Space-time multi-level modeling for zooplankton abundance employing double data fusion and calibration

Jorge Castillo-Mateo[1] · Alan E. Gelfand[2] · Christine A. Hudak[3] · Charles A. Mayo[3] · Robert S. Schick[4]

Abstract

An important objective for marine biologists is to forecast the distribution and abundance of planktivorous marine predators. To do so, it is critically important to understand the spatiotemporal dynamics of their prey. Here, the prey we study are zooplankton and we build a novel space-time hierarchical fusion model to describe the distribution and abundance of zooplankton species in Cape Cod Bay (CCB), MA, USA. The data were collected irregularly in space and time at sites within the first half of the year over a 17 year period, using two different sampling methods. We focus on *sea surface* zooplankton abundance and incorporate sea surface temperature as a primary driver, also collected with two different sampling methods. So, with two sources for each, we observe true abundance or true sea surface temperature with measurement error. To account for such error, we apply calibrations to align the data sources and use the fusion model to develop a prediction of daily spatial zooplankton abundance surfaces throughout CCB. To infer average abundance on a given day within a given year in CCB, we present a marginalization of the zooplankton abundance surface. We extend the inference to consider abundance averaged to a bi-weekly or annual scale as well as to make an annual comparison of abundance.

Keywords Bayesian melding · Gaussian process · Hierarchical model · Kriging · Markov chain Monte Carlo · Sea surface temperature

✉ Jorge Castillo-Mateo
  jorgecm@unizar.es

1 Department of Statistical Methods, University of Zaragoza, Zaragoza, Spain

2 Department of Statistical Science, Duke University, Durham, NC, USA

3 Center for Coastal Studies, Provincetown, MA, USA

4 Nicholas School of the Environment, Duke University, Durham, NC, USA

1 Introduction

Marine habitats are complex and can be difficult to systematically describe due in part to the difficulty of systematic sampling across relevant space and time scales (Everett et al. 2017). With regard to modeling resource distribution (Everett et al. 2017), while data are available from different sources, the use of different collection methods with different temporal and spatial scales confounds the process. Critical to the management of marine habitats are the distribution, abundance, and composition of marine zooplankton, a community that is central to the energy flow through the system from primary producers to higher trophic levels such as invertebrates, fish, and baleen whales (Lenz 2000; Everett et al. 2017). The diverse zooplankton groups have multiple functional roles in the marine food web including controlling the biomass by grazers, nutrient recycling via excrement, and vertical transport of material either by sinking, direct transport, or diel vertical migration (Andersen et al. 1993; Richardson 2008). Therefore, modeling the abundance and dynamics of zooplankton is critical in understanding the flow of energy to higher trophic levels (Mitra et al. 2014; Everett et al. 2017) and provides the motivation for the work presented here.

We use a zooplankton database that informs management of the Cape Cod Bay (CCB), MA, USA habitat, specifically related to those characteristics that impact the critically endangered North Atlantic right whales (*Eubalaena glacialis*; NARW; Cooke 2020). The study of the NARW prey resource in CCB suffers from intermittent sampling, both spatially and temporally. Despite the difficulty in describing the distribution and abundance of the zooplankton, such methodology is important for the development of conservation strategies related to the habitat use patterns of the NARW's.

We describe the distribution and abundance of *sea surface* zooplankton utilizing the following data. At irregularly spaced locations (in CCB) and times (in 2003–2019), we have two sources of space-time data collection for zooplankton, both of which inform about sea surface zooplankton abundance. These data are the product of two methods of collection, as described below. We offer a fusion of these two response sources which requires calibration. Further, an important environmental factor that affects sea surface zooplankton abundance is sea surface temperature (SST; Mauchline 1998; Richardson 2008). Here, we also have two different data sources for SST, also described below. We need to fuse these sources as well in order to develop an SST regressor. These sources also require calibration; hence we adopt the terminology of *double* fusion and *double* calibration for our modeling.

The notions of fusion and calibration arise here from the well-discussed issue of measurement error (or errors in variables) in the Statistics literature (see, e.g., Fuller 1987; Carroll et al. 2006). Substantial published literature is available on stochastic spatial or spatiotemporal data fusion modeling for environmental and ecological data. For example, Fuentes and Raftery (2005) consider fusion of air pollution data sources. In this regard, Liu et al. (2011) and Zidek et al. (2012) consider fusion for ozone sources. Foley and Fuentes (2008) present wind data

modeling fusion. Finley et al. (2011) offer a fusion for agricultural crop yields. Liu et al. (2016) investigate fusion to predict spatial tracks for marine mammals. Very recent work (Villejo et al. 2023) introduces the INLA-SPDE approach for spatial data fusion.

Our context is spatiotemporal data fusion, and neither the SST sources nor the zooplankton abundance sources observe the true value; they are observed with measurement error. However, there is a conceptual *true* SST which, in part, influences a conceptual *true* zooplankton abundance and we seek to learn about the latter surface for CCB over space and time. For each variable, we have two data sources, which inform us about the true values. For each variable, one source is essentially observed up to measurement error while the other requires calibration relative to the truth. An attractive generative modeling strategy (i.e., modeling which could actually have produced the observations) is to use a hierarchical specification, introducing a latent true surface and then connecting the observed data to the truth (Fuentes and Raftery 2005; Banerjee et al. 2014, Chapter 15).

The benefit of the fusion is to provide a more appropriate assessment of uncertainty, yielding prediction that is better centered. Ignoring measurement error underestimates uncertainty and leads to poorer prediction. This can be argued technically but is challenging to demonstrate with real data since the truth is not known. So, in the interest of illuminating the benefit, justifying the additional modeling effort to implement the fusion and calibration, we also provide a simplified simulation illustration where we know the truth.

Observed zooplankton abundance is also connected to the spatial variable, presence or absence of NARW's at the location of collection, i.e., where whales are present, we would expect higher zooplankton abundance. However, this presence/absence regressor is only recorded at sites where abundance data are collected. Further, data are collected at a combination of regular stations (i.e., spatially-fixed) and non-regular (spatially-varying) sites for varying number of days within the first half of the year over the 17 year period. The regular stations supply repeated measurements over time; the non-regular sites are essentially opportunistic, providing site measurements only at the unique time of collection and predominantly collected in the vicinity of feeding whales.

More precisely, we propose a continuous-space, discrete-time multi-level/hierarchical model, implemented via the foregoing fusion and calibration processes. We use this model to develop a prediction of daily spatial zooplankton abundance surfaces over CCB, enabling an assessment of spatial variability as well as inference regarding average abundance over CCB or subregions. The partial observation of the NARW presence/absence variable adds complication to such prediction. Is our prediction based on the assumption the indicator is 1 (presence) everywhere? Or should our prediction be based on the assumption the indicator is 0 (absence) everywhere? Neither surface would be what we want. Rather, we seek to provide an average surface, averaged over a spatial probability of presence surface. In fact, this averaging requires the estimation of the probability of whale presence at an arbitrary location on an arbitrary day in an arbitrary year. For this purpose, a further modeling component is needed; we develop a predictive spatiotemporal logistic regression for the probability of presence. The above modeling enables inference for spatial

abundance at selected temporal scales, e.g., averaged to bi-weekly or annual scale as well as to make inter-annual comparisons. As is evident, the modeling is complex and model fitting is challenging. However, again, the reward is prediction with better precision and more appropriate accuracy.

Returning to the data sources, for both the zooplankton data and the SST data, fusion and calibration are needed. As above, our spatial fusion approach assumes a latent true surface for the variable with each of the sources varying in some fashion around this truth. Calibration is required for the fusion. See, e.g., Fuentes and Raftery (2005) or Rundel et al. (2015) for an illustration of this so-called *Bayesian melding* or latent truth approach. For instance, below, we assume that the surface zooplankton measurements, since they are surface measurements, are observed as the true surface zooplankton plus error. However, since an oblique measurement is obtained by pulling the tow up from 19 ms, it is anticipated to *overestimate* surface zooplankton; as a result, it needs to be calibrated. We have a similar challenge with the temperature. Here, we also have two sources, one is highly accurate and available across all of CCB but is not observed exactly at the locations where the zooplankton is collected. The other is measured at the boat when the collection was made. It is expected to be less reliable due to variability in the depth of the measuring device on the boat and is available more sparsely over CCB. Consequently, we assume the former is observed as true SST plus error, while the latter may benefit from calibration.

The format of the paper is as follows. Section 2 offers the details on the datasets that were employed along with associated exploratory summaries. Section 3 demonstrates the benefit of our fusion and calibration through a simpler illustration. Section 4 develops the full spatiotemporal hierarchical specification with associated fitting details and model comparison. Section 5 presents a rich range of results from the model fitting. Finally, Sect. 6 concludes with a brief summary and directions for future work.

2 The dataset

Zooplankton abundances and temperature measurements were obtained from the Center for Coastal Studies (CCS) zooplankton database, which was established in 1981 (Hudak et al. 2023). The primary goal of CCS' long term monitoring is to assess the critically endangered NARW's prey resource within CCB, by using a 333-μm mesh filtration to mimic the filtration efficiency of the right whale's baleen (Mayo et al. 2001). While this database provides an invaluable resource, it presents several spatial and temporal challenges for statistical analysis, as sampling varied spatially over time, depending on the field objectives of a given research cruise.

The dataset contained 2012 observations collected within CCB over a 17-year period (2003–2019) in the first 6 months of each year (January through June). Each observation consisted of the following elements: (1) zooplankton abundance measurements from one or two types of plankton collection (surface and oblique net tows); (2) one or two types of SST measurements (CTD and thermistor, specified below); and (3) a binary variable indicating the presence/absence of

a NARW at the time of collection. The zooplankton collection methods included surface samples collected from a 30-cm diameter conical plankton net and flow meter towed at the surface, and oblique samples collected from a 60-cm diameter conical plankton net and flow meter dropped vertically to 19 ms then towed back to the surface obliquely. This deep drop associated with the oblique tows anticipates higher zooplankton per cubic meter than surface tows, motivating the need to calibrate the former. All samples were subsampled and each subsample was enumerated to obtain the abundance measurement (organisms/m^3) (Hudak et al. 2023). While the primary copepod prey species of NARW is *Calanus finmarchicus*, NARW feed on other species, including those from the *Pseudocalanus* complex (i.e., *Pseudocalanus*, *Paracalanus*, and *Clausocalanus* species) and *Centropages* genera (Waktins and Schevill 1976; Mayo and Marx 1990; Hudak et al. 2023). In CCB, the above mentioned taxa dominate the zooplankton resource in a successive pattern of three regimes (i.e., abundance dominance among the three taxa groups), with the turnover between the regimes within years at days 34 ± 3 and 92 ± 3 (Hudak et al. 2023); therefore, we use the total zooplankton abundance to account for the shift in dominance between the taxa through the study period.

