Environmental and Ecological Statistics
https://doi.org/10.1007/s10651-025-00653-x

A computationally efficient procedure for combining ecological datasets by means of sequential consensus inference

Mario Figueira[1] · David Conesa[1] · Antonio López-Quílez[1] · Iosu Paradinas[2]

Received: 13 June 2024 / Revised: 18 February 2025 / Accepted: 20 February 2025

Abstract
In ecology and environmental sciences, combining diverse datasets has become an essential tool for managing the increasing complexity and volume of ecological data. However, as data complexity and volume grow, the computational demands of previously proposed models for data integration escalate, creating significant challenges for practical implementation. This study introduces a sequential consensus Bayesian inference procedure designed to offer the flexibility of integrated models while significantly reducing computational costs. The method is based on sequentially updating some model parameters and hyperparameters, and combining information about random effects after the sequential procedure is complete. The implementation of the approach is provided through two different algorithms. The strengths, limitations, and practical use of the method are explained and discussed throughout the methodology and examples. Finally, we demonstrate the method's performance using two different examples with real ecological data, highlighting its strengths and limitations in practical ecological and environmental applications.

Keywords Geostatistics · INLA · Preferential sampling · Sequential inference · SPDE

Handling Editor: Luiz Duczmal.

✉ Mario Figueira
Mario.Figueira@uv.es

[1] Department of Statistics and Operations Research, University of Valencia, Carrer del Dr. Moliner, 50 Burjassot, 46100 Valencia, Spain

[2] AZTI, Txatxarramendi Ugartea z/g, 48395 Sukarrieta, Spain

Published online: 11 March 2025

1 Introduction

The field of ecology is undergoing a transformation fueled by the availability of diverse and abundant datasets. Historically, ecological research relied on limited data streams, often constrained by logistical challenges and disciplinary boundaries. However, recent advancements in technology and the proliferation of interdisciplinary collaborations have ushered in an era of data-driven ecology. While each dataset in isolation provides estimates of the ecological processes under investigation, their integration can yield more refined estimates, often resulting in reduced uncertainty (Fletcher Jr. et al. 2019; Rufener et al. 2021; Alglave et al. 2022; Paradinas et al. 2023). Different methods for combining data vary in their capacity to address sampling biases, establish linkages between divergent response variables across datasets, and effectively manage inherent uncertainties within the data (Fletcher Jr. et al. 2019).

The most straightforward method is *data pooling*, which involves aggregating data without explicitly accounting for their diverse sources and associated sampling biases. This approach assumes uniformity in the nature of the response variable across all datasets. In cases where the data sources differ in type, transformation of one dataset is necessary, potentially leading to information loss (Isaac et al. 2020; Fithian et al. 2015).

Another method is *ensemble modeling*, where multiple diverse models are created to predict a single outcome, either by applying various individual models to the same dataset or by using a single modeling setup across different datasets (Araújo and New 2007; Nisbet et al. 2018). However, the formal integration of parameter estimates may pose challenges unless the datasets have similar resolutions (Fletcher Jr. et al. 2019) and use consistent types of response variable.

A more formal approach for combining data involves modeling various datasets simultaneously. This approach is known as *integrated modeling*, where the model explicitly addresses the differences in their sampling processes. The strength of integrated models lies in their ability to combine information from diverse datasets, enabling the estimation of shared parameters across models through joint-likelihood procedures (Fletcher and Fortin 2019; Alglave et al. 2022; Rufener et al. 2021; Paradinas et al. 2023). Unlike data pooling and ensemble modeling techniques, integrated models offer a formal framework for combining different types of data and sampling procedures. As the complexity and scale of data increase, the computational demands of integrated models escalate proportionally, posing significant challenges in practical implementation.

An alternative and faster approach is to combine information using *sequential inference*, which is based on recursive Bayesian inference (Hooten et al. 2021) and sequential approaches (Doucet et al. 2010; Scott et al. 2016; Scott 2017), as implemented in other scientific fields such as neurology (Körding and Wolpert 2004), biometry (Zigler et al. 2013), machine learning (Nguyen et al. 2022) or quantum physics (Brakhane et al. 2012). In this case, information is sequentially incorporated by means of the posterior distributions of the parameters and

hyperparameters of a model as the new prior distributions of the next model. This procedure is repeated until all datasets have been fitted in a sequence.

Nevertheless, implementing this procedure for models with complex latent structures (such as spatio-temporal models) is not straightforward (Figueira et al. 2024). This limitation can be addressed by using approximate sequential inference procedures implemented within the Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009) framework, by exploiting the methodological underpinnings of the Laplace approximation and the Latent Gaussian Model (LGM) structure. As shown in this paper, the method can substantially reduce the computational burden of integrated models while maintaining a high degree of fidelity to the underlying data dynamics.

In particular, the objective of this study is to propose a computationally efficient framework for combining multiple sources of information, whether derived from different types of data or different sampling structures. The proposed method, called sequential consensus due to its resemblance to other consensus approaches (Doucet et al. 2010; Scott 2017), allows for a sequential update of information combined with consensus procedures, addressing the limitations of not sharing random effects information in the sequential procedure (Figueira et al. 2024). In this regard, consensus methods are typically considered (Doucet et al. 2010; Scott 2017) as procedures that allow data to be divided across multiple machines, with each machine independently carrying out a sampling process to implement Bayesian inference using MCMC. The purpose of this system is to provide the same flexibility as previous methods while reducing computational costs by analyzing different sources of information separately. We compare our method across simulated and real scenarios with the results obtained from a complete and simultaneous modeling, which serve as the gold standard for evaluating the performance of our proposed algorithm. The results show that both methodologies produce very similar or indistinguishable outcomes, suggesting that the sequential consensus approach provides a good approximation that significantly reduce computational costs.

After this Introduction, Section 2 provides a brief overview of spatio-temporal and integrated models, focusing on our selection from the extensive range of complex models, where incorporating additional information can lead to significant computational challenges. The section also includes a brief review of the INLA approach. Section 3 describes in detail our proposal to perform a sequential consensus approach, while in Section 4 we present the results of applying our approach in two real examples, an extra simulated example is included in the Supplementary material. We conclude in Section 5.

2 Inference and prediction in spatio-temporal modeling

Complex spatio-temporal models are the quintessential example of computational burden (Banerjee et al. 2015; Paradinas et al. 2017). In such cases, sequential inference can be particularly advantageous, potentially reducing computation time and making the analysis feasible. The computational challenge of spatio-temporal models is amplified as multiple datasets are considered, each contributing additional layers of information

and nuance. Therefore, the optimization of computational efficiency in spatio-temporal modeling is a great example for fully tapping the potential of sequential modeling in addressing complex real-world phenomena spanning domains such as species distribution models, climate science, public health, and a large etcetera.

At the forefront of this field lies geostatistics. Grounded in the principles of spatial dependence and variability, geostatistics provides a robust framework for characterising and predicting spatial processes through statistical inference. It yields reliable estimates when applied to randomly selected samples. However, when samples are preferentially gathered (as in citizen science data), it becomes crucial to address this inherent dependence in the analysis. Additionally, when temporal dependence appears in the data, the spatial domain must be expanded to include the spatio-temporal domain, enabling the joint evaluation of these effects. Furthermore, incorporating various sources of information that account for their unique characteristics (e.g., sampling designs) can be achieved through integrated modeling. All these analyses can be implemented using the well-known and extensively used `R-INLA` software.

2.1 Geostatistical, preferential and spatio-temporal models

Geostatistical models are designed to analyze spatially continuous data, assuming that the data $y(\mathbf{s})$ are realizations of a continuous spatial process defined over a domain $\mathcal{D}$ in a given set of sample locations $\mathbf{s}$ (Diggle et al. 1998). These models aim to describe the spatial variation that cannot be fully explained by covariates (Paradinas et al. 2023). Therefore, geostatistical models contain elements in their structure related to covariates and the spatial structure of the process. Typically, this spatial structure is built either by defining the variance-covariance matrix or directly specifying the precision matrix (Lindgren et al. 2011). Defining the precision matrix directly allows for sparser and more computationally efficient structures. In such cases, the covariance matrix is only implicitly available. A general characterization that can be provided for the different models we will discuss, given a set of n observations $\mathbf{y}(\mathbf{s}) = \{y(\mathbf{s}_1), ..., y(\mathbf{s}_n)\}$, is the following:

$$\begin{aligned} y(\mathbf{s}_i) \mid \eta_i, \boldsymbol{\theta} &\sim \ell(y_i \mid \eta_i, \boldsymbol{\theta}_\ell), \\ g(\mu_i) &= \eta_i = \mathbf{A}_i \mathbf{x}, \end{aligned} \tag{1}$$

where ℓ is the likelihood, and η_i is the linear predictor related to each observation $y(\mathbf{s}_i)$. $\mathbf{A} = (\mathbf{A}_{\boldsymbol{\beta}}, \mathbf{A}_{\mathbf{u}})$ is a projection matrix that can be decomposed into $\mathbf{A}_{\boldsymbol{\beta}}$, which contains the values of the covariates for the fixed effects $\boldsymbol{\beta}$, and $\mathbf{A}_{\mathbf{u}}$, which corresponds to the design matrices for the random effects $\mathbf{u}$. These random effects can include nonlinear effects related to certain covariates or spatial, temporal, or spatio-temporal effects. The latent field is denoted as $\mathbf{x} = (\boldsymbol{\beta}, \mathbf{u})$, consisting of both fixed and random effects, and follows a Gaussian Markov Random Field (GMRF) prior: $\mathbf{x} \sim \text{GMRF}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}_u))$ with zero mean and precision matrix $\mathbf{Q}(\boldsymbol{\theta}_u)$. Finally, $\boldsymbol{\theta} = (\boldsymbol{\theta}_\ell, \boldsymbol{\theta}_u)$ represents the set of hyperparameters, where $\boldsymbol{\theta}_\ell$ corresponds to those related to the likelihood, and $\boldsymbol{\theta}_u$ refers to the hyperparameters associated with the random effects.

Geostatistical models assume independence between sampling locations and the observed values. However, this assumption may fail under preferential sampling, where sampling locations are biased towards certain features (Diggle et al. 2010). Preferential models address this by jointly modeling the sampling process (e.g., using a log-Gaussian Cox process) and the observed values, allowing shared components between their predictors (Illian et al. 2012).

