Infering species distribution from aggregated spatial data

Going from coarse landings data to fine scale species distribution

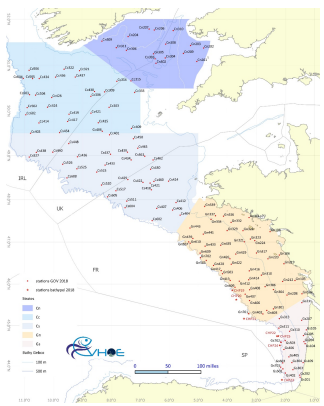
Baptiste Alglave, Marie-Pierre Etienne, Youen Vermard, Mathieu Woillez, Etienne Rivot

04/2022 - GdR Ecostats

Ifremer

decod
de la source à l'océan

Spatial data in ecology

	Survey data	Citizen science data	Declaration data
+	Standardized sampling plan High quality data	Access to more data Exact locations usually available	Mandatory declaration Massive data
−	Low amount of samples	Opportunistic (or even preferential) sampling	Aggregated at the scale of administrative units
Examples			
	EVHOE data, Bay of Biscay (marine ecology)	Ebird application (ornithology)	Harvest data, Wisconsin (hunting)

How to integrate all these datasources?

(especially when they do not have the same spatial resolution)

➠ Some specific proposition in the context of fishery science:

Alglave Baptiste, Rivot Etienne, Etienne Marie-Pierre, Woillez Mathieu, Thorson James T, Vermard Youen (2022). **Combining scientific survey and commercial catch data to map fish distribution.** ICES Journal of Marine Science IN PRESS. https://doi.org/10.1093/icesjms/fsac032

How to integrate all these datasources?

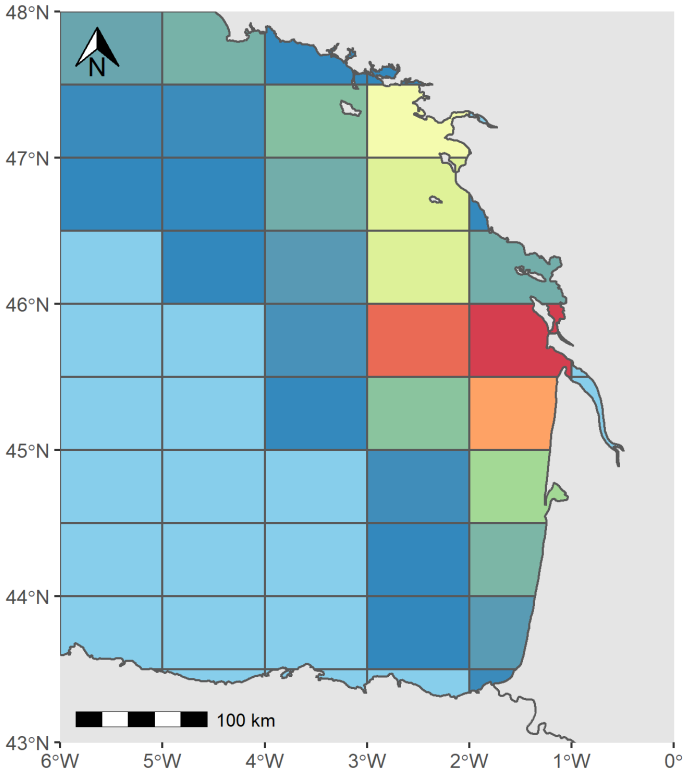
(especially when they do not have the same spatial resolution)

➠ Some specific proposition in the context of fishery science:

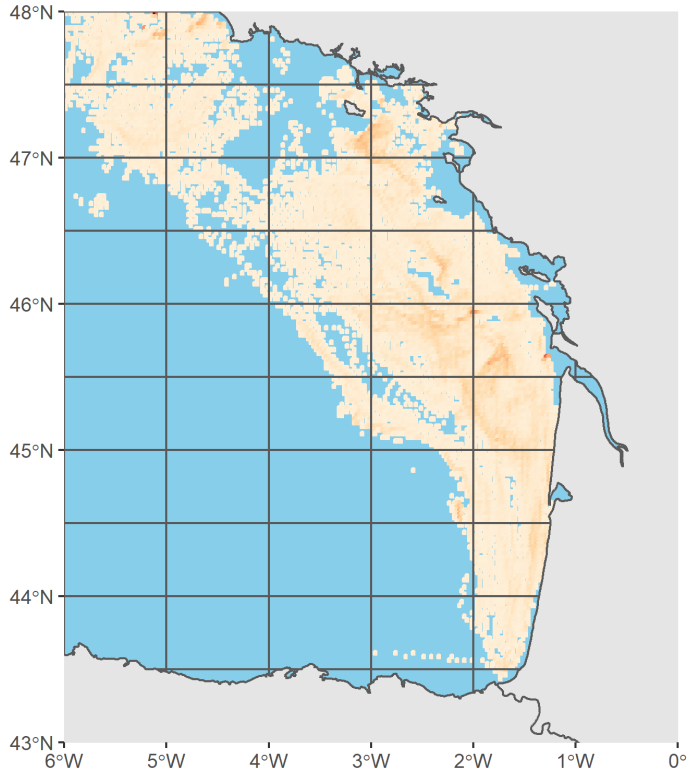
Alglave Baptiste, Rivot Etienne, Etienne Marie-Pierre, Woillez Mathieu, Thorson James T, Vermard Youen (2022). **Combining scientific survey and commercial catch data to map fish distribution.** ICES Journal of Marine Science IN PRESS. https://doi.org/10.1093/icesjms/fsac032

