Outils de Deep Learning pour le suivi d'oiseaux marins : approches supervisées et modèles génératifs

Amédée Roy[1,2], Ronan Fablet[2], Sophie Lanco Bertrand[1]

[1] Marbec, IRD, Sète, France [2] IMT Atlantique, Brest, France

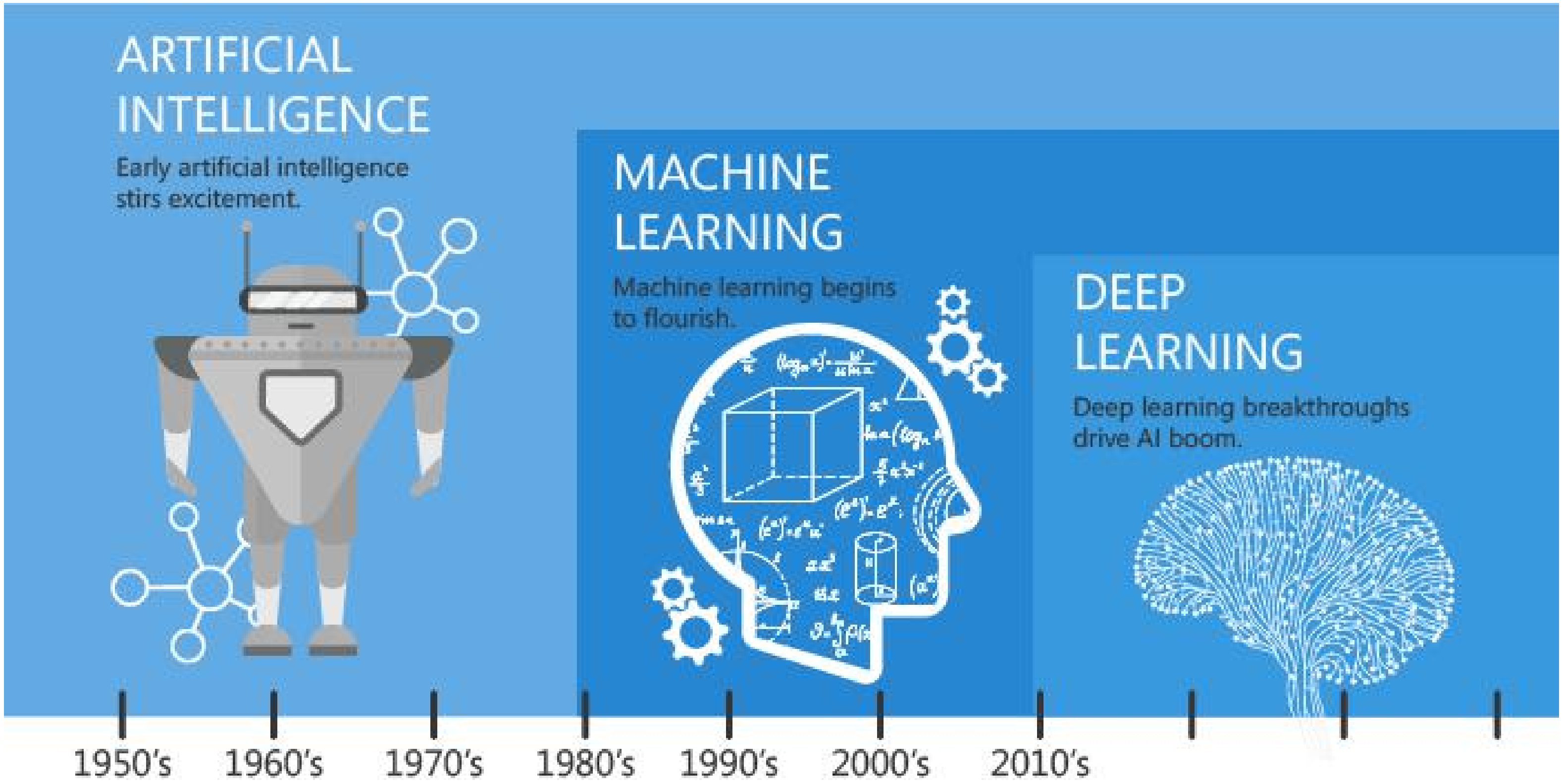
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

marbec
marine biodiversity
exploitation & conservation

Institut de Recherche
pour le Développement
FRANCE

Deep Learning

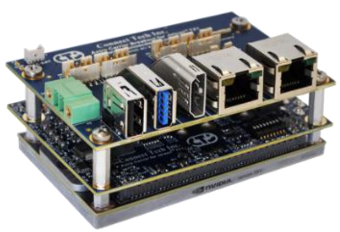
→ What is Deep Learning?



Deep Learning

→ Why Deep Learning?

High-Performance Computing



Large Datasets

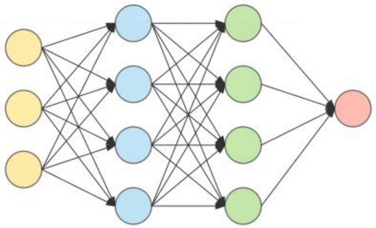
Efficient & easy-to-use libraries

Keras

TensorFlow

PYTORCH

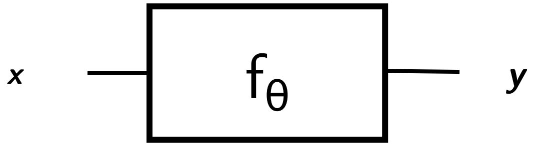
End-to-end Architectures

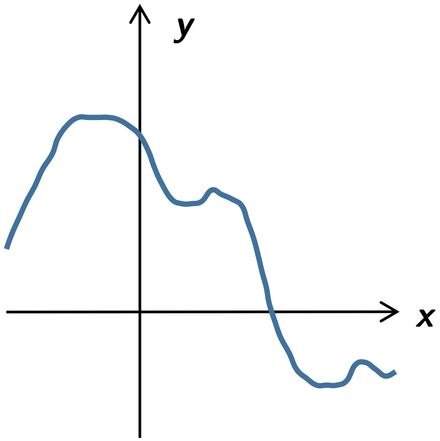


Neural Networks

→ What is a NN?

