Ecological Informatics 84 (2024) 102858

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Bayesian feedback in the framework of ecological sciences

Mario Figueira [a,*], Xavier Barber [b], David Conesa [a], Antonio López-Quílez [a], Joaquín Martínez-Minaya [c], Iosu Paradinas [d], Maria Grazia Pennino [e]

[a] Departamento de Estadística e Investigación Operativa. Universitat de València, Spain
[b] Centro de Investigación Operativa.Universidad Miguel Hernández de Elche, Spain
[c] Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Spain
[d] Azti, Txatxarramendi Ugartea z/g, 48395 Sukarrieta, Spain
[e] Instituto Español de Oceanografía (IEO-CSIC), Spain

ARTICLE INFO

Dataset link: https://github.com/MarioFigueiraP/Feedback_code_article

Keywords:
Hierarchical spatial models
INLA
Preferential sampling
Prior updating
Species distribution models

ABSTRACT

In ecological studies, it is not uncommon to encounter scenarios where the same phenomenon (e.g., species occurrence, species abundance) is observed using two different types of samplers. For example, species data can be collected from scientific sampling with a completely random sample pattern, but also from opportunistic sampling (e.g., whale watching from commercial fishing vessels or bird watching from citizen science), where observers tend to look for particular species in areas where they expect to find them.

Species Distribution Models (SDMs) are widely used tools for analysing this type of ecological data. In particular, two models are available for the aforementioned data: a geostatistical model (GM) for data collected where the sampling design is not directly related to the observations, and a preferential model (PM) for data obtained from opportunistic sampling.

The integration of information from disparate sources can be addressed through the use of expert elicitation and integrated models. This paper focuses on a sequential Bayesian procedure for linking two models by updating prior distributions. The Bayesian paradigm is implemented together with the integrated nested Laplace approximation (INLA) methodology, which is an effective approach for making inference and predictions in spatial models with high performance and low computational cost. This sequential approach has been evaluated through the simulation of various scenarios and the subsequent comparison of the results from sharing information between models using a variety of criteria. The procedure has also been exemplified on a real dataset.

The primary findings indicate that, in general, it is preferable to transfer information from the independent (with a completely random sampling) model to the preferential model rather than in the alternative direction. However, this depends on several factors, including the spatial range and the spatial arrangement of the sampling locations.

1. Introduction

Species distribution models (SDMs) have been widely used in ecological analysis for a multitude of purposes, such as making inference of ecological niches (Guisan and Zimmermann, 2000), the assessment of climate change impacts on habitats (Karp et al., 2023), the prediction of future invader and invasive species (Fournier et al., 2019; Landis et al., 2000), the suggestion of areas for protection (Paradinas et al., 2022), and the refinement of biodiversity inventories (Staniczenko et al., 2017).

In this framework, it is not uncommon to have multiple sources of information for the same ecological phenomenon. For instance, in fisheries ecology, data may originate from scientific surveys and/or commercial vessels (Braun et al., 2023). This raises the question of how to integrate information from disparate sources. Two distinct approaches exist for achieving this integration: integrated models and expert elicitation.

Integrated species distribution models (iSDMs) represent a suite of novel approaches that combines different sources of information to construct a unified model. In this approach, it is possible to jointly handle different data, which may originate from citizen science, scientific surveys, commercial fishing surveys, and so forth. This implies not only integrating disparate data sets, but also integrating various sampling structures to reduce the inherent biases associated with sampling

* Corresponding author.
E-mail address: Mario.Figueira@uv.es (M. Figueira).

https://doi.org/10.1016/j.ecoinf.2024.102858
Received 12 September 2023; Received in revised form 9 September 2024; Accepted 14 October 2024
Available online 24 October 2024

designs, thereby leveraging joint information to enhance inferential capacity (Koshkina et al., 2017; Fletcher, Jr. et al., 2019). In particular, Ahmad Suhaimi et al. (2021) propose various approaches for constructing a joint model that incorporates both presence-only and presence-absence data. Rufener et al. (2021), Alglave et al. (2022) and Paradinas et al. (2023) also present methods for integrating independent data and opportunistic or dependent data into a single model analysis overcoming the bias and the scarcity of information derived from isolated surveys or sample sets. In addition, Jung (2023) describes `ibis.iSDM`, a modelling framework for iSDMs, that allows the integration of different data sources into a single model, supporting parameter transformations, tuning and spatio-temporal projections.

In the Bayesian framework, expert elicitation involves gathering expert opinions and insights to incorporate additional information into the analysis. This is primarily achieved by defining prior distributions for certain parameters or by setting specific constraints on the values these parameters can take. This process involves multiple steps, including defining the format and model of the elicitation process and recruiting appropiate experts (O'Hagan, 2006; Dias et al., 2018). Expert elicitation has been shown to be particularly effective in the context of SDMs (Burgman et al., 2011; Vanhatalo et al., 2014; Nevalainen et al., 2018; LaMere et al., 2020; Kaurila et al., 2022), enabling the acquisition of information in unsampled areas, correcting biases and underreported values, and identifying the most reliable experts. In particular, Crawford et al. (2020) employed expert opinion to shape habitat suitability models for conservation planning in the US; Di Febbraro et al. (2018) evaluated the feasibility of monitoring habitat quality for bird communities in central Italy using a blend of survey data and expert-driven models; and Pearman-Gillman et al. (2020) leveraged expert elicitation to characterise the distribution of various species in New England, using a web questionnaire to extract species presence probabilities and insights into covariate effects on species occurrence.

A specific form of expert elicitation involves leveraging information from previous experiments or studies by defining prior distributions that incorporate insights from earlier results. If a new experiment shares a similar mathematical structure with a previous one, the parameter estimates from the earlier study can be used to enhance the accuracy of the new one. This method enables shared parameters between the models to use the knowledge gained from past experiments, effectively refining the conditions for analysis of the current experiment based on prior data.

The objective of this study is to propose a protocol for implementing a Bayesian feedback approach that considers two distinct types of sampling (independent and preferential), each of which is analyses with a specific model. In particular, independent sampling models (IMs) focus on geostatistical processes (Diggle et al., 1998), which can incorporate biotic (Barber et al., 2021), spatial and spatio-temporal effects (Paradinas et al., 2015; Martínez-Minaya et al., 2018). Preferential sampling models (PMs) are analyses using marked point pattern models, which consider that the sampled quantities of the ecological phenomenon of interest (i.e., species occurrence or abundance) are influenced by the sampling process (Diggle et al., 2010; Pennino et al., 2016, 2019). Both models can use Bayesian hierarchical modelling, and prior information can be integrated through informative prior distributions. The prior elicitation that occurs through Bayesian feedback between these two models can be performed in either direction, from the IM to the PM or vice versa.

This paper presents two methodologies for performing feedback between spatial Bayesian hierarchical models fitted with the integrated nested Laplace approximation (INLA, Rue et al. 2009, 2017) approach. The first methodology involves the full Bayesian updating of the marginal distributions, whereby the prior distributions of one model are replaced by the posterior distributions of the other. The second methodology updates the characteristic moments (mean, variance, mode, quantiles, etc.) of the prior distributions on one model based on the moments of the posterior distributions of the other model.

In order to validate the two procedures, we present here a set of simulated scenarios through which we compare the different directions of the two feedback procedures. The evaluation and assessment of the behaviour of both procedures is conducted through the analysis of residuals from predictive maps, encompassing metrics such as root mean square error (RMSE), bias, histograms of residuals, and residual plots against predicted values. These simulated environments also permit the identification of potential biases in parameter and hyperparameter estimation. Finally, we present an application of the proposed method in the context of a real fishery scenario. Specifically, the distribution of the European hake (*Merluccius merluccius*) off the southern French coast of the Bay of Biscay is studied. In the analysis, we combine information gathered from two distinct data sources: fishery independent samples collected through the French EVHOE fishery trawl survey (FI samples), and fishery dependent samples collected through onboard sampling of Basque pair trawlers.

2. Species distribution models

SDMs are statistical models that link biological data related to organism populations to some explanatory variables. The response variable of the model, biological population data, can be classified into these categories: presence-only data, presence/absence data, proportional data, discrete abundance data and continuous biomass data. Data may originate based on random or stratified field sampling. Observations may also originate opportunistically. Guisan and Zimmermann (2000), Guisan and Thuiller (2005), Elith and Leathwick (2009). The type of data related to explanatory variables can be either biotic (e.g. depredatory species distribution, tree cover density, etc.) or abiotic (environmental data such as soil data, temperature, salinity, etc.). These are selected to reflect the three main types of effects on the species population data: (i) *limiting factors* (or *regulators*), (ii) *disturbances* and (iii) *resources* (Guisan and Thuiller, 2005).

In this study, two statistical models will be employed: geostatistical (GM) and preferential (PM), in accordance with the notation proposed by Martínez-Minaya et al. (2018). The geostatistical model comprises a predictor with three elements: (i) an intercept, (ii) a linear effect for a spatial covariate, and (iii) a structured random spatial effect. Conversely, the preferential model will include the same three elements, but they will be structured in two submodels. The first submodel will be geostatistical, similar to the independent one, and the second will be a point process submodel for the sample structure. This submodel will include an additional spatial effect that modifies the preferential sampling process.

2.1. Geostatistical model

In general, a geostatistical model assumes that data are generated from a continuous spatial process and are constituted by the measurements of the phenomenon under study.

For the sake of simplicity, our geostatistical model will be formed by an intercept β_0, a linear effect β_1 for a spatial covariate X_i and a structured spatial random effect $\mathbf{u}$. The distribution for the response variable could be any distribution, but we exemplified the geostatistical model using a Gamma distribution because it is consistent with biomass or abundance distribution scenarios. Therefore, the response variable Y_i will follow a Gamma distribution with mean μ_i and variance ϕ, and the overall structure can be described as follows:

$$
\begin{aligned}
Y_i &\sim \text{Gamma}(\mu_i, \phi), \\
\log(\mu_i) &= \beta_0 + \beta_1 \cdot X_i + u_i, \\
\mathbf{u} &\sim N(\mathbf{0}, \mathbf{Q}(\rho, \sigma)), \\
\beta_0 &\sim N(0, \tau_0),\ \beta_1 \sim N(0, \tau_1), \\
\rho &\sim f_\rho(\rho \mid \boldsymbol{\alpha}_\rho),\ \sigma \sim f_\sigma(\sigma \mid \boldsymbol{\alpha}_\sigma),
\end{aligned}
\tag{1}
$$

where τ_0 and τ_1 are the precision for prior distributions of the intercept and the linear effect of the covariate, respectively, and f_ρ and f_σ are the prior distributions for the spatial range ρ and the marginal standard deviation σ of the spatial effect. The characteristic parameters of these prior distributions are $\boldsymbol{\alpha}_\rho$ and $\boldsymbol{\alpha}_\sigma$, which are known. A specific subsection will be devoted to priors, establishing the different options that can be used for spatial effects. From this point forward, $\boldsymbol{\theta}$ will be used to denote the complete set of fixed and random effects comprising the latent field, e.g. $\boldsymbol{\theta} = \{\beta_0, \beta_1, \mathbf{u}\}$, while $\boldsymbol{\psi}$ will be used to denote the complete set of the hyperparameters related to the latent field and the likelihood, e.g. $\boldsymbol{\psi} = \{\tau_0, \tau_1, \rho, \sigma\}$.

The simplicity of this structure will allow us to discern more clearly the variations in the fixed parameter estimates and to assess these differences according to the feedback mechanism.

