Approximating spatial processes with too many knots degrades the quality of probabilistic predictions

Eric J. Ward[1] and Sean C. Anderson[2]

[1]Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle WA, 98112, USA

[2]Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, V6T 6N7, Canada

[*]corresponding author: eric.ward@noaa.gov

Abstract

1. Spatial and spatiotemporal models are increasingly used in ecology for a range of purposes, such as tracking population change, assessing species distributions, and modelling spatial processes. Many models, including generalized additive models (GAMs) and Gaussian random fields (fit via the Stochastic Partial Differential Equation [SPDE] approach), approximate spatial surfaces using basis functions defined over a set of knots or mesh vertices. A common assumption is that higher resolution approximations with more knots yield better predictions, but this is rarely tested.

2. We develop three case studies and a simulation to understand how mesh resolution affects prediction quality and parameter estimates. Using temperature and groundfish data from U.S. West Coast fish surveys, we developed spatial and spatiotemporal models of ocean temperature and fish biomass density. We fit models with varying mesh complexity and assessed prediction quality using log predictive density (log score) with k-fold cross-validation. We also compared SPDE approaches to GAMs.

3. Our temperature models showed that the finest scale meshes decreased log predictive density. As mesh complexity increased, estimated spatial range, spatial variance, and observation error declined. Groundfish models showed similar patterns: more vertices led to smaller estimated spatial range and variance parameters. For most species, mesh resolution had little effect on area-weighted biomass indices, but it did affect scale for some. Simulations showed that predictive density declined with increasing mesh complexity, especially under fine-scale spatial variability, high spatial variance, and low observation error. Root mean square error remained stable, indicating that degraded predictive densities stemmed from poorly calibrated uncertainty estimates rather than reduced accuracy of predictions.

4. Our work highlights that practitioners should not assume predictive performance always improves with increased spatial complexity. Selecting appropriate spatial complexity is expected to improve parameter estimation accuracy and derived quantities when out-of-sample prediction is a focus of inference.

Key words: Spatial, spatiotemporal, Stochastic Partial Differential Equation (SPDE), Integrated Nested Laplace Approximation (INLA), Gaussian random field, mesh, predictive accuracy, sdmTMB, cross-validation

1 Introduction

Recent advances in computational methods have revolutionized the estimation of spatial and spatiotemporal models in ecology and related fields (Banerjee & Fuentes 2012, Thorson & Kristensen 2024). These models address key questions, such as tracking species abundance over time and space, identifying environmental predictors of occurrence, modelling disease spread, and quantifying range shifts in distributions (e.g., Ross *et al.* 2012, Miller *et al.* 2013, Thorson & Barnett 2017). Application of these methods to large georeferenced datasets has become increasingly important as ecological and environmental challenges demand precise and scalable statistical tools.

A range of statistical approaches are available for modelling spatial processes. Examples include generalized additive models (GAMs) (Murase *et al.* 2009, Forney *et al.* 2012, Wood 2017), non-parametric approaches (Mutanga *et al.* 2012), and applications of Gaussian random fields (Latimer *et al.* 2009, Datta *et al.* 2016). Fitting these models to large datasets can be computationally challenging (Banerjee *et al.* 2004), necessitating the use of approximations or smoothing. One solution to approximation is the Stochastic Partial Differentiation Equation (SPDE) approach (Lindgren *et al.* 2011). The SPDE approach approximates a continuous Gaussian random field with a Matérn covariance function using a Gaussian Markov random field (GMRF). By representing the Gaussian random field as the solution to an SPDE, and solving it numerically using finite element methods over a triangulated mesh, the method yields a GMRF whose sparse precision matrix reflects conditional dependence only between neighbouring mesh vertices (Lindgren *et al.* 2011). This enables computationally efficient inference compared to modelling a full-rank covariance matrix. Like other approximations, the SPDE approach can be seen as a form of smoothing (Yue *et al.* 2014, Miller *et al.* 2020). Miller *et al.* (2020) provides an excellent summary of the recent SPDE literature, and also shows the link between these approaches and smoothing splines.

Advances by Rue *et al.* (2009) enabled the SPDE approach to be implemented with the Integrated Nested Laplace Approximation (INLA) software as a fast and efficient approximation to estimating high-dimensional Gaussian random fields with the Laplace approximation to the posterior distribution (Lindgren & Rue 2015). These models have been extended in many ways (Bakka *et al.* 2018) including the ability to fit geographic barriers (such as islands or lakes) (Bakka *et al.* 2019), include anisotropic covariance functions (Haskard 2007), and fit non-Gaussian spatial fields (Wallin & Bolin 2015). Related software packages to increase the accessibility of these models include inlabru (Bachl *et al.* 2019) and PointedSDMs (Mostert & O'Hara 2023). As an alternative to INLA, the SPDE approach has been integrated with Template Model Builder (TMB, Kristensen *et al.* 2016) to enable fast marginal maximum likelihood estimation, also via the Laplace approximation. Software to implement the TMB-SPDE approach includes VAST (Thorson 2019), sdmTMB (Anderson *et al.* 2024), and tinyVAST (Thorson *et al.* 2024) and comparisons between the INLA- and TMB-based approaches can be found in Osgood-Zimmerman & Wakefield (2022).

Constructing spatiotemporal models with INLA or related software involves many of the same considerations as with simpler non-spatial models. For example, analysts frequently ask: "Which covariates are most important?" and "Does my model do a satisfactory job of making predictions?" Spatial or spatiotemporal models may include additional considerations, such as specifying the type and complexity of spatial or spatiotemporal variation (Anderson & Ward 2019). Because of the flexibility in these approaches, choices about how spatial processes are modelled can also affect inference about covariates. For example, fitting an overly complex spatial process may mask effects of important covariates (Dambly *et al.* 2023). One of the more complicated and important decisions—particularly for models implementing the SPDE approach—lies in specifying the dimensionality of the basis functions underlying the spatial fields (Lindgren & Rue 2015).

A critical yet often overlooked step in SPDE-based modelling is constructing the finite element spatial mesh or basis function representation—a triangular discretization that approximates the spatial field (Figure 1). Functionally, the mesh serves two purposes: (1) it defines the sparse matrices required to implement the SPDE approach and (2) it defines how the spatial values estimated at the vertices (also "nodes" or "knots") are projected to coordinates where data are observed or predictions are made (Lindgren *et al.* 2011, Lindgren & Rue 2015). Typically, these vertex values are projected with bilinear interpolation along the mesh triangles (Lindgren *et al.* 2011) but piecewise-constant approximations have also been used (e.g., Thorson *et al.* 2015). Decisions regarding the mesh, such as its boundary shape, maximum triangle edge length, and minimum triangle angles, significantly impact model accuracy and computational efficiency (Gómez-Rubio 2020). While increasing mesh resolution generally improves the approximation of spatial processes, it may also introduce overfitting, computational challenges, and inflated prediction uncertainty. Many previous applications and guidance around INLA have suggested that increasing the resolution of the mesh improves predictive accuracy (Godana *et al.* 2019, Wilson 2021, Williamson *et al.* 2021, Thorson 2019). A much smaller number of studies have noted that finer meshes do not always translate into better performance (Huang *et al.* 2017) and we are only aware of a small number of analyses examining the importance of mesh construction (Righetto *et al.* 2020, Røste 2020, Dambly *et al.* 2023).

The objective of this paper is to examine whether higher resolution spatial approximations translate to better probabilistic predictions, and to provide general recommendations for identifying models with good out-of-sample predictive performance. We develop three case studies based on a publicly available dataset of marine fishes in the Northeast Pacific Ocean. First, we develop a spatial model of ocean temperature from a single year, investigating the effect of mesh resolution on predictive performance and parameter estimation. Second, we develop spatial models of groundfish density for three species, examining the effect of mesh resolution on noisy data. Third, we extend the groundfish models to derive area-weighted population biomass indices for 27 species, and examine effects of mesh resolution on trend, scale, and uncertainty of biomass estimates. Finally, we conduct a simulation experiment to diagnose the causes of the performance we identify in our case studies. All data for our analysis are publicly available, and code and data to reproduce our analysis are available on GitHub (https://github.com/ericward-noaa/how-many-knots) and Zenodo (on publication).

