Supplementary material

Using integrated multispecies occupancy models to map co-occurrence between bottlenose dolphins and fisheries in the Gulf of Lion, French Mediterranean Sea

Comparison of precision between integrated model vs. single-dataset models

Valentin Lauret[1], Hélène Labach[2], Léa David[3], Matthieu Authier[4,5], Olivier Gimenez[1]

(1) CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France
(2) MIRACETI, Connaissance et conservation des cétacés, Place des traceurs de pierres, 13500 La Couronne, France
(3) EcoOcéan Institut, 18 rue des Hospices, 34090 Montpellier
(4) Observatoire Pelagis, UAR 3462, CNRS-La Rochelle Université, 17000 La Rochelle, France
(5) Adera, 162 Avenue Albert Schweitzer. CS 60040 – 33608 Pessac CEDEX, France

---

Data and codes are available on Github.

Introduction

In this supplementary document, we want to assess the benefit of the integration process by comparing the precision of ecological estimation associated with each model. Then, we ran multispecies occupancy models using :

- SAMM aerial line-transects data only, *ie* SAMM model.
- GDEGeM boat search-encounter data only, *ie* GDEGeM model.

and compared the precision between the models.

Methods

Model formulation

We used the (Rota et al. 2016b) formulation of multispecies occupancy models as described in the manuscript but we only fit one dataset.

Occupancy process

Latent ecological process

We followed Rota et al. (2016a) to formulate the ecological model describing the occupancy process. In grid-cell i, the latent occupancy state can take 4 values: $z = [1, 0, 0, 0]$ if neither dolphins nor trawlers use the grid-cell, $z = [0, 1, 0, 0]$ if dolphins use the grid-cell but trawlers do not, $z = [0, 0, 1, 0]$ if trawlers use the grid-cell but dolphins do not, and $z = [0, 0, 0, 1]$ if both dolphins and trawlers use the grid-cell. Then, ignoring the grid-cell index, our multispecies occupancy model estimated 4 occupancy probabilities.

- ψ^4 is the probability that both dolphins and trawlers use the grid-cell;
- ψ^3 is the probability that trawlers use the grid-cell and dolphins do not;
- ψ^2 is the probability that dolphins use the grid-cell and trawlers do not;
- ψ^1 is the probability that neither dolphins nor trawlers use the grid-cell, which corresponds to the probability that none of the previous events occurs, with $\psi^1 = 1 - \psi^2 - \psi^3 - \psi^4$.

We modeled the occupancy state of each grid-cell z as a multinomial logistic regression, z being draw in vector $\pi = [(1 - \psi^2 - \psi^3 - \psi^4), \psi^2, \psi^3, \psi^4]$:

$$z \sim Multinomial(1, \pi)$$

with π adjusted to sum to 1 using a generalized logit link function. We modeled occupancy probabilities ψ^2, ψ^3, and ψ^4 as a function of depth and non-parametric functions geographical coordinates of the grid-cell center X and Y with Generalized Additive Models (GAMs) (Wood 2006) using a multinomial-logit link. For grid-cell i:

$$\psi_i^1 = \frac{1}{1 + \exp(\delta_i^2) + \exp(\delta_i^3) + \exp(\delta_i^2 + \delta_i^3 + \delta_i^4)}$$
$$\psi^2 = \frac{\exp(\delta_i^2)}{1 + \exp(\delta_i^2) + \exp(\delta_i^3) + \exp(\delta_i^2 + \delta_i^3 + \delta_i^4)}$$
$$\psi^3 = \frac{\exp(\delta_i^3)}{1 + \exp(\delta_i^2) + \exp(\delta_i^3) + \exp(\delta_i^2 + \delta_i^3 + \delta_i^4)}$$
$$\psi^4 = \frac{\exp(\delta_i^2 + \delta_i^3 + \delta_i^4)}{1 + \exp(\delta_i^2) + \exp(\delta_i^3) + \exp(\delta_i^2 + \delta_i^3 + \delta_i^4)}$$

where, for $k \in \{2, 3, 4\}$, δ_i^k in grid-cell i is:

$$\delta_i^k = \alpha_0^k + \alpha_1^k \text{depth}_i + s^k(X_i, Y_i)$$

where $s^k(.)$ is a smooth function. Quantities $s^k(.)$, α_0^k and α_1^k were to be estimated.

