

Adapting the sampling design of research surveys to improve the biomass estimation of non-target species - the case study of *Raja clavata*

Daniela Silva¹, Raquel Menezes¹, Bárbara Serra-Pereira², Manuela Azevedo², Ivone Figueiredo²

¹Center of Mathematics (CMAT), University of Minho

²Portuguese Institute for the Sea and Atmosphere (IPMA)



Keywords: Spatial sampling design; Species distribution models; Two-part models, Research surveys, *Raja clavata*

Introduction

Research surveys used to monitor fishery resources of commercial interest constitute the main fishery-independent data source for stock assessment. IPMA conducts research surveys since 1980. The survey's sampling design was conceived with the focus on specific species but additionally they provide information to monitor other species. Rays and skates are important commercial by-catches whose exploitation is of generalized conservation concern given their biological and spatial dynamics characteristics.

The main aims of this study were to:

- Investigate **spatial sampling designs** for the Portuguese demersal surveys to improve the accuracy of the estimation of *Raja clavata* biomass without jeopardizing the estimation of *Merluccius merluccius* abundance.
- Propose alternative sampling designs that consider the estimated spatial distribution for both *R. clavata* biomass and *M. merluccius* abundance.

Data

The study used the demersal surveys conducted from **2013 to 2016**. Data concerns georeferenced fishing locations (hauls) performed **west off mainland Portugal**, and for each, the sampled biomass (in Kg per hour) of thornback ray (*R. clavata*) (Figure 1a). Additionally, for 2015 and 2016, total and juveniles abundance (in numbers per hour) of European hake (*M. merluccius*) caught at each fishing haul was also available (Figure 1b). We have considered two environmental variables, **bathymetry** and **seabed substratum**.

