Spatio-temporal Modeling of Zero-Inflated Count: Efficient Posterior Computation and Application to Monitoring Capelin Distribution in the Barents Sea

Shonosuke Sugasawa\textsuperscript{1}, Tomoyuki Nakagawa\textsuperscript{2}, Hiroko Kato Solvang\textsuperscript{3}, Sam Subbey\textsuperscript{4,5} and Salah Alrabeei\textsuperscript{5,4}

\textsuperscript{1}Center for Spatial Information Science, The University of Tokyo
\textsuperscript{2}Department of Information Sciences, Tokyo University of Science
\textsuperscript{3}Marine Mammals Research Group, Institute of Marine Research
\textsuperscript{4}Research Group on Fisheries Dynamics, Institute of Marine Research
\textsuperscript{5}Western Norway University of Applied sciences

Abstract

We consider a model for predicting the spatio-temporal distribution of a marine species based on zero-inflated count observation data that vary continuously over a specified survey region. The model is a mixture of two components; a one-point distribution at the origin and a Poisson distribution with spatio-temporal intensity, where both intensity and mixing proportions are related to some auxiliary information. We develop an efficient posterior computational algorithm for the model using a data augmentation strategy. An attractive feature of the modeling framework is that it accommodates scenarios where the auxiliary information is partially observed, or when the response variable is observed at spatially varying locations over non-uniform time intervals. We present results to show how utilizing the posterior distribution of the auxiliary information facilitates the successful prediction of future spatio-temporal distributions of an example marine species.

Key words: Predictive Gaussian process; Markov Chain Monte Carlo; Poisson distribution; Marine species, Spatio-temporal distribution
1 Introduction

Predicting how marine species distribution may change in response to changes in one or several exogenous factors (including from other species) in the physical environment is attractive for management and conservation. This quest has become even more timely due to recent variability in climate and the potential effect of the variability on biotic and abiotic components of the marine environment ([Dalpadado et al., 2020] [Ingvaldsen and Gjøsæter, 2013]). Scientific fisheries surveys are usually used to monitor the space-time evolution of species distributions over several years. For ecosystems with extensive spatial extent and/or highly variable spatial oceanographic conditions (e.g., partial ice coverage), complete survey coverage may be limited or even impossible. Survey points may be sparse, and the repeated use of the same survey grid (and stations) from year to year may also be impossible. The combination of limited spatial coverage and sparsity of observation points means inference about how spatial distribution of species may have changed with time cannot be based exclusively on survey observations.

The data we focus on in this paper is the estimate of size (in numbers) of Barents Sea capelin at spatial points along acoustic survey transects in August-October (see e.g., [Fall et al., 2018] for a description of the survey). We use data in the period 2014-2019, to derive the spatial distribution in numbers of capelin and, to predict its future distribution. A notable feature of the dataset is that acoustic estimates are zero in many sampling locations; such characteristics and potential spatio-temporal correlations should be taken into account in the statistical modeling of the dataset to achieve stable estimates and predictions.

There are several methods for spatio-temporal modeling of zero-inflated counts. Most of the existing methods are for area-level data (e.g. [Ver Hoef and Jansen, 2007] [Neelon et al., 2016] [Torabi, 2017] [Ghosal et al., 2020]), for which efficient computation algorithms are also available (e.g. [Rue et al., 2009]). However, such techniques cannot be applied in the current situation since our samples are observed at random locations known as point-referenced data ([Banerjee et al., 2014]). For spatio-temporal modeling of point-referenced zero-inflated count data, [Wang et al., 2015] proposed a Bayesian
approach using Gaussian predictive processes, assuming that the sampling locations are the same over the survey period. Since their algorithm for posterior Bayesian inference involves rejection sampling, the computational framework may be inefficient for large number of parameters. On the other hand, Bradley et al. (2018) and Bradley et al. (2020) developed a hierarchical spatio-temporal model using a new distribution theory. Although the appealing feature of the method is that the fitting procedure is computationally efficient, flexibility of the model seems limited and it is not trivial how to account for zero-inflation.

In this work, we propose a spatio-temporal zero-inflated Poisson model, which accommodates variability in the location of capelin survey stations over time. We consider a Bayesian estimation method that enables us to compute point estimates as well as uncertainty measures, such as, credible intervals. To enhance the efficiency of the posterior computation, we use a surrogate distribution for the Poisson distribution, regarded as a reparametrized version of the negative binomial distribution. We then apply the Polya-gamma data augmentation (Polson et al., 2013), and most of full conditional distributions are turned out to be of familiar forms. The proposed modeling approach can be seen as an extension of the spatial zero-inflated negative binomial model (Neelon, 2018). However, unlike in Neelon (2018), our method is not characterized by an unknown over-dispersion parameter. When the Poisson intensity includes random (spatial or time) effects, the marginal distribution of the observed response is over-dispersed. An explicit over-dispersion parameter is, therefore, not necessary in the proposed model. Hence, we employ the negative binomial distribution as an approximated model for the Poisson distribution to develop an efficient posterior computation algorithm. The proposed spatio-temporal zero-inflated modeling is applied to both simulation and capelin datasets. We will demonstrate that the proposed method can provide more stable estimation and prediction performance than a spatio-temporal model without zero-inflation or a zero-inflated Poisson without spatio-temporal correlations.

This paper is organized as follows. In Section 2, we introduce the proposed model and present a detailed description of the posterior computation algorithm. In Section 3, we check the performance of the developed posterior computation algorithm. In Section
the proposed method is applied to the capelin survey data.

