

# Exploration et classification automatique de comportements d'oiseaux migrateurs avec ARLAS

06/04/2020

---

*by*



**Gisaïa**  
Ethics Biotope  
avenue de l'Escadrille Normandie Niemen  
31700 Blagnac  
[www.gisaia.com](http://www.gisaia.com) [www.arlas.io](http://www.arlas.io)

# TABLE DES MATIERES



Introduction	3
1. Un jeu de données composé de cigognes	4
2. ARLAS, une exploration fluide et interactive	5
3. Vers la détection de la migration	7
Conclusion	13

## Introduction

Pour mieux comprendre les comportements des animaux, de nombreux chercheurs équipent certains individus de balises GPS et de capteurs afin de pouvoir suivre leurs déplacements. Ces nouvelles données disponibles permettent de grandes avancées dans la compréhension de notre biodiversité mais également du bouleversement climatique. Pour rendre possible l'analyse des données à la volumétrie est importante, il est nécessaire d'utiliser des outils performants.

Chez Gisaïa, nous sommes sensibles aux enjeux environnementaux et développons ARLAS Exploration, un logiciel open source d'exploration de données géo-spatiales capable de répondre aux problématiques de Big Data. Nous avons décidé de mettre ARLAS à l'épreuve de ces données animales pour rendre possible leur exploration interactive.

Les données d'une étude sur une population de cigognes ont été utilisées, et nous allons voir comment ARLAS permet de découvrir le comportement de ces oiseaux migrateurs. Nous verrons également comment il peut accompagner la mise en place d'algorithmes de Machine Learning pour détecter automatiquement certains de ces comportements.

# 1. Un jeu de données composé de cigognes

Le [Max Planck Institute of Animal Behavior](#), institut de recherche allemand qui étudie la faune, a lancé récemment [Movebank](#), une plateforme qui vise à centraliser les données d'études d'animaux des chercheurs du monde entier pour encourager la collaboration et rendre possible le croisement des données des différentes études. Cette initiative a également pour objectif de rendre ces données accessibles librement au grand public.



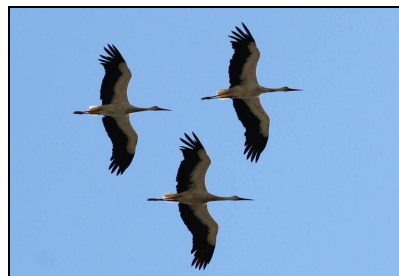
**MAX PLANCK INSTITUTE**  
OF ANIMAL BEHAVIOR



Pour une étude de Cheng et al. (2019) [1], les chercheurs ont équipé une population de 169 cigognes de balise GPS et ont recueilli des données entre 2013 et 2019. Ces données ont été récupérées sur la plateforme Movebank [2].



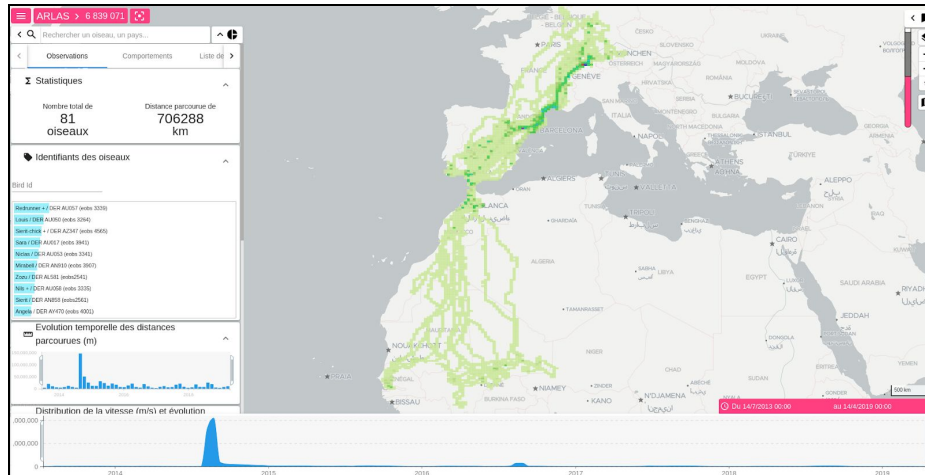
La cigogne blanche (*Ciconia ciconia*) est une grande espèce d'oiseaux échassiers de la famille des Ciconiidae. Son plumage est principalement blanc, avec du noir sur les ailes. Cette espèce a fait l'objet de programmes de protection et de réintégration et est principalement présente en Europe orientale et occidentale. La cigogne est une grande migratrice et hiverne en Afrique, ce qui rend ses déplacements particulièrement fascinants à étudier.



Les données ont subi un premier traitement afin notamment de regrouper et relier entre elles les observations successives d'un même oiseau et de calculer les distances parcourues. Ce genre de traitement est opéré grâce à ARLAS PROC/ML, notre plateforme de traitement massif distribué. Les données ainsi mises en forme sont ensuite intégrées dans ARLAS et nos cigognes sont alors prêtes à être explorées.