Commercial catch declarations data in fishery science

Logbook landings
(Sole, trawlers OTB-DEF, 2018)



Fishing locations (VMS)
(Trawlers OTB-DEF, 2018)



Resolution

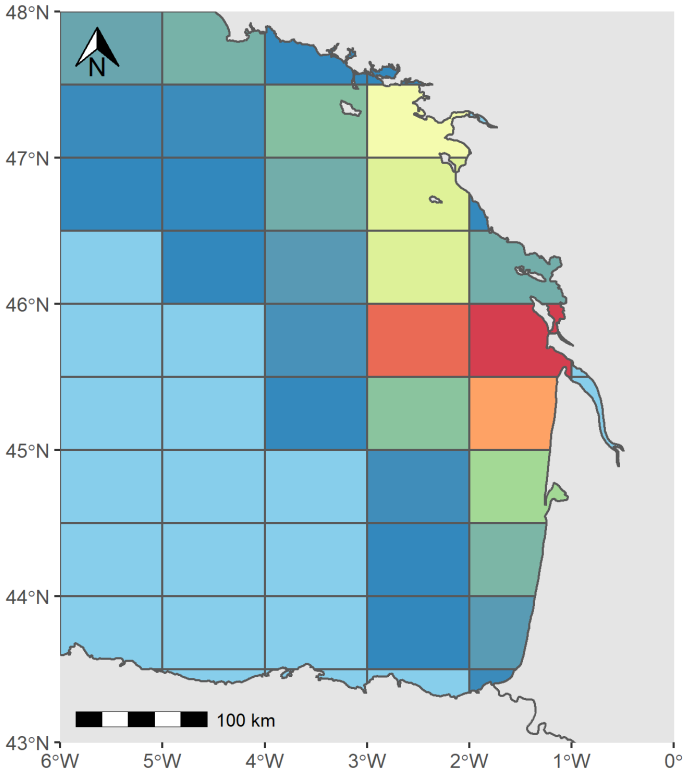
Catch are daily declared at the resolution of ICES rectangles

VMS pings are vessels GPS locations emitted each hour

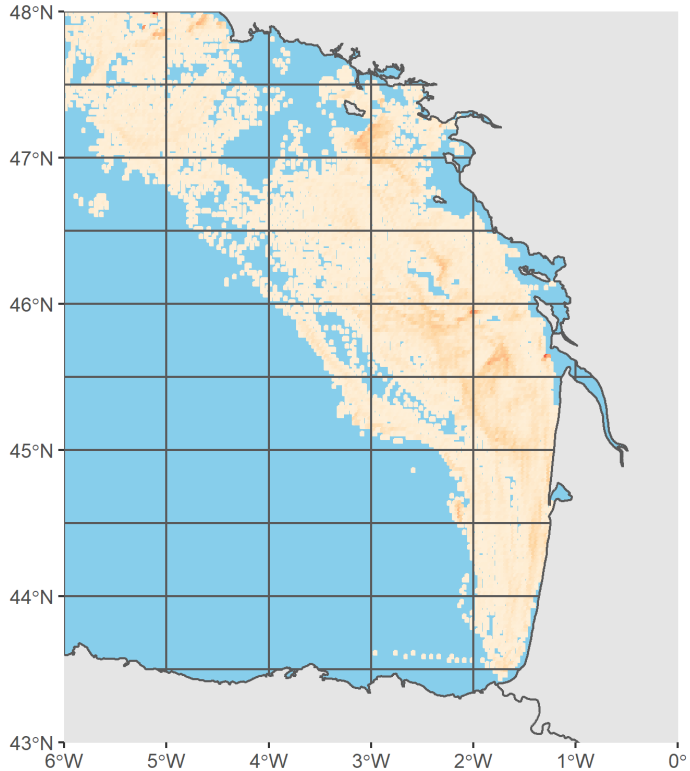
➠ Refine landings spatial resolution

Commercial catch declarations data in fishery science

Logbook landings
(Sole, trawlers OTB-DEF, 2018)



Fishing locations (VMS)
(Trawlers OTB-DEF, 2018)