2 Spatio-temporal modeling for zero-inflated count data

2.1 Model

Suppose that count data \( y_{it} \) is collected at the \( i \)th location in year \( t \), where \( i = 1, \ldots, N_t \) and \( t = 1, \ldots, T \). Let \( x_{it} \) be a vector of covariates associated with \( y_{it} \). We also suppose that location information \( s_{it} \) (typically two dimensional vector of longitude and latitude) is available. Note that this settings allow that the number of samples and sampling locations could be different over \( t \). To take account of potential zero-inflation structure of \( y_{it} \), we introduce a latent binary indicator \( z_{it} \in \{0, 1\} \), so the distribution of \( y_{it} \) is modeled as

\[
P(y_{it} = 0 | z_{it} = 1) = 1, \quad P(y_{it} = y | z_{it} = 0) = \text{Po}(y; \lambda_{it}), \quad y = 0, 1, \ldots,
\]

where \( \text{Po}(\cdot; \lambda) \) is the probability mass function of the Poisson distribution with intensity \( \lambda \). We model the latent variance \( z_{it} \) and intensity \( \lambda_{it} \) as follows:

\[
\lambda_{it} = \exp(x_{it}^T \beta + u(s_{it}) + v_t), \\
z_{it} = I(x_{it}^T \gamma + \xi(s_{it}) + \eta_t + e_{it} > 0),
\] (1)

where \( \beta \) and \( \gamma \) are unknown regression coefficients, \( u \) and \( \xi \) are spatial (location) effects, \( v_t \) and \( \eta_t \) are time effects, and \( e_{it} \sim N(0, 1) \). Let \( \Psi \) be the collection of the spatial and time effects. Note that under the model (1), the marginal distribution of \( y_{it} \) given the spatial and time effects is expressed as

\[
f(y_{it} | \Psi) = \Phi(g_{it}) \delta_0(y_{it}) + \{1 - \Phi(g_{it})\} \text{Po}(y_{it}; \lambda_{it}),
\]
with \( g_{it} = x_{it}^\top \gamma + \xi(s_{it}) + \eta_t \), thereby the expectation and probability being \( y_{it} = 0 \) (zero-count probability) are, respectively, given by

\[
E[y_{it}|\Psi] = \{1 - \Phi(g_{it})\} \lambda_{it},
\]

\[
P(y_{it} = 0|\Psi) = \Phi(g_{it}) + \{1 - \Phi(g_{it})\} \exp(-\lambda_{it}).
\]

Let \( u \) and \( \xi \) be a vector of \( u(s_{it}) \) and \( \xi(s_{it}) \), respectively. Since the dimension of \( u \) and \( \xi \) can be large in practice (the dimension is 2647 in our example in Section 4), the direct use of Gaussian processes would be computationally very demanding. To avoid the difficulty, we model the two spatial effects by the predictive process (Banerjee et al., 2008), that is,

\[
\begin{align*}
\mu_u &\sim N(0, \tau_u^{-1} H_u) \\
\mu_\xi &\sim N(0, \tau_\xi^{-1} H_\xi)
\end{align*}
\]

where \( D(s_{it}; b) = H^{-1} V(s_{it}; b) \) with \( V(s_{it}; b) = (\nu_h(s, s_1^i), \ldots, \nu_h(s, s_M^i)) \in \mathbb{R}^M \) and \( (H)_{kl} = \nu_h(s_k^i, s_l^i) \). \( \nu_h(s_1, s_2) \) is a correlation function with bandwidth \( b \), and \( s_1^i, \ldots, s_M^i \) be a set of knots. Here \( \mu_u \) and \( \mu_\xi \) are the Gaussian process for knots, namely, \( \mu_u \sim N(0, \tau_u^{-1} H_u) \) and \( \mu_\xi \sim N(0, \tau_\xi^{-1} H_\xi) \) with unknown precision parameters, \( \tau_u \) and \( \tau_\xi \), and bandwidth \( h_u \) and \( h_\xi \). We simply consider the exponential correlation function, \( \nu_h(s_1, s_2) = \exp\{- (s_1 - s_2)^2 / h^2\} \), where \( h \) is a bandwidth controlling the strength of spatial correlation. The number and locations of knots and the choice of the bandwidth will be discussed later. Regarding the time effects \( v_t \) and \( \eta_t \), we consider the random-walk process, namely, \( v_t|v_{t-1} \sim N(v_{t-1}, \sigma_v^2) \) and \( \eta_t|\eta_{t-1} \sim N(\eta_{t-1}, \sigma_\eta^2) \) with \( v_1 = \eta_1 = 0 \) and unknown variances \( \sigma_v^2 \) and \( \sigma_\eta^2 \). Note that the initial conditions, \( v_1 = 0 \) and \( \eta_1 = 0 \) are necessary for identifiability of the intercept terms in \( x_{it}^\top \beta \) and \( x_{it}^\top \gamma \), respectively.