Detection process

Each species T (for trawlers) and D (for dolphins) has a different detection probability, p_D and p_T, modeled as a logit-function of sampling effort

$$\text{logit}(p_d) = \beta_{0d} + \beta_{1d}\text{sampling effort}$$

where β_0 and β_1 are to be estimated.

Because we now consider only one detection process, 4 observation events can occur :

- 1 for none species detected
- 2 for species A detected
- 3 for species B detected
- 4 for both species detected

From the 4 ecological states and the 4 observation events, we have the observation process with the following 4x4 matrix.

$$t(\theta') = \begin{bmatrix} 1 & (1-p_A) & (1-p_B) & (1-p_B)(1-p_A) \\ 0 & p_A & 0 & p_A(1-p_B) \\ 0 & 0 & p_B & p_B(1-p_A) \\ 0 & 0 & 0 & p_A p_B \end{bmatrix}$$

Run with NIMBLE

We ran each model on NIMBLE with 200 000 iterations. See R codes for model comparison

Model comparison

From parameters α_1, α_0, and $s(.)$, we calculated the predicted space-use probability by bottlenose dolphins ψ_1, by trawlers ψ_2, and predicted co-occurrence probability ψ_3. We reported maps of mean and standard deviation of ψ_3

Results

Co-occurrence probability

Concerning co-occurrence probability ψ_3, integrated model exhibits a higher precision (i.e. lower standard deviation) that multispecies occupancy models using datasets in isolation (Figure 1 & 2).

