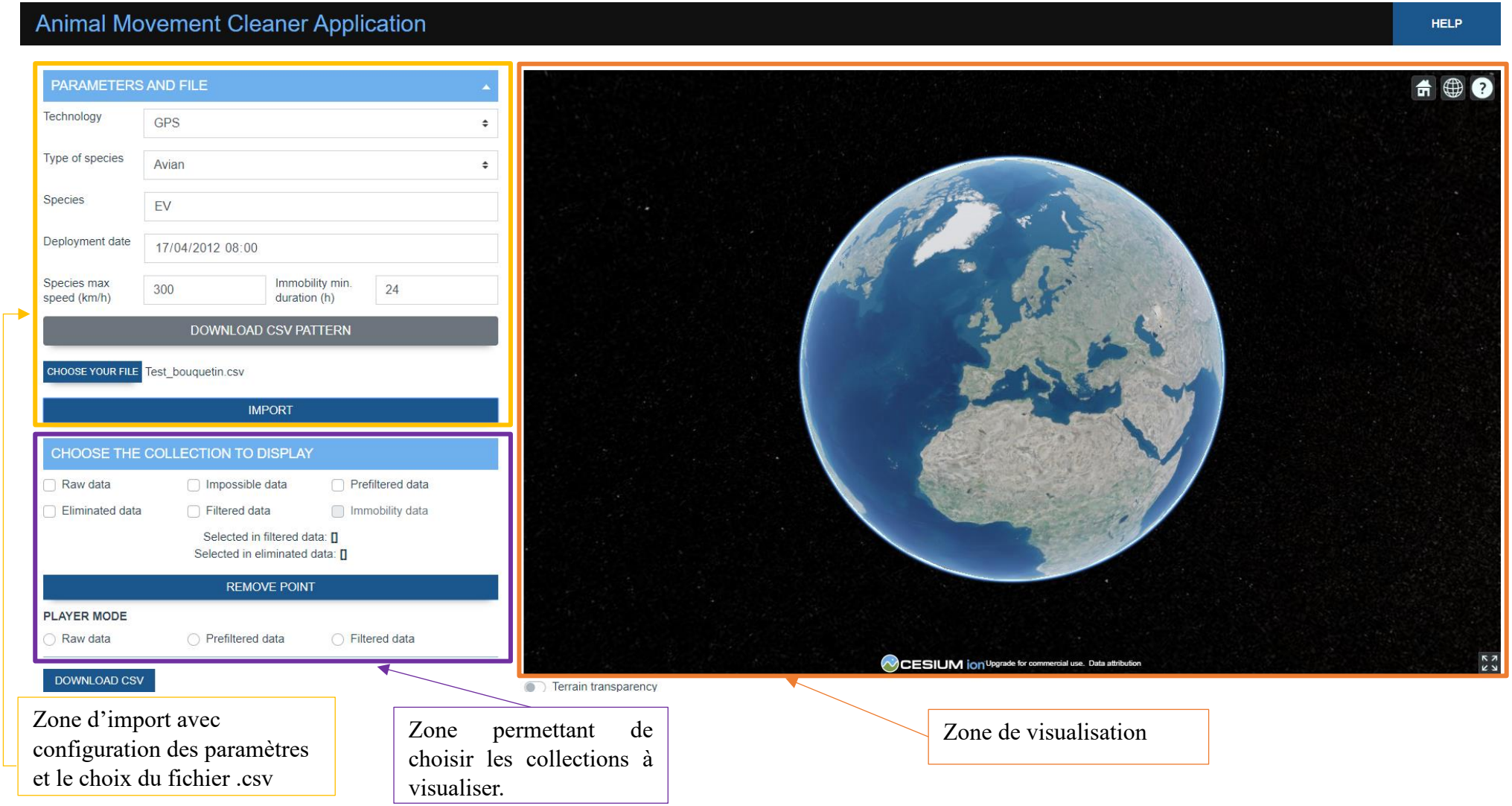
```
def Speed_algo(rawPointsAnnotated,points,MaxSpeed,deploymentDatestr):
      eliminatedSpeed =[]
      pointsfiltered = []
      deploymentDateobj = datetime.datetime.strptime(deploymentDatestr, '%Y-%m-
      %dT%H:%M')
      deploymentDateobj = deploymentDateobj.isoformat()
      L=len(points)
      start = 0
      alertDate = 0
      # Vérification que la date de déploiement soit bien avant la date de la
      dernière donnée
      if points[L-1]['date'] < deploymentDateobj:
            alertDate = 1
            return rawPointsAnnotated, eliminatedSpeed, pointsfiltered, alertDate
      # Recherche de l'indice de la donnée correspondant au déploiement pour
      commencer le filtre de vitesse à partir de cet indice
      if points[0]['date'] < deploymentDateobj:
            for d in range (L):
                  if points[d]['date'] >= deploymentDateobj:
                        start = d
                        break
                  else:
                        eliminatedSpeed.append(points[d])
                        for l in range (len(rawPointsAnnotated)):
                              if rawPointsAnnotated[l]['id'] == points[d]['id']:
                                    rawPointsAnnotated[l]['status']= 'before
                              deployment'
                        points[d]['distance1'] = 0
                        points[d]['speed'] = 0
      points[start]['distance1'] = 0
      points[start]['speed'] = 0
      i=start
      # Recherche des points valides sur la vitesse
      while i < L-1:
            for j in range (1,L-i):
                  # Calcul de la distance
                  points[i+j]['distance1'] =
                  vincenty((float(points[i]['LAT']),float(points[i]['LON'])),(float(
                  points[i+j]['LAT']),float(points[i+j]['LON'])))
                  # Calcul de la durée
                  diftimeS=datetime.datetime.strptime(points[i+j]['date'],'%Y-%m-
                  %dT%H:%M:%S') - datetime.datetime.strptime(points[i]['date'],'%Y-
                  %m-%dT%H:%M:%S')
                  diftimeH=diftimeS.total_seconds()/3600
                  # Calcul de la vitesse
                  speed=points[i+j]['distance1']/float(diftimeH)
                  points[i+j]['speed'] = speed
```

```
                # Comparaison à la vitesse maximale entrée en paramètre,
                # Si la vitesse est considérée aberrante on ajoute le point aux
                données éliminées et on l'annote dans les données brutes
          if speed > MaxSpeed:
                eliminatedSpeed.append(points[i+j])
                for l in range (len(rawPointsAnnotated)):
                      if rawPointsAnnotated[l]['id'] == points[i+j]['id']:
                            rawPointsAnnotated[l]['status']= 'speed outlier'
          else:
                i=i+j
                break
    # Elimination des points dont la vitesse a été jugée aberrante
    pointsfiltered = [x for x in points if x not in eliminatedSpeed]
    return rawPointsAnnotated, eliminatedSpeed, pointsfiltered, alertDate
```