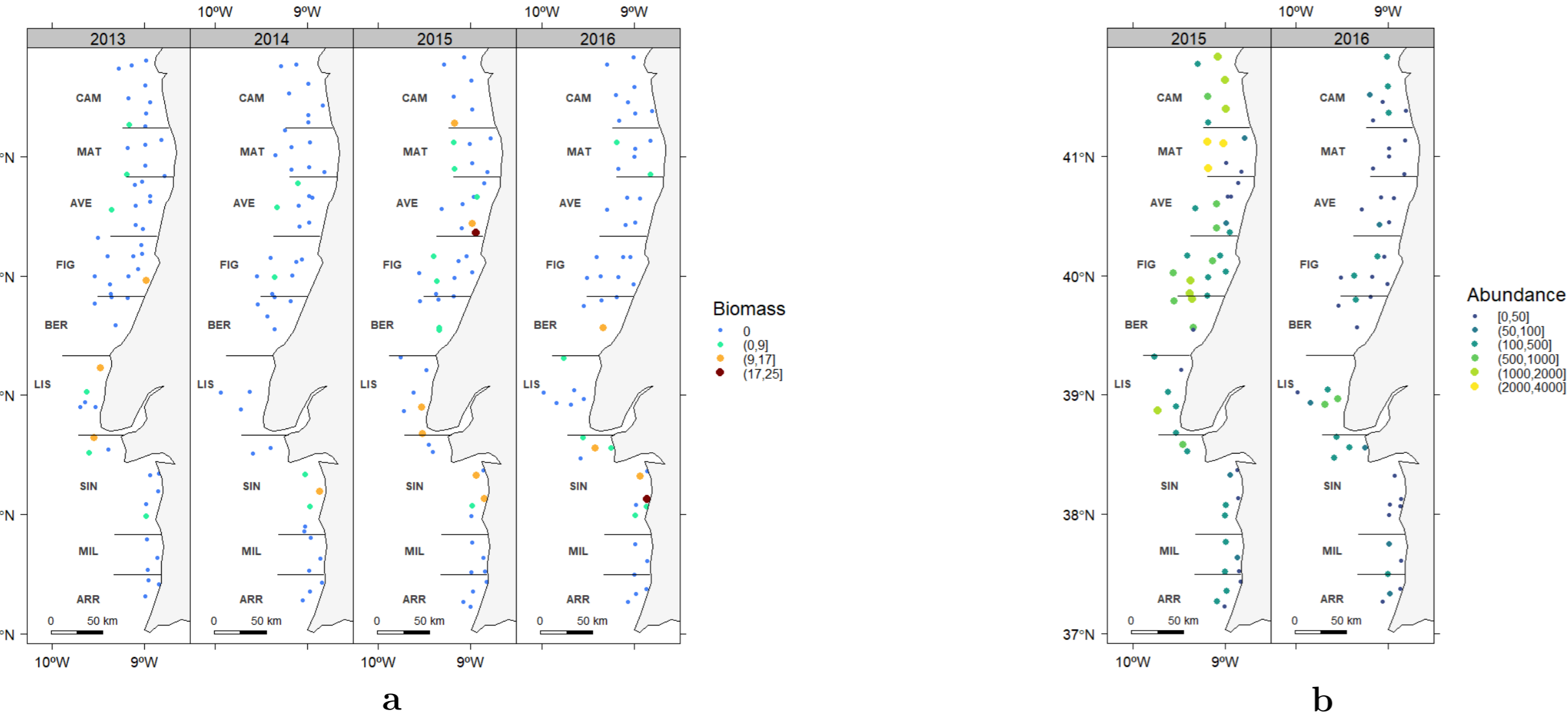


Figure 1: Exploitable biomass (Kg per hour) of *R. clavata* (left) and total abundance (number per hour) of *M. merluccius* (right) by year.

Species Distribution Model

The biomass of *R. clavata* and the abundance of *M. merluccius* were modelled with a **two-part model**: species occurrence was generated by a Bernoulli distribution and the biomass/abundance given the occurrence by a Gamma distribution.

Let $Y(\mathbf{s})$ be the spatially distributed biomass or abundance process (depending on the species) at location \mathbf{s} and $Z(\mathbf{s})$ denotes the occurrence sub-process. $Z(\mathbf{s})$ takes the value 0 if no species was observed in location \mathbf{s} and 1 otherwise. $Y(\mathbf{s})|Z(\mathbf{s}) = 1$ takes the positive value of biomass or abundance index observed in location \mathbf{s} . Therefore, $Z(\mathbf{s}) \sim \text{Bernoulli}(p(\mathbf{s}))$ and $Y(\mathbf{s})|Z(\mathbf{s}) = 1 \sim \text{Gamma}(a(\mathbf{s}), b(\mathbf{s}))$. Consequently, the two-part model can be defined by:

$$\text{logit}(p(\mathbf{s})) = \alpha_1 + \sum_{j=1}^{p_1} \beta_{i,j} f(X_{1,j}(\mathbf{s})) + W(\mathbf{s}) \quad (1)$$

and

$$\log(a(\mathbf{s})/b(\mathbf{s})) = \alpha_2 + \sum_{j=1}^{p_2} \beta_{i,j} f(X_{2,j}(\mathbf{s})) + kW(\mathbf{s}) \quad (2)$$

where $f(\cdot)$ denotes possible transformation functions (e.g., linear splines (Zuur et al., 2017) and logarithm) of environmental covariates, $X_{i,j}(\mathbf{s})$. α_1 and α_2 terms represent the intercepts, and $\beta_{i,j}$, $i = 1, 2$ are the regression coefficients (representing the impact of environmental conditions on response variable). $W(\mathbf{s})$ is a Gaussian Markov Random Field, an approximation via SPDE of an underlying Gaussian Field (Lindgren et al., 2011) and $[Y(\mathbf{s})] = [Z(\mathbf{s})][Y(\mathbf{s})|Z(\mathbf{s}) = 1]$ with $[\cdot]$ meaning “distribution of”.

Evaluation of alternative designs

The survey design to be proposed was defined according to the standard survey protocol, which includes a total of 65 stations, from which 54 were previously identified as fixed stations and 11 stations selected according to different sampling procedures.

Algorithm 1: Sampling procedure

1. Enumerate the biomass/abundance (y_s^i , $i = R, M$) and the standard deviation of spatial effects (σ_s^i , $i = R, M$) for each species i .
2. Select 11 locations according to w_s^m for each of the 8 measures m presented in Table 1.
3. Construct the alternative designs D^m merging the fixed locations and the locations selected in 2.

Table 1: Measures to select locations for the 11 fishing hauls and corresponding objectives.