### 2.2 Efficient posterior computation

The likelihood function of the latent and model parameters is given by

\[
\prod_{t=1}^{T} \prod_{i=1}^{N_t} \left\{ I(g_{it} > 0) \delta_0(y_{it}) \right\}^{z_{it}} \left\{ I(g_{it} \leq 0) \text{Po}(y_{it}; \lambda_{it}) \right\}^{1-z_{it}} \times \exp \left\{ -\frac{1}{2} \left( g_{it} - x_{it}^\top \gamma - \xi(s_{it}) - \eta_t \right)^2 \right\}
\]
Note that the Gaussian distribution for $u(s_{it})$ and $v_t$ are not conjugate under the Poisson likelihood, so the full conditional posterior distributions of these parameters are not familiar forms, which might make the sampling steps complicated and inefficient. We here solve the problem by using an approximate Poisson likelihood using the negative binomial distribution. To this end, we introduce an additional latent variable $\varepsilon_{it} \sim \text{Ga}(\delta, \delta)$ in the Poisson distribution, that is, $y_{it}|(z_{it} = 0, \varepsilon_{it}) \sim \text{Po}(\varepsilon_{it} \lambda_{it})$. Note that $E[\varepsilon_{it}] = 1$ and $\text{Var}(\varepsilon_{it}) = 1/\delta$, thereby $\varepsilon_{it}$ degenerates at $\varepsilon_{it} = 1$ when $\delta \to \infty$. In other words, the augmented model reduces to the Poisson model under $\delta \to \infty$. Then, the likelihood of $\lambda_{it}$ under the model integrating $\varepsilon_{it}$ is given by

$$
\text{Po}(y_{it}; \lambda_{it}) \approx \frac{\Gamma(y_{it} + \delta)}{\Gamma(\delta) y_{it}!} \left(\frac{\delta^{-1} \lambda_{it}}{\delta^{-1} \lambda_{it} + 1}\right)^{y_{it} + \delta}.
$$

(2)

We employ the above likelihood function with large $\delta > 0$ as an approximated likelihood of the Poisson likelihood. The main advantage of the approximated likelihood (2) is that the likelihood can be further augmented by the Polya-gamma random variable (Polson et al., 2013), that is, it holds that

$$
\frac{(\delta^{-1} \lambda_{it})^{y_{it}}}{(\delta^{-1} \lambda_{it} + 1)^{y_{it} + \delta}} = 2^{-(y_{it} + \delta)} e^{\kappa_{it} \psi_{it}} \int_0^\infty \exp \left(-\frac{1}{2} \omega_{it} \psi_{it}^2\right) p_{PG}(\omega_{it}; y_{it} + \delta, 0) d\omega_{it},
$$

where $\kappa_{it} = (y_{it} - \delta)/2$, $\psi_{it} = \log(\lambda_{it}/\delta) = x_{it}^\top \beta + u(s_{it}) + v_t - \log(\delta)$, and $p_{PG}(\cdot; b, c)$ denotes the density function of the Polya-gamma distribution. Here $\omega_{it}$ is an additional latent parameter and the above integral expression shows that the conditional distribution of $\psi_{it}$ given $\omega_{it}$ is Gaussian, which leads to a tractable posterior computation algorithm.

The joint posterior distribution of the latent variables and model parameters is given by

$$
\pi(\theta) \prod_{t=1}^T \prod_{i=1}^{N_t} \left\{I(g_{it} > 0)\delta_0(y_{it})\right\}^{z_{it}} \left\{I(g_{it} \leq 0) \exp \left(\kappa_{it} \psi_{it} - \frac{1}{2} \omega_{it} \psi_{it}^2\right) p_{PG}(\omega_{it}; y_{it} + \delta, 0)\right\}^{1 - z_{it}}
\times \phi(g_{it}; x_{it}^\top \gamma + \xi(s_{it}) + \eta, 1) \left\{\prod_{t=1}^T \phi(v_t; v_{t-1}, \sigma_v^2) \phi(\eta_t; \eta_{t-1}, \sigma_\eta^2)\right\}
\times \phi(\mu_u; 0, \tau_u^{-1} H_u) \phi(\mu_\xi; 0, \tau_\xi^{-1} H_\xi),
$$

6
where \( \pi(\theta) \) is a prior distribution for \( \theta \). For the prior distributions, we adopt \( \beta \sim N(0, D_\beta) \), \( \gamma \sim N(0, D_\gamma) \), \( \tau_u \sim \text{Ga}(d_{\tau_u}, d_{\tau_u}) \), \( \tau_\xi \sim \text{Ga}(d_{\tau_\xi}, d_{\tau_\xi}) \), \( \sigma_v^2 \sim \text{IG}(d_{\sigma_v}, d_{\sigma_v}) \), and \( \sigma_{\eta}^2 \sim \text{IG}(d_{\sigma_{\eta}}, d_{\sigma_{\eta}}) \) as default choices, which will be shown to be conditionally conjugate. We also introduce discrete uniform prior distributions for two bandwidth parameters, \( h_u \) and \( h_\xi \), that is, we put equal probabilities on the pre-specified candidates of knots, \( \{h_1, \ldots, h_L\} \). The Markov Chain Monte Carlo sampling algorithm is provided as follows.

- **Sampling of \( \omega_{it} \):** Given \( z_{it} = 0 \), the full conditional distribution of \( \omega_{it} \) is \( PG(y_{it} + \delta, \psi_{it}) \), where \( \psi_{it} = x_{it}^\top \beta + u(s_{it}) + v_t - \log \delta \). Since the Polya-gamma distribution \( PG(b, c) \) can be precisely approximated by a normal distribution under large \( b \) (Glynn et al., 2019) and \( \delta \) is set to a large value such as \( \delta = 10^5 \), we use \( \{\tilde{b}_{it}, \tilde{c}_{it}\} \) as the accurate proxy of the full conditional distribution, where

\[
\tilde{b}_{it} = \frac{y_{it} + \delta}{2\psi_{it}} \tanh \left( \frac{\psi_{it}}{2} \right), \quad \tilde{c}_{it} = \frac{y_{it} + \delta}{4\psi_{it}^3} \operatorname{sech}^2 \left( \frac{\psi_{it}}{2} \right) \left( \operatorname{sinh}(\psi_{it}) - \psi_{it} \right).
\]