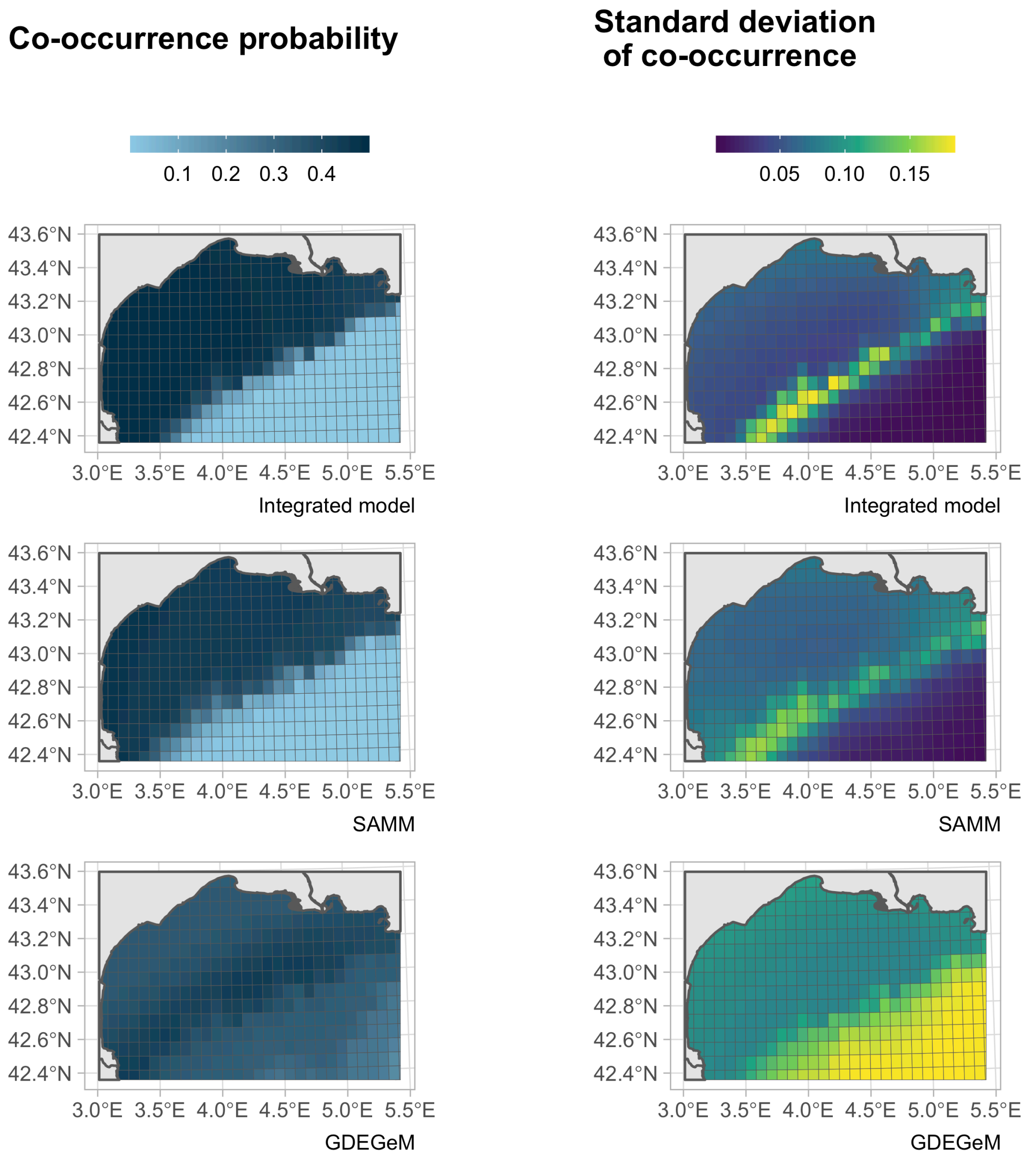

Figure 1: Maps of co-occurrence probability and standard deviation for each model

