L'Institut Agro Rennes-Angers
☐ **Site d'Angers** ☒ **Site de Rennes**

Année universitaire : 2024 – 2025 Spécialité : Sciences halieutiques et aquacoles, préparée à L'Institut Agro Rennes-Angers Spécialisation (et option éventuelle) : EcolH	**Mémoire de fin d'études** ☐ d'ingénieur de l'Institut Agro Rennes-Angers (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement) ☐ de master de l'Institut Agro Rennes-Angers (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement) ☐ de l'Institut Agro Montpellier (étudiant arrivé en M2) ☒ d'Ingénieur d'AgroParisTech

Evaluating the geographic transferability of deep learning models: the case of dugong detection in aerial images

Par : Caroline ROFFI

Soutenu à Rennes ***le 11/09/25***

Devant le jury composé de :

Président : Etienne RIVOT

Maître de stage : Laura MANNOCCI, chargée de recherche à l'IRD

Enseignant référent : Etienne Rivot

Autres membres du jury (Nom, Qualité)

Jérôme GUITTON (Enseignant-chercheur de l'Institut Agro Rennes-Angers)

Laeticia CHAPEL (Professeure à l'Institut Agro Rennes-Angers)

Confidentialité

☒ Non ☐ Oui si oui : ☐ 1 an ☐ 5 ans ☐ 10 ans

Pendant toute la durée de confidentialité, aucune diffusion du mémoire n'est possible [1].

Date et signature du **maître de stage** [2] : 25 aout 2025
(ou de l'étudiant-entrepreneur)

A la fin de la période de confidentialité, sa diffusion est soumise aux règles ci-dessous (droits d'auteur et autorisation de diffusion par l'enseignant à renseigner).

Droits d'auteur

L'auteur[3] **Roffi Caroline**
autorise la diffusion de son travail (immédiatement ou à la fin de la période de confidentialité)

☒ Oui ☐ Non

Si oui, il autorise

☐ la diffusion papier du mémoire uniquement(4)

☐ la diffusion papier du mémoire et la diffusion électronique du résumé

☐ la diffusion papier et électronique du mémoire (joindre dans ce cas la fiche de conformité du mémoire numérique et le contrat de diffusion)

☒ accepte de placer son mémoire sous licence Creative commons CC-By-Nc- Nd

Date et signature de l'**auteur** : 20/08/2025

Autorisation de diffusion par le responsable de spécialisation ou son représentant

L'enseignant juge le mémoire de qualité suffisante pour être diffusé (immédiatement ou à la fin de la période de confidentialité)

☐ Oui ☐ Non

Si non, seul le titre du mémoire apparaîtra dans les bases de données.
Si oui, il autorise

☐ la diffusion papier du mémoire uniquement(4)

☐ la diffusion papier du mémoire et la diffusion électronique du résumé

☐ la diffusion papier et électronique du mémoire

Date et signature de l'**enseignant** : 3/10/2025

(1) L'administration, les enseignants et les différents services de documentation de l'Institut Agro Rennes-Angers s'engagent à respecter cette confidentialité.

(2) Signature et cachet de l'organisme

(3) Auteur = étudiant qui réalise son mémoire de fin d'études

(4) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option)) sera signalée dans les bases de données documentaires sans le résumé

Avant-propos

Utilisation d'une Intelligence Artificielle générative (IAg) pour la rédaction du mémoire

Des outils d'intelligence artificielle générative ont-ils été utilisés ?

☐ Non ☒ Oui

Si oui, préciser pour quels usages :

J'ai utilisé l'IAg pour corriger mes codes Python et R (ChatGPT 4), pour m'assister dans la rédaction en anglais (Deepl), ainsi que pour ma bibliographie (Elicit).

ACKNOWLEDGEMENTS

Firstly, I would like to thank Laura Mannocci for giving me the opportunity to work on this project with her and for her support throughout the master thesis. Thank you to David Mouillot for lending me the equipment necessary for the proper conduct of my work and to the LIRMM laboratory for their advice. I would also like to thank Auguste Verdier and Martin Simonoviez for their help with the codes and the support, and my fellow interns from the MARBEC laboratory, Milena, Luka, Lily, Achille, Coline, Théo, Ema, Gaetan, Jeanne, Arthur and Loie who made this experience memorable. Finally, I would like to thank my referent teacher, Etienne Rivot, who helped me a lot during my master thesis, and generally throughout the year.

Table of contents

INTRODUCTION

An accurate knowledge of species occurrences is critical to reliably assess marine biodiversity. However, sampling biases toward more accessible regions compromise the ability to describe species' occurrences accurately (Menegotto, Rangel 2018). Data gaps are particularly apparent in the tropical waters of developing countries, reflecting the often-limited funds and capacities for marine research (Costello et al., 2010). To fill these data gaps, geographic transfers of Species Distribution Models (SDMs) developed from data-rich locations are valuable. Model transferability refers to the degree to which a model built in one place or time can successfully predict species distributions in a different place or time (Rousseau, Betts 2022). The geographic transferability of SDMs has received much attention in ecology (Sequeira et al., 2018; Yates et al., 2018). The transferability of a SDM is negatively affected by incomplete or biased sampling coverage and by model overfitting (Sequeira et al., 2018). It also decreases in case of environmental dissimilarity or when predictions are made in environments that are not represented in the training data (Yates et al., 2018). On the other hand, training a SDM on multiple regions can improve its performance, alleviating its dependency on local conditions (Sequeira et al., 2018). However, the geographic transferability of Convolutional Neural Networks (CNNs), a class of deep learning models that are used to extract species occurrences from images, has been comparatively less studied.

As biodiversity is declining at an unprecedented rate, deep learning models provide effective tools for the rapid assessment and monitoring of animal populations (Tuia et al., 2022). Applied to large amounts of remote sensing data from drones, aircraft or satellites, deep learning models allow detecting, localizing, and classifying species in images collected over increasingly large spatial scales (Tuia et al., 2022). Deep learning models aimed for detection are particularly key to reducing the time and labor associated with manually identifying and counting species in thousands of images. These models, particularly in large-scale biodiversity monitoring projects, dramatically decrease human effort (Delplanque et al., 2023; Willi et al., 2019). For example, in aerial wildlife surveys, deep learning models have been shown to cut down the manual image interpretation workload by up to 74% (Delplanque et al., 2023). Despite their achievements, detection models suffer from important limitations as they need to be trained on large amounts of annotated images to perform well (Kim et al., 2020). For the purpose of object detection, image annotation is especially expensive, as it requires not only identifying the objects but also precisely locating them with bounding boxes. These limitations have triggered a growing demand for transferable detection models, i.e. models capable of generalizing from data-rich regions to new regions with little to no training data (Kim et al., 2020). Generalizability ensures that the model can be effectively applied across diverse datasets and environments (Axford et al., 2024).

Transferring detection models is particularly challenging in the marine environment where conditions are highly dynamic. Changes in sea state, turbidity, and light conditions lead to variations in image characteristics across time and space, potentially impairing species detection (Walker et al., 2024; Minai et al., 2024). When a model trained in one environmental or geographic context is applied elsewhere, it faces a so-called domain shift (Kouw & Loog, 2021). Studies on fish detection (Ditria et al., 2020; Walker et al., 2024) have described such domain shifts, underscoring the influence of

environmental and habitat variability on model performance. However, no study to date has evaluated the geographic transferability of deep learning models for the detection of marine megafauna such as marine mammals, sea turtles and sharks.

Marine megafauna populations are increasingly monitored with aerial surveys from Unoccupied Aerial Vehicles (i.e. drones) or aircraft fitted with cameras. Such digital surveys have provided more accurate population counts than observer-based aerial surveys (Hodgson et al. 2023). Digital surveys also allow more frequent surveys thanks to their lower costs, while offering flexibility to study remote locations (Hodgson, Kelly & Peel, 2013; Christie et al., 2016; Hodgson et al., 2017; Rodofili et al., 2022). To automate image analyses from these surveys, deep learning detection models of marine megafauna have been developed (Mannocci et al., 2022; Woolcock et al., 2022; Gray et al., 2018). However, building databases to train these models is challenging for rare and threatened marine megafauna species that occur in low numbers in remote locations (Mannocci et al., 2022). A typical example is the dugong (Dugong dugon; Müller, 1776), an herbivorous marine mammal from the Indo-Pacific coastal waters globally listed as Vulnerable on the IUCN Red List of Threatened Species (Marsh and Sobtzick, 2019). While large to medium size populations remain in areas such as Australia, the Persian Gulf and New Caledonia, many populations are small and scattered (Marsh et al., 2024). Larger dugong populations from developed countries have been comparatively more studied than small populations from less developed countries like in South Asia (Panyawai et al., 2022). The cryptic and furtive nature of the dugong also makes it an inherently difficult species to detect at sea (Hodgson et al., 2004). This raises a critical question: To what extent can dugong detection models trained on data-rich areas be transferred to under-studied regions with little to no training data?

Based on the dugong case study, we investigated the potential of deep learning detection models to perform across geographic locations. Specifically, we tested the capacity of a dugong detection model trained in New Caledonia, hosting a medium size population, estimated between 649 (± 195 s.e.) and 1227 (± 296 s.e.) individuals from aerial surveys (Cleguer et al., 2017) to detect dugongs in West Papua (Indonesia) where populations are comparatively small and scattered (Purwanto et al., 2020). To do so, we performed two geographic transfers from a source region to a target region: (1) a short-distance transfer within New Caledonia and (2) a long-distance transfer from New Caledonia to West Papua, evaluating the impact of environmental contexts on dugong detection rates. To adapt the model to local environmental conditions, we relied on the principle of fine-tuning—that is providing the model with a limited number of annotated images from the target region. First, we explored how fine-tuning a model trained in the source region with varying amounts of local annotations from the target region affected its generalization capacity. Second, we analyzed how environmental and habitat contexts, both during model fine-tuning and inference, influenced dugong detection.

Recent studies showed that species detection models perform better when applied to locations with similar environmental conditions (Bravo-Diaz et al., 2024) and that fine-tuning with even a small number of local annotations can significantly improve species detection rates (Weinstein et al., 2022). Moreover, environmental variables such as water turbidity, sea state, sun glitter, the animal's depth in the water column, age class, group size and light affect detection rates of marine megafauna from aerial images (Fuentes et al., 2015; Aniceto et al., 2018; Dujon et al., 2021; Delplanque et al., 2023). Building on these findings, we formulated the following hypotheses:

(1) Dugong detection rate is higher for the short-distance transfer within New Caledonia than for the long-distance transfer to West Papua, due to greater environmental similarity.

(2) Increasing the amount of local annotations via fine-tuning improves dugong detection.

(3) Environmental and habitat contexts, both during fine-tuning and inference, significantly influence dugong detection.

MATERIALS AND METHODS

Dugong aerial surveys

Aerial surveys were conducted respectively in New Caledonia, an oversea territory of France located in the Coral Sea and in West Papua, an eastern province of Indonesia (Figure 1).

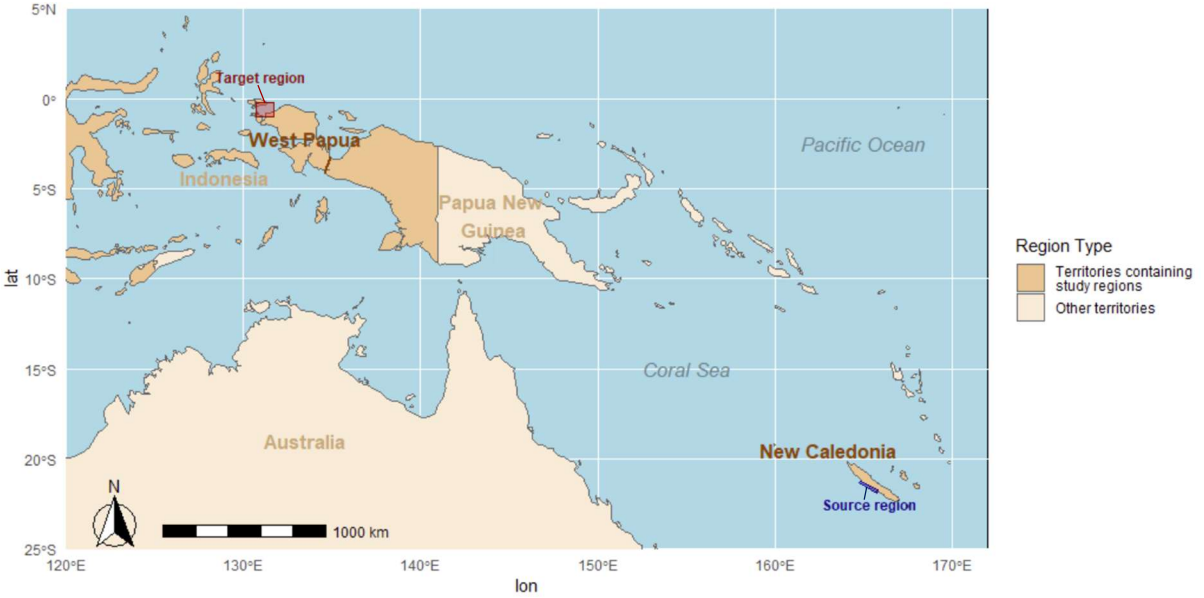


Figure 1. Map of the long-distance geographical transfer. The source region corresponds to the central west coast of New Caledonia, and the target region corresponds to the Bird's Head Seascape in West Papua, Indonesia.

The first survey was carried out from a manned ultra-light airplane equipped with a camera on the central west coast of New Caledonia (Mannocci et al., 2024). This area represents the core distribution range of dugongs in New Caledonia and hosts important aggregations in the austral winter (Mannocci et al., 2024; Cleguer et al. 2024), estimated between 649 (± 195 s.e.) and 1227 (± 296 s.e.) individuals from aerial surveys (Cleguer et al., 2017). The survey region included the area of Poé, as well as the wider west area divided into four sites: "Extreme North", "Mid-North", "Mid-South" and "Extreme South" (Figure 2). Each of the four sites was surveyed by at least two flights from June to November 2021 using an AirMax SeaMax airplane equipped with a GoPro Hero Black 7 camera. The camera was recording in video mode at a rate of 24 frames per second and with an image resolution of 2704 x 1520 pixels. Predefined transects were flown at a target altitude of 47 m and a speed of 110 km/h (Heudier et al., 2023). At this altitude, each image covered a surface of 89 m width (the 'strip width') × 50 m length, leading to an overall surveyed area of 62 km² (Mannocci et al., 2024). A total of 27 flights were flown. All flights are considered independent, as they took place on different days and/or at different locations.

From a geomorphological perspective, the central west coast of New Caledonia is primarily characterized by shallow lagoons, followed by the presence of reef flats (Appendix 1).

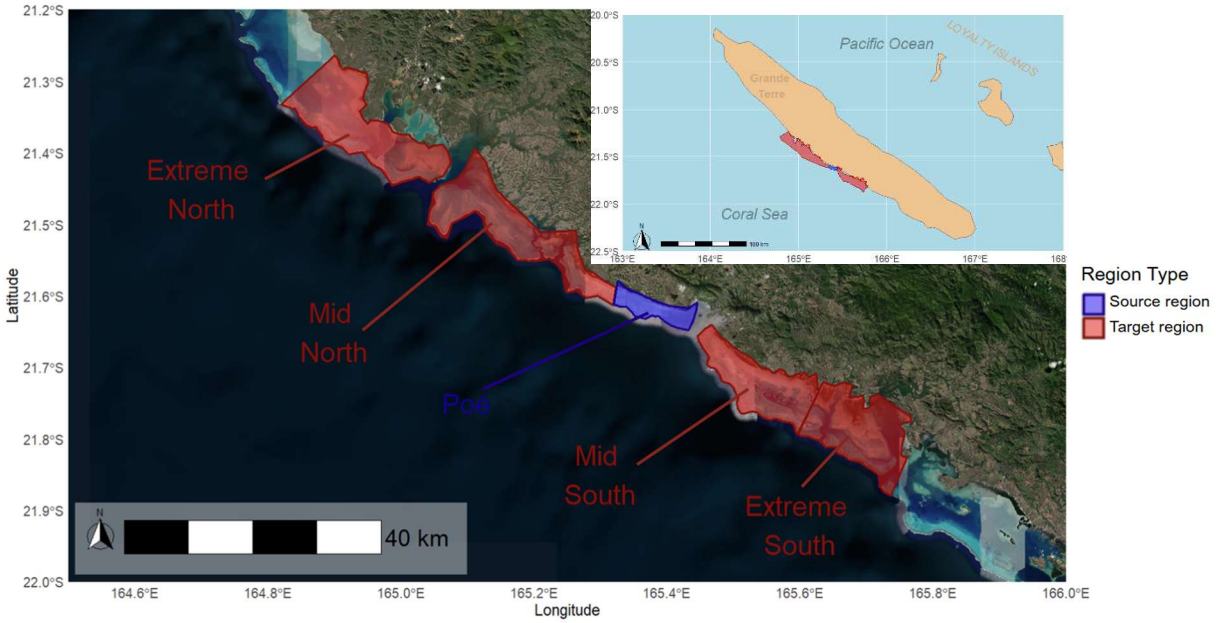


Figure 2. Map of the short-distance geographical transfer. The area surveyed by airplane included Poé (in blue), corresponding to the source region, and the wider west area (in red), corresponding to the target region. Satellite basemap from 'Esri.WorldImagery'.