- **Sampling of \( \beta \):** The full conditional distribution of \( \beta \) is \( N(\tilde{A}_\beta \tilde{B}_\beta, \tilde{A}_\beta) \), where

\[
\tilde{A}_\beta = \left( \sum_{t=1}^{T} \sum_{i=1}^{N_t} (1 - z_{it}) \omega_{it} x_{it} x_{it}^\top + D_\beta^{-1} \right)^{-1}, \quad \tilde{B}_\beta = \sum_{t=1}^{T} \sum_{i=1}^{N_t} (1 - z_{it}) \kappa_{it} - \omega_{it} (u(s_{it}) + v_t - \log \delta).
\]

- **Sampling of \( \gamma \):** The full conditional distribution of \( \gamma \) is \( N(\tilde{A}_\gamma \tilde{B}_\gamma, \tilde{A}_\gamma) \), where

\[
\tilde{A}_\gamma = \left( \sum_{t=1}^{T} \sum_{i=1}^{N_t} x_{it} x_{it}^\top + D_\gamma^{-1} \right)^{-1}, \quad \tilde{B}_\gamma = \sum_{t=1}^{T} \sum_{i=1}^{N_t} x_{it} (g_{it} - \xi(s_{it}) - \eta_t).
\]

- **Sampling of \( v_t \):** For \( t = 2, \ldots, T - 1 \), the full conditional distribution of \( v_t \) is \( N(\tilde{A}_v \tilde{B}_v, \tilde{A}_v) \), where

\[
\tilde{A}_v = \left( \sum_{i=1}^{N_t} (1 - z_{it}) \omega_{it} + \frac{2}{\sigma_v^2} \right)^{-1}, \quad \tilde{B}_v = \sum_{i=1}^{N_t} (1 - z_{it}) \left( \kappa_{it} - \omega_{it} (x_{it}^\top \beta + u(s_{it}) - \log \delta) \right) + \frac{v_{t-1} + v_{t+1}}{\sigma_v^2}.
\]
For $t = T$, $\tilde{A}_v$ is changed to $(\sum_{i=1}^{N_t} \omega_{it} + 1/\sigma^2_v)^{-1}$ and the same form of $\tilde{B}_v$ is used with $v_{T+1} = 0$.

- **Sampling of $\eta_t$:** The full conditional distribution of $\eta_t$ is $N(\tilde{A}_\eta \tilde{B}_\eta, \tilde{A}_\eta)$, where

$$
\tilde{A}_\eta = \left( N_t + \frac{2}{\sigma^2_{\eta}} \right)^{-1}, \quad \tilde{B}_\eta = \sum_{i=1}^{N_t} \left( g_{it} - x_{it}^\top \gamma + \xi(s_{it}) \right) - \frac{\eta_{t-1} + \eta_{t+1}}{\sigma^2_{\eta}}.
$$

For $t = T$, $\tilde{A}_\eta$ is changed to $(N_t + 1/\sigma^2_{\eta})^{-1}$ and the same form of $\tilde{B}_\eta$ is used with $\eta_{T+1} = 0$.

- **Sampling of $\mu_u$ and $u$:** The full conditional distribution of $\mu_u$ is $N(\tilde{A}_u \tilde{B}_u, \tilde{A}_u)$, where

$$
\tilde{A}_u = \left\{ \sum_{t=1}^{T} \sum_{i=1}^{N_t} (1 - z_{it}) \omega_{it} D(s_{it}) D(s_{it})^\top + \tau_u H^{-1} \right\}^{-1},
\tilde{B}_u = \sum_{t=1}^{T} \sum_{i=1}^{N_t} (1 - z_{it}) D(s_{it}) \left\{ \kappa_{it} - \omega_{it} (x_{it}^\top \beta + v_t - \log \delta) \right\}.
$$

Then the posterior sample of $u$ is obtained as $u(s_{it}) = D(s_{it})^\top \mu_u$.

- **Sampling of $\mu_\xi$ and $\xi$:** The full conditional distribution of $\mu_\xi$ is $N(\tilde{A}_\xi \tilde{B}_\xi, \tilde{A}_\xi)$, where

$$
\tilde{A}_\xi = \left\{ \sum_{t=1}^{T} \sum_{i=1}^{N_t} D(s_{it}) D(s_{it})^\top + \tau_\xi H^{-1} \right\}^{-1},
\tilde{B}_\xi = \sum_{t=1}^{T} \sum_{i=1}^{N_t} D(s_{it}) (g_{it} - z_{it}^\top \gamma - \eta_t).
$$

Then the posterior sample of $\xi$ is obtained as $\xi(s_{it}) = D(s_{it})^\top \mu_\xi$.

- **Sampling of $h_u$ and $h_\xi$:** The full conditional distribution of $h_u$ (where $\omega_{it}$ is marginalized out) is a discrete distribution on $\{h_1, \ldots, h_L\}$ whose probability mass function is proportional to

$$
\pi(h_u|H_u(h_u))^{-1/2} \exp \left\{ -\frac{1}{2} \tau_u \mu_u^\top H_u(h_u)^{-1} \mu_u \right\} \prod_{t=1}^{T} \prod_{i=1}^{N_t} \left\{ (\delta^{-1} \lambda_{it}) y_{it} (\delta^{-1} \lambda_{it} + 1)^{y_{it} + \delta} \right\}^{1 - z_{it}},
$$

Then the posterior sample of $h_u(s_{it}) = D(s_{it})^\top \mu_u$. 