The temperature collection methods included the Sea-Bird SBE-19 conductivity-temperature-depth (CTD) profiler, and temperature measurements recorded at the time of zooplankton collections from the negative temperature coefficient (NTC) thermistor sensor. Both temperature measurements are available in degrees Celsius (ºC). The CTD profiler was used intermittently throughout the years to sample ocean temperatures (Mayo et al. 2004). The profiler was mounted on a steel cage and lowered through the water column to the bottom or specific depths, with a scan rate of every 0.5 s. Only temperature measurements taken in less than 1.25 ms in depth for SST's were used. To increase the temperature data pool we included temperature measurements recorded at the time of zooplankton collections from the NTC thermistor sensor positioned on the keel of the vessel at ∼ 1 m depth (Mayo et al. 2004). The SST's obtained from the thermistor data may also benefit from calibration since they depend upon the depth of the thermistor in the water, which varies over time of collection.

All observations had the binary NARW sighting variable, at least one type of zooplankton measurement, and either one, both, or no temperature measurement associated with it. See Sect. S1 of the Supplementary Information (SI) for additional details on the proportion of non-missing measurements in the observations. The observations were partitioned into regular (1595, 79.3%) and non-regular stations (417, 20.7%). Figure 1 shows the position of the observations at non-regular stations (grey points) and regular stations (red crosses) within CCB. Regular stations had fixed coordinates and were repeatedly observed at these coordinates over time. A total of 20 regular stations were used of which 8 had more than 100 observations each over the study period. In contrast, non-regular stations had unique coordinate locations over time. Their collection was generally motivated by the presence of NARW's, as part of an effort to describe the environment selected by right whales. NARW presence at regular stations was 5.8% of the observations and 87.6% in non-regular stations, revealing the strong sampling bias for NARW's in non-regular stations. However, in order to employ as much
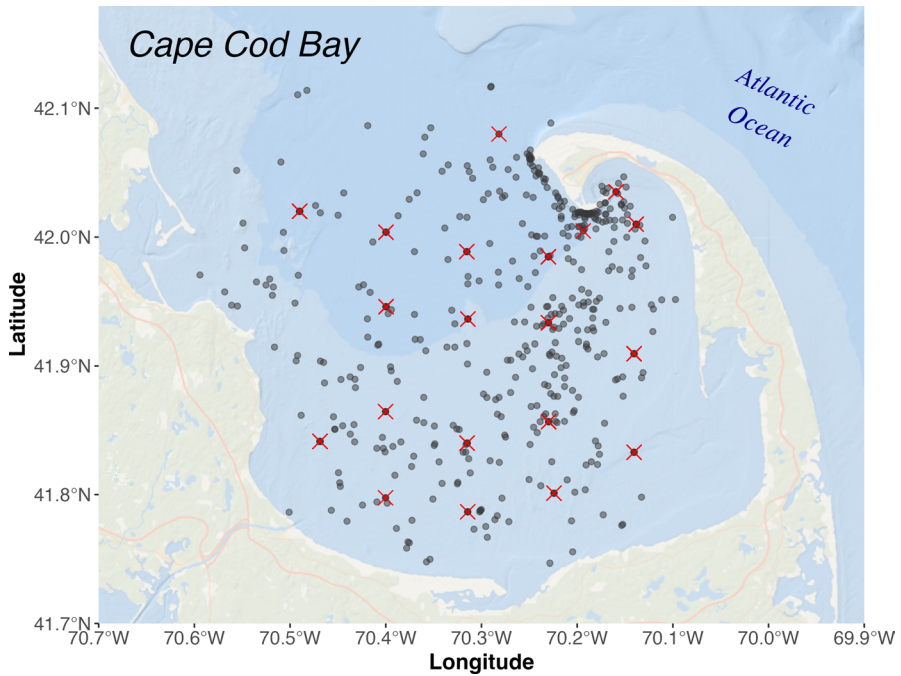
Fig. 1 Position of the observations at non-regular stations (grey points) and regular stations (red crosses) in CCB, MA, USA

zooplankton data as is available, we use the data from both the regular and non-regular stations, using the presence indicator variable as a regressor. That is, we expect more zooplankton (more prey) at the non-regular stations in response to whale presence.

Section S1.1 of the SI includes additional information about observation counts per year and day, as well as within year, by type of station, plankton collection method, and SST measurement.

2.1 Exploratory data analysis

We summarized an extensive exploratory data analysis of the zooplankton abundance per cubic meter, SST, and NARW presence data to establish the hierarchical structures in the model (see Sect. S1 of the SI for additional details). Kernel density estimation and normal quantile–quantile plots were plotted on the two measurement scales, original and logarithmic, for zooplankton measurements. Both zooplankton measurements on the original scale had a high positive skew, but, on the log scale, the variance stabilized (working at the log scale is usual for zooplankton data, see, e.g., Plourde et al. 2019; Record et al. 2019).

Figure 2 shows the seasonal pattern by days within years using moving averages (top) and the average annual long-term trend using annual averages (bottom) of the
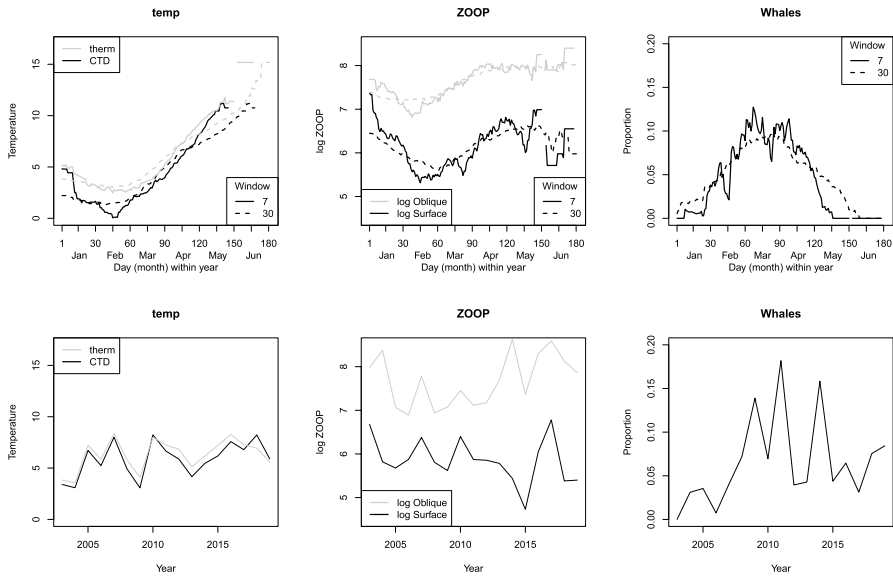
Fig. 2 For temperature (left), zooplankton (center), and proportion of NARW presence (right). Top: Moving average of data in a window of 7 and 30 days on each side (for days within years) and all years. Bottom: Average by year. (For NARW presence, we only average data for regular stations. At the bottom, for temperature and zooplankton we only average data where both collection methods are available)

two types of temperature measurement (left), the two types of zooplankton measurement (center), and the proportion of NARW presence (right). Both temperature measurements have the same strong seasonal component and a slight increase in the long-term trend, with the measurements from the thermistor differing by 1 °C above those from the CTD. Both zooplankton measurements have a similar seasonal pattern with log oblique shifted above log surface. The transition points of the regime effect are clear in the zooplankton plot, with three distinguished levels; two higher in January and later from April, and a lower one in February and March. The seasonal component in the NARW presence is strongly unimodal with a maximum in March and a NARW presence close to zero in January and June.

Section S1.2 of the SI includes additional exploration about the influential covariates in the zooplankton modeling; i.e., SST, the binary variable indicating the presence of NARW's, and regime. Linear regression models were fitted as exploratory tools to examine the components needed to capture the seasonal component of the temperature, zooplankton, and NARW's processes; these were generally modeled using harmonics (or regime for zooplankton).

Once the processes of interest and their relationships have been described, we delve into the relationship between the two types of measurement of zooplankton and of temperature, since they will be fused in the modeling. Section S1.3 of the SI includes additional exploration about the calibration of the zooplankton and temperature sources. In summary, the relationship between log surface and oblique measurements appears linear, but there is some variability associated with each type of measurement; this relationship is expected to vary in space and across years.

Thermistor and CTD measurements show an exceptionally strong linear relationship that could vary spatially, but does not appear to vary in time.

3 A simulation example to demonstrate the benefit of the fusion

As discussed in the Sect. 1, it is challenging to demonstrate the benefit of calibration and fusion with real data. Specifically, in the context of more than one source informing about the true value of a variable, neither source provides observations that are the true values. Each source has measurement error and, in the absence of external validation data, the true value of the variable is not observed. In a regression setting, whether a predictor or the response (or both) is observed with measurement error, learning about the true regression relationship is the objective. Therefore, in the absence of the true values, using the observed values implies increased uncertainty in this inference objective. Ignoring measurement errors leads to potentially inaccurate assessment of precision and accuracy. In order to demonstrate the improvement in inference by incorporating measurement error, in the form of calibration and fusion, into our modeling specification, we present a simple simulation example, in the spirit of our setting, where the truth is known.

We generated 100 points inside the unit square, as realizations from $Y_{true}(\mathbf{s})$. We captured spatial dependence with a mean $\mu = 4$ Gaussian process (GP) arising from an exponential covariance function with variance $\sigma^2_{true} = 4$ and decay parameter $\phi_{true} = 3/0.5$, so the spatial range is 0.5. Then, we sampled 80 locations at random. We randomly assigned half of these locations to source 1 and the other half to source 2. For source 1, we obtain the responses using $Y_1(\mathbf{s}) = Y_{true}(\mathbf{s}) + \epsilon_1(\mathbf{s})$ where the $\epsilon_1(\mathbf{s})$ are i.i.d. $N(0, \sigma^2_1 = 4)$. This is the measurement error sample. For source 2, we obtain the responses using $Y_2(\mathbf{s}) = \lambda_0 + \lambda_1 Y_{true}(\mathbf{s}) + \epsilon_2(\mathbf{s})$ where the $\epsilon_2(\mathbf{s})$ are i.i.d. $N(0, \sigma^2_2 = 1)$ for fixed $\lambda_0 = 2$ and $\lambda_1 = 2$. This is the calibration sample. The remaining 20 locations are employed as "out of sample" for both sources and are viewed for validation. We predict the $Y_{true}(\mathbf{s})$ for these points using kriging.

