Spatial Statistics 73 (2026) 100975

Contents lists available at ScienceDirect

Spatial Statistics

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

A multi-stage Bayesian approach to fit spatial point process models

Rachael Ren [a,*], Mevin B. Hooten [a], Toryn L.J. Schafer [b], Nicholas M. Calzada [a], Benjamin Hoose [b], Jamie N. Womble [c], Scott Gende [c]

[a] The University of Texas at Austin, United States of America
[b] Texas A&M University, United States of America
[c] National Park Service, United States of America

ARTICLE INFO

Keywords:
Spatial point process
Species distribution modeling
Presence-only data
Recursive Bayes
Parallel processing
MCMC

ABSTRACT

Spatial point process (SPP) models are commonly used to analyze point pattern data in many fields, including presence-only data in ecology. Existing exact Bayesian methods for fitting these models are computationally expensive because they require approximating an intractable integral each time parameters are updated and often involve algorithm supervision (i.e., tuning) in the Bayesian setting. We propose a flexible, efficient, and exact multi-stage recursive Bayesian approach to fitting SPP models that leverages parallel computing resources to obtain realizations from the joint posterior, which can then be used for inference on derived quantities. We outline potential extensions, including a framework for analyzing study designs with compact observation windows and a neural network basis expansion for increased model flexibility. We demonstrate this approach and its extensions using a simulation study and analyze data from aerial imagery surveys to improve our understanding of spatially explicit abundance of harbor seal (*Phoca vitulina*) pups in Johns Hopkins Inlet, a tidewater glacial fjord in Glacier Bay National Park, Alaska.

1. Introduction

Spatial point process (SPP) models are a class of generative stochastic models that give rise to a random number of irregularly spaced events in a spatial domain. They have been used to analyze point patterns in various fields, such as ecology (e.g., Haase, 1995; Law, 2009; Warton and Shepherd, 2010; Renner et al., 2015), seismology (e.g., Ogata, 1998), and social sciences (e.g., Mohler et al., 2011). Many models have been developed for SPP data including, but not limited to, the homogeneous and inhomogeneous Poisson point process, the log-Gaussian Cox process (LGCP; Møller et al., 1998), and the Hawkes process (Hawkes, 1971). This article focuses on a computational method for fitting the inhomogeneous Poisson Process (IPP), a point process model in which point intensity varies across space, and extensions that allow for increased model flexibility.

The stochasticity in SPP data arises from two sources: the random number of point events and their locations, which we refer to as n and $\{\mathbf{s}_i\}_{i=1}^n$, respectively. Although it is common in the species distribution modeling (SDM) literature to fit an IPP model using a likelihood conditioned on n (e.g., Johnson et al., 2013), we focus on a complete IPP likelihood that models both data types jointly, allowing spatial heterogeneity and total abundance (i.e., the total number of points in the study domain) to be modeled in a unified framework. This feature is especially useful when the data are partially observed (i.e., total abundance is unknown and may exceed n) and inference on total abundance and/or prediction in unobserved areas is desired. We extend the complete IPP

* Corresponding author.
E-mail address: rren@utexas.edu (R. Ren).

https://doi.org/10.1016/j.spasta.2026.100975
Received 2 September 2025; Received in revised form 30 January 2026; Accepted 17 March 2026
Available online 25 March 2026
2211-6753/© 2026 Published by Elsevier B.V.

likelihood to data observed in compact observation windows, a special case of partially observed data defined by Baddeley et al. (2015) that may arise naturally from certain study designs, such as aerial imagery surveys.

IPP likelihoods are often intensive to compute because they involve an integral over the spatial domain, which is analytically intractable and must be approximated using methods such as numerical quadrature. This is especially inefficient when using Monte Carlo Markov Chain (MCMC; Gelfand and Smith, 1990) methods because the integral is necessarily approximated each time the parameters are updated. Furthermore, conventional MCMC algorithms for fitting IPP models often require algorithmic tuning due to a lack of conjugate prior distributions. Various fast approximate Bayesian methods for fitting SPP models, such as the Integrated Nested Laplace Approximation (INLA; Rue and Martino, 2009) approach to fitting LGCP models (Illian et al., 2012), circumvent this issue by approximating marginal posterior distributions. However, obtaining exact Bayesian inference for the joint posterior distribution remains computationally challenging. Obtaining realizations from the joint posterior distribution, as opposed to approximating marginal distributions, may be beneficial for estimating derived quantities of interest, such as total abundance.

One alternative to fitting an exact Bayesian model using conventional MCMC is to use recursive Bayes approaches, also referred to as Bayesian filtering or sequential inference, to fit the model in stages (Särkkä, 2013). Although these approaches are widely applicable, recursive Bayes approaches for fitting various ecological models have been developed in recent years due to their natural application to hierarchical models and computational efficiency (e.g., Hooten et al., 2023, 2024; Johnson et al., 2022; Leach et al., 2022; McCaslin et al., 2021). These methods often implement a version of the prior-proposal recursive Bayes (PPRB) method described by Hooten et al. (2021). In a two-stage implementation of PPRB, the data are partitioned into two subsets, $\mathbf{y} := (\mathbf{y}_1, \mathbf{y}_2)'$. We let $\boldsymbol{\theta}$ represent a vector of model parameters and $[\boldsymbol{\theta}]$ represent its prior distribution. We use the bracket notation '$[\cdot]$' to represent probability distributions hereafter (Gelfand and Smith, 1990). In the first stage, we use $\mathbf{y}_1$ to find a transient posterior distribution $[\boldsymbol{\theta}|\mathbf{y}_1] \propto [\mathbf{y}_1|\boldsymbol{\theta}][\boldsymbol{\theta}]$. Next, in the second stage, we use the transient posterior distribution as a prior and proposal distribution in an MCMC algorithm to obtain a sample from the full posterior distribution

$$\begin{aligned}[\boldsymbol{\theta}|\mathbf{y}] &\propto [\mathbf{y}|\boldsymbol{\theta}][\boldsymbol{\theta}] \\ &\propto [\mathbf{y}_2|\boldsymbol{\theta}, \mathbf{y}_1][\boldsymbol{\theta}|\mathbf{y}_1].\end{aligned}$$

Critically, using the transient posterior distribution as both the prior and proposal distribution results in cancellations in the Metropolis–Hastings (M-H) ratio for the second stage. Furthermore, the multi-stage implementation allows computationally intensive calculations involved in evaluating the conditional likelihood $[\mathbf{y}_2|\boldsymbol{\theta}, \mathbf{y}_1]$ to be computed in parallel in an intermediate stage.

Conventionally, the partitions in PPRB comprise data of the same type, but alternative perspectives on partitioning can be helpful in some cases. For instance, PPRB approaches for fitting capture-recapture models partition the data such that the first partition comprises the observed binary capture histories and the second partition comprises the number of observed individuals (Hooten et al., 2023, 2024). Similarly, we show that it is natural to partition spatial point pattern data such that the first partition comprises the observed locations $\{\mathbf{s}_i\}_{i=1}^n$ and the second partition comprises the number of observed point events n. We can then fit the data model conditioned on n in the first stage and assimilate n in the second stage. This partitioning scheme allows us to take advantage of well-known strategies that use the conditional data model $[\{\mathbf{s}_i\}|\boldsymbol{\theta}, n]$, such as the logistic and Poisson regression approximations (Aarts et al., 2012; Baddeley et al., 2010; Warton and Shepherd, 2010). We propose a multi-stage MCMC method for fitting the IPP model using PPRB, along with a variety of strategies for first-stage sampling from the transient posterior distribution, resulting in an efficient and flexible approach to fitting an exact Bayesian SPP model.

In Section 2, we provide an overview for fitting an IPP model and potential extensions for additional model flexibility. Next, in Section 3, we construct a Bayesian IPP model, propose a recursive Bayes procedure for fitting the model, and compare various first-stage sampling strategies. We then demonstrate the multi-stage approach on simulated data in a compact window setting in Section 4. Finally, in Section 5, we apply our approach to data from aerial imagery surveys to learn spatially explicit abundance of harbor seals in Johns Hopkins Inlet, a protected glacial fjord in Glacier Bay National Park, Alaska, and describe strategies to account for additional spatial heterogeneity in the model.

2. Inhomogeneous Poisson process

We consider a space of interest, $\mathcal{S} \subset \mathbb{R}^2$, and a set of disjoint subspaces $\{A_l\}_{l=1}^L$ such that $\mathcal{S} = \cup_l A_l$. We let $\mathbf{s}_i \in \mathcal{S}$ be a two-dimensional vector denoting event location for $i = 1, \dots, n$ observed events. We also denote $N(\cdot)$ as a random counting measure and N as the total number of events in $\mathcal{S}$ (i.e., $N = N(\mathcal{S})$).

For an inhomogeneous Poisson process,

(i) $N(A_l)$ follows a Poisson distribution with rate parameter $\Lambda(A_l) := \int_{A_l} \lambda(\mathbf{s})\mathrm{d}\mathbf{s}$.

(ii) $N(A_1), \dots, N(A_L)$ are conditionally independent random variables for $l = 1, \dots, L$.

An IPP can also be interpreted as the continuous limit of an independent Poisson count model for an increasingly fine grid of subspaces (i.e., as L approaches ∞) over the spatial domain (Fithian and Hastie, 2013). Additional details and theoretical properties on the IPP are available in Illian et al. (2008) and Gelfand et al. (2010).

For p covariates associated with each point in $\mathcal{S}$, we can model the IPP intensity λ as a function of the spatially referenced covariates $\lambda(\mathbf{s}) = f\left(\beta_0 + \mathbf{x}'(\mathbf{s})\boldsymbol{\beta}\right)$, where β_0 represents the intercept, $\mathbf{x}$ and $\boldsymbol{\beta}$ denote $p \times 1$ vectors of covariates and their corresponding coefficients, respectively, and the function f represents a positive strictly increasing differentiable function. We assume the convention $f(x) = \exp(x)$. If desired, this assumption may be relaxed by using a different link function and/or transforming the covariates using a basis expansion. We discuss the latter in Section 5.

2.1. Complete likelihood

The IPP likelihood that jointly models the number of observed points and their locations is as follows

$$\left[\{\mathbf{s}_i\}, n|\beta_0, \boldsymbol{\beta}\right] = \frac{\prod_{i=1}^{n} \lambda(\mathbf{s}_i)}{n! \exp\left(\Lambda(\mathcal{S})\right)}, \tag{1}$$

for $i = 1, \dots, n$. We refer to this IPP likelihood in (1) as the complete likelihood hereafter.

In practice, the integral $\Lambda(\mathcal{S})$ in (1) is not analytically tractable and must be approximated. Using covariate information for grid centers $\mathbf{u}_l$ for $l = 1, \dots, L$, we can approximate $\Lambda(\mathcal{S})$ with numerical quadrature and use $\{\mathbf{u}_l\}_{l=1}^{L}$ as quadrature points, but this can be computationally intensive, especially for large L.

2.2. Conditional likelihood

The IPP likelihood conditional on n is as follows

$$\left[\{\mathbf{s}_i\}|n, \boldsymbol{\beta}\right] = \frac{\prod_{i=1}^{n} \lambda(\mathbf{s}_i)}{\left(\Lambda(\mathcal{S})\right)^n}, \tag{2}$$

for $i = 1, \dots, n$. We refer to this as the conditional likelihood hereafter. It is important to note that unlike in the complete likelihood, β_0 is non-identifiable because it cancels in the numerator and denominator in (2); however, inference on β_0 is necessary to estimate total abundance for partially observed domains. We let $\boldsymbol{\beta}$ denote the vector of coefficient values excluding the intercept hereafter for clarity.

The conditional likelihood (2) is commonly used in SDM literature and is appropriate for applications where the number of observed points n is treated as known and fixed *a priori*. For instance, in telemetric surveys for studying animal movement (Hooten et al., 2017), it is common to fit an SPP model to the observed spatially explicit points for a predetermined number of animal relocations (e.g., Johnson et al., 2013). The conditional likelihood suffers from similar computational challenges in approximating $\Lambda(\mathcal{S})$; however, useful likelihood approximations have been developed to circumvent this issue.