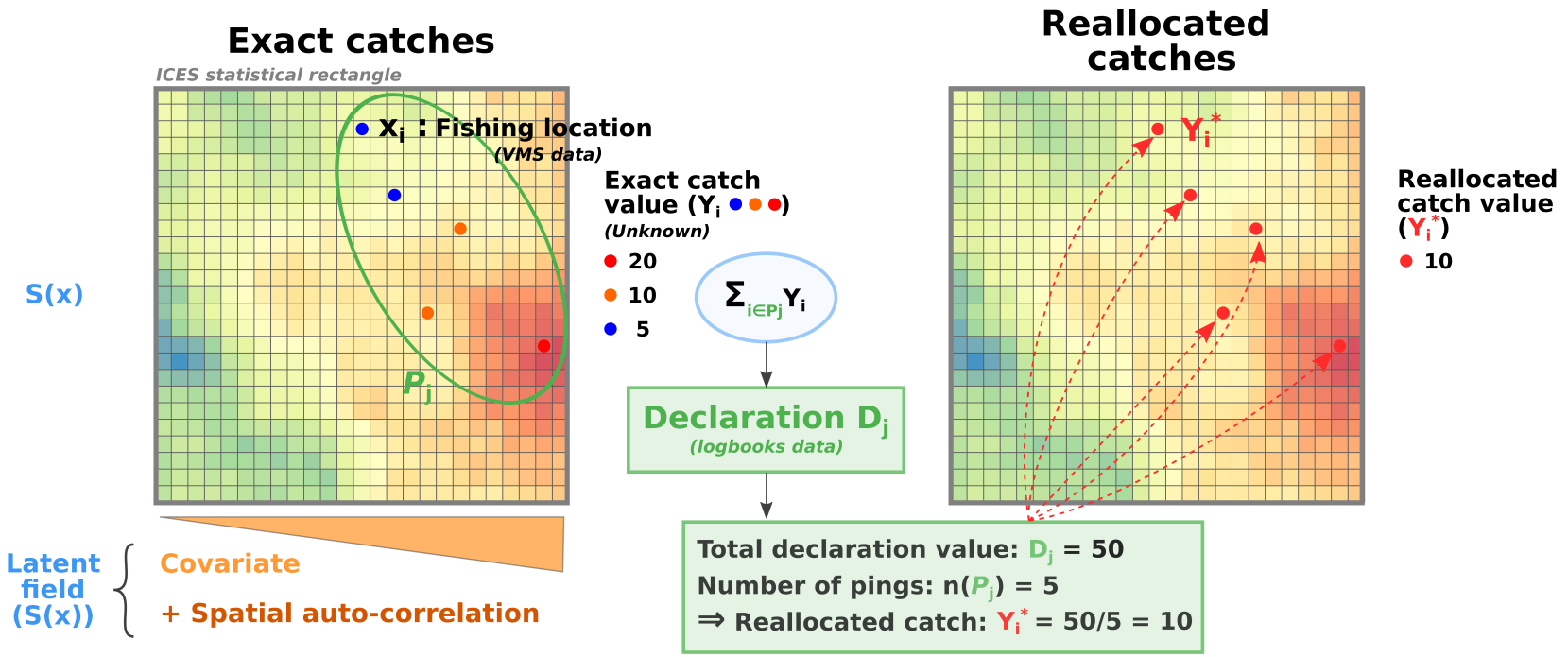
Resolution

Catch are daily declared at the resolution of ICES rectangles

VMS pings are vessels GPS locations emitted each hour

➠ **Refine landings spatial resolution**

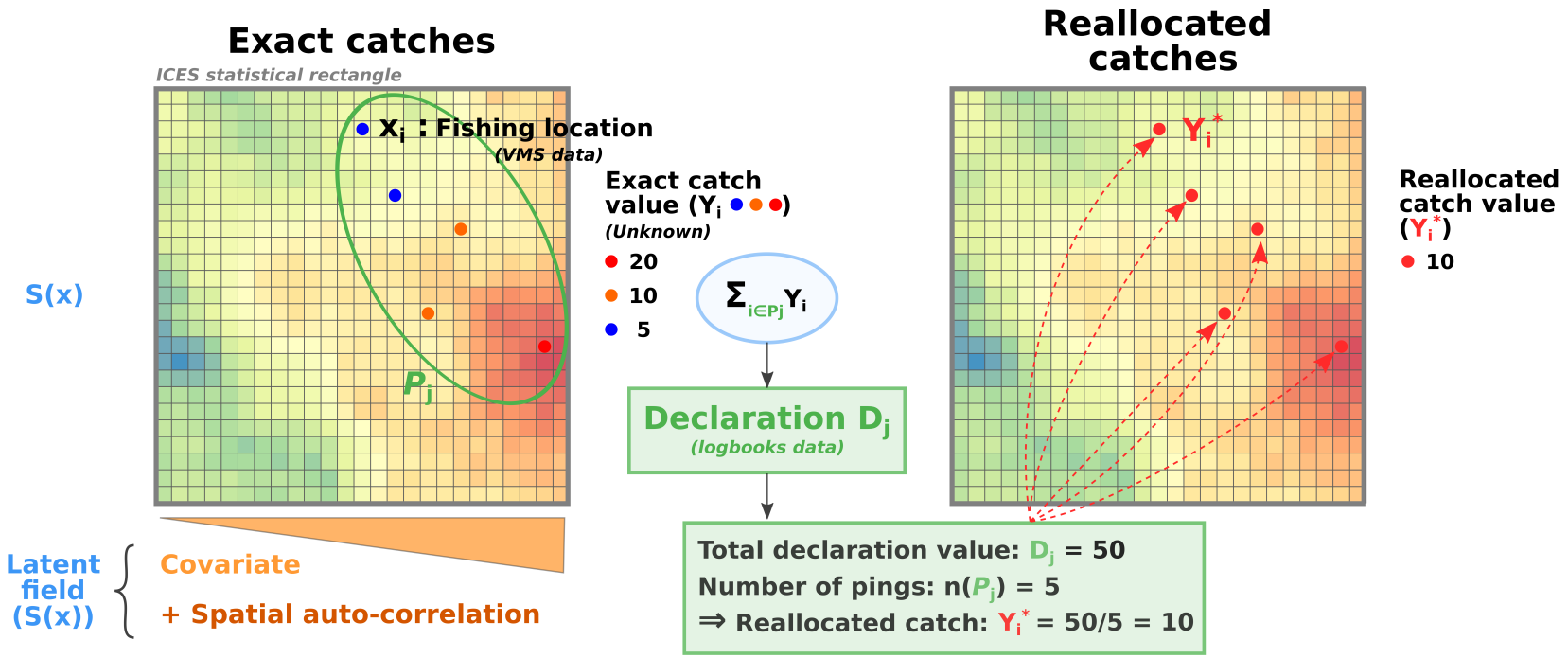
Modelling change of support



Actual situation $Y_i|S(x_i), x_i \sim \mathcal{L}_Y(S(x_i), \xi, \sigma)$ $Y_i = Y_i^*$