2 Methods

2.1 Data

To evaluate the effect of mesh resolution on the quality of probabilistic predictions of models using the SPDE approach, we used a publicly available scientific survey of marine fishes in the Northeast Pacific Ocean (https://www.nwfsc.noaa.gov/data). The West Coast Groundfish Bottom Trawl Survey (WCGBTS) data is from a fishery-independent survey collected annually over a 21-year period, 2003–2023 (no survey in 2020), covering the California Current ecosystem off the U.S. west coast (Keller *et al.* 2017). The survey is designed to collect information to inform estimate the trends in abundance and collect biological data including size and age compositions of commercially and recreationally targeted groundfish species found in near-bottom habitats. The fundamental methods of the survey, sampling intensity, gear, seasonal and spatial coverage, and random stratified design have remained relatively constant. In addition to collecting biological samples, the survey also measures environmental variables, such as water temperature at the depth of the net. Importantly, this survey shares similar characteristics to many other bottom trawl survey datasets used to monitor marine fishes around the world.

2.2 Modelling ocean bottom temperature

As a first case study, we constructed a spatial model of bottom temperature from the WCGBTS survey. To keep the initial application simple and avoid complications introduced by spatiotemporal variation, we used a single year of data (2018, n = 701 observations). Because the survey spans nearly five months (19 May–15 October), we included day of year as a quadratic predictor, to allow for seasonal changes in mean temperature. Extending notation used in linear mixed effect models, our model can be expressed as

$$u_i = b_0 + b_1 x_i + b_2 x_i^2 + \omega_i,$$

where ω_i is the estimated spatial field at the location of the ith observation and is modelled as $\boldsymbol{\omega} \sim \mathrm{MVN}(\mathbf{0}, \Sigma_\omega)$, b_0 is the intercept, x_i is the day of year, b_1 and b_2 are estimated coefficients, and u_i is the expected value incorporating fixed and random effects. Finally, we assumed error in the observed temperature values followed a Gaussian distribution. We varied mesh resolution by adjusting the cutoff distance—the distance below which two points are treated as a single vertex–because previous research has shown this parameter has the greatest influence on how the mesh affects model inference (Righetto *et al.* 2020). There is an inverse relationship between cutoff distance and mesh resolution that varies slightly dataset to dataset; we varied cutoff distances between 8–500 km (corresponding to 24–727 vertices for this example). For each mesh resolution, we performed k-fold cross validation, using 10 folds with points assigned randomly to each.

2.3 Groundfish spatial models

As a second example, we constructed a more complicated case study focused on modelling a biological response—groundfish density—from samples taken from the WCGBTS survey, 2003–2023. The primary objective of this case study is to examine whether mesh resolution impacts predictive performance when modelling a non-Gaussian response variable subject to relatively high levels of sampling or observation

error. We selected three species that are well observed by the survey, but also represent a range of occurrence rates and spatial ranges: sablefish (*Anoplopoma fimbria*), petrale sole (*Eopsetta jordani*), and arrowtooth flounder (*Atheresthes stomias*).

We applied the same form of the spatial model to all species so that for observation i,

$$g(u_i) = \boldsymbol{X}_i \boldsymbol{b} + \omega_i,$$

where for convenience the fixed effect design matrix and coefficients are collapsed into $\boldsymbol{X}$ and $\boldsymbol{b}$ respectively, and $\boldsymbol{\omega}$ represents spatial variation. For simplicity, we modelled catch per unit effort (CPUE, in kg/km^2) with a Tweedie distribution (allowing for both zeros and positive catches), where $g()$ represents a log link function. Fixed effect covariates included in the model were year (as a factor), allowing the average density to vary by year, and a quadratic effect of log(depth). Because of the shape of the California Current ecosystem, we estimated anisotropic spatial covariance, allowing effects of latitude and longitude to be asymmetric (Fuglstad *et al.* 2015).

We expanded on our initial case study and developed custom meshes. First, we used a non-convex hull to define a boundary, and incorporated the boundary into each mesh, along with a specified cutoff distance (12–175 km). We used 5% offsets for the mesh and boundary and set the maximum inner and outer triangle edges at 100 and 500 km, respectively. Meshes varied slightly by species, but the range of cutoff values resulted in meshes ranging in size from 51–893 vertices. High-resolution meshes were included as an upper extreme, though many applications of the SPDE approach to datasets with similar dimensions have used 500–600 knots (e.g., Thorson 2019, Tolimieri *et al.* 2020).

For each combination of species and mesh, we performed k-fold cross validation, using 10 folds and random fold assignments, as in our temperature case study. To reduce noise to the particular set of folds chosen, we repeated the 10-fold cross validation across five random seed values.

2.4 Groundfish biomass indices

As a third example, we examined the effect of mesh resolution on biomass indices generated from the WCGBTS survey data. A common use of bottom trawl survey data is to estimate annual indices of biomass, to be used in downstream integrated population models (stock assessments in fisheries). We used data from the same 27 species (Table S1) to generate standardized indices of biomass (Thorson *et al.* 2015, Thorson & Ward 2013). For each species, we constructed three meshes with cutoff distances of 15, 50, and 100 km, using the same process as described in our second case study. For consistency with status quo population models, catch per unit effort (CPUE, in kg/km^2) was used as a response variable, and because observations are both zero-inflated and skewed, we modelled CPUE as a two-part hurdle or delta model, with a binomial model used for encounter probability and a Gamma distribution used for positive catches (Pennington 1983). We only included year effects (as a factor variable); however, because of the hurdle response, two sets of fixed effect coefficients were estimated—one for each linear predictor. For random effects, we included both spatial fields and spatiotemporal fields (assumed independent by year), with a shared spatial range (Anderson *et al.* 2024). Standardized indices of biomass were generated by making model predictions on the survey design grid for each year (12000 3.7 km by 2.8 km cells of equal area; Keller *et al.* 2017) and then summing across grid cells within a year to get annual indices of total biomass.

2.5 Simulation testing

Finally, we tested the effects of increasing mesh complexity on the quality of probabilistic predictions using simulated data. We did this for two reasons: (1) to confirm that the patterns observed fitting to temperature and fish biomass data could be replicated with simulated data, and (2) to identify the parameter space where problems were most pronounced. We repeatedly simulated 1,000 observations from a spatial Gaussian random field with Gaussian observation error across a grid of parameter values for the spatial standard deviation, spatial range, and observation error standard deviation, and with a mean of zero. We then performed 10-fold cross validation and calculated the log predictive density based on the left-out observations and the predictive root mean squared error (RMSE) on the left-out known true values. In addition to predictive skill, we evaluated bias and uncertainty in parameter estimates.

2.6 Estimation

We fit all models in R (R Core Team 2024) using the package sdmTMB (Anderson *et al.* 2024), which combines the SPDE approach for spatial models with the speed of maximum marginal likelihood estimation using Template Model Builder (TMB, Kristensen *et al.* 2016). TMB leverages automatic differentiation to quickly converge on parameter estimates, and it applies the Laplace approximation to integrate out random effects, including spatial or spatiotemporal components. We used the $\mathrm{sdmTMB_cv}()$ function to implement k-fold cross validation for all examples. Model convergence was assessed using the $\mathrm{sanity}()$ function in sdmTMB; we only retained models that successfully converged (positive-definite Hessian matrix, maximum absolute log likelihood gradient < 0.001).

To test whether the observed patterns were consistent across modelling platforms, we repeated the fitting and cross validation with mgcv (Wood 2017) and inlabru (Bachl *et al.* 2019) for the temperature case study. For mgcv, we tested a sequence of basis dimensions (k in mgcv) ranging from three to 167 (approximately the number of SPDE mesh vertices divided by three). A major difference between estimation in inlabru (Bachl *et al.* 2019) and sdmTMB (Anderson *et al.* 2024) is that the former implicitly includes a penalized complexity (PC) prior on the Matérn parameters (Fuglstad *et al.* 2019), while in sdmTMB the PC prior is optional. To test whether the choice of PC prior affected results, we repeated the analysis with a sequence of possible PC bivariate priors on the spatial range and spatial variance.