## 2. ARLAS, une exploration fluide et interactive

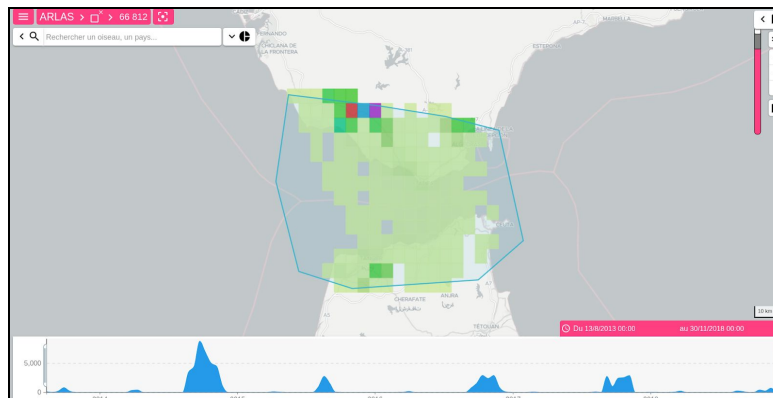
ARLAS Exploration est une application carto-centrique qui permet avant tout d'apprécier la dispersion spatiale des données. Une barre de graphiques sur la gauche de l'application permet également de visualiser et filtrer les autres dimensions de la donnée, ainsi on peut observer les distributions de différentes grandeurs telles que la distance parcourue, mais aussi l'altitude et la vitesse. En bas, la frise temporelle permet de voir la répartition temporelle des mesures réalisées.



*Vue d'ensemble des données de l'étude dans ARLAS Exploration*

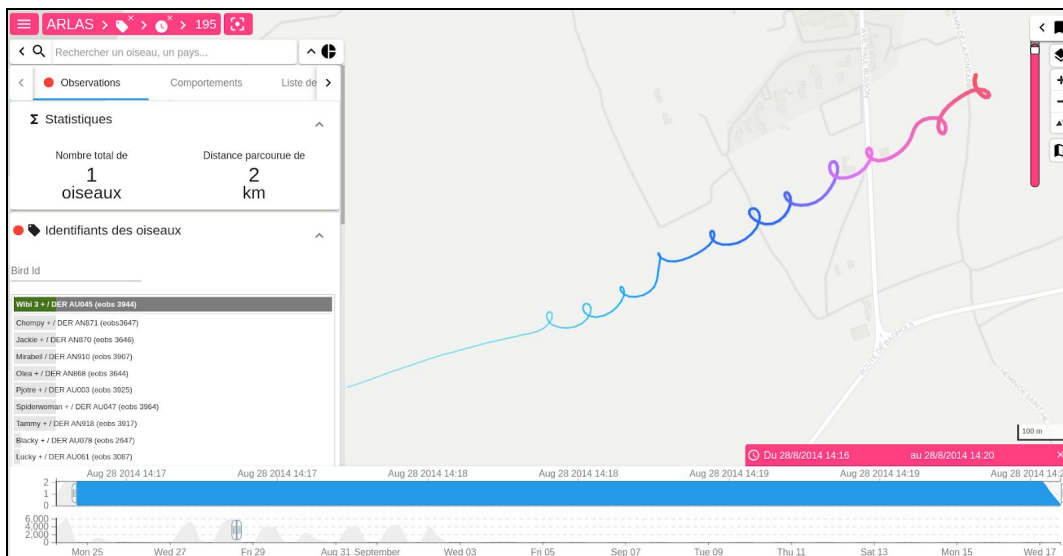
On observe bien que les cigognes tracées évoluent au sein d'un périmètre allant du sud de l'Allemagne à l'Afrique de l'Ouest. On remarque un pic du nombre d'observations au mois d'Août 2014. De plus, on dispose en réalité des données de 81 oiseaux pour un total d'environ 7 millions de positions.

Les différents graphiques permettent de filtrer les données sur les dimensions représentées et l'ensemble de l'application se met à jour instantanément à chaque nouvelle sélection. On peut également naviguer dans la carte de manière classique pour se focaliser sur certaines zones et filtrer selon des sélections géographiques dessinées grâce à un outil situé à droite de l'application.



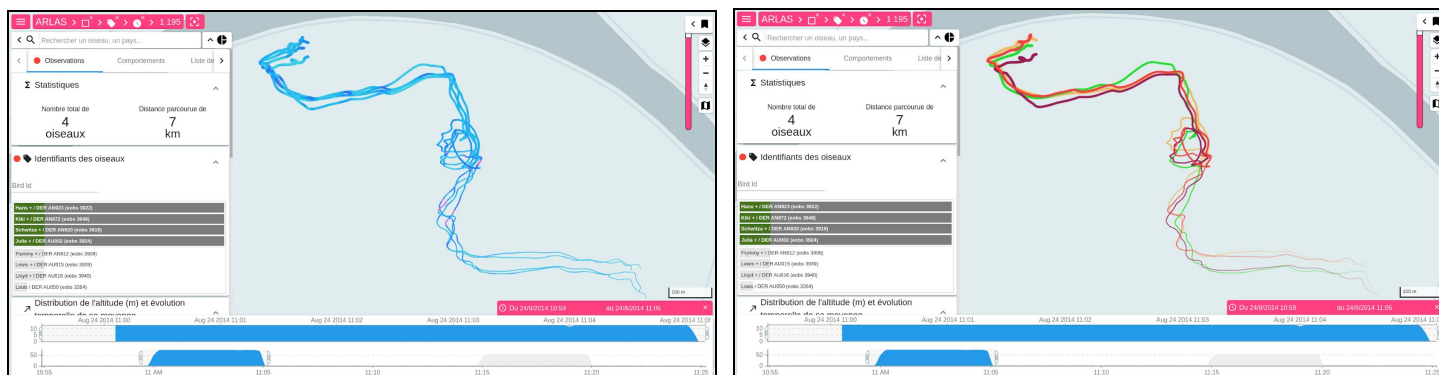
*Exemple de sélection au détroit de Gibraltar*