The spatial models described above can be extended to include temporal terms alongside spatial terms, resulting in spatio-temporal models. These models can be structured either as independent spatio-temporal components or through interaction terms between spatial and temporal structures. Independent spatio-temporal models include separate terms for spatial and temporal trends, while interaction-based models account for dependencies between these dimensions. The interaction can be separable, where the precision matrix for the spatio-temporal effect is expressed as the Kronecker product of pure spatial and pure temporal precision matrices, or non-separable, where the spatial and temporal components are jointly modeled without assuming independence. These extensions allow for more nuanced analyses by capturing both spatial and temporal variations and their interactions.

2.2 Integrated models

Integrated models allow different sources of information to be combined by sharing components in the same way as described above for preferential models. These models can combine information from samples with different structures (Alglave et al. 2022), such as completely random samples, stratified random samples or preferential samples. But they can also be used to combine information from different types of data on variables that share components in the latent structure (Koshkina et al. 2017; Fletcher Jr. et al. 2019; Jung 2023; Paradinas et al. 2023).

The following two situations in ecological and environmental contexts, respectively, could be handled with integrated models: combining biomass and abundance observations of the same species collected through different surveys; and combining information on the presence/absence of a toxin along a river with other measures of its concentration at different locations. In both cases, by combining the two sources of information in a joint model, it becomes possible to analyse both variables simultaneously and to use common elements of the latent field for a more accurate and robust estimation.

In both examples, we have two different likelihoods whose linear predictors can be connected by either scaling a shared component with a parameter α or by sharing the effect without scaling it ($\alpha = 1$):

$$\begin{aligned} y_1(\mathbf{s}_i) \mid \eta_{1i}, \boldsymbol{\theta}_1 &\sim \ell_1(y_{1i} \mid \eta_{1i}, \boldsymbol{\theta}_1), \\ y_2(\mathbf{s}_j) \mid \eta_{2j}, \boldsymbol{\theta}_2 &\sim \ell_2(y_{2j} \mid \eta_{2j}, \boldsymbol{\theta}_2), \\ g_1(\mu_{1i}) = \eta_{1i} &= \beta_{10} + \mathbf{A}_{1i}\boldsymbol{\beta}_1 + \textstyle\sum_{l_1=1}^{L_1} f_{1l_1}(z_{1il_1}) + u(s_i), \\ g_2(\mu_{2j}) = \eta_{2j} &= \beta_{20} + \mathbf{A}_{2j}\boldsymbol{\beta}_2 + \textstyle\sum_{l_2=1}^{L_2} f_{2l_2}(z_{2il_2}) + \alpha \cdot u(s_j), \end{aligned} \tag{2}$$

where ℓ_1 and ℓ_2 are different likelihood functions for $\mathbf{y}_1$ $(y_1(\mathbf{s}_i) : i = 1, .., n_1)$ and $\mathbf{y}_2$ $(y_2(\mathbf{s}_j) : j = 1, .., n_2)$, and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the set of hyperparameters associated with each likelihood. Additionally, the model includes two different intercepts β_{10} and β_{20}, while $\mathbf{A}_1$ and $\mathbf{A}_2$ respectively denote the matrices for the explanatory variables with their corresponding linear coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. The terms $f_{1l_1}(z_{1il_1})$ and $f_{2l_2}(z_{2il_2})$ represent random effects. The number of fixed and random effects can differ between the two linear predictors. Here, $l_1 = \{1, ..., L_1\}$ denotes the different random effects for the first predictor, while $l_2 = \{1, ..., L_2\}$ does the same for the second. Both linear predictors share the spatial component $u(\mathbf{s})$, scaled by an α parameter in the second linear predictor. This illustrates the flexibility of the integrated modeling approach to analyze multiple sources of information together.

2.3 INLA

The methodology proposed in this paper for combining information from different sources has been implemented within the framework of the Integrated Nested Laplace Approximation (Rue et al. 2009) in the `R-INLA` software (Martins et al. 2013). INLA is a deterministic approximation approach deeply rooted in GMRF theory (Rue and Leonhard 2005) for Bayesian inference (Rue et al. 2009; Van Niekerk et al. 2023) and Latent Gaussian Models (LGM). INLA focuses on estimating the marginal posterior distributions of the model parameters, allowing the computation of marginal likelihoods and standard goodness-of-fit metrics such as Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002), Watanabe-Akaike Information Criterion (WAIC) (Watanabe 2013) and Conditional Predictive Ordinates (CPO) (Pettit 1990), all of which are essential for model evaluation and comparison.

3 Methodology

This section presents a computationally efficient approach for fitting computationally intensive models. The proposed method partitions the dataset into subsets and approximates the full model sequentially by updating the information from one subset to the next, ultimately obtaining the posterior distributions of the latent field and hyperparameters for the entire dataset. We leverage the advantages of working with latent Gaussian fields within the INLA framework to perform sequential inference.

3.1 Sequential inference

The underlying idea of a sequential analysis is to divide the dataset into n different subsets $\mathbf{y} = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$, and fit a model for each subset $\mathbf{y}_i$ by sequentially updating the joint prior distribution of the latent field and hyperparameters with the posterior distributions from the previous model. Sequential inference relies on the conditional independence property $\pi(\mathbf{y}_i \mid \mathbf{y}_{-i}, \eta_i, \boldsymbol{\theta}) = \pi(\mathbf{y}_i \mid \eta_i, \boldsymbol{\theta})$ (Scott et al. 2016), given the linear predictor η_i and the hyperparameters $\boldsymbol{\theta}$. This conditional independence property is highly relevant and widely used in Bayesian filtering (Särkkä 2013),

state-space models (Triantafyllopoulos 2021), graphical models (Jordan 2004), and is also a fundamental principle of the INLA approach (Rue et al. 2009).

The joint posterior distribution of the latent field $\mathbf{x} = \{x_1, ..., x_K\}$ can be decomposed by using this property, where the nodes of the latent field are denoted by x_j, $j = 1, ..., K$, and the hyperparameters are presented as $\boldsymbol{\theta}$:

$$\begin{aligned} \pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) &\propto \pi(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})\, \pi(\mathbf{x}, \boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \pi(\mathbf{y}_i \mid \mathbf{x}, \boldsymbol{\theta})\, \pi(\mathbf{x}, \boldsymbol{\theta}) \\ &\propto \prod_{i'=i}^{n} \pi(\mathbf{y}_{i'} \mid \mathbf{x}, \boldsymbol{\theta}) \pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}_1, ..., \mathbf{y}_i), \end{aligned} \tag{3}$$

where each subset $\mathbf{y}_i$ of the data is related to one model and $\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}_1, ..., \mathbf{y}_i)$ stands for the joint posterior distribution of the latent field and hyperparameters until the i-th subset.

INLA focuses on the marginals of the latent field and the hyperparameters, thus it is not possible to obtain the joint distribution of the latent field and hyperparameters $\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})$. However, instead of sharing the joint posterior distribution, it is possible to share information only between the different common fixed parameters and hyperparameters (Figueira et al. 2024).

In particular, using the reasoning for the joint posterior distribution, and in line with Figueira et al. (2024), we propose the following approximation to calculate the marginal posteriors of the fixed effects $\boldsymbol{\beta} = \{\beta_1, ..., \beta_K\}$, given a partition of the dataset $\mathbf{y} = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$:

$$\begin{aligned} \pi(\beta_k \mid \mathbf{y}) &\propto \pi(\mathbf{y} \mid \beta_k)\, \pi(\beta_k) \\ &\approx \prod_{i=1}^{n} \pi(\mathbf{y}_i \mid \beta_k)\, \pi(\beta_k) \\ &= \prod_{i'=i}^{n} \pi(\mathbf{y}_{i'} \mid \beta_k)\, \pi(\mathbf{y}_1, ..., \mathbf{y}_i \mid \beta_k)\, \pi(\beta_k) \\ &\propto \prod_{i'=i}^{n} \pi(\mathbf{y}_{i'} \mid \beta_k)\, \pi(\beta_k \mid \mathbf{y}_1, ..., \mathbf{y}_i), \end{aligned} \tag{4}$$

where the posterior distribution from the step $i - 1$, $\pi(\beta_k \mid \mathbf{y}_1, ..., \mathbf{y}_{i-1})$, is used as the prior for the following step i for the fixed effect β_k. It is important to highlight that this approach assumes that the fixed effects are independent of each other.

With respect to the marginal posterior of the hyperparameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$, we propose a similar approximation:

$$\begin{aligned} \pi(\theta_m \mid \mathbf{y}) &\propto \pi(\mathbf{y} \mid \theta_m)\, \pi(\theta_k) \\ &\propto \prod_{i=2}^{n} \pi(\mathbf{y}_i \mid \theta_m)\, \pi(\theta_m \mid \mathbf{y}_1) \qquad : \quad m = 1, ..., M, \end{aligned} \tag{5}$$

where again the marginal posterior of θ_m is used as the prior for the following step.

It is worth noting that this sequential inference procedure does not provide an update of the random effects of the latent field. To overcome this deficiency, in what follows we present two different consensus procedures for updating the random effects by combining the information of the latent field random effects between the different modeling outputs along the data subsets. The first one is based on marginal weighted averages, while the second one focuses on the distribution of each random effect.

3.2 Marginal weighted averages

Our first proposal for combining the information of the latent field random effects $\mathbf{x}_{-\boldsymbol{\beta}}$ (all those nodes in the latent field $\mathbf{x}$ that are not fixed effects $\boldsymbol{\beta}$) is based on averaging their marginal distributions. Given that each node in a latent Gaussian field is a random variable following a normal distribution $x_j \sim \mathcal{N}(\mu_j, \tau_j)$, $j = 1, ..., K$, with mean μ_j and precision τ_j, it is possible to combine information from a random effect structure $\mathcal{X} = \{x_1, ..., x_K\}$ that is used in several models into which the dataset has been divided.