Measure m	Weight w_s^m	Objective
1	$v_s^R \times v_s^M$	Minimize the uncertainty of what was not explained by the models applied to <i>R. clavata</i> and <i>M. merluccius</i> .
2	$v_s^R (1 - u_s^R) \times v_s^M (1 - u_s^M)$	Same objective of measure 1 maximizing , at same time, the biomass/abundance of both species.
3	$v_s^R (1 - u_s^R) \times v_s^M$	Same objective of measure 1 maximizing the biomass of <i>R. clavata</i> .
4	$v_s^R \times v_s^M \times q_s$	Same objective of measure 1 giving importance to sectors of the study area.
5	$\hat{\sigma}_s^R$	Minimize the uncertainty of what was not explained by the model for <i>R. clavata</i> .
6	$\hat{\sigma}_s^R (1 - u_s^R)$	Same objective of measure 5 maximizing , at same time, the biomass of <i>R. clavata</i> .
7	$\hat{\sigma}_s^R \times q_s$	Same objective of measure 5 giving importance to sectors of the study area.
8	$\sqrt{(\sqrt{u_s^R} - \sqrt{u_s^M})^2}$	Maximize the similarity between the distributions of the two species.

where $u_s^i = \frac{\hat{y}_s^i}{\max\{\hat{y}_s^i\}}$ and $v_s^i = \frac{\hat{\sigma}_s^i}{\max\{\hat{\sigma}_s^i\}}$. q_s means the sector weight for location \mathbf{s} . The sector weight takes integers values from 1 to 9, where the sector with 1 corresponds to sector with the highest estimated mean of *R. clavata* biomass and the sector identified by 9 is the sector with the lowest estimated mean.

Algorithm 2: Performance of sampling designs

1. Estimate the biomass of *R. clavata* and the abundance of *M. merluccius*, in the locations observed in 2015, based on the estimates of each D^m obtained in Algorithm 1.
2. Compute Root Mean Squared Error $RMSE^i(m)$, Mean Absolute Error $MAE^i(m)$, stratified mean $\bar{y}_{str}^i(m)$ and respective stratified variance $s_{y_{str}^i}^2(m)$.
3. Repeat 1 and 2 based on 200 alternative designs where the locations of the 11 fishing hauls were selected according to a spatial randomness method.

Main results

The highest biomass values of *R. clavata* were observed in the area near Lisbon, particularly in the last two years. In most fishing hauls, the percentage of zero catches of adults varied between 72% and 88% depending on the year. The 2015 survey abundance of *M. merluccius* was higher than in 2016, namely in the north. The results of the Species Distribution Models show that:

- **the probability of occurrence of *R. clavata* increases** for locations corresponding to the mixed sediment substratum, but **decreases** with bathymetry in this substratum;
- ***R. clavata* biomass decreases** with bathymetry;
- ***M. merluccius* abundance increases** with bathymetry, by 15 specimens until 92 m depth and 31 specimens in deeper waters (both for an increase of about 2.7 m in bathymetry);
- ***M. merluccius* is more than twice abundant** in substratum mud to muddy sand locations.