2.2.1. Logistic regression approximation

One widely adopted approximation for the conditional likelihood (2) in both point process and SDM literature is the logistic regression approximation, which involves augmenting m background points, also referred to as pseudo-absences, to the data to represent locations where point events were absent (Baddeley et al., 2010; Warton and Shepherd, 2010; Fithian and Hastie, 2013). Warton and Shepherd (2010) showed that if the number of observed points n is fixed and the number of background points m grows infinitely large, then the logistic regression coefficients (intercept excluded) converge to the maximum likelihood estimates of the IPP coefficients in the conditional likelihood (2).

To implement this approach, we use the Berman-Turner device (Berman and Turner, 1992), where we define an indicator variable, $y(\mathbf{s}) = \mathbb{1}(\mathbf{s} \in \{\mathbf{s}_i\}_{i=1}^{n})$, which indicates whether a location $\mathbf{s}$ is within our set of observed points. By definition, we have $y(\mathbf{s}_i) = 1$ for $i = 1, \dots, n$. To obtain points where $y(\mathbf{s}) = 0$, we generate a uniform random sample of size m within $\mathcal{S}$ (see Northrup et al., 2013 for guidance on choosing the number of background points m in practice). For m background points, we order $y(\mathbf{s}_i)$ such that $y(\mathbf{s}_i) = 1$ for $i = 1, \dots, n$ and $y(\mathbf{s}_i) = 0$ for $i = n+1, \dots, \tilde{n}$ where $\tilde{n} = n+m$. We then fit $y(\mathbf{s}_i) \sim \text{Bern}(p_i)$ where $\text{logit}(p_i) = \beta_0 + \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta}$. Although the logistic regression intercept is not relevant for the IPP model, note that its inclusion is necessary to adjust for the background sample size (Fieberg et al., 2021).

2.3. Compact observation window extension

In some scenarios, point events are only observed within compact observation windows. To extend the likelihood to a compact window setting, we denote each window using $\mathcal{S}_j$ for $j = 1, \dots, J$. The remaining unobserved area is defined as $\mathcal{S}_0 := \mathcal{S} \setminus \cup_j \mathcal{S}_j$ (see Fig. 2b for a visual representation of a compact window setting). We focus on the case where the probability of detecting point events in $\cup_j \mathcal{S}_j$ is 1 (i.e., perfect detection) and the probability of detecting point events in $\mathcal{S}_0$ is 0 (i.e., no detection). In such cases, we are often interested in inference on N, the total number of points in $\mathcal{S}$.

The joint likelihood of $\{\mathbf{s}_{ij}\}$, the set of event locations for each individual i in the jth window, and $\{n_j\}$, the set of point event counts within each window j, becomes

$$\left[\{\mathbf{s}_{ij}\}, \{n_j\}|\beta_0, \boldsymbol{\beta}\right] = \frac{\prod_{\{\forall j, n_j > 0\}} \prod_{i=1}^{n_j} \lambda(\mathbf{s}_{ij})}{n! \exp\left(\Lambda(\cup_j \mathcal{S}_j)\right)}, \tag{3}$$

for $i = 1, \dots, n$ and $j = 1, \dots, J$ where $n = \sum_j n_j$. We refer to (3) as the windowed complete likelihood hereafter. Similar to the conditional likelihood (2), the windowed conditional likelihood can be derived by conditioning on $\{n_j\}$.

3. Bayesian model

Using the complete likelihood (1), we can estimate and quantify uncertainty for the parameters of interest, β_0 and $\boldsymbol{\beta}$, using a Bayesian model. The full posterior distribution associated with our model is as follows:

$$\left[\beta_0, \boldsymbol{\beta} | \{\mathbf{s}_i\}, n\right] \propto \left[\{\mathbf{s}_i\}, n | \beta_0, \boldsymbol{\beta}\right][\beta_0][\boldsymbol{\beta}]. \tag{4}$$

Recall that, for the complete likelihood (1), $\Lambda(\mathcal{S})$ is analytically intractable. Using numerical quadrature to approximate this integral often involves computing a large sum that scales with the number of quadrature points. Furthermore, because $\Lambda(\mathcal{S})$ depends on β_0 and $\boldsymbol{\beta}$, the integral must be approximated each time the parameters are updated.

Secondly, we are unable to obtain closed form representations for the full-conditional distributions for β_0 and $\boldsymbol{\beta}$. If we use a normal random walk proposal for β_0 and $\boldsymbol{\beta}$, we must perform tuning, which incurs additional computational costs. Unsupervised MCMC algorithms, such as Hamiltonian Monte Carlo (HMC; Neal, 2011) with the No-U-Turn Sampler (NUTS; Hoffman and Gelman, 2014), may also be implemented, but require overhead calculations.

Instead of evaluating the full posterior distribution (4) using a conventional single-stage MCMC algorithm, we propose a multi-stage MCMC algorithm for fitting the model using the complete likelihood that does not require parameter tuning and allows for the $\Lambda(\mathcal{S})$ approximation to be computed in parallel.

3.1. Recursive Bayes algorithm

To construct our multi-stage MCMC algorithm, we apply PPRB following Hooten et al. (2021). Our algorithm follows a similar structure of the two-stage PPRB algorithm summarized in Section 1 with the addition of an intermediate stage for parallel computation.

We first show that the complete likelihood (1) can be decomposed into two terms:

$$\left[\{\mathbf{s}_i\}, n | \beta_0\ \boldsymbol{\beta}\right] = \frac{\prod_{i=1}^{n} \lambda(\mathbf{s}_i)}{n! \exp\left(\Lambda(\mathcal{S})\right)} \tag{5}$$

$$= \left(\frac{\prod_{i=1}^{n} \lambda(\mathbf{s}_i)}{\left(\Lambda(\mathcal{S})\right)^n}\right)\left(\frac{\left(\Lambda(\mathcal{S})\right)^n}{n! \exp\left(\Lambda(\mathcal{S})\right)}\right) \tag{6}$$

$$= \left[\{\mathbf{s}_i\} | \boldsymbol{\beta}, n\right][n | \beta_0, \boldsymbol{\beta}]. \tag{7}$$

Moreover, note that the first term in (7) is equivalent to the conditional likelihood (2). The second term, the likelihood for n, follows a Poisson distribution:

$$[n | \beta_0, \boldsymbol{\beta}] = \text{Pois}\left(\Lambda(\mathcal{S})\right). \tag{8}$$

The full posterior distribution can then be decomposed as

$$\left[\beta_0, \boldsymbol{\beta} | \{\mathbf{s}_i\}, n\right] \propto \left[\{\mathbf{s}_i\} | \boldsymbol{\beta}, n\right][n | \beta_0, \boldsymbol{\beta}][\beta_0][\boldsymbol{\beta}]. \tag{9}$$

Evaluating the decomposed posterior distribution (9) using PPRB results in two stages. In the first stage, we evaluate the transient posterior using the conditional likelihood (2) and the prior for $\boldsymbol{\beta}$

$$\left[\boldsymbol{\beta} | \{\mathbf{s}_i\}, n\right] \propto \left[\{\mathbf{s}_i\} | \boldsymbol{\beta}, n\right][\boldsymbol{\beta}] \tag{10}$$

to obtain an MCMC sample, $\left\{\boldsymbol{\beta}^{(k)} | \{\mathbf{s}_i\}, n\right\}_{k=1}^{K}$. Recall that in Section 2.2.1, we show that fitting the model using the conditional likelihood (2) can be approximated using logistic regression, which permits a variety of ways to obtain a sample from the transient posterior distribution. We describe and compare first-stage sampling strategies in the context of the multi-stage algorithm in the following section.

In the intermediate stage, we use the transient posterior realizations to compute corresponding values of $\Lambda(\mathcal{S})$ in parallel and store them for recall in the following stage. Approximating $\Lambda(\mathcal{S})$ is the most computationally intensive step in evaluating the likelihood for n and parallelization results in significant speedup for the final stage of the algorithm.

In the second and final stage, we use the transient posterior distribution as a prior and proposal distribution in an MCMC algorithm to account for the stochasticity associated with n in the full posterior distribution (4). Using the transient posterior distribution (10) as a prior and proposal distribution for $\boldsymbol{\beta}$ results in convenient cancellations involving the data model conditioned on n in the M-H acceptance ratio:

$$\alpha = \frac{\left[\{\mathbf{s}_i\}, n | {\beta_0}^{(*)}, \boldsymbol{\beta}^{(*)}\right]\left[{\beta_0}^{(*)}\right]\left[{\beta_0}^{(k-1)}\right]_*\left[\boldsymbol{\beta}^{(*)}\right]\left[\boldsymbol{\beta}^{(k-1)}\right]_*}{\left[\{\mathbf{s}_i\}, n | {\beta_0}^{(k-1)}, \boldsymbol{\beta}^{(k-1)}\right]\left[{\beta_0}^{(k-1)}\right]\left[{\beta_0}^{(*)}\right]_*\left[\boldsymbol{\beta}^{(k-1)}\right]\left[\boldsymbol{\beta}^{(*)}\right]_*} \tag{11}$$

$$= \frac{\left[\{\mathbf{s}_i\} | {\beta_0}^{(*)}, \boldsymbol{\beta}^{(*)}\right]\left[n | {\beta_0}^{(*)}, \boldsymbol{\beta}^{(*)}\right]\left[{\beta_0}^{(*)}\right]\left[{\beta_0}^{(k-1)}\right]_*\left[\boldsymbol{\beta}^{(*)}\right]\left[\boldsymbol{\beta}^{(k-1)}\right]_*}{\left[\{\mathbf{s}_i\} | {\beta_0}^{(k-1)}, \boldsymbol{\beta}^{(k-1)}\right]\left[n | {\beta_0}^{(k-1)}, \boldsymbol{\beta}^{(k-1)}\right]\left[{\beta_0}^{(k-1)}\right]\left[{\beta_0}^{(*)}\right]_*\left[\boldsymbol{\beta}^{(k-1)}\right]\left[\boldsymbol{\beta}^{(*)}\right]_*} \tag{12}$$

$$= \frac{\left[n | {\beta_0}^{(*)}, \boldsymbol{\beta}^{(*)}\right]\left[{\beta_0}^{(*)}\right]\left[{\beta_0}^{(k-1)}\right]_*}{\left[n | {\beta_0}^{(k-1)}, \boldsymbol{\beta}^{(k-1)}\right]\left[{\beta_0}^{(k-1)}\right]\left[{\beta_0}^{(*)}\right]_*}. \tag{13}$$

The $(*)$ superscript denotes a proposed value, the $(k-1)$ superscript denotes the value of a parameter at the $(k-1)$th step of the MCMC chain, and $[\cdot]_*$ denotes a proposal distribution. Critically, we do not need to perform additional calculations using the point data $\{\mathbf{s}_i\}$ at this stage.

Recall that we are unable to learn about β_0 using the first-stage logistic regression approximation. It is, however, crucial to efficiently propose β_0 for the M-H correction in the second stage to obtain reasonable MCMC mixing rates and prevent sample degeneracy. Conveniently, if we transform the intercept such that $\zeta = e^{\beta_0}$, we can obtain a Gibbs sampler using a Gamma(a, b) prior for ζ. The Gibbs update for ζ is based on the full-conditional distribution

$$[\zeta|\cdot] = \text{Gamma}\big(a+n, b+\Lambda(\mathcal{S})\big). \tag{14}$$

We then log-transform our posterior sample for ζ to obtain a sample from the full-conditional distribution $[\beta_0|\cdot]$. For a full derivation of the Gibbs update for ζ, see Appendix A.

The Gibbs sampler for β_0 also allows for the M-H ratio (13) to simplify to

$$\alpha = \big[n|{\beta_0}^{(k)}, \boldsymbol{\beta}^{(*)}\big] / \big[n|{\beta_0}^{(k)}, \boldsymbol{\beta}^{(k-1)}\big]. \tag{15}$$

Thus, each MCMC iteration in the second stage only requires updating β_0 using the Gibbs update (14) and evaluating the Poisson likelihood $\big[n|{\beta_0}^{(k)}, \boldsymbol{\beta}^{(*)}\big]$, both of which are fast using the pre-computed values of $\Lambda(\mathcal{S})$ from the intermediate stage. The resulting MCMC realizations from the second stage are realizations from the full posterior distribution (4). The full multi-stage algorithm is summarized in Algorithm 1.