In particular, our proposal is to approximate the posterior of the marginals for each node $x_j \in \mathcal{X}$ by a weighted averaging (Huang and Gelman 2005; Scott et al. 2016) along the different n models, corresponding to each step of the algorithm:

$$x_j \approx \sum_{i=1}^{n} w_{ji} x_{ji}, \tag{6}$$

where $x_{ji} \sim \mathcal{N}(\mu_{ji}, \tau_{ji})$ represents the marginal random variable for j-th node from the i-th model, related to the i-th subset, with mean μ_{ji} and precision τ_{ji}; and $w_{ji} = \tau_{ji} / \sum_{i=1}^{n} \tau_{ji}$ are the optimal weights for Gaussian random variables (Huang and Gelman 2005; Scott et al. 2016), such that $\sum_{i=1}^{n} w_{ji} = 1$. As a result, each node x_j is Gaussian distributed with mean and precision:

$$\begin{aligned} \mu_j &= \sum_{i=1}^{n} w_{ji} \mu_{ji}, \\ \tau_j &= \left(\sum_{i=1}^{n} w_{ji}^2 / \tau_{ji} \right)^{-1} = \sum_{i=1}^{n} \tau_{ji}. \end{aligned} \tag{7}$$

Performing this approximation along the whole set of nodes $\{x_1, ..., x_k\}$ related to $\mathcal{X}$, we can combine the information for the latent field random effects. Furthermore, the weighted averaging approach for the marginal distributions of the random effects also allows the use of other weights, $\mathbf{w}_e$ such as $\sum_{i=1}^{n} w_i^{(e)} = 1$, that can be proposed by experts. Note that this expert elicitation of weights mimics the weighted likelihood approach (Fletcher Jr. et al. 2019), allowing in both cases to fit weighted joint models by incorporating several data sources of different quality. The expert weights can be directly used in Eq. (6), or combined with the optimal weights proposed $w_j = \tau_j / \sum_{i=1}^{n} \tau_{ji}$. For example, we can redefine the weights as:

$$w_{ji}^{*} = \frac{w_{ji}^{(e)} \cdot w_{ji}}{\sum_{i=1}^{n} w_{ji}^{(e)} \cdot w_{ji}}, \tag{8}$$

where we blend expert suggested weights with the optimal weights for Gaussian random variables. With the introduction of the new weights the precision for the averaged node x_j is determined by the more general expression $\tau_j = \left(\sum_{i=1}^{n} w^{*2}_{ji}/\tau_{ji}\right)^{-1}$.

3.3 Product of multivariate Gaussian densities

In this second approach, our emphasis lies in integrating comprehensive information regarding the structure of each random effect $\mathcal{X}$, rather than solely focusing on the marginal distribution of each node. This can be done using the properties of multivariate Gaussian distributions that allow us to combine the latent field information obtained in the fitting of each subset of the full dataset (Huang and Gelman 2005). In particular, we can approximate the density of the multivariate posterior distribution $\pi(\mathbf{x} \mid \mathbf{y})$ for a specific random effect as:

$$\pi(\mathbf{x} \mid \mathbf{y}) \approx \prod_{i=1}^{n} \pi(\mathbf{x} \mid \mathbf{y}_i), \tag{9}$$

where $\pi(\mathbf{x} \mid \mathbf{y}_i)$ represents the multivariate Gaussian density of the latent field $\mathbf{x}$ with mean $\boldsymbol{\mu}_i$ and precision matrix $\mathbf{Q}_i$. Consequently, the product $\pi(\mathbf{x} \mid \mathbf{y})$ is another multivariate Gaussian density, denoted as $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q})$. The corresponding precision matrix and mean of $\pi(\mathbf{x} \mid \mathbf{y})$ can be easily calculated leveraging Gaussian properties:

$$\begin{aligned} \mathbf{Q} &= \sum_{i=1}^{n} \mathbf{Q}_i, \\ \boldsymbol{\mu} &= \mathbf{Q}^{-1} \sum_{i=1}^{n} \mathbf{Q}_i \boldsymbol{\mu}_i \,. \end{aligned} \tag{10}$$

Although this approximated multivariate posterior distribution contains all the required information from each latent field random effect, the fact that we can have direct access to its marginals allows us to describe it better. Indeed, we can compute the marginal distribution for each node $x_j \sim \mathcal{N}(\mu_j, \Sigma_{jj})$, $j = \{1, \ldots, K\}$, of the multivariate distribution $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q})$ related to the random effect $\mathcal{X}$ as:

$$\begin{aligned} \mu_j &= \boldsymbol{\mu}_j, \\ \tau_j &= (\boldsymbol{\Sigma}_{jj})^{-1}. \end{aligned} \tag{11}$$

With this new proposal, we are able to reconstruct the multivariate posterior distributions and the marginal distributions of different random effects (such as spatial effects $\mathcal{X}_s$, temporal effects $\mathcal{X}_t$, or other nonlinear random effects $\mathcal{X}_f$) by combining the information from different models.

Figure 1 summarizes the complete sequential consensus framework for the partition $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$, where the posterior marginal distributions of the fixed effects and hyperparameters are used as priors for modeling the next element $\mathbf{y}_i$ of the partition. The figure shows that once the sequential procedure is completed, the information related to the posterior distributions of the random effects is integrated, either through marginal weighted averages or through the product of multivariate Gaussian densities. Finally, it shows that after these steps, the results are the final posterior distributions from the sequential consensus Bayesian inference procedure.

3.4 Sharing latent field components

Until now we have shown a general framework to implement the sequential consensus in any context. In what follows, we present its implementation in the context of integrated models. Our proposal to implement sequential consensus with models sharing effects takes into account the following issue with the hyperparameters of the shared random effects. If a random effect, $\mathbf{x} \sim \text{GMRF}(\boldsymbol{\mu}, \mathbf{Q}(\boldsymbol{\theta}))$, is scaled in another model as $\mathbf{x}^* = \alpha \cdot \mathbf{x}$, then the shared random effect is distributed as a $\mathbf{x}^* \sim \text{GMRF}(\alpha \cdot \boldsymbol{\mu}, \alpha^{-2} \cdot \mathbf{Q})$. Since this scaling implies also modifying the precision structure, we must be aware that it is not possible to perform the previously proposed sequential updating for every hyperparameter of that random effect. In fact, if we express the precision matrix of $\mathbf{x}$ as $\mathbf{Q}(\tau, \boldsymbol{\theta}) = \tau \mathbf{R}(\boldsymbol{\theta})$, where τ is the marginal precision, for $\mathbf{x}^*$ we have $\mathbf{Q}^* = \alpha^{-2}\tau \mathbf{R}(\boldsymbol{\theta}) = \tau^* \mathbf{R}(\boldsymbol{\theta})$. A

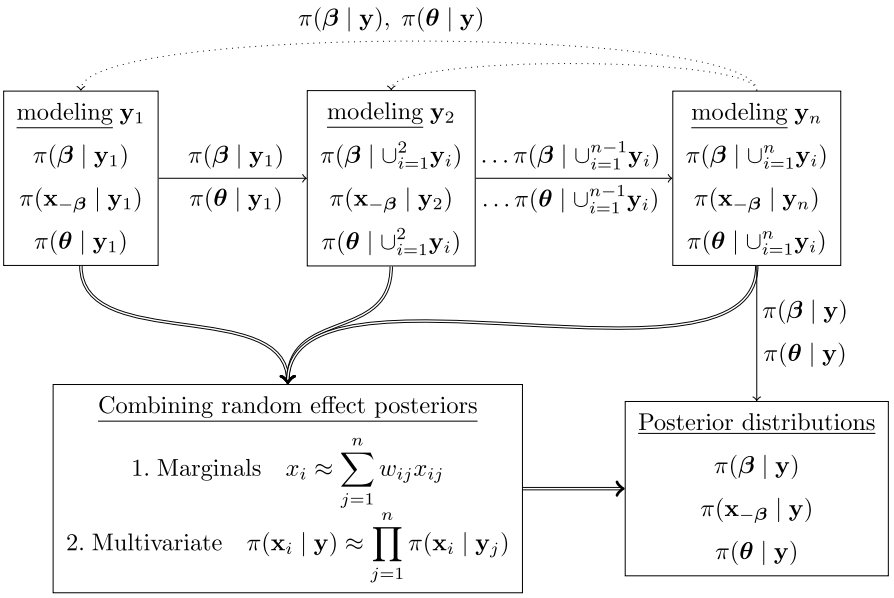

Fig. 1 Scheme of the sequential consensus framework for the data partition $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$: updating in sequence the fixed effects and hyperparameters and performing a consensus for the random effects after the sequential updating

consequence of α and τ not being simultaneously identifiable is that it becomes impossible to perform an update on τ. As a result, we can only update the remaining hyperparameters $\boldsymbol{\theta}$ of the GMRF.

In line with this, we propose two approaches to estimate the sharing parameter α: a Gaussian approximation of its posterior distribution and a point estimate of it. The first consists in combining the approximation of the ratio of the marginal distributions for each node. In particular, each of the quotients $\alpha_j = x_j^*/x_j$, where $x_j^* \sim \mathcal{N}(\mu_j^*, \sigma_j^*)$ and $x_j \sim \mathcal{N}(\mu_j, \sigma_j)$, can be considered as a ratio of two Gaussian random variables. Then, following Hayya et al. (1975), the quotient can be approximated as $\alpha_j \sim \mathcal{N}(\mu_j^{(\alpha)}, \tau_j^{(\alpha)})$ by means of a second-order Taylor expansion, resulting in a Gaussian distribution with mean and variance respectively

$$\begin{aligned} \mu_j^{(\alpha)} &\approx \frac{\mu_j^*}{\mu_j} + \frac{\mu_j^*}{\tau_j \mu_j^3} - \frac{\rho}{\mu_j^2 \sqrt{\tau_j^* \tau_j}}, \\ \tau_j^{(\alpha)} &\approx \frac{\mu_j^{*2}}{\tau_j^2 \mu_j^4} + \frac{1}{\tau_j^* \mu_j^2} - 2\frac{\rho \mu_j^*}{\mu_j^3 \sqrt{\tau_j^* \tau_j}}, \end{aligned} \tag{12}$$

where ρ is the correlation between x_j^* and x_j. From these approximations of the α_i's, our first proposal to approximate the posterior distribution of the shared effect α is $\pi(\alpha) = \prod_i^n \pi(\alpha_i)$. It is worth noting that this approximation tends to be closer to the distribution computed by the integrated model when $\rho = 0$.