2.7 Quantifying performance

To quantify the predictive performance of models with different mesh configurations, we calculated the log likelihood of the training and test data in each fold, as these quantities incorporate both accuracy and precision of estimates. We refer to the likelihood associated with the test data as the log predictive density, which is also known as the log-score (Draper 1995); test log-likelihood, predictive log-likelihood, or test log-predictive (Deshpande *et al.* 2024); and is related to expected log predictive density in the Bayesian literature (Vehtari *et al.* 2017). For the simulated data, we also calculated the root mean square error (RMSE) of the predictions against the true values; this provided a measure of predictive performance that ignored prediction uncertainty.

3 Results

3.1 Temperature case study

Our spatial models of ocean temperature (Figure 1a) demonstrated that average log density of training data generally increased with mesh resolution (Figure 2). However, average log density for out-of-sample testing (log predictive density) data peaked at an intermediate mesh complexity of about 400 vertices and declined thereafter (Figure 2). For higher resolution meshes, our results demonstrate a decline in predictive performance for out-of-sample data. We observed similar patterns between mesh complexity or basis dimensions and log predictive density when fitting our models with TMB, INLA, or mgcv (Figure 2).

Examining the relationship between estimated model parameters and vertices in the mesh, we find that the intercept, spatial range, and spatial standard deviation were poorly estimated at coarser meshes (under $\approx$ 200 vertices) (Figure 3). Beyond a peak at approximately 100 vertices, the spatial range declined with increasing mesh complexity. The spatial variance and observation error variance both decreased with increasing mesh complexity (Figure 3).

3.2 Groundfish spatial models

Results from our models of groundfish density are largely consistent with the pattern described in our first case study. For our groundfish density (CPUE) results, all three species showed evidence of consistent increases in log predictive density calculated for the training data as the mesh resolution increased (Figure 4). All three species also showed similar patterns of declining log predictive density at the finest mesh resolutions. Arrowtooth flounder and petrale sole show peak predictive performance occurring at moderate mesh complexity ($\approx$ 600 mesh vertices), while sablefish shows a peak at a coarser resolution ($\approx$ 200–300 mesh vertices) and declines strongly thereafter (Figure 4).

3.3 Groundfish biomass indices

For the 27 species to which we tried fitting spatiotemporal models of groundfish biomass, 10 passed convergence criteria across all three cutoff distances (15, 50, 100 km; Figure S3). Most species followed a pattern similar to that observed for lingcod or petrale sole, where the scale of biomass indices increased as meshes became more coarse (Figure 5); however for some species—including shortspine thornyhead—this pattern was reversed. Across species, the average standard errors of biomass estimates were qualitatively similar and did not depend on cutoff distances (Figure S4). For most species, there were high correlations between biomass indices produced with different cutoff distances—though for several rockfish species (canary rockfish, stripetail rockfish), there appeared to be less agreement between indices (Figure S4).

3.4 Simulation testing

We were able to replicate the pattern of declining log predictive density with cross validation for simulated data (Figure 6). The pattern was strongest given a combination of low observation error, low spatial range, and high spatial variance (Figure 6c, Figure S5). However, our simulations showed that this was not due to the mean prediction becoming over-fit to the data given that predictive RMSE against the true values continued to improve at higher mesh complexity (Figure S6). The pattern appeared to be driven by decreasing

estimated observation error at higher mesh complexities (Figure S7) combined with an underestimated spatial variance (Figure S9). At some point, the observation error was estimated at a small enough value that the Gaussian log likelihood started penalized the lack of fit to the data resulting in decreasing log predictive density with increasing mesh complexity.

4 Discussion

Our results demonstrate that increasing spatial complexity, whether through finer SPDE meshes or higher basis dimensions in GAMs, does not universally improve the quality of probabilistic predictions. Across case studies and modeling frameworks (TMB, INLA, and mgcv), we consistently observed that in-sample fit improved with increasing spatial resolution, but out-of-sample predictive performance (as measured by log predictive density) often peaked at intermediate levels of complexity and declined thereafter. This pattern was especially pronounced when spatial variance was high, spatial range was small, and observation error was low. These findings highlight a general principle applicable across modeling approaches: overfitting in the latent spatial process can lead to poorly calibrated uncertainty and degraded probabilistic predictions, even when point estimates appear accurate. Though several examples and best practices papers have been written to demonstrate how to construct models or specify hyperparameters (Lindgren & Rue 2015, Gómez-Rubio 2020), general advice on specifying the spatial resolution of these approaches has remained limited because it varies dataset to dataset. As such, we recommend that a key step in the application of these models is to assess predictive performance using cross-validation and explore sensitivity to spatial complexity. These steps are particularly critical when interpreting fixed effects, spatial parameters, or uncertainty.

4.1 Benefits and limitations of high spatial resolution

In some situations, fitting SPDE models with higher resolution meshes is advantageous. Finer meshes can provide a better approximation of the latent spatial process by capturing smaller-scale variation and improving model fit to training data. Consistent with previous work (Røste 2020), our results show that increasing mesh resolution up to a point stabilizes parameter estimates. However, as resolution increases, the spatial fields gain flexibility and may begin to absorb variability that could otherwise be explained by fixed effects. This redistribution of variability can obscure the relationships between covariates and the response, particularly when the scale of the covariates aligns poorly with the scale of the mesh. For studies focused on identifying precise relationships between environmental covariates and species' occupancy or density, we recommend exploring a range of mesh resolutions with cross validation, as finer meshes may not always be most appropriate. For applications related to index standardization, where the goal is often to derive area-weighted annual indices of abundance or biomass, fitting the highest resolution meshes that are computationally feasible is reasonable for most species. A caveat from our results is that for a small number of species examined, the choice of mesh resolution did impact the scale and uncertainty of index estimates.

Aside from predictive performance, there are several potential downsides to fitting models with high resolution meshes. If latent spatial variables have finer spatial variation than the measured covariates, spatial confounding bias is introduced (Paciorek 2010). In these cases, fitting a high resolution meshes lets the model fit to this finer scale spatial variation potentially exasperating the isssue leading to biased

fixed effects and an understimation of uncertainty (Paciorek 2010). This has also been illustrated with point process models where overly complex spatial fields may dominate potential fixed effects of interest, making coefficients for the latter smaller in magnitude (Dambly *et al.* 2023). Second, our results highlight a trade-off between mesh complexity and estimated spatial parameters. As the spatial field becomes more complex, it can better fit localized spatial variation (otherwise explained by fixed effects, observation error, or other sources of variability), resulting in smaller estimates of spatial range. As the spatial range becomes smaller, variability may be redistributed into the spatiotemporal variance component.

We found that log predictive density usually decreased beyond an intermediate mesh complexity, but not necessarily RMSE or other metrics that ignore the uncertainty in the observation error when predicting. Log predictive density, or the log score (Draper 1995), reflects the likelihood used when fitting the model to the observed data and is a commonly used metric to quantify model predictive performance when some other domain-specific metric is not available. Log predictive density is also closely tied to the commonly used metric expected log predictive density (ELPD) in the Bayesian literature (Vehtari *et al.* 2017). In cases where a modeller is only interested in the mean prediction and not the probabilistic prediction, the issues with high mesh complexity demonstrated here may be less relevant.

Our results suggests the phenomenon of declining prediction quality at high mesh complexity is caused by an underestimation of observation error uncertainty (given model predictions) in cases of low range (high wiggliness), high spatial variance, and low observation error. However, exactly what "high"' and "low" values need to be to exacerbate the phenomenon in applied situations is not known. The work of Paciorek (2010) suggests that "high" wiggliness is defined as finer spatial variability in the latent covariates than the observed covariates. Unfortunately, it is not always obvious when this condition is met—particularly given that the latent effects are by definition not observed. Our results suggest the effect is apparent across both environmental and biological data—modelled with a variety of likelihoods and with spatial and possibly spatiotemporal latent effects and with isotropic or anisotropic correlation—and across software and statistical frameworks. We have no reason to expect the datasets used in our case studies are unique, but analysts should check with their own datasets and study systems.