Alternative solution $D_j = \sum_{i \in \mathcal{P}_j} Y_i$ $D_j|S_{\mathcal{P}_j}, \mathcal{P}_j \sim \mathcal{L}_D(S_{\mathcal{P}_j}, \xi, \sigma)$ ➠ Match $\mathcal{L}_D$ and $\mathcal{L}_Y$ moments

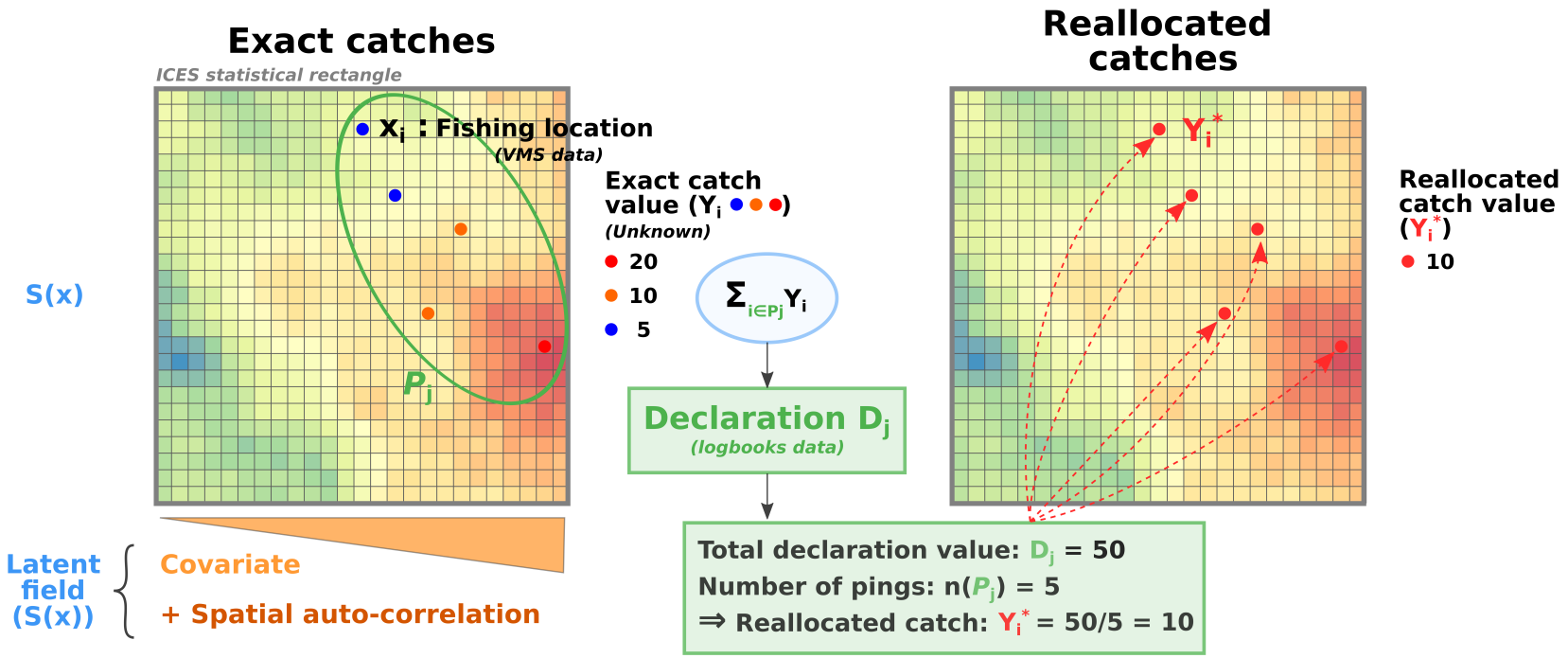
Modelling change of support



Actual situation $Y_i|S(x_i), x_i \sim \mathcal{L}_Y(S(x_i), \xi, \sigma)$ $Y_i = Y_i^*$