<u>Neural Network = a composition of many elementary functions</u>

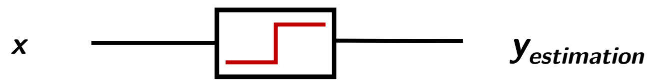


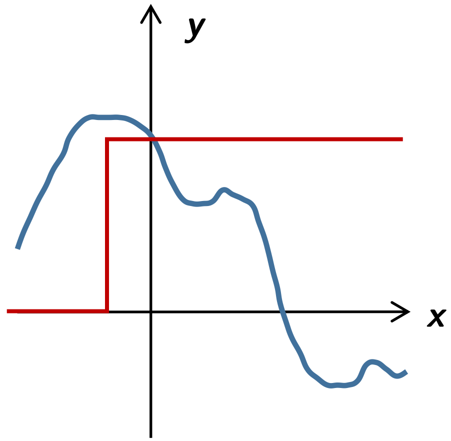


Neural Networks

→ What is a NN?

Neural Network = a composition of many elementary functions

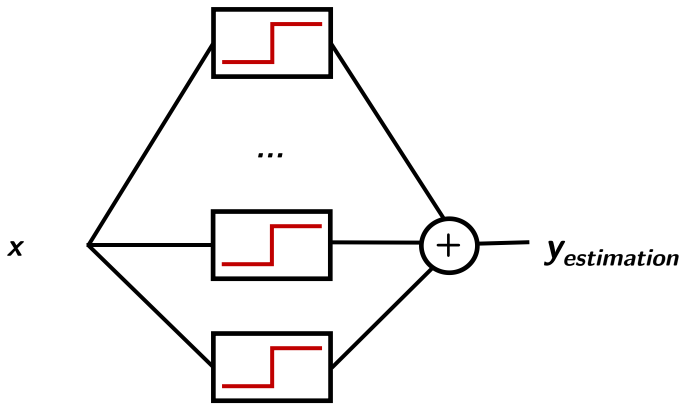


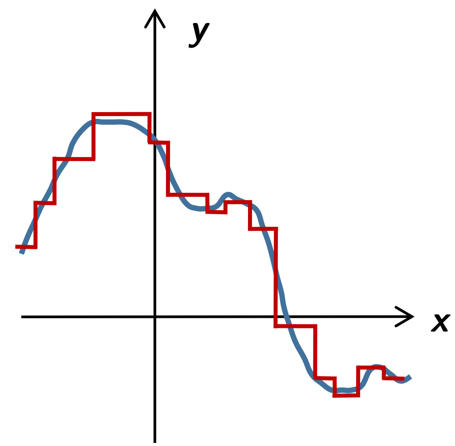


Neural Networks

→ What is a NN?

Neural Network = a composition of many elementary functions





Neural Networks

$\longrightarrow$ How to train a NN?

<u>Neural Network Training procedure</u>

01 Prediction

$$x \longrightarrow \boxed{f_\theta} \longrightarrow y_{estimation}$$

Neural Networks

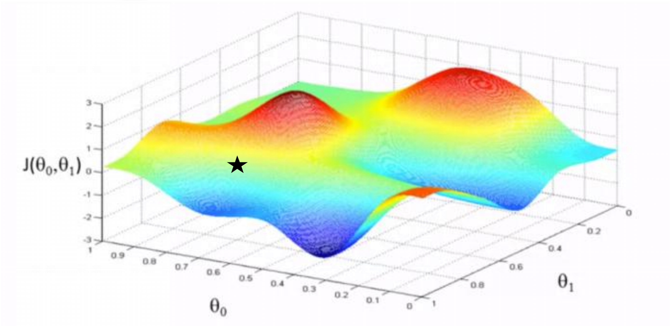
→ How to train a NN?

Neural Network Training procedure

01 Prediction

x → $\boxed{f_\theta}$ → $y_{estimation}$ y_{true}

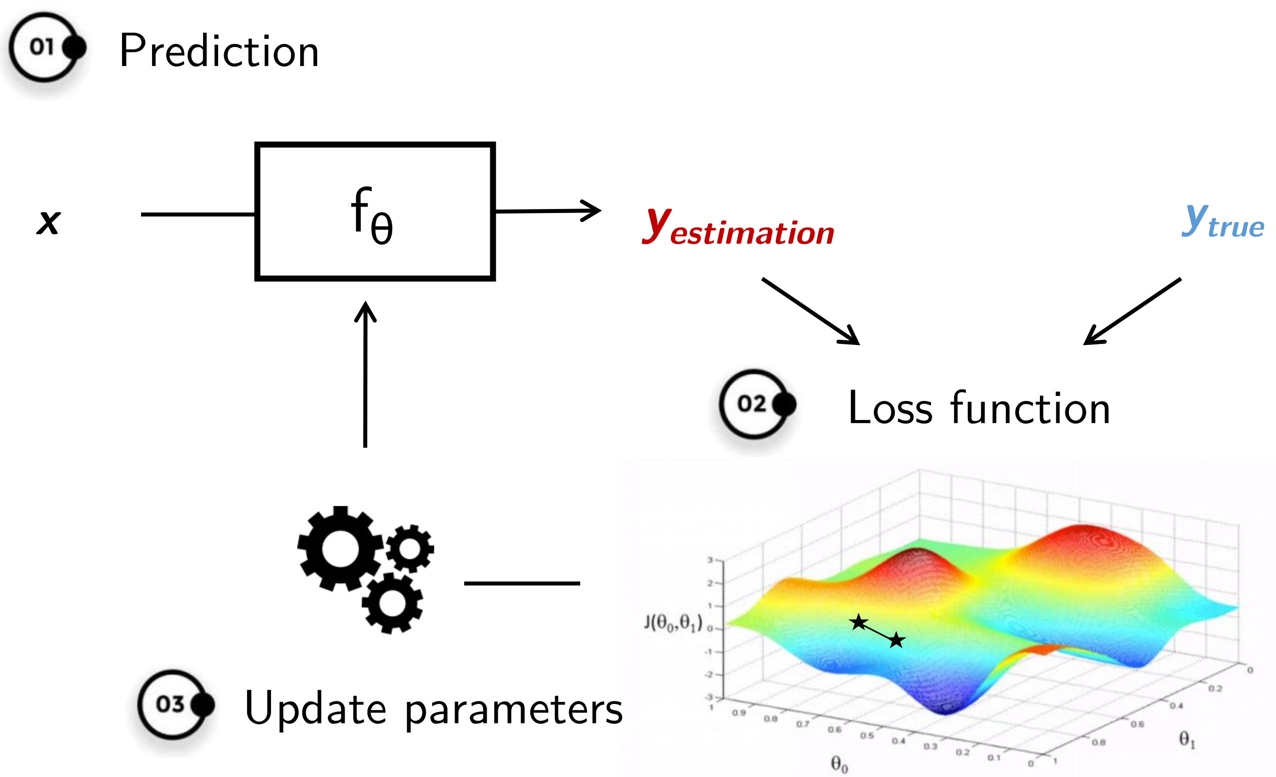
02 Loss function



Neural Networks

→ How to train a NN?