8
where \( h_u \in \{h_1, \ldots, h_L\} \). The full conditional distribution of \( h_\xi \) is proportional to
\[
\pi(h_\xi | H_\xi(h_\xi))^{-1/2} \exp \left\{ -\frac{1}{2} \tau_\xi \mu_\xi^\top H_\xi(h_\xi)^{-1} \mu_\xi + \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N_t} (g_{it} - x_{it}^\top \gamma - D(s_{it})^\top \mu_\xi - \eta_t)^2 \right\},
\]
where \( h_\xi \in \{h_1, \ldots, h_L\} \).

- **Sampling of** \( \tau_u \) **and** \( \tau_\xi \): The full conditional distributions of \( \tau_u \) and \( \tau_\xi \) are
  \( \text{Ga}(\delta_{\tau_u} + \sum_{t=1}^{T} N_t, \delta_{\tau_u} + \mu_u^\top H(h_u)^{-1} \mu_u / 2) \) and \( \text{Ga}(\delta_{\tau_\xi} + \sum_{t=1}^{T} N_t, \delta_{\tau_\xi} + \mu_\xi^\top H(h_\xi)^{-1} \mu_\xi / 2) \), respectively.

- **Sampling of** \( \sigma_v^2 \) **and** \( \sigma_\eta^2 \): The full conditional distributions of \( \sigma_v^2 \) and \( \sigma_\eta^2 \) are
  \( \text{IG}(\delta_{\sigma_v} + T/2, \delta_{\sigma_v} + \sum_{t=1}^{T} (v_t - v_{t-1})^2 / 2) \) and \( \text{IG}(\delta_{\sigma_\eta} + T/2, \delta_{\sigma_\eta} + \sum_{t=1}^{T} (\eta_t - \eta_{t-1})^2 / 2) \), respectively.

- **Sampling of** \( g_{it} \): The full conditional distribution of \( g_{it} \) is \( N_+ (x_{it}^\top \gamma + \xi(s_{it}) + \eta_t, 1) \) if \( z_{it} = 1 \) and \( N_- (x_{it}^\top \gamma + \xi(s_{it}) + \eta_t, 1) \) if \( z_{it} = 0 \).

- **Sampling of** \( z_{it} \): The full conditional distribution of \( z_{it} \) is the Bernoulli distribution with success probability \( 1/(1 + d_{it}) \), where
  \[
d_{it} = \frac{1 - \Phi(x_{it}^\top \gamma + \xi(s_{it}) + \eta_t))}{\Phi(x_{it}^\top \gamma + \xi(s_{it}) + \eta_t) \delta_0(y_{it})}.\]

It should be noted that all the above sampling steps are simply generated from some familiar distributions. Consequently, no rejection steps are required in generating posterior samples from the full conditional distributions, which would prevent high serial correlations of the posterior samples.

To complete description of the proposed method, we specify the tuning parameters. First, we recommend the number of knots, \( M \), should be set to a moderate value to ensure low computational cost. For specification of the locations of knots, we suggest using the \( k \)-means clustering algorithm with \( M \) clusters, and then set \( s_1^*, \ldots, s_M^* \) to be the centers of the \( M \) clusters.
3 Simulation study

We investigate the estimation performance of the proposed method using simulated data that is generated according to the model given by (1). We set $T = 6$ (the number of time periods) and $N_i (= N) = 400$ (the number of samples in each time period), where the setting is similar to our application in Section 4. The location information $s_{it}$ was generated from the uniform distribution on $[-2, 2] \times [-2, 2]$, and then two spatial effects $u(s_{it})$ and $\xi(s_{it})$ were generated from the Gaussian processes with covariance functions given by $0.5 \exp(-\|s_{it} - s'_{i't'}\|^2/h^2)$ with $h = 0.5$ for $u(s_{it})$ and $h = 0.9$ for $\xi(s_{it})$. For the time effect, we set $(v_2, \ldots, v_6) = (0.4, 0.8, 1.2, 1.6, 2.0)$ and $(\eta_2, \ldots, \eta_6) = (0.5, 1, 1, 0.5, 0)$. In this study, we use a single covariate $x_{it}$ generated from $N(0, (0.5)^2)$, and the fixed regression coefficients are set as $\beta = (0.5, 0.5)$ and $\gamma = (-1.5, -1)$. Under the settings, the count response $y_{it}$ was generated, and the ratio of zero-count was 33.6%.

In application of our proposed method, we set $M = 100$ (the number of knots in predictive processes), $\delta = 10000$ (tuning parameter for Poisson likelihood approximation) and employed default priors for the unknown parameters. We use the Gibbs sampling procedure in Section 2.2 to draw 40000 posterior samples, after discarding the first 5000 as burn-in samples. The true values, posterior means and 95% credible intervals of fixed regression coefficients and time effects are presented in Table 1. We observe that the posterior means of the parameters are close to the true values and credible intervals seem to have reasonable lengths such that all the credible intervals cover the true values.

In Figure 1, we present the true values and posterior means of the two spatial effects, where the $M = 100$ knots for the predictive processes are shown by cross-marks. It is worth noting that the spatial characteristics are adequately captured by the estimation strategy even though the proposed method adopts predictive processes that are not exactly equal to the full Gaussian processes used in the data generation. Finally, in Figure 2, we reported the posterior means and 95% credible intervals of the average counts $E[y_{it}]$ and zero-count probability $P(y_{it} = 0)$. It can be seen that the posterior means are accurate estimates of the true values, and the 95% credible intervals provide...
reasonable inference results. To be more precise, the empirical coverage rates for average count and zero-count probability are 94.6% and 91.9%, respectively, both of which are close to the nominal level.