We fit Model 1 using just the source 1 data in a standard spatial linear regression model, $Y_1(\mathbf{s}) = \mu + w(\mathbf{s}) + e(\mathbf{s})$ where $w(\mathbf{s})$ is a mean-zero spatial GP (with known decay parameter) and $e(\mathbf{s})$ is pure error. So, this model has three parameters, $(\mu, \sigma^2_w, \sigma^2_e)$. We fit Model 2 using the source 1 and source 2 data. This model specifies $Y_1(\mathbf{s})$ and $Y_2(\mathbf{s})$ as above. It has six parameters, $(\mu, \sigma^2_{true}, \sigma^2_1, \lambda_0, \lambda_1, \sigma^2_2)$. Under each model, we do posterior prediction for the 20 hold out true values. We repeat the

Table 1 Performance metrics (RMSE, MAE, and CRPS) for the two models summarized by their average across 100 simulations

	RMSE	MAE	CRPS
Model 1	1.72	1.40	1.02
Model 2	1.45	1.17	0.85

entire simulation 100 times. Table 1 summarizes across simulations, using the predictive performance metrics described in Sect. 4.5 below, RMSE, MAE, and CRPS. Model 2 improves Model 1 by roughly 15–20% based on all metrics, demonstrating the benefits of the fusion in terms of predictive performance.

4 Modeling

We propose a hierarchical model for zooplankton abundance per cubic meter that requires double calibration using fusion of the two SST sources as a regressor and the two zooplankton measurement methods as a response. The model operates over continuous space and two discrete temporal scales, years and days within years. The fundamental assumption underlying the calibration modeling is that there are "true" unobserved daily spatial processes for zooplankton and for temperature, and that the two sources of observations are related to these processes via measurement error models. As we elaborate below, the measurement error models assume that one data source is the truth plus error while the other needs *spatiotemporal* calibration. The models for true zooplankton and true temperature capture spatial dependence through spatial process modeling of yearly or constant-in-time intercepts, respectively. The amount of zooplankton is, presumably, associated with the presence of whales, but here we only attempt to model abundance of the former, introducing presence/absence of the latter at a site as a predictor. Prediction under this modeling of zooplankton abundance at unobserved locations requires averaging/marginalization over this unobserved binary regressor at arbitrary locations. In order to address this, we propose a simple logistic regression response model for the presence of a whale at any zooplankton location of interest which operates only on temporal scales. We detail the calibration modeling below and then discuss model fitting, validation and model comparison. We also present details of prediction under the calibration and the marginalization using a binary response model.

We specify geostatistical, i.e., point level models. While there is a SST at every location in CCB, conceptually there is no zooplankton abundance at a point. The recorded abundance arises from the tow collection procedure and is scaled to a per cubic meter value. That value is assigned to the geo-coded location of the collection site. This does not prevent us from inferring a zooplankton abundance surface; rather, it clarifies how to interpret the value at a selected location. Further, if we seek an average zooplankton abundance over CCB for a given day in a given year, we will interpret it as a block average of these point level abundances.

4.1 The temperature model

We begin with the space-time calibration modeling for true temperature; in fact, since we seek sea surface zooplankton, we confine this predictor to "true" SST. Let t denote year and let j denote day within year, with years running from 2003

to 2019 and days from 1 (January 1) to 181 (June 30, except for leap years). Let $temp_{t,j}(\mathbf{s})$ denote the true SST for day j in year t at location $\mathbf{s}$ in CCB. We have two data sources for temperature to fuse: CTD, which we take as truth up to error, and the thermistor data, which we expect to benefit from calibration. We propose the following joint model to implement the fusion of the two sources:

$$
\begin{aligned}
CTD_{t,j}(\mathbf{s}) &= temp_{t,j}(\mathbf{s}) + \epsilon_{t,j}^{CTD}(\mathbf{s}), & \epsilon_{t,j}^{CTD}(\mathbf{s}) &\sim \text{ i.i.d. } N\big(0, \sigma_{CTD}^2\big), \\
Therm_{t,j}(\mathbf{s}) &= \tilde{\alpha}_0(\mathbf{s}) + \tilde{\alpha}_1(\mathbf{s})temp_{t,j}(\mathbf{s}) + \epsilon_{t,j}^{Therm}(\mathbf{s}), & \epsilon_{t,j}^{Therm}(\mathbf{s}) &\sim \text{ i.i.d. } N\big(0, \sigma_{Therm}^2\big),
\end{aligned}
\tag{1}
$$

where

$$
\begin{aligned}
\tilde{\alpha}_0(\mathbf{s}) &= \alpha_0 + \alpha_0(\mathbf{s}), \\
\tilde{\alpha}_1(\mathbf{s}) &= \alpha_1 + \alpha_1(\mathbf{s}),
\end{aligned}
\tag{2}
$$

and the $\epsilon_{t,j}(\mathbf{s})$'s are pure errors at sites for days within years for both data sources. They provide learning about the uncertainty associated with the CTD and thermistor measurement processes.

Here, α_0 and α_1 provide the overall slope and intercept bias of the thermistor model, while $\alpha_0(\mathbf{s})$ and $\alpha_1(\mathbf{s})$ provide local adjustments, respectively. The spatially-varying coefficients $\alpha_0(\mathbf{s})$ and $\alpha_1(\mathbf{s})$ are in turn modeled as bivariate mean-zero spatial GP's using the method of coregionalization (Banerjee et al. 2014, Chapter 9). Therefore, we suppose that there exist two mean-zero unit variance independent GP's, $v_0(\mathbf{s})$ and $v_1(\mathbf{s})$, with exponential covariance functions having decay parameters ϕ_{v_0} and ϕ_{v_1}, respectively. Then, we specify $\alpha_0(\mathbf{s}) = a_{00}v_0(\mathbf{s})$ and $\alpha_1(\mathbf{s}) = a_{10}v_0(\mathbf{s}) + a_{11}v_1(\mathbf{s})$, where a_{10} captures the dependence between the two processes.

Next, the true temperature for day j, year t, and location $\mathbf{s}$ is modeled additively in space and time by

$$
temp_{t,j}(\mathbf{s}) = \gamma_0 + \gamma_{0t} + \gamma_0(\mathbf{s}) + \gamma_1\sin(2\pi j/365) + \gamma_2\cos(2\pi j/365)
\tag{3}
$$

in which γ_0 is a global intercept, and the sin and cos terms are introduced to provide a semi-annual seasonal component. Based on the exploratory analysis, additional harmonic terms are not necessary. We consider an autoregression in years with $\gamma_{0t} \sim N(\rho_\gamma \gamma_{0t-1}, \sigma_\gamma^2)$ and $\gamma_{0,2002} \equiv 0$, providing annual intercepts. This specification can help to capture factors yielding correlation across years, such as naturally occurring ocean–atmosphere oscillations or the slight increase in temperatures observed in the exploratory analysis. In addition, $\gamma_0(\mathbf{s})$ provides local spatial adjustment to the intercept and it is modeled by a mean-zero spatial GP with an exponential covariance function having variance parameter $\sigma_{\gamma_0(\mathbf{s})}^2$ and decay parameter $\phi_{\gamma_0(\mathbf{s})}$. With this specification the data models consider pure errors but the true temperature has no pure error and, in accord with the behavior of the true process, it is continuous over space.

4.2 The zooplankton model

Turning to the zooplankton data, let $Y_{t,j}^{sur}(\mathbf{s})$ denote the logarithm of the surface collection at day j in year t at location $\mathbf{s}$ in CCB, similarly, for the logarithm of the oblique collection, $Y_{t,j}^{obl}(\mathbf{s})$. Let $\mathbf{1}_{t,j}(\mathbf{s})$ be the indicator as to whether there was a whale sited when zooplankton was collected at day j, year t, site $\mathbf{s}$. Let r_j indicate the regime (one of three regimes) associated with day j. Following Hudak et al. (2023), the first regime corresponds to $j \in [1, 32]$, the second to $j \in [33, 89]$, and the third to $j \in [90, 181]$. Then, with the goal of obtaining zooplankton abundance at the sea surface, we have that for $\mathrm{ZOOP}_{t,j}^{true}(\mathbf{s})$, the true surface zooplankton log abundance per cubic meter is additive in space and time; i.e., at day j, year t, site $\mathbf{s}$:

$$\log \mathrm{ZOOP}_{t,j}^{true}(\mathbf{s}) = \beta_0 + \beta_1 temp_{t,j}(\mathbf{s}) + \delta \mathbf{1}_{t,j}(\mathbf{s}) + r_j + \eta_t(\mathbf{s}). \tag{4}$$

Here, β_0 is a global intercept, β_1 is the coefficient that accompanies the true temperature modeled earlier, and δ accompanies the binary variable indicating NARW presence (i.e., the presence of a whale increases in δ units the logarithm of zooplankton abundance). Again, r_j indicate regime and is modeled with a dummy variable having three categories where the first regime is the reference. Finally, $\eta_t(\mathbf{s})$ is a yearly spatial process providing local adjustment, centered around η_t which introduces a global annual effect. We could consider dynamics as an autoregression in the η_t's. However, with a fairly small number of years we found no evidence of such autoregression. Consequently, we model $\eta_t \sim$ i.i.d. $N(0, \sigma_\eta^2)$. As for the $\eta_t(\mathbf{s})$, for each t, they follow a GP with mean parameter η_t and a common exponential covariance function having a variance parameter $\sigma_{\eta_t(\mathbf{s})}^2$ and decay parameter $\phi_{\eta_t(\mathbf{s})}$.