The second survey was carried out by drone in West Papua, specifically in the Bird's Head Seascape known as the epicenter of global marine biodiversity (Purwanto et al., 2021). Although the dugong is known to occur in this area (Purwanto et al., 2021), no information is available on the population abundance or distribution. The survey was conducted from April – May 2023 and in June 2024, using drones, at four sites: Gam Island, Friwen Island, Manta Sandy and Um Island (Figure 3). Surveys were flown in both automated and manual modes using DJI Phantom 4 (P4) and DJI Phantom 4 Pro (P4P) drones. Automated flights consisted in surveying transects to fully cover a given area, here ranging from 0.2 to 1.5 km². They were conducted at a speed of 7 m/s, an altitude of 50 m and with 10 % side and front image overlap. Manual flights consisted in remotely tracking dugongs first spotted during exploratory search flights. Their altitudes ranged from 35 to 50 m and their speed ranged from 0 to 10 m/s. In both automated and manual flights, the camera was pointing straight downward (i.e. had a nadir position). Image resolution was 3,648 x 5,472 pixels in flights recorded with the P4P drone and 3,000 x 4,000 pixels in flights recorded with the P4 drone. Flights tracking the same individuals or taking place at the same time and place were grouped under a specific "mission". All missions are considered independent as they show different individuals in different places and/or at different times.

From a geomorphological perspective, Manta Sandy and Um Island are primarily characterized by reef slopes, while Gam and Friwen Islands are dominated by reef flats. All four sites are predominantly classified as reef slopes. Um Island is the only site that also features areas of shallow lagoon (Appendix 2).

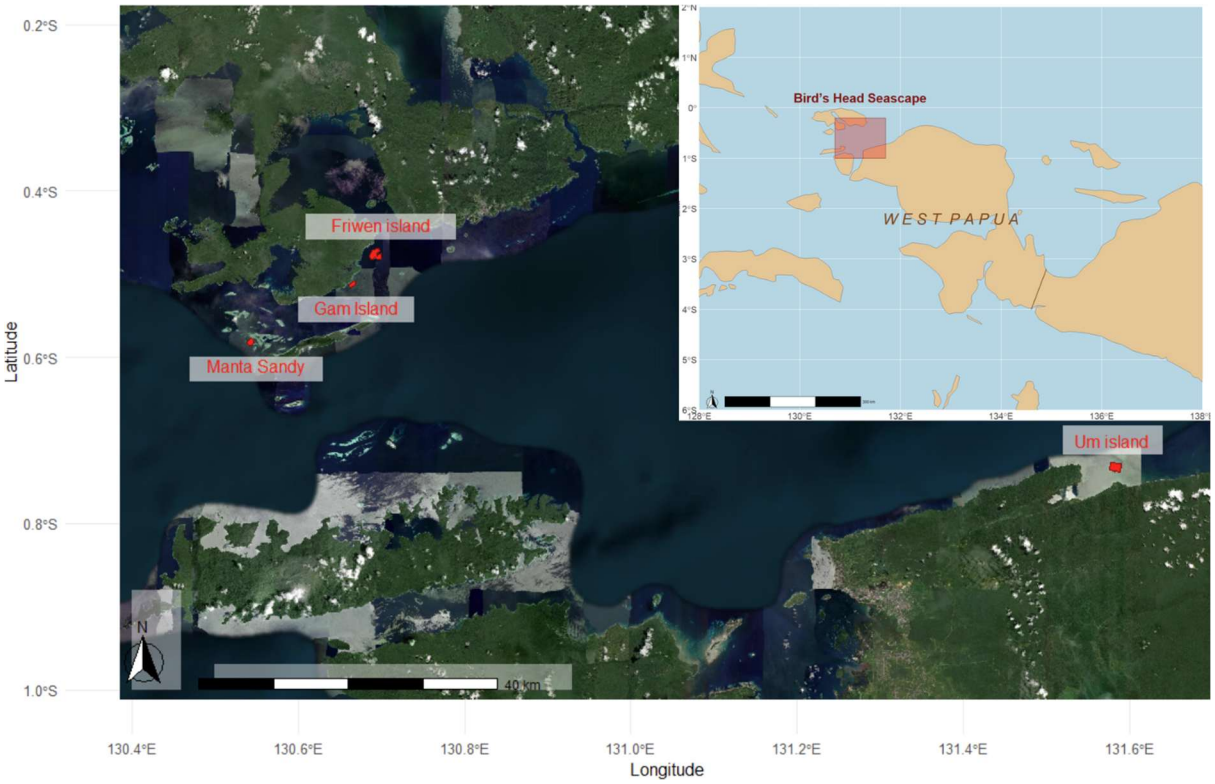
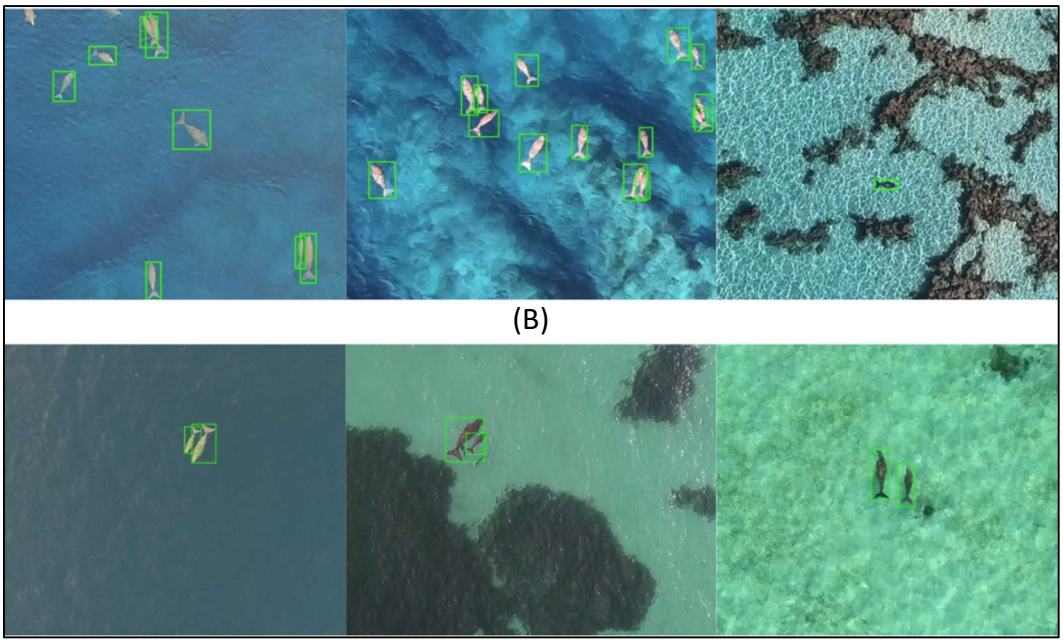


Figure 3. Map of the study region and sites surveyed by drone in West Papua. Satellite basemap from 'Esri.WorldImagery'.

Dugong annotations

Videos and images were imported into 'WAVE' (https://megafauna-project.ird.fr/), a customonline software for marine megafauna annotation. Videos were first converted to images at a rate of 1 frame per second. The annotation procedure consisted in manually drawing a rectangle bounding box around each dugong, provided at least 75% of its body was visible, and associating it a label (Figure 4) (Mannocci et al., 2024). Dugongs were initially labelled as 'probable' and 'certain' depending on how clearly they appeared on images. However, to maximize the dataset, certain and probable dugongs were lumped together under the 'Dugong' label. All annotations were checked by an independent expert observer.

(A)



(B)

Figure 4. Examples of annotated dugongs in images from (A) New Caledonia and (B) West Papua, highlighting the diversity of environmental contexts.

Context variables

In order to determine how the environment influenced model transfer, we defined context variables associated with each annotated dugong and each image (detailed in Table 1). These variables described both the individual-scale contexts (e.g., dugong's position in the water column) and the image-scale contexts (e.g., sea state). Each individual dugong was treated as a distinct instance, allowing us to capture fine-scale context variations within an image, including cases where multiple dugongs appear under different conditions.

Table 1. Description of context variables.

Variable	Levels	Explanations	Scale
habitat_type	coral	major habitat type (coverage) present on the image's background	image
	sand		
	dense_seagrass		
	sparse_seagrass		
	deep_water		
background_complexity	low	flat monotone uniform background (McSpadden et al., in prep)	
	medium	mostly flat heterogeneous background with scattered protrusions (McSpadden et al., in prep)	
	high	heterogeneous background with important changes in topography (McSpadden et al., in prep)	
sea_state	glassy	glassy smooth sea surface	
	rippled	sea surface altered by wind	
sun_glitter	0	percentage of image covered by sun glitter i.e. by the sparkling reflection of the sun on the water (Hodgson et al., 2013)	
	0-25		
	25-50		
	>50		
cloud_reflection	0	percentage of image covered by cloud reflection on the water	
	0-25		
	25-50		
	>50		
calf_presence	yes	at least one calf is present on the image (Hodgson et al., 2023)	
	no	no calf is present on the image (Hodgson et al., 2023)	
group	alone	only one individual is present on the image	
	sparse	two adults or 3 individuals are present on the image	
	school	4 or more individuals are present on the image	
	mother_calf	one mother and its calf (only) are present on the image	
turbidity_global	1	shallow and bottom clearly visible (Hodgson et al., 2013)	
	2	shallow and bottom obscured by turbidity (Hodgson et al., 2013)	
	3	deep and clear water (bottom not visible) (Hodgson et al., 2013)	
	4	deep and turbid water (bottom not visible) (Hodgson et al., 2013)	
bottom_visible	yes	derived from turbidity_global, corresponds to level 1, bottom clearly visible	
	no	derived from turbidity_global, corresponds to levels 2, 3 and 4, bottom not clearly visible	
depth	shallow	derived from turbidity_global, corresponds to levels 1 and 2, bottom distinguishable	
	deep	derived from turbidity_global, corresponds to levels 3 and 4, bottom not visible and deep	
coral	P	presence of coral on the image	
	A	absence of coral on the image	
sand	P	presence of sand on the image	
	A	absence of sand on the image	
dense_seagrass	P	presence of dense seagrass (only UM island) on the image	
	A	absence of dense seagrass on the image	
sparse_seagrass	P	presence of sparse seagrass on the image	
	A	absence of sparse seagrass on the image	
background_dugong	coral	dugong's local background	individual
	sand		
	dense_seagrass		
	sparse_seagrass		
	deep_water		
	dugong		
deep_water	P	presence of areas where the bottom is not visible because of depth	
	A	absence of areas where the bottom is not visible because of depth	
dugong_superposition	yes	the indivudal is positioned on top of another dugong	
	no	the individual is not positioned on top of another dugong	
dugong_posititon	surface	dugong's body clearly visible, at the surface	
	water_col	dugong's body quite visible, but colour or silhouette altered by the water column	
	bottom	dugong's body difficult to delimit beacause of overlying water column	
calf	yes	the individual is a calf	
	no	the individual is an adult	

Deep learning models

General approach

The study aims to train a dugong detection model (referred to as the **baseline model**) in a **source region** and to assess its capacity to detect dugongs in a **target region**. The baseline model is then **fine-tuned** on the **target region**, and its performance improvement is evaluated as it adapts to local conditions.

Two geographical transfers were done: (1) A **short-distance transfer** within New Caledonia, including the Poé area as the source region and the wider west area as the target region (Figure 2) and (2) A **long-distance transfer** including the whole west area of New Caledonia as the source region and West Papua as the target region (Figure 1). We hypothesized that the short distance transfer would be more successful as the target and source regions are geographically, geomorphologically and environmentally similar.

YOLO detection model

We used the You Only Look Once (YOLO) model (Redmon et al., 2016) implemented in PyTorch within the Ultralytics package (https://github.com/ultralytics/ultralytics). YOLO is an open-source detection model (Appendix 5) pre-trained on the Common Objects in COntexts (COCO) dataset (https://github.com/ultralytics/ultralytics/blob/main/ultralytics/cfg/datasets/coco.yaml). YOLO unifies the region proposal and classification stages of CNNs into a single, end-to-end architecture, simultaneously predicting bounding boxes and class probabilities for those boxes. YOLO trains on full images and directly optimizes detection performance. Indeed, this holistic approach allows it to learn from the overall context of the objects and their relationships within the image, maximizing its ability to accurately detect objects in a single pass (Redmon et al., 2016). Several subsequent versions of YOLO have been introduced; in this study we used version 8 medium (named YOLOv8m). During training, YOLO automatically saves the set of weights that yields the best performance on the validation data. The model selected is the one with the weights maximizing the mAP50 (mean Average Precision), a standard evaluation metric where detections are considered correct if the predicted bounding box overlaps with the ground-truth box by at least 50% of their area.

Cross-validation

To make the best use of our limited data and reduce bias introduced by any single split, **a k-fold cross-validation** was implemented. The target data was divided into three sets: a **training** set, a **validation** set to tune hyperparameters during model training while avoiding overfitting, and a **test** set to evaluate the model's performance on unseen images. The target data was divided into several partitions, ensuring that each partition is used once as validation set and once as test set while the remaining partitions are used for training (Figure 5). To avoid model biases from unbalanced datasets, flights (in New Caledonia) and missions (in West Papua) were partitioned to obtain equal numbers of

dugong annotations. This approach helped us ensure fair and consistent model evaluation across all folds, minimizing the risk of overfitting to a particular data subset.

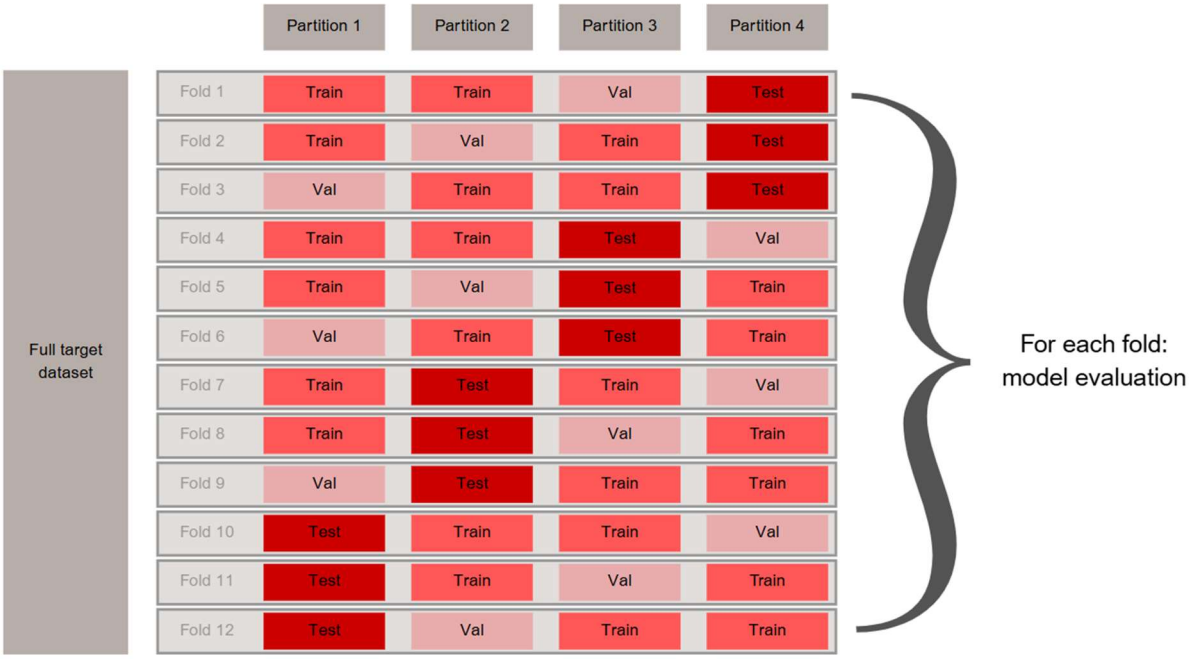


Figure 5. K-fold Cross Validation principles for an example with 4 partitions for the target region. A fold corresponds to a combination of partitions used for the training set, the validation set and the test set.