Neural Network Training procedure



Neural Networks

→ How to train a NN?

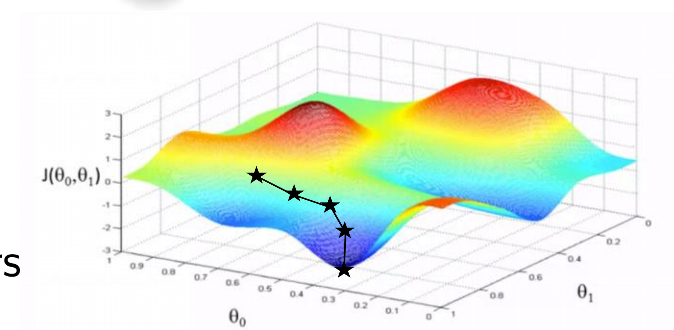
Neural Network Training procedure

01 Prediction

x → f_θ → $y_{estimation}$

y_{true}

02 Loss function



03 Update parameters

Roadmap

01 **Supervized Learning**

→ Identification of breeding habitat from satellite data

→ Segmentation of GPS tracks to detect dives

02 **Generative Models**

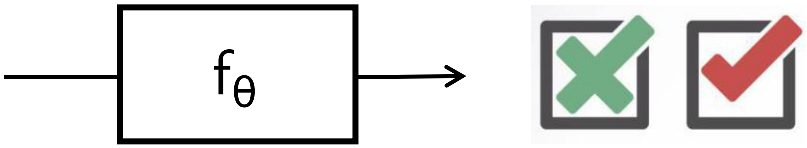
→ Simulation of seabird foraging trajectories

→ Climate-based prediction of trajectories

Identification of breeding habitat

→ Problem Overview

Antonio Garcia-Quintas
(IRD, Universidad de La Habana)



Habitat description :

- satellite
- 10 km radius around site

Breeding site presence/absence

Identification of breeding habitat

→ Data Overview

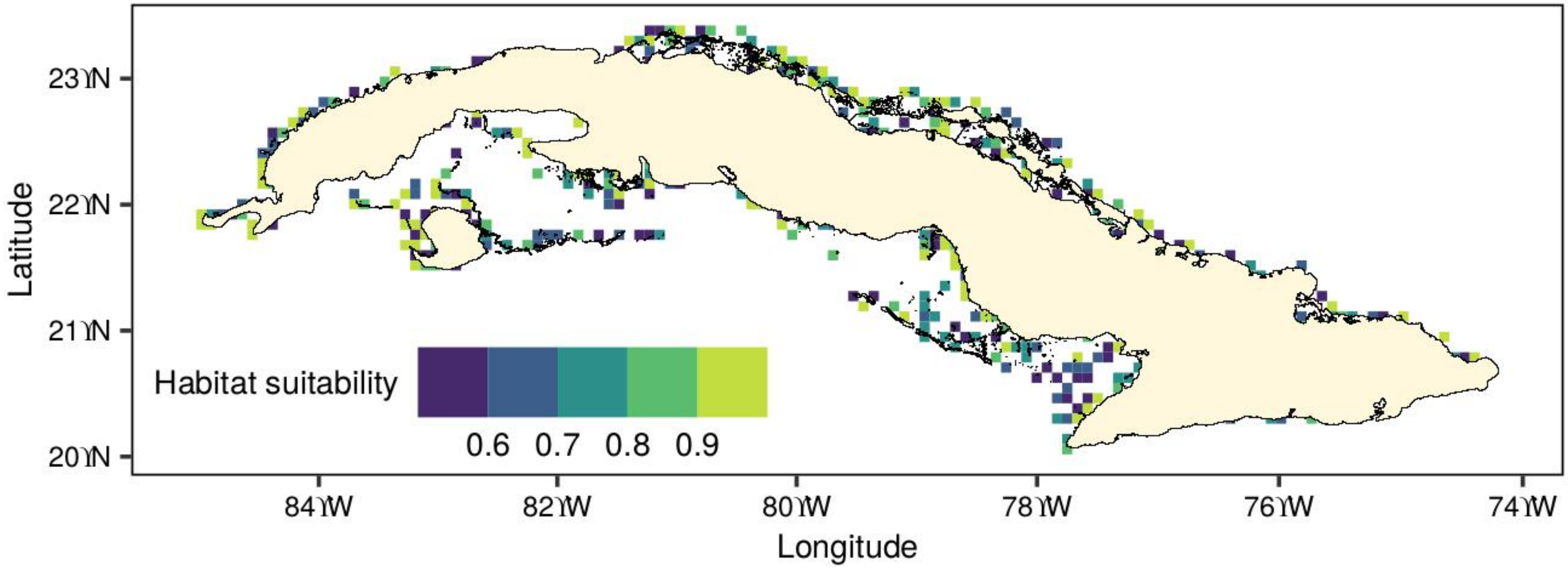
Available area with the potential breeding localities