Among other tested options, our second approach to estimate the sharing parameter α is to use the empirical median of the following set of values $\{\mu_j^*/\mu_j; j = 1, \ldots, n\}$, where μ_j^* and μ_j are the mean of x_j^* and x_j, respectively.

Both proposals are based on their relatively good performance in simulated and real scenarios, showing particular accuracy when the shared random effect behaves proportionally across the different models it is shared between. However, when the shared random effects have posterior distributions that are not proportional, according to a scaling parameter, across the different models, the sequential consensus method may yield slightly different results than the integrated model. This discrepancy could suggest that these random effects do not share information about the processes in which they are involved, at least not linearly.

Once we have a method for estimating the sharing parameter, we can integrate it in the main procedure of the sequential consensus framework. In particular, in order to combine the information of two shared random effects $\mathbf{x}$ and $\mathbf{x}^*$, we have to scale the second effect $\mathbf{x}^*$ by an estimate $\tilde{\alpha}$ of the scaling parameter, that is, $\mathbf{x}^*/\tilde{\alpha}$. This $\tilde{\alpha}$ can be either the point estimate of the second method or the mean of the Gaussian approximation of the first method in Eq. (12). The final step is then to perform the consensus between the two random effects $\mathbf{x}^*/\tilde{\alpha}$ and $\mathbf{x}$ as presented in the previous sections.

3.5 Sequential consensus algorithms

The sequential process of integrating the information by updating the priors of the fixed effects and hyperparameters, as well as combining the random effects information stored throughout the different sequential inference steps, can be synthesised in the Sequential Consensus algorithm (SC, from now on). The algorithm starts with a split dataset or a set of datasets, and allows to perform inference in sequence for those subsets, updating the marginal prior distributions in the step i of the sequence by using the marginal posterior related to each fixed effect and hyperparameter from the previous step $i-1$. In addition, at each step of the sequence, the information related to the random effects is stored. This can be either the marginal posterior distribution, if the weighted marginal averaging approach is applied, or the multivariate posterior distribution for each random effect, if the multivariate Gaussian density product approach is used.

Algorithm 1 Sequential consensus (SC)

$\mathbf{y} = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$ ▷ dataset partition
for $i = 1$ to n **do**
  if $i = 1$ **then**
    Inference of the first dataset $\mathbf{y}_1$:
    $\pi_{\boldsymbol{\beta}}[i] \leftarrow \pi(\boldsymbol{\beta} \mid \mathbf{y}_i) \propto \pi(\mathbf{y}_i \mid \boldsymbol{\beta})\pi(\boldsymbol{\beta})$ ▷ Store fixed effect posteriors
    $\pi_{\boldsymbol{\theta}}[i] \leftarrow \pi(\boldsymbol{\theta} \mid \mathbf{y}_i)$ ▷ Store hyperparameter posteriors
    $\pi_{\mathbf{x}}[i] \leftarrow \pi(\mathbf{x}_{-\boldsymbol{\beta}} \mid \mathbf{y}_i, \boldsymbol{\theta})$ ▷ Store random effect posteriors
  end
  else
    Inference of the $i-th$ dataset, using $\{\pi_{\boldsymbol{\beta}}[i-1], \pi_{\boldsymbol{\theta}}[i-1]\}$ as prior distributions:
    $\pi_{\boldsymbol{\beta}}[i] \leftarrow \pi(\boldsymbol{\beta} \mid \cup_{j=1}^{i}\mathbf{y}_j) \propto \pi(\boldsymbol{\beta} \mid \mathbf{y}_i)\pi_{\boldsymbol{\beta}}[i-1]$
    $\pi_{\boldsymbol{\theta}}[i] \leftarrow \pi(\boldsymbol{\theta} \mid \cup_{j=1}^{i}\mathbf{y}_j) \propto \pi(\boldsymbol{\theta} \mid \mathbf{y}_i)\pi_{\boldsymbol{\theta}}[i-1]$
    $\pi_{\mathbf{x}}[i] \leftarrow \pi(\mathbf{x}_{-\boldsymbol{\beta}} \mid \mathbf{y}_i, \boldsymbol{\theta})$
  end
end
if Sharing latent field components **then**
  Compute an approximation of the scaling parameters α_i using either the Gaussian approximation or the point estimate.
end
if Marginal weighted averages **then**
  Weighted sum of the random variables x_{ji} from marginal posterior distributions for each j latent field node:
  $x_j \mid \mathbf{y} \approx \sum_{i=1}^{n} w_{ji} x_j[i] \ : \ x_j[i] \sim \pi_{x_j}[i]$

$$\pi(x_j \mid \mathbf{y}) = \mathcal{N}(\mu_j, \tau_j) \ : \ \begin{cases} \mu_j = \sum_{i=1}^{n} w_i \mu_{ji} \\ \tau_j = \left(\sum_{i=1}^{n} w_i^2/\tau_{ji}\right)^{-1} \end{cases}.$$

end
else if Product of multivariate Gaussian densities **then**
  Product of multivariate densities for each random effects structured $\mathcal{X}_i$ stored:
  $\pi(\mathbf{x}_j \mid \mathbf{y}) \approx \prod_{i=1}^{n} \pi_{\mathbf{x}_j}[i]$

$$\pi(\mathbf{x}_j \mid \mathbf{y}) = \text{GMRF}(\boldsymbol{\mu}_j, \mathbf{Q}_j) \ : \ \begin{cases} \mathbf{Q}_j = \sum_{i=1}^{n} \mathbf{Q}_{ji}, \\ \boldsymbol{\mu}_i = \mathbf{Q}_j^{-1} \sum_{j=1}^{n} \mathbf{Q}_{ji}\boldsymbol{\mu}_{ji}. \end{cases}.$$

end

However, the SC algorithm has a shortcoming that can be very important in certain cases. This is due to the fact that the random effects of the latent field estimated in the first step lack the information on fixed effects and hyperparameters that would be available in the last step of the algorithm. Thus, if a partition of the latent field is also done to reduce the computational burden of each step, similar to the partition of the latent field implemented in (Orozco-Acosta et al.

2023; Vicente et al. 2023), we will not be able to perform the consensus procedure between these non-common parts of the latent field. This means that we cannot correct for the lack of information in the estimates of the non-common random effects in the initial steps of the algorithm.

For those situations, we propose to use the following algorithm that avoids that deficiency, the Sequential Consensus for latent field Partitions (SCP, from now on). This second algorithm allows us to leverage all the information obtained in the first algorithm by performing a second pass through the partition, fixing the posterior distributions of the hyperparameters and re-evaluating the posterior distributions of the fixed effects to avoid using duplicated information in each step. With this new algorithm, we are able to obtain better estimations of the latent field effects by leveraging the computations done in SC.

The new algorithm SCP starts fixing the posterior distribution of the hyperparameters resulting from the application of algorithm SC. This can be achieved by leveraging INLA's own methodology, given that fixing the support points that INLA uses for calculating the marginal posteriors of the latent field is equivalent to fixing those posteriors of the hyperparameters.

The second step in this new proposed algorithm involves the re-evaluation of the fixed effects, as their posterior distribution cannot be fixed. The underlying idea of this step is to avoid the use of duplicated information, and consists of computing the posterior distribution of the fixed effects $\pi(\boldsymbol{\beta} \mid \mathbf{y}_{-i})$ for the full dataset excluding the data corresponding at each step i of the algorithm that will be used as the prior distribution at that same step. It can be shown that this distribution $\boldsymbol{\beta} \mid \mathbf{y}_{-i}$ is Gaussian with variance and mean respectively

$$\begin{aligned} \tau^* &= \tau_{i-1} + \tau_n - \tau_i, \\ \mu^* &= (\tau_{i-1} + \tau_n - \tau_i) \times (\tau_{i-1}\mu_{i-1} + \tau_n\mu_n - \tau_i\mu_i). \end{aligned} \tag{13}$$

In order to show this result in Eq. (13), we need to take into account that

$$\pi(\boldsymbol{\beta} \mid \mathbf{y}) \propto \prod_{j=1}^{i} \left[\pi(\mathbf{y}_j \mid \boldsymbol{\beta})\, \pi(\boldsymbol{\beta})\right] \times \prod_{j'=i+1}^{n} \pi(\mathbf{y}_{j'} \mid \boldsymbol{\beta}), \tag{14}$$

where $\prod_{j=1}^{i} \left[\pi(\mathbf{y}_j \mid \boldsymbol{\beta})\, \pi(\boldsymbol{\beta})\right]$ is also proportional to the posterior $\pi(\boldsymbol{\beta} \mid \cup_{j=1}^{i} \mathbf{y}_j)$.

Then, $\pi(\boldsymbol{\beta} \mid \mathbf{y}_{-i})$ can be obtained as:

$$\begin{aligned} \pi(\boldsymbol{\beta} \mid \mathbf{y}_{-i}) &\propto \prod_{j=1}^{i-1} \left[\pi(\mathbf{y}_j \mid \boldsymbol{\beta})\, \pi(\boldsymbol{\beta})\right] \times \prod_{j'=i+1}^{n} \pi(\mathbf{y}_{j'} \mid \boldsymbol{\beta}) \\ &\propto \pi(\boldsymbol{\beta} \mid \cup_{j=1}^{i-1} \mathbf{y}_j) \times \frac{\pi(\boldsymbol{\beta} \mid \mathbf{y})}{\pi(\boldsymbol{\beta} \mid \cup_{j'=1}^{i} \mathbf{y}_{j'})}. \end{aligned} \tag{15}$$

Assuming now that $\boldsymbol{\beta} \mid \mathbf{y} \sim \mathcal{N}(\mu_n, \tau_n)$, $\boldsymbol{\beta} \mid \cup_{j=1}^{i} \mathbf{y}_j \sim \mathcal{N}(\mu_i, \tau_i)$ and $\boldsymbol{\beta} \mid \cup_{j=1}^{i-1} \mathbf{y}_j \sim \mathcal{N}(\mu_{i-1}, \tau_{i-i})$, that is, that these posterior distributions are Gaussian, then the distribution of $\boldsymbol{\beta} \mid \mathbf{y}_{-i}$ is also Gaussian, $\boldsymbol{\beta} \mid \mathbf{y}_{-i} \sim \mathcal{N}(\mu^*, \tau^*)$, with mean and variance in Eq. (13).