4.2 Methodological recommendations

While there are several alternative ways to measure model performance, we recommend using out-of-sample methods including k-fold cross-validation to quantify predictive accuracy across a range of meshes if computationally feasible. Tools such as AIC (Akaike 1973) have also been used to discriminate among these types of models; however, AIC is known to be problematic when used as a metric to compare mixed effects models (Harrison *et al.* 2018) or as a tool to estimate spatial complexity (Velasco & González-Salazar 2019). While WAIC is available in INLA, higher resolution meshes generally correspond to better WAIC values (Røste 2020). Spatially blocked or spatially buffered leave-one-out approaches with blocks or buffers larger than or roughly equivalent to the range, respectively, are likely to provide the most accurate measures of predictive performance (Roberts *et al.* 2017). Experiments we ran with spatially blocked cross-validation showed similar results to the random k-fold cross-validation shown here (not shown). Thus, we focussed random cross-validation to avoid the complexity of showing results across multiple block sizes.

Comparing the performance of models with alternative spatial complexity introduces additional computational burdens, both in performing cross-validation, and comparing models with different numbers of

knots. This challenge is not unique to models using the INLA-SPDE approach. Selecting both the location and number of knots is a general problem for knot-based spatial models. Other examples of spatial statistics models that have encountered this issue include machine learning methods (reviewed in Banerjee *et al.* 2012), Gaussian Process models (Garton *et al.* 2020) and P-splines (Ruppert 2002). Identifying a range of knot values that yield good predictive performance will be specific to individual datasets (e.g., scale of latent spatial variation, magnitude of observation error, number of observations, type of response). The optimal mesh resolution for one species may be informative for similar species from the same dataset—potentially reducing the computational overhead. Though some work has suggested vertices be chosen relative to the estimated spatial range of a dataset (Røste 2020), this advice creates a somewhat circular framework in that the range is usually known only after a reasonable spatial model is constructed.

If spatial regression modelling tools are being used to compare models with different covariates, or to compare spatial parameters (ranges, variances) across models, it is critical to ensure that model performance is not confounded by arbitrary choices in mesh resolution. We recommend fitting models across a range of mesh resolutions to identify those that balance goodness of fit to the training data with predictive performance on out-of-sample data. Where spatial parameters or covariate effects are being compared across models, consistency across mesh specifications—or at a minimum, sensitivity analyses to mesh complexity—should be used to avoid drawing spurious conclusions based on modelling artifacts. While such sensitivities add computational overhead, they will ensure the robustness and interpretability of statistical inference from these spatial models.

5 Acknowledgements

We express our gratitude to Dave Miller, Andy Seaton, Ole Shelton, and Chantel Wetzel for helpful comments that improved our analysis and presentation of results.

6 Figures

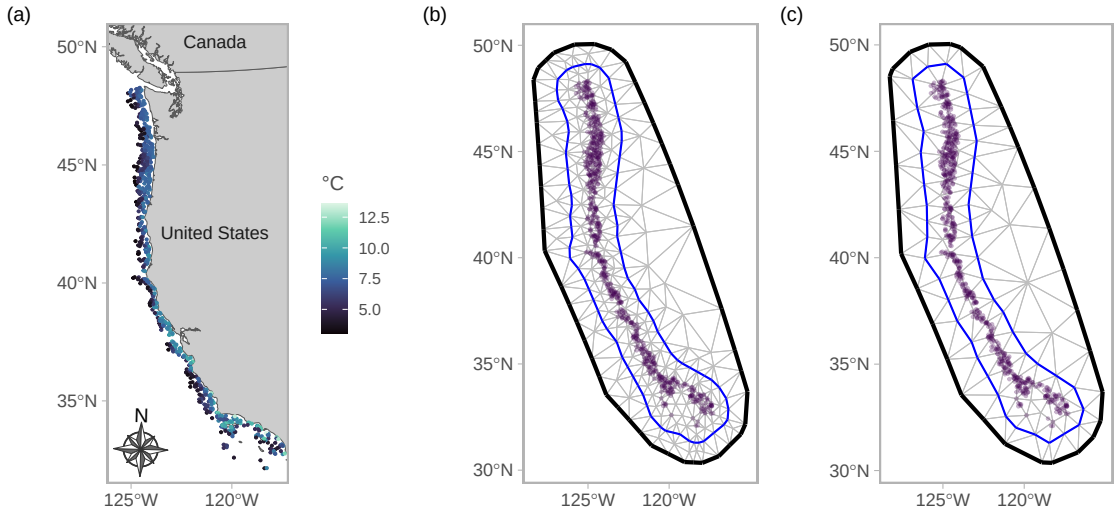

Figure 1: (a) Sampled ocean temperature from the 2018 WCGBTS survey. Sample locations further offshore are typically colder and at greater depth. Panels (b) and (c) represent example meshes constructed using the 2018 WCGBTS survey data. Cutoff values of 20 and 80 km are used, translating into 394 and 110 knots (vertices) respectively. The blue polygon illustrates the boundary between the inner and outer mesh where the outer mesh has a larger maximum triangle edge length.

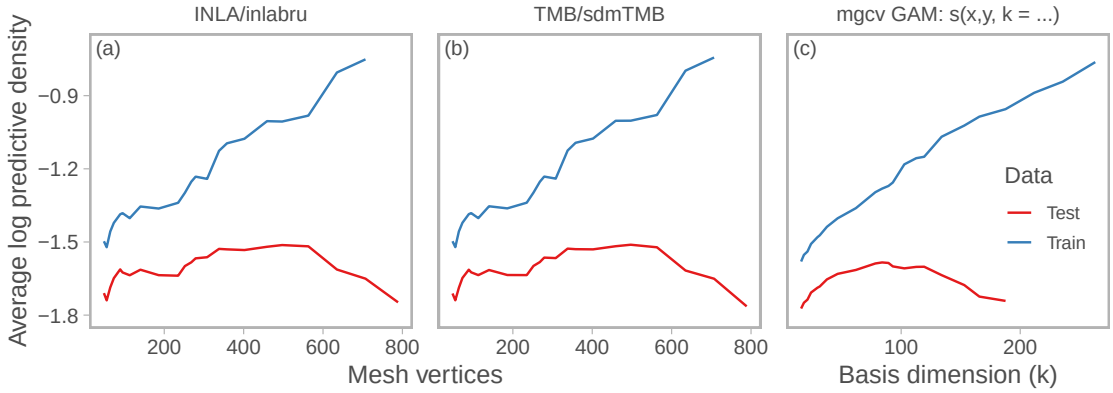

Figure 2: Average pointwise log predictive density for in-sample (train) and out-of-sample (test) data within 10-fold cross validation for the bottom temperature example across spatial approximation complexity. Panels (a) and (b) illustrate results for INLA and TMB across number of vertices in the SPDE finite-element mesh. Panel (c) illustrates a generalized additive model (GAM) fit with mgcv using a penalized bivariate thin-plate regression spline on the spatial coordinates across a sequence of basis dimensions (k).