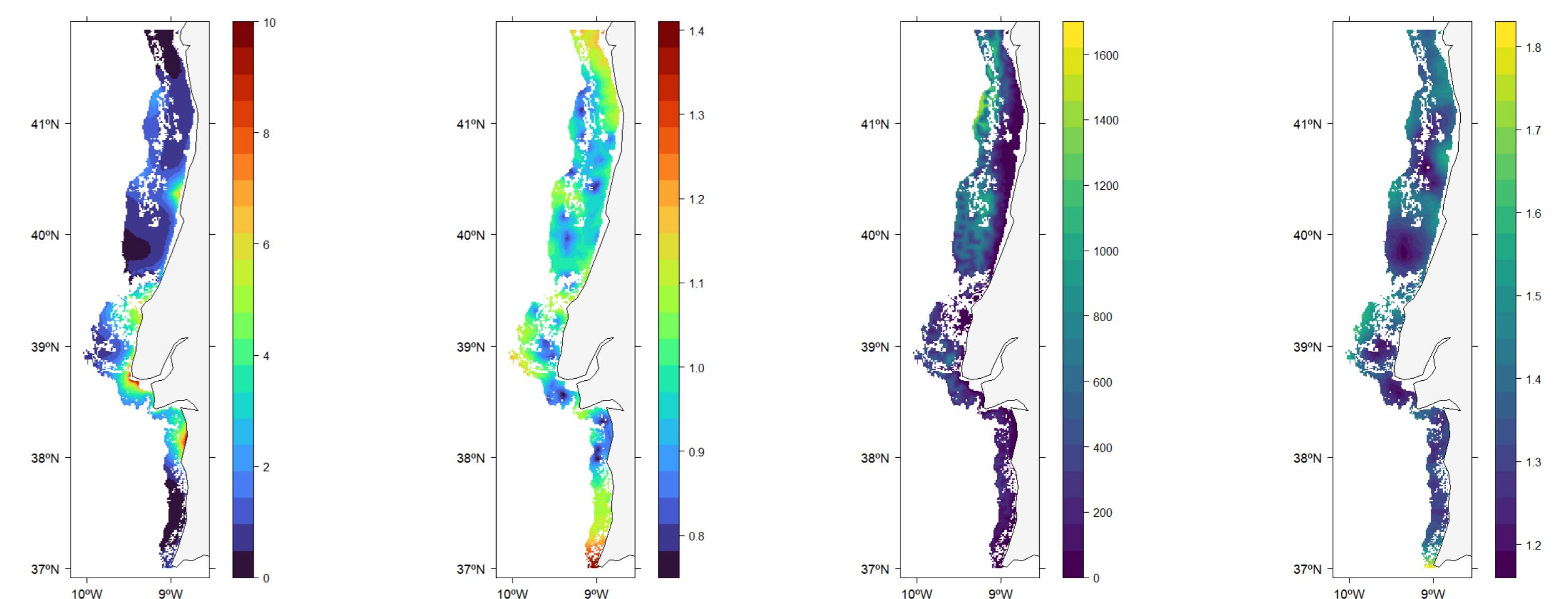


Figure 2: Posterior median of the predictive distribution of biomass (left panel) and posterior standard deviations (right panel) of spatial effects of *R. clavata*.

Figure 3: Posterior median of the predictive distribution of abundance (left panel) and posterior standard deviations (right panel) of spatial effects of *M. merluccius*.

As an example, Figure 4 shows the densities of 200 values of MAE and RMSE resulting from the estimation of *R. clavata* biomass and *M. merluccius* total abundance using the random samples and the 3 best sampling designs.