After re-evaluating the distributions of the fixed parameters and fixing the joint posterior distribution of the hyperparameters, the final step of the SCP algorithm is to compute the posterior distribution of the latent field for each step. Note that this algorithm allows a better estimation of the random effects of the latent field, particularly when there are non-common random effects among the partition elements, by taking advantage of all the calculations performed in the SC algorithm.

Algorithm 2 Sequential consensus for latent field partitions (SCP)

$\mathbf{y} = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$ ▷ dataset partition
for $i = 1$ to n **do**
  Perform the sequential process as in SC Algorithm.
end
for $i = 1$ to n **do**
  Perform the second-pass sequential process.
  Use the design support points for the hyperparameters $\pi(\boldsymbol{\theta} \mid \mathbf{y})$.
  Re-calculating the posterior distribution for the fixed effects:

$$\pi(\boldsymbol{\beta} \mid \mathbf{y}_{-i}) = \pi(\boldsymbol{\beta} \mid \cup_{j=1}^{i-1} \mathbf{y}_j) \prod_{j'=i+1}^{n} \pi(\mathbf{y}_{j'} \mid \boldsymbol{\beta})$$

  Use $\pi(\boldsymbol{\beta} \mid \mathbf{y}_{-i})$ as the prior distribution for the fixed effects.
  Compute and store the latent random effects $\pi(\mathbf{x}_i \mid \mathbf{y}_i)$.
end
if Sharing latent field components **then**
  Compute an approximation of the scaling parameters α_i using either the Gaussian approximation or the point estimate.
end
if Marginal weighted averages **then**
  Weighted sum of the random variables x_{ji} from marginal posterior distributions for each j latent field node:
  $x_j \mid \mathbf{y} \approx \sum_{i=1}^{n} w_{ji} x_j[i] \ : \ x_j[i] \sim \pi_{x_j}[i]$ where $\pi(x_j \mid \mathbf{y}) = \mathcal{N}(\mu_j, \tau_j)$.
end
else if Product of multivariate Gaussian densities **then**
  Product of multivariate densities for each random effects structured $\mathcal{X}_i$ stored:
  $\pi(\mathbf{x}_j \mid \mathbf{y}) \approx \prod_{i=1}^{n} \pi_{\mathbf{x}_j}[i]$ where $\pi(\mathbf{x}_j \mid \mathbf{y}) = \text{GMRF}(\boldsymbol{\mu}_j, \mathbf{Q}_j)$.
end

An essential aspect of the sequential consensus approach, common to all distributed inference methods, is the establishment of an appropriate data partitioning scheme to facilitate the sequential process. Nevertheless, we offer brief guidelines and criteria for effectively data partitioning and implementing the sequential consensus method, acknowledging that this is a substantial topic deserving of its own detailed analysis.

To implement an appropriate data partitioning scheme, the first step is to perform a descriptive analysis, examining dependencies in the temporal and spatial structure and conducting graphical evaluations of the aggregated data. Spatial dependence can also be assessed through visual analysis or quantitative measures

such as the Moran index. Generally, partitioning spatio-temporal data is more straightforward when based on its temporal structure, ensuring a balanced distribution that facilitates the accurate identification of the model's various parameters and hyperparameters. Alternatively, partitioning can be done based on likelihoods, where the data are grouped and analysed using a sequential consensus process. However, this approach may produce less optimal results compared to partitioning while preserving the likelihoods intact. In this context, descriptive analysis remains a crucial initial step, assessing whether the effects, given the explanatory variables or their spatio-temporal structure, are consistent across different likelihoods, and whether it is reasonable to share certain effects among the linear predictors of those likelihoods.

4 Examples

In this section, we illustrate the application of the sequential consensus inference methodology for analyzing different data sets. We present two examples of real data: one from fisheries science, where the integrated model is complex and involves the integration of different sources of information (biomass, abundance and presence-absence), and another involving a large amount of temperature data. The second example is of particular interest because it demonstrates that the sequential consensus inference procedure can effectively fit large databases using spatio-temporal models, which would otherwise be infeasible. Additionally, we compare the sequential consensus procedure with an ordinary inference process applied to a subset of the data. Moreover, in the material we present an additional simulated example using a spatio-temporal model with two different sampling designs (stratified random sampling and preferential sampling) to show how sequential consensus can be used to combine data from different sampling processes.

4.1 Analysing hake distribution in the bay of Biscay

This case study shows how different sources of information can be combined. In particular, with the ultimate aim of describing the spatial distribution of hake, we combine data from the EVHOE trawl survey and two commercial fishing fleets, sampled by on-board observers, for the years 2003 to 2021 in the Bay of Biscay (Fig. 2).

The scientific survey collected discrete abundance data, while one commercial fishing fleet targeting hake collected continuous biomass data and the other commercial fleet recorded presence-absence data. The commercial fleet targeting hake carried out a preferential exploration of the sea in order to maximize the hake biomass catch. This implies a preferential sampling process, so the integrated model that takes into account all this information, including the one under preferential sampling, can be expressed as follows:

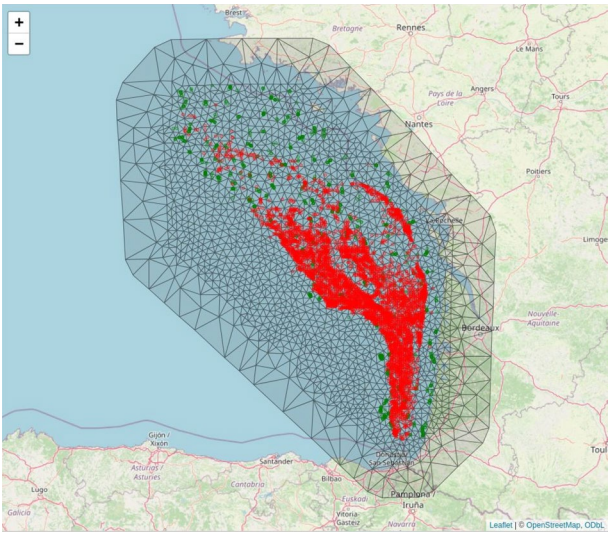

Fig. 2 Hake example. Data from non-preferential random sampling (green dots) and preferential sampling (red dots) obtained in surveys during 2003 and 2021 along the Bay of Biscay

$$
\begin{aligned}
y_{1i} \mid \eta_{1i} &\sim \text{Po}(\lambda_i) \\
\log(\lambda_i) &= \beta_{10} + f_{1d}(z_{di}) + f_{1y}(z_{yi}) + \alpha_{s1} \cdot u_i, \\
y_{2j} \mid \eta_{2j}, \tau &\sim \text{Gamma}(y_{2j} \mid \eta_{2j}, \tau), \\
\log(\mu_j) &= \beta_{20} + f_{2d}(z_{dj}) + f_{2y}(z_{yj}) + u_j, \\
y_{3k} \mid \eta_{3k} &\sim \text{Ber}(\pi_k), \\
\text{logit}(\pi_k) &= \beta_{30} + \alpha_{d3} \cdot f_{2d}(z_{dk}) + f_{3y}(z_{yk}) + \alpha_{s3} \cdot u_k, \\
y(s_j) \mid \lambda(s_j) &\sim \text{LGCP}(\lambda(s_j)), \\
\log(\lambda(s_j)) &= \beta_{40} + \alpha_{d4} \cdot f_{2d}(z_{dj}) + u_j^*,
\end{aligned}
\tag{16}
$$

where the count data from the scientific survey ($\mathbf{y}_1$) is Poisson distributed, the preferentially sampled biomass ($\mathbf{y}_2$) follows a Gamma distribution and the presence/absence ($\mathbf{y}_3$) follows a Bernoulli distribution. The point pattern of the preferential sampling pattern ($\mathbf{s}$) is modelled using a log-Gaussian Cox process (LGCP). The $\boldsymbol{\beta}$ parameters are intercepts associated with each of the response variables, $\boldsymbol{\alpha}$ components represent scaling parameters for the shared effects, f_d represents a structured random effect related to the depth covariate (a second-order random walk for f_{1d} and a one-dimensional SPDE for f_{2d}), and f_y is a first-order random walk for years. Finally, $\mathbf{u}$ is a separable type III spatio-temporal effect (Knorr-Held 2000), with a precision matrix derived from a two-dimensional SPDE with Mátern covariances (Lindgren et al. 2011) for the spatial part and an iid precision matrix for the temporal part.

In implementing the sequential consensus approach, we can consider different ways of splitting the integrated model, e.g. we could consider splitting it into as many parts as there are likelihoods, or we could separate all parts except the Gamma distribution of the preferential data and the LGCP. This latter approach

would facilitate a clearer evaluation of the preferential structure while minimising the discrepancy of information by keeping these likelihoods together. In this case, we decompose the integrated model according to the second proposal and obtain three partitions: a preferential model with biomass data, a Poisson model for counts and a Bernoulli model for presence-absences.

The results for the latent field are shown in the following figures. In Fig. 3 we show the posterior distribution for the fixed effects, where the main discrepancy between the integrated model and the sequential consensus comes from the intercept for the Bernoulli response variable. In Fig. 4 we show the mean and the 95% credible interval (CI) for the effect of depth. Figure 5 shows the mean and 95% CI for the temporal effects. To compute the pure temporal structure, a post-consensus correction is applied to the spatio-temporal component, since the difference in spatial aggregation for each temporal node after consensus induces a change in the pure temporal component. This means that this correction is applied depending on the two types of consensus proposed, marginal or multivariate. Finally, in Figs. 6 and 7 we show the posterior distribution for the spatio-temporal component for the shared spatio-temporal term and the spatio-temporal effect for the LGCP, respectively.

This example demonstrates that even with smaller datasets, the sequential consensus inference process yields results similar to those obtained from the joint model. In terms of computational time, the integrated model takes 62.12 minutes, while the sequential consensus approach performed by SC takes 13.81 minutes, with all computations performed on a server with 63 cores and 157 GB of RAM.