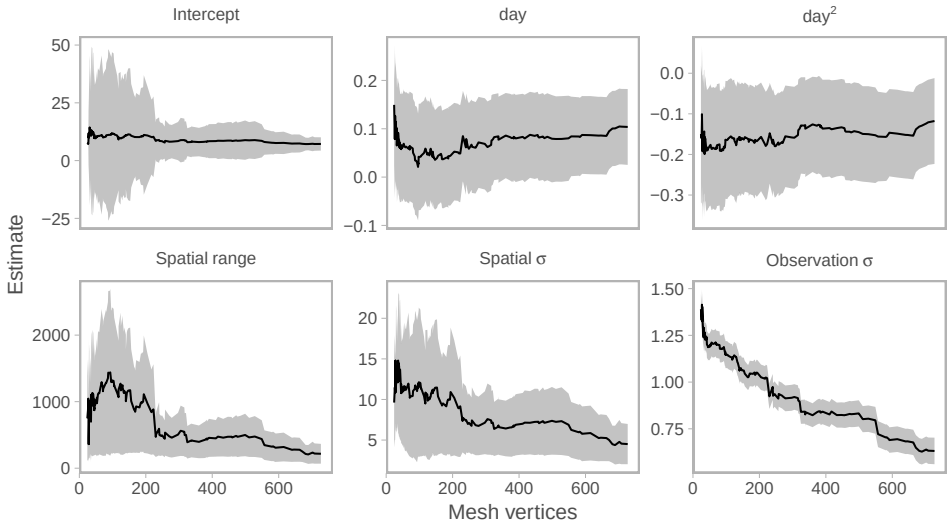

Figure 3: Estimated fixed effect parameters (top row) and spatial parameters (bottom row) as a function of mesh vertices (more vertices translates to a cutoff distance in creating the mesh). The panels "day" and "day^2" represent the linear and quadratic effects of day of year; σ represents the standard deviation. Solid lines represent mean estimates, and shaded ribbons represent 95% confidence intervals.

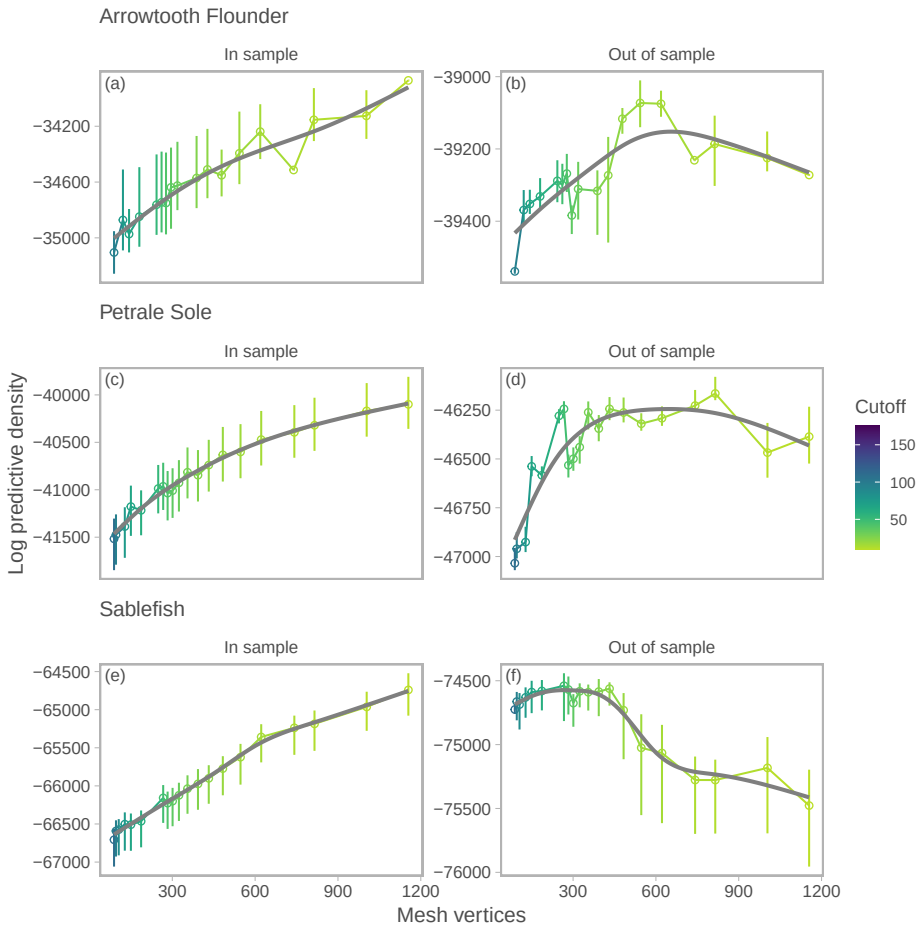

Figure 4: Log predictive density from modelling spatial biomass density of three example groundfish species across a range of mesh complexities. Coloured dots and lines indicate means and ranges of log predictive densityacross five iterations of 10-fold cross validation. Thick grey lines represent generalized additive model smoothers to highlight the pattern. "Cutoff" refers to the minimum triangle edge length in the mesh—higher cutoffs correspond to fewer mesh vertices.

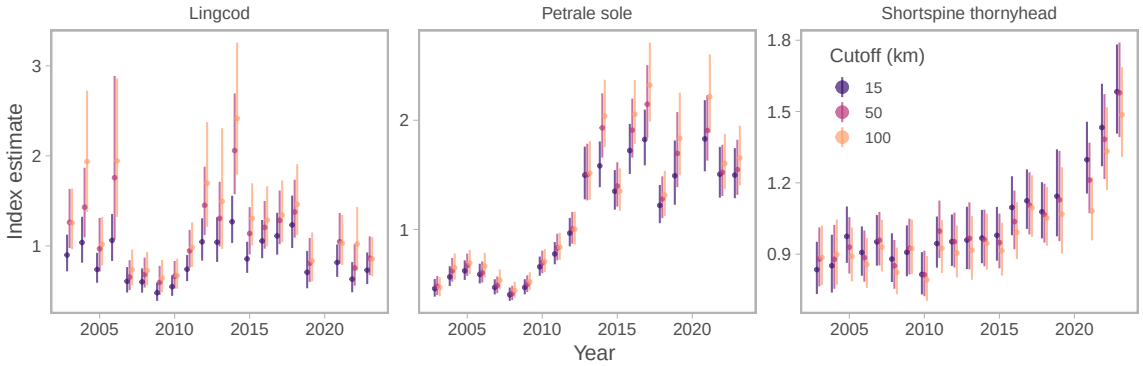

Figure 5: Time series of biomass estimates estimated for three species, across three mesh resolutions (larger cutoff distances result in coarser meshes). Lingcod indices show a pattern of relatively large increases in scale with increasing cutoff distances, petrale sole indices show a relatively small change in scale with increasing cutoff distances, and shortspine thornyhead indices show an opposite pattern where scale decreases with increasing cutoff distances.

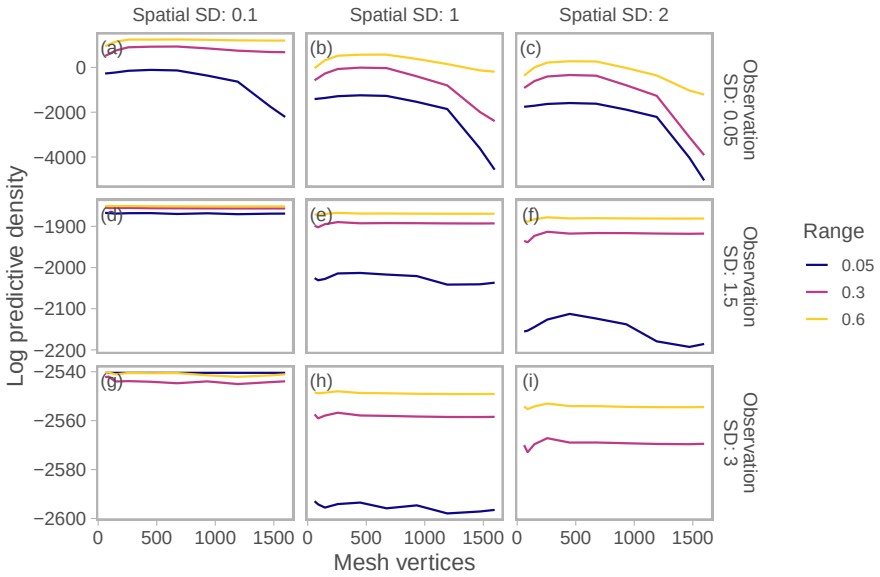

Figure 6: Log predictive density of models fit to simulated Gaussian random fields across a range of mesh vertices (x-axis), spatial standard deviations (columns), observation error standard deviations (rows), and spatial ranges (colours). The scale of the spatial coordinates ranged from zero to one. Log predictive density declines at high mesh complexities in cases of comparatively low range (high wiggliness), high spatial variance, and low observation error.