★ 50 Registered breeding localities

Identification of breeding habitat

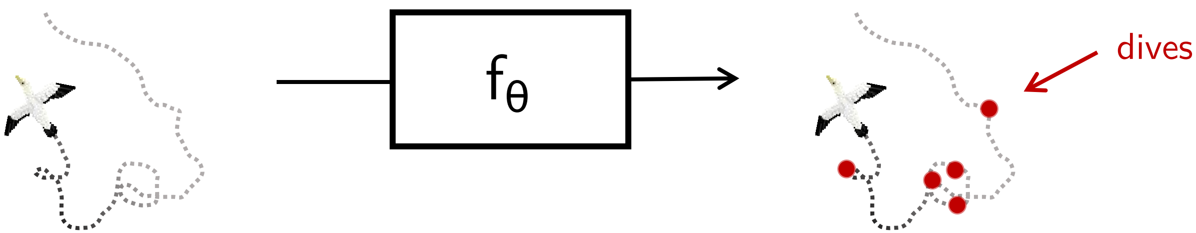
→ Results

Precision = 0.90
Recall = 0.99
F-score = 0.95



Segmentation of GPS tracks

→ Problem Overview



GPS track :

- step speed / turning angle
- 20 steps window

Dive occurences :

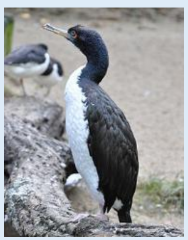
- TDR data threshold

Segmentation of GPS tracks

→ Data Overview

01 Network training

Guanay cormorant ☆
(Leucocarbo bougainvilli)



Nb of trips : **76**
Dives : **23 %**
Dive duration: **18 s**
Resting : **22 %**

Peruvian booby ☆
(Sula variegata)

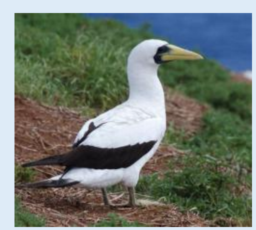
Nb of trips : **133**
Dives : **1 %**
Dive duration: **2 s**
Resting : **7 %**

Deep Networks

(UNet, CNNet, FCNet)

02 Network transfer

Masked booby △
(Sula dactylatra)