To evaluate the model's performance improvement as it adapts to local conditions, we fine-tuned (i.e. trained) the baseline model on **subsets of the target datasets**, containing 0%, 20%, 40%, 60%, 80%, and 100% of images taken randomly from available flights/missions (Figure 6).

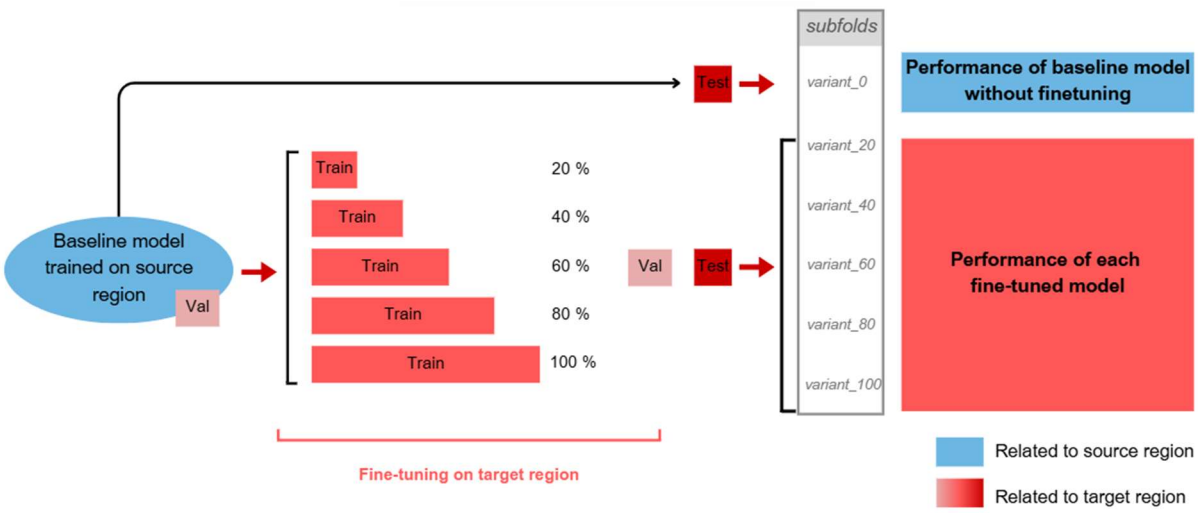


Figure 6. Fine-tuning process for a given fold.

Consequently, for a target dataset with n partitions, we obtained $\binom{n}{n-2} \times 2 \times 6 = n \times (n-1) \times 6$ **subfolds**. The optimal number of partitions for a particular target dataset was the number that led to equal dugong annotations between partitions. For the short-distance transfer (within New

Caledonia), the optimal number of partitions was 4 so we had 72 folds (meaning that we trained, validated and tested 72 models). For the long-distance transfer, the optimal number of partitions was 5 so we had 120 folds (meaning that we trained, validated and tested 120 models).

The detection models were trained on the training datasets and their hyperparameters, that control the training process, were simultaneously tuned on the validation dataset. The tuning of hyperparameters, including data augmentation, batch size, and initial learning rate was guided by prior empirical work on dugong detection in New Caledonia (Mannocci et al., in prep) (Appendix 3). The input image size of YOLO, i.e. the resolution to which input images are resized during training, was set to 1280 x 1280 pixels. This value represented a compromise between computational time and detection accuracy, as dugongs appeared as relatively small objects in image. Non-Max Suppression (MNS) was set at 0.5, meaning that predicted bounding boxes overlapping over 50 % were considered the same, and only the one with the highest confidence score was retained, ensuring that the most likely detection was kept for each annotated object. Training was performed for a maximum of 400 epochs, where one epoch corresponds to a complete pass through the training dataset. We applied an early stopping criterion with a patience of 100 epochs, meaning training would stop if the validation performance failed to improve for 100 consecutive epochs. The final learning rate, i.e. the factor by which the initial learning rate is multiplied at the end of training, was set at 0.01 and 0.001 respectively for training models in New Caledonia and West Papua. This empirically chosen setting allowed the model sufficient time to converge while avoiding unnecessary overfitting. Model optimization was guided by three YOLO loss functions: the box loss, the class loss, and the Distributed Focal (DF) loss. The box loss evaluated how closely the predicted bounding boxes matched annotated objects. The class loss quantified the model's ability to correctly classify objects within those boxes. The DF loss, a refined version of the box loss, emphasized difficult-to-learn bounding box edges, improving precision by assigning greater weight to hard examples.

Baseline models

Baseline models were first trained for each transfer scenario (long-distance and short-distance). For a given scenario, the entire corresponding source dataset was used for training. Each model was simultaneously validated against a non-overlapping partition of the target dataset (Figure 7). For the short-distance transfer, 4 baseline models were trained on Poé and validated against the 4 distinct target partitions in the wider west coast of New Caledonia. For the long-distance transfer, 5 baseline models were trained in New Caledonia and validated against the 5 distinct target partitions in West Papua.

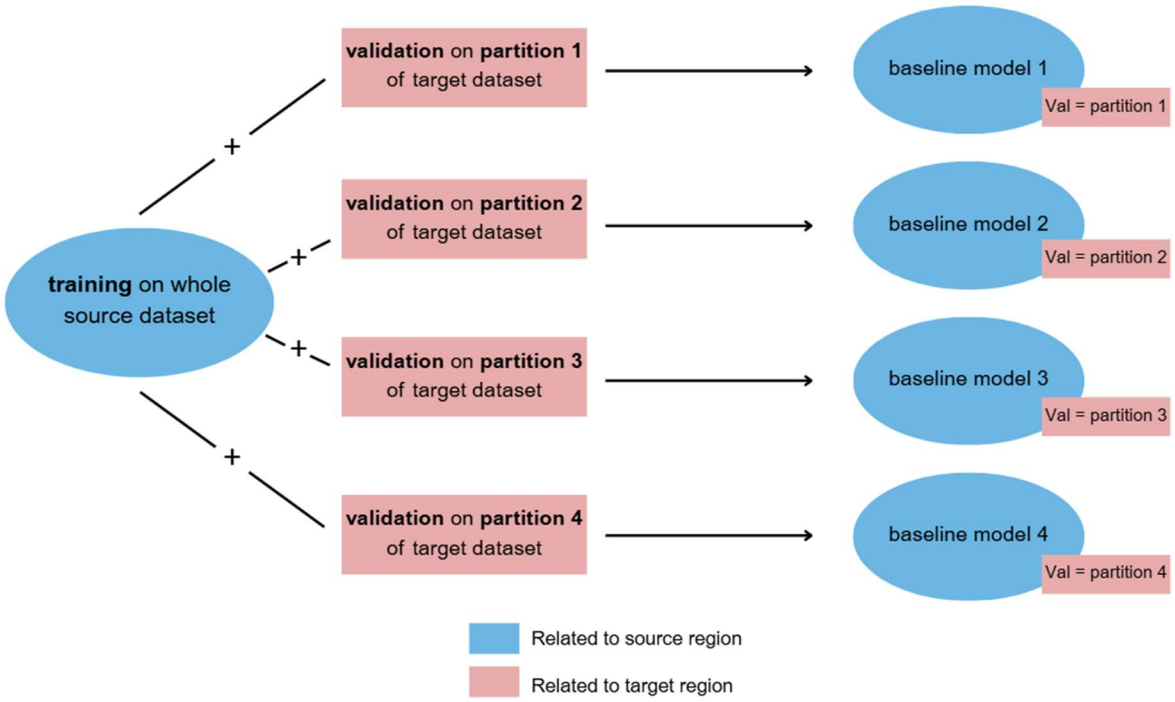


Figure 7. Training and validation of the baseline models. In this example, the target dataset is divided in 4 partitions, leading to 4 baseline models.

Fine-tuned models

Each baseline model was then **fine-tuned on the target dataset**, i.e. each model was trained and validated according to its subfold (Figure 6). The first 10 layers of the models, corresponding to the backbone of YOLO were frozen. Model inference was then done on all subfolds of the test set, resulting in 72 inferences for the short-distance transfer and 120 inferences for the long-distance transfer. The Intersection Over Union (IoU) (Figure 8) threshold was set to 0.5. The confidence threshold, i.e. the model's probability that a detected object actually belongs to a given class (e.g., here a dugong) was set to 0.1. We chose lenient thresholds for both the IoU and the confidence score to minimize false negatives, given the rarity of dugongs and the importance of retaining even uncertain detections for expert post-validation. So, in our study, a dugong was correctly predicted if its IoU was ≥ 0.5 and its confidence score was ≥ 0.1.

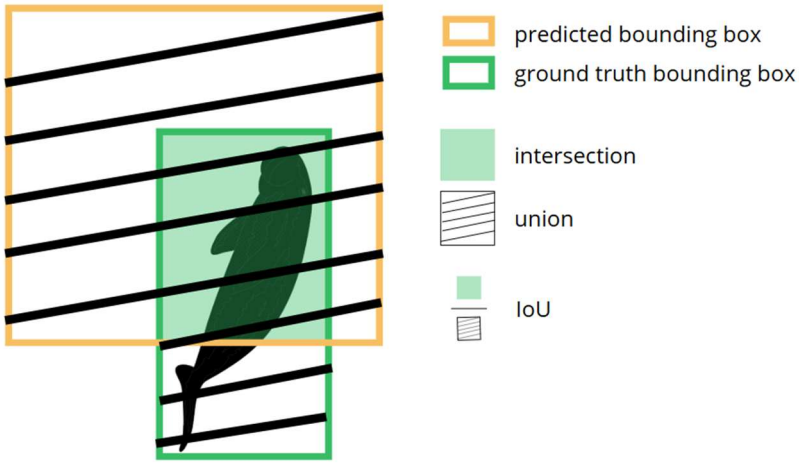


Figure 8. Illustration of the Intersection over Union (IoU) i.e. the ratio of the intersection between predicted bounding box and the annotated bounding box (green area) to the union of the two bounding boxes (striped area).

All detection models successfully converged during training. The number of epochs allowed sufficient time for loss minimization across all three loss functions, and validation metrics remained stable throughout the training process.

Performance evaluation

Dugongs correctly predicted by our model were **True Positives** (TP) (Figure 9). Conversely, annotated dugongs that were not predicted by our model were **False Negatives** (FN) (Figure 9).

(A) (B)

Figure 9. Example of a True Positive and a True Negative. (A) Example of a True Positive. The green rectangle represents the bounding box predicted by the model. (B) Example of a False Negative.

To assess the model's performance, the **recall** was calculated for each subfold, i.e. model (n= 72 for the short-distance transfer and n= 120 for the long-distance transfer). The recall measures a model's ability to detect all relevant objects within a dataset, i.e. the detection rate. It quantifies the proportion of annotated objects that were correctly predicted by the model (Equation 1).

Equation 1

$$Recall = \frac{TP}{TP + FN}$$

Statistical models

Statistical models were used to assess (1) **the recall at the subfold scale** and (2) the **probability of correct** detection at the individual scale as a function of the environmental contexts characterizing the training data and the dugongs on inference images.

Modelling of recall

The **recall** was modeled as a function of the **number of training annotations** and **the number of training sites** in the target region, as well as the frequencies of context variables (Table 1) in each training set of each subfold. The recall is a continuous variable that ranges between 0 and 1, so it was modeled using a beta regression with a logit link function. The context variables modalities were recoded into binary variables (Appendix 4) for the sake of simplification and interpretability, grouping similar modalities together. The frequency of one modality in the training set of a subfold, whether the variable is at the individual scale or at the image scale, is calculated for individuals as (Equation 2).

Equation 2

$$f_{modality} = \frac{number\ of\ individuals\ with\ modality}{number\ of\ individuals}$$

We refer to this model as "Recall model". At the subfold scale, as the context variables frequencies account directly for the environmental dependency between train sets and indirectly between test sets, we did not add random effects for the train set ID and the test set ID.

Recall model:

recall ~ Beta(μ. ϕ, (1-μ).ϕ)

$$logit(\mu) = \beta_0 + \ \mathrm{poly}\left(\mathrm{nb}_{\mathrm{annotations}_{\mathrm{train}}}, \mathrm{n1}\right) \ + \ \mathrm{poly}(nb\_sites\_train, n2) + \sum_j freq_{var_j}$$

μ represents the expected mean of the recall. ϕ is a constant precision parameter estimated across all observations. β_0 is the intercept, i.e. it represents the baseline logit-odds of recall when no fine-tuning was done. The polynomials degrees of the number of annotations and the number of sites are respectively n1 and n2 ∈ {0, 1 or 2}.

Modeling of probability of detection

The probability of a correct dugong detection was modeled as a function of the context variables at the individual scale (Table 1) characterizing the dugong on which the inference was done, as well as the numbers of annotations and sites used for training. The response variable, 'detection', is binary, taking the value of 1 if an annotated dugong is detected, 0 if not. Thus, a logistic regression with a binomial family and a logit link was used. Interactions between context variables were added to account for potential combined effects.

The train set ID was added as a random effect in the model to account for the environmental dependency of the images used in the training. Here the dependency between the train set and the test set is no longer relevant, as they are at different scales in this model; the train information is still

at the subfold scale and the test variables are at the individual scale. We refer to this model as "Probability of detection model".

Probability of detection model, with k categorical context variables, each taking levels (modalities) $i_1,i_2,\ldots,i_k$ respectively, modeled the probability of detection $p_{i1,\,i2,\,\ldots,\,ik}$ as follows:

Probability of detection model:

$$\text{detection}_{i1,\,i2,\,\ldots,\,ik} \sim \text{Bernoulli}(p_{i1,\,i2,\,\ldots,\,ik})$$

$$logit(p_{i1,i2,\ldots,ik}) = \beta_0 + \text{poly}(\text{nb\_annotations\_train}, \text{n1}) + \text{poly}(nb\_sites\_train, n2) + \sum_{j=1}^{k} \beta_{i_j}^{(j)} + \sum_{1 \le l < j \le k} \beta_{i_j i_l}^{(j,l)} + (1 \mid train)$$

In this model, we consider k categorical context variables. Each variable j∈{1,…,k} takes a value i_j corresponding to one of its levels. The terms $\beta_{i_j}^{(j)}$ represent the effects of level i_j of the j-th environmental variable. The interaction terms $\beta_{i_j i_l}^{(j,l)}$ capture the joint effect of the combination of levels i_j and i_l for the j-th and l-th variables, respectively. (1 | train) represents the random intercept for the training set.

General modeling procedure

Generalized linear mixed models (GLMMs) were fitted for both recall and probability of detection models using the package glmmTMB (McGillycuddy et al., 2025) in R. Variable selection followed a three-step process. First, only context variables with a balanced factorial design were included to ensure proper representation across all factor levels. Second, the Variance Inflation Factor (VIF) was computed for each variable to address potential multicollinearity. Variables with VIF values > 5, indicating strong multicolinearity, were excluded. Third, a stepwise variable selection algorithm was applied, combining forward and backward procedures based on p-values, only keeping significant effects (parsimony principle). At each step, variables were added or removed according to their statistical significance (P-value < 0.05 for main effects, P-value < 0.001 for interaction effects). The best model was selected as the one minimizing model deviance, ensuring the best possible fit to the data.

Variable importance

We used the increase of deviance due to the removal of each variable from the full model as a proxy of variable importance, the greater the deviance increase, the more important the variable. Deviance measures the quality of model fit. It is defined as the difference of log-likelihood of the fitted full model and a saturated model that perfectly fits the data (Equation 3).

Equation 3

$$D_{model\_full} = -2 \times (logLik_{model_{full}} - logLik_{mode\ _{saturated}})$$

For main effects in models 2 that included with interactions, we measured the increase of deviance when removing the main effect and any associated interactions.

For beta regressions, i.e. models 1, the saturated model is not defined. So, the increase of pseudo deviance (Equation 4) due to the removal of each variable from the full model was used as a proxy of variable importance.