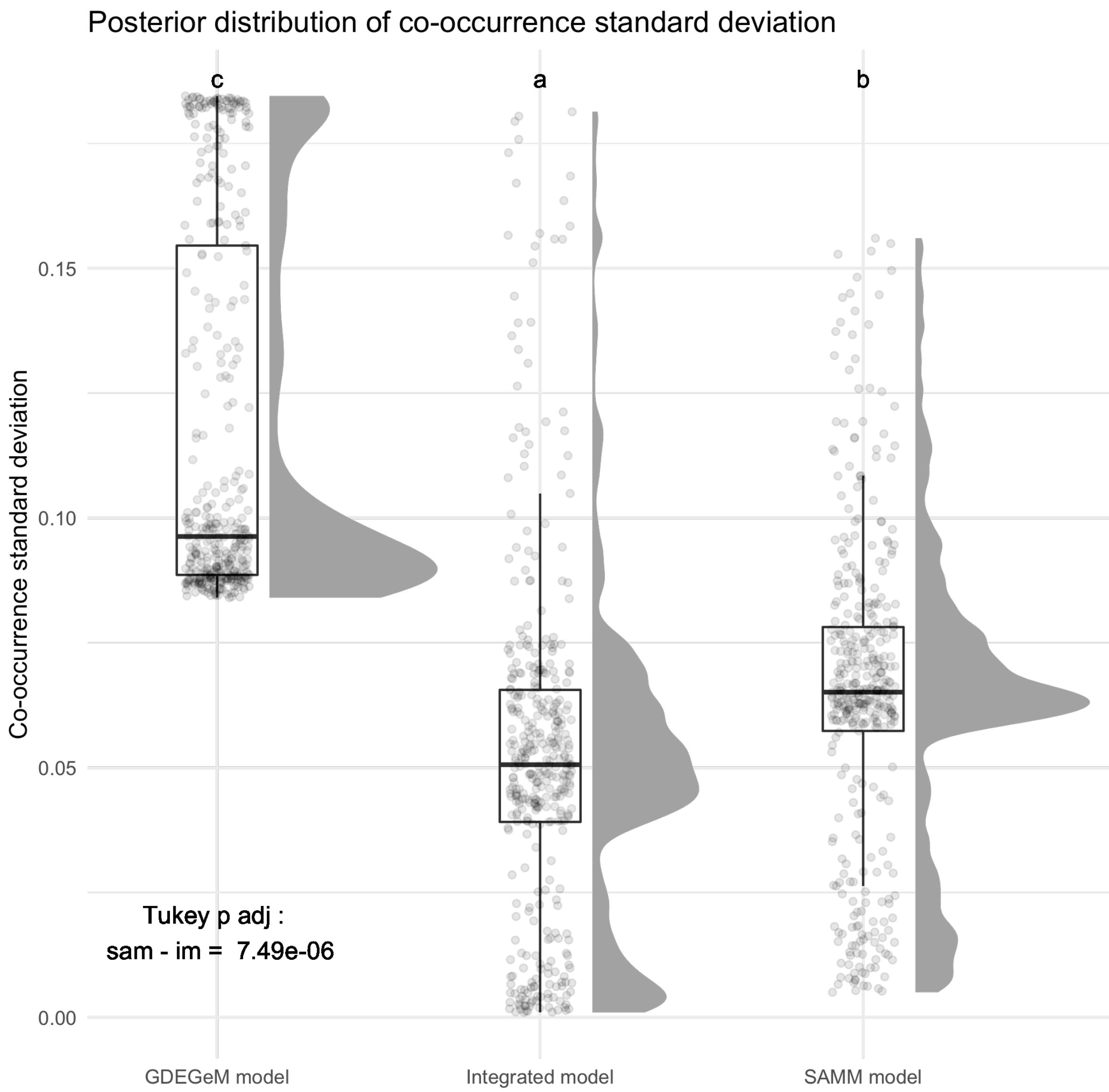

Figure 2: Standard deviation associated with co-occurrence probability. We tested for statistical differences between posterior distribution. 'sam' and 'im' respectively refer to SAMM and integrated model

Concerning parameters ψ_1 and ψ_2, we observed an increase in precision for integrated model but this difference is not significant compared to the SAMM model (Figure 3).