4.2 Spatio-temporal temperature modeling

Our final example illustrates the use of sequential consensus to deal with large databases. In particular, we analyse temperature measurements at 308 geolocated locations collected over 480 months in the coastal area of Alicante, Spain, as shown in Fig. 8. It can be seen that the temperature values have a spatial structure, but this

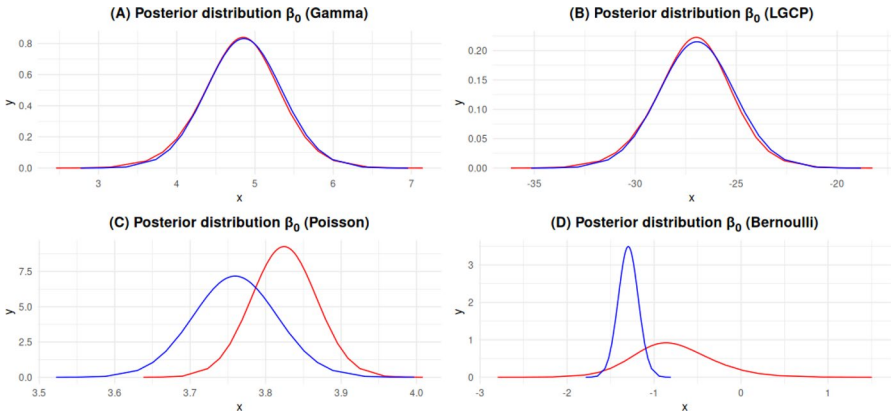

Fig. 3 Hake example. Fixed effects. In red the posterior distributions from the integrated model and in blue the posterior distribution from the sequential consensus approach

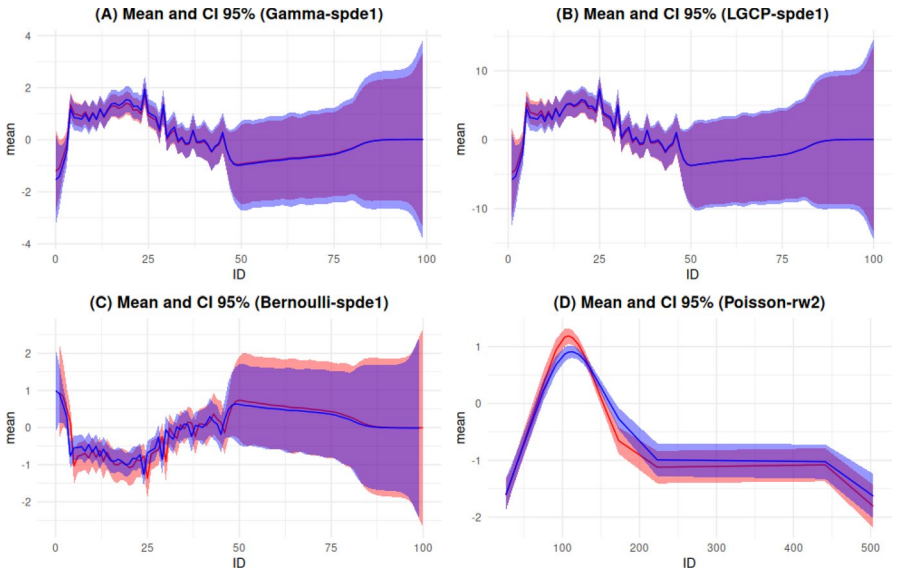

Fig. 4 Hake example. Mean and 95% credible intervals for the posterior distribution of depth. Distributions estimated from the integrated model are shown in red, while those estimated from the sequential consensus are shown in blue

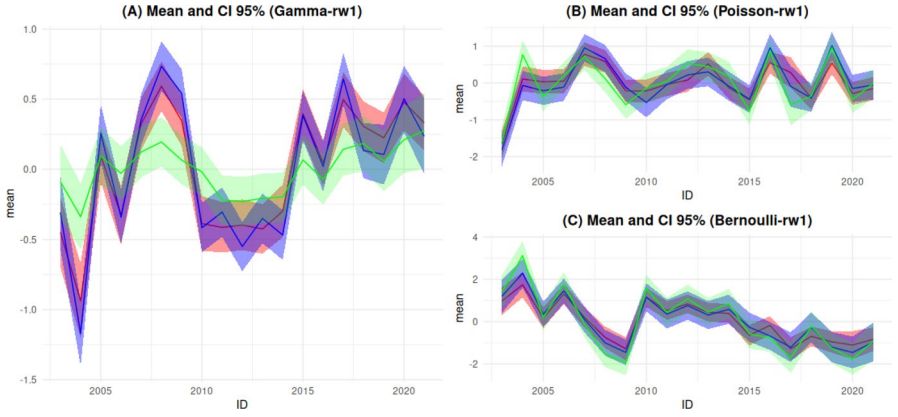

Fig. 5 Hake example. Mean and 95% credible intervals for the posterior distribution for the time effect. Distributions estimated from the integrated model are shown in red, while those estimated from the sequential consensus with the correction from the multivariate consensus are shown in blue and those estimates with the correction from the marginal consensus are shown in green

varies over the months without there appearing to be any real pattern that repeats systematically.

The aim of this example is to show how the sequential consensus approach can be used when the partitioning of the data breaks the latent field to reduce its

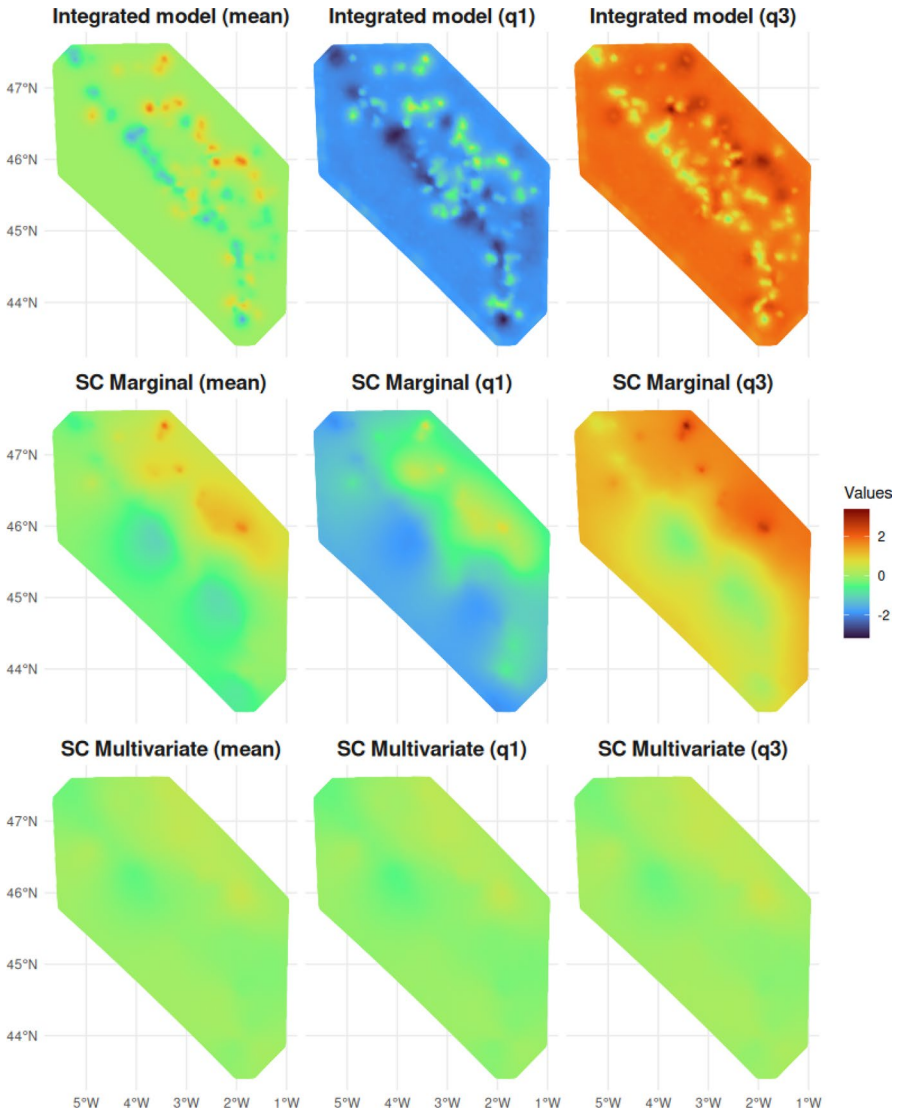

Fig. 6 Hake example. Spatial-temporal effect for the first temporal node of the shared spatio-temporal component between the Gamma, Poisson and Bernoulli models. The plots display the mean values, as well as the quantiles at 0.025 and 0.975, of the posterior distributions, comparing also the marginal and the multivariate consensus

computational burden. In fact, the results of the complete model are compared with the results obtained with the two sequential consensus algorithms (SC and SCP) applied to the partitioned data sets. The comparison is based on the posterior distributions of the latent field nodes, the posterior of the hyperparameters, and the computational cost of the different modeling approaches.