En fonction de la quantité de données à afficher dans la fenêtre de l'application, on bascule d'un mode "agrégé", des cartes de densités, idéales pour la visualisation générale des flux, à un mode "features", le détail des données réelles pour observer le trajet effectif des cigognes. On peut ainsi isoler des comportements intéressants, et observer que certaines cigognes semblent utiliser les courants d'air chaud pour gagner en altitude par exemple :



*Exemple de trajet de la cigogne Wibi 3 colorée par altitude*

On peut colorer le trajet réel affiché de différentes manières, dépendant de la vitesse ou bien de l'identifiant de l'oiseau par exemple :



*Trajetoires de 4 cigognes colorées par vitesse (à gauche), par identifiant de l'oiseau (à droite)*

Ces deux représentations permettent de comprendre que les quatre cigognes sélectionnées, nommées Hans, Schwitza, Kiki et Julia, se déplacent ensemble aux mêmes allures sur cette période, tout en pouvant bien distinguer les quatre cigognes.

Ainsi ARLAS Exploration est un outil particulièrement adapté pour interagir de manière intuitive et interactive avec les données de positions d'oiseaux, même lorsque le volume des données devient important, ce qui en fait un solide partenaire pour les chercheurs.

### 3. Vers la détection de la migration

Les cigognes observées ont tendance à migrer sur des distances importantes pour changer de lieux d'habitat. On peut alors distinguer deux attitudes : rester dans la même zone ('local') ou voyager pour changer de zone ('travel'). Ces deux comportements sont visibles localement à l'œil nu dans ARLAS Exploration, mais une détection automatique de ces activités pourrait permettre d'étudier à grande échelle les migrations de l'ensemble des oiseaux de manière extrêmement efficace. C'est pourquoi nous avons choisi d'utiliser des algorithmes de Machine Learning pour automatiser cette identification. Un apprentissage supervisé a été réalisé pour entraîner le modèle de classification.

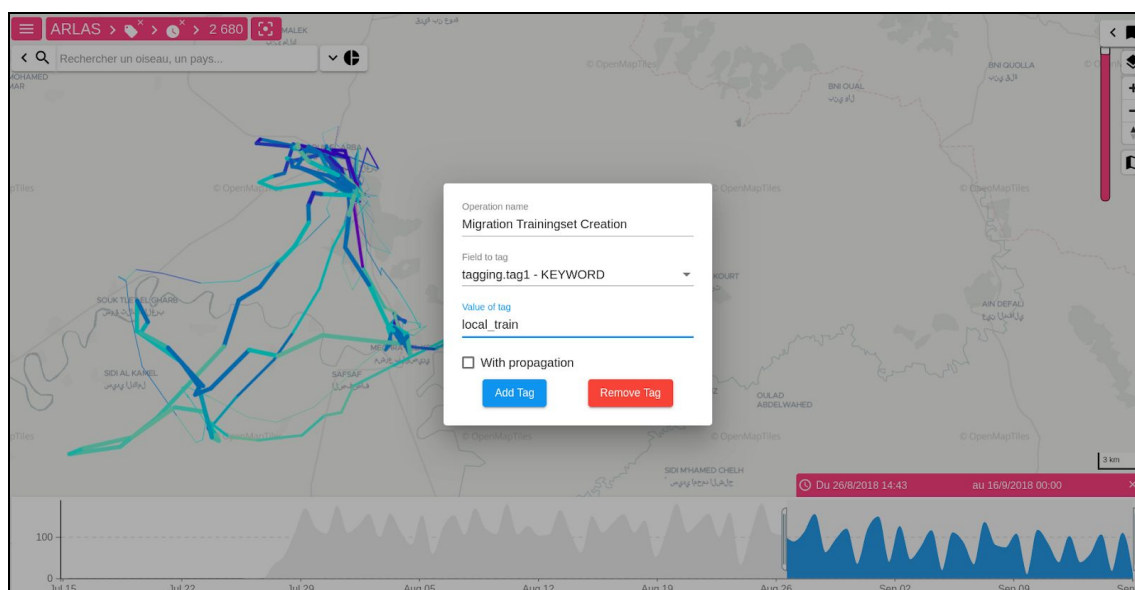
Ce processus s'est donc effectué en plusieurs étapes :

- Construction d'un jeu d'entraînement
- Calcul de nouveaux indicateurs
- Choix du modèle de classification
- Visualisation du résultat

#### Construction d'un jeu d'entraînement

Un modèle de classification supervisé a besoin de données d'entraînement pour apprendre à reconnaître les comportements ciblés. Dans notre cas, il est nécessaire de pouvoir annoter nos données, ainsi chaque fraction de déplacement doit être identifiée comme 'travel' ou 'local'.