Equation 4

$$\Delta D_{var_i} = -2 \times (logLik_{model_{full}} - logLik_{model_{without\_var_i}})$$

Marginal effects

Marginal effects of each variable were calculated for an average subfold (models 1 and 2) or an average individual (probability of detection model), ignoring the random effects if present. The quantitative variables were fixed at their mean corresponding to $\beta_j \times \overline{variable_j}$ in the logit scale. For each categorical variable env_j, a weighted average of the levels' estimated effects was computed. This approach allowed us to summarize the effect of each categorical variable while accounting for the prevalence of its levels. The weights corresponded to the empirical frequencies $f_{i_j}^{(j)}$ of each level i_j in the dataset. The resulting marginal effect for variable env_j was given by the mean of the effects of each modality i_j, weighted by their empirical frequencies (Equation 5).

Equation 5

$$\overline{\beta^{(j)}} = \sum_{i_j} f_{i_j}^{(j)} \times \beta_{i_j}^{(j)}$$

RESULTS

Dataset overview

A total of 342,875 images was collected, of which 2,724 contained dugongs, corresponding to 7,109 dugong bounding boxes. The data repartition is given in Table 2.

Table 2. Data repartition between New Caledonia and West Papua.

	New Caledonia	West Papua
Total number of images	340,300	2,575
Nb. of images with dugongs	716	2,008
Nb. of dugong bounding boxes	3,546	3,563

In New Caledonia, 62% of the images were collected over the Poé area, and 38% over the wider west area, representing respectively 2,183 and 1,363 dugong bounding boxes.

Comparisons of transfers

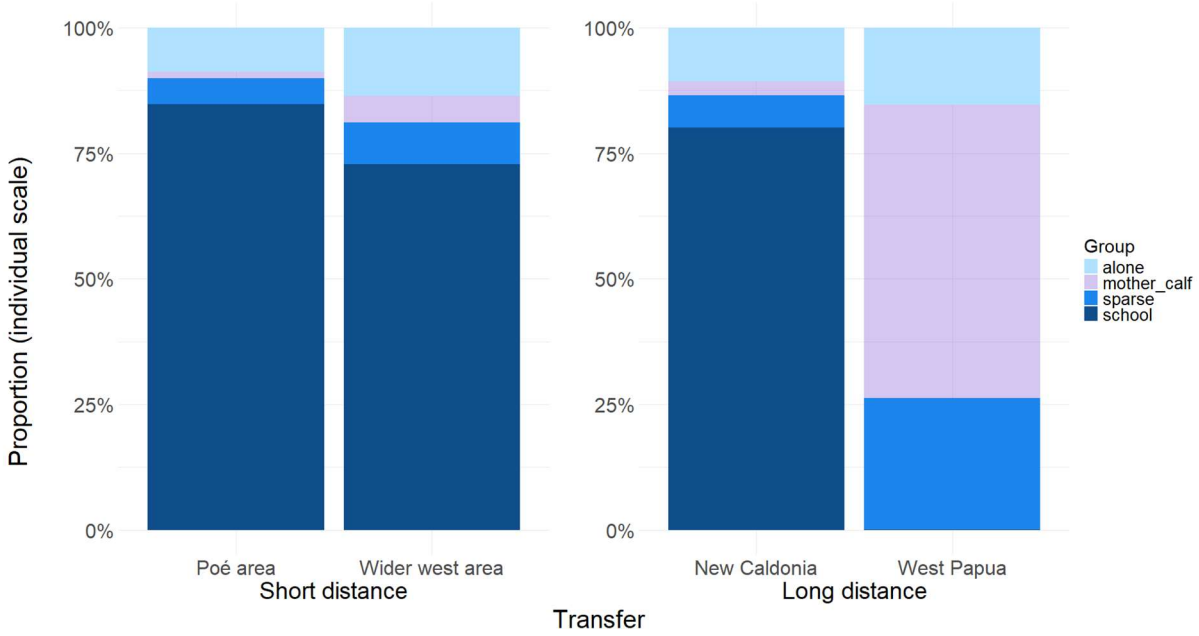


Figure 10. Proportion of observed dugong group compositions in short-distance and long-distance transfers. Levels are described in Table 1.

In the short-distance transfer within New Caledonia, group compositions were similar between source (Poé area) and target (Wider west area) regions (Figure 10). In both regions, over 70 % of dugongs were part of a school (n > 4 individuals), with schools reaching up to 40 dugongs on some images. In contrast, group composition markedly differed between source (New Caledonia) and target (West Papua) regions in the long-distance transfer (Figure 10). Indeed, over 75 % of dugongs in New Caledonia were part of a school whereas most dugongs observed in West Papua occurred as 2 to 3 individuals. Over half of dugongs observed in West Papua were part of a mother – calf pair (Figure 10).

The proportion of calves in West Papua was also twice higher than in New Caledonia (0.33 vs 0.16, χ-squared = 282.05, P-value < 2.2e-16). Conversely, the proportion of calves was slightly lower in the Poé area than in the wider west area (0.18 vs 0.13, χ-squared = 18.267, P-value = 1.92e-05).

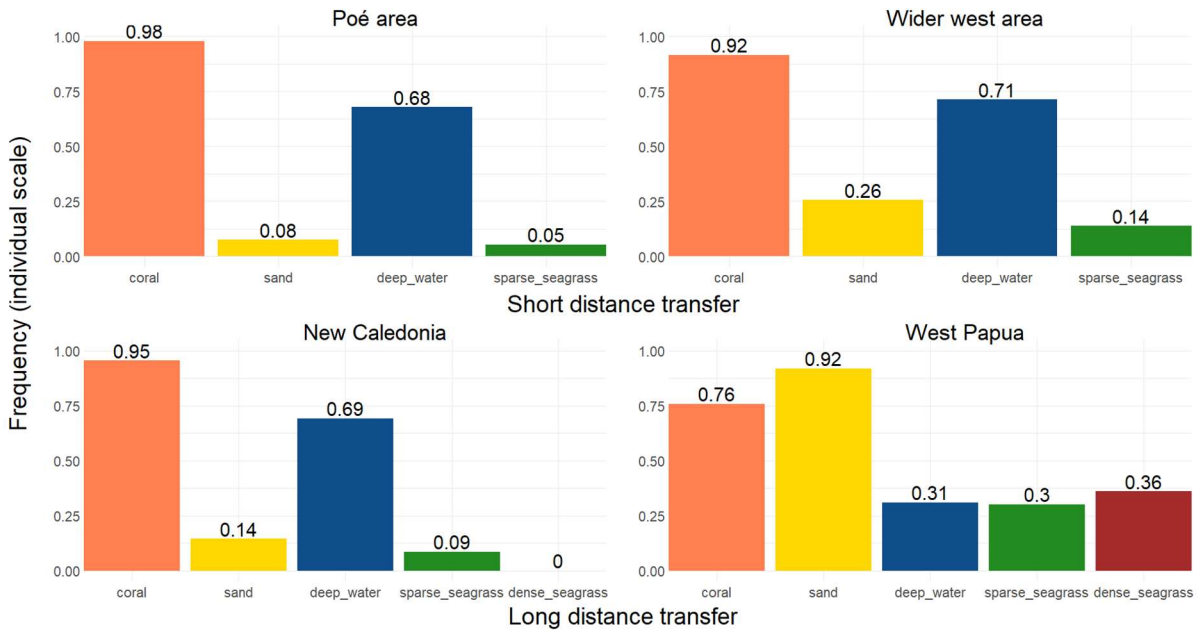


Figure 11. Frequencies of habitat contexts of individuals in the short-distance and long-distance transfers. In the images analyzed, habitats were not mutually exclusive, and several types can be present simultaneously.

In the short-distance transfer, habitat contexts were similar across the source and target regions, with over 90 % of individuals associated with coral and nearly 70 % of individuals associated with deep water (Figure 11). These individuals corresponded to the ones found in schools on barrier reefs (Figure 10). This suggests that the dugongs from both regions occupied comparable habitats.

Conversely, habitat contexts differed widely between the source and target regions in the long-distance transfer (Figure 11). In both regions, more than 75% of individuals were associated with coral (Figure 11). However, in West Papua, over one-third of individuals were associated with dense seagrass while this habitat was absent in New Caledonia (Figure 11). New Caledonia showed strong disparities in occurrence frequencies: most dugongs were associated with coral and deep-water, while fewer than 15% were associated with sand or sparse seagrass. In contrast, West Papua showed a more balanced distribution, with the majority of dugongs associated with sand and coral (Figure 11).

Generalization capacity of detection models

In the short-distance transfer, the recall of the baseline model was 0.61 (± sd 0.14), meaning that on average a model trained in Poé area was able to detect 61 % of the dugongs in the wider west area. Fine-tuning the model in the target region allowed us to reach a recall of 0.66 (± sd 0.11) with 100% of the train set, meaning that the model was able to detect 5 % more dugongs when locally trained. The mean recalls of the baseline and fine-tuned models with 100 % of the train set did not differ significantly (Welch modified Two-Sample t-Test, P-value = 0.27).

In the long-distance transfer, the recall of the baseline model was 0.52 (± sd 0.22), meaning that on average a model trained in New Caledonia was able to detect 52 % of the dugongs in West Papua. Fine-tuning the model in the target region allowed us to reach a recall of 0.70 (± sd 0.22) with

100% of the train set, meaning that the model was able to detect 18 % more dugongs when locally trained. The mean recall was significantly higher in the fine-tuned models with 100 % of the train set (Welch modified Two-Sample t-Test, P-value = 0.011).

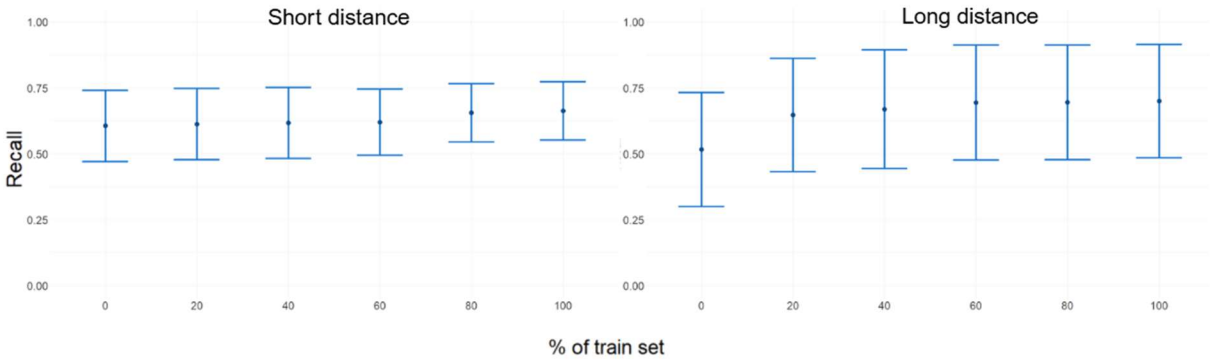


Figure 12. Means (central dot) and standard deviations (error bars) of the recalls for the short-distance and long-distance transfers for different proportions of images of the train sets. Note that the train sets have different sizes in each transfer.

The baseline models of the short-distance transfer (0.61) detected on average 9 % more dugongs than the baseline models of the long-distance transfer (0.52) (Figure 12) However, the two recalls did not differ significantly (Welch modified Two-Sample t-Test, P-value = 0.16). On the other hand, the fine-tuned models with 100 % of the train set performed better in the long-distance transfer (0.70), detecting 4 % more dugongs than the fine-tuned models in the short distance transfer (0.66) (Figure 12). Similarly, the recalls of fine-tuned models did not differ significantly between transfers (Welch modified Two-Sample t-Test, P-value = 0.53).

Influence of train set environmental contexts

To understand how environmental context influences model training, we first modeled the recall as a function of environmental context variables in the target region while controlling for the number of sites and annotations.

The model was fitted on 60 points out of 72, as the fine-tuning environmental context frequencies of the points referring to the baseline models were coded NA. For the short-distance transfer, the best model included only the frequency of dugongs with **visible sea bottom** on the image (Appendix 6), explaining 18.2 % of the variance. The frequency of dugongs with visible sea bottom on the image had a significant positive effect on the recall (P-value = 3.2e-4) (Figure 13).

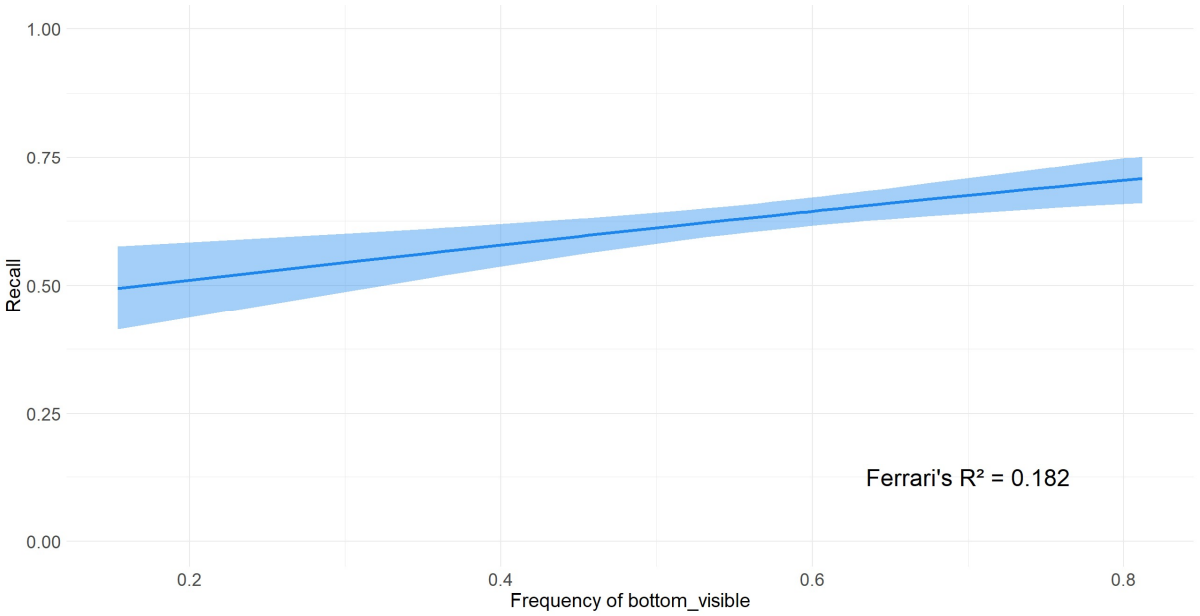


Figure 13. Results of the recall model for the short-distance transfer. Marginal effect of the frequency of dugongs in images with the bottom visible. on the recall. (the shaded area corresponds to the 95 % confidence interval) Because there are only fixed effect in the model, the marginal R^2 is equal to the conditional R^2, summarized under the Ferrari's R^2.

For the long-distance transfer, the best model included the frequency of dugongs with **coral** present on the image and the frequency of **dugongs superposed with another**, explaining about 25.1% of the variance (Appendix 7). The frequency of dugongs with coral present on the image had a significant positive effect (P-value = 8.2e-09) while the frequency of dugongs superposed with another set had a significant negative effect (P-value = 1.4e-3) on the recall (Figure 14).

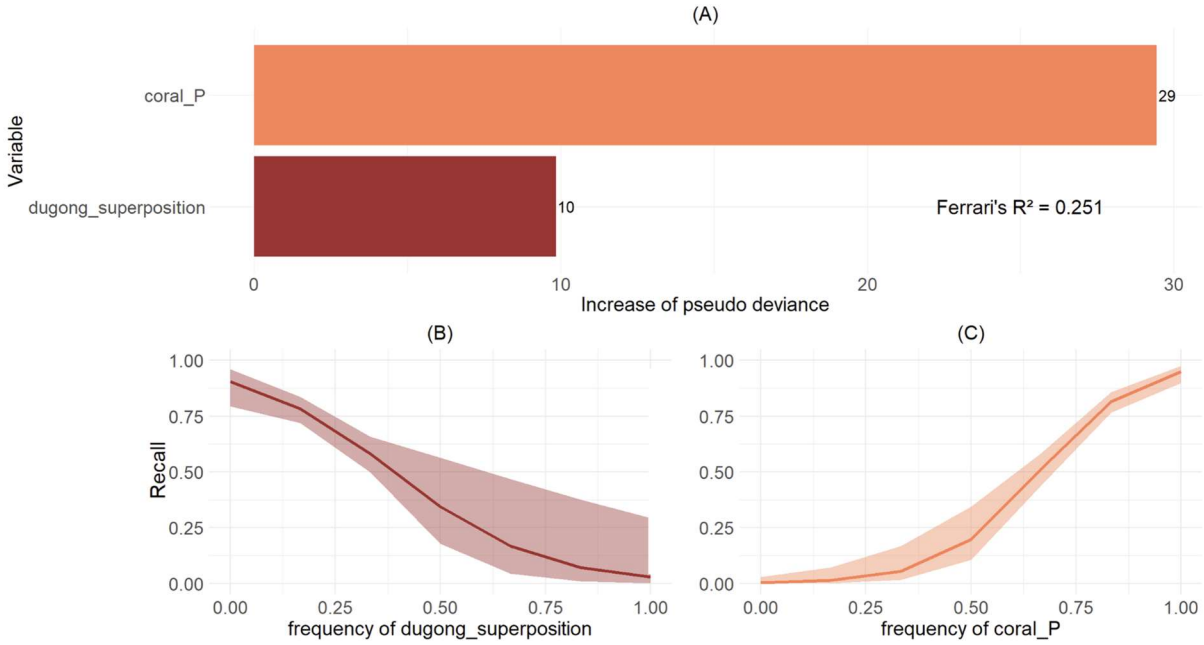


Figure 14. Results of the recall model for the long-distance transfer. (A) Increase of deviance due to the removal of each variable as a proxy of variable importance. (B) Marginal effect of the frequency of dugongs with coral present in the image in the training set on the recall. (C) Marginal effect of the frequency of dugongs superposed on top of another dugong in the training set on the recall. The shaded area corresponds to the 95 % confidence intervals. Because there are only fixed effects in the model, the marginal R^2 is equal to the conditional R^2, summarized under the term "Ferrari's R^2".