2.2. Preferential sampling

The second model employed in this analysis is a preferential model. As with the geostatistical model, we have a series of locations, denoted by $\mathbf{s} = \{s_1, \ldots, s_n\}$. However, in contrast to the geostatistical model, which assumes that the locations are random and do not share information with the marks of the sampling points, the locations in the preferential model are generated by a non-homogeneous Poisson process with intensity λ, which is a log-Gaussian Cox process (LGCP) relative to the geostatistical marking process. Therefore, preferential sampling is a specific case of the broader point process class referred to as a *marked point pattern* (Diggle, 2013).

The preferential model can be regarded as a two-stage model that shares information through some common components (Krainski et al., 2018). In particular, if Y_i represents the response variable of the quantity of interest and it is Gamma distributed with mean μ_i and variance ϕ, then the structure of the preferential model can be expressed as follows:

$$\begin{aligned} Y_i | s_i &\sim \text{Gamma}(\mu_i, \phi), \\ s_i &\sim \text{LGCP}(\lambda_i), \\ \log(\mu_i) &= \beta_0 + \beta_1 \cdot X_i + u_i, \\ \log(\lambda_i) &= \gamma \cdot (\beta_0 + \beta_1 \cdot X_i + u_i) + u_i^*. \end{aligned} \tag{2}$$

In this equation, most of the parameters have already been discussed in the geostatistical model, except for u_i^* and γ. The u_i^* represents a specific spatial effect associated with the generating location process. It enables the consideration of other spatially structured elements that may influence the point process but are not directly linked to the response variable of the geostatistical process. Meanwhile, the γ parameter enables the establishment of a linear scaling in the shared components (Gómez-Rubio, 2020). In our case, this scale refers to the transformation of the intensity of the coefficients of the covariates and the spatial effect with respect to the geostatistical process, while maintaining the same spatial range of the spatial effect.

3. Inference

In this work, the analyses are performed using Bayesian hierarchical models and the Integrated Nested Laplace Approximation (INLA, Rue et al. 2009, 2017) approach. This approach, implemented in the `R-INLA` software (Martins et al., 2013), has become a well-established tool for Bayesian inference in numerous research fields (Blangiardo and Cameletti, 2015; Lindgren and Rue, 2015), including ecology (Cosandey-Godin et al., 2015; Pennino et al., 2019; Paradinas et al., 2017), epidemiology (Blangiardo et al., 2013; Moraga, 2019) and econometrics (Gómez-Rubio et al., 2021), due to its versatility and high performance. This will allow us to carry out protocols for the sharing of information between models. In particular, the Stochastic Partial Differential Equation (SPDE) approach is applied in conjunction with *finite element methods* (FEMs) in order to evaluate the spatial structure, thereby allowing for a fast and cost-effective computational resolution of spatial latent effects.

In INLA, two principal methods are employed for the formulation of the prior distributions of the spatial effect. One approach uses exponential transformations of Normal distributions (EN-priors), while the other utilises penalised complexity distributions (PC-priors).

The initial step in the first type of prior distributions is to define an initial value for the range and variance, as well as the precision of a null-mean Normal distribution that is exponentially transformed (Lindgren and Rue, 2015):

$$\begin{aligned} \rho &= \rho_0 \cdot \exp(\theta_1), \quad \theta_1 \sim N(0,1), \\ \sigma &= \sigma_0 \cdot \exp(\theta_2), \quad \theta_2 \sim N(0,1). \end{aligned} \tag{3}$$

Consequently, as the mean is fixed to zero (which implies that the mode of $p(\rho)$ and $p(\sigma)$ are ρ_0 and σ_0, respectively) and the Normal distribution is transformed by an exponential function, we have a positive definite distribution for the hyperparameters of the spatial effect. Moreover, this allows us to define an uninformative prior by setting a low precision while preserving the positive definite condition. However, the posterior distributions of these hyperparameters would lack interpretability, as they would be provided with respect to the characteristic parameters of the Normals. As a result, direct interpretation of the hyperparameters (ρ and σ) of the spatial effect is not feasible unless the exponential transformation illustrated in Eq. (3) is performed.

The second type of prior distributions was designed with interpretability in mind. The penalised complexity prior distributions, proposed in Simpson et al. (2017) and extended in Sørbye et al. (2018) and Fuglstad et al. (2019), are defined by two elements: (i) a fixed value ρ_0 or σ_0, and (ii) a probability value p_0, which indicates the probability above or below the initial value (ρ_0, σ_0). Consequently, the prior distributions for ρ and σ are defined as follows

$$\begin{aligned} \rho &\sim \text{PC-prior}(\rho_0, p_0), \\ \sigma &\sim \text{PC-prior}(\sigma_0, p_0), \end{aligned} \tag{4}$$

where this PC-prior differs in the definition of tail probability depending on whether it refers to the spatial range (ρ) or to the marginal standard deviation (σ). In particular,

$$\begin{aligned} \text{PC-prior}(\rho_0, p_{\rho 0}) &\equiv P(\rho < \rho_0) = p_0, \\ \text{PC-prior}(\sigma_0, p_{\sigma 0}) &\equiv P(\sigma > \sigma_0) = p_0. \end{aligned} \tag{5}$$

For instance, defining $p_0 = 0.5$ means that the probability of ρ being less than ρ_0 and σ being greater than σ_0 are equal to 0.5, setting the median of the prior distribution.

4. Bayesian feedback

This section presents a proposal for implementing a Bayesian feedback procedure within the context of species distribution models. The basic scheme starts with the fitting of model M_1 (either a geostatistical or a preferential model) with data $\mathbf{y}_1$, resulting in posterior distributions of parameters and hyperparameters $\pi(\theta, \boldsymbol{\psi} | M, \mathbf{y}_1)$. These distributions are then used to feed back the inferential process of fitting another model M_2 (the corresponding contrary model considered in M_1) for a new data set $\mathbf{y}_2$. For instance, when analysing the distribution of European hake (*Merluccius merluccius*) off the southern French coast of the Bay of Biscay, as described later, we could first analyse the information gathered from fishery independent samples collected by a trawl survey, and then use the posterior distribution of the parameters of the geostatistical model to feed back the fishery dependent samples collected by onboard sampling of Basque pair trawlers. Or vice versa.

The remainder of this section will first introduce two different updating protocols, and then examine the two possible feedback situations that can arise when having two different models: the feeding-back of a geostatistical model with information from a preferential model, and vice versa.

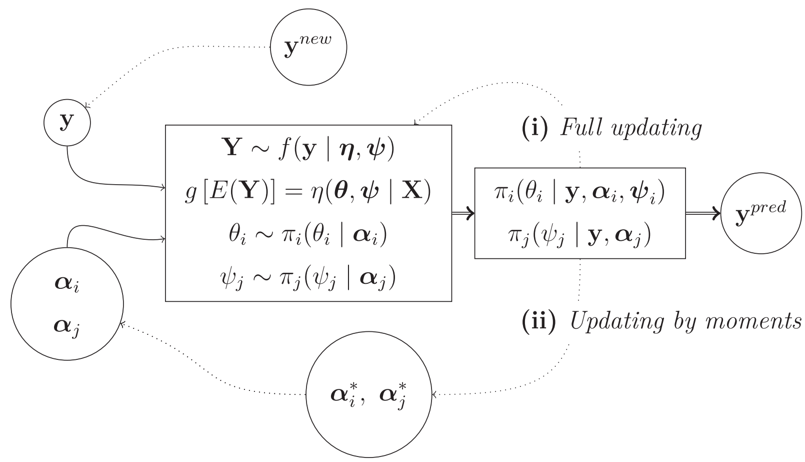


Fig. 1. Illustration of the two proposals for incorporating information gathered from a previous fitting in a new analysis. The full updating procedure can be applied to hyperparameters, as the latent field is defined as a multivariate Gaussian distribution. Meanwhile, the updating by moments procedure can be applied to both fixed parameters and hyperparameters.

4.1. Updating structures for feed-back

In Fig. 1 two proposals for Bayesian feedback are represented: (i) full updating and (ii) updating by moments. In both procedures, we start with a data set $\mathbf{y}$ that is analysed by a model determined by the likelihood $f(\mathbf{y} \mid \boldsymbol{\eta}, \boldsymbol{\psi})$, the model for the mean $E(\mathbf{Y})$ linked to the linear predictor $\eta(\theta, \boldsymbol{\psi} \mid \mathbf{X})$ by the link function $g(\cdot)$, where $\mathbf{X}$ is the design matrix of the model related to the covariates, and the prior distribution of the latent field $\boldsymbol{\theta}$ and hyperparameters $\boldsymbol{\psi}$. The prior distribution of the latent field $\pi(\theta_i \mid \boldsymbol{\alpha}_i)$ and the hyperparameters $\pi(\boldsymbol{\psi}_j \mid \boldsymbol{\alpha}_j)$ are determined by some characteristic parameters $\boldsymbol{\alpha}_i$ and $\boldsymbol{\alpha}_j$. Once the posterior distributions of the latent field and the hyperparameters are obtained, we can apply two procedures for Bayesian feedback to fit a new data set, taking advantage of the available information from the posterior distributions.

The two procedures to perform Bayesian feedback are as follows:

- **(i) Full updating:** this protocol simply involves replacing the prior distributions with the posterior distributions. It is conceptually and naturally the simplest feedback procedure in Bayesian statistics. This procedure can be implemented by providing the distribution in tabular form (x, y) with enough points covering a sufficiently wide interval for the exploration of the distribution. However, this must be provided in the internal parameterisation used by INLA.
- **(ii) Updating by moments:** this approach assumes that prior distributions can be analytically defined by updating the characteristic parameters of the kernel in the prior distributions with their estimates from the posteriors. For example, if a distribution is characterised by its mean and variance, both parameters could be updated with the corresponding values from the posterior distribution.

The first procedure can be implemented in INLA for the hyperparameters, but not for the elements of the latent field, as the latent field is defined as Gaussian. Therefore, for the fixed parameters, the second procedure must be implemented, where the mean and precision of the Gaussian distribution associated with each fixed parameter will be updated. Evidently, the second procedure can also be applied to the hyperparameters. This means that if we want to implement Bayesian feedback on a complex model, the fixed parameters are restricted to being updated through moment updating, while for the hyperparameters, we can implement either of the two procedures; even applying one procedure to a subset of hyperparameters and the other to the rest.

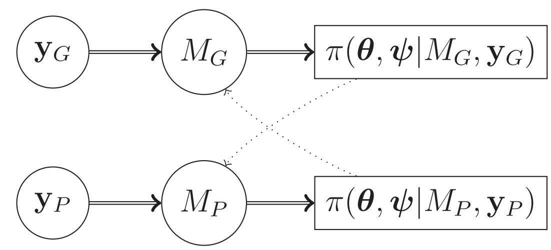


Fig. 2. Scheme for Bayesian feedback between geostatistical and preferential models. In this case, the feedback process is asymmetric, because thegeostatistical model and preferential model have different parameters and hyperparameters.

It is important to note that regardless of the approach used, it is not possible to simultaneously report the random effects of the latent field and the hyperparameters associated with the prior distributions of these random effects. Therefore, throughout this study, we do not consider including information about the random effects of the latent field. Only information about the fixed effects and the hyperparameters of the latent field are included.

Our focus is on describing the two possible feedback situations mentioned above (feeding-back a geostatistical model with information from a preferential model and vice versa). Fig. 2 describes scheme for the Bayesian feedback for the two datasets $\mathbf{y_G}$ and $\mathbf{y_P}$, given two model structures M_G and M_P, and the posterior distributions $\pi(\theta, \boldsymbol{\psi} | M_G, \mathbf{y_G})$ and $\pi(\theta, \boldsymbol{\psi} | M_P, \mathbf{y_P})$.