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. B. Petrov & F. Csaki, eds., *Second International Symposium on Information Theory*, pp. 267–281. Akademiai Kiado, Budapest.

Anderson, S.C. & Ward, E.J. (2019) Black swans in space: modeling spatiotemporal processes with extremes. *Ecology*, **100**, e02403. https://dx.doi.org/10.1002/ecy.2403.

Anderson, S.C., Ward, E.J., English, P.A., Barnett, L.A.K. & Thorson, J.T. (2024) sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. *In press at Journal of Statistical Software*. https://dx.doi.org/10.1101/2022.03.24.485545.

Bachl, F.E., Lindgren, F., Borchers, D.L. & Illian, J.B. (2019) inlabru: an r package for bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, **10**, 760–766. https://dx.doi.org/10.1111/2041-210x.13168.

Bakka, H., Rue, H., Fuglstad, G.A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D. & Lindgren, F. (2018) Spatial modeling with R-INLA: A review. *WIREs Computational Statistics*, **10**, e1443. https://dx.doi.org/10.1002/wics.1443.

Bakka, H., Vanhatalo, J., Illian, J., Simpson, D. & Rue, H. (2019) Non-stationary Gaussian models with physical barriers. *arXiv*. https://dx.doi.org/10.48550/arXiv.1608.03787.

Banerjee, A., Dunson, D.B. & Tokdar, S.T. (2012) Efficient Gaussian process regression for large datasets. *Biometrika*, **100**, 75–89. https://dx.doi.org/10.1093/biomet/ass068.

Banerjee, S., Carlin, B. & Gelfand, A. (2004) *Hierarchical Modeling and Analysis of Spatial Data*. Chapman & Hall/CRC.

Banerjee, S. & Fuentes, M. (2012) Bayesian modeling for large spatial datasets. *WIREs Computational Statistics*, **4**, 59–66. https://dx.doi.org/10.1002/wics.187.

Dambly, L.I., Isaac, N.J.B., Jones, K.E., Boughey, K.L. & O'Hara, R.B. (2023) Integrated species distribution models fitted in inla are sensitive to mesh parameterisation. *Ecography*, **2023**. https://dx.doi.org/10.1111/ecog.06391.

Datta, A., Banerjee, S., Finley, A.O. & Gelfand, A.E. (2016) Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, **111**, 800–812. https://dx.doi.org/10.1080/01621459.2015.1044091.

Deshpande, S.K., Ghosh, S., Nguyen, T.D. & Broderick, T. (2024) Are you using test log-likelihood correctly?

Draper, D. (1995) Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B (Methodological)*, **57**, 45–97.

Forney, K.A., Ferguson, M.C., Becker, E.A., Fiedler, P.C., Redfern, J.V., Barlow, J., Vilchis, I.L. & Ballance, L.T. (2012) Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. *Endangered Species Research*, **16**, 113–133. https://dx.doi.org/10.3354/esr00393.

Fuglstad, G.A., Lindgren, F., Simpson, D. & Rue, H. (2015) Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy. *Statistica Sinica*, **25**, 115–133.

Fuglstad, G.A., Simpson, D., Lindgren, F. & Rue, H. (2019) Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, **114**, 445–452. https://dx.doi.org/10.1080/01621459.2017.1415907.

Fuglstad, G.A., Simpson, D., Lindgren, F. & Rue, H. (2018) Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, **114**. https://dx.doi.org/10.1080/01621459.2017.1415907.

Garton, N., Niemi, J. & Carriquiry, A. (2020) Knot selection in sparse Gaussian processes with a variational objective function. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, **13**, 324–336. https://dx.doi.org/10.1002/sam.11459.

Godana, A.A., Mwalili, S.M. & Orwa, G.O. (2019) Dynamic spatiotemporal modeling of the infected rate of visceral leishmaniasis in human in an endemic area of Amhara regional state, Ethiopia. *PLOS ONE*, **14**, e0212934. https://dx.doi.org/10.1371/journal.pone.0212934.

Gómez-Rubio, V. (2020) *Bayesian inference with INLA*. Chapman & Hall/CRC Press.

Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E.D., Robinson, B.S., Hodgson, D.J. & Inger, R. (2018) A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, **6**, e4794. https://dx.doi.org/10.7717/peerj.4794.

Haskard, K.A. (2007) *An Anisotropic Matern Spatial Covariance Model: REML Estimation and Properties*. Thesis.

Huang, J., Malone, B.P., Minasny, B., McBratney, A.B. & Triantafilis, J. (2017) Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping. *The Science of the Total Environment*, **609**, 621–632. https://dx.doi.org/10.1016/j.scitotenv.2017.07.201.

Keller, A.A., Wallace, J.R. & Methot, R.D. (2017) The Northwest Fisheries Science Center's West Coast Groundfish Bottom Trawl Survey: History, Design, and Description. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-136. DOI: 10.7289/V5/TM-NWFSC-136, Seattle, WA.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H. & Bell, B.M. (2016) TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, **70**, 1–21. https://dx.doi.org/10.18637/jss.v070.i05.

Latimer, A.M., Banerjee, S., Jr, H.S., Mosher, E.S. & Jr, J.A.S. (2009) Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern United States. *Ecology Letters*, **12**, 144–154. https://dx.doi.org/10.1111/j.1461-0248.2008.01270.x.

Lindgren, F. & Rue, H. (2015) Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, **63**, 1–25. https://dx.doi.org/10.18637/jss.v063.i19.

Lindgren, F., Rue, H. & Lindström, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**, 423–498. https://dx.doi.org/10.1111/j.1467-9868.2011.00777.x.

Miller, D.L., Burt, M.L., Rexstad, E.A. & Thomas, L. (2013) Spatial models for distance sampling data: recent developments and future directions. *Methods in Ecology and Evolution*, **4**, 1001–1010. https://dx.doi.org/10.1111/2041-210X.12105.

Miller, D.L., Glennie, R. & Seaton, A.E. (2020) Understanding the stochastic partial differential equation approach to smoothing. *Journal of Agricultural, Biological and Environmental Statistics*, **25**, 1–16. https://dx.doi.org/10.1007/s13253-019-00377-z.

Mostert, P.S. & O'Hara, R.B. (2023) PointedSDMs: An R package to help facilitate the construction of integrated species distribution models. *Methods in Ecology and Evolution*, **14**, 1200–1207. https://dx.doi.org/10.1111/2041-210x.14091.

Murase, H., Nagashima, H., Yonezaki, S., Matsukura, R. & Kitakado, T. (2009) Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: a case study in Sendai Bay, Japan. *ICES Journal of Marine Science*, **66**, 1417–1424. https://dx.doi.org/10.1093/icesjms/fsp105.

Mutanga, O., Adam, E. & Cho, M.A. (2012) High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, **18**, 399–406. https://dx.doi.org/10.1016/j.jag.2012.03.012.

Osgood-Zimmerman, A. & Wakefield, J. (2022) A statistical review of Template Model Builder: a flexible tool for spatial modelling. *International Statistical Review*, **91**, 318–342. https://dx.doi.org/10.1111/insr.12534.

Paciorek, C.J. (2010) The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, **25**. https://dx.doi.org/10.1214/10-STS326.

Pennington, M. (1983) Efficient estimators of abundance, for fish and plankton surveys. *Biometrics*, **39**, 281. https://dx.doi.org/10.2307/2530830.

R Core Team (2024) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Righetto, A.J., Faes, C., Vandendijck, Y. & Ribeiro, P.J. (2020) On the choice of the mesh for the analysis of geostatistical data using R-INLA. *Communications in Statistics - Theory and Methods*, **49**, 203–220. https://dx.doi.org/10.1080/03610926.2018.1536209.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. & Dormann, C.F. (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, **40**, 913–929. https://dx.doi.org/10.1111/ecog.02881.

Ross, B.E., Hooten, M.B. & Koons, D.N. (2012) An accessible method for implementing hierarchical models with spatio-temporal abundance data. *PLOS ONE*, **7**, e49395. https://dx.doi.org/10.1371/journal.pone.0049395.