Influence of test set environmental context

To understand how environmental context influences inference on the test sets, we then modeled dugong detection probability as a function of the environmental context of annotated dugongs in the test sets while accounting for the number of sites and annotations in the target region.

For the short-distance transfer, the best model retained **"calf", "group", "dugong_position", "cloud_reflection", "sea_state", "sun glitter" and "bottom"** as most important variables and had an explained deviance of 25.5 % (Figure 15). The fixed effects accounted for most of the variance (marginal R^2 ≈ conditional R^2 ≈ 0.44). The model achieved an Area Under the Curve (AUC) of 0.83, indicating a good ability to discriminate between positive and negative dugong detections.

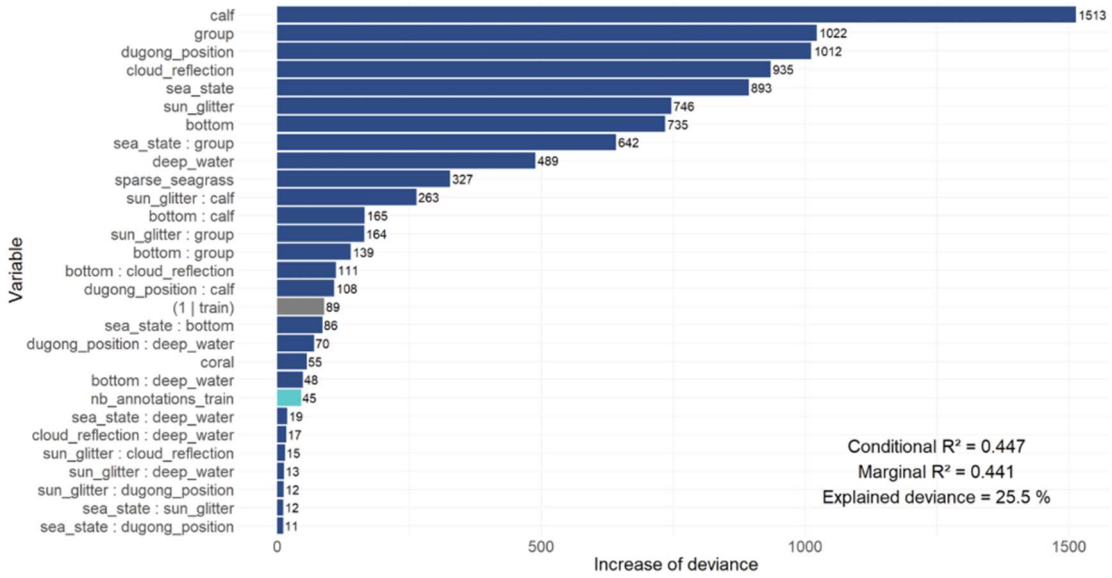


Figure 15. Results of probability of detection model for the short-distance transfer. Increase of deviance due to the removal of each variable as a proxy of variable importance. For the sake of interpretability, we kept only the interaction effects that had a P-value lower than 0.001. Grey is for random effects, fixed effects are in different shades of blue. Dark blue is for environmental context variables, blue lagoon is for the number of annotations used for fine-tuning.

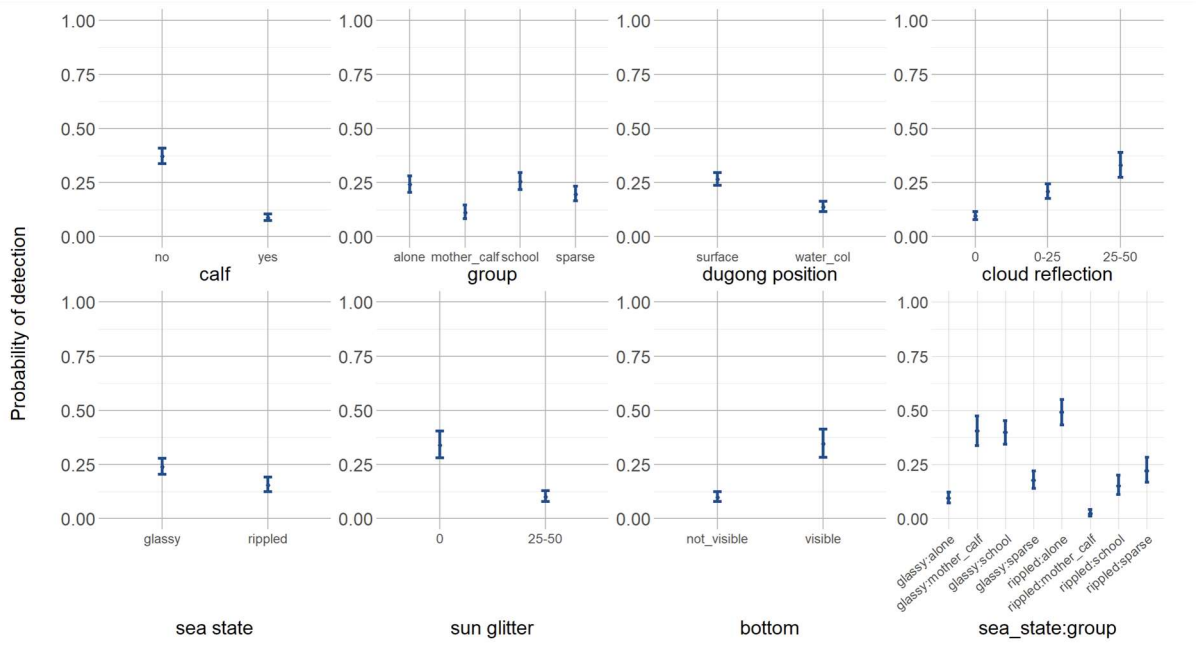


Figure 16. Marginal effects of top 8 most important variables of the probability of detection model in the short-distance transfer (showing an increase of deviance > 500 following their removal).

Calves were poorly detected compared to adults, as on average one out of ten was detected (Figure 16). Dugongs located in the water column were two times less detected than surfacing dugongs. Similarly, dugongs were more poorly detected when the sea surface was rippled. Cloud reflection positively affected dugong detection whereas sun glitter decreased dugong detection. Dugongs were also better detected when the bottom was visible. Generally, the dugongs belonging to mother – calf

pairs (particularly calves)) were more poorly detected than those that occurred alone, within sparse groups or schools. When combining the effects of sea state and group, individuals that occurred alone (single) in calm waters were poorly detected and individuals from a mother–calf pair were almost not detected at all. In contrast, single individuals associated with a rippled sea state were better detected and individuals from of a mother–calf pair or a school were better detected when associated with a glassy sea state, as on average almost half of them were detected.

For the long-distance transfer, the best model retained **"calf", "dugong_position", "sea_state", "group" and "sun_glitter"** as most important variables and had an explained deviance of 34.7 % (Figure 17). The fixed effects accounted for most of the variance (marginal R^2 ≈ conditional R^2 ≈ 0.68). The model achieved an Area Under the Curve (AUC) of 0.89.

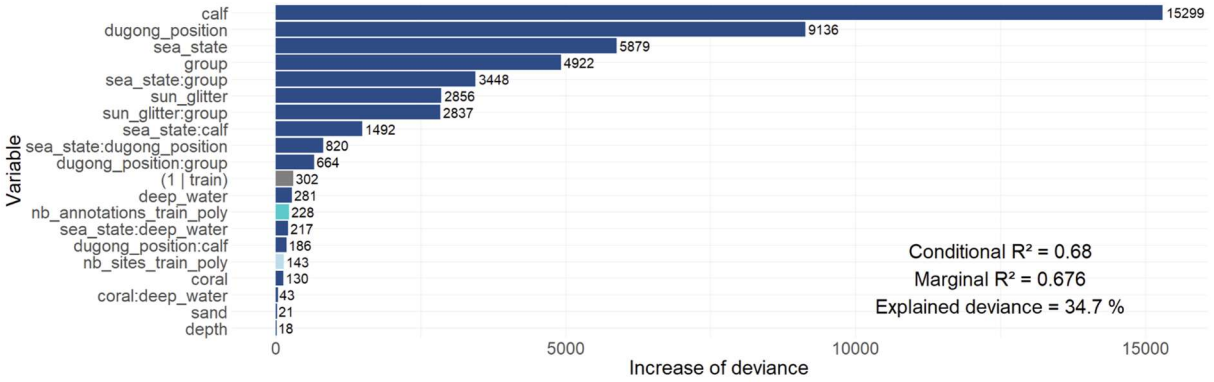


Figure 17. Results of the probability of detection model for the long-distance transfer. Increase of deviance due to the removal of each variable as a proxy of variable importance. All fixed effects retained were significant (P-value < 0.05). Grey is for random effects, fixed effects are in different shades of blue. Dark blue is for environmental context variables, blue lagoon and light blue are respectively for the number of annotations and the number of sites used for fine-tuning.

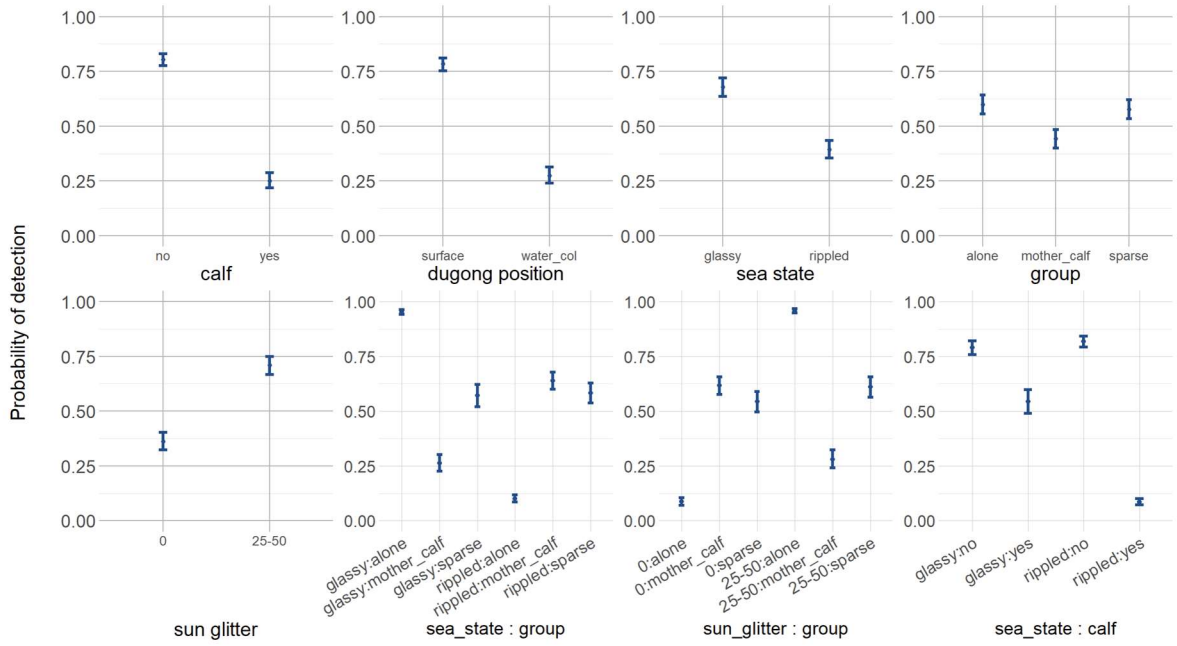


Figure 18. Marginal effects of top 8 most important variables of the probability of detection model in the long-distance transfer. Top 8 effects are selected according to the increase of deviance due to their removal; they are selected if ΔDeviance > 1000.

Calves were more poorly detected than adults, with only one in four calves detected on average (Figure 18). Dugongs in the water column were three times less detected than surfacing ones. Dugongs were better detected when the sea state was glassy, with nearly 70% of dugongs detected, whereas only 40% of annotated dugongs were detected in case of rippled sea state. In general, dugongs belonging to mother–calf pairs were more poorly detected than those that occurred alone or in sparse groups. Interestingly, dugongs present in images with sun glitter were better detected than those in images without sun glitter.

When examining the interaction between sea state and group composition, single individuals were almost always detected in glassy waters, while only one in ten individuals was detected in rippled conditions. Conversely, mother–calf pairs were better detected under rippled conditions. Sea state had little influence on the detection of sparse individuals. A similar pattern was found regarding the interaction between sun glitter and group composition. Single individuals were rarely detected in the absence of sun glitter but were consistently detected when sun glitter was present. On the other hand, mother–calf pairs were better detected in the absence of sun glitter. Sun glitter had little impact on the detection of sparse individuals.

Finally, the interaction between sea state and calf presence revealed a strong effect: calves were detected over four times more frequently in glassy waters than in rippled waters.

DISCUSSION

Transferable detection models, i.e. models trained in source regions that are capable of generalizing to new target regions, are particularly valuable in the context of population monitoring and conservation. Such models minimize the need for extensive local annotations, making it possible to monitor understudied populations more efficiently. To our knowledge, our study is the first to assess the geographic transferability of deep learning models for the detection of marine megafauna, specifically the vulnerable dugong. To do so, we trained deep learning models to detect dugongs in New Caledonia and assessed their capacity to detect dugongs in new target regions, in a nearby region of New Caledonia (short-distance transfer) and in the Indonesian province of West Papua (long-distance transfer). To adapt the models to local conditions, we relied on fine-tuning—that is providing the models with a limited number of annotated images from the target regions. We then evaluated the effect of the environmental contexts on detection performance, both during model fine-tuning and inference. In this study, we hypothesized that 1 – Dugong detection rates would be higher for the short-distance transfer than for the long-distance transfer; 2 – Increasing the amount of local annotations via fine-tuning improved dugong detection rates and 3 – Environmental and habitat contexts, both during model fine-tuning and inference, influenced significantly dugong detection.

Generalization capacity of baseline models

Baseline models trained in New Caledonia were able to detect 52% of the dugongs in the distant environments of West Papua. When transferred to a nearby region of New Caledonia, baseline models detected 62 % of the dugongs. These modest recall values reflect partial detection capacity in both transfer scenarios. Although the difference was not statistically significant (P-value = 0.16), the baseline models tended to better detect dugongs in nearby region. This finding aligns with hypothesis 1 and may be attributable to the greater similarity in environmental contexts between the source and target regions in the short-distance transfer. In particular, both regions within New Caledonia exhibited comparable dugongs group structures and habitat types, with high densities of dugongs, often occurring in schools, associated with coral and deep-water habitats. Such ecological consistency likely facilitated model transfers, as deep learning models generally achieve higher detection accuracy when applied to individuals of similar body size, exhibiting consistent behaviors, and occupying uniform habitats (Dujon et al., 2021). In contrast, the long-distance transfer involved regions with markedly different group structures and habitat types, which may have reduced the models' ability to generalize.

Influence of fine-tuning

Fine-tuning was crucial to improve dugong detection in new target regions. Increasing the amount of local annotations through fine-tuning led to a slight improvement in recall for the short-distance transfer (from 0.62 to 0.66) and a significant improvement for the long-distance transfer (from 0.52 to 0.70). This result is consistent with hypothesis 2, reflecting the model's adaptation to local

conditions in the target region (Weinstein et al., 2022). The larger improvement in the long-distance transfer may be due to greater environmental differences between source and target regions, providing more scope for the model to benefit from local adaptation. However, the larger performance of the long-distance transfer may also be explained by the slightly higher number of annotations in West Papua. Considering equal numbers of annotations led to similar performances in both transfers (recall = 0.66 for both; P-value = 0.92).