The main difference between the two schemes is that the feedback occurs between two distinct models. There are parameters common to both models that can be updated using the posterior distributions from the other model. This results in an asymmetrical process because, given the posterior distributions of the preferential model, all parameters and hyperparameters of the geostatistical model can be updated, but not vice versa.

4.2. Geostatistical model feedback

Firstly, we will study the feedback of the geostatistical models by assuming the fit of the preferential model. As previously stated with respect to feedback schemes, the feedback will be conducted by replacing the prior distributions with the posterior ones or by updating the characteristic parameters of the prior distributions in accordance with the estimation of the same from the posteriors of the common parameters and hyperparameters.

Therefore, upon consideration of the parameters and hyperparameters of the geostatistical model and those of the preferential model, it becomes evident that the feedback relationship between the two is asymmetric. This is due to the fact that, starting from the PM, all the parameters and hyperparameters of the IM can be updated, but not vice versa. In Eq. (6), we can compare the whole set of parameters and hyperparameters.

$$\text{GM} \equiv \left\{ \begin{array}{c} \boldsymbol{\theta} \\ \hline \beta_0 \sim N(0, \tau_{(\beta_0)}), \\ \beta_1 \sim N(0, \tau_{(\beta_1)}). \\ \hline \boldsymbol{\psi} \\ \hline \rho \sim \text{PC-prior}(\rho_0, p_\rho), \\ \sigma \sim \text{PC-prior}(\sigma_0, p_\sigma), \\ \log(\phi) \sim \log-\text{Gamma}(\mu_{(\phi)}, \phi_{(\phi)}). \end{array} \right\},$$
$$\text{PM} \equiv \left\{ \begin{array}{c} \boldsymbol{\theta} \\ \hline \beta_0 \sim \text{N}(0, \tau_{(\beta_0)}), \\ \beta_1 \sim \text{N}(0, \tau_{(\beta_1)}), \\ \hline \boldsymbol{\psi} \\ \hline \rho \sim \text{PC-prior}(\rho_0, p_\rho), \\ \sigma \sim \text{PC-prior}(\sigma_0, p_\sigma), \\ \alpha \sim \text{N}(0, \tau_{(\alpha)}), \\ \log(\phi) \sim \log-\text{Gamma}(\mu_{(\phi)}, \phi_{(\phi)}). \end{array} \right\}. \tag{6}$$

In short, the feedback process involves updating the set of fixed parameters $\{\mu_{(\beta_0)}, \tau_{(\beta_0)}, \mu_{(\beta_1)}, \tau_{(\beta_1)}, \rho_0, p_\rho, \sigma_0, p_\sigma, \mu_{(\phi)}, \phi_{(\phi)}\}$ with their estimated analogues from the posterior distributions of the latent effects and PM hyperparameters, common in GMs. Regarding the prior distributions of the spatial effect hyperparameters, we will employ PC-priors for the base models (the model without any feedback), as they allow easy definition of vague prior distributions. However, for the feedback, we will define Normal distributions for the logarithmic transformation of these hyperparameters, as they are more flexible for defining informative prior distributions:

$$\begin{aligned} \log(\rho) &\sim \log(\rho_0) + \text{N}(0, \tau_\rho), \\ \log(\sigma) &\sim \log(\sigma_0) + \text{N}(0, \tau_\sigma). \end{aligned} \tag{7}$$

When applying this feedback approach, it is important to consider potential identification issues. These challenges would typically stem from those encountered when analysing each model separately, rather than being new problems introduced by the approach itself. In principle, as long as both individual models are free from identification problems, no additional issues should arise. However, another concern is the potential for bias in the models for specific datasets, which could be transferred to subsequent models through updated priors. Still, if identification problems are expected in certain parameters of one model, providing informative priors from the other model can help alleviate these issues.

4.3. Preferential model feedback

We now explain how to implement the feedback of the preferential models by assuming the fit of the geostatistical model. As previously noted, this feedback mechanism differs from that of the geostatistical model. In particular, it is not feasible to directly incorporate all the parameters associated with the likelihood of the point process or the hyperparameter γ solely from the results of the geostatistical model. Consequently, to integrate the full set of parameters and hyperparameters into the feedback loop, it is necessary to perform a point process fitting, which would yield the latent effects and the hyperparameters of range and marginal standard deviation. For γ, an estimate could be made assuming normality and using uncertainty propagation, typically through the Laplace approximation (see Appendix). However, since this approach does not appear to significantly improve the fitting result, we have chosen to automate the process of integrating feedback from the preferential process, thereby circumventing the need for an additional fitting procedure. In essence, the feedback process of the preferential model involves incorporating all the parameters of the geostatistical model, except for the random effects of the latent field, as mentioned earlier.

Table 1
Values for the parameters and hyperparameters to simulate the different scenarios.

β_0	β_1	ρ	σ	ϕ
−1	2	(0.2, 0.5, 0.8)	(0.5, 1)	10

5. A simulation study of the Bayesian feedback

This section presents a set of simulated scenarios through which we compare the behaviour of the different directions of the proposed Bayesian feedback procedure. In particular, we first present a scheme for performing the biomass/abundance simulation, and then how to mimic the real sampling processes by sampling from the simulated scenarios. A more detailed explanation of the simulation can be found in the Appendix.

5.1. Spatial abundance/biomass simulation

In order to simulate a spatial abundance/biomass simulation with a latent spatial process, the first step is to simulate a continuous covariate (X_i), multiply it by its linear coefficient β_1, and then add the latent spatial effect (u_i) in a study region defined as a 10×10 square. These components are then combined with the global mean effect (β_0) to form the linear predictor. Next, the data are simulated following a Gamma distribution, consistent with biomass or abundance distribution scenarios, where the mean is the exponential of the sum of these effects. In addition, the precision of the Gamma (ϕ) is set sufficiently high to mitigate significant variability from the data distribution and prevent from overshadowing the structure of the linear predictor. The mathematical structure of the model for simulation is then as follows:

$$\begin{aligned} Y_i &\sim \text{Gamma}(y_i \mid \mu_i, \phi), \\ \log(\mu_i) &= \beta_0 + \beta_1 \cdot X_i + u_i. \end{aligned} \tag{8}$$

In Table 1, we provide the values of the parameters and hyperparameters used to define the different scenarios. These scenarios were generated by considering all possible combinations of these values, together with those related to the sampling schemes. In addition, for each resulting scenario, we conducted 10 simulation replicates.

5.2. Sampling from the spatial abundance/biomass field

Once we have the simulated scenarios, we need to mimic the usual sampling procedures (independent and preferential):

1. *Independent sampling*. Within the region of study, we conducted a uniform random generation for both the X and Y dimensions.
2. *Preferential sampling*. The simulation of preferential samples was performed by defining the intensity of an LGCP along the study region. To replicate the structure of the preferential model outlined in Eq. (2), we scaled the geostatistical linear predictor and introduced a spatial effect (u^*) specific to sample generation: $\lambda = \exp\left[\gamma \cdot (\beta_0 + \beta_1 \cdot x + u) + u^*\right]$, where we set γ to 0.5 across the different scenarios. However, since we needed to control the number of generated samples —a random quantity in the LGCP — we optimised the total expected number of points Λ through the following objective function, in which we integrate the new element a to control the expected number of points within the study region (Ω):

Table 2
Number of observations for the different balanced and unbalanced scenarios.

Symmetric	50, 100, 250, 1000	
	Small	Huge
Asymmetric	50, 75, 150	1000

Table 3
Different values for the range and stde. of the spatial effect.

Range	0.2	0.5	0.8
Sigma	0.5	1	2

$$\Lambda = \int_{s\in\Omega} \lambda \cdot ds = \int_{s\in\Omega} \exp\left[\gamma \cdot (\beta_0 + \beta_1 \cdot x + u) + u^* + a\right] ds \approx \sum \exp\left[\gamma \cdot (\beta_0 + \beta_1 \cdot x_i + u_i) + u_i^* + a\right] \Delta,$$

where Δ is the area related to a minimum size for each data location generated along the study region Ω. This last step is carried out because of the need to discretise the space for computationally generating the samples from the LGCP.

5.3. Set of scenarios

To assess the efficacy of the feedback processes, it is necessary to determine the various scenarios that will be used. Specifically, two distinct sampling designs have been employed: a balanced design (with equal quantities for independent and preferential sampling) and an unbalanced or asymmetric design (where one sampling method has a large amount of data, while the other has one of the smaller sizes indicated). Table 2 presents the different sample sizes chosen for each design.

In the balanced design, the basic idea is to explore the effects of feedback when an equal number of observations are available in both sampling designs. This approach simulates real-world scenarios where similar amounts of information are available for both sampling processes. In this context, it is of paramount importance to ensure that the samples are balanced. Therefore, these quantities represent the expected values for the LGCP process and not the definitive values. In other words, the number of samples for the preferential process defined as an LGCP is a random quantity whose expectation can be controlled. Once the number of samples has been simulated in accordance with the previously explained procedure, the same number of samples for the independent sampling is then simulated. To sum up, for balanced samples, the final number of scenarios is the combination of the parameters that characterise the response variable along with the size of the samples and the ten replicas for each of these combinations, assuming 240 scenarios.

In the unbalanced design, the primary focus is on comparing scenarios with significantly different sample sizes, one with limited information and the other with a substantial amount of data. These scenarios are intended to assess situations where information availability is highly asymmetric and to evaluate the impact of feedback, similar to real cases where preferential designs have a large amount of data. Accordingly, the procedure entailed combining the parameters that configure the response variable with the different possible values for the small samples for one of the two types of sampling, while fixing the larger sample for the other. The same procedure was followed for the alternative sampling design, with the additional step of drawing ten replicas for each of these combinations. This procedure yielded in 360 distinct scenarios.

The different scenarios result from all possible combinations of the parameters that define each situation. These parameters are presented in Tables 2 and 3. The aim is to evaluate scenarios where the spatial effect has a smoother structure as the range increases and where more extreme values are observed as the standard deviation increases. Additionally, the sample size reflects the amount of information available, with more data being gathered as the study region is more extensively explored.

6. Results

This section presents the results of the four fitting structures across the different scenarios for both geostatistical and sampling processes. In other words, the aforementioned protocols are considered as previously described using the updating by moments approach for all the parameters and hyperparameters, with the exception of the precision of the Gamma distribution, which were updated through the full updating protocol. In order to obtain these results, the PC priors were used in the base models for the spatial hyperparameters. Subsequently, these distributions were updated in the feedback models using the Normal distribution for the re-parameterisation of the spatial hyperparameters.

The following sections will first illustrate the quality of the approximation in the updating by moments protocol and the change in the posterior distribution between the base model and its feedback counterpart. The distributions of the updated parameters and hyperparameters will also be analyses. Subsequently, we demonstrate how three quantiles are distributed in conjunction with the replicas for the parameters and hyperparameters. Finally, the quantitative results of the out-of-sample validation analysis will be presented. This includes the mean global root mean square error (RMSE) and bias for the four models (the two base models and their feedback counterparts).