Alternative solution $D_j = \sum_{i \in \mathcal{P}_j} Y_i$ $D_j|S_{\mathcal{P}_j}, \mathcal{P}_j \sim \mathcal{L}_D(S_{\mathcal{P}_j}, \xi, \sigma)$ ➠ Match $\mathcal{L}_D$ and $\mathcal{L}_Y$ moments

Modelling change of support



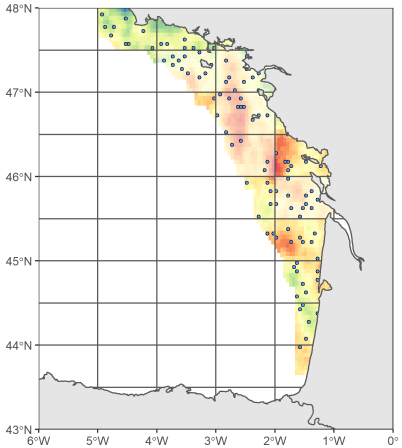
Actual situation $Y_i | S(x_i), x_i \sim \mathcal{L}_Y(S(x_i), \xi, \sigma)$ $Y_i = Y_i^*$

Alternative solution $D_j = \sum_{i \in \mathcal{P}_j} Y_i$ $D_j | S_{\mathcal{P}_j}, \mathcal{P}_j \sim \mathcal{L}_D(S_{\mathcal{P}_j}, \xi, \sigma)$

➠ Match $\mathcal{L}_D$ and $\mathcal{L}_Y$ moments

Simulation-estimation and case study

Simulation/estimation



Simulation

- **Latent field** (covariate + spatial random effect)
- **Commercial data** (3000 samples over 2/3 of the area)
- **Reallocation process** (10 locations per declaration)
- **Scientific data** (100 samples over the whole the area)

Model evaluation

1/ Mean square prediction error

$$MSPE = \frac{\sum_x (S(x) - \hat{S}(x))^2}{n}$$

2/ Covariate effect (or species-habitat relationship):

$$\beta_S = 2 \text{ versus } \hat{\beta}_S$$

Estimation

Comparison of 3 model configurations:

1/ Model fitted to scientific data (reference data)

2/ Integrated model with commercial likelihood built on Y_i^*

3/ Integrated model with commercial likelihood built on D_j

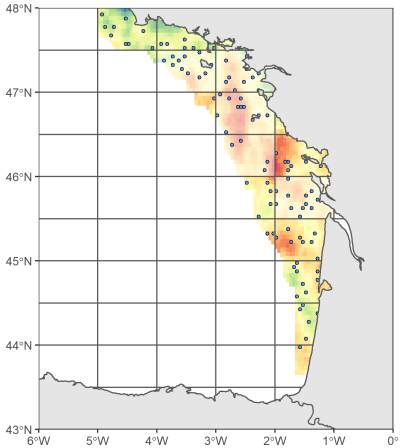
Estimation realized through TMB (Template Model Builder)
100 runs of simulation-estimation

Case study: Sole in the Bay of Biscay

Survey data: Orhago
Commercial data: OTB-DEF traulers (to ease convergence onboard observer data were integrated in fitting)
Fitted models: same as simulation
Covariate: substrate

Simulation-estimation and case study

Simulation/estimation



Simulation

- **Latent field** (covariate + spatial random effect)
- **Commercial data** (3000 samples over 2/3 of the area)
- **Reallocation process** (10 locations per declaration)
- **Scientific data** (100 samples over the whole the area)

Model evaluation

1/ Mean square prediction error

$$MSPE = \frac{\sum_x (S(x) - \hat{S}(x))^2}{n}$$

2/ Covariate effect (or species-habitat relationship):

$$\beta_S = 2 \text{ versus } \hat{\beta}_S$$

Estimation

Comparison of 3 model configurations:

1/ Model fitted to scientific data (reference data)

2/ Integrated model with commercial likelihood built on Y_i^*

3/ Integrated model with commercial likelihood built on D_j

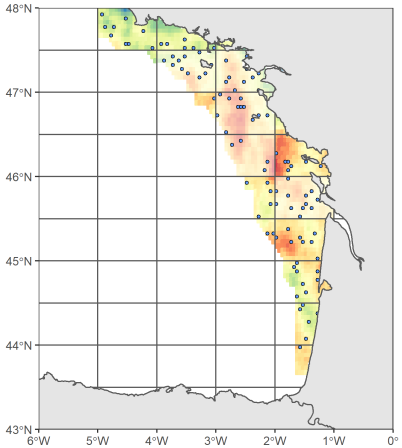
Estimation realized through TMB (Template Model Builder)
100 runs of simulation-estimation