Influence of environmental contexts

Environmental contexts influenced dugong detection during both fine-tuning and inference, which is consistent with hypothesis 3.

Variability of environments during fine-tuning

During fine-tuning, the environmental contexts that positively influenced recall differed between the two transfer scenarios but led to comparable outcomes. In the short-distance transfer, the recall improved with a higher frequency of dugongs observed in images with **visible sea bottom**. In contrast, in the long-distance transfer, the recall only increased with the frequency of dugongs occurring in **coral** environments. In West Papua, coral habitats were highly **heterogeneous**, ranging from small, scattered coral patches to large, dense coral aggregations (Figure 19), providing a wide diversity of backgrounds. Conversely, in New Caledonia, coral habitats were predominantly **barrier reef** structures (Figure 19), offering lower diversity in bottom types. In New Caledonia, background diversity was captured by the “bottom visible” variable, encompassing coral reefs, sandy substrates, and sparse seagrass meadows. These results suggest that **greater background diversity** in the training data positively influenced recall during fine-tuning. These findings align with previous studies showing that training set diversity generally enhances detection performance and improves model generalization (Norman et al., 2022; Bravo-Diaz et al., 2024).

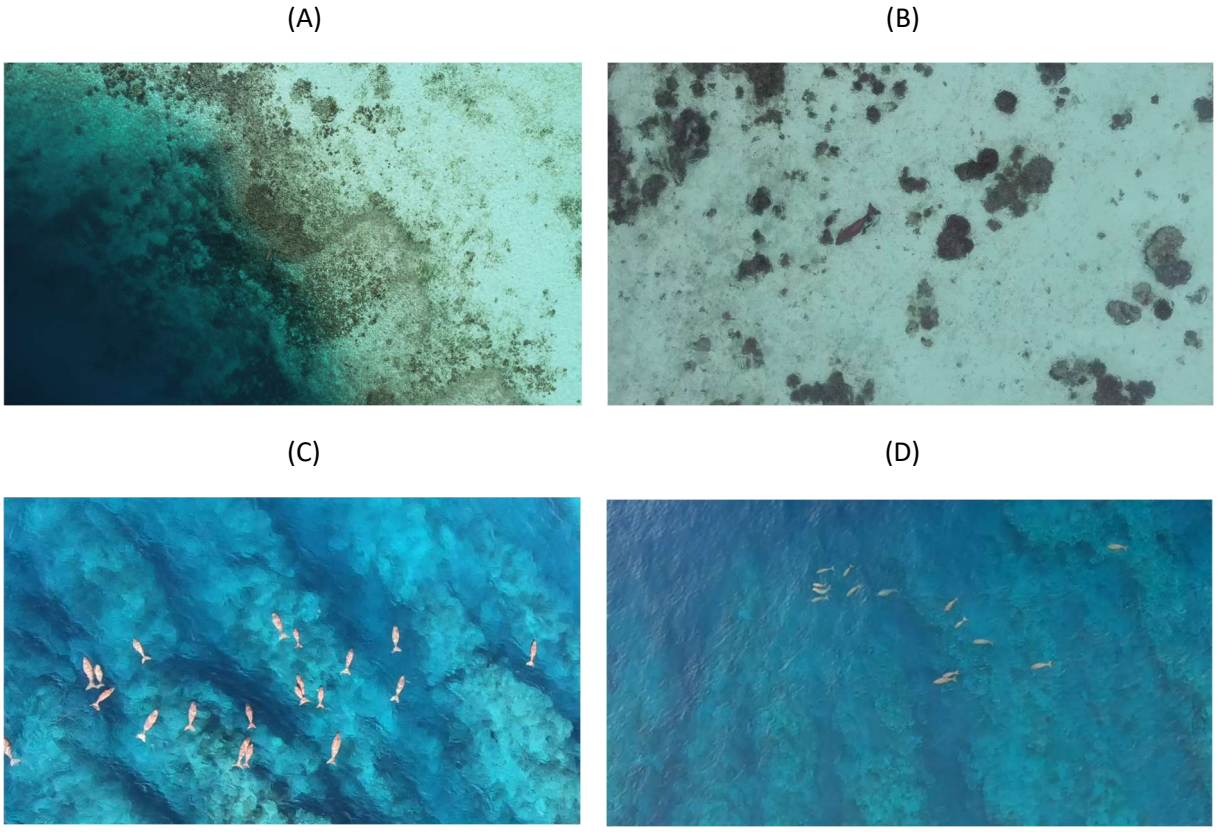


Figure 19. Examples of coral structures. (A) Dense coral colony in West Papua. (B) Patches of coral in West Papua. (C) and (D) Examples of barrier reef in New Caledonia

Occlusion effects

During both fine-tuning and inference, **occlusion**, i.e. superposed individuals, negatively influenced dugong detection. During fine-tuning in the long-distance transfer, the frequency of superposed dugongs in the training set negatively affected the recall. The presence of superposed individuals in the training set seems to cause the model to gradually "**unlearn**" the features of a dugong (Figure 14), particularly when the shape of the top individual becomes unclear (see Figure 20).

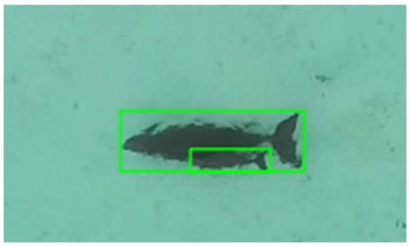


Figure 20. Example of a calf superposed to its mother in West Papua.

During inference, the occlusion effect was particularly visible for calves. In the long-distance transfer, mother-calf pairs were particularly impacted, with **calves** usually **superposed** to their mother, causing their probability of detection to drop significantly. In the short-distance transfer, occlusion occurred mainly in **groups**, with calves occurring close to their mothers exhibiting a lower detection probability. Such occlusion effects negatively impacting detection were previously described in other taxa. Delplanque et al. (2023), found a decrease in detection rates in dense elephant herds, leading to

mutual occlusions and close-by bodies. Dujon et al. (2021) noted missed detections were mostly represented by juveniles partially hidden (overlapping) by their parents in gannet colonies and by aggregated pups in seal colonies. These findings suggest that individuals occurring in close proximity, whether overlapping or aggregated, are more difficult to detect due to poor visibility of their body outlines.

Distortion and fading colors effects

Detection probability during inference was also negatively affected by **sea state** and **dugong position** in the **water column** for both transfers. Dugongs were more likely to be detected at the surface than underwater The detection of underwater dugongs was impaired by distortion effects and fading colors (Figure 21). Detection was also poorer in rippled than in glassy sea states, causing distortion effects (Figure 21).

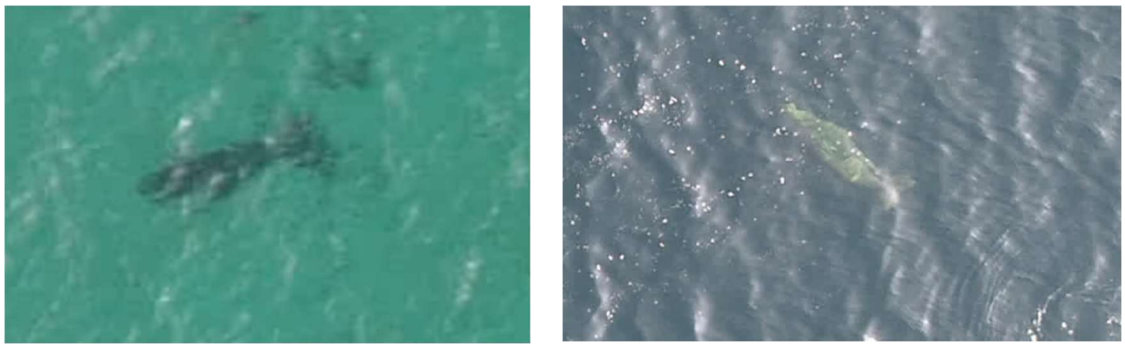


Figure 21. Example of dugongs underwater with rippled sea state conditions.

Similarly, in aerial surveys with human observers, detection probability is known to decrease with animal depth (Fuentes et al., 2015) and poorer sea states (Aniceto et al., 2018; Hodgson et al., 2023). Similar effects were also reported in CNN-based detection models. For example Dujon et al. (2021) demonstrated that, for sea turtles, detection probability decreased with depth, particularly in turbid waters. These findings suggest that CNN-based detection models are subject to the same detectability constraints as human observers.

Surface glare effects

Surface glare, including **sun glitter** and **cloud reflection**, influenced the probability of dugong detection. In the short-distance transfer, sun glitter decreased detection probability, consistent with the findings of Dujon et al. (2021), who also reported poorer detection of sea turtles with sun glitter. This effect can be explained by the reflection of sunrays masking the body outline of dugongs and reducing their underwater visibility. However, in the long-distance transfer, sun glitter appeared to positively affect detection probability; and in the short-distance transfer, cloud reflection also positively affected detection. A potential explanation is that sun glitter is often associated with higher luminescence, which has been shown to improve detection rates for human observers on drone images (Aniceto et al., 2018). In West Papua, sun glitter frequently co-occurred with the absence of dense seagrass and glassy sea states, so unbalanced sampling may also be responsible for this apparent

positive effect. Dense seagrass, a type of seagrass only occurring at the Um Island site in West Papua was noted to reduce dugong detection during manual footage review, supporting the hypothesis that the interaction between dense seagrass, glare, and sea state underlies the positive impact of sun glitter on detection. Similarly, in the short-distance transfer, cloud reflection positively affected detection. High cloud reflection (>25 %) was often associated with glassy sea states and the absence of calves, potentially contributing to the apparent positive effect.

Study limits

Although our study is the first to assess the transferability of deep learning models for dugong detection, it suffers from several limitations. First, our study is constrained by the achieved **sampling** in each region, meaning that not all environmental conditions are equally represented in the datasets. Certain factors tend to co-occur (e.g., sea state with seagrass cover or glare), preventing the balanced sampling of environmental conditions. In addition, aerial surveys were conducted over only a few days, limiting the **range of conditions** that could be observed. As a result, our models may not fully capture detection conditions that would occur across broader spatial and temporal scales. **Differing aerial survey protocols** (pre-defined transects versus 'manual' flights) and sensors (ultralight airplane versus drones) between New Caledonia and West Papua may also have affected detection results, especially in the long-distance transfer. Nevertheless, the fact that the New Caledonia-trained model detected over half of the dugongs observed in West Papua stressed partial generalization capacities to new protocols and sensors types.

Another more technical limit relates to the parameterization of **Non-Maximum Suppression** (NMS) during model validation. With the IoU (Intersection over Union) threshold set at 0.5, bounding boxes overlapping by more than 50% were discarded during validation. Such parametrization is problematic in cases of occlusion, particularly for calves that often occur very close to or directly above their mothers. In such cases, valid detections may be removed during validation, leading to an underestimation of the recall for calves and potentially preventing the selection of the weights maximizing the mAP50, i.e. the average detection accuracy with an IoU of 0.5, during training/validation. A different choice of NMS threshold or specific parameterization for mother–calf pairs would improve the model's ability to retain calf detections.

Implications for population monitoring

In regions with limited imagery and small, understudied populations, training a detection model from scratch is often impractical. A more efficient strategy is to develop a baseline model using data from a well-monitored, data-rich region and then fine-tune it with imagery from the target region. Such transfer learning approaches enhance generalizability and can be particularly valuable in places where dugongs are rare, such as Southeast Asia (Panyawai et al., 2022), or in areas like Mayotte and Mozambique. Once a fine-tuned model provides reliable detections, it can be used to estimate abundance and body size distributions, which are essential for understanding population trends (Trotzuk et al., 2021). To ensure accuracy, duplicate detections must be removed either through post-processing (human-in-the-loop correction or algorithms based on movement speed and spatial

overlap) or automated tracking (Lalgudi et al., 2025). Finally, abundance estimates should be corrected for availability bias, for example by incorporating depth-dependent detection probabilities (Hagihara et al., 2018). Together, these steps will enable reliable large-scale monitoring tools to support the conservation of threatened marine megafauna.

Perspectives

Two complementary approaches appear promising to address insufficient detection rates. Firstly, **human in the loop frameworks** could be implemented, whereby experts selectively review images with higher detection uncertainty such as those containing groups or mother – calf pairs. Studies have shown that such hybrid frameworks significantly decrease working time (Thums et al., 2018). Secondly, **adaptive confidence thresholds** could be applied depending on the environmental context, particularly by lowering thresholds in regions and conditions where dugongs are likely to occur to maximize recall. This method has been documented by Qi et al. (2018) for vehicle detection and showed an improvement in detection.

CONCLUSION AND PERSPECTIVES

In conclusion, this study demonstrated both the potential and the limitations of the geographic transfer of deep learning models applied to marine megafauna detection. Using the dugong as a case study, we found modest detection performances of model transfers both within New Caledonia and to West Papua. The slightly higher detection performance in the short-distance transfer was likely due to similar environmental contexts between source and target regions in New Caledonia. Fine-tuning with small amounts of local annotations was key to improve dugong detection, particularly in the long-distance transfer.

Our findings also highlighted the strong influence of environmental contexts, both during fine-tuning and inference, on detection performance. Occlusion effects consistently reduced detection rates, especially for calves, while individuals observed under favorable conditions, with clear outlines or visible sea bottom, were better detected. Models trained on more diverse datasets tended to generalize better, confirming the value of environmental variability within the training data.

Beyond the case of dugongs, this study emphasizes the importance of combining local annotations and diversified training datasets to build more robust and transferable detection models. Future work should integrate detection with removal of individual duplicates and correction for detection availability to obtain accurate population counts. Such approaches will pave the way for reliable large-scale monitoring tools to support the conservation of threatened marine megafauna.

BIBLIOGRAPHY

ANICETO, Ana S., BIUW, Martin, LINDSTRØM, Ulf, SOLBØ, Stian A., BROMS, Fredrik and CARROLL, JoLynn, 2018. Monitoring marine mammals using unmanned aerial vehicles: quantifying detection certainty. *Ecosphere*. March 2018. Vol. 9, no. 3, p. e02122. DOI 10.1002/ecs2.2122.

AXFORD, Daniel, SOHEL, Ferdous, VANDERKLIFT, Mathew A and HODGSON, Amanda J, 2024. Collectively advancing deep learning for animal detection in drone imagery: Successes, challenges, and research gaps. *Ecological Informatics*. November 2024. Vol. 83, p. 102842. DOI 10.1016/j.ecoinf.2024.102842.

BRAVO-DIAZ, A., MORENO, S. and LOPATIN, J., 2024. Evaluating the ability of convolutional neural networks for transfer learning in Pinus radiata cover predictions. *Ecological Informatics*. September 2024. Vol. 82, p. 102684. DOI 10.1016/j.ecoinf.2024.102684.

CHRISTIE, K.S., GILBERT, S.L., BROWN, C.L., HATFIELD, M. & HANSON, L., 2016. Unmanned aircraft systems in wildlife research: current and future applications of a transformative technology. *Frontiers in Ecology and the Environment*, 14(5), pp.241–251. https://doi.org/10.1002/fee.1281

CLEGUER, C., GARRIGUE, C., FUENTES, M., EVERINGHAM, Y., HAGIHARA, R., HAMANN, M., PAYRI, C. & MARSH, H., 2017. Drivers of change in the relative abundance of dugongs in New Caledonia. *Wildlife Research*, 44, pp.365–376. https://doi.org/10.1071/WR16133

CLEGUER, C., GARRIGUE, C., FUENTES, M.M.P.B., HAMANN, M., PAYRI, C. & MARSH, H., 2024. Temporal changes in habitat use by dugongs in a spatially restricted coral reef environment. *Pacific Conservation Biology*, 30(2), PC23036. https://doi.org/10.1071/PC23036

COSTELLO, Mark John, COLL, Marta, DANOVARO, Roberto, HALPIN, Pat, OJAVEER, Henn and MILOSLAVICH, Patricia, 2010. A Census of Marine Biodiversity Knowledge, Resources, and Future Challenges. HUMPHRIES, Stuart (ed.), *PLoS ONE*. 2 August 2010. Vol. 5, no. 8, p. e12110. DOI 10.1371/journal.pone.0012110.