Rue, H., Martino, S. & Chopin, N. (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**, 319–392. https://dx.doi.org/10.1111/j.1467-9868.2008.00700.x.

Ruppert, D. (2002) Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, **11**, 735–757.

Røste, J. (2020) *The importance of mesh resolution when using the SPDE approach.* Ph.D. thesis, Department of Mathematics, Norwegian University of Science and Technology, Oslo, Norway.

Thorson, J.T. (2019) Guidance for decisions using the vector autoregressive spatio-temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*, **210**, 143–161. https://dx.doi.org/10.1016/j.fishres.2018.10.013.

Thorson, J.T., Anderson, S.C., Goddard, P. & Rooper, C.N. (2024) tinyVAST: R package with an expressive interface to specify lagged and simultaneous effects in multivariate spatio-temporal models. *arXiv*. https://dx.doi.org/10.48550/arXiv.2401.10193.

Thorson, J.T. & Barnett, L.A.K. (2017) Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. *ICES Journal of Marine Science*, **74**, 1311–1321. https://dx.doi.org/10.1093/icesjms/fsw193.

Thorson, J.T. & Kristensen, K. (2024) *Spatio-Temporal Models for Ecologists.* CRC Press, Boca Raton, FL.

Thorson, J.T., Shelton, A.O., Ward, E.J. & Skaug, H.J. (2015) Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for west coast groundfishes. *ICES Journal of Marine Science*, **72**, 1297–1310. https://dx.doi.org/10.1093/icesjms/fsu243.

Thorson, J.T. & Ward, E.J. (2013) Accounting for space–time interactions in index standardization models. *Fisheries Research*, **147**, 426–433. https://dx.doi.org/10.1016/j.fishres.2013.03.012.

Tolimieri, N., Wallace, J. & Haltuch, M. (2020) Spatio-temporal patterns in juvenile habitat for 13 groundfishes in the California Current Ecosystem. *PLOS ONE*, **15**, e0237996. https://dx.doi.org/10.1371/journal.pone.0237996.

Vehtari, A., Gelman, A. & Gabry, J. (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, **27**, 1413–1432. https://dx.doi.org/10.1007/s11222-016-9696-4.

Velasco, J.A. & González-Salazar, C. (2019) Akaike information criterion should not be a "test" of geographical prediction accuracy in ecological niche modelling. *Ecological Informatics*, **51**, 25–32. https://dx.doi.org/https://doi.org/10.1016/j.ecoinf.2019.02.005.

Wallin, J. & Bolin, D. (2015) Geostatistical modelling using non-Gaussian Matérn fields. *Scandinavian Journal of Statistics*, **42**, 872–890. https://dx.doi.org/10.1111/sjos.12141.

Williamson, L.D., Scott, B.E., Laxton, M.R., Bachl, F.E., Illian, J.B., Brookes, K.L. & Thompson, P.M. (2021) Spatiotemporal variation in harbor porpoise distribution and foraging across a landscape of fear. *Marine Mammal Science*, **38**, 42–57. https://dx.doi.org/10.1111/mms.12839.

Wilson, B. (2021) Evaluating the INLA-SPDE approach for Bayesian modeling of earthquake damages from geolocated cluster data. *EarthArXiv*. https://dx.doi.org/10.31223/osf.io/64whm.

Wood, S. (2017) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Second edition.

Yue, Y.R., Simpson, D., Lindgren, F. & Rue, H. (2014) Bayesian adaptive smoothing splines using stochastic differential equations. *Bayesian Analysis*, **9**, 397–424. https://dx.doi.org/10.1214/13-BA866.

Table S1: Table of groundfish species included in our modelling.

Common name	Scientific name	Proportion of positive tows
Arrowtooth flounder	*Atheresthes stomias*	0.36
Aurora rockfish	*Sebastes aurora*	0.13
Big skate	*Raja binoculata*	0.16
Blackgill rockfish	*Sebastes melanostomus*	0.06
Bocaccio	*Sebastes paucispinis*	0.08
Canary rockfish	*Sebastes pinniger*	0.08
Chilipepper rockfish	*Sebastes goodei*	0.15
Curlfin sole	*Pleuronichthys decurrens*	0.12
Darkblotched rockfish	*Sebastes crameri*	0.18
Flathead sole	*Hippoglossoides elassodon*	0.08
Greenspotted rockfish	*Sebastes chlorostictus*	0.06
Greenstriped rockfish	*Sebastes elongatus*	0.26
Lingcod	*Ophiodon elongatus*	0.33
Longnose skate	*Raja rhina*	0.60
Longspine thornyhead	*Sebastolobus altivelis*	0.35
Petrale sole	*Eopsetta jordani*	0.43
Redbanded rockfish	*Sebastes babcocki*	0.08
Rex sole	*Glyptocephalus zachirus*	0.62
Rosethorn rockfish	*Sebastes helvomaculatus*	0.08
Sablefish	*Anoplopoma fimbria*	0.67
Sharpchin rockfish	*Sebastes zacentrus*	0.07
Shortspine thornyhead	*Sebastolobus alascanus*	0.51
Splitnose rockfish	*Sebastes diploproa*	0.20
Stripetail rockfish	*Sebastes saxicola*	0.22
Widow rockfish	*Sebastes entomelas*	0.04
Yelloweye rockfish	*Sebastes ruberrimus*	0.02
Yellowtail rockfish	*Sebastes flavidus*	0.07

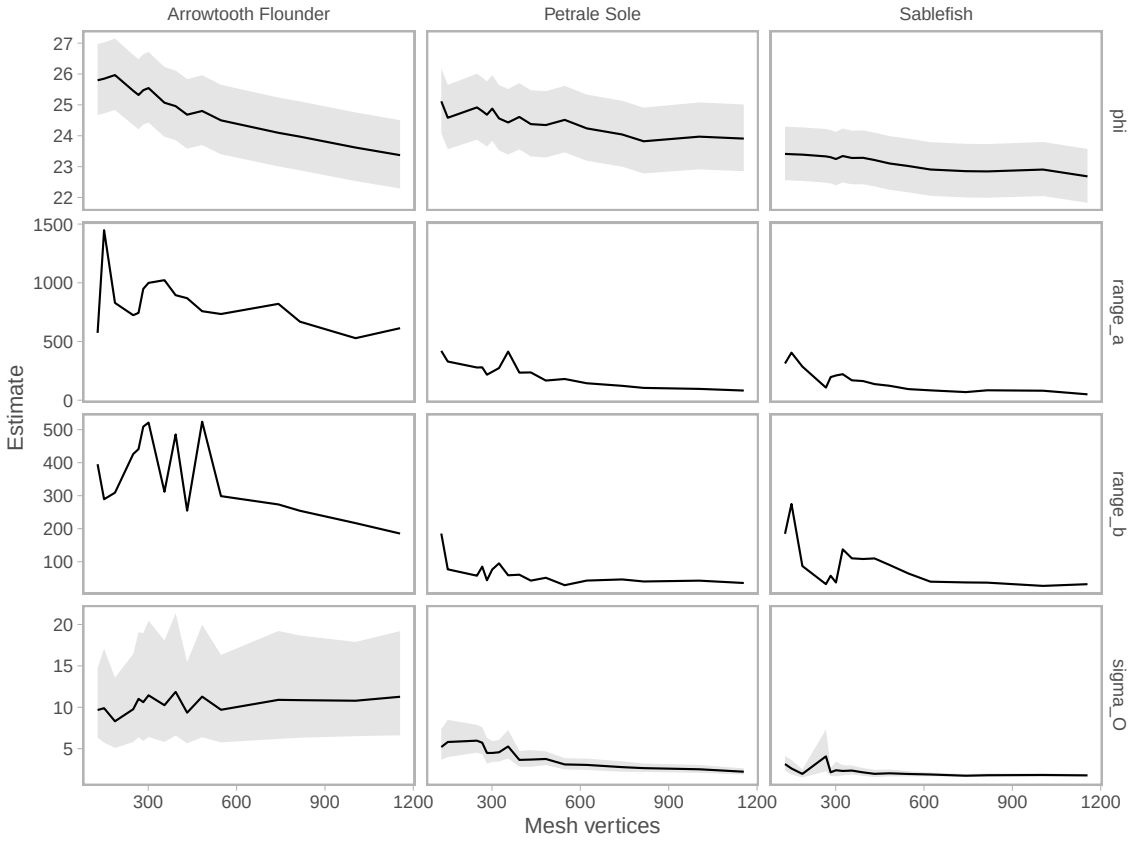

Figure S1: Parameter estimates from the models shown in Figure 4. "Phi" refers to the observation error standard deviation, "range_a" and "range_b" correspond to the ranges within the anisotropic spatial correlation, and "sigma_O" refers to the spatial GMRF standard deviation. Lines indicate mean parameter estimates and shaded ribbons, where shown, represent 95% confidence intervals.