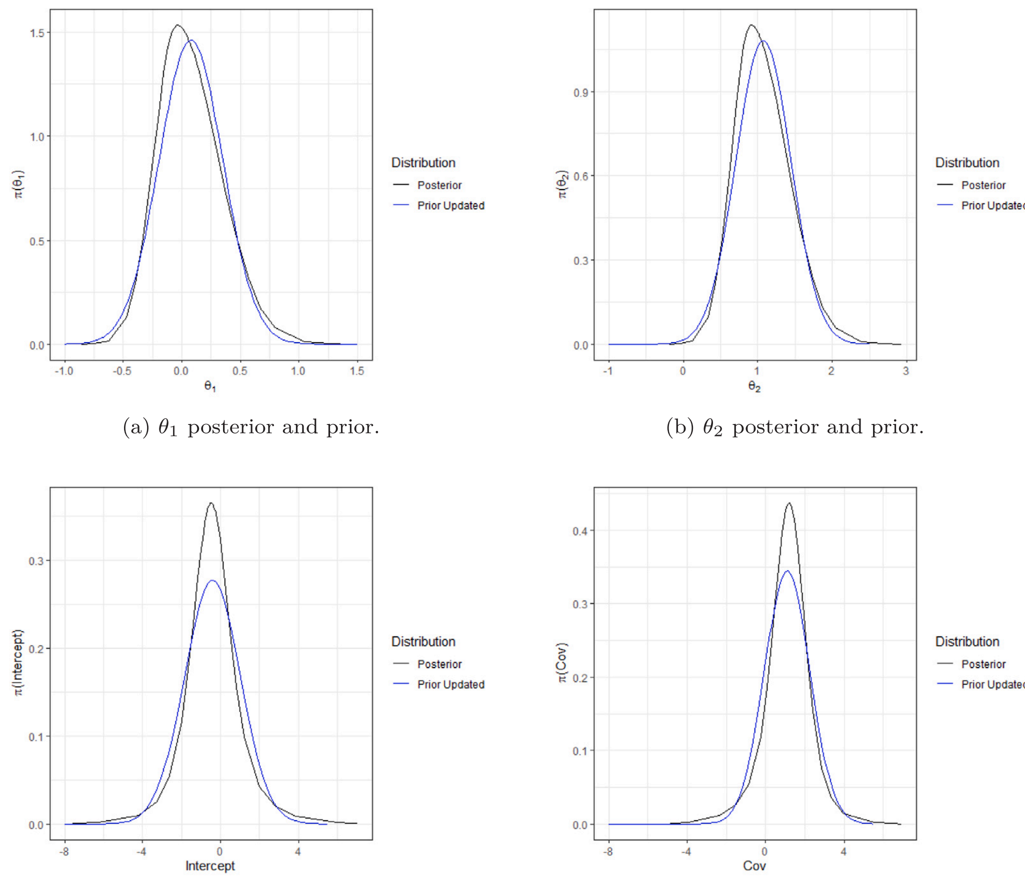
6.1. Updating by moments vs reference posteriors

In this section, we show the difference between the posterior distribution and the prior distribution constructed through moment updating, given that the full updating protocol will logically replicate the complete information of the marginal posterior distribution. Fig. 3 illustrates several plots comparing the posterior distribution and the prior distribution constructed by moment updating, extracted from the simulated scenarios. This figure presents the posterior and prior distributions for the internal parameters of the spatial effect, θ_1 and θ_2, along with the two fixed parameters, intercept β_0 and the linear coefficient associated with the covariate β_1. The prior distributions defined by moment updating replicate relatively well the information from the posterior distribution.

6.2. Comparison of posteriors between base and feedback models

The impact of feedback on posterior distributions varies when assessing either fixed parameters or hyperparameters. In general, the analysis of the posteriors reveals an improvement in parameter identification, resulting in reduced variances and enhanced accuracy compared to the values used in the simulation. This improvement is particularly notable when prior information is provided about the fixed parameters. With regard to hyperparameters, the distinction between models with and without feedback is not readily apparent. However, variations are observed based on the type of prior employed and which of its moments will be updated subsequently. Notably, when the log transformation of the spatial hyperparameters is performed using the Normal distribution, more precise values are obtained. In contrast, the PC-prior exhibits greater variance, resulting in a less accurate alignment with the simulated value.

Fig. 4 illustrates the change on the posterior distributions impact of feedback on two elements of the PM. A more comprehensive comparison of posterior distributions for this instance can be found in the Appendix.



(a) θ_1 posterior and prior.

(b) θ_2 posterior and prior.

(c) Intercept posterior and prior.

(d) Covariate posterior and prior.

Fig. 3. Prior distributions constructed by moment updating (blue) compared with the posterior distributions (black) from which the moments are updated. The prior distributions replicate the posterior distribution relatively well, in particular for the spatial hyperparameters in the internal parameterisation used by INLA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6.3. Analysis and comparison of posterior distributions

In our analysis of the posterior distributions, we assess the following three quantiles: $q_1 = 0.025$, $q_2 = 0.5$ (the median), and $q_3 = 0.975$. While the value of q_2 provides insight into the central tendency of the distribution, the values of q_1 and q_3 offer an understanding of the variability at the extremes, in conjunction with the replicates of the posterior distributions, which are associated with the fixed parameters and hyperparameters of the model. This will allow us to compare the distribution of the values obtained for these three quantiles across all simulated scenarios and evaluate changes in the central tendency (q_2) and the extremes (q_1 and q_3) of the distributions for the different fixed parameters and hyperparameters.

Figs. 5 and 6 illustrate the distribution of the three quantiles, comparing the base and feedback geostatistical models and the base and feedback preferential models for symmetric data sizes. Figs. 7 and 8 present analogous results for asymmetric data sizes. The distributions of these quantiles for the fixed parameters, when considering the entire set of symmetric simulations, show a consistent pattern for the q_2 quantile, with no discernible difference between feedback and non-feedback models. However, the q_1 and q_3 quantiles tend to align more closely with the median values for the feedback model. This suggests that the feedback protocol enhances accuracy by reducing uncertainty in the estimates. This improvement is particularly evident in the case of unbalanced samples, where a low-data model is updated with the posteriors of a big-data model. Regarding the hyperparameters, the distribution of the quantiles does not consistently outperform or improve the estimation provided by the feedback models, whether in balanced or unbalanced cases.

6.4. Out-of-sample analysis

The assessment of the out-of-sample analysis is conducted through the use of two statistical measures: the Root Mean Square Error (RMSE) and the bias. The RMSE is defined as

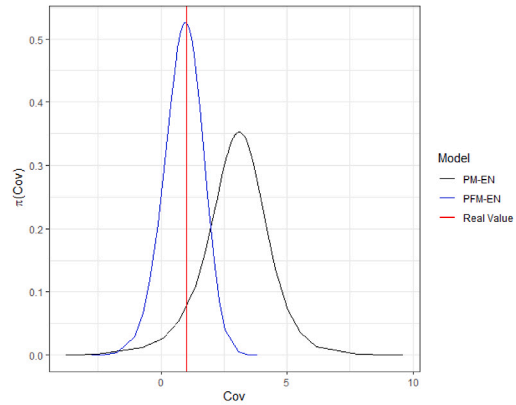
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{9}$$

while the bias is defined for the global predictive results as

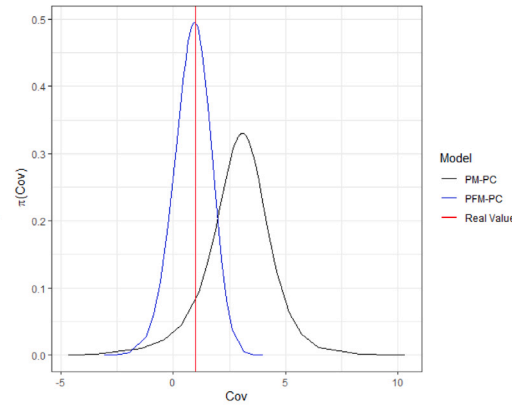
$$bias = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{10}$$

In both formulae, n represents the number of simulated values, y_i denotes the simulated values, and $\hat{y}_i$ refers to the predicted values generated by the model based on a sample set that differs from the one used to calculate the RMSE. Specifically, RMSE and bias are computed using predicted values and simulated values on a grid that covers the study area, avoiding the coincide with the sampling locations.

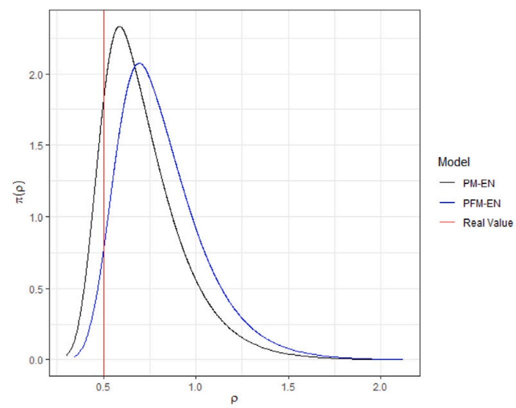
Fig. 9 illustrates the proportion of models that exhibit a lower global RMSE compared to the alternative set of models. From the figure, it can be seen that 72% of the preferential models under feedback outperform the geostatistical base models, and 71% outperform the base preferential models in terms of RMSE. The global results show that the preferential model under feedback tends to perform better than the alternative models. However, based on the results, it would be prudent to perform the feedback process for the geostatistical model as well, given that the geostatistical feedback model outperforms the RMSE of the geostatistical base model by 68%. Furthermore,



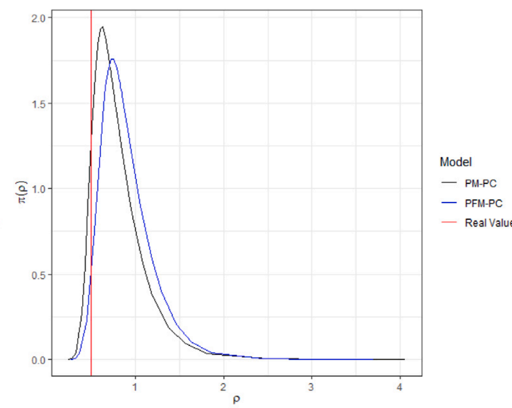
(a) Covariate posteriors for PM with EN priors.



(b) Covariate posteriors for PM with PC-priors.



(c) Spatial range posteriors for PM with EN priors.



(d) Spatial range posteriors for PM with PC-priors.

Fig. 4. Comparison of the posterior distributions in the preferential model (PM) for the parameter β_1 and the spatial range hyperparameter ρ using EN-priors (black) or PC-priors (blue) for the spatial effect, considering the fixed value for the simulations (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 11 and Table 4 present the density of the distribution of proportions relative to the global estimates shown in Fig. 9, comparing the following pairs of models: (A) geostatistical feedback/geostatistical base, (B) preferential base/geostatistical base, (C) preferential feedback/geostatistical base, (D) preferential base/geostatistical feedback, (E) preferential feedback/geostatistical feedback and (F) preferential feedback/preferential base. In particular, Table 4 provides key statistics for the assessment of centrality and deviation. In addition, it includes the values of five quantiles, which serve to illustrate the variability of these proportions and, consequently, the distribution in which a model exhibits greater or lesser RMSE compared to the one being evaluated.

Fig. 10 illustrates the proportion of models that exhibit lower global bias compared to the alternative set of models. The results indicate that the feedback process generally results in lower biases in the models. This effect is particularly evident in the case of the geostatistical feedback model. In particular, we observe that the bias in the predicted values for the geostatistical feedback model is lower in 90% of cases compared to the geostatistical base model, in 95% of cases compared to the preferential base model, and in 80% of cases compared to the preferential feedback model.