There is no error term in (4). Again, the true surface for any day within any year is assumed to be smooth. Error terms appear when we connect the observed collection to the true. That is:

$$
\begin{aligned}
Y_{t,j}^{sur}(\mathbf{s}) &= \log \mathrm{ZOOP}_{t,j}^{true}(\mathbf{s}) + \epsilon_{t,j}^{sur}(\mathbf{s}), & \epsilon_{t,j}^{sur}(\mathbf{s}) &\sim \text{ i.i.d. } N\!\left(0, \sigma_{sur}^2\right), \\
Y_{t,j}^{obl}(\mathbf{s}) &= \tilde{\lambda}_{0t}(\mathbf{s}) + \tilde{\lambda}_{1t}(\mathbf{s}) \log \mathrm{ZOOP}_{t,j}^{true}(\mathbf{s}) + \epsilon_{t,j}^{obl}(\mathbf{s}), & \epsilon_{t,j}^{obl}(\mathbf{s}) &\sim \text{ i.i.d. } N\!\left(0, \sigma_{obl}^2\right),
\end{aligned}
\tag{5}
$$

calibrating the oblique net tows. The $\epsilon_{t,j}(\mathbf{s})$'s are pure errors where we attempt to learn about the uncertainty associated with the surface and oblique measurement processes. Finally, the space-time calibration coefficients are given by

$$
\begin{aligned}
\tilde{\lambda}_{0t}(\mathbf{s}) &= \lambda_0 + \lambda_{0t} + \lambda_0(\mathbf{s}), \\
\tilde{\lambda}_{1t}(\mathbf{s}) &= \lambda_1 + \lambda_{1t} + \lambda_1(\mathbf{s}),
\end{aligned}
\tag{6}
$$

where λ's and $\lambda(\mathbf{s})$'s are defined analogously to (2) changing the notation from a's to b's and from v's to w's. In addition, $\lambda_t = (\lambda_{0t}, \lambda_{1t})^\top$ are yearly adjustments to the intercept and slope bias, respectively. These coefficients are modeled dynamically in time (West and Harrison 1997) by $\lambda_t \sim N_2(\mathbf{P}\lambda_{t-1}, \mathbf{V}_\lambda)$ and $\lambda_{2002} \equiv \mathbf{0}$ where $\mathbf{P}$ has ρ_{λ_0} and ρ_{λ_1} in the diagonal and 0's otherwise.

As an aside, the true *surface* zooplankton is of direct interest with regard to connection to whale abundance (observed at the surface). However, we could be interested in overall zooplankton *biomass*, a more 3-dimensional view. In this regard, we could consider the oblique zooplankton observations as an appropriate measure since they collect zooplankton down to 19 ms. If so, then we would reverse the roles of the calibration above; we would view the oblique observations as varying around the truth and calibrate the surface observations.

We refer to the above specifications as a "double calibration" model and fit it to the data using all sites observed for all days within years. We fit the temperature specification and the zooplankton specification jointly as one fully hierarchical model. At any time and location, there are, potentially, four observable measurements, two temperature measurements and two zooplankton measurements. If at least one of the four measurements is observed, the observation is employed in the model fitting, i.e., we use all of the data. For a missing zooplankton or temperature measurement, we simply remove that value from the corresponding likelihood.

4.3 Model fitting details

4.3.1 Priors

The double fusion and calibration model is proposed as a Bayesian hierarchical formulation and is completely specified given priors for all the parameters. In general, these priors are weak or diffuse and, when available, conjugate prior distributions are chosen.

We consider the parameters of the temperature and zooplankton models in (3) and (4), respectively. We assign to each of the regression coefficients independent and diffuse normal priors with mean 0 and variance 100^2, i.e., for γ_0, γ_1, γ_2, β_0, β_1, δ, and the two non-reference categories in r_j. Additionally, the variance parameters of the spatial and temporal random effects are assigned inverse gamma priors with scale parameter and mean equal to 1, i.e., for σ_γ^2, $\sigma_{\gamma_0(\mathbf{s})}^2$, σ_η^2, $\sigma_{\eta_t(\mathbf{s})}^2$. It is not possible to estimate consistently all of the exponential covariance parameters under noninformative priors (Zhang 2004). However, with the decay parameter equal to 3/range, we set $\phi \equiv \phi_{\gamma_0(\mathbf{s})} = \phi_{\eta_t(\mathbf{s})} = 3/d_{max}$, where $d_{max} \approx 46.6$ km is the maximum distance between any pair of spatial locations. With concern regarding prior sensitivity, we have also explored model performance over a 2-dimensional grid of ϕ's taking values $3/d_{max}$, $6/d_{max}$, $12/d_{max}$, $21/d_{max}$, and $30/d_{max}$. These values correspond to approximate ranges of 46.6, 23.3, 11.6, 6.7, and 4.7 km, respectively. We have found that the model performance metrics described in Sect. 4.5 are essentially insensitive to the choice of the decay parameters and the selected combination of decay choices yields the same or slightly better results than the others. In addition, the spatial interpolation using Bayesian methods becomes computationally much more manageable if ϕ is kept fixed during model fitting. Finally, for the autoregressive term ρ_γ we assign a uniform prior over the interval $(-1, 1)$.

Regarding priors for the parameters in the calibration specifications (1) and (5), we adopt inverse gamma priors as above for all the variance terms, i.e., for σ_{CTD}^2,

σ^2_{Therm}, σ^2_{sur}, σ^2_{obl}. Global calibration of thermistor and log oblique is captured by $(\alpha_0, \alpha_1)^\top$ and $(\lambda_0, \lambda_1)^\top$. For these, we employ independent bivariate normal priors with prior mean equal to $(0, 1)^\top$ and a diagonal prior covariance matrix with large diagonal entries, 100^2, corresponding to vague prior variances. For the coregionalization coefficients, we place vague half/folded normal priors with scale parameter 10 on a_{00}, a_{11}, b_{00} and b_{11}, as they are related to the variances of the local adjustments; we add a vague zero-mean normal prior with variance 100^2 on a_{10} and b_{10}, as they are related to the correlation. Noting the identifiability problem in the decay parameters mentioned above, we employ an informative uniform over the interval $(3/d_{max}, 30/d_{max})$ for the decay parameters ϕ_{v_0}, ϕ_{v_1}, ϕ_{w_0} and ϕ_{w_1} (Banerjee et al. 2014, Chapter 6). For the dynamic coefficients, $\mathbf{V}_\lambda$ is assigned an inverse Wishart prior with 2 degrees of freedom and $\mathbf{I}_2$ as the positive definite scale matrix. For the autoregressive terms, ρ_{λ_0} and ρ_{λ_1}, we assign a uniform over $(-1, 1)$.

4.3.2 Computational details

Markov chain Monte Carlo (MCMC) is used to fit the model and obtain realizations from the joint posterior distribution of all the model parameters; we employ a Metropolis-within-Gibbs version. The regression coefficients have normal full conditional distributions, the variance parameters have inverse gamma full conditionals, the autoregressive coefficients have truncated normal full conditionals, the coregionalization coefficients, as well as the covariance matrix from the dynamic specification are also conjugate. The only parameters that are not conditionally conjugate are the decay parameters that have not been fixed. They are sampled using a random-walk Metropolis–Hastings step with truncated normal proposals. We implement hierarchical centering reparameterizations to improve the mixing of the algorithm (Gelfand et al. 1995; Pitt and Shephard 1999).

4.4 Prediction to obtain zooplankton abundance

Here, we show how we implement space-time interpolation to produce zooplankton abundance maps over CCB for a given day within a given year. This is a post model fitting activity. Furthermore, and perhaps more usefully, afterward, we can create bi-weekly or yearly maps by averaging in time. In addition, averaging in space, we can obtain daily, bi-weekly or yearly time series of averages in CCB.

4.4.1 Spatial zooplankton surfaces

In order to obtain the posterior predictive distribution for true zooplankton for day j within year t at an unobserved site $\mathbf{s}_0$, $\text{ZOOP}^{true}_{t,j}(\mathbf{s}_0)$, we need to introduce into (4) posterior samples of the model parameters, posterior realizations of the spatial processes, the true temperature at that time and site, $temp_{t,j}(\mathbf{s}_0)$, and a value of the indicator function, $\mathbf{1}_{t,j}(\mathbf{s}_0)$. To obtain predictions on the original scale, we exponentiate the predictive realizations drawn from (4). Posterior samples of the model

parameters are obtained from model fitting and posterior realizations of the GP's are obtained through Bayesian kriging (Banerjee et al. 2014, Chapter 6) using posterior samples of the parameters. In a similar manner, the fusion model for the CTD data and thermistor data enables a posterior predictive distribution for true temperature at any location in CCB. For prediction, we introduce posterior predictive realizations for the true temperature at the location.

As noted above, the $\mathbf{1}_{t,j}(\mathbf{s}_0)$ are supplied only for the actual observations. They are not known for the desired grid of prediction locations across the 181 days within the 17-year period. To address this, we can create a conditional zooplankton abundance map assuming the indicator is 0 everywhere or assuming the indicator is 1 everywhere. Since conditional abundance surfaces given a value of the indicator are of limited interest, we seek to marginalize over the indicator. That is, in principle, we need the probability surface, $P(\mathbf{1}_{t,j}(\mathbf{s}_0) = 1)$, hence $P(\mathbf{1}_{t,j}(\mathbf{s}_0) = 0)$, to obtain a marginal abundance for any location on a given day in a given year. Such a task is more than the currently available data can support. Instead, we focus on modeling temporal dependence for these probabilities, but we assume that they are constant over space for any given time. That is, we obtain a day within year weighting, but the weighting is global. The maps will still be spatial because the zooplankton abundance maps are; it is only the weighting for the indicator that is global.

Specifically, then, we model $p_{t,j}$, the probability of a sighting on day j within year t on the logit scale as:

$$\text{logit } p_{t,j} = \omega_0 + \omega_{0t} + \omega_1 \sin(2\pi j/365) + \omega_2 \cos(2\pi j/365). \tag{7}$$

Again, ω_0 is the intercept, and the sin and cos terms are introduced to provide an annual seasonal component. Exploratory analysis shows that one harmonic is enough to capture the seasonal component of the data. The ω_{0t}'s are i.i.d. annual random effects $\sim N(0, \sigma_\omega^2)$. We note that this specification is additive in day and year. As above, the priors are $\omega_0, \omega_1, \omega_2 \sim N(0, 100^2)$ and $\sigma_\omega^2 \sim IG(2, 1)$. We fit this model using only the data from the regular stations due to concerns regarding introducing sampling bias.