Algorithm 1: Multi-stage PPRB MCMC

First Stage:
- **Data:** $\{s_i\}_{i=1}^n$
- **Parameters:** $\boldsymbol{\beta}$
- **for** $k = 1, \dots, K$ **do**
  - Sample from $\big[\boldsymbol{\beta}|\{\mathbf{s}_i\}, n\big] \propto \big[\{\mathbf{s}_i\}|\boldsymbol{\beta}, n\big][\boldsymbol{\beta}]$.
- **end**

Output: $\big\{\boldsymbol{\beta}^{(k)}|\{\mathbf{s}_i\}, n\big\}_{k=1}^K$

Intermediate Stage:
- **Input:** $\big\{\boldsymbol{\beta}^{(k)}|\{\mathbf{s}_i\}, n\big\}_{k=1}^K$
- **do in parallel**
  - **for** $k = 1, \dots, K$ **do**
    - Approximate $\Lambda^{(k)}(\mathcal{S})$ using first-stage $\boldsymbol{\beta}^{(k)}$.
  - **end**
- **end**

Output: $\big\{\Lambda^{(k)}(\mathcal{S})\big\}_{k=1}^K$

Second Stage:
- **Data:** n
- **Parameters:** β_0, $\boldsymbol{\beta}$
- Initialize β_0, $\boldsymbol{\beta}$ and compute corresponding $\Lambda(\mathcal{S})$.
- **for** $k = 1, \dots, K$ **do**
  - Sample $\zeta^{(k)} \sim \text{Gamma}\big(a+n, b+\Lambda^{(k)}(\mathcal{S})\big)$.
  - Set ${\beta_0}^{(k)} = \log\big(\zeta^{(k)}\big)$.
  - Sample uniformly with replacement $\boldsymbol{\beta}^{(*)} \sim \big\{\boldsymbol{\beta}^{(k)}|\{\mathbf{s}_i\}, n\big\}_{k=1}^K$.
  - Fetch corresponding $\Lambda^{(*)}(\mathcal{S})$.
  - Set $\boldsymbol{\beta} = \boldsymbol{\beta}^{(*)}$ and $\Lambda(\mathcal{S}) = \Lambda^{(*)}(\mathcal{S})$ w.p. $\alpha = \big[n|{\beta_0}^{(k)}, \boldsymbol{\beta}^{(*)}\big] / \big[n|{\beta_0}^{(k)}, \boldsymbol{\beta}^{(k-1)}\big]$.
  - Set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}$ and $\Lambda^{(k)}(\mathcal{S}) = \Lambda(\mathcal{S})$.
- **end**

Output: $\big\{{\beta_0}^{(k)}, \boldsymbol{\beta}^{(k)}|\{\mathbf{s}_i\}, n\big\}_{k=1}^K$

3.2. First-stage sampling strategies

One advantage to fitting the Bayesian SPP model in multiple stages is the increased flexibility facilitated by the ability to approximate the first stage using a logistic regression model. There are a variety of approaches for fitting a Bayesian logistic regression model such as the Pólya-Gamma data augmentation method described by Polson et al. (2013), which results in a fully Gibbs algorithm and alleviates the need to perform parameter tuning. Other unsupervised Monte Carlo methods, such as HMC with NUTS, for fitting logistic regression can be performed using existing software, such as `rstanarm` (Gabry and Goodrich, 2016) or

Table 1
Comparison of strengths between various first-stage sampling methods.

Property	PG	HMC	GLM-E	GLM-A
Unsupervised	✓	✓	✓	✓
Fully Bayesian	✓	✓	×	×
M-H cancellations	✓	✓	×	✓
Exact Bayesian inference	✓	✓	✓	×
Independent second stage proposals	×	×	✓	✓

`brms` (Bürkner et al., 2025) in R. We refer to fitting the multi-stage procedure with the Pólya-Gamma data augmentation and HMC approaches as the Pólya-Gamma (PG) and HMC methods, respectively, in reference to their first-stage sampling strategies.

An alternative to fitting the multi-stage algorithm as fully Bayesian is to fit the first stage using maximum likelihood estimation. This can be attractive because fitting a Bayesian logistic regression model in the first stage can be slow, especially after performing large data augmentations, such as the background points generated from the Berman-Turner device and/or the latent variables required for the Pólya-Gamma Gibbs sampler. This creates a computational bottleneck because the first stage cannot be parallelized and thus does not scale with the number of computational cores. Johnson et al. (2022) proposed a multi-stage PPRB algorithm for a different class of models in which the transient posterior distribution is approximated using non-Bayesian likelihood maximization in the first stage. A sample from the approximated transient posterior distribution can then be corrected using an approximate M-H ratio in later stages (Johnson et al., 2022).

Applying a similar approach to fit SPP models, we can perform fast maximum likelihood estimation for $\boldsymbol{\beta}$ in the first stage using the `glm` function in R to fit the logistic regression model described in Section 2.2.1. Assuming a flat prior for $\boldsymbol{\beta}$ and standard regularity conditions (i.e., the likelihood is continuously twice-differentiable), we can apply large sample theory to obtain the following approximation:

$$\left[\boldsymbol{\beta} \middle| \left\{y(\mathbf{s}_i)\right\}, n\right] \approx \mathrm{N}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}_{\beta}\right), \tag{16}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}_{\beta}$ are the maximum likelihood estimates and inverse observed Fisher information, respectively. We assume that the sample size is large enough for $\hat{\boldsymbol{\beta}}$ to be identifiable and well-defined. Next, we draw an i.i.d. sample from the approximate Gaussian distribution in (16) as an approximate sample from the transient posterior distribution.

Johnson et al. (2022) showed that the approximate transient posterior distribution (16) allows for approximate cancellations in the later stage M-H ratio and that parameter estimates obtained using this method are close to those obtained through a conventional single-stage MCMC algorithm. This method can be applied in our multi-stage procedure to obtain an approximate and fast sample from the posterior distribution for $\boldsymbol{\beta}$. Using the Gibbs sampler in (14), the M-H ratio in the second stage again reduces to (15). We refer to this multi-stage method as the GLM approximate (GLM-A) method hereafter.

3.2.1. GLM exact approach

If exact inference is desired, we propose an adjustment to the approximate approach developed by Johnson et al. (2022) to make it exact. After performing maximum likelihood estimation and obtaining an approximate transient posterior distribution as in (16), we can obtain an exact M-H correction for the complete likelihood (1) that retains the contribution from the original prior for $\boldsymbol{\beta}$ and the approximate transient posterior proposal distribution. Using the Gibbs sampler for ζ based on the full-conditional distribution in (14) results in the following M-H ratio:

$$\alpha = \frac{\left[\{\mathbf{s}_i\}, n | {\beta_0}^{(*)}, \boldsymbol{\beta}^{(*)}\right]\left[\boldsymbol{\beta}^{(*)}\right]\left[\boldsymbol{\beta}^{(k-1)}\right]_*}{\left[\{\mathbf{s}_i\}, n | {\beta_0}^{(k-1)}, \boldsymbol{\beta}^{(k-1)}\right]\left[\boldsymbol{\beta}^{(k-1)}\right]\left[\boldsymbol{\beta}^{(*)}\right]_*}. \tag{17}$$

Note that this approach differs from the previously described PPRB strategies (PG, HMC, and GLM-A) because we use the naive (i.e., noninformative) prior as opposed to the transient posterior distribution as the prior distribution for $\boldsymbol{\beta}$. However, this approach still differs from a traditional MCMC algorithm in that it uses a PPRB-style strategy to obtain an efficient proposal distribution and a multi-stage construction to leverage parallel computing resources.

To accelerate the evaluation of the log-likelihood for $\left[\{\mathbf{s}_i\}, n | {\beta_0}^{(*)}, \boldsymbol{\beta}^{(*)}\right]$ in the second stage, we compute $\sum_{i=1}^{n} \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta}$ in parallel in the intermediate stage in addition to $\Lambda(\mathcal{S})$ using the first stage transient posterior sample. Although the intermediate and second stages involve more calculations compared to the previous multi-stage methods (PG, HMC, and GLM-A), the overall algorithm results in significant speedup for this model compared to a conventional single-stage MCMC algorithm because of the fast first-stage evaluation using maximum likelihood estimation and the ability to parallelize large computations over k in the intermediate stage. We refer to this multi-stage method as the GLM exact (GLM-E) method hereafter.

Both the GLM-A and GLM-E methods result in significant speedup when compared to both the single-stage MCMC algorithm and fully Bayesian multi-stage methods (PG and HMC) due to fast maximum likelihood estimation in the first stage using the `glm` function in R. Furthermore, the independent samples from the first stage can help prevent sample degeneracy, a common issue that arises in multi-stage MCMC and other Bayesian filtering algorithms (Barreto and Hooten, 2025; Scharf, 2025; Taylor et al., 2025). Fig. 1 provides a flow chart that compares the various first-stage sampling strategies and Table 1 compares the strengths of each approach.

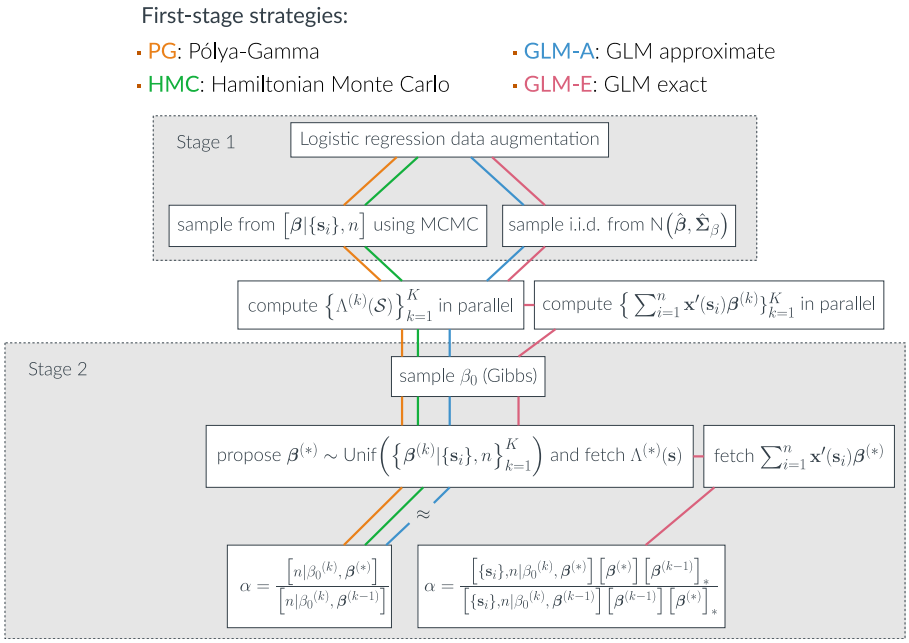

Fig. 1. Flowchart of multi-stage algorithm with various first-stage sampling strategies (PG, HMC, GLM-A, GLM-E).

3.3. Compact window prediction

3.3.1. Posterior prediction for total abundance

In the compact window setting, recall that we do not observe events that occur outside of the compact observation windows. We denote the number of events outside of the observation windows as n_0. If we assume that the point process across the entire spatial domain is an IPP, then by definition, n_0 also follows a Poisson distribution:

$$[n_0|\beta_0, \boldsymbol{\beta}, \{\mathbf{s}_{ij}\}, \{n_j\}] = \mathrm{Pois}(\Lambda(\mathcal{S}_0)). \tag{18}$$

The posterior predictive distribution for n_0 is expressed using

$$[n_0|\{s_{ij}\}, \{n_j\}] = \int\int [n_0|\beta_0, \boldsymbol{\beta}, \{s_{ij}\}, \{n_j\}][\beta_0, \boldsymbol{\beta}|\{s_{ij}\}, \{n_j\}]\mathrm{d}\beta_0\mathrm{d}\boldsymbol{\beta}. \tag{19}$$

After obtaining a posterior sample of β_0 and $\boldsymbol{\beta}$, we obtain a sample from the posterior predictive distribution for n_0 (19) via composition sampling by drawing ${n_0}^{(k)} \sim \mathrm{Pois}(\Lambda^{(k)}(\mathcal{S}_0))$, where $\Lambda^{(k)}(\mathcal{S}_0)$ is computed using ${\beta_0}^{(k)}$ and $\boldsymbol{\beta}^{(k)}$. Approximating $\Lambda^{(k)}(\mathcal{S}_0)$ still requires numerical quadrature, but can be computed in parallel for $k = 1, \ldots, K$. The posterior predictive sample for n_0 can then be used to perform a finite population correction for the total number of events in the spatial domain, $\mathcal{S}$: $N^{(k)} = n_0^{(k)} + n$ for $k = 1, \ldots, K$, which provides both a point estimate and uncertainty quantification for total abundance.