Table 1: Posterior summary of the static regression coefficients and time effects in the simulation study

|                | Intensity part | Zero-inflation part |
|----------------|----------------|---------------------|
|                | true PM CI-l CI-u | true PM CI-l CI-u   |
| $\beta_0$     | 0.5  0.502  0.335  0.683 | $\gamma_0$ -1.5 -1.147 -1.800 -0.616 |
| $\beta_1$     | 0.5  0.514  0.471  0.556 | $\gamma_1$ -1.0 -1.059 -1.216 -0.903 |
| $\nu_2$       | 0.3  0.311  0.207  0.410 | $\eta_2$  0.4  0.365  0.074  0.663 |
| $\nu_3$       | 0.6  0.620  0.525  0.717 | $\eta_3$  0.8  0.704  0.429  0.996 |
| $\nu_4$       | 0.9  0.988  0.897  1.077 | $\eta_4$  0.8  0.572  0.302  0.855 |
| $\nu_5$       | 1.2  1.246  1.161  1.332 | $\eta_5$  0.4  0.202 -0.076  0.494 |
| $\nu_6$       | 1.5  1.503  1.421  1.584 | $\eta_6$  0.0 -0.065 -0.347  0.226 |

4 Application: monitoring capelin distribution in Barents Sea

4.1 Data description

We apply the proposed method to the Barents Sea capelin survey data in August-October in the years 2014-2019. Maturation in the Barents Sea capelin is assumed to be length-dependent, and fish of at least 14 cm is considered to be mature [Jourdain et al., 2021]. Our data is derived from approximately 400 sampling stations per year, with information about number of matured/immature capelin per station. Both the locations of the sampling stations and the survey dates vary from year to year.

As auxiliary information, we use sea surface temperature (SST) at 20m depths as the averaged values of SSTs at depth less than 20m at each sampling location. We defined 10m SST (similarly for 20m SST). We transformed the date information to cumulative days from August 1st. In Figure 3, we show a scatter plot of observed counts against cumulative days and 20m SST. It can be seen that the observed count tends to be high at 40 $\sim$ 60 cumulative days (i.e. the sampled date is in September), and the effect of cumulative days seems nonlinear. Furthermore, it is observed that the observed counts and SST seem positively correlated.
Figure 1: The true values and posterior means of the two spatial effects, $u(s_{it})$ and $\xi(s_{it})$, in the simulation study. The locations of knots in the predictive processes are shown by ‘×’.

4.2 Model with nonparametric component

Let $C_{it}$ be the cumulative day and $SST_{it}$ be the 10m or 20m SST at the $i$th location in the $t$th year. We consider the following model for the Poisson intensity and zero-inflation indicator:

$$
\log \lambda_{it} = f_1(C_{it}) + \alpha_1 SST_{it} + u(s_{it}) + v_t,
$$

$$
z_{it} = I\left(f_2(C_{it}) + \alpha_2 SST_{it} + \xi(s_{it}) + \eta_t + e_{it} > 0\right),
$$

(3)
Figure 2: The posterior means with 95% credible interval bars of the intensity (upper) and zero-inflation probability (lower) against the true values.

where \( f_1(\cdot) \) and \( f_2(\cdot) \) are completely unknown functions of \( C_{it} \), \( \alpha_1 \) and \( \alpha_2 \) are regression coefficients of SST. We employ the P-spline method for estimating \( f_k(\cdot) \) (\( k = 1, 2 \)), that is, \( f_k(\cdot) \) is estimated by the following form:

\[
f_k(x) = a_{k0} + a_{k1}x + \cdots + a_{kq}x^q + \sum_{\ell=1}^{K} a_{k,q+\ell}(x - \kappa_\ell)^q_+, \]

where \((x - c)_+ = \max(x - c, 0)\), \(\kappa_1, \ldots, \kappa_K\) are knots and \(K\) are the number of knots. In this analysis, we first scaled \(C_{it}\) to lie on the interval \([0, 1]\), and we set \(q = 2\), \(K = 9\) and \(\kappa_\ell = \ell/10\). To avoid overfitting of the above model, we assume that \(a_{k,q+\ell} \sim N(0, \tau_{P_k}^{-1})\) with unknown precision parameter \(\tau_{P_k}\).
We let $\beta = (\alpha_1, a_{10}, \ldots, a_{1,q+K})^\top$, $\gamma = (\alpha_2, a_{20}, \ldots, a_{2,q+K})^\top$, and

$$
x_{it} = (\text{SST}_{it}, 1, C_{it}, \ldots, C_{it}^{q}, (C_{it} - \kappa_1)^{q+\ell}, \ldots, (C_{it} - \kappa_K)^{q+\ell})^\top.
$$

The model [3] can then be written in the original form [1]. Thus, we can employ almost the same posterior computation algorithm as in Section 2.2. However, sampling steps for $\beta$ and $\gamma$ are slightly different due to the shrinkage priors for $\alpha_{k,q+\ell}$. Under assumptions of non-informative priors for the rest of the parameters in $\beta$ and $\gamma$ (denoted by $N(0, D_\beta^*)$ and $N(0, D_\gamma^*)$, respectively) the sampling procedures replaces $D_\beta$ and $D_\gamma$ respectively, with blockdiag($D_\beta^*, \tau_{p1}^{-1}I_K$) and blockdiag($D_\gamma^*, \tau_{p2}^{-1}I_K$). With gamma priors $\text{Ga}(d_{\tau_{p1}}, d_{\tau_{p1}})$ for $\tau_{p1}$, the full conditional distribution of $\tau_{pk}$ is given by $\text{Ga}(d_{\tau_{p1}} + K/2, d_{\tau_{p1}} + \sum_{\ell=1}^{K} a_{k,q+\ell}/2)$ for $k = 1, 2$, which is incorporated into the sampling algorithm in Section 2.2.