7. Real data example

This section presents an example based on empirical data from the field of fishery sciences. The aim is to illustrate the impact of the different feedback procedures presented in this paper on the predictive results. In particular, two sources of information on hake abundance are available. The first source comes from a random sampling from the EVHOE scientific survey, which was conducted from 2003 to 2021. The second source of information is derived from commercial data collected by observers on board fishing vessels during the same time interval. The region where all the locations of the sampling were performed is the southern French coast of the Atlantic Ocean, as shown in Fig. 12.

Table 4
Evaluation of the variability in the proportion of the global RMSE between different models. Each row contains the name of the index associated with the pair of models compared.

	Mean	Q. 0.025	Q. 0.25	Q. 0.5	Q. 0.75	Q. 0.975
A	16.74	0.03	0.21	0.56	1.23	14.56
B	1.23	0.08	0.74	0.99	1.31	4.54
C	1.62	0.01	0.17	0.42	1.14	12.92
D	20.09	0.07	0.71	1.63	4.22	42.01
E	1.55	0.07	0.59	0.84	1.01	9.74
F	1.79	0.01	0.18	0.46	1.13	13.21

The scientific survey was analysed via a hurdle model (Martin et al., 2005; Paradinas et al., 2017) in order to address the issue of locations with zero catches. In other words, two likelihoods are employed in conjunction to model the data. One is a Bernoulli distribution, which is used to model the data with respect to the presence/absence of abundance, and the other is a Gamma distribution, which is utilised for the data with positive non-zero abundance. The model is expressed as follows:

$$
\begin{aligned}
Z_i &\sim \text{Bernoulli}(\pi_i),\\
\text{logit}(\pi_i) &= \beta_0' + \beta_1' \cdot \text{depth} + \gamma_t \cdot f_t(t_i) + \gamma_u \cdot u_i,\\
Y_i \mid Z_i = 1 &\sim \text{Gamma}(\mu_i, \phi),\\
\log(\mu_i) &= \beta_0 + \beta_1 \cdot \text{deph} + f_t(t_i) + u_i,
\end{aligned}
\tag{11}
$$

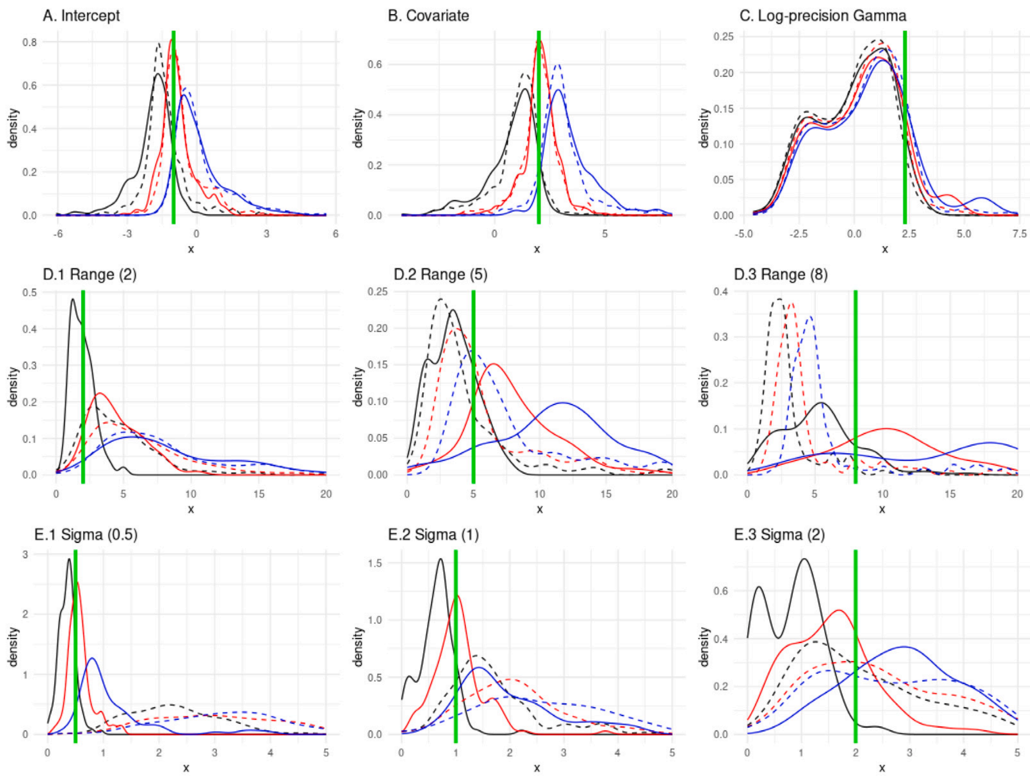


Fig. 5. Distribution of the values of the three quantiles for the base geostatistical model (solid line) and the feedback geostatistical model (dashed lines) when sample sizes are balanced or symmetric. The distribution of the values for the quantile q_1 is depicted in black, that for the second quantile q_2 in red, that for the third quantile q_3 in blue, and the real values for the parameters and hyperparameters are indicated in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

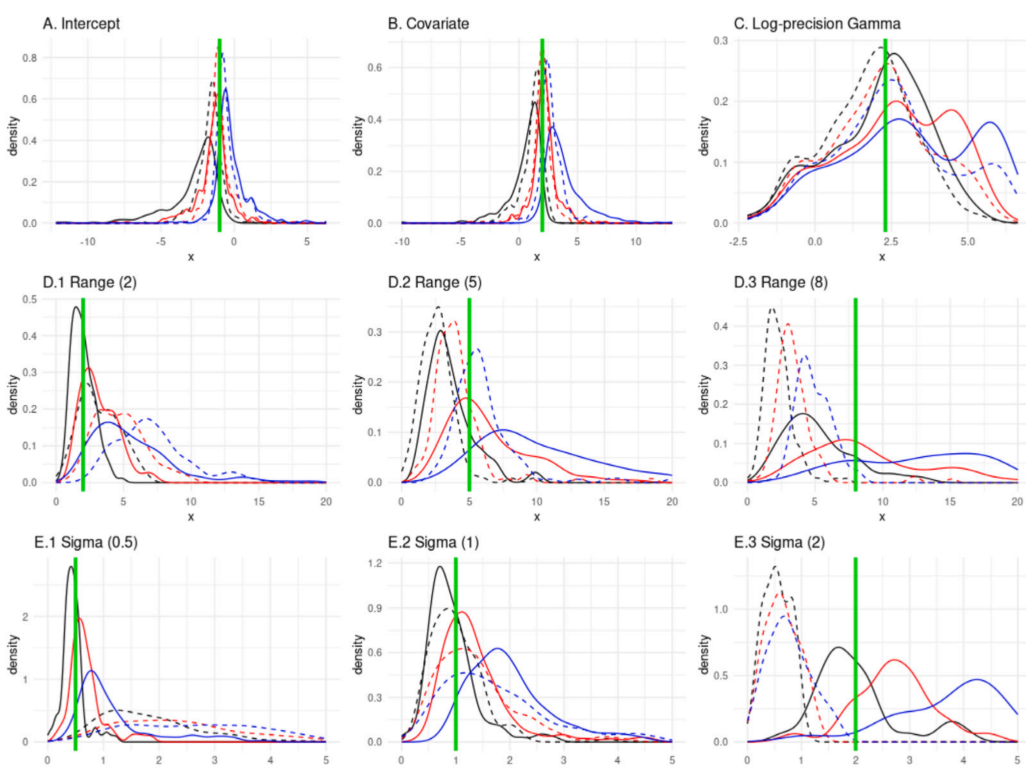


Fig. 6. Distribution of the values of the three quantiles for the base preferential model (solid line) and the feedback preferential model (dashed lines) when we have balanced or symmetric sample sizes. The distribution of the values for the quantile q_1 is depicted in black, that for the second quantile q_2 in red, that for the third quantile q_3 in blue, and the real values for the parameters and hyperparameters are indicated in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

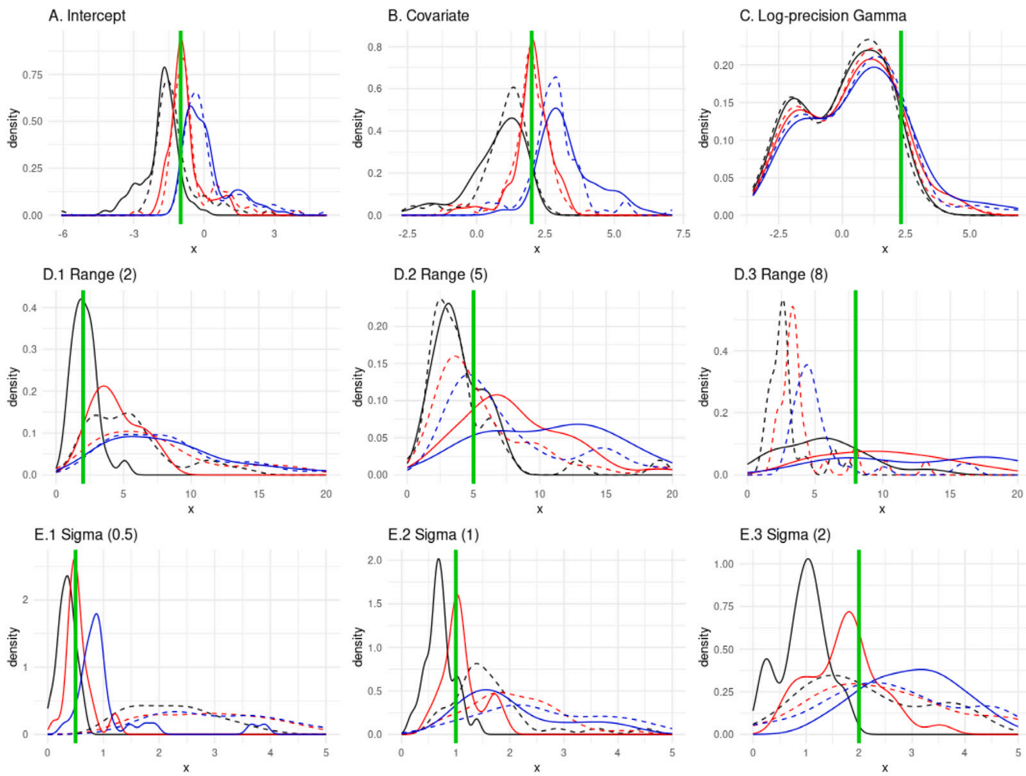


Fig. 7. Distribution of the values of the three quantiles for the base geostatistical model (solid line) and the feedback geostatistical model (dashed lines) when we have asymmetric sample sizes. The distribution of the values for the quantile q_1 is depicted in black, that for the second quantile q_2 in red, that for the third quantile q_3 in blue, and the real values for the parameters and hyperparameters are indicated in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