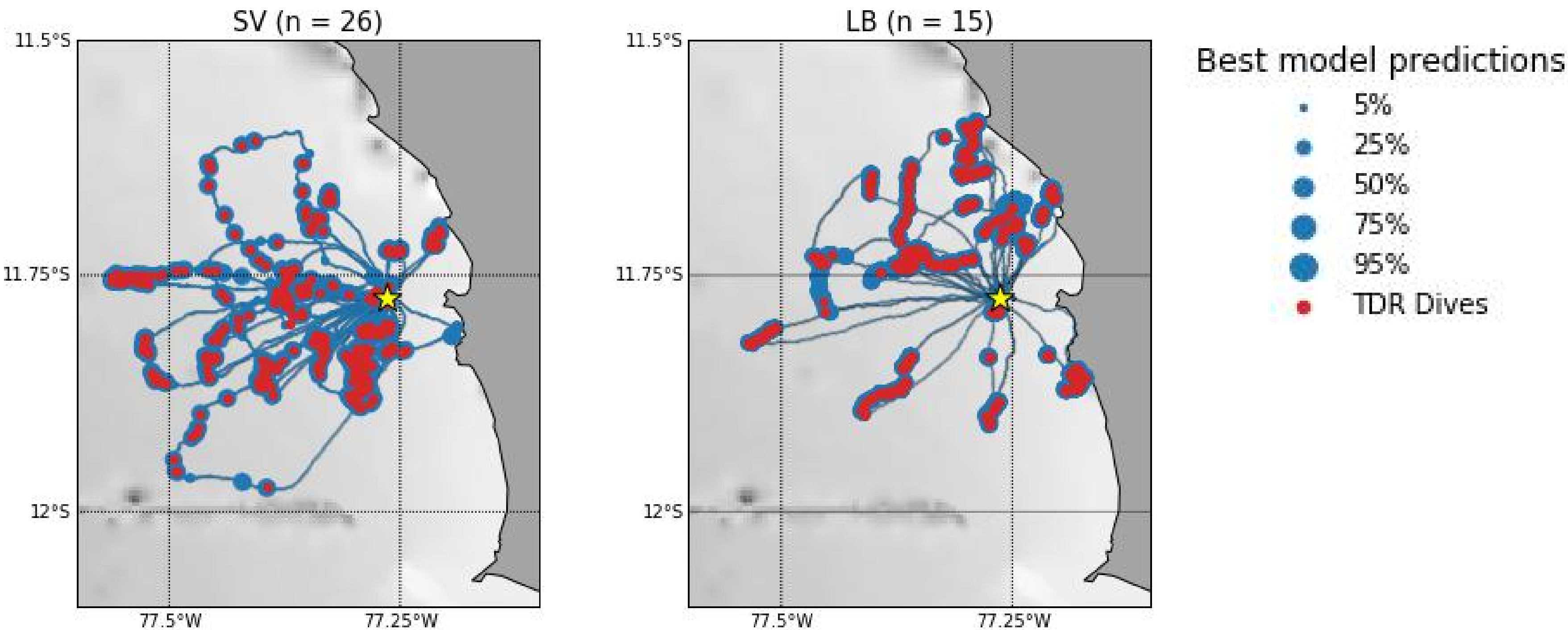


Nb of trips : **64**
Dives : **0.2 %**
Dive duration: **2 s**
Resting : **33 %**

Segmentation of GPS tracks

→ Results

01 Network training



AUC = 0.96
F-score = 0.91

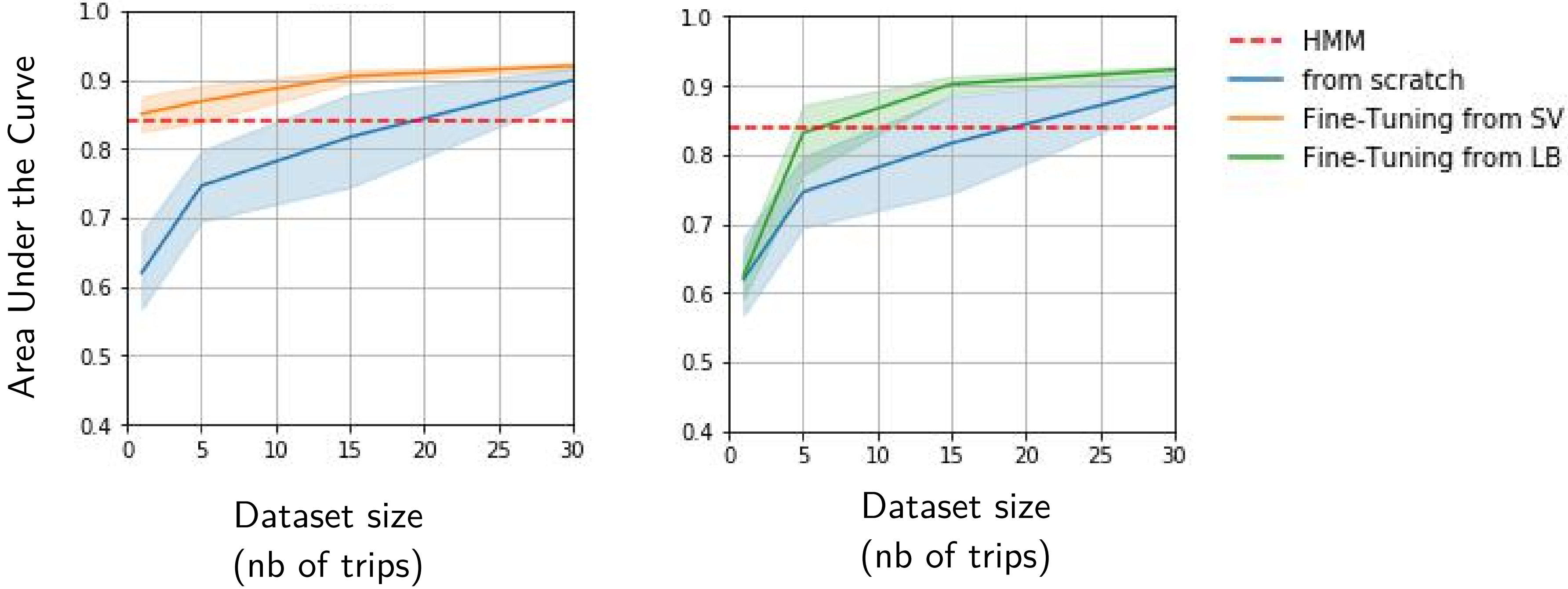
AUC = 0.93
F-score = 0.85

Segmentation of GPS tracks

→ Results

Masked booby
(Sula dactylatra)

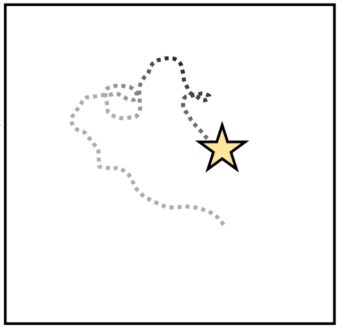
02 Network transfer



Simulation of foraging trips

$\rightarrow$ Generative Adversarial Networks

$z \sim \mathcal{U}(0,1) \longrightarrow \boxed{G_\theta} \longrightarrow$



Random noise :

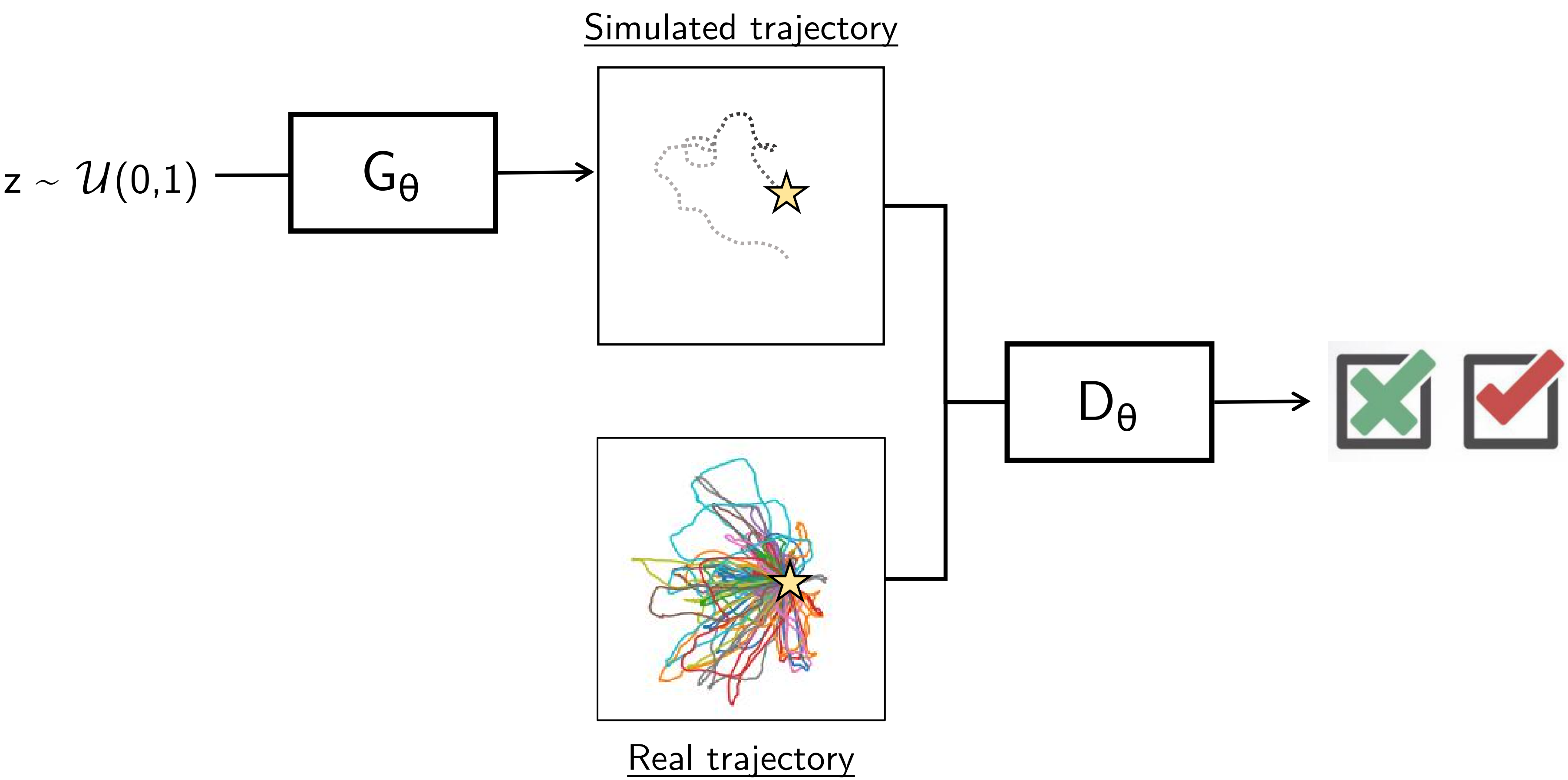
- samples from a pre-defined distribution