For model comparison, we also applied two sub-models of [1]; a spatio-temporal Poisson model (STP) without zero-inflation structure, and a static zero-inflated Poisson model (ZIP) without spatio-temporal effects. Their MCMC algorithms can be easily obtained by slight modification of the algorithm given in Section 2.2.
models, we generated 40000 posterior samples after discarding the first 5000 samples. We then computed posterior predictive loss (PPL; Gelfand and Ghosh 1998) for each model, which are reported in Table 2. The results show that the proposed STZIP is the best among the three models, indicating that both spatio-temporal effects and zero-inflation structures can improve overall fitting to the dataset. The posterior means (PM) and point-wise 95% credible intervals of the regression coefficient of SST are also reported in Table 2. The effect of 10m SST is not so significant while that of 20m SST seems significantly positive. The latter result would be more trustworthy since the PPL with 20m SST is smaller. In the discussion that follows we focus on the results with the 20m SST dataset.

The posterior means and posterior standard deviations of the time effects are reported in Table 3. There seems no significant effects of the time effects on zero-inflation components, but time effects in the Poisson intensity seems strongly significant in 2016, 2017 and 2019. We next visualize the nonparametric effect of the cumulative days. By fixing the spatial and time effects to 0 and setting $\alpha_1 = \alpha_2 = 0$ in the proposed model, the expected count and zero-count probability are given by $\{1 - \Phi(f_2(C))\} \exp\{f_1(C)\}$ and $\Phi(f_2(C)) + \{1 - \Phi(f_2(C))\} \exp(-\exp\{f_1(C)\})$, respectively, as functions of the cumulative days $C$. The posterior means and point-wise 95% credible intervals of these quantities are shown in Figure 4. The estimated effect in average counts seems consistent with the scatter plots of observed count given in Figure 3, namely, observed counts tend to be large around $C = 60$. On the other hand, when $C > 70$, the average count decreases toward 0 and zero-count probability rapidly increases, which is also consistent with Figure 3 in that most of observed counts are 0 when $C > 70$.

We computed posterior means of two spatial effects in entire regions (including locations without observations) using posterior samples of $\mu_u$ and $\mu_\xi$. Then, using the time effect in 2019, we computed posterior means of average counts (expected number of capelin) as well as zero-count probability (probability that there is no capelin) at 6 dates, which are presented in Figures 5 and 6, respectively. Figure 5 indicates that some “hot-spots” existed in the eastern part of the survey area until mid-October, which disappeared at the end of October, indicating that the average count is very small across
the entire region. Figure 6 shows that there is no capelin in the south-western region since the zero-count probability is constantly almost 1. It is also observed that the zero-count probability is high across the entire region at the end of October, which is consistent with the results in Figure 5.

Table 2: Posterior mean (PM) and 95% credible intervals of regression coefficients of and posterior predictive loss (PPL) of the three models.

| regression coefficient | PPL |
|------------------------|-----|
| SST                    |     |
| PM                     | 0.009 |
| CI-l                   | -0.014 |
| CI-u                   | 0.032 |
| STZIP                   | 1793 |
| STP                     | 2080 |
| ZIP                     | 1827 |
| 10m                    |     |
| PM                     | 0.066 |
| CI-l                   | 0.046 |
| CI-u                   | 0.087 |
| STZIP                   | 1763 |
| STP                     | 2075 |
| ZIP                     | 1828 |

Table 3: Posterior means (PM) and standard deviation (SD) of the time effects.

| Year | Poisson intensity | Zero-inflation |
|------|-------------------|----------------|
|      | PM                | SD             |
|      | PM                | SD             |
| 2015 | -0.031            | 0.023          |
|      | 0.085             | 0.094          |
| 2016 | -0.684            | 0.032          |
|      | 0.190             | 0.107          |
| 2017 | -0.096            | 0.025          |
|      | -0.050            | 0.097          |
| 2018 | -0.038            | 0.028          |
|      | -0.051            | 0.110          |
| 2019 | 0.107             | 0.024          |
|      | 0.124             | 0.096          |

Figure 4: Point-wise posterior means and 95% credible intervals of the nonparametric effect of cumulative days, $f_1(\cdot)$ and $f_2(\cdot)$. 

Figure 5: Spatial distributions of posterior means of average counts (expected number of capelin) at 6 dates. Note that the average count is truncated at 50.
Figure 6: Spatial distributions of posterior means of zero-count probability (probability that there is no capelin) at 6 dates.
4.3 Validation

Finally, we compared the prediction performance of the STZIP and its two sub-models, STP and ZIP. To this end, we left the dataset in 2019 as a validation set (or test data), and the three models with 20m SST are estimated based on the data from 2014 to 2018. By generating 20000 posterior samples after discarding the first 5000 samples, we computed point prediction of the data in 2019. The performance of the prediction is evaluated using the mean absolute error (MAE) and two-types of mean absolute percentage error (MAPE) given by

$$\text{MAE} = \frac{1}{m} \sum_{j=1}^{m} |y_j - \hat{\lambda}_j|, \quad \text{MAPE}_1 = \frac{1}{m} \sum_{j=1}^{m} \frac{|y_j - \hat{\lambda}_j|}{y_j + 1},$$

$$\text{MAPE}_2 = \left( \sum_{j=1}^{m} I(y_j > 0) \right)^{-1} \sum_{j=1: y_j > 0}^{m} \frac{|y_j - \hat{\lambda}_j|}{y_j},$$

where $m$ is the number of test samples, i.e. $m = 499$. The results are reported in Table 4 which shows that the STZIP model provides considerably better prediction performance than the others regardless of the three different types of errors.