After fitting the model, from the true zooplankton model in (4), according to the value of $\mathbf{1}_{t,j}(\mathbf{s})$, we can obtain conditional posterior predictive realizations of $\text{ZOOP}_{t,j}^{true}(\mathbf{s})$ for any location at any day within any year. In order to develop a *marginal* posterior predictive distribution at unobserved locations and times, we average these conditioned realizations using the global weights, $p_{t,j,b}$ for $b = 1, \ldots, B$, yielding

$$\text{ZOOP}_{t,j,b}^{true}(\mathbf{s}) = \text{ZOOP}_{t,j,b}^{true}(\mathbf{s}) \mid \mathbf{1}_{t,j}(\mathbf{s}) = 1)p_{t,j,b} + \text{ZOOP}_{t,j}^{true}(\mathbf{s}) \mid \mathbf{1}_{t,j}(\mathbf{s}) = 0)(1 - p_{t,j,b}). \tag{8}$$

With kriging, these samples can be used to provide marginal posterior predictive inference at any $\mathbf{s}$. In particular, we can obtain the posterior predictive mean zooplankton at $\mathbf{s}$ on day j in year t.

Continuing with (8), researchers have documented critical zooplankton density thresholds needed to promote feeding in NARW (cf. Table 5 in

Gavrilchuk et al. 2021). To investigate the nature of such patches, we can use the above modeling to create the posterior predictive probability of exceedance of a specified zooplankton threshold, i.e., $P(\text{ZOOP}_{t,j}^{true}(\mathbf{s}) > c \mid data)$. Hence, we can develop a posterior predictive probability of exceedance surface for any threshold of interest. In this regard, the realized exceedance surface is a binary map, i.e., a 0 or 1 at each location in CCB. We can not show this but rather, we can show the posterior expectation of the map, i.e., $E(\mathbf{1}(\text{ZOOP}_{t,j}^{true}(\mathbf{s}) > c) \mid \text{data}) = P(\text{ZOOP}_{t,j}^{true}(\mathbf{s}) > c \mid \text{data})$, the posterior predictive probability surface.

4.4.2 Zooplankton time series through block averaging

It is of key interest to infer about average zooplankton abundance over CCB for a given day j within a given year t. This becomes a block average (Banerjee et al. 2014, Chapter 7), which is a stochastic integral defined by

$$\text{ZOOP}_{t,j}^{true}(CCB) = \frac{1}{|CCB|} \int_{CCB} \text{ZOOP}_{t,j}^{true}(\mathbf{s})\, d\mathbf{s}, \tag{9}$$

where $|CCB|$ denotes the area of CCB. We approximate this integral by Monte Carlo integration of the form

$$\widehat{\text{ZOOP}}_{t,j}^{true}(CCB) = \frac{1}{|G|} \sum_{\mathbf{s}_k \in G} \text{ZOOP}_{t,j}^{true}(\mathbf{s}_k), \tag{10}$$

where G is a sufficiently fine spatial grid of CCB and $|G|$ is the number of grid cells in G. In this way, with arbitrarily many posterior predictive realizations, we can learn arbitrarily well about the posterior predictive distribution for $\text{ZOOP}_{t,j}^{true}(CCB)$. We can plot the posterior mean of these integrals versus time to see daily trends in zooplankton abundance. In addition, we can average zooplankton at a site over days within a week, a month, or within a year. Lastly, implementing the block average can lead to annual time series. We denote the yearly averages by $\text{ZOOP}_{t.}^{true}(CCB)$, with analogous notation for their Monte Carlo version.

Instead of integrating over all of CCB in (9), a subregion of CCB could also be considered, so that the time trends in zooplankton abundance of different subregions could be compared. In a different direction, the idea of block averaging could be extended to the concept of extent (Cebrián et al. 2022). Defining the block average of the indicator $\mathbf{1}(\text{ZOOP}_{t,j}^{true}(\mathbf{s}) > c)$, we can formally define the spatiotemporal extent of the exceedance of a specified zooplankton threshold. Precisely, it is the proportion of CCB which is exceeding the specified zooplankton threshold c on day j within year t.

4.5 Metrics for validation and model comparison

Model validation can only be done with regard to observed measurements. In that sense, we can validate the temperature calibration using solely the CTD and

thermistor data. The zooplankton validation can be done with both the surface and oblique measurements but must be done using the full model. Since zooplankton prediction is our primary objective, we confine our validation to the full model.

For model assessment, we leave out 402 (20%) observations as test data and we fit the models with the remaining 1610 (80%) observations. Once MCMC samples of the model parameters are obtained, subsequent posterior predictive realizations of the true temperature are sampled for the test data. These samples are used to obtain the posterior predictive realizations for CTD and thermistor, and also to obtain posterior predictive realizations for true zooplankton. Again, these zooplankton samples are used to obtain posterior predictive realizations from surface and oblique.

We conduct model validation and comparison employing the following metrics (Castillo-Mateo et al. 2022): (1) root mean square error (RMSE), (2) mean absolute error (MAE), (3) continuous ranked probability score (CRPS), and (4) 90% coverage (CVG). The smaller the RMSE, MAE, and CRPS values, the better the model performance, and the target for CVG is proximity to 0.90. In the next section, we investigate comparison of four models.

5 Analysis

This section summarizes the results of the double calibration model in Sect. 4 fitted to the dataset described in Sect. 2. First, we validate and compare the proposed model with some parsimonious versions of it. Then, after preferring the full model (as we shall see in Sect. 5.1), we show the posterior inference for the model parameters as well as the posterior predictive spatial surfaces and time series of average zooplankton abundance.

For each fitting in the validation and comparison part as well as the fitting for the final model we use the MCMC described in Sect. 4.3. For each model, we ran two chains of the Gibbs sampling algorithm, each with different initial values, running 500,000 iterations for each chain, to obtain samples from the joint posterior distribution. The first 250,000 samples were discarded as burn-in and the remaining 250,000 samples were thinned to retain 1000 samples from each chain for posterior inference. Convergence of the algorithms (not shown) was monitored by usual trace plots and the potential scale reduction factors (Brooks and Gelman 1998).

5.1 Validation and model comparison

We compare the predictive performance of several models, according to the specification of the calibration coefficients in (2) and (6). We denote models with $\mathcal{M}$ and one subscript. The choices are: 1 indicates constant parameters over space and time, i.e., $\tilde{\alpha}_k(\mathbf{s}) \equiv \alpha_k$ and $\tilde{\lambda}_{kt}(\mathbf{s}) \equiv \lambda_k$ ($k = 0, 1$); 2 indicates spatially varying intercepts but constant slope parameters; 3 indicates all four coefficients spatially varying but fixed in time; and 4 indicates the full model for the calibration coefficients described in Sect. 4.

Table 2 Performance metrics (RMSE, MAE, CRPS and CVG) for the four elements of interest (log surface, log oblique, CTD and thermistor) in the models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$

	RMSE				MAE				CRPS				CVG			
	Y^{sur}	Y^{obl}	CTD	Therm	Y^{sur}	Y^{obl}	CTD	Therm	Y^{sur}	Y^{obl}	CTD	Therm	Y^{sur}	Y^{obl}	CTD	Therm
$\mathcal{M}_1$	1.81	0.98	0.99	1.09	1.47	0.74	0.75	0.85	1.03	0.53	0.54	0.61	0.88	0.91	0.88	0.90
$\mathcal{M}_2$	1.81	0.97	1.00	1.08	1.47	0.74	0.76	0.84	1.03	0.53	0.55	0.61	0.88	0.91	0.86	0.90
$\mathcal{M}_3$	1.81	0.97	0.99	1.08	1.47	0.74	0.75	0.83	1.03	0.53	0.54	0.60	0.89	0.90	0.87	0.88
$\mathcal{M}_4$	1.74	0.92	0.99	1.08	1.39	0.69	0.75	0.83	0.98	0.50	0.54	0.60	0.89	0.91	0.88	0.88
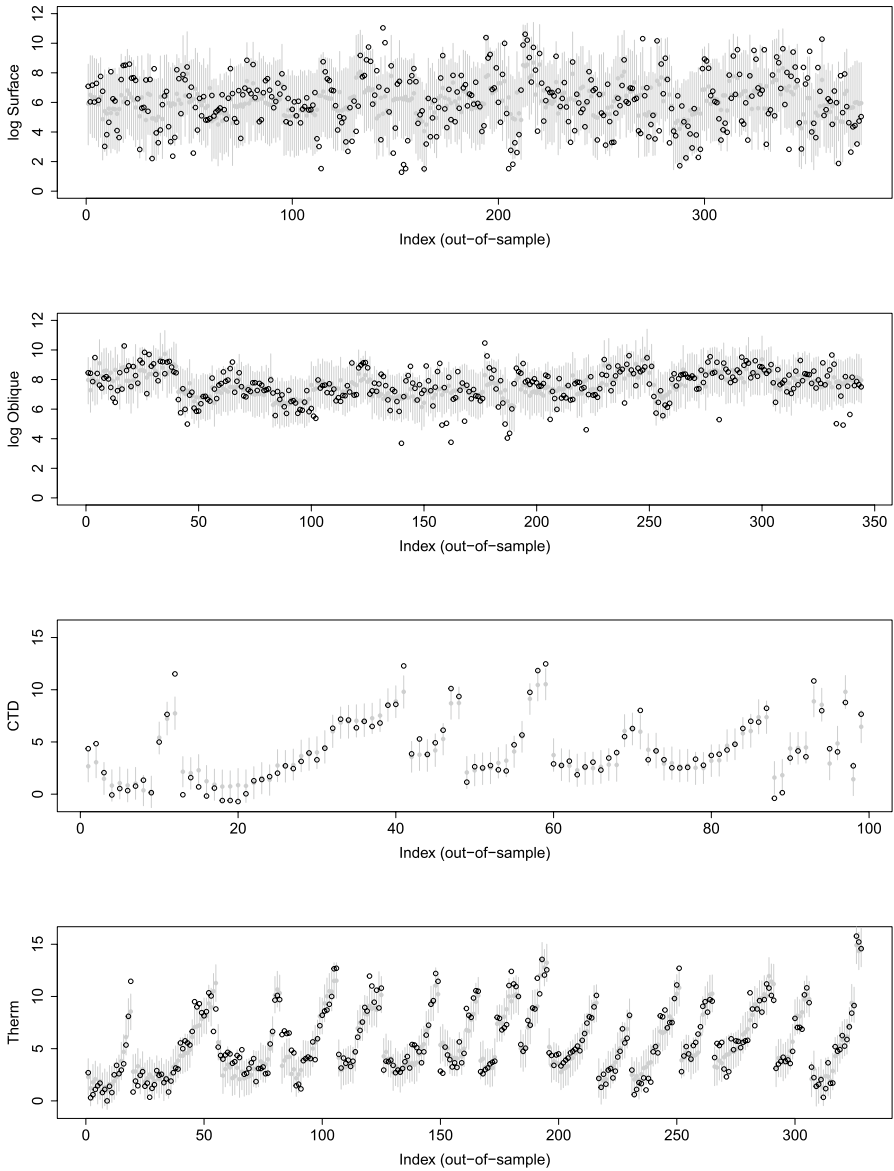
Fig. 3 Posterior predictive mean and 90% intervals (gray) for test log surface, log oblique, CTD and thermistor measurements (black), respectively

Following Sect. 4.5, Table 2 shows the performance metrics RMSE, MAE, CRPS and CVG for the test data for the four models described above. According to this table, all models perform similarly, with the full model presented in Sect. 4, $\mathcal{M}_4$, being the best. So, from here on we present the results from $\mathcal{M}_4$.