ARLAS Exploration est également un outil particulièrement adapté à la création de jeu d'entraînement puisqu'il permet d'attribuer un label à la sélection de données en cours. On peut donc manuellement identifier les parties de trajectoires correspondant à un grand déplacement ('travel') ou à une activité locale ('local') et les labelliser comme tel.



Interface de labellisation



Exemple de trajectoire du jeu d'entraînement, 'local' (rouge) et 'travel' (vert)

En pratique, 4 oiseaux ont été labellisés ainsi :

Nom de l'oiseau	Période	Nombre de fragments	Proportion 'travel'	Proportion 'local'
Yolo	26/07/13 - 06/08/14 (12 mois)	77.487	8.5%	91.5%
Redrunner	10/06/15 - 27/08/16 (14 mois)	95.862	8.4%	91.6%
Niclas	20/04/15 - 01/09/15 (16 mois)	87.066	1.7%	98.3%
Mirabell	27/04/15 - 21/07/17 (15 mois)	73.957	5.0%	95.0%

Pour chaque oiseau, une période d'environ un an a été utilisée pour capturer au moins un aller-retour dans la migration. On a donc un total de 334.372 fragments (intervalle entre deux mesures) qui vont être exploitables pour l'entraînement du modèle.

## Calcul de nouveaux indicateurs

La qualité d'un algorithme de Machine Learning dépend avant tout de la qualité de ses données d'entraînement. Une fois que l'on a labellisé un nombre suffisant de fragments représentatifs des comportements à détecter, il est nécessaire de bien choisir les grandeurs que l'on va donner en entrée du modèle. Tout d'abord, la récupération des données taguées peut s'effectuer grâce à une API d'ARLAS, disponible notamment en Python, permettant le téléchargement des données d'entraînement.

```
import arlas_api_python as arlas_api

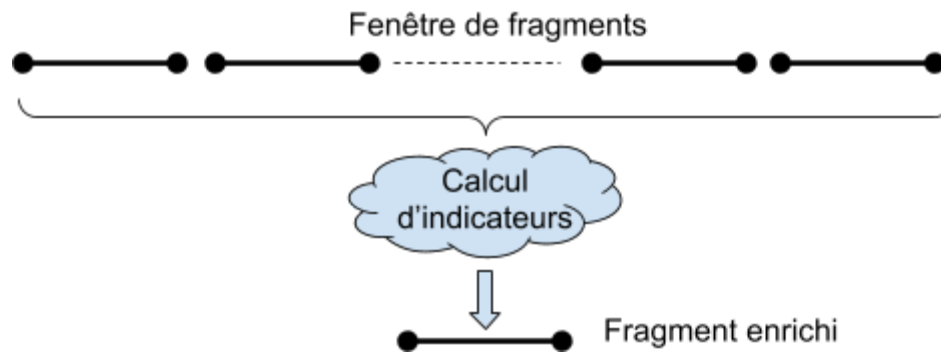
# Build request to get manually labelled data
request_body = {"filter":
    {"f": [
        {"field": "training.set.path",
         "op": "eq",
         "value": "local,travel"}
    ]}
}

# Create and call the Arlas Explore API instance
arlas_api_instance = arlas_api.ExploreApi(arlas_api.ApiClient("http://arlas_server_url"))
request_result = arlas_api_instance.search_post(collection="birdtracking_collection_name",
    body=request_body)
```

Récupération de données grâce à l'API (Python) d'ARLAS



Dans notre cas, les grandeurs choisies illustrant le déplacement de ces oiseaux sont basées sur des distances parcourues et des distances "à vol d'oiseau" calculées sur différentes fenêtres temporelles. Ces grandeurs ne sont pas présentes dans les données brutes et sont calculées en utilisant ARLAS PROC/ML, notre plateforme de processing adaptée aux gros volumes de données.



Une fois ces nouvelles grandeurs calculées, elles pourront être utilisées pour entraîner les différents modèles de Machine Learning choisis.

## Choix du modèle de classification

Plusieurs modèles de classification ont été testés. Afin de pouvoir comparer la qualité de ces classifieurs, le calcul d'indicateurs est nécessaire. Comme les classes sont disproportionnées (~6% de 'travel' dans le jeu d'entraînement), plusieurs métriques ont été utilisées pour évaluer correctement la qualité de détection du 'travel'. Une "cross-validation" est réalisée pour éviter le sur-apprentissage en partitionnant le jeu d'entraînement et en mesurant les performances de classification sur des données n'ayant pas servi à l'entraînement du modèle.