3.3.2. Posterior point simulation

We can simulate posterior predictive event locations in the unobserved region $\mathcal{S}_0$ after obtaining estimates for $\lambda(\mathbf{u})$ $\forall \mathbf{u} : \mathbf{u} \in \mathcal{S}_0$ using the Lewis-Shedler method for simulating from an IPP (Lewis and Shedler, 1979). Alternatively, a new point realization for the entire study domain $\mathcal{S}$ can be simulated using posterior estimates for $\lambda(\mathbf{u})$ $\forall \mathbf{u} : \mathbf{u} \in \mathcal{S}$. These results can then be used to characterize posterior predictive derived quantities or functions, such as the Ripley's K function or pair correlation function.

4. Simulation study

We demonstrate the multi-stage MCMC approach using a simulation study in a compact window setting. The data were generated from an IPP with an intensity function that is log-linear in terms of three known parameters across the entire study domain (Fig. 2). Only the events simulated within the predefined compact windows were used to fit the model with the windowed complete likelihood (3).

We fit our model to the simulated data using the first-stage sampling strategies described in Section 3.2 and a conventional single-stage MCMC algorithm for comparison on a 24-core machine with 3.68 GHz processors and 192 GB of RAM. Fig. 3 compares the posterior samples for β_0 and $\boldsymbol{\beta}$ obtained using the single-stage and the multi-stage MCMC algorithms. It is evident that the marginal posterior distributions for β_0 and $\boldsymbol{\beta}$ are almost identical among implementation approaches, barring Monte Carlo error

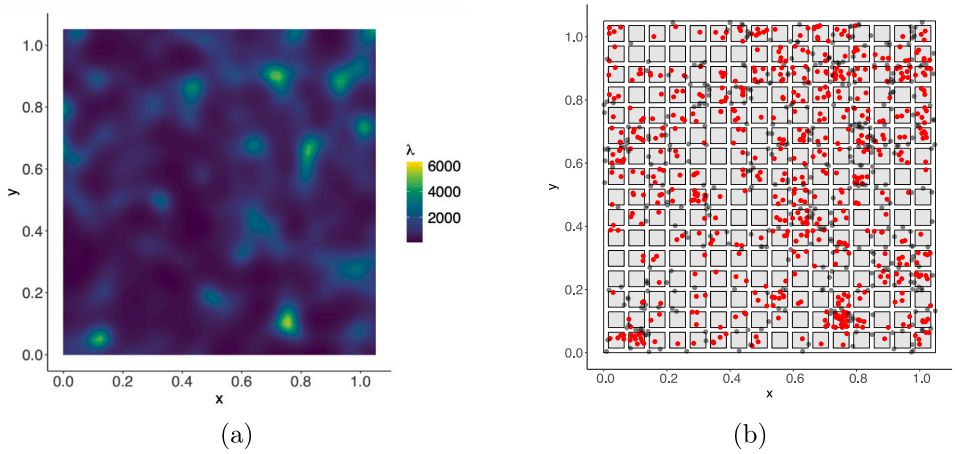

Fig. 2. (a) Intensity heat map for simulated data. (b) Realization of simulated points. Points located within the compact windows (red; $n = 580$) were used to fit the model.

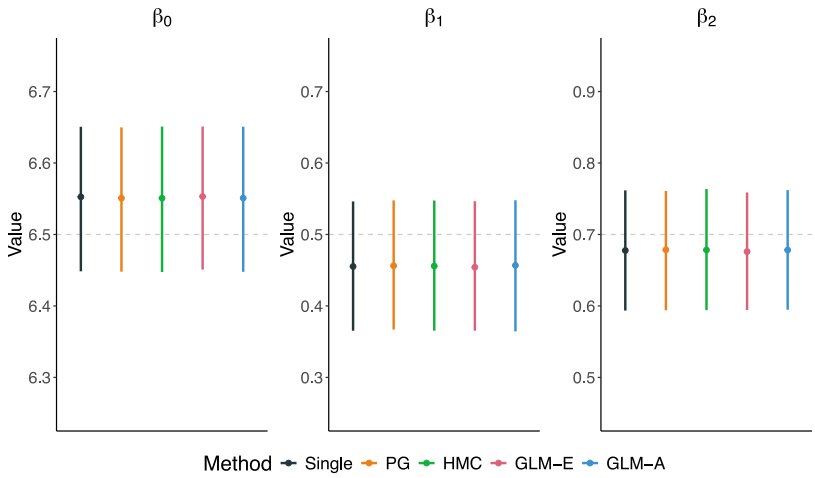

Fig. 3. Comparison of marginal posterior 95% credible intervals for the single-stage method and various multi-stage methods. The true values of the parameters are shown in dashed gray.

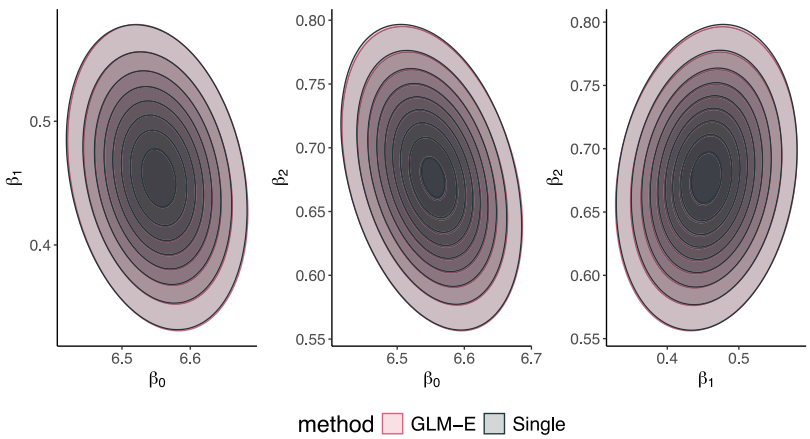

Fig. 4. Comparison of bivariate joint posterior distributions for each pair of model parameters. Contours represent kernel density estimates of the joint posterior distributions. The GLM-E contours (pink) are overlaid on top of the single-stage contours (gray).

Table 2
Seconds per effective sample for each coefficient and method. Lower values indicate more efficient sampling.

	β_0	β_1	β_2
Single-stage	0.132	0.149	0.158
PG	0.027	0.028	0.037
HMC	0.014	0.014	0.018
GLM-E	0.001	0.001	0.001
GLM-A	0.001	0.001	0.001

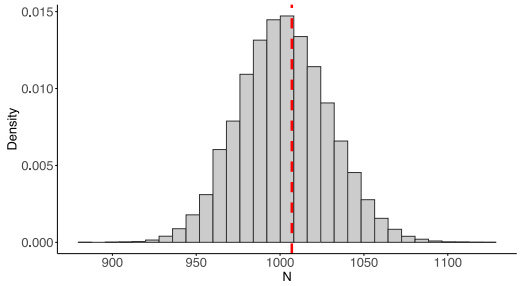

Fig. 5. Posterior predictive distribution for N using posterior realizations obtained using the GLM-E method. The true value of N is marked in dashed red.

and the approximation error from the logistic regression approximation. Similarly, the joint posterior distributions for each pair of model parameters appear nearly indistinguishable between the single-stage method and the GLM-E method (Fig. 4).

It is worth noting that while the PG, HMC, and GLM-A methods experience some, but often negligible, approximation error from the logistic regression approximation in the first stage, the GLM-E method is unaffected because it does not approximate the conditional likelihood (2) with logistic regression in the M-H correction. Increasing the number of background points m can decrease the logistic regression approximation error, but this incurs significant computational costs for the PG and HMC methods.

Table 2 compares the seconds per effective sample size between different methods for each parameter. The time recorded for the single-stage method does not include *a priori* parameter tuning, while the remaining multi-stage methods did not require tuning. The Bayesian first-stage methods (PG and HMC) showed an order of magnitude improvement, while the non-Bayesian first-stage methods (GLM-E and GLM-A) showed two orders of magnitude improvement. These results were due to both an improvement in runtime and MCMC effective sample size. All algorithms were fit using $L = 640,000$ quadrature points and the results scale with L and the number of computational cores. Fig. 5 shows the posterior predictive distribution for N obtained using the posterior samples from the GLM-E multi-stage method. Similar results were obtained for three additional simulations with varying compact window sizes and latent intensity fields (Appendix B).

5. Application: Harbor seal pups in Johns Hopkins inlet

Johns Hopkins Inlet (Tsalxaan Niyaadé Wool'éex'i Yé; Fig. 6) in Glacier Bay National Park is one of several tidewater glacial fjords in southeastern and south central Alaska that collectively host some of the largest seasonal aggregations of harbor seals (*Phoca vitulina*) in the world (Jansen et al., 2015; Womble et al., 2020). Harbor seals use the icebergs in these tidewater glacial fjords as haul out sites for resting, pupping, molting, and avoiding predators (Womble et al., 2020). Harbor seal abundance and space use in Johns Hopkins Inlet are of particular interest to management at Glacier Bay National Park because they help inform policy that aims to minimize potential seal disturbance while also allowing park visitors access to tidewater glaciers, one of the founding mandates of the park (National Park Service, 2010, 2025).

Harbor seal locations in Johns Hopkins Inlet were observed using aerial imagery surveys in June and August, the height of pupping and molting, respectively (Womble et al., 2020, 2021). Non-overlapping photos were taken along flight transects to avoid double-counting seals and provide systematic sampling of the entire inlet. Each image covered approximately 80 m × 120 m of surface area. We refer to the corresponding geo-referenced region as an image footprint hereafter. After imagery data were collected, a trained observer reviewed each digital image and annotated the spatial locations of adult and pup seals hauled out on ice. Further details on the survey methods and data processing are presented in Womble et al. (2020).

Previous studies analyzing these data have focused on modeling harbor seal abundance for the entire inlet or using coarse spatial grids. For example, Womble et al. (2020) incorporated the aerial imagery survey data along with other sources of data to obtain estimates and uncertainty quantification for harbor seal abundance. Womble et al. (2021) fit a multivariate conditional autoregressive (MCAR) model to summarize spatially explicit abundance on a 200 m × 300 m grid ($L = 434$ total grid cells). Jansen et al. (2015) used harbor seal data from a similar aerial survey to fit a generalized additive model (GAM) and summarized spatially explicit abundance on a grid of 400 m × 400 m cells ($L = 779$ total grid cells). Previous studies on spatially explicit abundance of harbor seals in Johns Hopkins Inlet have not explored modeling the locations as a spatial point process directly. Modeling these

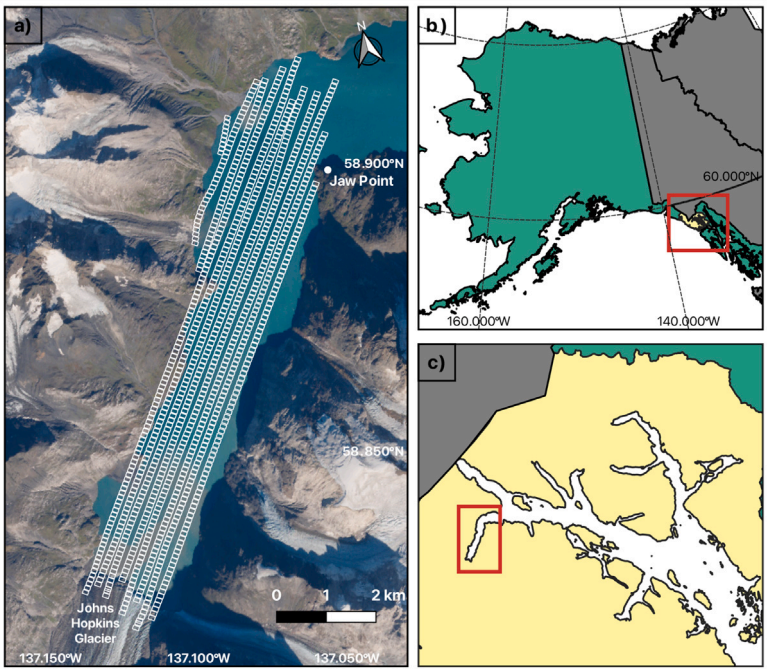

Fig. 6. (a) A Sentinel-2 satellite image of Johns Hopkins Inlet in 2018. Each white rectangle represents an image footprint from the 21 June 2007 survey. (b) and (c) show the location of Glacier Bay National Park and Johns Hopkins Inlet, respectively, in red.

data as an SPP provides unique advantages because it allows for informed simulation of potential seal locations outside the image footprints using an established statistical framework. Furthermore, the multi-stage algorithm makes fitting the model on a fine spatial grid more computationally flexible and feasible, especially with the availability of distributed computing resources.