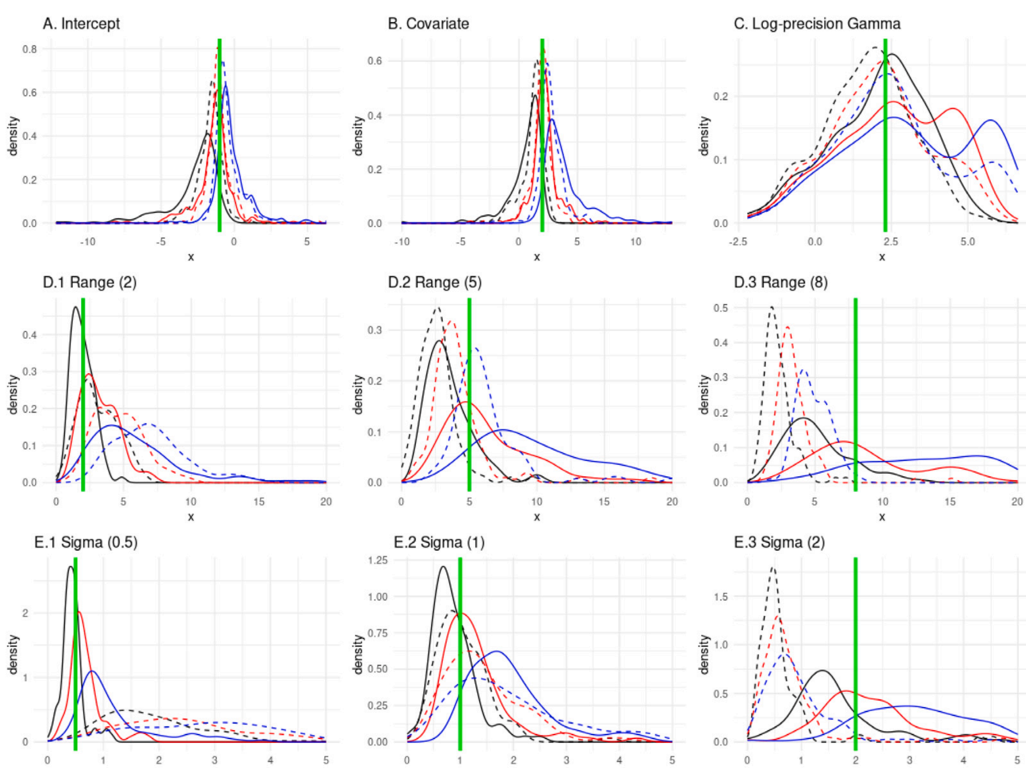


Fig. 8. Distribution of the values of the three quantiles for the base preferential model (solid line) and the feedback preferential model (dashed lines) when we have asymmetric sample sizes. The distribution of the values for the quantile q_1 is depicted in black, that for the second quantile q_2 in red, that for the third quantile q_3 in blue, and the real values for the parameters and hyperparameters are indicated in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 9. The proportion of models with a lower RMSE (in rows) compared to those for the same scenarios (in columns) for the balanced sample size analysis. In the figure, it can be observed that models with feedback show a higher proportion of scenarios with lower RMSE.

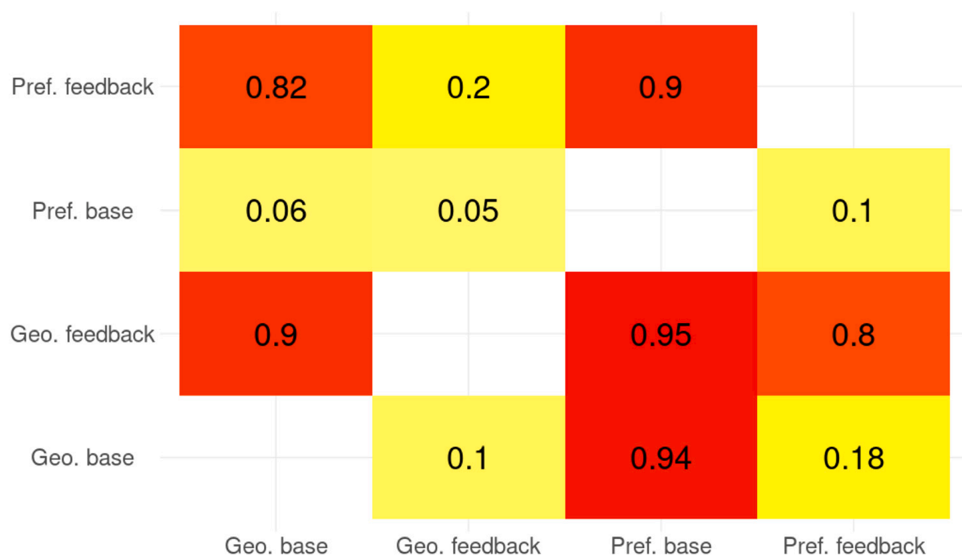


Fig. 10. The proportion of models with a lower bias (in rows) compared to those for the same scenarios (in columns) for the balanced sample size analysis. In this case, models with Bayesian feedback tend to exhibit less bias, particularly the geostatistical feedback model, including when compared proportionally to the preferential model with feedback.

where β_0 and β_1 are coefficients associated with the intercept and the explanatory variable depth. The term $f_t(t_i)$ is associated with each year t_i, such that its prior is that of a first-order random walk (RW1), with τ_t hyperparameter, and is shared between the predictor of the Gamma distribution and the Bernoulli distribution scaled by γ_t. Finally, u_i is the spatial effect term, with ρ_s and σ_s hyperparameters, which is also shared between both linear predictors and scaled by γ_u.

With respect to the commercial survey, we have used a preferential model (Pennino et al., 2019) that incorporates identical components for the two linear predictors, with the distinction that here we use a likelihood for the log-Gaussian Cox process and a Gamma likelihood for the abundance data. The model is expressed as follows:

$$\begin{aligned} Y_i \mid s_i &\sim \text{Gamma}(\mu_i, \phi), \\ \log(\mu_i) &= \beta_0 + \beta_1 \cdot \text{deph} + f_t(t_i) + u_i, \\ s_i &\sim \text{LGCP}(\lambda_i), \\ \log(\lambda_i) &= \beta_0' + \beta_1' \cdot \text{deph} + \gamma_t \cdot f_t(t_i) + \gamma_u \cdot u_i, \end{aligned} \tag{12}$$

where we have the same components with the same meaning as those in the model related to the scientific survey.

Feedback between the models involves setting prior distributions for the fixed parameters (intercept and depth) and the prior distributions for the hyperparameters related to the random effect of years, modelled as a first-order random walk (RW1), and the spatial effect (range and standard deviation). In this way, information is shared between the fixed parameters β_0 and β_1, and the hyperparameters τ_t, ρ_s, and σ_s.

After fitting both models, Fig. 13 shows the predictive results provided by the base and feedback models. This figure illustrates that the spatial predictions made by the base and feedback models are similar. Nevertheless, discernible differences emerge in the spatial patterns, particularly within the preferential model. Fig. 14 shows the results for the distributions of the fixed parameters and the hyperparameters. The results for the models with feedback indicate that the posterior distributions are indistinguishable and have lower uncertainty. With the exception of the posterior distributions for the hyperparameter of the precision of the Gaussian distribution of the data, which has not shared information throughout the models; being identical to the original models without feedback.

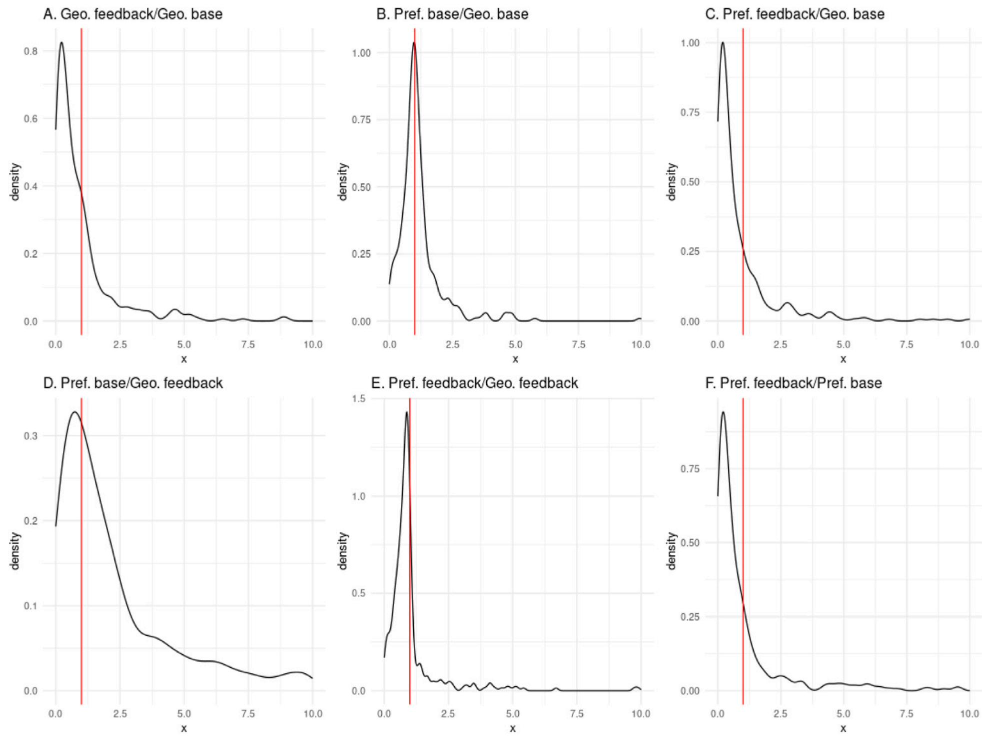


Fig. 11. The distribution of the RMSE proportions for the different pairs of models considered throughout the study. The red vertical line indicates the value of 1. Plots where the density presents higher values to the left of this line show a higher proportion of scenarios with a lower RMSE for the numerator relative to the denominator of the models considered. Conversely, plots where the density is to the right of the line indicate a higher proportion of scenarios with a lower RMSE for the denominator relative to the numerator of the models considered. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

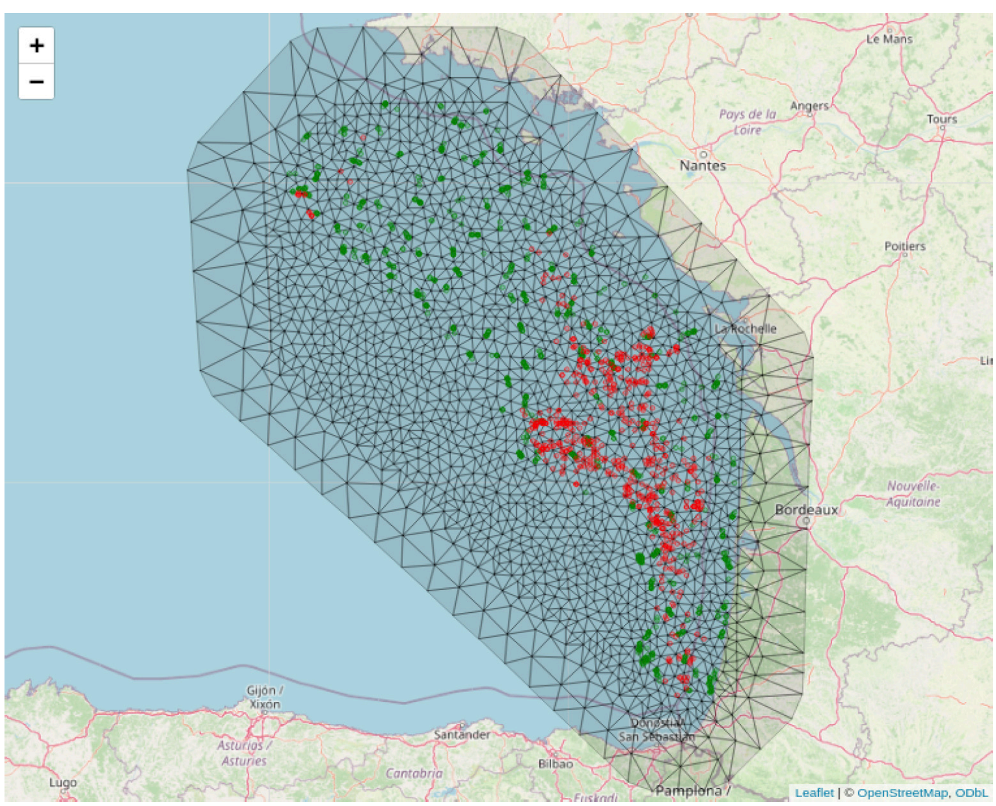


Fig. 12. The western French coast area together with sampling locations from the scientific survey (green) and samples from the commercial surveys (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