Realization of a stochastic process :

- 200 steps trajectories
- longitude / latitude / dives

Simulation of foraging trips

$\rightarrow$ Generative Adversarial Networks



Simulation of foraging trips

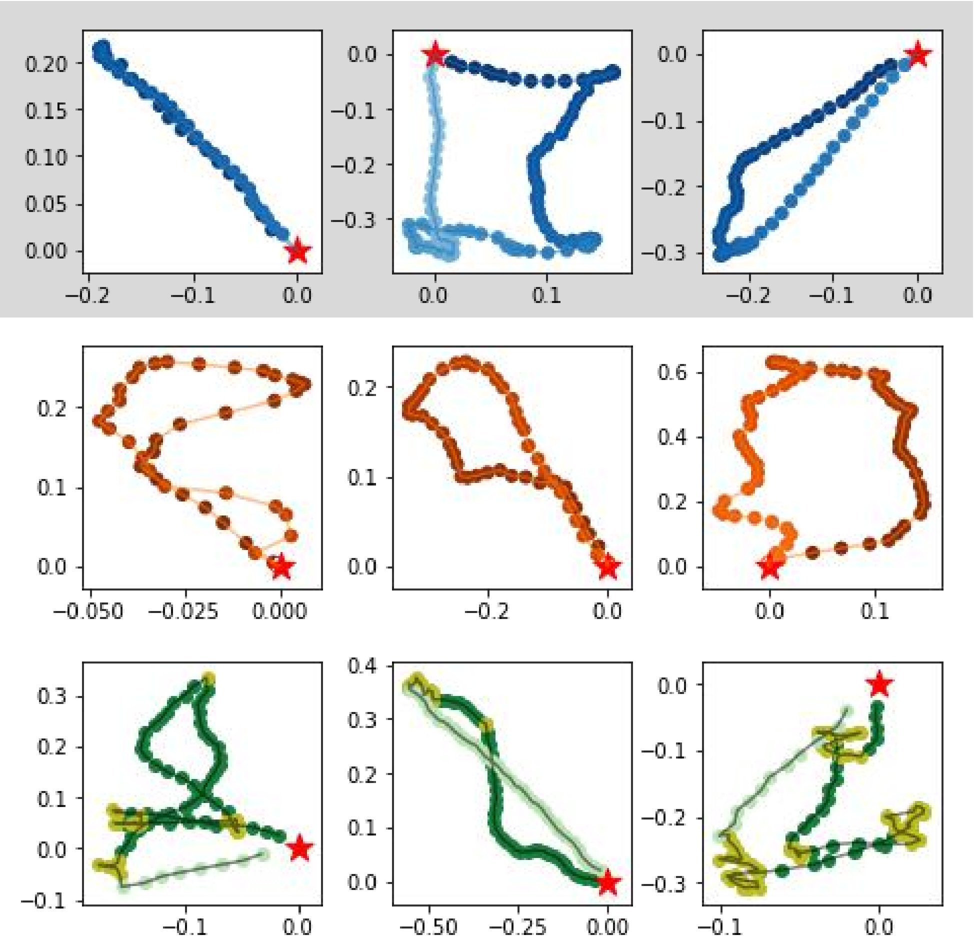
→ Result

Peruvian booby
(Sula variegata)