Case study: Sole in the Bay of Biscay

Survey data: Orhago
Commercial data: OTB-DEF traulers (to ease convergence onboard observer data were integrated in fitting)
Fitted models: same as simulation
Covariate: substrate

Simulation-estimation and case study

Simulation/estimation



Simulation

- **Latent field** (covariate + spatial random effect)
- **Commercial data** (3000 samples over 2/3 of the area)
- **Reallocation process** (10 locations per declaration)
- **Scientific data** (100 samples over the whole the area)

Model evaluation

1/ Mean square prediction error

$$MSPE = \frac{\sum_x (S(x) - \hat{S}(x))^2}{n}$$

2/ Covariate effect (or species-habitat relationship):

$$\beta_S = 2 \text{ versus } \hat{\beta}_S$$

Estimation

Comparison of 3 model configurations:

1/ Model fitted to scientific data (reference data)

2/ Integrated model with commercial likelihood built on Y_i^*

3/ Integrated model with commercial likelihood built on D_j

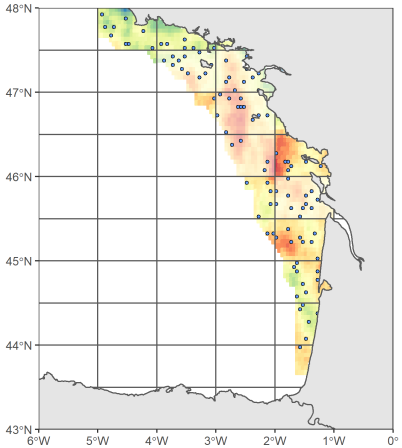
Estimation realized through TMB (Template Model Builder)
100 runs of simulation-estimation

Case study: Sole in the Bay of Biscay

Survey data: Orhago
Commercial data: OTB-DEF traulers (to ease convergence onboard observer data were integrated in fitting)
Fitted models: same as simulation
Covariate: substrate

Simulation-estimation and case study

Simulation/estimation



Simulation

- Latent field (covariate + spatial random effect)
- **Commercial data** (3000 samples over 2/3 of the area)
- Reallocation process (10 locations per declaration)
- Scientific data (100 samples over the whole the area)

Model evaluation

1/ Mean square prediction error

$$MSPE = \frac{\sum_x (S(x) - \hat{S}(x))^2}{n}$$

2/ Covariate effect (or species-habitat relationship):

$$\beta_S = 2 \text{ versus } \hat{\beta}_S$$

Estimation

Comparison of 3 model configurations:

1/ Model fitted to scientific data (reference data)

2/ Integrated model with commercial likelihood built on Y_i^*

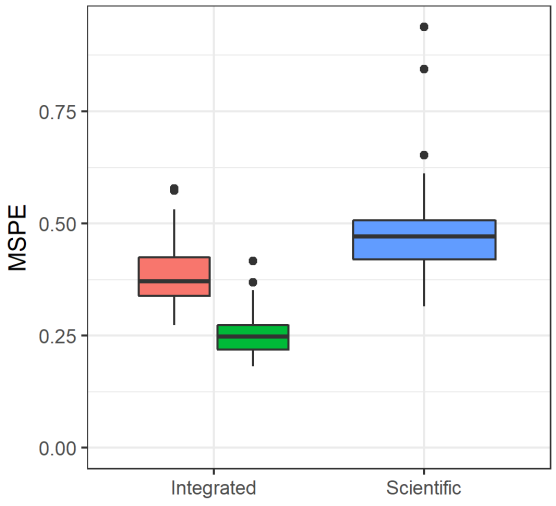
3/ Integrated model with commercial likelihood built on D_j

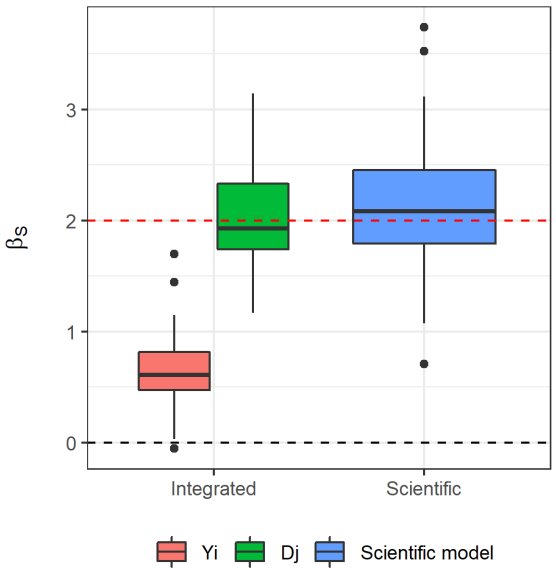
Estimation realized through TMB (Template Model Builder)
100 runs of simulation-estimation