Les métriques utilisées se basent sur la matrice de confusion de la prédiction :

Matrice de confusion		Classe réelle	
		- (local)	+ (travel)
Classe Prédite	- (local)	Vrais Négatifs	Faux Négatifs
	+ (travel)	Faux positifs	Vrais positifs

On a :

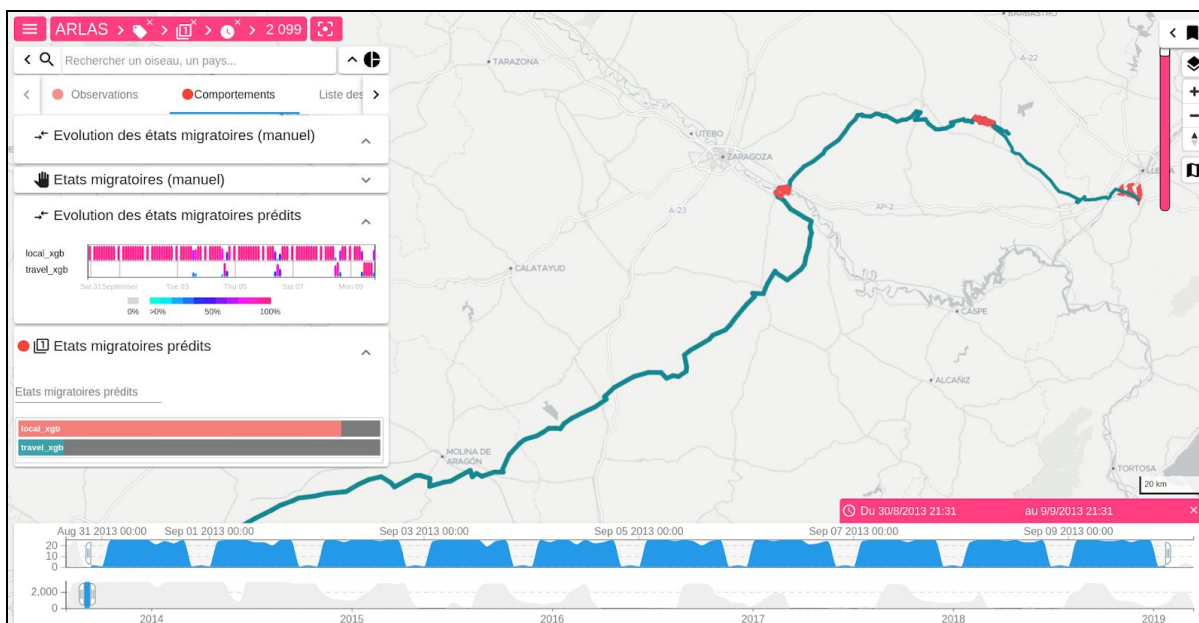
- Justesse (Accuracy) : Proportion générale de bonnes classifications
- Rappel (Recall) : Part des 'travel' réels correctement détectés
- Précision (Precision) : Proportion des 'travel' détectés effectivement réels
- Spécificité (Specificity) : Part des 'local' réels correctement détectés
- F1-score : Moyenne harmonique (compromis) entre Rappel et Précision

Toutes les expériences étant réalisées dans les mêmes conditions, les modèles peuvent être comparés entre eux, notamment grâce à l'outil "MLFlow", utilisé pour consigner les résultats. Finalement, après de nombreuses expérimentations, c'est un classifieur de type "XGBoost" qui a été retenu, à la fois pour ses performances et sa rapidité d'entraînement.

## Visualisation des résultats

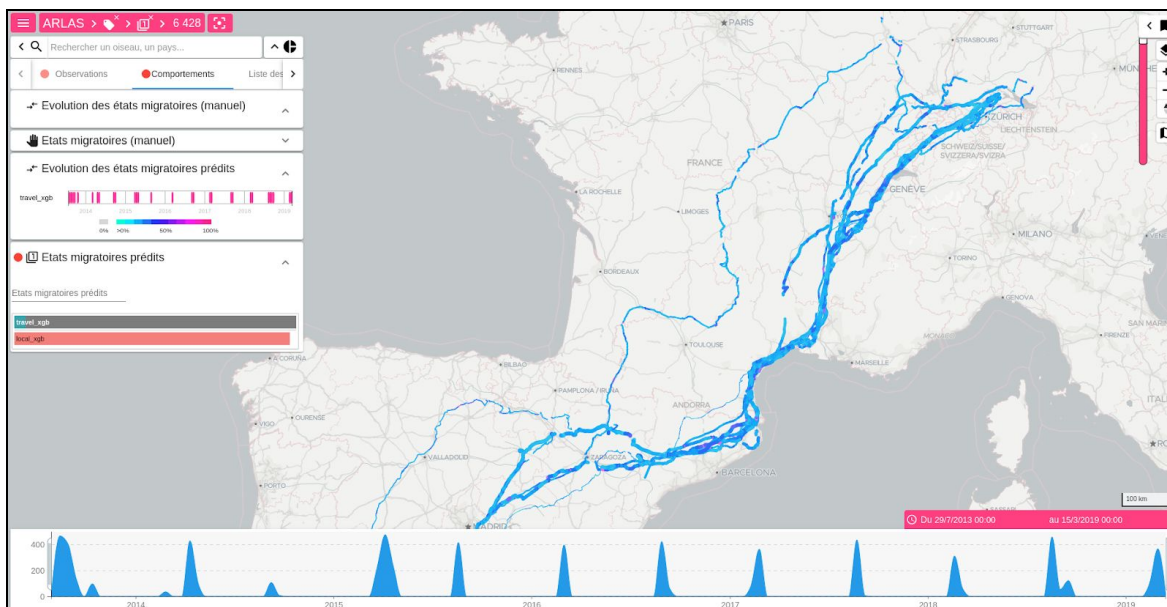
Une fois le modèle choisi et entraîné, il va pouvoir être appliqué à d'autres oiseaux et les résultats de cette détection de migration pourront être exploités dans ARLAS Exploration grâce au système de tagging (également disponible par API). Les résultats sont directement visualisables dans ARLAS Exploration. Ce qui permet de mieux comprendre notre modèle en identifiant rapidement sur quelle partie des données les prédictions seraient défaillantes. Ce qui donne également la possibilité de valider ou corriger les résultats, ce qui rend possible l'augmentation du training set pour entraîner de nouveau le modèle sur davantage de données. Dans le cas de nos cigognes, le modèle a été appliqué sur 26 oiseaux, correspondant à 2.650.000 fragments.