We fit a Bayesian SPP model to harbor seal pup haul out location data from 21 June 2007 ($n = 452$; Fig. 7) using the aforementioned multi-stage MCMC approaches. We modeled harbor seal pups because they are the most vulnerable segment of the population and are especially susceptible to disturbance from vessels (Mathews et al., 2016). Furthermore, pups are hauled out on ice and observable with relatively high probability at the time the survey was conducted (i.e., in the month of June and in the afternoon) based on secondary analysis of wet-dry tag data of harbor seal pups in Disenchantment Bay, Alaska. Our study domain spanned from the Johns Hopkins Glacier terminus to Jaw Point, which is approximately 9 km long and 1.5–2 km wide (Fig. 6a). The image footprints for this survey cover approximately half of the study domain. We assume perfect detection of hauled out pups within the image footprints and no detection outside the image footprints, giving rise to compact observation windows. Additionally, due to the systematic surveying within the study domain, we assume the unobserved areas in the inlet to be ignorable missing data (i.e., the covariate values and relationship between covariates and pup locations do not differ significantly between observed and unobserved areas). Using the pup locations as point events, we fit the SPP model using the windowed complete likelihood (3).

We fit the model using three covariates: ice proportion, bathymetry, and distance to glacier terminus (Appendix C). The ice data were represented as iceberg outlines within the image footprints and summarized as ice proportions to allow for spatial prediction outside the image footprints (Appendix D). Each covariate was summarized on the same raster grid with a spatial resolution of approximately 10 m × 10 m ($L = 304{,}945$ total grid cells). The grid centers were used as quadrature points throughout the analysis.

5.1. Accounting for additional spatial heterogeneity

From preliminary analysis and model checking, we found that using a log-linear combination of the three covariates to model the IPP intensity was insufficient in capturing the heterogeneity of the observed harbor seal pup data (Appendix E). We first discuss how we modified the model to account for additional spatial heterogeneity for the harbor seal pup model in addition to alternative strategies and their compatibility with the multi-stage algorithm.

5.1.1. Neural network basis expansion

To increase model flexibility and allow for more complex relationships between covariates and intensity, we constructed a basis representation of $\mathbf{X}$, the $p \times L$ design matrix consisting of the covariate values at each grid center. Such practice is common in spatial modeling, with the popularity of fixed-rank (Cressie and Johannesson, 2008) and radial (Buhmann, 2000) basis functions. Of the basis functions typically used to account for spatial heterogeneity (Hefley et al., 2017; Wikle and Zammit-Mangion, 2023), we found that a single-hidden layer feedforward neural network (SLFNN) was able to capture the relationship between the covariates

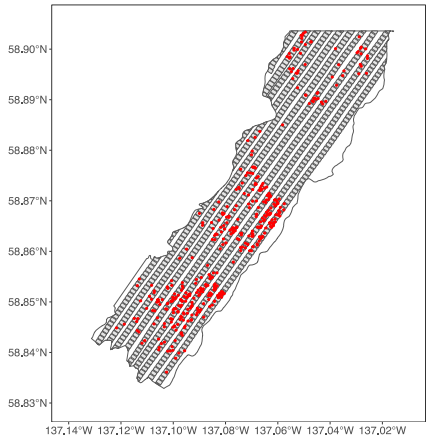

Fig. 7. Observed harbor seal pup haul out locations (red) on 21 June 2007 ($n = 452$). The large polygon represents the survey domain and each gray rectangle represents an image footprint.

and pup space use. Additionally, we found that the Extreme Learning Machine (ELM) scheme (Huang et al., 2006) was sufficient for constructing the neural network (NN) weights and resulted in efficient computation in the multi-stage algorithm.

To incorporate the NN basis expansion in the multi-stage algorithm, we first trained the NN in the first stage when fitting the logistic regression approximation. Using our samples generated from the Berman-Turner device $\left\{\left(\mathbf{x}(\mathbf{s}_i), y(\mathbf{s}_i)\right)\right\}_{i=1}^{\tilde{n}}$, we modeled the relationship between $\mathbf{x}(\mathbf{s}_i)$ and $\tilde{y}_i = \text{logit}(p_i)$ as a linear relationship using q hidden nodes

$$\sum_{j=1}^{q} \tilde{\beta}_j g(\boldsymbol{\omega}_j \cdot \mathbf{x}_i) := \tilde{y}_i, \tag{20}$$

where $g(\cdot)$ represents a user-specified activation function, $\boldsymbol{\omega}_j$ is a $p \times 1$ vector of weights, and $\tilde{\beta}_j$ represents the coefficient associated with the jth hidden node. The quantity in (20) can be summarized compactly as

$$\mathbf{W}\tilde{\boldsymbol{\beta}} = \tilde{\mathbf{y}}_i, \tag{21}$$

where

$$\mathbf{W} = \begin{bmatrix} g(\boldsymbol{\omega}_1 \cdot \mathbf{x}_1) & \cdots & g(\boldsymbol{\omega}_q \cdot \mathbf{x}_1) \\ \vdots & \cdots & \vdots \\ g(\boldsymbol{\omega}_1 \cdot \mathbf{x}_{\tilde{n}}) & \cdots & g(\boldsymbol{\omega}_q \cdot \mathbf{x}_{\tilde{n}}) \end{bmatrix}_{\tilde{n} \times q}. \tag{22}$$

The matrix $\mathbf{W}$ is referred to as the hidden layer output matrix. Instead of learning the weights, $\boldsymbol{\omega}_j$ for $j = 1, \ldots, q$ using gradient-based methods, the ELM scheme samples random weights from a user-specified distribution for efficient computation (Huang et al., 2006). After we computed $\mathbf{W}$ using the randomly generated weights, we used it as a basis expansion for $\mathbf{X}$ (i.e., each hidden node represents a basis vector), which then served as a new design matrix for the first-stage logistic regression problem. Aside from replacing $\mathbf{X}$ with $\mathbf{W}$, the intermediate and second stages in the multi-stage algorithm remain unchanged.

For the 21 June 2007 harbor seal pup data, we used the Gaussian Error Linear Units (GELU) activation function (Hendrycks and Gimpel, 2016) defined as $g(\mathbf{x}) = \mathbf{x}\boldsymbol{\Phi}(\mathbf{x})$, where $\boldsymbol{\Phi}(x)$ is a standard Gaussian cumulative distribution function, and generated 100 hidden layer output matrices for $q = 5$ and weights sampled from a standard Gaussian distribution. We selected the $\mathbf{W}$ matrix with the optimal AIC based on (20) for the basis expansion. We then used the GLM-E multi-stage algorithm to obtain $K = 100{,}000$ posterior realizations for $\tilde{\boldsymbol{\beta}}$ on the same machine as the simulation study (24-core, 3.68 GHz processors and 192 GB of RAM), which required approximately three minutes for the entire multi-stage algorithm. The posterior realizations were used to simulate points across the entire study domain using the method described in Section 3.3.2. The simulated points were summarized in a grid of posterior mean abundance (Fig. 8). We validated our model using posterior predictive model-checking and did not observe large discrepancies between the simulated and observed data (Appendix E).

When compared to the observed point pattern in Fig. 7, Fig. 8 indicates that greater abundance values were simulated in areas of high observed pup concentrations in addition to surrounding areas that were originally unobserved. Areas of greater simulated abundance values are also consistent with areas of higher ice proportions and more negative bathymetry values (i.e., areas of deeper water), supporting previous studies on harbor seal behavior in Southeast Alaska (Montgomery et al., 2007; Kaluzienski et al., 2023). Additionally, although we did not include this as a covariate, greater abundance values were simulated in areas where eddies typically form in the inlet, which is consistent with findings that harbor seals prefer relatively slow-moving icebergs during the pupping season (Kaluzienski et al., 2025). The posterior predictive distribution for total abundance N was generated using the procedure in Section 3.3.1 (Fig. 9). The posterior mean and standard deviation of N were 835.2 and 27.7, respectively.

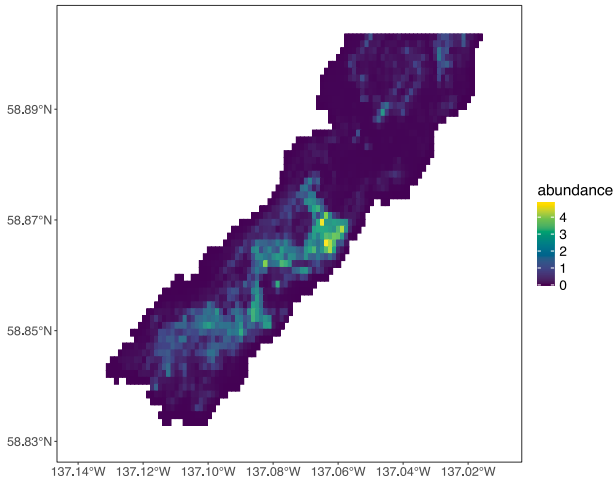

Fig. 8. Spatially explicit posterior mean abundance of simulated harbor seal pup haul out locations on 21 June 2007.

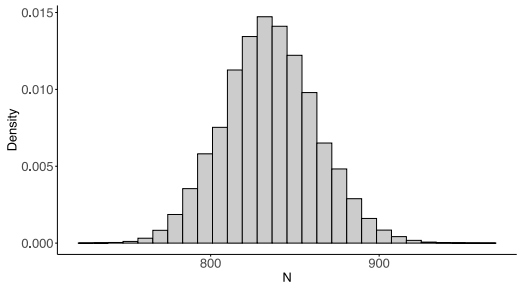

Fig. 9. Posterior predictive distribution for total abundance of hauled out seal pups N obtained using the GLM-E posterior realizations.

5.1.2. Alternative strategies for increased model flexibility

Although we found the NN basis expansion approach to be both sufficient for capturing complex spatial patterns, such as the harbor seal pup data, and computationally efficient within the multi-stage computing algorithm, alternative strategies for increased model flexibility exist and may be applicable. For instance, fixed-rank or radial basis functions may be used instead of the NN basis functions. Dovers et al. (2024) showed that a similar basis function approach can be used to approximate an LGCP model and only requires fitting a generalized linear mixed model (GLMM) in R. Fitting the multi-stage algorithm using an SPP model with a true latent Gaussian process to capture additional spatial heterogeneity as in an LGCP model (Brix and Diggle, 2001) is achievable, but introduces additional computational complexities associated with learning parameters for a high-dimensional latent field, which is the subject of ongoing research.

6. Discussion

Recent innovation in technology and survey techniques have led to significant improvements in the availability and spatial resolution of datasets, making the development of computationally-efficient and scalable SPP models especially crucial. Many useful methods have been developed to facilitate flexible and efficient SPP modeling (Baddeley et al., 2010; Warton and Shepherd, 2010; Aarts et al., 2012; Fithian and Hastie, 2013) for conditional IPP likelihoods as in (2). Furthermore, developments in using the Laplace approximation to approximate Bayesian SPP model marginal distributions have led to significant computational improvements (Rue and Martino, 2009; Illian et al., 2012) for fitting LGCP models. However, fitting an SPP model that both uses the complete likelihood (1), which accounts for the stochasticity associated with n, and provides exact Bayesian inference for the joint posterior distribution has remained computationally challenging.

More recently, recursive Bayes methods for fitting ecological statistical models have gained traction and have been shown to substantially improve computational efficiency (Hooten et al., 2021, 2023, 2024; Leach et al., 2022; McCaslin et al., 2021). We showed that the complete likelihood for an IPP model (1) can be similarly fit in stages where the coefficients are first estimated using the likelihood conditioned on n (2), then corrected with the Poisson likelihood for n (8). Although sample degeneracy is a commonly cited issue for recursive Bayes and other sequential filtering algorithms (Barreto and Hooten, 2025; Scharf, 2025; Taylor et al., 2025), our multi-stage algorithm did not suffer from this phenomenon due to only having two stages and the ability to sample the additional learned parameter in the second stage, β_0, using a Gibbs sampler. This is especially true for the non-Bayesian first stage

strategies, GLM-A and GLM-E, because the transient posterior realizations from (16) are independent. In practice, the acceptance rate for $\boldsymbol{\beta}$ in the second stage ranged from 0.25 to 0.7 in our simulations and case study.