DELPLANQUE, Alexandre, FOUCHER, Samuel, THÉAU, Jérôme, BUSSIÈRE, Elsa, VERMEULEN, Cédric and LEJEUNE, Philippe, 2023. From crowd to herd counting: How to precisely detect and count African mammals using aerial imagery and deep learning? *ISPRS Journal of Photogrammetry and Remote Sensing*. March 2023. Vol. 197, p. 167–180. DOI 10.1016/j.isprsjprs.2023.01.025.

DELPLANQUE, Alexandre, LAMPREY, Richard, FOUCHER, Samuel, THÉAU, Jérôme and LEJEUNE, Philippe, 2023. Surveying wildlife and livestock in Uganda with aerial cameras: Deep Learning reduces the workload of human interpretation by over 70%. *Frontiers in Ecology and Evolution*. 17 November 2023. Vol. 11, p. 1270857. DOI 10.3389/fevo.2023.1270857.

DITRIA, E.M., SIEVERS, M., LOPEZ-MARCANO, S., JINKS, E.L. & CONNOLLY, R.M., 2020. Deep learning for automated analysis of fish abundance: the benefits of training across multiple habitats. *Environmental Monitoring and Assessment*, 192, p.698

DUJON, Antoine M., IERODIACONOU, Daniel, GEESON, Johanna J., ARNOULD, John P. Y., ALLAN, Blake M., KATSELIDIS, Kostas A. and SCHOFIELD, Gail, 2021. Machine learning to detect marine animals in

UAV imagery: effect of morphology, spacing, behaviour and habitat. SCALES, Kylie and BOUCHET, Phil (eds.), *Remote Sensing in Ecology and Conservation*. September 2021. Vol. 7, no. 3, p. 341–354. DOI 10.1002/rse2.205.

FUENTES, M.M.P.B., BELL, I., HAGIHARA, R., HAMANN, M., HAZEL, J., HUTH, A., SEMINOFF, J.A., SOBTZICK, S. and MARSH, H., 2015. Improving in-water estimates of marine turtle abundance by adjusting aerial survey counts for perception and availability biases. *Journal of Experimental Marine Biology and Ecology*. October 2015. Vol. 471, p. 77–83. DOI 10.1016/j.jembe.2015.05.003.

GRAY, P.C., FLEISHMAN, A.B., KLEIN, D.J., MCKOWN, M.W., BÉZY, V.S., LOHMANN, K.J. & JOHNSTON, D.W., 2019. A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, 10(3), pp.345–355. https://doi.org/10.1111/2041-210X.13132

HAGIHARA, R., JONES, R. E., SOBTZICK, S., CLEGUER, C., GARRIGUE, C. et MARSH, H., 2018. Compensating for geographic variation in detection probability with water depth improves abundance estimates of coastal marine megafauna. *PLoS ONE*, 13(1), e0191476. DOI 10.1371/journal.pone.0191476.

HEUDIER, Mickael, MOUILLOT, David and MANNOCCI, Laura, 2023a. Assessing the effects of coral reef habitat and marine protected areas on threatened megafauna using aerial surveys. *Aquatic Conservation: Marine and Freshwater Ecosystems*. March 2023. Vol. 33, no. 3, p. 286–297. DOI 10.1002/aqc.3923.

HEUDIER, Mickael, MOUILLOT, David and MANNOCCI, Laura, 2023b. Assessing the effects of coral reef habitat and marine protected areas on threatened megafauna using aerial surveys. *Aquatic Conservation: Marine and Freshwater Ecosystems*. March 2023. Vol. 33, no. 3, p. 286–297. DOI 10.1002/aqc.3923.

HODGSON, Amanda Jane. Dugong behaviour and responses to human influences. PhD thesis, James Cook University, 2004.

HODGSON, Amanda, KELLY, Natalie and PEEL, David, 2013. Unmanned Aerial Vehicles (UAVs) for Surveying Marine Fauna: A Dugong Case Study. FENTON, Brock (ed.), *PLoS ONE*. 4 November 2013. Vol. 8, no. 11, p. e79556. DOI 10.1371/journal.pone.0079556.

HODGSON, Amanda J., KELLY, Nat and PEEL, David, 2023a. Drone images afford more detections of marine wildlife than real-time observers during simultaneous large-scale surveys. *PeerJ*. 3 November 2023. Vol. 11, p. e16186. DOI 10.7717/peerj.16186.

HODGSON, Amanda, PEEL, David, et KELLY, Natalie. Unmanned aerial vehicles for surveying marine fauna: assessing detection probability. *Ecological Applications*, 2017, vol. 27, n° 4, p. 1253-1267. DOI: https://doi.org/10.1002/eap.1519

KIM, G, JUNG, H. -G, LEE, S.-W, 2020. Few-Shot Object Detection via Knowledge Transfer. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Toronto, ON, Canada, 2020, pp. 3564-3569. DOI 10.1109/SMC42975.2020.9283497.

KOUW, W.M. & LOOG, M, 202. A review of domain adaptation without target labels. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43, 766–785.

LALGUDI, C. K., LEONE, M. E., CLARK, J. V., MADRIGAL-MORA, S. et ESPINOZA, M., 2025. Zero-shot shark tracking and biometrics from aerial imagery. Methods in Ecology and Evolution, 00, p. 1-13. DOI 10.1111/2041-210X.70116.

MANNOCCI, Laura, DERVILLE, Solène, SEGUIN, Raphael and MOUILLOT, David, 2024. Aerial Video Surveys and Spatial Prioritization Reveal Conservation Opportunities for Endangered Dugongs in New Caledonia. *Aquatic Conservation: Marine and Freshwater Ecosystems*. August 2024. Vol. 34, no. 8, p. e4237. DOI 10.1002/aqc.4237.

MANNOCCI, Laura, VILLON, Sébastien, CHAUMONT, Marc, GUELLATI, Nacim, MOUQUET, Nicolas, IOVAN, Corina, VIGLIOLA, Laurent and MOUILLOT, David, 2022. Leveraging social media and deep learning to detect rare megafauna in video surveys. *Conservation Biology*. February 2022. Vol. 36, no. 1, p. e13798. DOI 10.1111/cobi.13798.

MARSH, H.D., BLAIR, D., MCKENZIE, L. & SCHRAMM, L., 2024. A global assessment of dugong status and conservation needs. Bonn: United Nations Environment Program.

MCGILLYCUDDY, M. WARTON, DI. POPOVIC, G. BOLKER, BM., 2025. Parsimoniously Fitting Large Multivariate Random Effects in glmmTMB. *Journal of Statistical Software*, *112*(1), 1-19. https://doi.org/10.18637/jss.v112.i01.

MENEGOTTO, André and RANGEL, Thiago F., 2018. Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. *Nature Communications*. 9 November 2018. Vol. 9, no. 1, p. 4713. DOI 10.1038/s41467-018-07217-7.

MINAI, G., BOBELDYK, D. et LEIDIG, J. P., 2024. Evaluating the Impact of Diverse Marine Environments on Image Classification Performance. In: *OCEANS 2024 - Halifax*. Halifax, NS, Canada: IEEE, pp. 1-4. DOI 10.1109/OCEANS55160.2024.10753966.

NORMAN, Danielle L., BISCHOFF, Philipp H., WEARN, Oliver R., EWERS, Robert M., ROWCLIFFE, J. Marcus, EVANS, Benjamin, SETHI, Sarab, CHAPMAN, Philip M. and FREEMAN, Robin, 2023. Can CNN-based species classification generalise across variation in habitat within a camera trap survey? *Methods in Ecology and Evolution*. January 2023. Vol. 14, no. 1, p. 242–251. DOI 10.1111/2041-210X.14031.

PANYAWAI, Janmanee and PRATHEP, Anchana, 2022. A Systematic Review of the Status, Knowledge, and Research Gaps of Dugong in Southeast Asia. *Aquatic Mammals*. 15 May 2022. Vol. 48, no. 3, p. 203–222. DOI 10.1578/AM.48.3.2022.203.

PURWANTO, ANDRADI-BROWN, D.A., MATUALAGE, D., RUMENGAN, I., AWALUDINNOER, PADA, D., HIDAYAT, N.I., AMKIELTIELA, FOX, H.E., FOX, M., MANGUBHAI, S., HAMID, L., LAZUARDI, M.E., MAMBRASAR, R., MAULANA, N., MULYADI, TUHAREA, S., PAKIDING, F. & AHMADIA, G.N., 2021. The Bird's Head Seascape Marine Protected Area network—Preventing biodiversity and ecosystem service loss amidst rapid change in Papua, Indonesia. *Conservation Science and Practice*, 3(6), e393. https://doi.org/10.1111/csp2.393.

QI, S., LI, J. and SUN, Z., 2018. Adaptive Confidence Threshold Algorithm for Vehicle Detection by Employing Temporal Information. *2018 10th International Conference on Intelligent Human-Machine*

Systems and Cybernetics (IHMSC). Hangzhou, China: IEEE, pp. 348–352. DOI 10.1109/IHMSC.2018.00087.

RODOFILI, Esteban N., LECOURS, Vincent and LARUE, Michelle, 2022. Remote sensing techniques for automated marine mammals detection: a review of methods and current challenges. *PeerJ*. 20 June 2022. Vol. 10, p. e13540. DOI 10.7717/peerj.13540.

ROUSSEAU, Josée S. and BETTS, Matthew G., 2022. Factors influencing transferability in species distribution models. *Ecography*. July 2022. Vol. 2022, no. 7, p. e06060. DOI 10.1111/ecog.06060.

SEQUEIRA, Ana M. M., BOUCHET, Phil J., YATES, Katherine L., MENGERSEN, Kerrie and CALEY, M. Julian, 2018a. Transferring biodiversity models for conservation: Opportunities and challenges. MCPHERSON, Jana (ed.), *Methods in Ecology and Evolution*. May 2018. Vol. 9, no. 5, p. 1250–1264. DOI 10.1111/2041-210X.12998.

SEQUEIRA, Ana M. M., BOUCHET, Phil J., YATES, Katherine L., MENGERSEN, Kerrie and CALEY, M. Julian, 2018b. Transferring biodiversity models for conservation: Opportunities and challenges. MCPHERSON, Jana (ed.), *Methods in Ecology and Evolution*. May 2018. Vol. 9, no. 5, p. 1250–1264. DOI 10.1111/2041-210X.12998.

THUMS, Michele, JENNER, Curt, WAPLES, Kelly and SALGADO-KENT, Chandra. 2018. Humpback whale use of the Kimberley: understanding and monitoring spatial distribution.

TUIA, Devis, KELLENBERGER, Benjamin, BEERY, Sara, COSTELLOE, Blair R., ZUFFI, Silvia, RISSE, Benjamin, MATHIS, Alexander, MATHIS, Mackenzie W., VAN LANGEVELDE, Frank, BURGHARDT, Tilo, KAYS, Roland, KLINCK, Holger, WIKELSKI, Martin, COUZIN, Iain D., VAN HORN, Grant, CROFOOT, Margaret C., STEWART, Charles V. and BERGER-WOLF, Tanya, 2022. Perspectives in machine learning for wildlife conservation. *Nature Communications*. 9 February 2022. Vol. 13, no. 1, p. 792. DOI 10.1038/s41467-022-27980-y.

WEINSTEIN, Ben G., GARNER, Lindsey, SACCOMANNO, Vienna R., STEINKRAUS, Ashley, ORTEGA, Andrew, BRUSH, Kristen, YENNI, Glenda, MCKELLAR, Ann E., CONVERSE, Rowan, LIPPITT, Christopher D., WEGMANN, Alex, HOLMES, Nick D., EDNEY, Alice J., HART, Tom, JESSOPP, Mark J., CLARKE, Rohan H., MARCHOWSKI, Dominik, SENYONDO, Henry, DOTSON, Ryan, WHITE, Ethan P., FREDERICK, Peter and ERNEST, S. K. Morgan, 2022. A general deep learning model for bird detection in high-resolution airborne imagery. *Ecological Applications*. December 2022. Vol. 32, no. 8, p. e2694. DOI 10.1002/eap.2694.

WALKER, Joseph L., ZENG, Zheng, WU, Chengchen L., JAFFE, Jules S., FRASIER, Kaitlin E. and SANDIN, Stuart S., 2024. Underwater Object Detection Under Domain Shift. *IEEE Journal of Oceanic Engineering*. October 2024. Vol. 49, no. 4, p. 1209–1219. DOI 10.1109/JOE.2024.3425453.

WILLI, Marco, PITMAN, Ross T., CARDOSO, Anabelle W., LOCKE, Christina, SWANSON, Alexandra, BOYER, Amy, VELDTHUIS, Marten and FORTSON, Lucy, 2019. Identifying animal species in camera trap images using deep learning and citizen science. GAGGIOTTI, Oscar (ed.), *Methods in Ecology and Evolution*. January 2019. Vol. 10, no. 1, p. 80–91. DOI 10.1111/2041-210X.13099.

WOOLCOCK, A.B., COTTON, S. & COTTON, A.J., 2022. Effectiveness of using drones and convolutional neural networks to monitor aquatic megafauna. *African Journal of Ecology*, 60(3), pp.544–556. https://doi.org/10.1111/aje.12950

YATES, Katherine L., BOUCHET, Phil J., CALEY, M. Julian, MENGERSEN, Kerrie, RANDIN, Christophe F., PARNELL, Stephen, FIELDING, Alan H., BAMFORD, Andrew J., BAN, Stephen, BARBOSA, A. Márcia, DORMANN, Carsten F., ELITH, Jane, EMBLING, Clare B., ERVIN, Gary N., FISHER, Rebecca, GOULD, Susan, GRAF, Roland F., GREGR, Edward J., HALPIN, Patrick N., HEIKKINEN, Risto K., HEINÄNEN, Stefan, JONES, Alice R., KRISHNAKUMAR, Periyadan K., LAURIA, Valentina, LOZANO-MONTES, Hector, MANNOCCI, Laura, MELLIN, Camille, MESGARAN, Mohsen B., MORENO-AMAT, Elena, MORMEDE, Sophie, NOVACZEK, Emilie, OPPEL, Steffen, ORTUÑO CRESPO, Guillermo, PETERSON, A. Townsend, RAPACCIUOLO, Giovanni, ROBERTS, Jason J., ROSS, Rebecca E., SCALES, Kylie L., SCHOEMAN, David, SNELGROVE, Paul, SUNDBLAD, Göran, THUILLER, Wilfried, TORRES, Leigh G., VERBRUGGEN, Heroen, WANG, Lifei, WENGER, Seth, WHITTINGHAM, Mark J., ZHARIKOV, Yuri, ZURELL, Damaris and SEQUEIRA, Ana M.M., 2018a. Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*. October 2018. Vol. 33, no. 10, p. 790–802. DOI 10.1016/j.tree.2018.08.001.

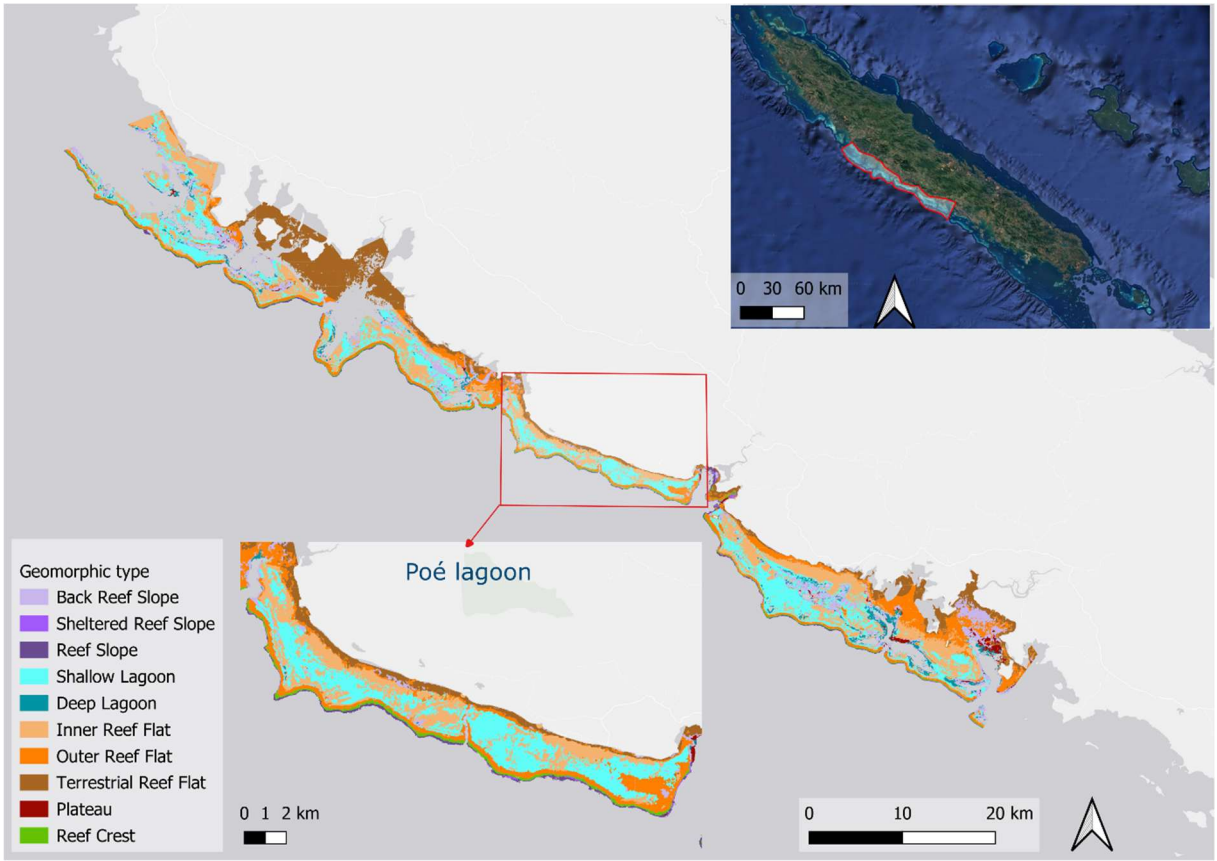
SITOGRAPHY

ALLEN CORAL ATLAS, 2025. The Allen Coral Atlas – mapping the world's coral reefs. [online]. Available at: https://allencoralatlas.org/ [Accessed June 2025].