Annexe 7 : Code python de l'algorithme d'immobilité

```
def Immobility_algo(rawPointsAnnotated,points,immo_range, immo_time):
      pointsAlive = []
      detected_immo = []
      L=len(points)
      points_for_circle = []
      K = len(points_for_circle)
      r=0
      # Boucle en partant de la fin
      for record in reversed(points):
      # Conversion des coordonnées en degrés en coordonnées cartésiennes
      x,y,zone,p= utm.from_latlon(float(record['LAT']),float(record['LON']))
      # Ajout du point à la liste permettant de faire le cercle minimum
      points_for_circle.append((x,y))
      # Création du cercle minimum
      cx,cy,r = make_circle(points_for_circle)
      detected_immo.append(record)
      # Vérification du rayon du cercle : s'il est supérieur à l'erreur de
      localisation,
      # alors le dernier point ajouté ne faisait pas partie de l'immobilité
      if r > immo_range :
      del points_for_circle[-1]
      del detected_immo[-1]
      break
      K = len(points_for_circle)
      # Calcul de la durée de la potentielle immobilité détectée
      diftimeS=datetime.datetime.strptime(points[L-1]['date'],'%Y-%m-%dT%H:%M:%S') -
      datetime.datetime.strptime(points[L-K]['date'],'%Y-%m-%dT%H:%M:%S')
      diftimeH=diftimeS.total_seconds()/3600
      # Comparaison de la durée calculée diftimeH à la durée minimale entrée en
      paramètre immo_time,
      # Si diftimeH >= immo_time, l'immobilité est validée
      if diftimeH >= immo_time:
      print('Immobility detected from',points[L-K]['date'])
      pointsAlive = [x for x in points if x not in detected_immo]
      # Annotation des points de l'immobilité dans la collection des données brutes
      for i in range (len(rawPointsAnnotated)):
      if rawPointsAnnotated[i]['date']>= points[L-K]['date']:
      rawPointsAnnotated[i]['status']='immobility'
      else:
      detected_immo = []
      pointsAlive = points
      print("Aucune immobilité n'a été détectée")
      return rawPointsAnnotated,detected_immo, pointsAlive
```

Annexe 8 : Capture d'écran de l'interface de l'application



Annexe 9 : Guide d'installation de l'application AMCA publié sur github

Prerequisites:

Nodejs 10.16.0 (https://nodejs.org/en/), Python 3.7.3 (https://www.python.org/downloads/)

Everything can be done from cmde

Installation:

1/ Create a directory for the application : here it will be named 'application'

2/ Go in your new directory and execute $ git clone https://github.com/AJambon/stageNS.git

Back installation:

3/Still in the same directory, execute $ pip install virtualenv

4/ Go in Back directory C:\application\stageNS\Back and

- create virtual environment with python $ -m venv virtualenvironmentname
- activate virtual environment with $.\virtualenvironmentname\Scripts\activate
- Add the path of the virtual environment folder in the gitignore file

5/ Go in the directory named telemetry2 with $ cd telemetry2

- Dependencies installation with $ python setup.py develop
- execute $ pserve development.ini

Front installation:

6/ Go in src directory with $ cd C:\application\stageNS\Front\stage_front\src

- Open the file config.js
- Replace the apiUrl by your server name

```javascript
export const myConfig = {
   apiUrl : 'http://localhost:6543'
}
```

7/ Go in Front directory with $ cd C:\application\stageNS\Front\stage_front

- Execute $ npm install
- Execute $ npm run dev

Then you can open your browser and enter http://localhost:8080

Annexe 10 : Guide d'utilisation

User guide/ Guide d'utilisation

The application aims at running pre-treatments on position data to remove impossible data, duplicates, speed outliers and immobility data. Please be acknowledged that it has been developed to be optimized on Google Chrome, other browsers can show errors or differences.