The 90% prediction intervals for the test measurements with this model are plotted in Fig. 3. Beyond the summary table of RMSE's, MAE's, CRPS's, and CVG's; it is useful to show a few posterior predictive distributions for test surface, oblique, CTD and thermistor data with the observed values overlaid, to give a visual illustration of the predictive performance. Out of the 402 observations for each type of measurement to be validated, 26 were missing for oblique, 58 for surface, 303 for CTD, and 74 for thermistor. Among the available measurements, 88.6%, 90.7%, 87.9%, and 88.1% of the 90% predictive intervals correctly included the observed values, indicating that the model is performing as it should for out-of-sample predictions.

5.2 Parameter estimates

Section S2.1 of the SI gives a detailed description with tables and figures of each of the parameters and processes of the model $\mathcal{M}_4$. Here, we give a summary of the main conclusions regarding zooplankton modeling in (4) and the calibration of measurements with (1) and (5). The most important parameters are summarized in Table 3.

Regarding the zooplankton modeling from (4), since β_1 is significantly positive, higher SST explains higher concentrations of zooplankton. The significance of δ asserts a larger intercept in the presence of whales than without it. This increases expected zooplankton abundance when a whale is near a site, in agreement with the expectation that whales identify areas of rich prey. However, whales are not continuously feeding and may be displaying other behaviors, e.g., engaging in surface active groups or traveling. In these instances we would not necessarily expect zooplankton abundance to be higher in the vicinity of whales. Distinguishing between these behaviors is beyond the scope of the present work. Further parameter inference is supplied in the SI, where we see that only the second regime is significant with the first as the base. However, the annual random effects show temporal variability

Table 3 Posterior mean and 90% credible intervals for some model parameters

		Mean	90% credible interval
ZOOP	β_0 (intercept)	5.80	(5.51, 6.10)
	β_1 (true temp)	0.06	(0.04, 0.09)
	δ (whale sighting)	1.25	(1.09, 1.42)
	σ_η^2	0.24	(0.12, 0.43)
	$\sigma_{\eta_t(s)}^2$	0.27	(0.18, 0.39)
Calibration	α_0 (thermistor)	1.00	(0.57, 1.35)
	α_1 (thermistor)	0.95	(0.89, 1.01)
	λ_0 (oblique tow)	4.62	(3.22, 6.10)
	λ_1 (oblique tow)	0.52	(0.20, 0.82)
	σ_{sur}^2	2.71	(2.55, 2.87)
	σ_{obl}^2	0.82	(0.77, 0.87)

True ZOOP top, calibration bottom

across years. The estimates of the variance components show that there is more unexplained variability in the pure errors than variability in the spatiotemporal effects. It is also seen that the surface measurement process has approximately three times more variance than the oblique measurement process on the log scale.

Focusing on the calibration of measurements with (1) and (5), the main result is given by the global coefficients of the calibrations, i.e., $\alpha_0, \alpha_1, \lambda_0, \lambda_1$. On average, thermistor measurements are shifted by 1 °C from true temperature and for every degree that it increases, the thermistor measurements increase by almost 1 °C. On the other hand, the log oblique measurements have a shift of four and a half units (organisms/m^3 on a log scale) with respect to the true log zooplankton; each unit that this increases results in an increase in the log oblique measurements by an average of half a unit.

5.3 Prediction of zooplankton surfaces and time series

For prediction, we need the probabilities of NARW presence, and the parameter estimates associated with its temporal model (7) are given in Sect. S2.2 of the SI. We also need a fine grid of G points within CCB. We choose G with resolution of 1 km × 1 km yielding $|G| = 1580$ grid cells. This grid is shown in Fig. S14 of the SI. Altogether, the entire posterior predictive time series results in very large data files from which maps and time series are computed. For instance, we have 17 years by 181 days by 1580 grid centroids by 2000 replicates yielding approximately 10 billion points.

With regard to space, we illustrate daily predictions of spatial surfaces for three illustrative days: 29 April 2011, 21 March 2017, and 17 April 2019. The predicted mean surfaces of zooplankton abundance and their standard deviation at these days are given in Fig. 4, at the top and the bottom, respectively. At the top, there are large differences with regard to the locations of highs and lows between the three
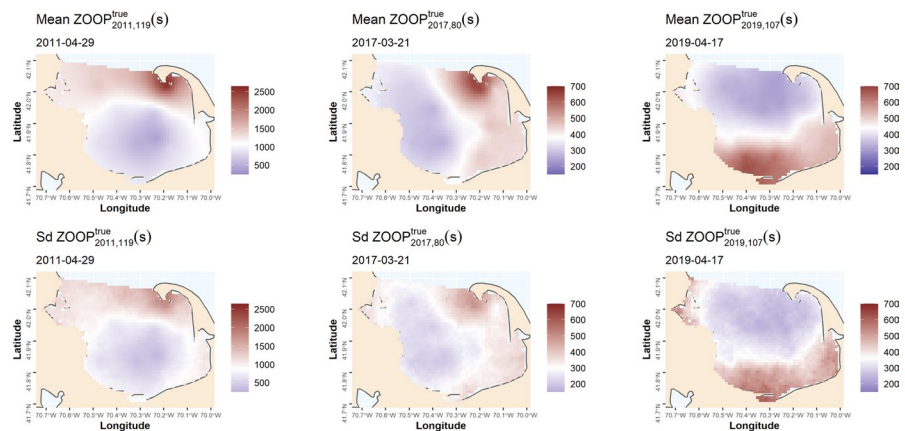
Fig. 4 Maps of the mean (top) and standard deviation (bottom) of the posterior predictive distribution of zooplankton abundance at 29 April 2011, 21 March 2017, and 17 April 2019, respectively
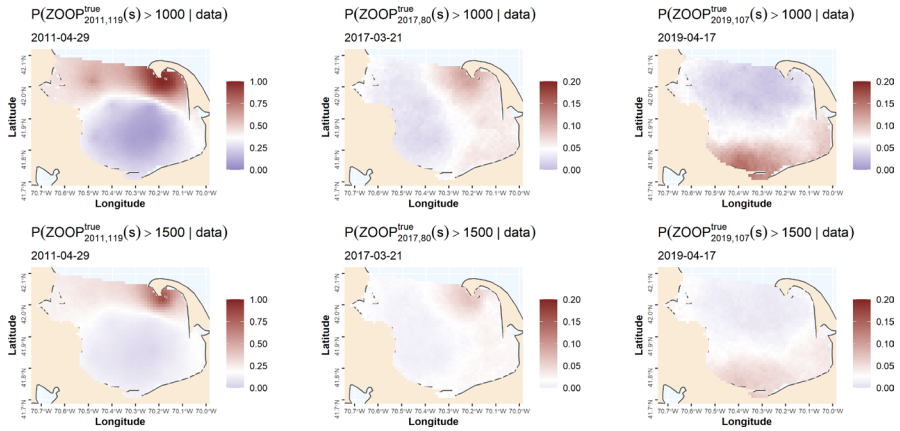
Fig. 5 Maps of the posterior probability of exceeding zooplankton thresholds $c = 1000$ (top) and 1500 (bottom) at 29 April 2011, 21 March 2017, and 17 April 2019, respectively

maps. Also, there is more spatial variation in the predicted means for day 29 April 2011 than for days 21 March 2017 and 17 April 2019. The maps of the standard deviations show similar spatial patterns. They tend to be larger where mean levels are larger. Because the mean levels were larger for day 29 April 2011, the standard deviations are larger for day 29 April 2011. The block averages for these days are 1090.3, 383.0, and 442.3 organisms/m^3, respectively.

In addition to the average zooplankton of CCB, the probability of threshold exceedance is of interest because it provides inference on the response of NARW to higher prey availability (Gavrilchuk et al. 2021). We started with a threshold value of 1000 organisms/m^3 (Mayo and Marx 1990), with recent analyses suggesting both higher and habitat-specific threshold values (Sorochan et al. 2021). Figure 5 show $P(\text{ZOOP}_{t,j}^{true}(\mathbf{s}) > 1000 \mid data)$ at the top and $P(\text{ZOOP}_{t,j}^{true}(\mathbf{s}) > 1500 \mid data)$ at the bottom. Note that the scale is different for each plot and the "white" is set to the mean level of the bay for that particular day. The spatial pattern is the same as in the previous figure, the day 29 April 2011 has a much higher average probability of exceeding the thresholds than the other 2 days, whose probabilities do not exceed 0.1.

To show inference for a longer period of time we represent the posterior mean of the annual prediction surface in Fig. 6. Illustrative bi-weekly predictions are shown in Sect. S2.2. Here, we observe both, variation across space and time of zooplankton abundance.