Fitting the IPP model in stages provides many benefits including increased flexibility in the first stage. We showed that the first stage is approximately equivalent to fitting a model using the conditional likelihood (2), which is well approximated using logistic regression. This approximation provides various ways to obtain realizations from the transient posterior distribution, including user-friendly GLM R packages, making fitting Bayesian SPP models more accessible to a broader audience. Additionally, the multi-stage methods we presented do not require algorithm supervision and result in improved mixing for MCMC chains when compared to a single-stage MCMC algorithm, especially for the non-Bayesian first-stage approaches, GLM-E and GLM-A. Furthermore, all aforementioned multi-stage methods leverage parallel computing resources to pre-compute computationally intensive values, allowing the method to scale with the number of computer cores and increase the feasibility of fitting an IPP model using a large number of quadrature points.

The multi-stage construction of the algorithm also allows practitioners to take a meta-analytic approach to adjust model output that was originally fit using the conditional IPP likelihood (2) by incorporating data that are already available. The original MCMC sample (or approximate Gaussian distribution if non-Bayesian) can be treated as the transient posterior realizations in the multi-stage algorithm. Proceeding with the intermediate and second stages adjusts the estimates to account for the stochasticity associated with n and provides learning about the β_0 parameter, which was non-identifiable during the original model fitting. The resulting model output provides exact Bayesian inference using the complete likelihood in (1).

Restricting the IPP intensity function to be log-linear in the available covariates may not be sufficient to capture spatial heterogeneity in complex spatial patterns, such as the harbor seal pup data in Fig. 7. In Section 5.1, we outlined ways to account for additional spatial heterogeneity in the model, including using an NN basis expansion, for which training is compatible and efficient within the multi-stage algorithm. Furthermore, including x- and y-coordinates in the basis expansion mimics the LGCP in that they account for spatial heterogeneity that is unexplained by the observed covariates (Dovers et al., 2024). One main limitation of modeling the points as an IPP or LGCP is that points are assumed conditionally independent given the spatial covariates (and latent field for the LGCP). Basis functions and/or latent fields can allow for flexible estimation of complex intensity fields, but additional dependence among points may be necessary to reasonably capture extreme forms of spatial heterogeneity such as clustering due to social structure. Exploratory data analysis on the observed point pattern and model validation techniques concerned with the second-order properties of a point pattern (Illian et al., 2008) should be explored during analysis. Future directions of this work include extending the multi-stage procedure for self-exciting point processes such as the Hawkes process (Hawkes, 1971).

Finally, we showed how this method can be extended to the compact window setting where a posterior predictive distribution for the total abundance N can be obtained. Furthermore, this method allows for intensity estimation and point simulation that lie outside the observation windows, which can be useful for downstream analyses.

Acknowledgments

This research was funded by the National Science Foundation Graduate Research Fellowship Program and the National Park Service, United States. The authors thank Paul Conn, Brett McClintock, Justin Van Ee, Daniel Barreto, and Myungsoo Yoo for helpful discussions and insights. We also thank Jason Amundson and John Jansen for providing relevant data.

Appendix A

Recall that the complete likelihood is given by

$$\left[\{\mathbf{s}_i\}, n|\beta_0, \boldsymbol{\beta}\right] = \frac{\prod_{i=1}^n \lambda(\mathbf{s}_i)}{n!\exp\left(\Lambda(\mathcal{S})\right)}.$$

We can rewrite the complete likelihood using $\zeta := e^{\beta_0}$:

$$\begin{aligned}\left[\{\mathbf{s}_i\}, n|\beta_0, \boldsymbol{\beta}\right] &= \frac{\prod_{i=1}^n \zeta \exp\{\mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta}\}}{n!\exp\left\{\zeta\Lambda(\mathcal{S})\right\}} \\ &= \frac{1}{n!}\zeta^n \exp\left\{-\zeta\Lambda(\mathcal{S}) + \exp\{\mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta}\}\right\}.\end{aligned}$$

Then, using a vague Gamma(a, b) prior with shape-rate parameterization, we can obtain a closed-form full-conditional distribution for ζ:

$$\begin{aligned}\left[\zeta|\{\mathbf{s}_i\}, n, \boldsymbol{\beta}\right] &\propto [\{\mathbf{s}_i\}, n|\zeta, \boldsymbol{\beta}][\zeta] \\ &\propto \zeta^n \exp\left\{-\zeta\Lambda(\mathcal{S})\right\}\zeta^{a-1}\exp\{-b\zeta\} \\ &\propto \text{Gamma}\left(a+n, b+\Lambda(\mathcal{S})\right).\end{aligned}$$

Recall that $\Lambda(\mathcal{S})$ in the full-conditional update can be pre-computed in the previous stage in the multi-stage MCMC algorithm. Thus, sampling from the full-conditional distribution for ζ in the second stage is fast.

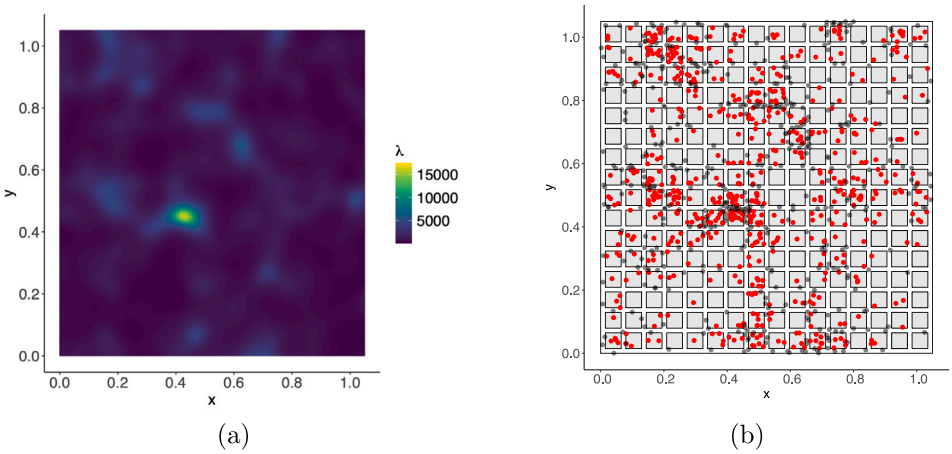

Fig. 10. (a) Intensity heat map for simulated data. (b) Realization of simulated points. Points located within the compact windows (red; n = 617) were used to fit the model.

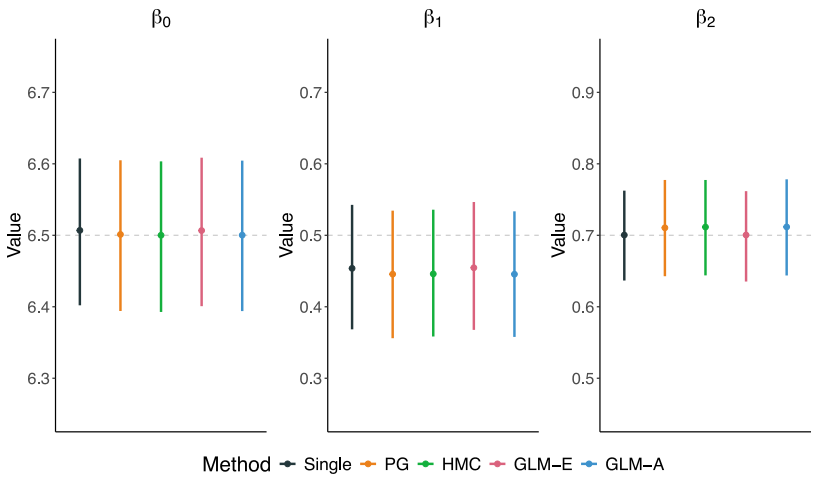

Fig. 11. Comparison of marginal posterior 95% credible intervals for the single-stage method and various multi-stage methods. The true values of the parameters are shown in dashed gray.

Appendix B

We conducted simulations using two distinct compact windowed constructions in which the windows cover approximately half of the study domain area. For each of the windowed constructions, we fit the IPP model with the windowed complete likelihood (3) using a single-stage MCMC algorithm and each multi-stage method to data generated from two distinct latent intensity fields. The simulation study with small windows and point pattern 1 is presented in Section 4, while the remaining three simulations are shown below.

Small windows, point pattern 2

See Figs. 10–13 and Table 3.

Large windows, point pattern 1

See Figs. 14–17 and Table 4.

Large windows, point pattern 2

See Figs. 18–21 and Table 5.

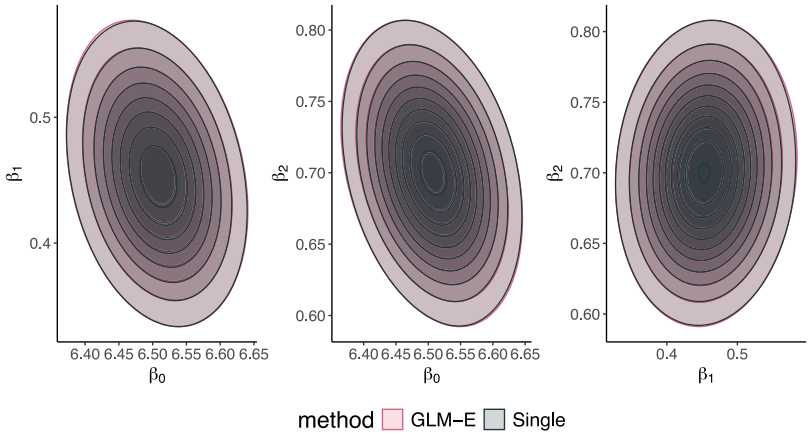

Fig. 12. Comparison of bivariate joint posterior distributions for each pair of model parameters. Contours represent kernel density estimates of the joint posterior distributions. The GLM-E contours (pink) are overlaid on top of the single-stage contours (gray).

Table 3
Seconds per effective sample for each coefficient and method. Lower values indicate more efficient sampling.

	β_0	β_1	β_2
Single-stage	0.170	0.167	0.146
PG	0.038	0.032	0.045
HMC	0.017	0.015	0.020
GLM-E	0.001	0.002	0.002
GLM-A	0.001	0.001	0.002

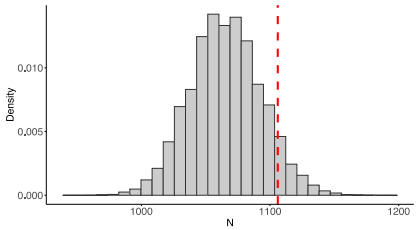

Fig. 13. Posterior predictive distribution for N using posterior realizations obtained using the GLM-E method. The true value of N is marked in dashed red.

Table 4
Seconds per effective sample for each coefficient and method. Lower values indicate more efficient sampling.

	β_0	β_1	β_2
Single-stage	0.128	0.139	0.137
PG	0.028	0.029	0.035
HMC	0.015	0.015	0.018
GLM-E	0.001	0.001	0.001
GLM-A	0.001	0.001	0.001

Table 5
Seconds per effective sample for each coefficient and method. Lower values indicate more efficient sampling.

	β_0	β_1	β_2
Single-stage	0.162	0.174	0.157
PG	0.040	0.037	0.048
HMC	0.023	0.021	0.027
GLM-E	0.002	0.002	0.002
GLM-A	0.001	0.001	0.001

Appendix C

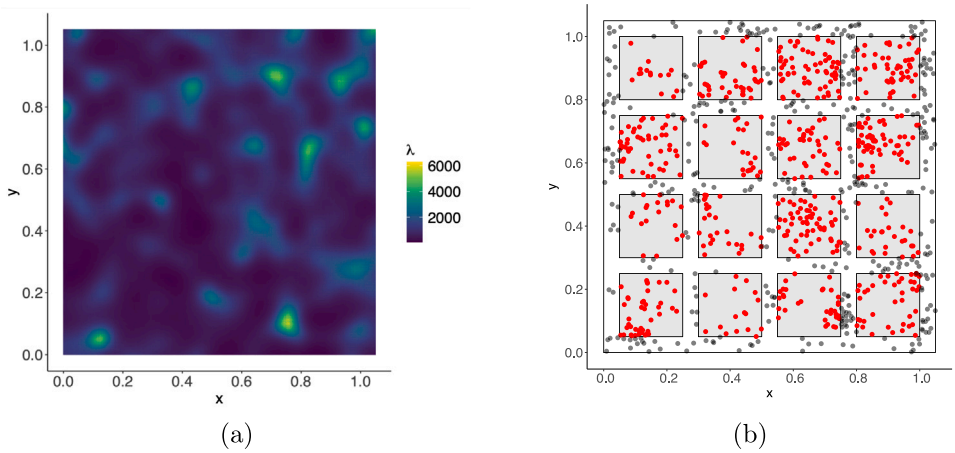

Fig. 14. (a) Intensity heat map for simulated data. (b) Realization of simulated points. Points located within the compact windows (red; n = 528) were used to fit the model.