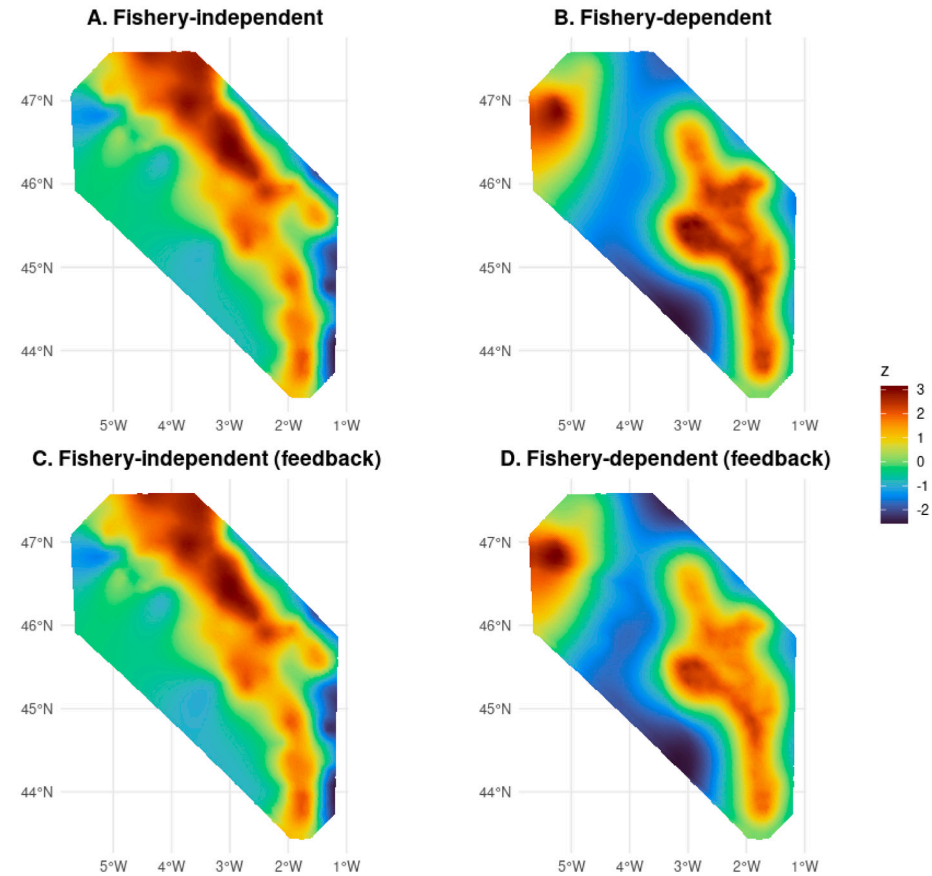


Fig. 13. Predictive maps (temporally aggregated) obtained by the Hurdle model without feedback and with feedback (top-left and bottom-left respectively), and by the preferential model without feedback and with feedback (top-right and bottom-right respectively).

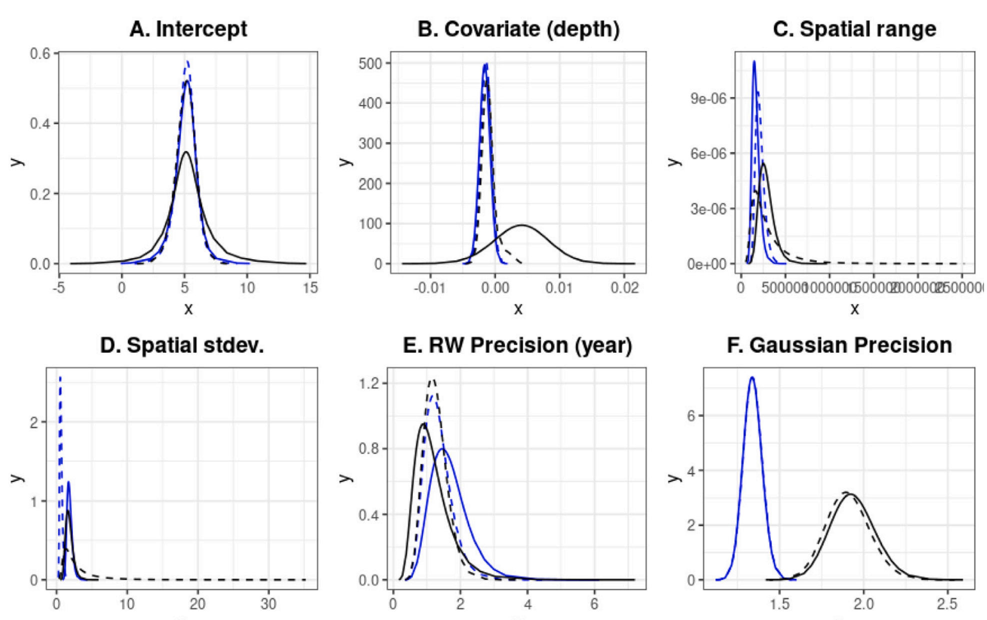


Fig. 14. Posterior distributions for fixed parameters and hyperparameters. In blue the distributions for the models with independent data are plotted, while in black the distributions for the preferred data are plotted. Dotted line represents the posterior distributions associated with models with feedback and solid line represents models without feedback. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

8. Conclusions

SDMs are a widely used tool for analysing spatially distributed data. As previously stated, there are scenarios in which several sources of information are available for the same phenomenon. In this study, we have analysed a particular case in which we managed to share information between geostatistical and preferential models in both directions. However, we have also provided a generalisation of the Bayesian feedback protocol, which allows for the feedback of any models that fall within the designed framework. It is important to note that if all available information is updated (i.e., fixed parameters, random effects, and hyperparameters), the resulting outcome should not be influenced by the direction in which the procedure is carried out. However, in the present case, the random effects have not been updated, and the procedure has been carried out exclusively with the fixed effects and hyperparameters.

In this paper, two methods for performing the feedback between the posterior and prior distributions of two given models are presented: full update and updating by moments. The applicability of these methods is contingent upon the feasibility of implementing either approach. In additional, these two distinct approaches to providing feedback between two models were contrasted: one geostatistical model and a preferential model. The four resulting situations were then evaluated through residual analysis, with commentary provided on the proportion of RMSE and bias. Moreover, all the procedures presented here can be easily implemented and used by non-experts via the BAYSPINS app (Figueira et al., 2024).

The principal findings can be summarised as follows. The feedback process is generally effective in improving the accuracy of the latent field parameters, while this is not always the case for the hyperparameter field. The feedback procedure appears to enhance the prediction of preferential and geostatistical models, with the preferential feedback model showing the best performance in terms of RMSE. Furthermore, the incorporation of feedback increases the robustness of the preferential model, which may encounter computational challenges in specific scenarios (Conn et al., 2017). This study addressed the issue of robustness in preferential modelling by adjusting certain internal parameters of **INLA**, while models employing feedback required no adjustments.

In light of these findings, it can be concluded that providing feedback to the preferential model generally leads to improvements in result accuracy, computational efficiency, and robustness. This last point is particularly important in preferential models (Conn et al., 2017; Diggle et al., 2010). While the improvement in accuracy tends to be more pronounced compared to its alternative, the difference becomes negligible when highly precise results are not achieved. However, the other two improvements consistently occur, reducing the time required for modelling and streamlining computational processes. Therefore, it is recommended to use a simpler model to provide feedback to a more complex one, with the goal of enhancing computational stability, reducing computational costs, and potentially yielding more accurate outcomes. Regarding the real data example, it has been shown that models with feedback produced nearly indistinguishable posterior distributions for elements where information was shared, and these posterior distributions were more accurate (lower uncertainty). Together with the simulated scenarios, this indicates an improvement in the identification of parameters and hyperparameters among models where feedback was shared.

To the best of the authors' knowledge, this is the first study to employ elicitation on a preferential model. The current proliferation of opportunistic data sources available in the field of ecological sciences makes this study particularly pertinent, as it offers a straightforward approach to improving their use in terms of robustness and accuracy. Other studies have demonstrated that elicitation helps reduce uncertainty in species habitat predictions when implementing Bayesian models. However, combining expert information to improve geostatistical models is the most common approach (Pearce et al., 2001; Di Febbraro et al., 2018). Despite the differences in the modelling approach, our findings align well with these earlier studies. The combination of multiple data sources helps to increase the robustness and reliability of the results. This is of particular importance in contexts such as fisheries, as there is often a need for alternative methods to increase confidence in fisheries-dependent data in order to achieve the efficient management of marine resources (Vanhatalo et al., 2014). In conclusion, the results of this study support the wider use of these methods and could be adapted to larger-scale applications, such as environmental management. In fact, the majority of climate change projections are currently conducted on a global scale, making it a challenge to translate them to a regional scale. Elicitation could assist in achieving the priority objectives of the Ocean Decade.

CRediT authorship contribution statement

Mario Figueira: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Xavier Barber:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **David Conesa:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Antonio López-Quílez:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Joaquín Martínez-Minaya:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Iosu Paradinas:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Maria Grazia Pennino:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is part of the ThinkInAzul program and is financed by the MCIN with funds from the European Union (NextGenerationEU-PRTR-C1711) and by the Generalitat Valenciana, Spain GVA-THINKINAZUL/2021/021. DC acknowledges Grant CIAICO/2022/165 funded by Generalitat Valenciana, Spain. XB, DC and ALQ acknowledge the support of Grant PID2022-136455NB-I00, funded by Ministerio de Ciencia, Innovación *y* Universidades of Spain (MCIN/AEI/10.13039/501100011033/FEDER, UE) and the European Regional Development Fund. MGP also thanks the project FRESCO (PID2022-140290OB-I00), funded by Ministerio de Ciencia, Innovación *y* Universidades of Spain.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102858.

Code and data availability

The link to the code and the data used can be found in the following GitHub repository: https://github.com/MarioFigueiraP/Feedback_code_article.

References

Ahmad Suhaimi, S.S., Blair, G.S., Jarvis, S.G., 2021. Integrated species distribution models: A comparison of approaches under different data quality scenarios. Diversity and Distributions 27 (6), 1066–1075. http://dx.doi.org/10.1111/ddi.13255.

Alglave, B., Rivot, E., Etienne, M.P., Woillez, M., Thorson, J.T., Vermard, Y., 2022. Combining scientific survey and commercial catch data to map fish distribution. ICES J. Mar. Sci. 79 (4), 1133–1149. http://dx.doi.org/10.1093/icesjms/fsac032.

Barber, X., Conesa, D., López-Quílez, A., Martínez-Minaya, J., Paradinas, I., Pennino, M.G., 2021. Incorporating Biotic information in species distribution models: A coregionalized approach. Mathematics 9 (4), http://dx.doi.org/10.3390/math9040417.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Bayesian Models with R - INLA. Wiley, http://dx.doi.org/10.1002/9781118950203.

Blangiardo, M., Cameletti, M., Baio, G., Rue, H., 2013. Spatial and spatio-temporal models with R-INLA. Spatial Spatio-Temporal Epidemiol. 4, 33–49. http://dx.doi.org/10.1016/j.sste.2012.12.001.