Real data

GAN simulations

HMM simulations



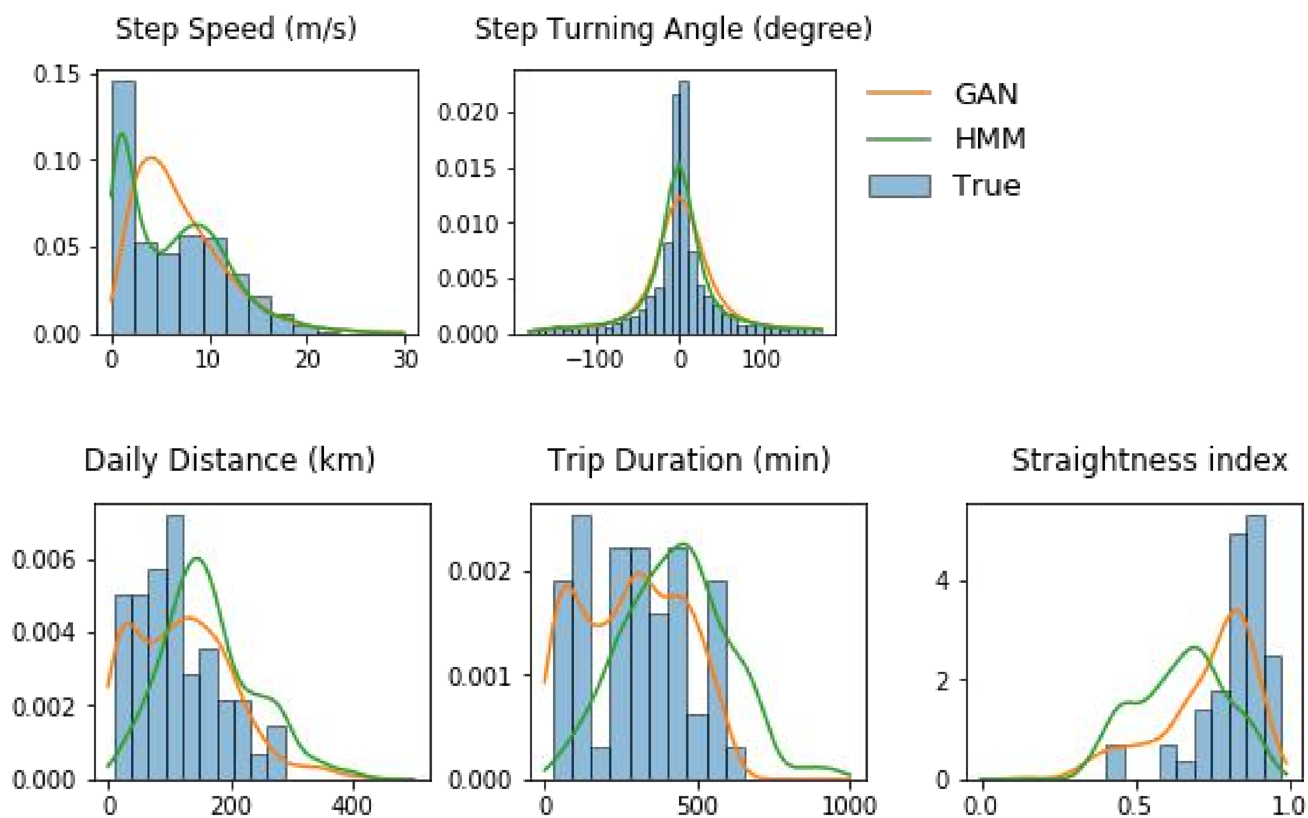
Simulation of foraging trips

→ Result

Peruvian booby
(Sula variegata)

Local statistics

Global statistics



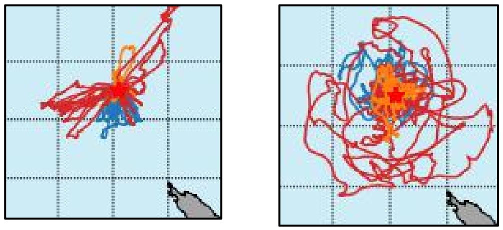
Simulation of foraging trips

→ Example of application

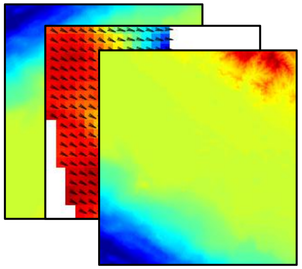
Andreas Ravache
(IRD, UMR Entropie)

<u>Generation of Pseudo-Absence</u>

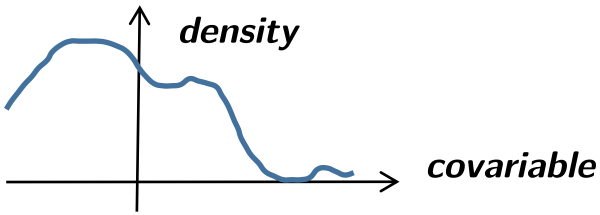
Presence/Absence occurrences



Environmental Data



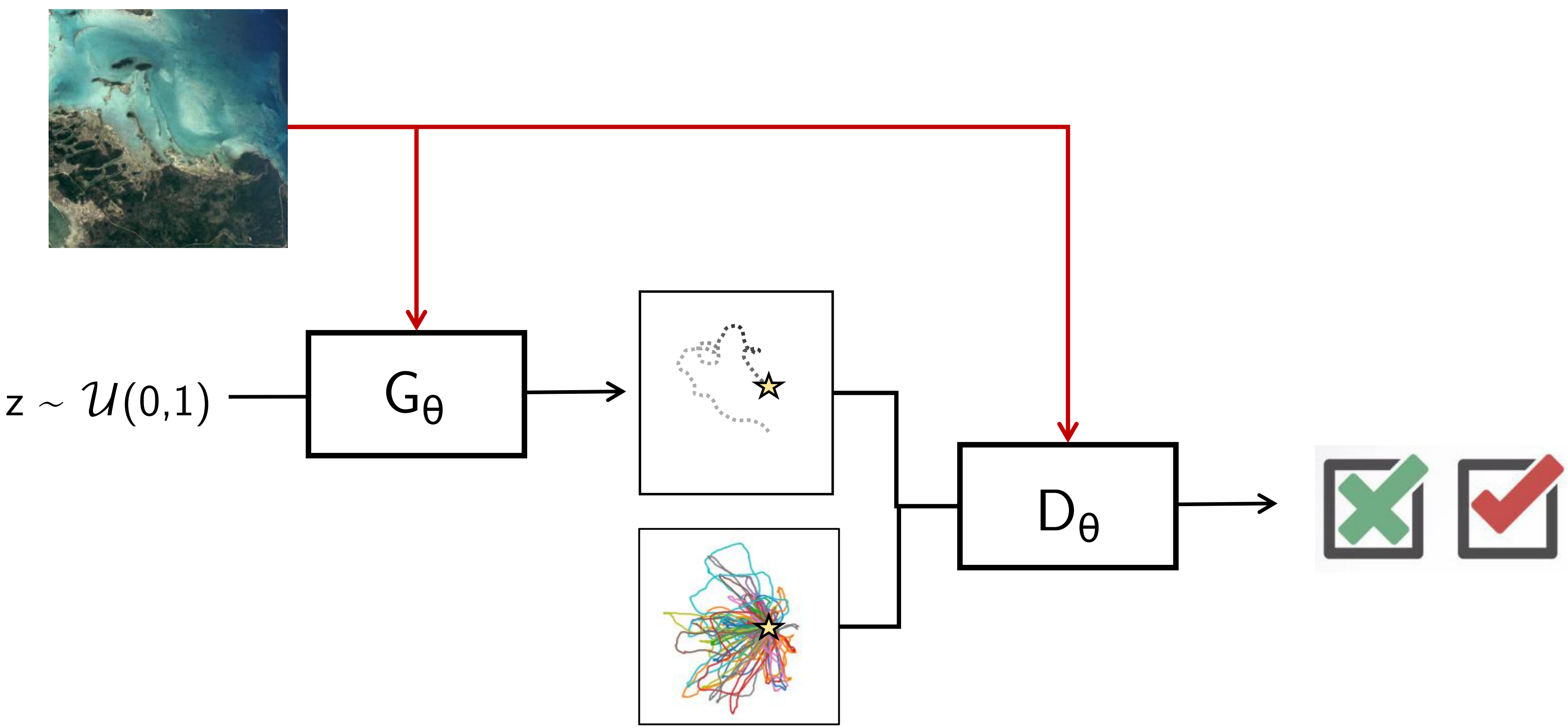
Habitat modelling



Climate-based prediction of trajectories

→ Problem Overview

Environmental data

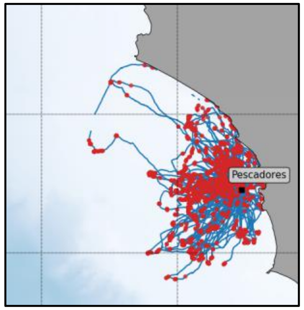


Climate-based prediction of trajectories

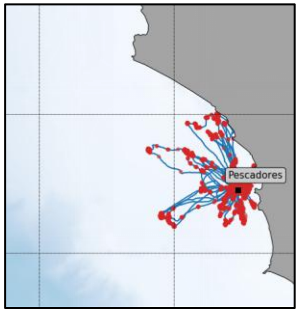
→ Data Overview

Peruvian booby
(Sula variegata)