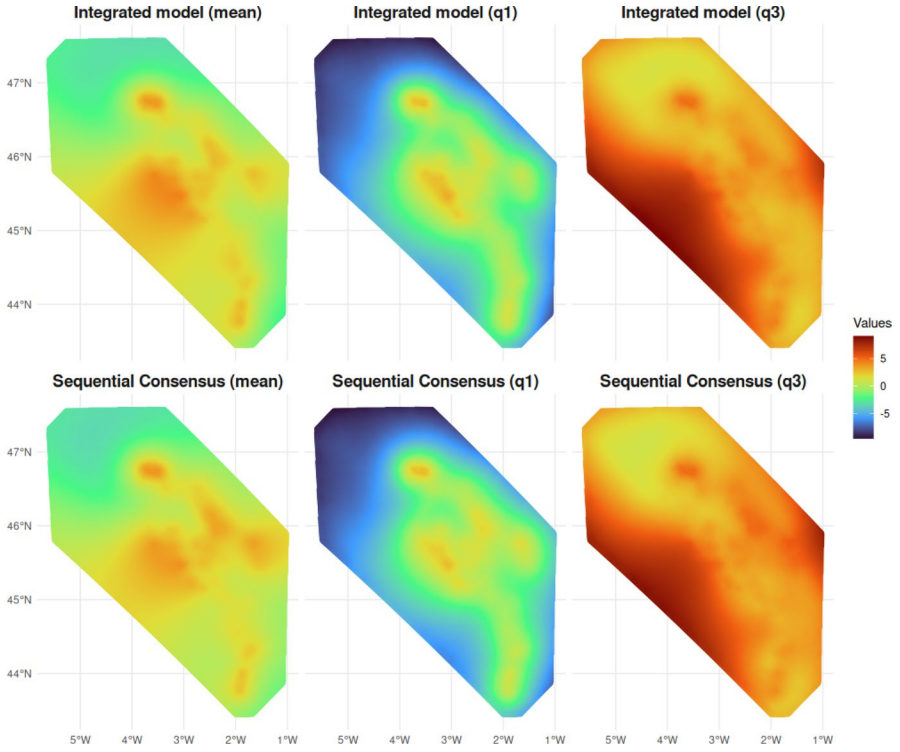

Fig. 7 Hake example. Spatial-temporal effect for the first temporal node of the spatio-temporal component for the LGCP. The plots display the mean values, as well as the quantiles at 0.025 and 0.975, of the posterior distributions

In this case we use a separable spatio-temporal interaction model $\mathbf{Q}_{st} = \mathbf{Q}_s \otimes \mathbf{Q}_t$. The precision matrix of the spatial structure is constructed according to a two-dimensional SPDE effect with Matérn's covariance function (Lindgren et al. 2011), while the precision matrix associated with the temporal structure is defined according to an autoregressive effect of order 1 structure. The spatio-temporal model can be written as:

$$\begin{aligned} y_i \mid \eta_i, \tau &\sim \mathcal{N}(y_i \mid \eta_i, \tau), \\ \mu_i &= \beta_0 + u_i, \\ u_i &\sim \text{GRMF}(\mathbf{0}, \mathbf{Q}_{st}), \end{aligned} \tag{17}$$

where μ_i is the mean of the normal distribution $\mu_i = \eta_i$, τ is the precision of the normal distribution, β_0 is a global intercept and u_i is the spatio-temporal interacting effect with $\mathbf{Q}_{st}(\theta_1, \theta_2, \rho)$ precision matrix, being θ_1 and θ_2 the reparametrization of the spatial range and the standard marginal deviation of the spatial effect, and ρ the autocorrelation hyperparameter.

The model for the complete data set does not run without the process being terminated on a server with 63 cores and 157 GB of RAM. Therefore, we use a

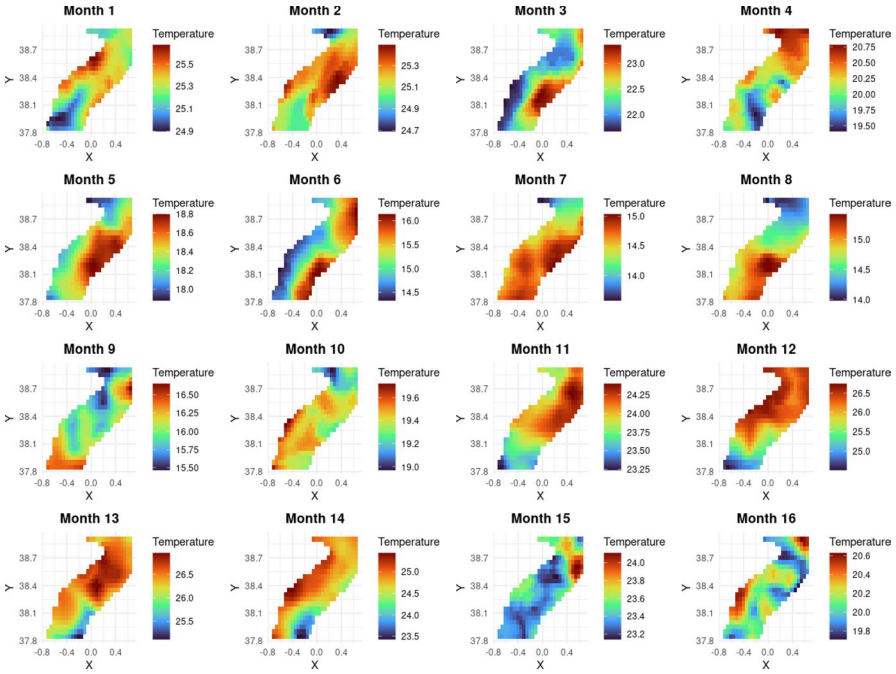

Fig. 8 Temperature big data example. Representation of the first 16 months for temperature values

subset of the first 120 months to allow for model comparison. This subset is used to run a model on the full data and a sequential consensus approach using the two algorithms. To perform the sequential consensus process, the subset of the 120 months is divided into 6 groups, each consisting of the data associated with 20 consecutive time nodes. This partitioning allows each segment to adequately identify the model without imposing excessive computational costs, considering the model's structure and the volume of data in each partition.

In Fig. 9, we can compare the mean and standard deviation of the latent field for the spatio-temporal effect of the first month (as an example) between the complete model and the sequential consensus approach performed according to the SCP Algorithm. It can be observed that both results are very similar in terms of mean and standard deviation, with the standard deviation obtained by the sequential consensus SCP Algorithm being slightly lower. In Fig. 10 we compare an arbitrary node from each group into which the sequential consensus data have been split with the corresponding node from the complete model. It can be seen that the SCP Algorithm produces the same results as the complete model, while the SC Algorithm shows a progressive approximation to the complete model as it progresses through the sequence, obtaining the same results as the SC for the last sequence. Figure 11 presents the posterior distributions of the fixed effect and hyperparameters, showing discrepancies in the latter between the complete model and the sequential consensus approach.

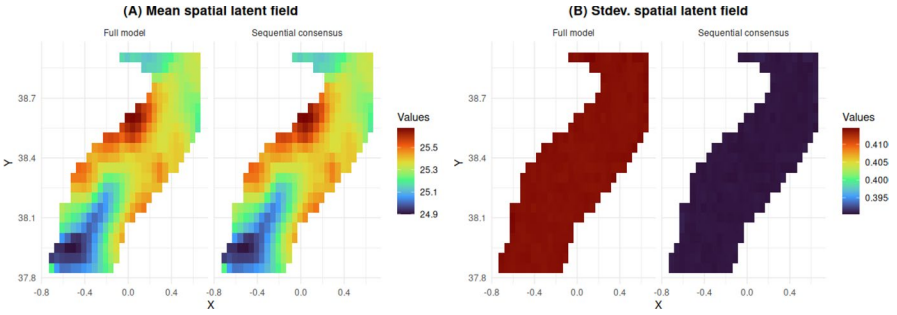

Fig. 9 Temperature big data example. Mean and standard deviation of the posterior distribution for the spatio-temporal effect for the first month (temporal node), using the SCP Algorithm

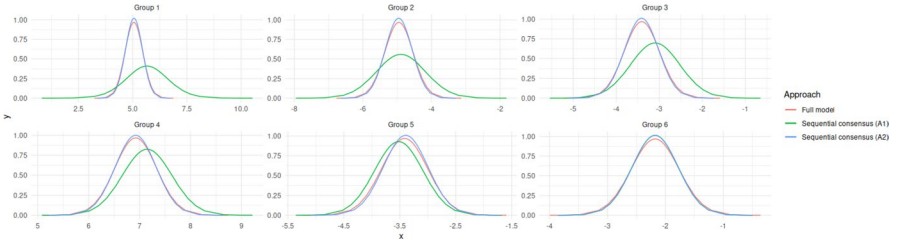

Fig. 10 Temperature big data example. The marginal posterior distribution of an arbitrary node within each group, into which the data and the latent field have been divided for the sequential consensus approach, is compared according to the corresponding posterior distribution for the model with the complete data and latent field

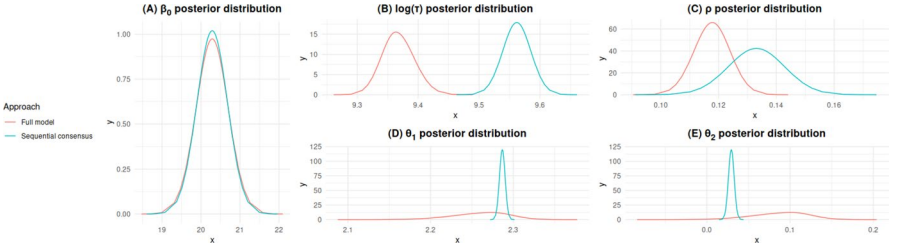

Fig. 11 Temperature big data example. Posterior distribution for the fixed effect and hyperparameters

The computational cost for the complete model is 47.13 minutes, while for the sequential consensus using the SC Algorithm takes 6.57 minutes, and using the SCP Algorithm 11.02 (6.57 + 4.45) minutes. These computations are performed on a server with the previously mentioned specifications: 63 cores and 157 GB of RAM. More importantly, as mentioned above, the model does not run with the full set of 480 time nodes, but with the implementation of the sequential consensus approach, we are able to analyse it in 25.48 minutes and 37.55 minutes using the SC Algorithm and the SCP Algorithm, respectively.

5 Conclusions

Ecological research has been transformed by the availability of diverse datasets, which offer valuable insights into ecological systems (Fletcher Jr. et al. 2019). Integrated models provide a way to combine different types of data, but as data complexity and volume grow, so do the computational demands. To address this challenge, this study proposes a more efficient approximation using sequential inference, leveraging the INLA (Rue et al. 2009; Van Niekerk et al. 2023) method and the Latent Gaussian Model (LGM) framework.

The case studies included in this study focus on complex spatio-temporal modeling scenarios to showcase the potential of sequential modeling in addressing complex real-world phenomena. Spatio-temporal models are quintessential example of computational burden, where the complexity of these models increases as multiple datasets are integrated, each contributing additional layers of information and nuance. The computational improvement is evident in both examples. The first example combined three different spatial datasets and computational time was reduced from 61.02 min to 13.81 min. The second example used a large spatiotemporal dataset that was too large to fit in a dedicated server. However, through the proposed sequential consensus approach, we were able to analyze it in 25.48 min and 37.55 min using the SC Algorithm and the SCP Algorithm, respectively. We also considered a reduced dataset of the second example in which computation time was reduced from 47.13 min to 4.45 min using the SC and an additional 6.57 min for the SCP).