Braun, C.D., Arostegui, M.C., Farchadi, N., Alexander, M., Afonso, P., Allyn, A., Bograd, S.J., Brodie, S., Crear, D.P., Culhane, E.F., Curtis, T.H., Hazen, E.L., Kerney, A., Lezama-Ochoa, N., Mills, K.E., Pugh, D., Queiroz, N., Scott, J.D., Skomal, G.B., Sims, D.W., Thorrold, S.R., Welch, H., Young-Morse, R., Lewison, R.L., 2023. Building use-inspired species distribution models: Using multiple data types to examine and improve model performance. Ecol. Appl. 33 (6), e2893. http://dx.doi.org/10.1002/eap.2893.

Burgman, M., Carr, A., Godden, L., Gregory, R., McBride, M., Flander, L., Maguire, L., 2011. Redefining expertise and improving ecological judgment. Conserv. Lett. 4 (2), 81–87. http://dx.doi.org/10.1111/j.1755-263X.2011.00165.x.

Conn, P.B., Thorson, J.T., Johnson, D.S., 2017. Confronting preferential sampling when analysing population distributions: diagnosis and model-based triage. Methods Ecol. Evol. 8 (11), 1535–1546.

Cosandey-Godin, A., Krainski, E.T., Worm, B., Flemming, J.M., 2015. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian arctic. Can. J. Fish. Aquat. Sci. 72 (2), 186–197. http://dx.doi.org/10.1139/cjfas-2014-0159.

Crawford, B.A., Maerz, J.C., Moore, C.T., 2020. Expert-informed habitat suitability analysis for at-risk species assessment and conservation planning. J Fish Wildlife Manag. 11 (1), 130–150. http://dx.doi.org/10.3996/092019-JFWM-075.

Di Febbraro, M., Sallustio, L., Vizzarri, M., De Rosa, D., De Lisio, L., Loy, A., Eichelberger, B., Marchetti, M., 2018. Expert-based and correlative models to map habitat quality: Which gives better support to conservation planning? Glob. Ecol. Conserv. 16, e00513. http://dx.doi.org/10.1016/j.gecco.2018.e00513.

Di Febbraro, M., Sallustio, L., Vizzarri, M., De Rosa, D., De Lisio, L., Loy, A., Eichelberger, B., Marchetti, M., 2018. Expert-based and correlative models to map habitat quality: Which gives better support to conservation planning? Glob. Ecol. Conserv. 16, e00513.

Dias, L., Quigley, J., Morton, A., 2018. Elicitation. The Science and Art of Structuring Judgement, first ed. Springer.

Diggle, P.J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, third edition http://dx.doi.org/10.1201/b15326.

Diggle, P.J., Menezes, R., Su, T.l., 2010. Geostatistical inference under preferential sampling. J. R. Stat. Soc. Ser. C. Appl. Stat. 59 (2), 191–232. http://dx.doi.org/10.1111/j.1467-9876.2009.00701.x.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. J. R. Stat. Soc. Ser. C. Appl. Stat. 47, http://dx.doi.org/10.1111/1467-9876.00113.

Elith, J., Leathwick, J.R., 2009. Species distribution models: Ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159.

Figueira, M., Conesa, D., López-Quílez, A., 2024. A shiny R app for spatial analysis of species distribution models. Ecol. Inform. 80, 102542. http://dx.doi.org/10.1016/j.ecoinf.2024.102542, URL https://www.sciencedirect.com/science/article/pii/S1574954124000840.

Fletcher, Jr., R.J., Hefley, T.J., Robertson, E.P., Zuckerberg, B., McCleery, R.A., Dorazio, R.M., 2019. A practical guide for combining data to model species distributions. Ecology 100 (6), e02710. http://dx.doi.org/10.1002/ecy.2710.

Fournier, A., Penone, C., Pennino, M.G., Courchamp, F., 2019. Predicting future invaders and future invasions. Proc. Natl. Acad. Sci. 116 (16), 7905–7910. http://dx.doi.org/10.1073/pnas.1803456116.

Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H., 2019. Constructing priors that penalize the complexity of Gaussian random fields. J. Amer. Statist. Assoc. 114 (525), 445–452. http://dx.doi.org/10.1080/01621459.2017.1415907.

Gómez-Rubio, V., 2020. Bayesian Inference with INLA. Chapman & Hall/CRC Press, http://dx.doi.org/10.1201/9781315175584.

Gómez-Rubio, V., Bivand, R.S., Rue, H., 2021. Estimating spatial econometrics models with integrated nested Laplace approximation. Mathematics 9, http://dx.doi.org/10.3390/math9172044.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8 (9), 993–1009. http://dx.doi.org/10.1111/j.1461-0248.2005.00792.x.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135 (2), 147–186. http://dx.doi.org/10.1016/S0304-3800(00)00354-9.

Jung, M., 2023. An integrated species distribution modelling framework for heterogeneous biodiversity data. Ecol. Inform. 76, 102127. http://dx.doi.org/10.1016/j.ecoinf.2023.102127.

Karp, M.A., Brodie, S., Smith, J.A., Richerson, K., Selden, R.L., Liu, O.R., Muhling, B.A., Samhouri, J.F., Barnett, L.A.K., Hazen, E.L., Ovando, D., Fiechter, J., Jacox, M.G., Pozo Buil, M., 2023. Projecting species distributions using fishery-dependent data. Fish Fish. 24 (1), 71–92. http://dx.doi.org/10.1111/faf.12711.

Kaurila, K., Kuningas, S., Lappalainen, A., Vanhatalo, J., 2022. Species distribution modeling with expert elicitation and Bayesian calibration. arXiv:2206.08817.

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R.M., White, M., Stone, L., 2017. Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection. Methods Ecol. Evol. 8 (4), 420–430. http://dx.doi.org/10.1111/2041-210X.12738.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., Rue, H., 2018. Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. http://dx.doi.org/10.1201/9780429031892.

LaMere, K., Mäntyniemi, S., Vanhatalo, J., Haapasaari, P., 2020. Making the most of mental models: Advancing the methodology for mental model elicitation and documentation with expert stakeholders. Environ. Model. Softw. 124, 104589. http://dx.doi.org/10.1016/j.envsoft.2019.104589.

Landis, D.A., Wratten, S.D., Gurr, G.M., 2000. Habitat management to conserve natural enemies of arthropod pests in agriculture. Annu. Rev. Entomol. 45, 175–201.

Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. J. Stat. Softw. 63 (19), 1–25. http://dx.doi.org/10.18637/jss.v063.i19.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., Possingham, H.P., 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. 8 (11), 1235–1246.

Martínez-Minaya, J., Cameletti, M., Conesa, D., Pennino, M.G., 2018. Species distribution modeling: a statistical review with focus in spatio-temporal issues. Stoch. Environ. Res. Risk Assess. 32 (11), 3227–3244. http://dx.doi.org/10.1007/s00477-018-1548-7.

Martins, T.G., Simpson, D., Lindgren, F., Rue, H., 2013. Bayesian computing with INLA: New features. Comput. Statist. Data Anal. 67, 68–83. http://dx.doi.org/10.1016/j.csda.2013.04.014, URL https://www.sciencedirect.com/science/article/pii/S0167947313001552.

Moraga, P., 2019. Geospatial Health Data. Chapman and Hall/CRC, http://dx.doi.org/10.1201/9780429341823.

Nevalainen, M., Helle, I., Vanhatalo, J., 2018. Estimating the acute impacts of arctic marine oil spills using expert elicitation. Mar. Pollut. Bull. 131, 782–792. http://dx.doi.org/10.1016/j.marpolbul.2018.04.076.

O'Hagan, A., 2006. Uncertain Judgements: Eliciting Experts' Probabilities. Wiley, p. 321.

Paradinas, I., Conesa, D., López-Quílez, A., Bellido, J.M., 2017. Spatio-temporal model structures with shared components for semi-continuous species distribution modelling. Spatial Stat. 22, 434–450. http://dx.doi.org/10.1016/j.spasta.2017.08.001.

Paradinas, I., Conesa, D., Pennino, M.G., Muñoz, F., Fernández, A.M., López-Quílez, A., Bellido-Millán, J.M., 2015. Bayesian spatio-temporal approach to identifying fish nurseries by validating persistence areas. Mar. Ecol. Prog. Ser. http://dx.doi.org/10.3354/meps11281.

Paradinas, I., Giménez, J., Conesa, D., López-Quílez, A., Pennino, M.G., 2022. Evidence for spatiotemporal shift in demersal fishery management priority areas in the western Mediterranean. Can. J. Fish. Aquat. Sci. 79 (10), 1641–1654. http://dx.doi.org/10.1139/cjfas-2021-0327.

Paradinas, I., Illian, J.B., Alonso-Fernändez, A., Pennino, M.G., Smout, S., 2023. Combining fishery data through integrated species distribution models. ICES J. Mar. Sci. 80 (10), 2579–2590. http://dx.doi.org/10.1093/icesjms/fsad069.

Pearce, J.L., Cherry, K., S, F., Whish, G., 2001. Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. J. Appl. Ecol. 38 (2), 412–424.

Pearman-Gillman, S.B., Katz, J.E., Mickey, R.M., Murdoch, J.D., Donovan, T.M., 2020. Predicting wildlife distribution patterns in new England USA with expert elicitation techniques. Glob. Ecol. Conserv. 21, e00853. http://dx.doi.org/10.1016/j.gecco.2019.e00853.

Pennino, M.G., Conesa, D., López-Quílez, A., Muñoz, F., Fernández, A., Bellido, J.M., 2016. Fishery-dependent and -independent data lead to consistent estimations of essential habitats. ICES J. Mar. Sci. 73 (9), 2302–2310. http://dx.doi.org/10.1093/icesjms/fsw062.

Pennino, M.G., Paradinas, I., Illian, J.B., Muñoz, F., Bellido, J.M., López-Quílez, A., Conesa, D., 2019. Accounting for preferential sampling in species distribution models. Ecol. Evol. 9 (1), 653–663. http://dx.doi.org/10.1002/ece3.4789.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392. http://dx.doi.org/10.1111/j.1467-9868.2008.00700.x.

Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., Lindgren, F.K., 2017. Bayesian computing with INLA: A review. Annu. Rev. Stat. Appl. 4 (1), 395–421. http://dx.doi.org/10.1146/annurev-statistics-060116-054045.

Rufener, M.-C., Kristensen, K., Nielsen, J.R., Bastardie, F., 2021. Bridging the gap between commercial fisheries and survey data to model the spatiotemporal dynamics of marine species. Ecol. Appl. 31 (8), e02453.

Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H., 2017. Penalising model component complexity: A principled, practical approach to constructing priors. Statist. Sci. 32, 1–28. http://dx.doi.org/10.1214/16-STS576.

Sørbye, S.H., Illian, J.B., Simpson, D.P., Burslem, D., Rue, H., 2018. Careful prior specification avoids incautious inference for log-Gaussian cox point processes. J. R. Stat. Soc. Ser. C. Appl. Stat. 68 (3), 543–564. http://dx.doi.org/10.1111/rssc.12321.

Staniczenko, P.P., Sivasubramaniam, P., Suttle, K.B., Pearson, R.G., 2017. Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks. Ecol. Lett. 20 (6), 693–707. http://dx.doi.org/10.1111/ele.12770.

Vanhatalo, J., Vetemaa, M., Herrero, A., Aho, T., Tiilikainen, R., 2014. By-catch of grey seals (*Halichoerus grypus*) in baltic Fisheries—A Bayesian analysis of interview survey. PLoS One 9 (11), 1–16. http://dx.doi.org/10.1371/journal.pone.0113836.