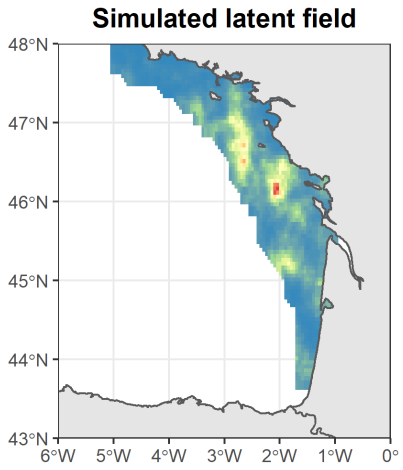
Case study: Sole in the Bay of Biscay

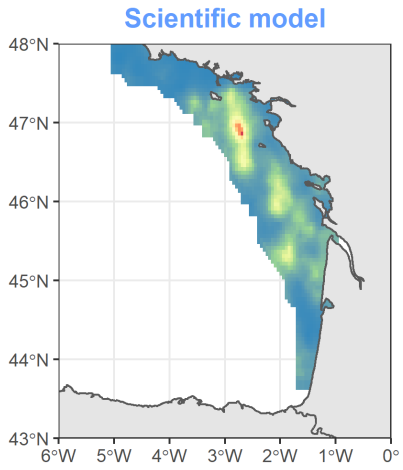
Survey data: Orhago
Commercial data: OTB-DEF traulers (to ease convergence onboard observer data were integrated in fitting)
Fitted models: same as simulation
Covariate: substrate