With regard to time, we illustrate block average daily and annual predictions, i.e., $\widehat{\text{ZOOP}}_{t,j}^{true}(CCB)$ and $\widehat{\text{ZOOP}}_{t.}^{true}(CCB)$. We also consider $\widehat{\text{ZOOP}}_{t,r_j=3}^{true}(CCB)$, i.e., rather than averaging zooplankton abundance across all the period, only predictions from the third regime are considered. In addition, we show extents of the event $\mathbf{1}(\text{ZOOP}_{t,j}^{true}(\mathbf{s}) > 1000)$ across days and years. The predicted time series and extents on a daily scale are given in Fig. 7. Figure 8 shows annual dynamics,
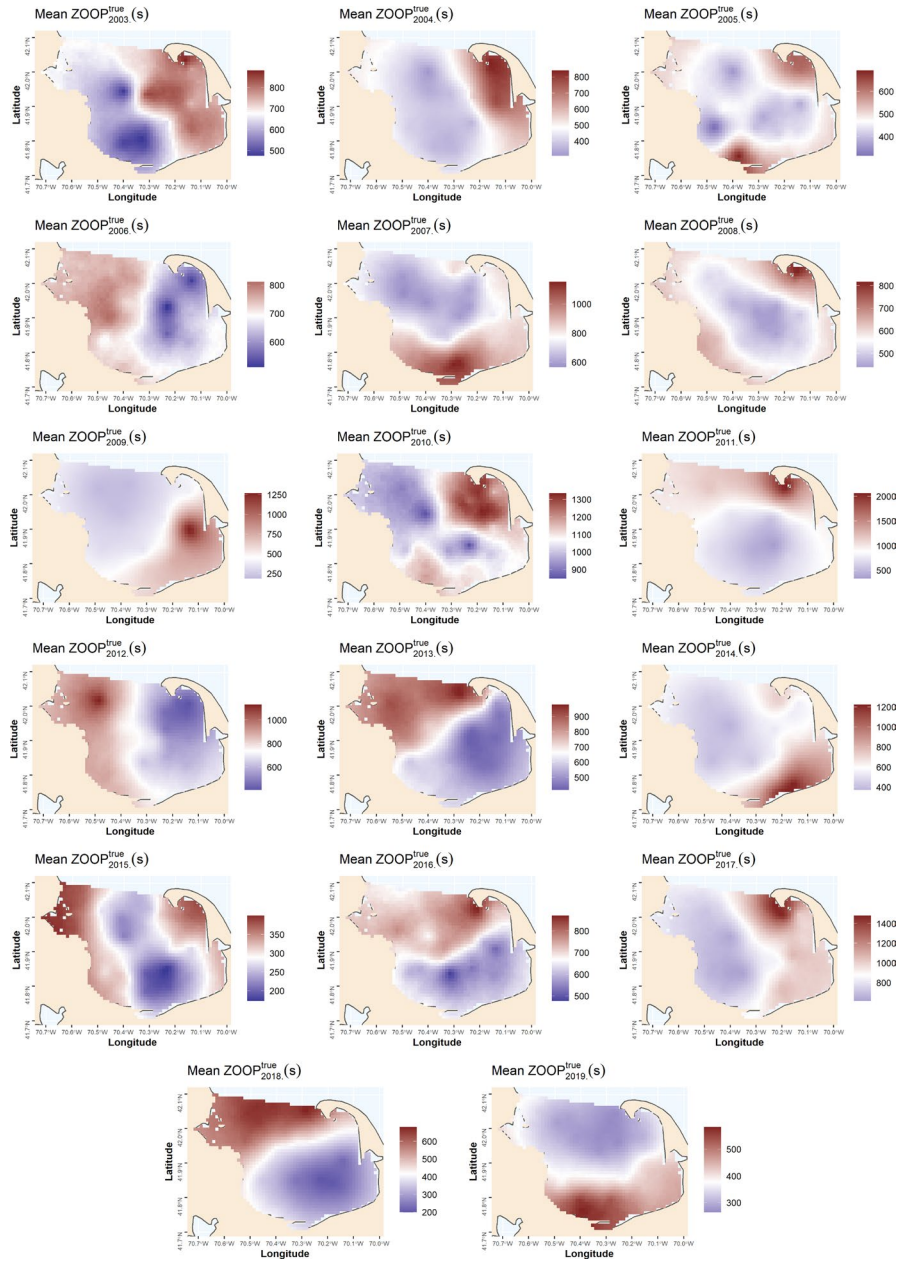
Fig. 6 Map of the posterior mean of the spatial predictions of annual average zooplankton abundance, $\text{ZOOP}_{t.}^{true}(\mathbf{s}) = \frac{1}{181} \sum_{j=1}^{181} \text{ZOOP}_{t,j}^{true}(\mathbf{s})$, for $t = 2003, \ldots, 2019$
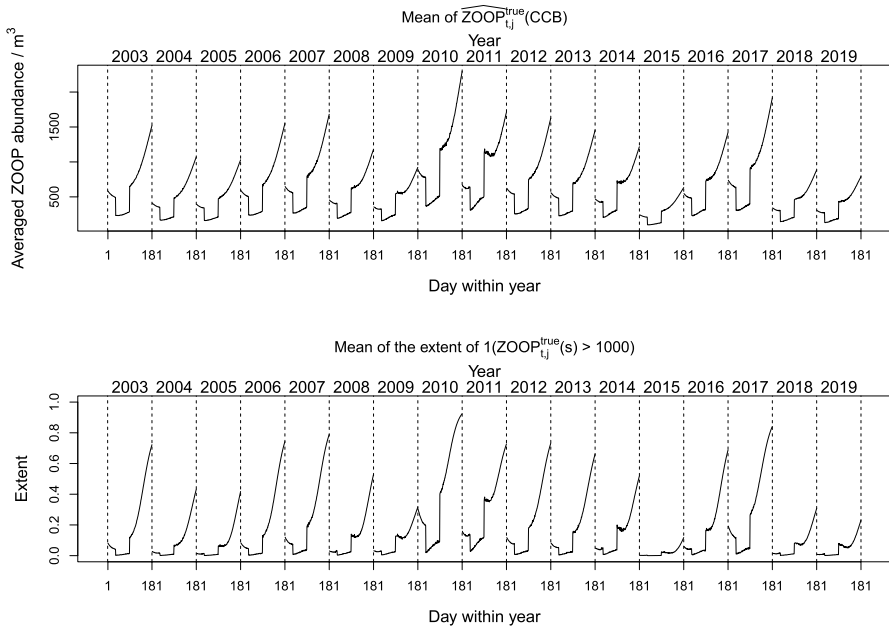
Fig. 7 Posterior mean of the time series of zooplankton abundance per cubic meter (top) and of the extent of the event $\mathbf{1}(\mathrm{ZOOP}_{t,j}^{true}(\mathbf{s}) > 1000)$ (bottom) in CCB by day through block average
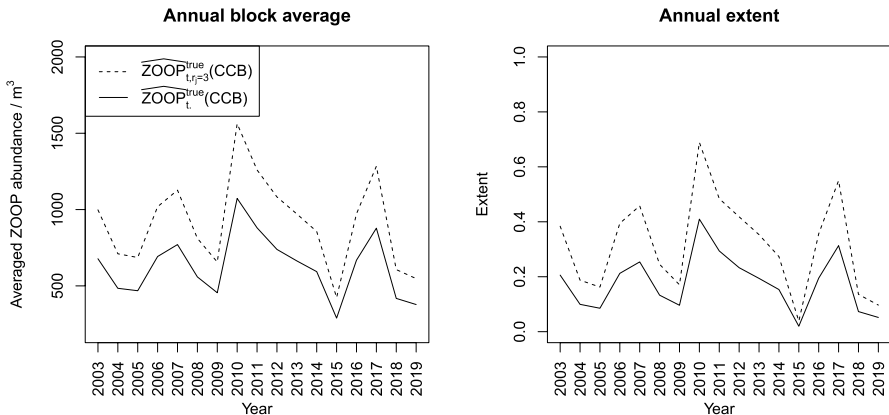
Fig. 8 Posterior mean of the time series of zooplankton abundance per cubic meter (left) and of the extent of the event $\mathbf{1}(\mathrm{ZOOP}_{t,j}^{true}(\mathbf{s}) > 1000)$ (right) in CCB averaged by year through block average

showing years with higher abundance of zooplankton than others, and consequently, years with higher average extents than others. The years 2010 and 2017 have the greatest abundance and extent; 2015 is the year with the least. Some of this variability is due to the NARW-focused sampling, e.g., in 2010 there was a considerable amount of sampling in the vicinity of skim-feeding NARW's. These

higher observed abundances likely increased the estimates for 2010. In 2017, while the zooplankton abundance followed the typical increasing-decreasing pattern seen in other years during the winter/spring period, higher concentrations were observed throughout the seasonal period. In addition, high concentrations of *Calanus finmarchicus* resulted in an extended zooplankton enriched resource.

In general, variability of the mean total zooplankton concentrations can be attributed to the high spatial variability with zooplankton found in very high concentrations in relatively small patches and extensive areas of low concentration. We assume that the spatial variability is due to environmental factors that form these patches along with the seasonal and behavioral responses of the zooplankton resource that varies by taxon (Folt et al. 1999; Hudak et al. 2023). In addition, the annual estimates of zooplankton abundance fit well with observations of NARW abundance (results not shown).

6 Summary, discussion, and future work

Zooplankton are an important food source for fish and whale species. To better understand their dynamics in a critical habitat for planktivorous NARW's, we have developed a multi-level model to learn about the spatial and temporal sea surface abundance of this prey. While a better understanding of the foraging ecology of NARW has motivated much of the data collection in CCB, here we have focused primarily on understanding the spatial and temporal distribution of zooplankton. We have used two sources for zooplankton collection along with two sources for SST, as a regressor to build this model. The modeling required fusion and calibration for each of the two data sources. The resulting model was applied to data collected over 17 years from CCB, MA, USA, and enables us to develop space-time abundance surfaces as well as probability exceedance surfaces over CCB. Using block averaging, the abundance surfaces were employed to yield average daily abundance for CCB. Also averaging to weeks, months, or years enables a useful, coarser picture of average abundance.