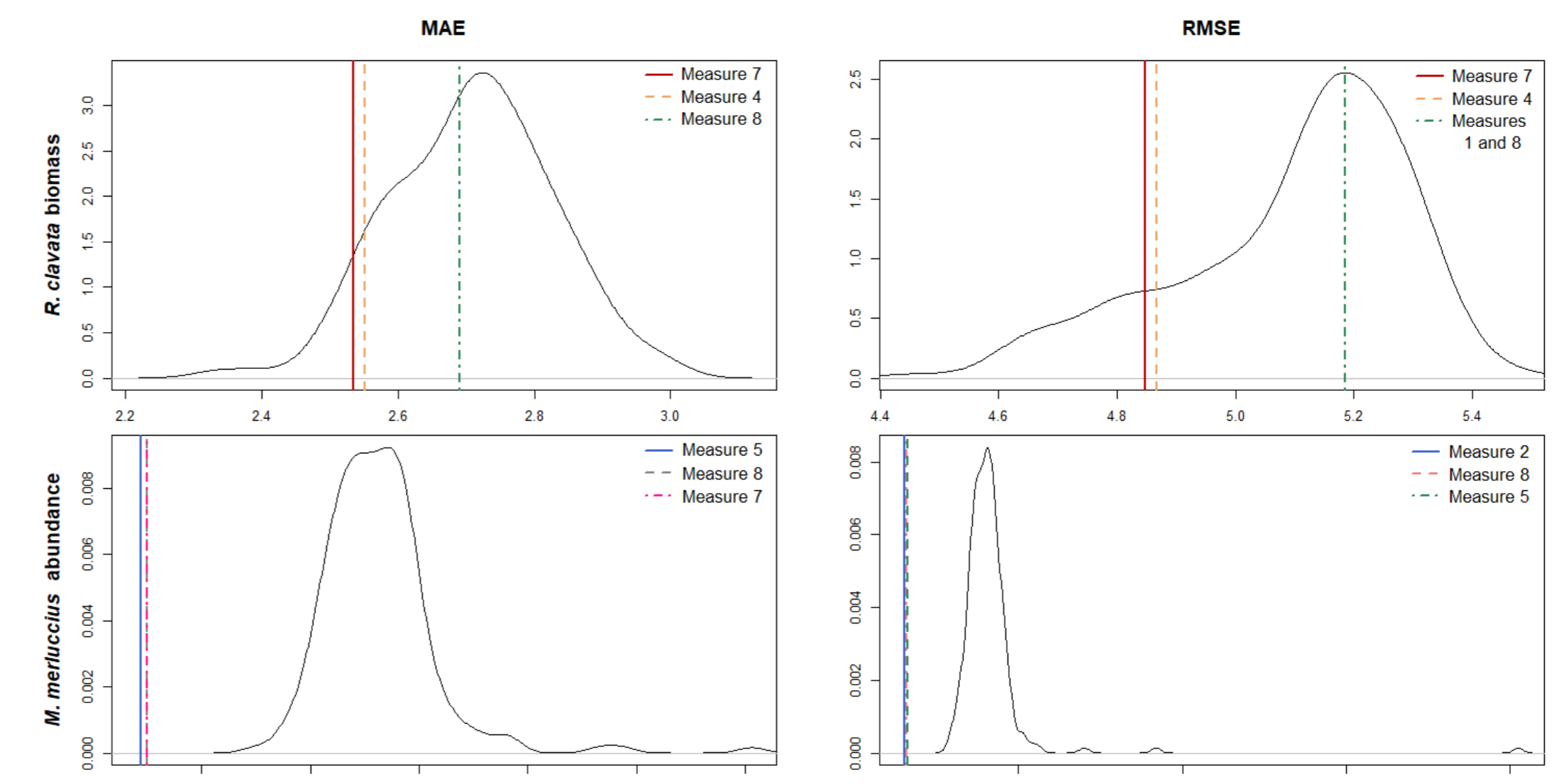


Figure 4: Densities of 200 values of MAE (left panels) and RMSE (right panels) resulting from estimation for *R. clavata* biomass (top panels) and *M. merluccius* abundance (bottom panels) using the random samples. Vertical lines represent the 3 best proposed surveys (with lowest values of MAE and RMSE).

Conclusions

- Sampling designs 4, 7 and 8 (highlighted in Table 1) resulted in higher accuracy of *R. clavata* biomass and *M. merluccius* abundance estimates than that obtained with a random selection of fishing stations.
- Sampling design 7 maximizing the accuracy of *R. clavata* biomass estimates shows an acceptable trade-off between bias and variance of *M. merluccius* abundance estimates.
- The approach presented in this study can be easily replicated to analyse other sampling efforts and/or other group of species caught by the research survey.

References

- Consortium, E. B. (2020). EMODnet Digital Bathymetry (DTM 2020).
- Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Statistical Methodology*, 73(4):423–498.
- Zuur, A. F., Ieno, E. N., and Saveliev, A. A. (2017). *Beginner's Guide to Spatial, Temporal, and Spatial-Temporal Ecological Data Analysis with R-INLA*, volume I: Using GLM and GLMM, chapter 20: GAM with correlation and zero-inflation in R-INLA using owl data. Highland Statistics Ltd.
- Acknowledges:** Supported by Portuguese funds through the Centre of Mathematics and the Portuguese Foundation for Science and Technology (FCT), within projects UIDB/00013/2020, UIDP/00013/2020 and “PREFERENTIAL”-PTDC/MAT-STA/28243/2017. The first author also acknowledges was The authors acknowledge FCT for funding this research through Individual Scholarship PhD. PD/BD/150535/2019. The data used in this work was collected under the European Commission's Data Collection Framework - PNAB/EU-DCF Programa Nacional de Amostragem Biológica, (Reg. EC 2008/199). Data used in this publication was made available by the EMODnet Bathymetry project www.emodnet-bathymetry.eu, and the EMODnet Geology project, <http://www.emodnet-geology.eu>, funded by the European Commission Directorate General for Maritime Affairs and Fisheries. The data layer of bathymetry used for this work was Consortium (2020). These seabed substrate data were collected by Geological Survey of Finland GTK.