On peut désormais de manière beaucoup plus efficace suivre la trace d'un oiseau en particulier et identifier les différents lieux de pause de son trajet. Pour la cigogne nommée Zozu par exemple, on obtient les résultats suivants :



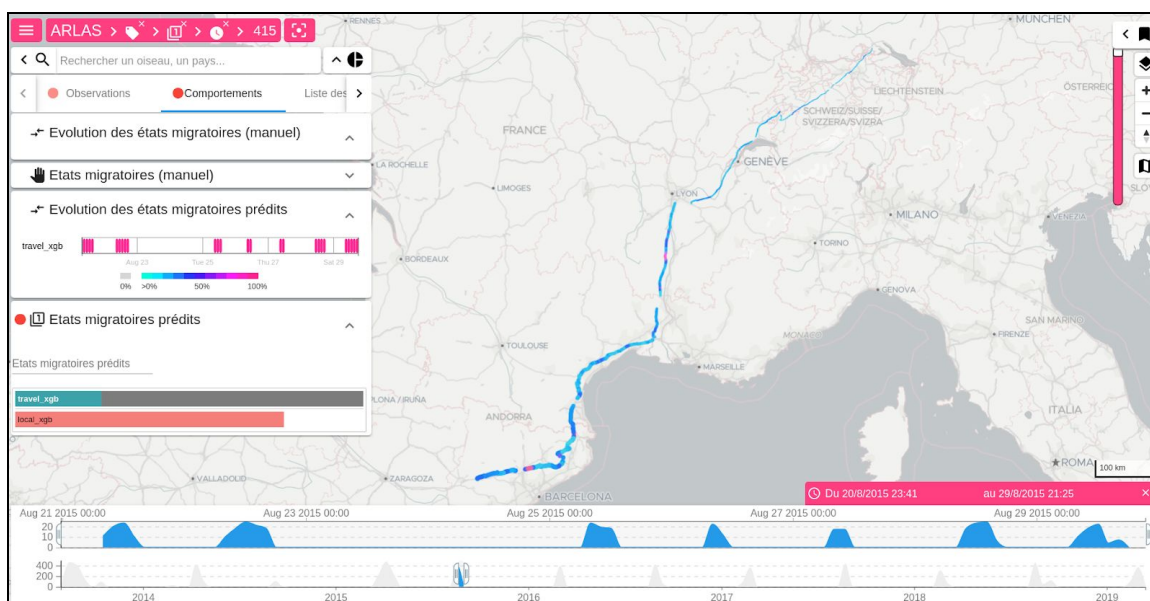
*Sélection d'une trace de la cigogne nommée Zozu*

On peut également dater les grands déplacements de ces cigognes. Si l'on considère les fragments prédits en 'travel' de la cigogne nommée Zozu par exemple, on observe les différents pics sur la frise temporelle, ce qui permet d'identifier les périodes de l'année où l'oiseau a migré :