The issue of scale for foraging NARW is critical, and there are likely several scales over which they make foraging decisions. The spatial and temporal scales at which right whales actually open their mouths to feed are smaller than we are able to capture here; there is likely local increase/decline in abundance that is critical to NARW, which we cannot capture. However, with estimates of exceedance probability (Fig. 5) we can make initial comparisons of the prey resources to CCB-wide aerial observations of feeding NARW. Nevertheless, CCB remains a critical foraging habitat for right whales due to the phenological cycle of their prey and limited inter-annual changes in prey abundance (Hudak et al. 2023).

Acknowledging substantial seasonal variation in abundance behavior for the primary species included in the zooplankton database, future work will explore joint space-time modeling for copepods and NARW. Also, there is an additional SST data source available, HYCOM (Cummings and Smedstad 2013) that has wide coverage over the bay, but is recorded at a coarse spatial resolution relative to the spatial extent of CCB. Future work will assess whether this additional

source of temperature data can help the fusion for a SST regressor. Also, there is an additional zooplankton catch data source, vertical pump sampling up through the water column. Fusion with this source will also require careful calibration, particularly because it is never collected jointly with the surface and oblique tow net collection. A further opportunity will move our modeling to a larger spatial area to attempt to extract a zooplankton gradient over the spatial domain, driven by SST, e.g., through joint modeling of the ECOMON dataset (Hare 2021).

An exciting future challenge involves the consideration of bioenergetics. That is, can we demonstrate an association between high zooplankton abundance and high NARW abundance. This will require joint modeling for both abundances. In a generative sense, we will attempt the joint modeling through a conditional times marginal specification, i.e., a marginal model for zooplankton abundance (using some of what we have learned from the present work) and then a conditional whale abundance model given zooplankton abundance. For this modeling, we will have several whale abundance data sources, including distance sampling data (Ganley et al. 2019), passive acoustic monitoring data (Clark et al. 2010), and opportunistic data collection (i.e., the whale sighting data noted herein), resulting in a very demanding space-time data fusion challenge.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Andersen V, Fransz HG, Frost BW et al (1993) Modelling zooplankton. In: Evans GT, Fasham MJR (eds) Towards a model of ocean biogeochemical processes, vol 10. Springer, Berlin, Heidelberg, pp 177–191. https://doi.org/10.1007/978-3-642-84602-1_8

Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data, 2nd edn. Chapman and Hall/CRC, New York. https://doi.org/10.1201/b17115

Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7(4):434–455. https://doi.org/10.1080/10618600.1998.10474787

Burns CW, Folt CL, Folt CL et al (1999) Biological drivers of zooplankton patchiness. Trends Ecol Evol 14(8):300–305. https://doi.org/10.1016/S0169-5347(99)01616-X

Carroll RJ, Ruppert D, Stefanski LA et al (2006) Measurement error in nonlinear models, 2nd edn. Chapman and Hall/CRC, New York. https://doi.org/10.1201/9781420010138

Castillo-Mateo J, Lafuente M, Asín J et al (2022) Spatial modeling of day-within-year temperature time series: an examination of daily maximum temperatures in Aragón, Spain. J Agric Biol Environ Stat 27(3):487–505. https://doi.org/10.1007/s13253-022-00493-3

Cebrián AC, Asín J, Gelfand AE et al (2022) Spatio-temporal analysis of the extent of an extreme heat event. Stoch Environ Res Risk Assess 36(9):2737–2751. https://doi.org/10.1007/s00477-021-02157-z

Clark CW, Brown MW, Corkeron P (2010) Visual and acoustic surveys for North Atlantic right whales, Eubalaena glacialis, in Cape Cod Bay, Massachusetts, 2001–2005: management implications. Mar Mamm Sci 26(4):837–854. https://doi.org/10.1111/j.1748-7692.2010.00376.x

Cooke JG (2020) *Eubalaena glacialis*. The IUCN red list of threatened species 2020. e.T41712A162001243. https://doi.org/10.2305/IUCN.UK.2020-2.RLTS.T41712A162001243.en. Accessed 23 Feb 2023

Cummings JA, Smedstad OM (2013) Variational data assimilation for the global ocean. In: Park SK, Xu L (eds) Data assimilation for atmospheric, oceanic and hydrologic applications, vol II. Springer, Berlin, Heidelberg, pp 303–343. https://doi.org/10.1007/978-3-642-35088-7_13

Everett JD, Baird ME, Buchanan P et al (2017) Modeling what we sample and sampling what we model: challenges for zooplankton model assessment. Front Mar Sci 4(77):1–19. https://doi.org/10.3389/fmars.2017.00077

Finley AO, Banerjee S, Basso B (2011) Improving crop model inference through Bayesian melding with spatially varying parameters. J Agric Biol Environ Stat 16(4):453–474. https://doi.org/10.1007/s13253-011-0070-x

Foley KM, Fuentes M (2008) A statistical framework to combine multivariate spatial data and physical models for hurricane surface wind prediction. J Agric Biol Environ Stat 13(1):37–59. https://doi.org/10.1198/108571108X276473

Fuentes M, Raftery AE (2005) Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics 61(1):36–45. https://doi.org/10.1111/j.0006-341X.2005.030821.x

Fuller WA (1987) Measurement error models. Wiley, New York

Ganley LC, Brault S, Mayo CA (2019) What we see is not what there is: estimating North Atlantic right whale *Eubalaena glacialis* local abundance. Endanger Species Res 38:101–113. https://doi.org/10.3354/esr00938

Gavrilchuk K, Lesage V, Fortune SME et al (2021) Foraging habitat of North Atlantic right whales has declined in the Gulf of St. Lawrence, Canada, and may be insufficient for successful reproduction. Endanger Species Res 44:113–136. https://doi.org/10.3354/esr01097

Gelfand AE, Sahu SK, Carlin BP (1995) Efficient parametrisations for normal linear mixed models. Biometrika 82(3):479–488. https://doi.org/10.1093/biomet/82.3.479

Hare J (2021) EcoMon plankton counts (10-m2) from Bongo tows, including MARMAP data from multiple cruises in the Northeast Continental Shelf of the United States from 1977–2015 (EcoMon Zooplankton project). https://doi.org/10.26008/1912/bco-dmo.3327.3.1

Hudak CA, Stamieszkin K, Mayo CA (2023) North Atlantic right whale *Eubalaena glacialis* prey selection in Cape Cod Bay. Endanger Species Res 51:15–29. https://doi.org/10.3354/esr01240

Lenz J (2000) 1–Introduction. In: Harris R, Wiebe P, Lenz J et al (eds) ICES zooplankton methodology manual. Academic Press, London, pp 1–32. https://doi.org/10.1016/B978-012327645-2/50002-5

Liu Z, Le ND, Zidek JV (2011) An empirical assessment of Bayesian melding for mapping ozone pollution. Environmetrics 22(3):340–353. https://doi.org/10.1002/env.1054

Liu Y, Zidek JV, Trites AW et al (2016) Bayesian data fusion approaches to predicting spatial tracks: application to marine mammals. Ann Appl Stat 10(3):1517–1546. https://doi.org/10.1214/16-AOAS945

Mauchline J (1998) Advances in marine biology 33: the biology of calanoid copepods. Academic Press, London

Mayo CA, Marx MK (1990) Surface foraging behaviour of the North Atlantic right whale, *Eubalaena glacialis*, and associated zooplankton characteristics. Can J Zool 68(10):2214–2220. https://doi.org/10.1139/z90-308

Mayo CA, Letcher BH, Scott S (2001) Zooplankton filtering efficiency of the baleen of a North Atlantic right whale, *Eubalaena glacialis*. J Cetacean Res Manag 2:225–229. https://doi.org/10.47536/jcrm.vi.286

Mayo CA, Nichols OC, Bessinger MK et al (2004) Surveillance, monitoring and management of North Atlantic right whales in Cape Cod Bay and adjacent waters—2004. Final report submitted to the Commonwealth of Massachusetts, Division of Marine Fisheries. Technical report, Center for Coastal Studies, Provincetown, MA

Mitra A, Castellani C, Gentleman WC et al (2014) Bridging the gap between marine biogeochemical and fisheries sciences; configuring the zooplankton link. Prog Oceanogr 129:176–199. https://doi.org/10.1016/j.pocean.2014.04.025

Pitt MK, Shephard N (1999) Analytic convergence rates and parameterization issues for the Gibbs sampler applied to state space models. J Time Ser Anal 20(1):63–85. https://doi.org/10.1111/1467-9892.00126

Plourde S, Lehoux C, Johnson CL et al (2019) North Atlantic right whale (*Eubalaena glacialis*) and its food: (I) a spatial climatology of Calanus biomass and potential foraging habitats in Canadian waters. J Plankton Res 41(5):667–685. https://doi.org/10.1093/plankt/fbz024

Record N, Runge J, Pendleton D et al (2019) Rapid climate-driven circulation changes threaten conservation of endangered North Atlantic right whales. Oceanography 32(2):162–169. https://doi.org/10.5670/oceanog.2019.201

Richardson AJ (2008) In hot water: zooplankton and climate change. ICES J Mar Sci 65(3):279–295. https://doi.org/10.1093/icesjms/fsn028

Rundel CW, Schliep EM, Gelfand AE et al (2015) A data fusion approach for spatial analysis of speciated PM2.5 across time. Environmetrics 26(8):515–525. https://doi.org/10.1002/env.2369

Sorochan KA, Plourde S, Baumgartner MF et al (2021) Availability, supply, and aggregation of prey (*Calanus* spp.) in foraging areas of the North Atlantic right whale (*Eubalaena glacialis*). ICES J Mar Sci 78(10):3498–3520. https://doi.org/10.1093/icesjms/fsab200

Villejo SJ, Illian JB, Swallow B (2023) Data fusion in a two-stage spatio-temporal model using the INLA-SPDE approach. Spat Stat 54(100):744. https://doi.org/10.1016/j.spasta.2023.100744

Waktins W, Schevill W (1976) Right whale feeding and baleen rattle. J Mammal 57(1):58–66. https://doi.org/10.2307/1379512

West M, Harrison J (1997) Bayesian forecasting and dynamic models, 2nd edn. Springer, New York. https://doi.org/10.1007/b98971

Zhang H (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J Am Stat Assoc 99(465):250–261. https://doi.org/10.1198/016214504000000241

Zidek JV, Le ND, Liu Z (2012) Combining data and simulated data for space-time fields: application to ozone. Environ Ecol Stat 19(1):37–56. https://doi.org/10.1007/s10651-011-0172-1