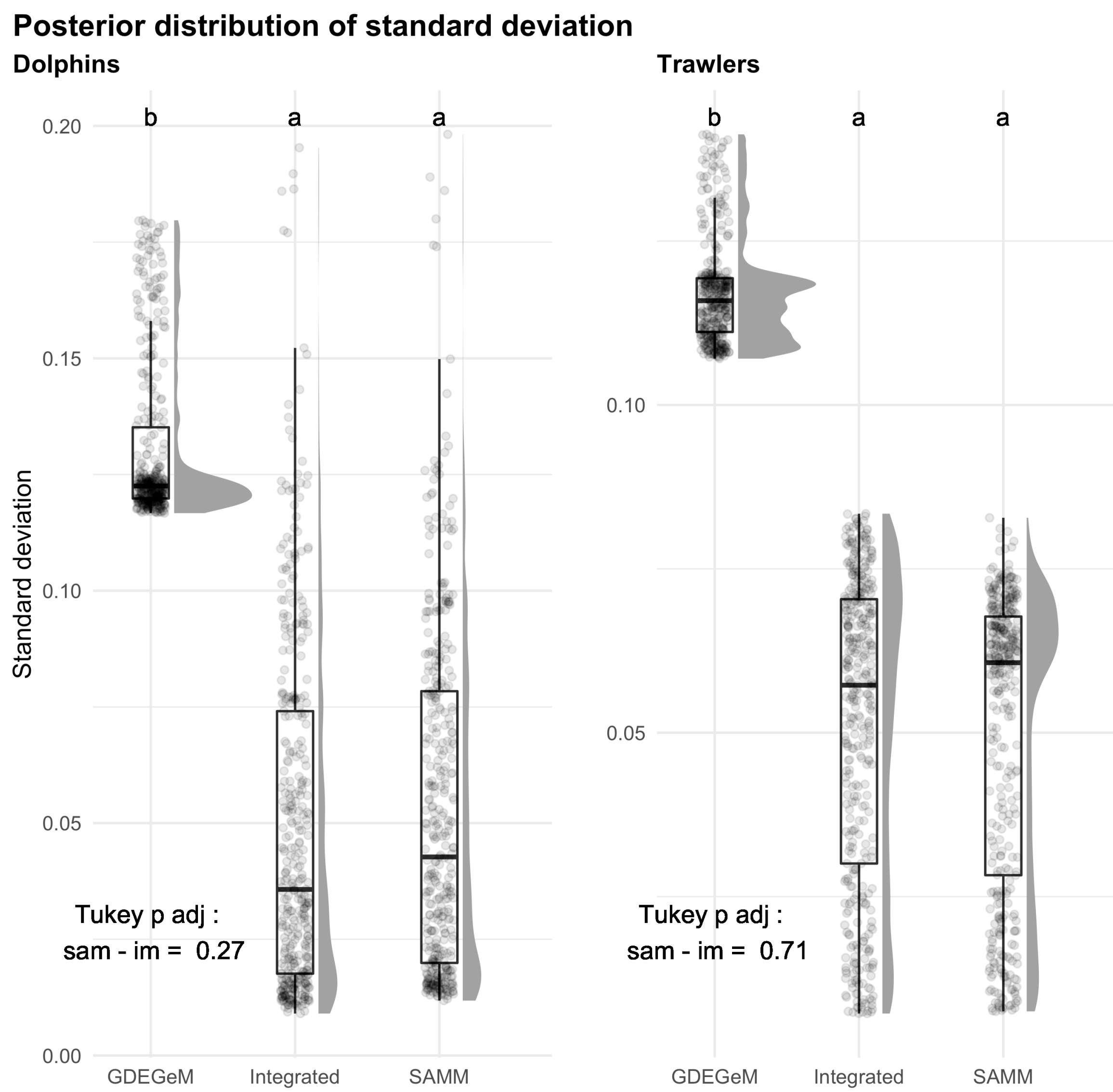

Figure 3: Standard deviation associated with space-use probability for dolphins and trawlers. We tested for statistical differences between posterior distribution. 'sam' and 'im' respectively refer to SAMM and integrated model.

Discussion

Comparing integrated model with other models using datasets in isolation put emphasis on the benefit of integrated approaches to overcome data scarcity (Zipkin et al. 2019). Integrated model exhibits more precise estimated of co-occurrence probability that that of GDEGeM and SAMM model separately. However, our ecological estimations are limited by the small quantity of data available in this case study about bottlenose dolphins.

Nevertheless, we presented a potential illustration of data integration benefit. Integrated models constitute an asset for multispecies occupancy as these models are known to require a substantial amount of data to fit (Clipp et al. 2021). Through our case study on bottlenose dolphins, we provided an promising framework to extend multispecies occupancy models to integrated mutliple datsets.

References

Clipp H L, Evans AL, Kessinger BE, Kellner K, Rota CT (2021) A Penalized Likelihood for Multi-Species Occupancy Models Improves Predictions of Species Interaction. Ecology In press.

Rota CT, Ferreira MAR, Kays RW, Forrester TD, Kalies EL, McShea WJ, Parsons AW, Millspaugh JJ (2016a) A Multispecies Occupancy Model for Two or More Interacting Species. Methods in Ecology and Evolution 7:1164–1173.

Rota CT, Wikle CK, Kays RW, Forrester TD, McShea WJ, Parsons AW, Millspaugh JJ (2016b) A Two-Species Occupancy Model Accommodating Simultaneous Spatial and Interspecific Dependence. Ecology 97:48–53.

Wood SN (2006) Generalized Additive Models: An Introduction with R. Chapman; Hall/CRC, New York.

Zipkin EF, Inouye BD, Beissinger SR (2019) Innovations in Data Integration for Modeling Populations. Ecology:e02713.