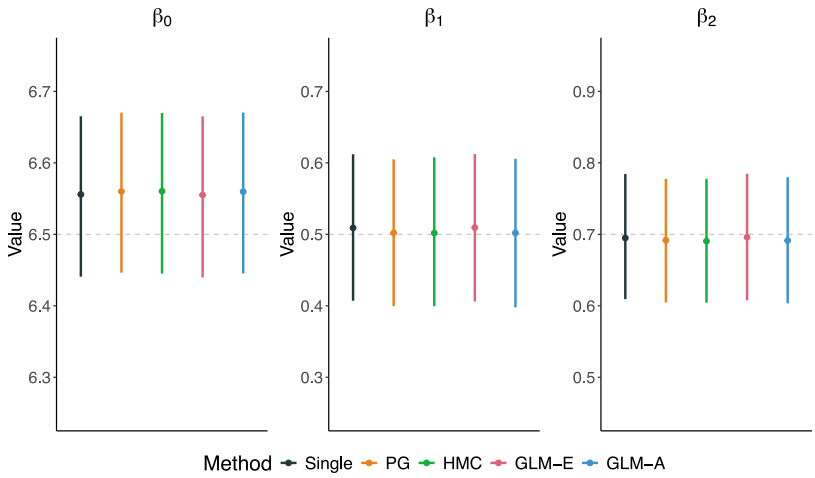

Fig. 15. Comparison of marginal posterior 95% credible intervals for the single-stage method and various multi-stage methods. The true values of the parameters are shown in dashed gray.

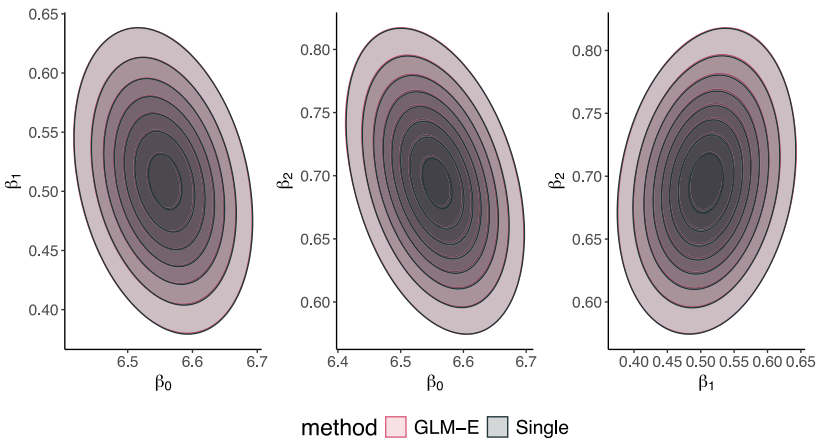

Fig. 16. Comparison of bivariate joint posterior distributions for each pair of model parameters. Contours represent kernel density estimates of the joint posterior distributions. The GLM-E contours (pink) are overlaid on top of the single-stage contours (gray).

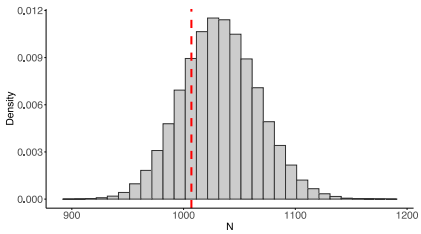

Fig. 17. Posterior predictive distribution for N using posterior realizations obtained using the GLM-E method. The true value of N is marked in dashed red.

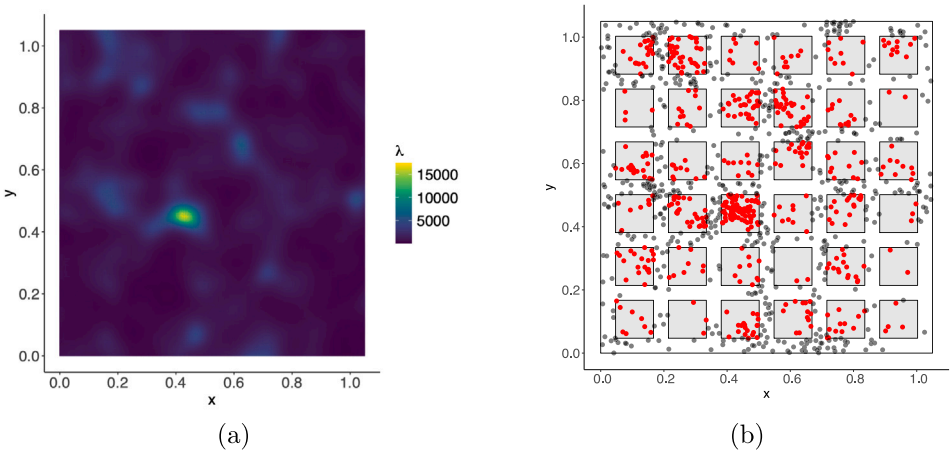

Fig. 18. (a) Intensity heat map for simulated data. (b) Realization of simulated points. Points located within the compact windows (red; n = 483) were used to fit the model.

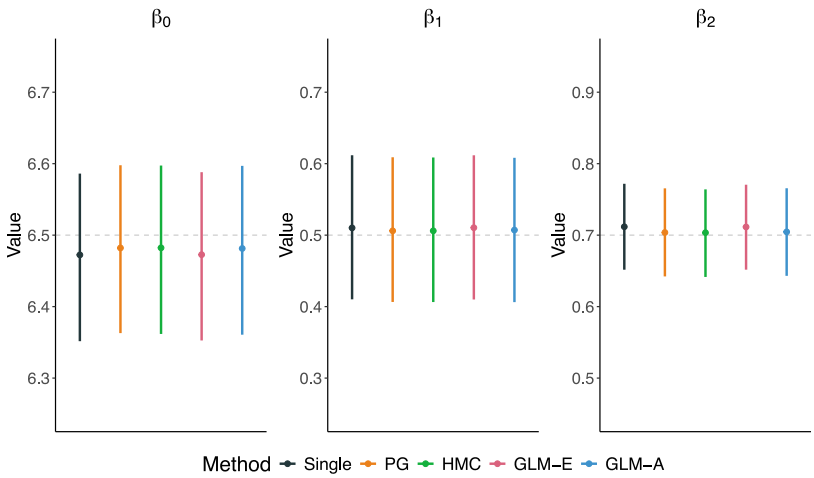

Fig. 19. Comparison of marginal posterior 95% credible intervals for the single-stage method and various multi-stage methods. The true values of the parameters are shown in dashed gray.

We used three covariates (bathymetry, ice proportion, and distance to glacier) to fit the IPP model to harbor seal pup data on 21 June 2007. For the bathymetry covariate, we downloaded a raster with cell size 1/3 arc-second (≈10 m) from the National Oceanic and Atmospheric Administration (NOAA) Hydrographic Survey Metadata Database (HSMD). The raster values are relative to Mean Lower Low Water, with more negative values corresponding to deeper waters. We used the bathymetry raster grid to construct the remaining covariate rasters. Ice data were collected using the same aerial imagery survey described in Section 5. Kaluzienski et al. (2023) conducted an ice segmentation approach to obtain iceberg outlines within each image footprint. We used the iceberg outlines

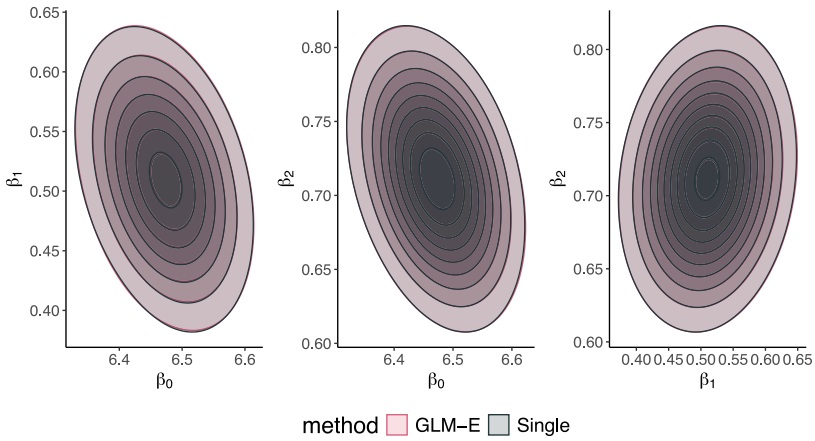

Fig. 20. Comparison of bivariate joint posterior distributions for each pair of model parameters. Contours represent kernel density estimates of the joint posterior distributions. The GLM-E contours (pink) are overlaid on top of the single-stage contours (gray).

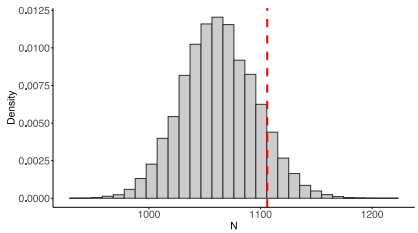

Fig. 21. Posterior predictive distribution for N using posterior realizations obtained using the GLM-E method. The true value of N is marked in dashed red.

to compute iceberg proportions and performed kriging to complete the ice proportion raster (Appendix D). To construct the distance to glacier terminus raster, we computed the Euclidean distance from each grid center to the closest point on the glacier terminus.

Appendix D

The goals for utilizing the aerial ice imagery were to quantify ice availability from the raw images by creating a rasterized map of ice proportions and produce a map that has the same resolution and spatial extent as the other covariates used in the model (i.e., bathymetry and distance from glacier terminus) across the entire study domain. To ensure an appropriate grid to account for the varying locations of observed ice proportions, the geometries of each image footprint in the survey date were first intersected with a pre-specified raster (Fig. 9A). The resulting output was a series of grid cell geometries and their corresponding centroid coordinates.

Using these grid cell geometries, as well as a binarized version of the observed icebergs, an iterative masking method was used to find the proportion of ice within each grid cell. For each image in the survey date, the raster grid was superimposed onto the image and a logical mask was used to iterate through the grid cell and image intersections. Within a grid cell, the total number of iceberg pixels was divided by the total number of pixels in the grid cell, resulting in an ice proportion value. This process was repeated for each grid cell in an image, and for each image in the survey date (Fig. 22B). Because ice was only observed within the areas of the image footprints, the calculated ice proportions and the `LatticeKrig` R package (Nychka et al., 2023) were used to produce an interpolated surface of ice proportions across the entire survey domain. After performing kriging, a raster was produced, where ice proportion estimates were available at the same resolution as the other covariates used, and across the entire survey domain.

Appendix E

To check whether the observed data reasonably arises from the fitted IPP model, we performed a posterior predictive check based on the L-function, a function used to summarize the level of clustering in a spatial point pattern at varying distances. For each $\beta_0^{(k)}$ and $\boldsymbol{\beta}^{(k)}$ posterior realization obtained from the multi-stage GLM-E method, we simulated points within the image footprints and computed the corresponding L-function using an isotropic edge correction. These were then compared to the L-function computed for the observed harbor seal pup locations.

We performed this model checking technique for the model with and without the NN basis expansion. For the model without the NN basis expansion, we found that the L-function computed from the observed pup data did not lie within the realm of posterior

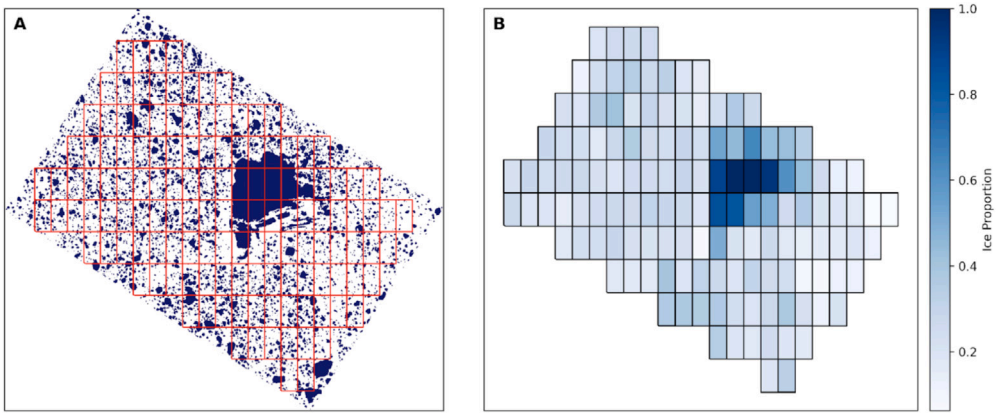

Fig. 22. (A) Pre-processed iceberg outlines for one image footprint in the 21 June 2007 survey. (B) Computed ice proportions on the pre-specified covariate grid.