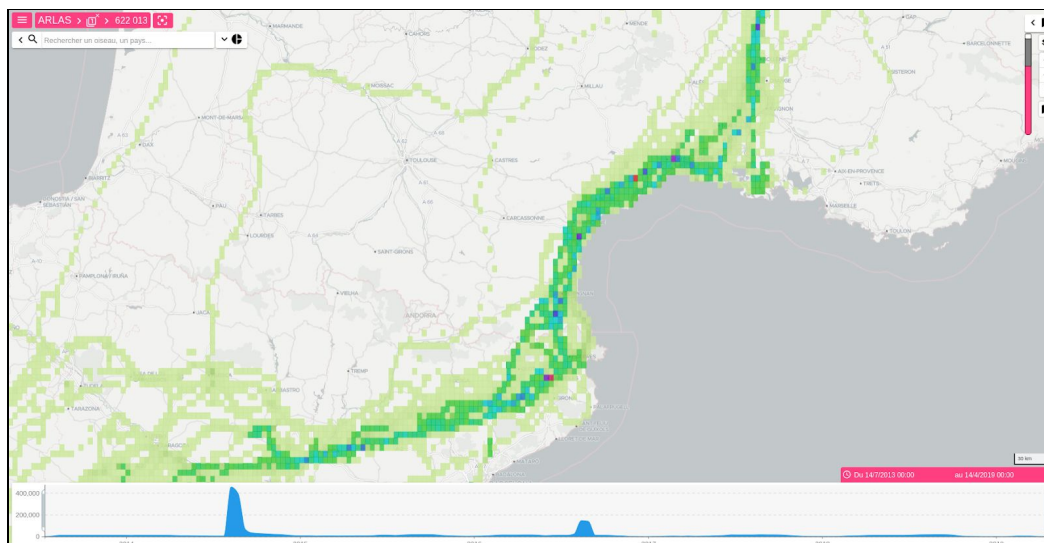
Sélection des fragment 'travel' de la cigogne nommée Zozu

Si l'on s'intéresse à un de ces pic en particulier, on peut dater très précisément ces migrations et voir les différentes étapes du trajet, ainsi ce trajet entre la Suisse et le nord de l'Espagne réalisé entre le 21/08/2015 et le 29/08/2015 s'est effectué en 7 étapes :



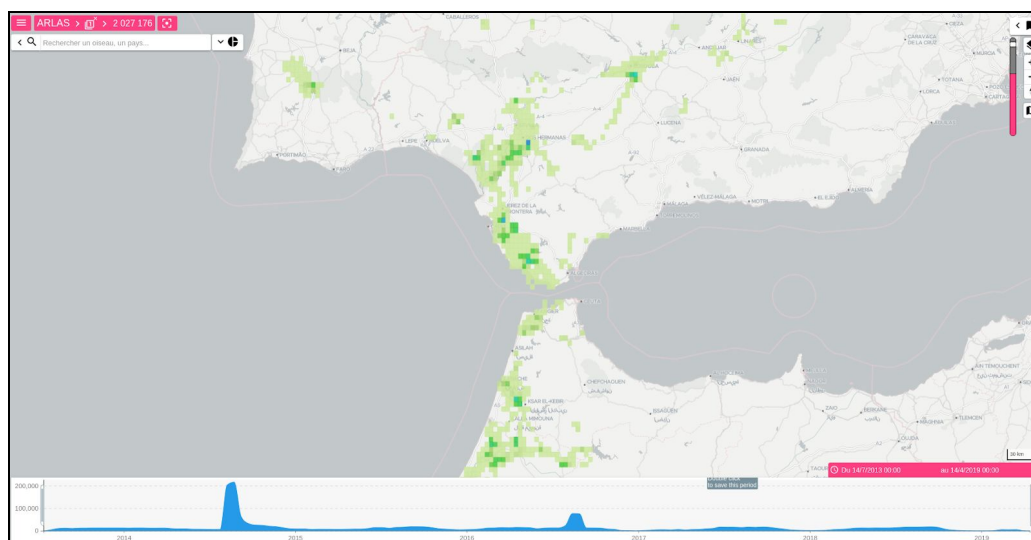
Sélection des fragments 'travel' de la cigogne nommée Zozu (21/08/2015 - 29/08/2015)

On peut également explorer la migration de ces oiseaux à plus grande échelle. Si on sélectionne l'ensemble des fragments en 'travel' pour tous les oiseaux labellisés, on voit se dessiner un couloir de migration suivant la côte méditerranéenne en direction de l'Espagne dans le sud de la France :



*Sélection des fragments identifiés en 'travel'*

Enfin on peut aussi identifier les différents lieux de vie que privilégient les cigognes lors de leur périple aux alentours du détroit de Gibraltar en sélectionnant les fragments prédits en 'local' pour l'ensemble des cigognes :



*Sélection des fragments identifiés en 'local'*

La détection automatique des migrations a grandement facilité leur analyse, permettant aux experts ornithologues de se concentrer sur les variations de dates et de destinations des migrations entreprises par les cigognes. La possibilité de croiser ces informations avec d'autres données, météorologiques par exemple, pourrait permettre d'encore mieux comprendre le comportement de ces grands oiseaux migrateurs ainsi que leur adaptation aux changements climatiques.

## Conclusion

Nous avons vu qu'ARLAS est particulièrement bien adapté à l'exploration des données de positions des cigognes. La navigation cartographique interactive permet d'en tirer des informations précieuses sur le comportement de ces oiseaux migrateurs. ARLAS permet également d'accompagner la production d'algorithmes de Machine Learning en facilitant la création de jeu d'entraînement et la visualisation des résultats de classification. Enfin, une fois les modèles de Machine Learning entraînés, il est possible de les appliquer aux données à grande échelle avec à ARLAS PROC/ML et d'en voir les résultats dans ARLAS Exploration. Tous ces résultats sont accessibles dans une démonstration que vous pouvez retrouver à [demo](#).