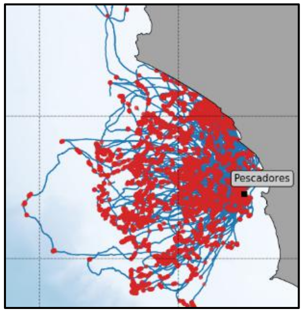
2008
144 trips



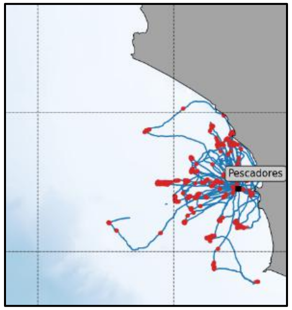
2009
44 trips



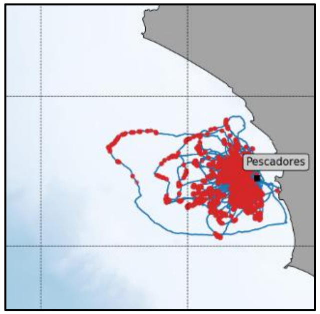
2010
152 trips



2011
36 trips



2012
107 trips



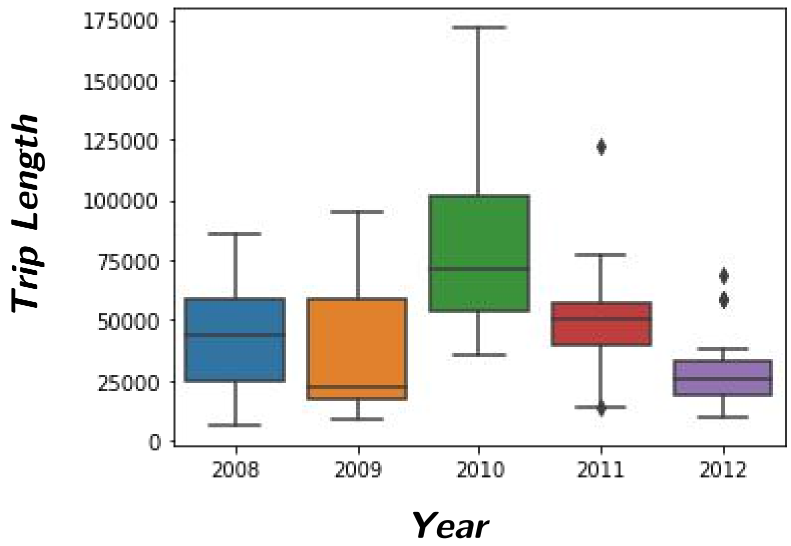
	2008	2009	2010	2011	2012
sst	=	+	-	=	+
chl	+	=	+	=	=
wind	+	=	+	-	=
oxy depth	-	+	=	=	+
biomass	+	-	=	+	-

Climate-based prediction of trajectories

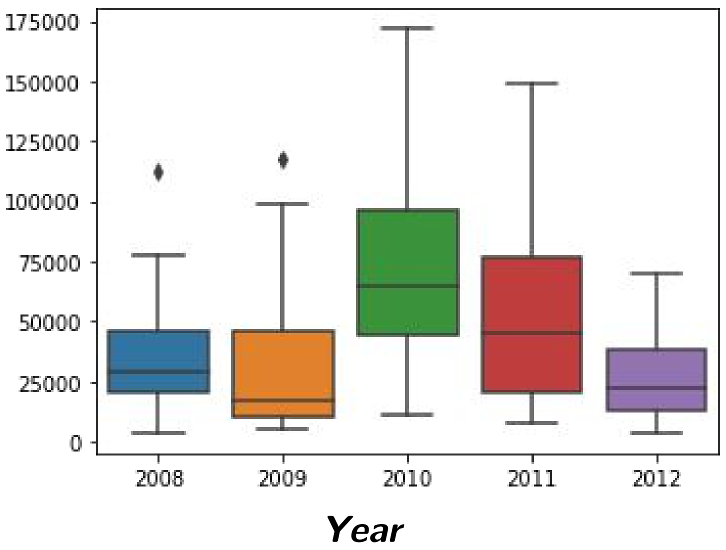
Peruvian booby
(Sula variegata)

→ Results

True data



Simulated data



Perspectives

→ Supervized Learning

- Landscape analysis (e.g. classification of habitat)
- Behavioural studies (e.g. track segmentation)
- Population monitoring (e.g. counting individual)

→ Generative models

- Movement modeling (e.g. landscape, social interactions)
- Data interpolation (e.g. upsampling, filling gaps)
- Scenario simulation (e.g. impact of climate change)

Challenges

→ Supervized Learning

- Need of labelled data
- Generability: can I transfer my network for a similar task?
- Explainability: what does my network understand?

→ Generative models

- Oscillation: non-convergence
- Mode collapse: G produces the same few patterns
- Evaluation metrics: how good G approximates the data?

Take-home message

01 **Deep Learning** is a family of tools based on deep networks

02 **Deep Networks** are parametric models that are trained to minimize a certain loss function

03 **Supervized Learning** is used to map an input to output based on input/output pairs

04 **Generative Models** are used to learn data distribution

05 Many applications for **Seabird Ecology**

Thank you for your attention!

Sophie Lanco Bertrand

Ronan Fablet

Antonio Garcia-Quintas

Guilherme Tavares Nunes

Andreas Ravache