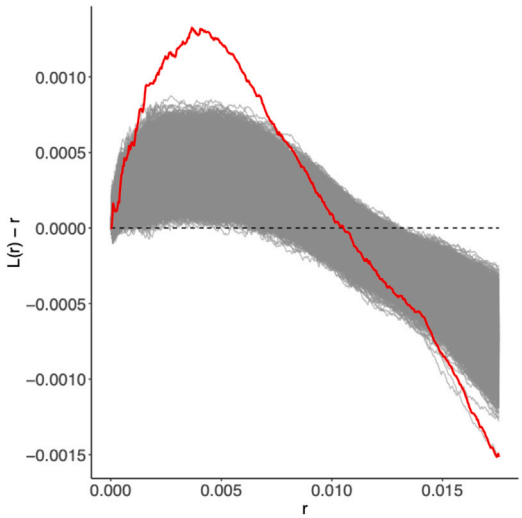

Fig. 23. Posterior predictive L-functions from the model without NN basis expansion (gray) compared to the observed L-function (red).

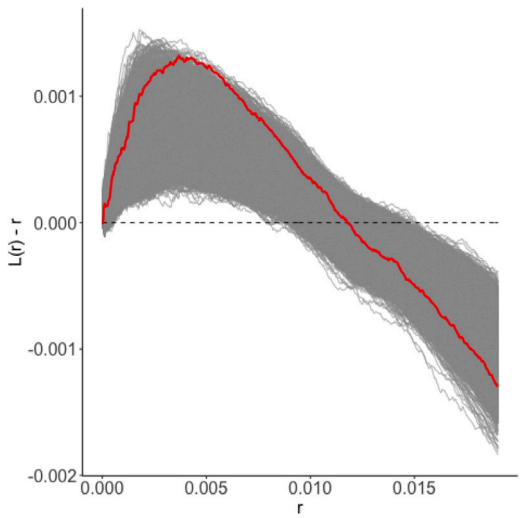

Fig. 24. Posterior predictive L-functions from the model with NN basis expansion (gray) compared to the observed L-function (red).

predictive L-functions (Fig. 23), indicating that the model did not sufficiently account for clustering in the observed point pattern. However, after refitting the model with the inclusion of the NN basis expansion, we found that the L-function computed from the observed pup data lie within the realm of posterior predictive L-functions (Fig. 24), indicating a reasonable model fit.

References

Aarts, G., Fieberg, J., Matthiopoulos, J., 2012. Comparative interpretation of count, presence–absence and point methods for species distribution models. Methods Ecol. Evol. 3, 177–187.

Baddeley, A., Berman, M., Fisher, N.I., Hardegen, A., Milne, R.K., Schuhmacher, D., Shah, R., Turner, R., 2010. Spatial logistic regression and change-of-support in Poisson point processes. Electron. J. Stat. 4, 1151–1201.

Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications with R. Chapman & Hall/CRC Interdisciplinary Statistics.

Barreto, D.W., Hooten, M.B., 2025. Recursive adaptive importance sampling with optimal replenishment. arXiv preprint arXiv:2509.08102.

Berman, M., Turner, T.R., 1992. Approximating point process likelihoods with GLIM. J. R. Stat. Soc. 41, 31–38.

Brix, A., Diggle, P.J., 2001. Spatiotemporal prediction for log-Gaussian cox processes. J. R. Stat. Soc. 63, 823–841.

Buhmann, M.D., 2000. Radial basis functions. Acta Numer. 9, 1–38.

Bürkner, P.-C., Gabry, J., Weber, S., Johnson, A., Modrak, M., Badr, H.S., Weber, F., Vehtari, A., Ben-Shachar, M.S., Rabel, H., Mills, S.C., Wild, S., Popov, V., Kosmidis, I., Schneider, B., Kallioinen, N., Silva, W.J., Carvalho, L., Pekin, S., 2025. brms: Bayesian regression models using 'stan'. R package version 2.23.0.

Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. 70, 209–226.

Dovers, E., Stoklosa, J., Warton, D.I., 2024. Fitting log-Gaussian cox processes using generalized additive model software. Amer. Statist. 78, 418–425.

Fieberg, J., Signer, J., Smith, B., Avgar, T., 2021. A 'how to' guide for interpreting parameters in habitat-selection analyses. J. Anim. Ecol. 90, 1027–1043.

Fithian, W., Hastie, T., 2013. Finite-sample equivalence in statistical models for presence-only data. Ann. Appl. Stat. 7, 1917–1939.

Gabry, J., Goodrich, B., 2016. rstanarm: Bayesian applied regression modeling via stan. R package version 2.21.1.

Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M., 2010. Handbook of Spatial Statistics. Chapman & Hall/CRC Interdisciplinary Statistics.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Amer. Statist. Assoc. 85, 398–409.

Haase, P., 1995. Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction. J. Veg. Sci. 6, 575–582.

Hawkes, A.G., 1971. Spectra of some self-exciting and mutually exciting point processes. Biometrika 58, 83–90.

Hefley, T.J., Broms, K.M., Brost, B.M., Buderman, F.E., Kay, S.L., Scharf, H.R., Tipton, J.R., Williams, P.J., Hooten, M.B., 2017. The basis function approach for modeling autocorrelation in ecological data. Ecology 98, 632–646.

Hendrycks, D., Gimpel, K., 2016. Gaussian error linear units (GELUs). arXiv preprint arXiv:1606.08415.

Hoffman, M.D., Gelman, A., 2014. The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res. 15, 1593–1623.

Hooten, M.B., Johnson, D.S., Brost, B.M., 2021. Making recursive Bayes inference accessible. Amer. Statist. 75, 185–194.

Hooten, M.B., Johnson, D.S., McClintock, B.T., Morales, J.M., 2017. Animal Movement: Statistical Models for Telemetry Data. CRC Press.

Hooten, M.B., Schwob, M.R., Johnson, D.S., Ivan, J.S., 2023. Multistage hierarchical capture-recapture models. Environmetrics 34, e2799.

Hooten, M.B., Schwob, M.R., Johnson, D.S., Ivan, J.S., 2024. Geostatistical capture-recapture models. Spat. Stat. 59, 100817.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley & Sons.

Illian, J.B., Sørbye, S.H., Rue, H., 2012. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). Ann. Appl. Stat. 6, 1499–1530.

Jansen, J.K., Boveng, P.L., Ver Hoef, J.M., Dahle, S.P., Bengston, J.L., 2015. Natural and human effects on harbor seal abundance and spatial distribution in an Alaskan glacial fjord. Mar. Mam. Sci. 31, 66–89.

Johnson, D.S., Brost, B.M., Hooten, M.B., 2022. Greater than the sum of its parts: Computationally flexible Bayesian hierarchical modeling. J. Agric. Biol. Environ. Stat. 27, 382–400.

Johnson, D.S., Hooten, M.B., Kuhn, C.E., 2013. Estimating animal resource selection from telemetry data using point process models. J. Anim. Ecol. 82, 1155–1164.

Kaluzienski, L.M., Amundson, J.M., Womble, J.M., Bliss, A.K., Pearson, L.E., 2023. Impacts of tidewater glacier advance on iceberg habitat. Ann. Glaciol. 64, 44–54.

Kaluzienski, L.M., Amundson, J.M., Womble, J.N., Bliss, A.K., Pearson, L.E., 2025. Fine-scale variability in iceberg velocity fields and implications for an ice-associated pinniped. Cryosphere 19, 2197–2211.

Law, R., 2009. Ecological information from spatial patterns of plants: insights from point process theory. J. Ecol. 97, 616–628.

Leach, C.B., Williams, P.J., Eisaguirre, J.M., Womble, J.N., Bower, M.R., 2022. Recursive Bayesian computation facilitates adaptive optimal design in ecological studies. Ecology 103, e03573.

Lewis, P.A.W., Shedler, G.S., 1979. Simulation of nonhomogeneous Poisson process by thinning. Nav. Res. Logist. Q. 26, 403–413.

Mathews, E.A., Jemison, L.A., Pendleton, G.W., Blejwas, K.M., Hood, K.E., Raum-Suryan, K.L., 2016. Haul-out patterns and effects of vessel disturbance on harbor seals (phoca vitulina) on glacial ice in tracy arm, Alaska. Fish. Bull. 114 (2), 186–202.

McCaslin, H.M., Feuka, A.B., Hooten, M.B., 2021. Hierarchical computing for hierarchical models in ecology. Methods Ecol. Evol. 12, 245–254.

Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., Tita, G.E., 2011. Self-exciting point process modeling of crime. J. Amer. Statist. Assoc. 106, 100–108.

Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian cox processes. Scand. J. Stat. 25, 451–482.

Montgomery, R.A., Ver Hoef, J.M., Boveng, P.L., 2007. Spatial modeling of haul-out site use by harbor seals in cook inlet, alaska. Mar. Ecol. Prog. Ser. 341, 257–264.

National Park Service, 2010. Glacier Bay National Park and Preserve Foundation Statement. Foundation Statement, National Park Service, Alaska Regional Office; Denver Service Center, April 2010.

National Park Service, 2025. Tidewater glacier viewing and seals. Alaska Park Science, https://www.nps.gov/articles/000/aps-23-1-4-tidewater-glacier-viewing-and-seals.htm.

Neal, R.M., 2011. MCMC using Hamiltonian dynamics. In: Brooks, S., Gelman, A., Jones, G.L., Meng, X.-L. (Eds.), Handbook of Markov Chain Monte Carlo. Chapman & Hall / CRC Press, pp. 113–162.

Northrup, J.M., Hooten, M.B., Anderson, J., Wittemeyer, G., 2013. Practical guidance on characterizing availability in resource selection functions under a use–availability design. Ecology 94, 1456–1463.

Nychka, D., Gerber, F., Li, B., 2023. LatticeKrig: Multi-resolution kriging based on Markov random fields. R package version 9.8.

Ogata, Y., 1998. Statistical models for earthquake occurrences and residual analysis for point processes. J. Amer. Statist. Assoc. 83, 366–379.

Polson, N.G., Scott, J.G., Windle, J., 2013. Bayesian inference for logistic models using Pólya–Gamma latent variables. J. Amer. Statist. Assoc. 108, 1339–1349.

Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G., Warton, D.I., 2015. Point process models for presence-only analysis. Methods Ecol. Evol. 6, 366–379.

Rue, H., Martino, S., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. 71, 319–392.

Särkkä, S., 2013. Bayesian Filtering and Smoothing. Cambridge University Press.

Scharf, H.R., 2025. A strategy to avoid particle depletion in recursive Bayesian inference. arXiv preprint arXiv:2508.01572.

Taylor, I., Kaplan, A., Betancourt, B., 2025. Generative filtering for recursive Bayesian inference with streaming data. J. Comput. Graph. Statist. 34, 1263–1275.

Warton, D.I., Shepherd, L.C., 2010. Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. Ann. Appl. Stat. 4, 1383–1402.

Wikle, C.K., Zammit-Mangion, A., 2023. Statistical deep learning for spatial and spatiotemporal data. Annu. Rev. Stat. Appl. 10, 247–270.

Womble, J.N., Ver Hoef, J.M., Gende, S.M., Mathews, E.A., 2020. Calibrating and adjusting counts of harbor seals in a tidewater glacier fjord to estimate abundance and trends 1992 to 2017. Ecosphere 11, e03111.

Womble, J.N., Williams, P.J., McNabb, R.W., Prakash, A., Gens, R., Sedinger, B.S., Acevedo, C.R., 2021. Harbor seals as sentinels of ice dynamics in tidewater glacier fjords. Front. Mar. Sci. 8, 66–89.