Si nous nous sommes focalisés ici sur la migration, de nombreux autres comportements animaliers pourraient faire l'objet de telles études. Evidemment, ARLAS peut se décliner à toutes sortes d'animaux géo-tracés, mais également à toutes données géoréférencées. N'hésitez pas à jeter un œil à d'autres exemples d'application sur [demo.arlas.io](http://demo.arlas.io).

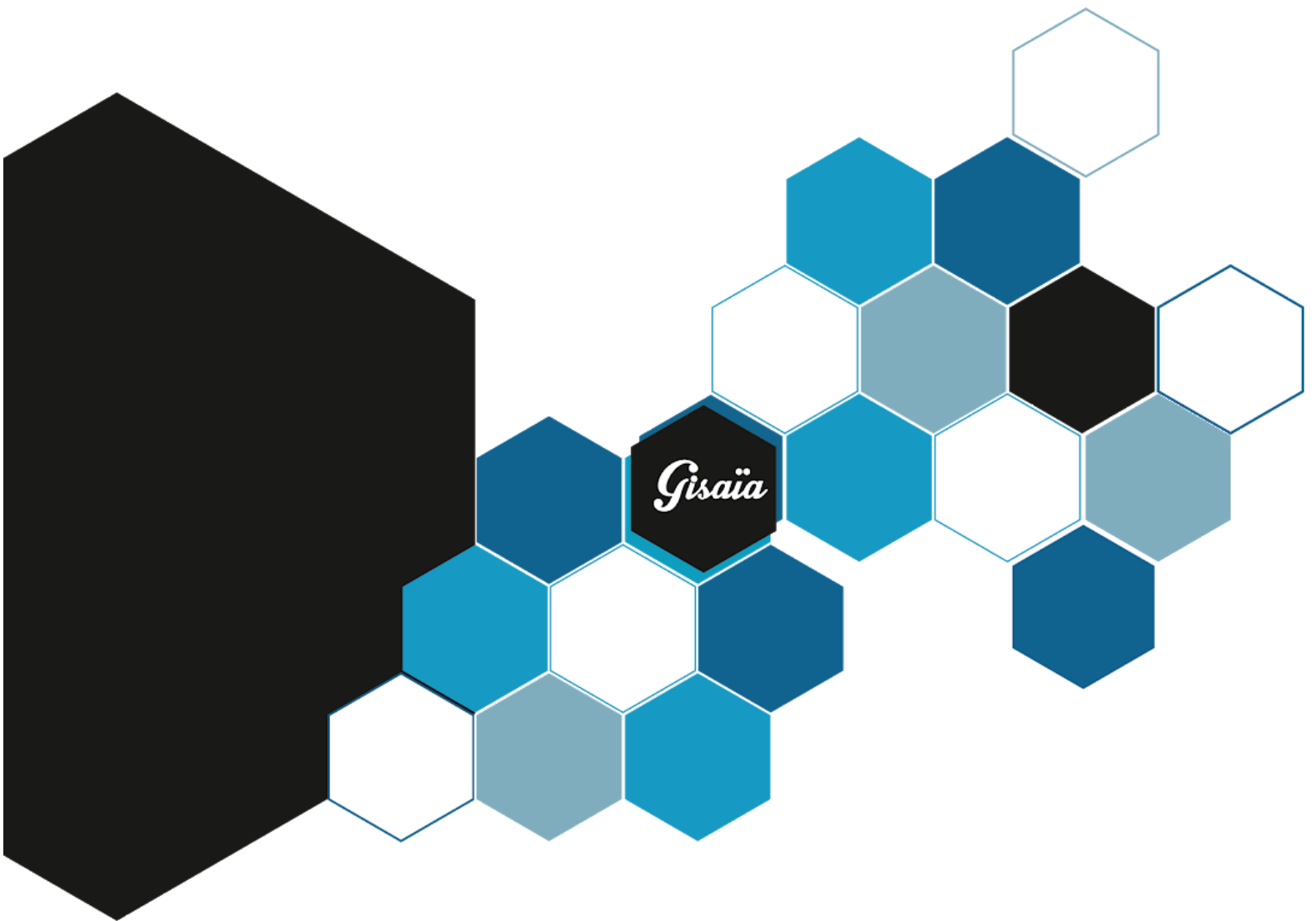
### Remerciements :

Nous tenons à remercier le directeur du Max Planck Institute of Animal Behavior, Dr Martin Wikelski, et son équipe pour la mise à disposition de ces données.

### Références :

[1] Cheng Y, Fiedler W, Wikelski M, Flack A (2019) "Closer-to-home" strategy benefits juvenile survival in a long-distance migratory bird. *Ecology and Evolution*. doi:10.1002/ece3.5395

[2] Fiedler W, Flack A, Schäfle W, Keeves B, Quetting M, Eid B, Schmid H, Wikelski M (2019) Data from: Study "LifeTrack White Stork SW Germany" (2013-2019). Movebank Data Repository. doi:10.5441/001/1.ck04mn78



**Gisaia**

Ethics Biotope  
2 avenue de l'Escadrille Normandie Niemen  
31700 Blagnac  
[www.gisaia.com](http://www.gisaia.com) [www.arlas.io](http://www.arlas.io)  
All Rights Reserved.