Table 4: Three prediction errors of the three models.

|       | STZIP | STP  | ZIP  |
|-------|-------|------|------|
| MAE   | 25.5  | 31.2 | 29.9 |
| MAPE$_1$ | 3.95 | 10.3 | 9.10 |
| MAPE$_2$ | 2.33 | 6.27 | 5.69 |

5 Concluding remarks

Developing a model to capture, e.g., space-time variation in species distribution is challenging. This is because several factors (including, but not limited to logistics, financial, human, and environmental) limit the applicability of the same survey grid (transect) from year-to-year, and the time at which repeat grid points (survey stations) are sampled. This challenge is exacerbated by the fact that data from one or several sampling stations are usually zero-inflated.
This paper has presented an algorithm for effective prediction of the spatio-temporal distribution of a marine species using zero-inflated (species) count observation data and auxiliary environmental information. The algorithm has been validated with results (based on simulated and empirical data), which demonstrate the efficacy of the modeling framework.

One major significance of this paper is that it presents a computationally effective framework that facilitates the study of how changes in space-time species distributions may be linked to variability in environmental factors. The results in the paper are also timely, as establishing a link between spatio-temporal species dynamics and changes (e.g., temperature rise) in the physical environment, is an active field of research in the marine sciences.

Acknowledgement

S. Sugasawa is supported by Japan Society for Promotion of Science (KAKENHI) grant number 18H03628, 20H00080, and 21H00699. T. Nakagawa is supported by Japan Society for Promotion of Science (KAKENHI) grant number 19K14597 and 21H00699. S. Subbey and H.K. Solvang have been supported by grants 84126, Management Strategy for the Barents Sea, and GA19-NOR-081 from the Sasakawa foundation (Norway).

References

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2014). Hierarchical modeling and analysis for spatial data. CRC press.

Banerjee, S., A. E. Gelfand, A. O. Finley, and H. Sang (2008). Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70(4), 825–848.

Bradley, J. R., S. H. Holan, and C. K. Wikle (2018). Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion). Bayesian Analysis 13, 253–310.
Bradley, J. R., S. H. Holan, and C. K. Wikle (2020). Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. *Journal of the American Statistical Association* 115(532), 2037–2052.

Dalpadado, P., K. R. Arrigo, G. L. van Dijken, H. R. Skjoldal, E. Bagøien, A. V. Dolgov, I. P. Prokopchuk, and E. Sperfeld (2020). Climate effects on temporal and spatial dynamics of phytoplankton and zooplankton in the Barents Sea. *Progress in Oceanography* 185, 102320.

Fall, J., L. Ciannelli, G. Skaret, and E. Johannesen (2018). Seasonal dynamics of spatial distributions and overlap between northeast arctic cod (gadus morhua) and capelin (mallotus villosus) in the barents sea. *PloS one* 13(10), e0205921.

Gelfand, A. E. and S. K. Ghosh (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* 85(1), 1–11.

Ghosal, S., T. S. Lau, J. Gaskins, and M. Kong (2020). A hierarchical mixed effect hurdle model for spatiotemporal count data and its application to identifying factors impacting health professional shortages. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 69(5), 1121–1144.

Glynn, C., S. T. Tokdar, B. Howard, and B. L. Banks (2019). Bayesian analysis of dynamic linear topic models. *Bayesian Analysis* 14, 53–80.

Ingvaldsen, R. B. and H. Gjøsæter (2013). Responses in spatial distribution of Barents Sea capelin to changes in stock size, ocean temperature and ice cover. *Marine Biology Research* 9(9), 867–877.

Jourdain, N., E. Fuglebakk, and S. Subbey (2021). Maturation in the barents sea capelin—contrasting length-and gonad-based metrics. *Fisheries Research* 237, 105880.

Neelon, B. (2018). Bayesian zero-inflated negative binomial regression based on polya-gamma mixtures. *Bayesian Analysis*, to appear.
Neelon, B., H. H. Chang, Q. Ling, and N. S. Hastings (2016). Spatiotemporal hurdle models for zero-inflated count data: Exploring trends in emergency department visits. *Statistical Methods in Medical Research* 6, 2558–2576.

Polson, N. G., J. G. Scott, and J. S. Windle (2013). Bayesian inference for logistic models using polya-gamma latent variables. *Journal of the American Statistical Association* 108, 1339–1349.

Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* 71(2), 319–392.

Torabi, M. (2017). Zero-inflated spatio-temporal models for disease mapping. *Biometrical Journal* 59(3), 430–444.

Ver Hoef, J. M. and J. K. Jansen (2007). Space—time zero-inflated count models of harbor seals. *Environmetrics: The official journal of the International Environmetrics Society* 18(7), 697–712.

Wang, X., M. Chen, R. C. Kuo, and D. K. Dey (2015). Bayesian spatial-temporal modeling of ecological zero-inflated count data. *Statistica Sinica* 25, 189–204.