1. Enter the parameters corresponding to your dataset.
   They will permit to run the algorithms.
   For now, the parameters are:
   - the **geolocation technology** used: permits to get location error depending on the technology (for now only GPS and Argos). It is used in the algorithm to detect immobility.
   - **Deployment date**. That parameter is important for the speed algorithm which considers the first location to be valid, so corresponding to a position when the individual was free. Then if the dataset contains testing positions, they will be removed.
   - **Species**
   - **Species type**. Here there are 3 possibilities: terrestrial, avian or marine animal. That information is mainly used for the visualization functionality. Actually, the elevation data found in the datasets (if present), is either incorrect or not corresponding to the data included in the 3D viewer used, resulting in flying animals for ones not able to fly, or underground data. Here are the cases are considered:
     - Terrestrial: all the positions are clamped to the ground.
     - Avian: only positions with elevation inferior to the viewer elevation data are clamped to the ground. Superior elevation data are considered as 'flying' state and aren't touched, even if they can be incorrect.
     - Marine: same as terrestrial.

     Later, this parameter will also be used to detect outliers depending on the type of habitat: a terrestrial species cannot be found in the ocean and a marine species cannot be found outside the ocean.
   - **Species maximum speed**: used in the algorithm to detect outliers. When setting this parameter, please consider the frequency of your data. Indeed, most of the time the maximum speed to enter is the mean speed of a species. (for instance, some mammals can run up to 60km/h but can't keep that speed one hour long.)
   - **Duration to consider an immobility** is true: the tag is lost, or the individual is dead. By default, the algorithm is set to consider an immobility if its duration > 24h.

2. Import your data:
   - Verify if the columns used by the application have matching names (you will find the information by clicking on 'Download pattern' button which will download the typical csv architecture needed). You just have to change the names of the targeted columns, no need to remove the other ones. (This is not developed yet for the test,

so here are the matching columns' names: 'event-id', 'timestamp', 'location-lat', 'location-long')
- Choose a csv file on your computer
- Click on "Import" button. Please be aware that the bigger the file is, the longer the application will be to analyse it. For files with more than 8000 positions, the application can crash. Be patient while it is downloading, it can take up to 1 minute.

You will see a new section appearing that will permit to visualize your data on the map:

- If you check a checkbox in the first line of checkboxes, you will see all the points displayed on the globe.
- If you check a checkbox in the second line named 'Player mode', you will be able to activate the player mode and see points displaying depending on their timestamp.

Checkboxes and corresponding data found by checking them:

Raw data: in this collection you will find the data as you have given it, without any modifications except a new column named 'status' where you will find the reason why a point has been removed. If the point is kept, the cell will be empty[1].

Impossible data: this collection gathers the points detected as impossible because of localization or transmission problem (ex: all values equal to 0, not existing coordinates (|latitude| > 90° and/or |longitude| > 180°, date in the future, or unappropriated habitat (ex: ibex in the ocean)). Those data will be annotated as *'impossible'* in the raw data's status column.

Prefiltered data: result of "raw data" collection without "impossible data" and duplicates.

Eliminated data: data can be eliminated because out of species speed range or eliminated by hand from the map (see in 'remove by hand' section to see how to proceed), or because the timestamp is inferior than the one entered in the 'deployment date' parameter. Then, in the Raw data's status column you will find those data with the status *"Speed outlier*", "*before deployment*" or "*removed by hand*".

Immobility data: data detected as immobility. Found in Raw data's status column as '*immobility*'.

Filtered data: Result of "prefiltered data" collection without "Eliminated data" and "Immobility data" (if present).

NB: If you want to compare filtered data with other collections, choose 2D planisphere mode by clicking on the icon below. Otherwise points won't be superposed because of elevation correction on filtered data.

3. <u>Remove data by hand (not possible in the player mode for now):</u>

To go further in the cleaning of your data you can remove some points by hand or reintroduce eliminated points in the filtered collection. To do so you have to:

- Check "Filtered data" and/or "Eliminated data".
- Click on the points you want to remove.

    - If you select a point from “filtered data”, it will turn red with white outline.
    - If you select a point from “Eliminated data”, it will turn white with black outline.
- Click on “Remove point” button. You will see them disappearing from the collection of origin.
- You will notice points you have removed from ‘filtered data’ have appeared in “Eliminated data” coloured in red with yellow outline, and points removed from “Eliminated data” will be in “filtered data” coloured in green with yellow outline.

4. Download clean dataset:

Once you are fine with the cleaning results, you can download the new dataset in csv format by clicking on “Download csv”. The csv file is filled up with the data found in the ‘Filtered data’ collection[1].

5. Other

To change terrain transparency

Help button is not working for now.