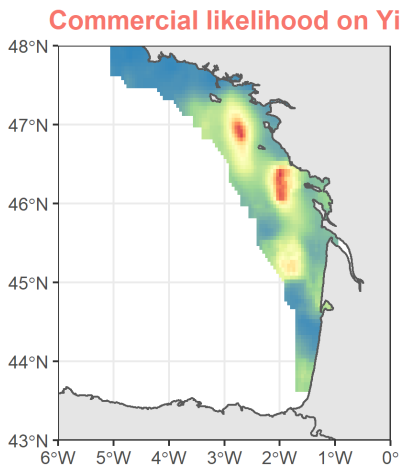
Simulation-estimation

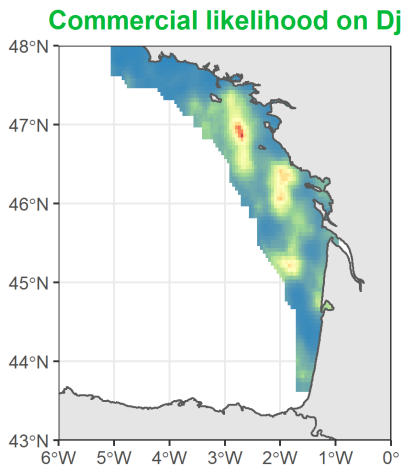




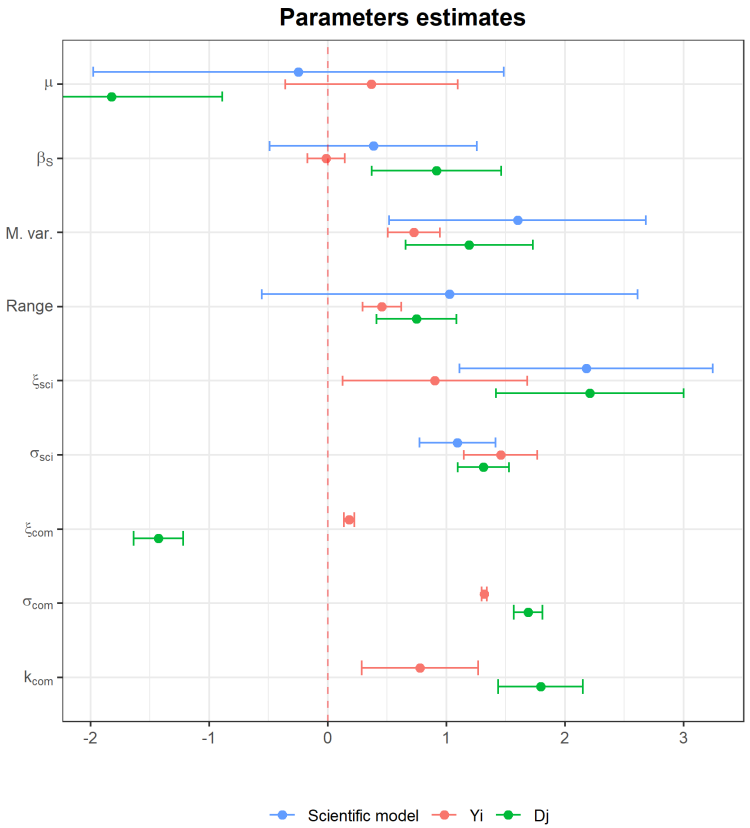


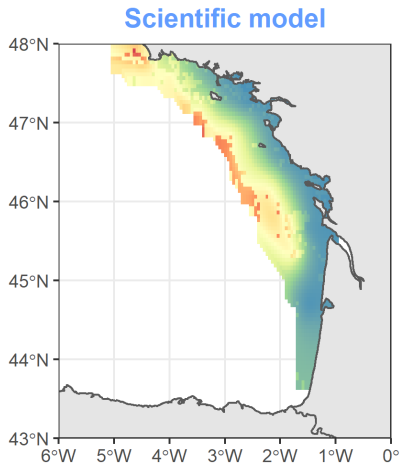


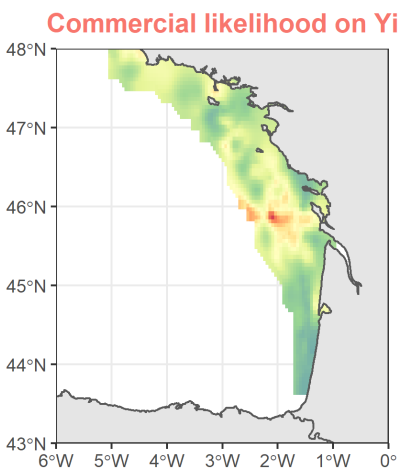


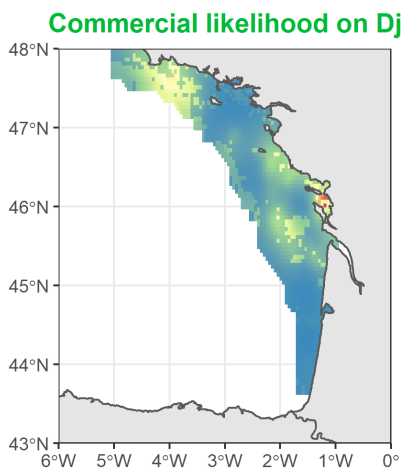


Case study: sole in the Bay of Biscay









*n.b.: punctual commercial data (i.e. observer data) needed to be integrated in inference in order to make the model converge and some parmaters needed to be step estimated (e.g. range)

Discussion

- Framework that integrates **catch declarations data** (rough resolution of catch) and **scientific/onboard observer data** (exact locations of catch)
  - **Still question:** how to ease convergence ?

- **Change of support issues**
  - not so frequent in spatial ecology,
  - but increasing amount of data available that may face such issues

- **Is it a generic framework ?**
  - the overall approach is,
    (i.e. modelling observed aggregagetd catches as a convolution of latent exact catches)
  - but need to adapt the observation model to the data
    (here zeroinflated continuous data)
  - and make the hypothesis that fishing locations ($\mathcal{P}_j$) are known
    (in reality $\mathcal{P}_j$ is a trajectory and the set of fishing points are a realization of this trajectory)

Discussion

- Framework that integrates **catch declarations data** (rough resolution of catch) and **scientific/onboard observer data** (exact locations of catch)
  - **Still question:** how to ease convergence ?

- **Change of support issues**
  - not so frequent in spatial ecology,
  - but increasing amount of data available that may face such issues

- **Is it a generic framework ?**
  - the overall approach is,
    (i.e. modelling observed aggregagetd catches as a convolution of latent exact catches)
  - but need to adapt the observation model to the data
    (here zeroinflated continuous data)
  - and make the hypothesis that fishing locations ($\mathcal{P}_j$) are known
    (in reality $\mathcal{P}_j$ is a trajectory and the set of fishing points are a realization of this trajectory)

Discussion

- Framework that integrates **catch declarations data** (rough resolution of catch) and **scientific/onboard observer data** (exact locations of catch)
  - **Still question:** how to ease convergence ?

- **Change of support issues**
  - not so frequent in spatial ecology,
  - but increasing amount of data available that may face such issues

- **Is it a generic framework ?**
  - the overall approach is,
    (i.e. modelling observed aggregagetd catches as a convolution of latent exact catches)
  - but need to adapt the observation model to the data
    (here zeroinflated continuous data)
  - and make the hypothesis that fishing locations ($\mathcal{P}_j$) are known
    (in reality $\mathcal{P}_j$ is a trajectory and the set of fishing points are a realization of this trajectory)

Discussion

- Framework that integrates **catch declarations data** (rough resolution of catch) and **scientific/onboard observer data** (exact locations of catch)
  - ➠ **Still question:** how to ease convergence ?

- **Change of support issues**
  - ➠ not so frequent in spatial ecology,
  - ➠ but increasing amount of data available that may face such issues

- **Is it a generic framework ?**
  - ➠ the overall approach is,

(i.e. modelling observed aggregagetd catches as a convolution of latent exact catches)

  - ➠ but need to adapt the observation model to the data

(here zeroinflated continuous data)

  - ➠ and make the hypothesis that fishing locations ($\mathcal{P}_j$) are known

(in reality $\mathcal{P}_j$ is a trajectory and the set of fishing points are a realization of this trajectory)

Thank you for your attention!