The sequential consensus has been addressed using different approximations for the latent field, treating the nodes of the latent field according to whether they represent fixed or random effects, as well as the hyperparameters. For the fixed effects we define their prior distributions according to the posteriors of the sequence's previous inferential process. For the random effects of the latent field we have applied two consensus process approaches, one based on the marginals for each random effect and another based on the multivariate distribution of the random effects. Both consensus proposals lead to similar results, but the multivariate one takes correlations into account and therefore produces better results when correlations have a high impact. The accuracy and precision of estimates may not be very good in the first steps of the sequence given that not all data has been used at that stage. Therefore, two algorithms have been proposed, a simpler one applicable when our interest is in the final step of the sequence (e.g. last years abundance estimates) and a refined version that performs a second pass over the partitions, improving the estimation of all random effect estimations. This is clearly demonstrated in the results of Fig. 10, where it can be observed that the estimation of the marginal distributions of the random effects was worse in the initial steps of the sequence when the SC Algorithm was applied. In contrast, this deficiency is corrected with the SCP Algorithm, with both algorithms converging to the same result by the final step of the sequential process.

In general, both algorithms gave good estimates of the latent field, including both fixed and random effects. However, the SC Algorithm gives worse estimates

of the random effects in the early steps of the sequence if these are not compensated by estimates from later steps as in the proposed SCP. However, in terms of reproducing the posterior distributions of the hyperparameters, neither algorithm succeeds in reproducing those obtained by the integrated model. This is because the updating of the marginals requires the assumption of no correlation between the hyperparameters, which is generally not true, although it is sufficiently accurate to allow the correct estimation of the latent field. In addition, the approach used for scaling parameters between shared effects can be problematic if the different posterior distributions in the sequence have non-proportional values. Therefore, when using the sequential consensus approach to mimic integrated models with a large number of likelihoods, it is more effective to recompose the problem using multiple joint models rather than a single large joint model or many single likelihood models.

Note that this sequential consensus allows us to perform a sequential inference updating the priors of the inferential step i with the posteriors obtained in the inferential step $i-1$, and combining the information from the random effects by a consensus strategy. However, since we are updating the marginal distributions, it is expected that for the hyperparameters, this may not be the best approximation compared to updating their joint distribution, due to the high correlation they may exhibit in the posterior joint distribution.

Future developments should improve the updating of hyperparameter information throughout the sequence, together with the implementation of a method to estimate these scaling parameters within the sequential modeling itself. This would minimise the impact of model splitting and allow the integrated model to be reconstructed from the simpler models without loss of fidelity.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10651-025-00653-x.

Acknowledgements DC, MF and ALQ thank support by the 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. DC also acknowledges Grant CIAICO/2022/165 funded by Generalitat Valenciana and Grant RED2022-134202-T also funded by Ministerio de Ciencia, Innovación y Universidades of Spain.

Author contributions M.F. developed the methodology, wrote the code and did the analysis, prepared the figures and wrote and reviewed the main document. D.C. wrote and revised the main document. A.LQ. wrote and revised the main document. I.P. wrote and revised the main document.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability The data and code used in this study are publicly available in the following repository: https://github.com/MarioFigueiraP/Secuencial_Consensus_Inference. This repository contains all the necessary materials to reproduce the analyses and results presented in our paper, including detailed instructions for use.

Declarations

Competing interests The authors declare no competing interests.

References

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

Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends Ecol Evol 22(1):42–47. https://doi.org/10.1016/j.tree.2006.09.010

Banerjee S, Carlin BP, Gelfand AE (2015) Hierarchical modeling and analysis for spatial data, 2ed edn. Chapman & Hall/CRC, New York

Brakhane S, Alt W, Kampschulte T, Martinez-Dorantes M, Reimann R, Yoon S, Widera A, Meschede D (2012) Bayesian feedback control of a two-atom spin-state in an atom-cavity system. Phys Rev Lett 109:173601. https://doi.org/10.1103/PhysRevLett.109.173601

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

Diggle PJ, Tawn JA, Moyeed RA (1998) Model-based geostatistics. J R Stat Soc: Ser C Appl Stat 47(3):299–350. https://doi.org/10.1111/1467-9876.00113

Doucet A, Freitas N, Gordon N, Smith A (2010) Sequential Monte Carlo methods in practice, softcover reprint of hardcover. Information science and statistics, 1st edn. Springer, New York

Figueira M, Barber X, Conesa D, López-Quílez A, Martínez-Minaya J, Paradinas I, Pennino MG (2024) Bayesian feedback in the framework of ecological sciences. Ecol Inf 84:102858. https://doi.org/10.1016/j.ecoinf.2024.102858

Fithian W, Elith J, Hastie T, Keith DA (2015) Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol Evol 6(4):424–438. https://doi.org/10.1111/2041-210X.12242

Fletcher RJ Jr, Hefley TJ, Robertson EP, Zuckerberg B, McCleery RA, Dorazio RM (2019) A practical guide for combining data to model species distributions. Ecology 100(6):02710. https://doi.org/10.1002/ecy.2710

Fletcher R, Fortin M-J (2019) Spatial ecology and conservation modeling: applications with R. Springer, Cham

Hayya J, Armstrong D, Gressis N (1975) A note on the ratio of two normally distributed variables. Manag Sci 21(11):1338–1341. https://doi.org/10.1287/mnsc.21.11.1338

Hooten MB, Johnson DS, Brost BM (2021) Making recursive Bayesian inference accessible. Am Stat 75(2):185–194. https://doi.org/10.1080/00031305.2019.1665584

Huang Z, Gelman A (2005) Sampling for Bayesian computation with large datasets. Tech Rep. https://doi.org/10.2139/ssrn.1010107

Illian JB, Sørbye SH, Rue H (2012) A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). Ann Appl Stat 6(4):1499–1530. https://doi.org/10.1214/11-AOAS530

Isaac NJ, Jarzyna MA, Keil P, Dambly LI, Boersch-Supan PH, Browning E, Freeman SN, Golding N, Guillera-Arroita G, Henrys PA et al (2020) Data integration for large-scale models of species distributions. Trends Ecol Evol 35(1):56–67. https://doi.org/10.1016/j.tree.2019.08.006

Jordan MI (2004) Graphical models. Stat Sci. https://doi.org/10.1214/088342304000000026

Jung M (2023) An integrated species distribution modelling framework for heterogeneous biodiversity data. Ecol Inf 76:102127. https://doi.org/10.1016/j.ecoinf.2023.102127

Knorr-Held L (2000) Bayesian modelling of inseparable space-time variation in disease risk. Stat Med 19(17–18):2555–2567

Körding KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427(6971):244–247. https://doi.org/10.1038/nature02169

Koshkina V, Wang Y, Gordon A, Dorazio RM, 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. https://doi.org/10.1111/2041-210X.12738

Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc: Ser B (Stat Methodol) 73(4):423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Martins TG, Simpson D, Lindgren F, Rue H (2013) Bayesian computing with INLA: new features. Comput Stat Data Anal 67:68–83. https://doi.org/10.1016/j.csda.2013.04.014

Nguyen CV, Ho LST, Xu H, Dinh V, Nguyen BT (2022) Bayesian active learning with abstention feedbacks. Neurocomputing 471:242–250. https://doi.org/10.1016/j.neucom.2021.11.027

Nisbet R, Miner GD, Yale K (2018) Handbook of statistical analysis and data mining applications. Academic Press, Boston

Orozco-Acosta E, Adin A, Ugarte MD (2023) Big problems in spatio-temporal disease mapping: methods and software. Comput Methods Programs Biomed 231:107403. https://doi.org/10.1016/j.cmpb.2023.107403

Paradinas I, Illian JB, Alonso-Fernández A, Pennino MG, Smout S (2023) Combining fishery data through integrated species distribution models. ICES J Mar Sci 80(10):2579–2590. https://doi.org/10.1093/icesjms/fsad069

Paradinas I, Illian J, Smout S (2023) Understanding spatial effects in species distribution models. PLoS ONE 18(5):0285463

Paradinas I, Conesa D, López-Quílez A, Bellido JM (2017) Spatio-temporal model structures with shared components for semi-continuous species distribution modelling. Spat Stat 22:434–450. https://doi.org/10.1016/j.spasta.2017.08.001

Pettit LI (1990) The conditional predictive ordinate for the normal distribution. J R Stat Soc: Ser B (Methodol) 52(1):175–184. https://doi.org/10.1111/j.2517-6161.1990.tb01780.x

Rue H, Leonhard H (2005) Gaussian Markov random fields. Chapman and Hall/CRC, New York

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. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rufener M-C, Kristensen K, Nielsen JR, Bastardie F (2021) Bridging the gap between commercial fisheries and survey data to model the spatiotemporal dynamics of marine species. Ecol Appl 31(8):02453. https://doi.org/10.1002/eap.2453

Särkkä S (2013) Bayesian filtering and smoothing. Institute of mathematical statistics textbooks. Cambridge University Press, Cambridge

Scott SL, Blocker AW, Bonassi FV, Chipman HA, George EI, McCulloch RE (2016) Bayes and big data: the consensus Monte Carlo algorithm. Int J Manag Sci Eng Manag 11(2):78–88. https://doi.org/10.1080/17509653.2016.1142191

Scott SL (2017) Comparing consensus Monte Carlo strategies for distributed Bayesian computation. Brazilian J Prob Stat 31(4):668–685. https://doi.org/10.1214/17-BJPS365

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc: Ser B (Stat Methodol) 64(4):583–639. https://doi.org/10.1111/1467-9868.00353

Triantafyllopoulos K (2021) Bayesian inference of state space models: Kalman filtering and beyond. Springer, Cham

Van Niekerk J, Krainski E, Rustand D, Rue H (2023) A new avenue for Bayesian inference with INLA. Comput Stat Data Anal 181:107692. https://doi.org/10.1016/j.csda.2023.107692

Vicente G, Adin A, Goicoa T, Ugarte MD (2023) High-dimensional order-free multivariate spatial disease mapping. Stat Comput 33(5):104. https://doi.org/10.1007/s11222-023-10263-x

Watanabe S (2013) A widely applicable Bayesian information criterion. J Mach Learn Res 14(1):867–897

Zigler CM, Watts K, Yeh RW, Wang Y, Coull BA, Dominici F (2013) Model feedback in Bayesian propensity score estimation. Biometrics 69(1):263–273. https://doi.org/10.1111/j.1541-0420.2012.01830.x