MARSH, H., SOBTZICK, S. Dugong dugon. *The IUCN Red List of Threatened Species*, 2019, e.T6909A160756767. DOI: https://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T6909A160756767.en . Accessed 15 May 2025.

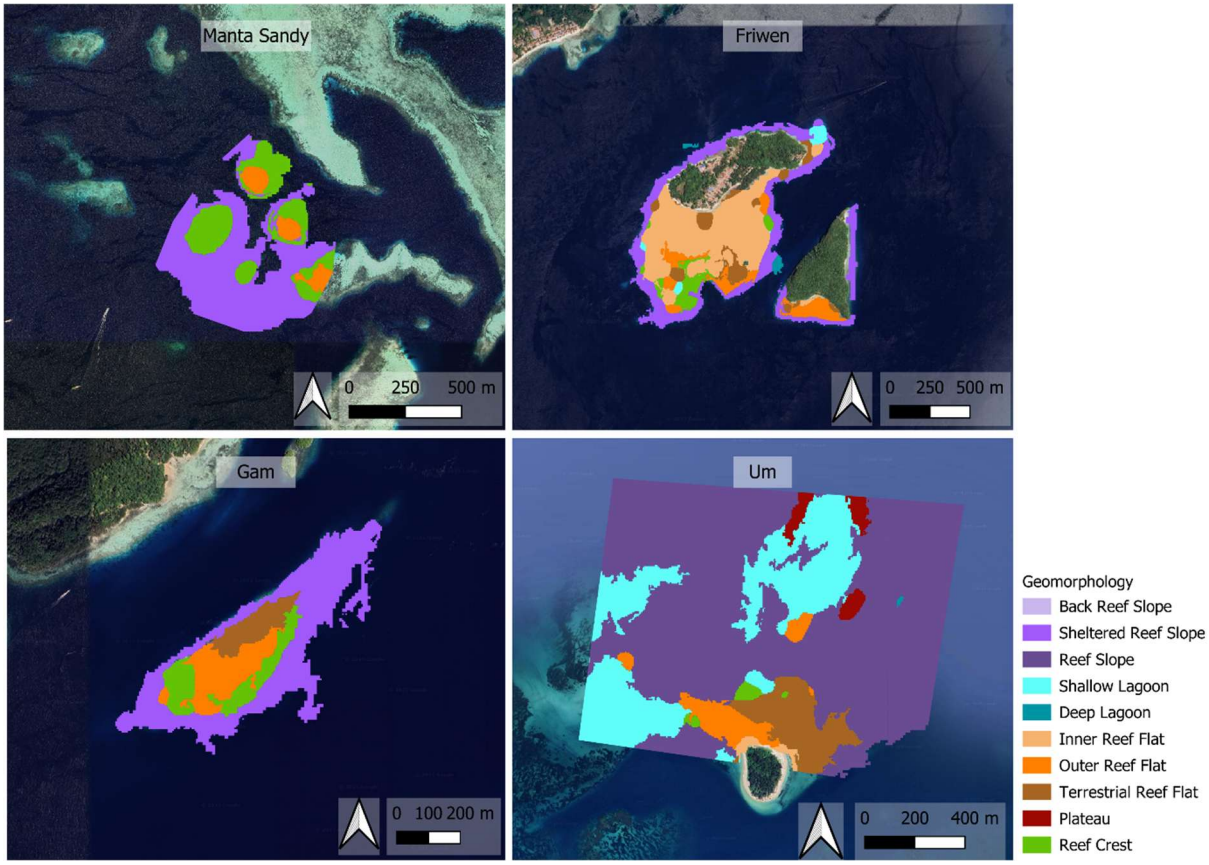
APPENDIX

Appendix 1. Geomorphological context of New Caledonia.



Geomorphological data was downloaded from the ALLEN Coral Atlas 2025.

Appendix 2. Geomorphological context the West Papua sites.



Geomorphological data was downloaded from the ALLEN Coral Atlas 2025.

Appendix 3. Hyperparameters used for fine-tuning.

Hyperparameter	Fine-tuning on short-distance transfer's target region	Fine-tuning on long-distance transfer's target region
Lr0	0.0001	0.001429
Lrf	0.001	0.01
Freeze	10	
Number of epochs	400	
Batch size	16	
Image resizing	1280	
Scale	0.1	
Translate	0.1	
IoU	0.5	

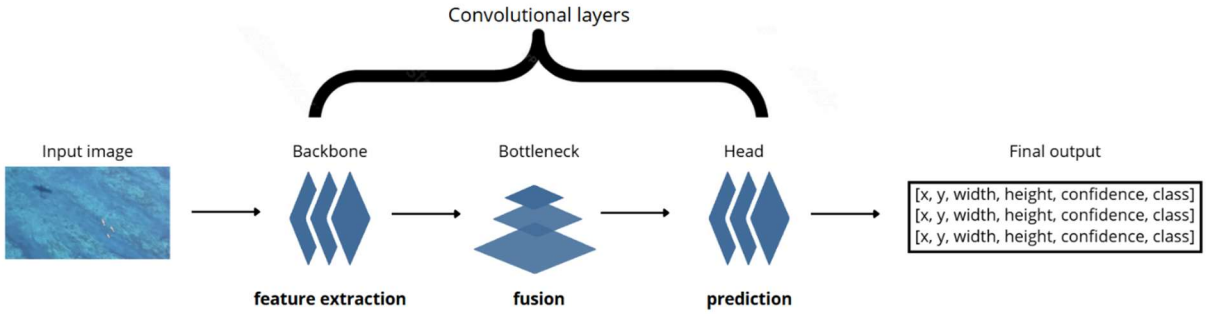
All other hyperparameters are default hyperparameters from YOLOv8m Ultralytics.

Appendix 4. Table of context variables recoded into binary variables as part of recall models.

Variables	Initial levels	Binary levels for short distance transfer	Binary levels for long distance transfer	Scale
background_complexity	low	low_medium	low	image
	medium		complex	
	high	high		
sea_state	glassy	glassy	glassy	
	rippled	rippled	rippled	
sun_glitter	0	absence	absence	
	0-25	presence	presence	
	25-50			
	>50			
cloud_reflection	0	absence	absence	
	0-25	presence	presence	
	25-50			
	>50			
calf_presence	yes	yes	yes	
	no	no	no	
group	alone	alone	alone	
	sparse	group	group	
	school			
	mother_calf			
turbidity_global	1	/	/	
	2			
	3			
	4			
bottom_visible	yes	yes	/	
	no	no		
depth	shallow	/	shallow	
	deep		deep	
coral	Presence	presence	presence	
	Absence	absence	absence	
sand	Presence	presence	presence	
	Absence	absence	absence	
dense_seagrass	Presence	presence	presence	
	Absence	absence	absence	
sparse_seagrass	Presence	presence	presence	
	Absence	absence	absence	
deep_water	Presence	presence	presence	individual
	Absence	absence	absence	
dugong_superposition	yes	/	yes	
	no		no	
background_dugong	coral	/	/	
	sand			
	dense_seagrass			
	sparse_seagrass			
	deep_water			
	dugong			
dugong_position	surface	surface	surface	
	water_col	water_col	water_col	
	bottom			
calf	yes	yes	yes	
	no	no	no	

'/' means that the variable couldn't be recoded with binary levels as the levels differ too greatly to be grouped together (e.g. background_dugong) or the factorial design of the initial variable was not balanced enough (e.g. dugong_superposition).

Appendix 5. Structure of YOLO detection model.



Appendix 6. Simplified model selection for the recall model in the short-distance transfer.

Model name	Model formula	Pseudo-deviance	AIC
Initial model	nb_annotations_train + bottom_visible_train + cloud_reflection_train + dugong_position_water_col_train + calf_train + group_train	-103,8	-87,8
Final model	bottom_visible_train	-100,1	-94,1

The number of sites was rank deficient; therefore, it couldn't be added to the model.

Appendix 7. Simplified model selection for the recall model in the long-distance transfer.

Model name	Model formula	Pseudo-deviance	AIC
Initial model	poly(nb_annotations_train, 2) + poly(nb_sites_train, 2) + sea_rippled_train + dugong_superposition_train + group_train + coral_P_train	-89.0	-69.0
Final model	dugong_superposition_train + coral_P_train	-84.9	-76.9

The recall model was fitted on 100 points out of 120, as the fine-tuning environmental context frequencies of the points referring to the baseline models were coded NA.

Appendix 8. Simplified model selection for the probability of detection model in the short-distance transfer.

Model name	Model formula	Deviance	AIC
Initial model	nb_annotations_train + poly(nb_sites_train, 2) + sea_state + bottom + sun_glitter + cloud_reflection + dugong_position + calf + group + coral + sparse_seagrass + deep_water + sea_state:bottom + sea_state:sun_glitter + sea_state:dugong_position + sea_state:group + sea_state:deep_water + bottom:cloud_reflection + bottom:calf + bottom:group + bottom:deep_water + sun_glitter:cloud_reflection + sun_glitter:dugong_position + sun_glitter:calf + sun_glitter:group + sun_glitter:deep_water + sun_glitter:sparse_seagrass + cloud_reflection:deep_water + dugong_position:calf + dugong_position:deep_water + (1 | train)	21, 178	21, 286
Final model	nb_annotations_train + sea_state + bottom + sun_glitter + cloud_reflection + dugong_position + calf + group + coral + sparse_seagrass + deep_water + sea_state:bottom + sea_state:sun_glitter + sea_state:dugong_position + sea_state:group + sea_state:deep_water + bottom:cloud_reflection + bottom:calf + bottom:group + bottom:deep_water +sun_glitter:cloud_reflection + sun_glitter:dugong_position + sun_glitter:calf + sun_glitter:group + sun_glitter:deep_water + cloud_reflection:deep_water + dugong_position:calf + dugong_position:deep_water + (1 | train)	21, 182	21, 288

The sample size was 21,636. All fixed effects retained in the final model were significant (P-value < 0.05), except *bottom* (P-value = 0.13) that was retained as it was involved in 5 significant interactions (*bottom:cloud_reflection P-value* < 2.2e-16; *bottom:calf* P-value < 2.2e-16; *bottom:group* P-value < 2.2e-16; bottom:deep_water P-value = 2.453e-11; and *sea_state:bottom* P-value < 2.2e-16).

Appendix 9. Simplified model selection for the probability of detection model in the long-distance transfer.

Model name	Model formula	Deviance	AIC
Initial model	poly(nb_sites_train, 2) + poly(nb_annotations_train_s, 2) + sea_state + depth + dugong_position + calf + group + coral + sand + sun_glitter + deep_water + sea_state:dugong_position + sea_state:calf + sea_state:group + sea_state:deep_water + depth:sand + sun_glitter:dugong_position + sun_glitter:group + dugong_position:calf + dugong_position:deep_water + dugong_position:group + coral:deep_water + (1 | train)	66, 194	66, 278
Final model	poly(nb_sites_train, 2) + poly(nb_annotations_train_s, 2) + sea_state + depth + dugong_position + calf + group + coral + sand + sun_glitter + deep_water + sea_state:dugong_position + sea_state:calf + sea_state:group + sea_state:deep_water + sun_glitter:group + dugong_position:calf + dugong_position:group + coral:deep_water + (1 | train)	66, 172	66, 294

Because the initial outlier test from the DHARMa package on residuals was significant (P-value = 1.4e-4), we removed outliers, i.e. the observations outside the 2.5th and 97.5th percentile range, from the original dataset. After refitting the model, the outlier test was no longer significant (P-value = 1), and the sample size was 86,616. All fixed effects retained in the final model were significant (P-value < 0.05).

L'INSTITUT agro Rennes Angers	Diplôme : Ingénieur Spécialité : Sciences halieutiques et aquacoles, préparée à L'Institut Agro Rennes-Angers Spécialisation / option : EcolH Enseignant référent : Etienne RIvot

Auteur(s) : **Caroline Roffi** Date de naissance* : 15/10/2001	Organisme d'accueil : Laboratoire MARBEC Université de Montpellier Adresse : 90 Place Eugène Bataillon, 34090 Montpellier Maître de stage : Laura MANNOCCI

Nb pages : 30	Annexe(s) : 9
Année de soutenance : 2025	

Titre français :	Evaluation de la transférabilité géographique de modèles de deep learning : cas de la détection du dugong sur des images aériennes
Titre anglais :	Evaluating the geographic transferability of deep learning models: the case of dugong detection in aerial images

Résumé (1600 caractères maximum) :
Cette étude a exploré comment des modèles de deep learning entraînés à détecter la mégafaune marine sur des images aériennes peuvent être transférés d'une région source vers une région cible, en prenant le dugong comme cas d'étude. Nous avons comparé le succès de la détection de dugongs dans le cadre d'un transfert à courte distance au sein de la Nouvelle-Calédonie et d'un transfert à longue distance de la Nouvelle-Calédonie vers la Papouasie occidentale (Indonésie). Nous avons analysé l'influence des conditions environnementales ainsi que l'apport d'un petit nombre d'annotations locales issues de la région cible sur les performances de détection. Les modèles entraînés uniquement sur la région source ont été légèrement plus performants lors du transfert à courte distance, dû à une plus grande similarité environnementale entre régions source et cible. L'ajout d'annotations locales a amélioré la détection des dugongs, en particulier dans le transfert à longue distance. Le succès de la détection a été fortement influencé par des contextes environnementaux tels que l'occlusion entre individus, l'état de la mer, la position du dugong dans la colonne d'eau, les reflets de surface, ainsi que la diversité des arrière-plans des images d'entraînement. Nous recommandons une approche hybride combinant deep learning, validation manuelle par des experts et ajustements adaptés aux contextes environnementaux. Au-delà du cas du dugong, cette étude apporte des perspectives pour le développement d'outils de suivi transférables, afin de soutenir la conservation de la mégafaune marine menacée.

Abstract (1600 caractères maximum) :
This study explored how deep learning models trained to detect marine megafauna in aerial images can be transferred from a source region to a target region, using the dugong as a case study. We compared the success of dugong detection in aerial images in a short-distance transfer within New Caledonia and a long-distance transfer from New Caledonia to West Papua (Indonesia). We investigated the impact of environmental conditions and the addition of a small number of local annotations from the target region on detection. Models trained solely on the source region performed slightly better in short-distance transfers due to higher environmental similarity between the source and target regions. Training the model with local annotations from the target region improved dugong detection, particularly in the long-distance transfer. The success of dugong detection was strongly influenced by environmental and ecological factors such as occlusion between individuals, sea state, position in the water column, and surface glare as well as diversity of backgrounds in the training. We recommend a hybrid approach to dugong detection, combining deep learning, manual human review, and adjustments tailored to environmental and ecological contexts. Beyond dugongs, this study provides valuable insights for developing transferable monitoring tools to support the conservation of threatened marine megafauna.

Mots-clés : Analyse d'images ; mégafaune marine ; études aériennes ; détection d'espèces ; contexte environnemental

Key Words: Image analysis; marine megafauna; aerial surveys; species detection; environmental context