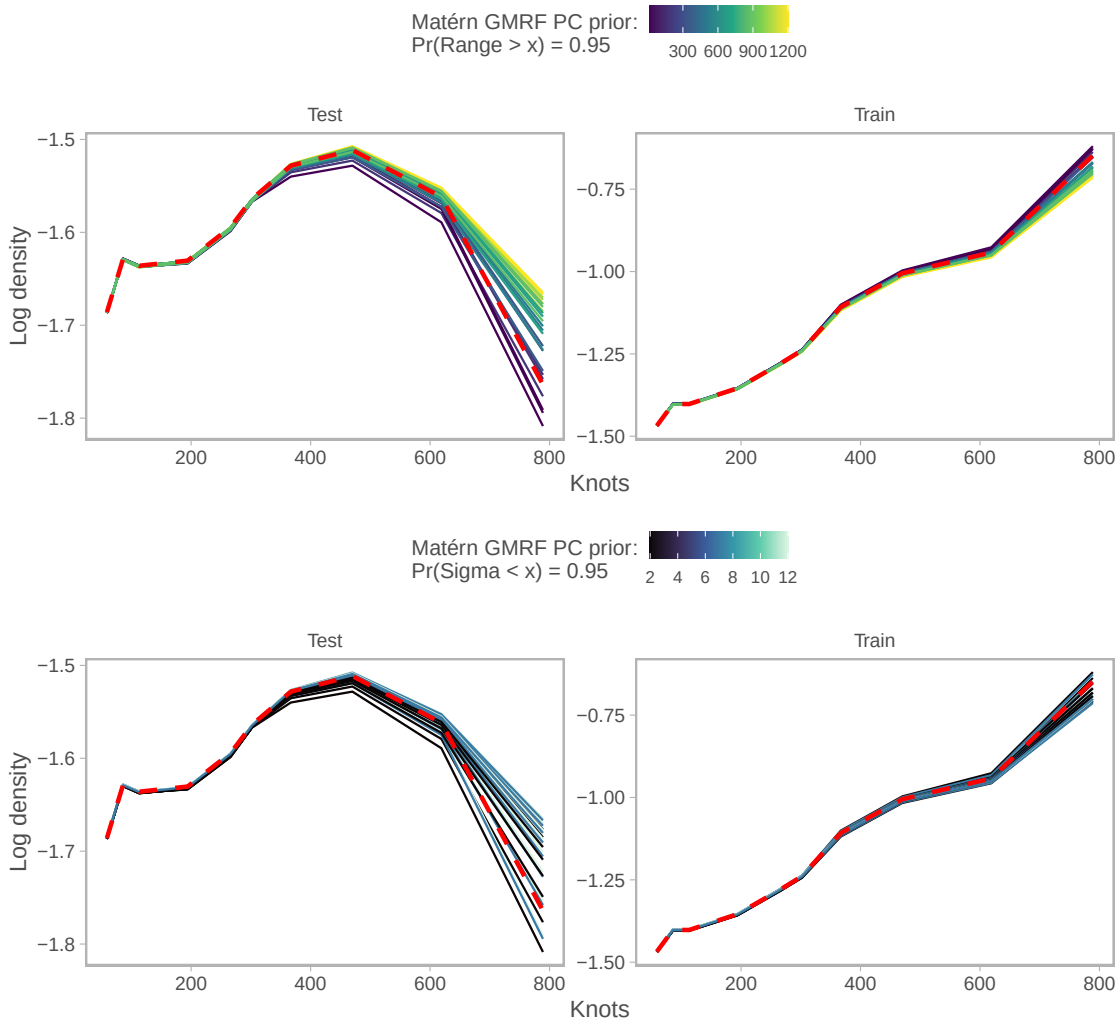

Figure S2: Same as Figure 4 but with a series of penalized complexity (PC) bivariate priors (penalties) (Fuglstad *et al.* 2018) applied to the Matérn correlation parameters. The dashed thick red line represents the log predictive density in a model without a prior applied.

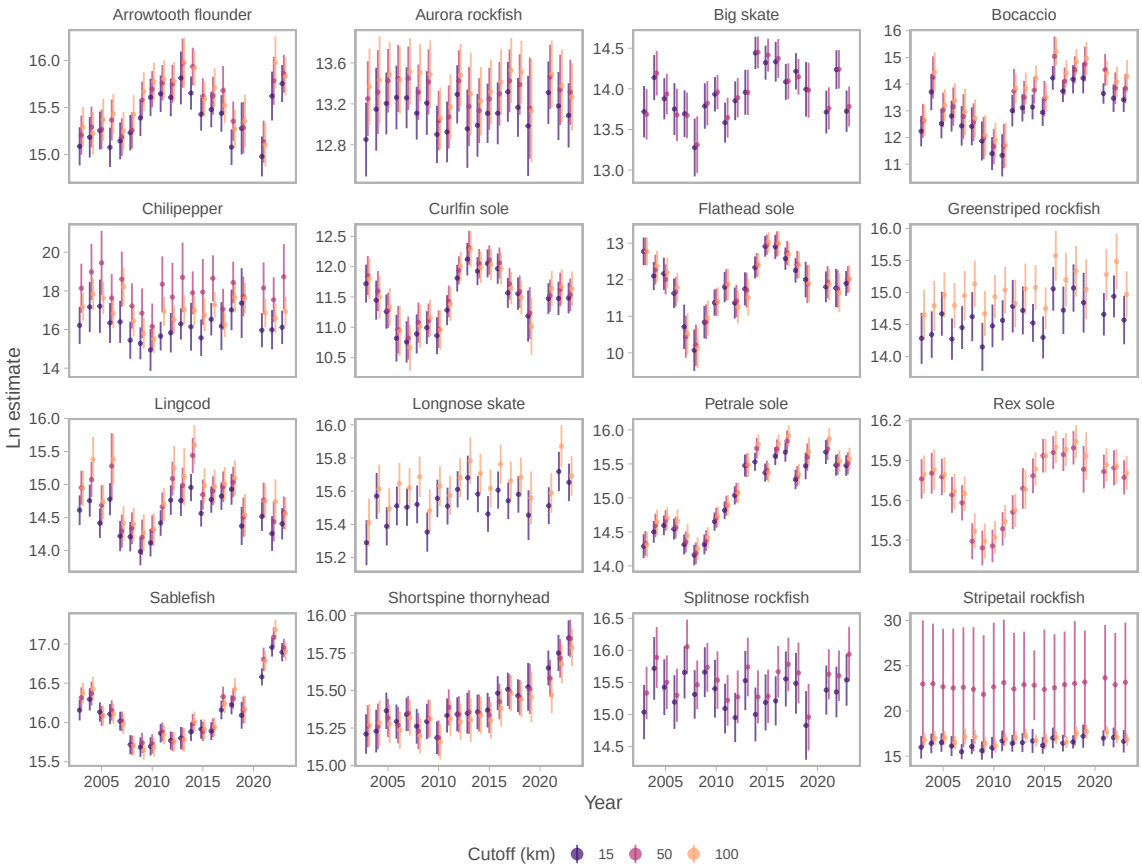

Figure S3: Time series of biomass estimates estimated for all species, across three mesh resolutions (larger cutoff distances result in coarser meshes). Only species with more than 2 converged models are shown for comparison.

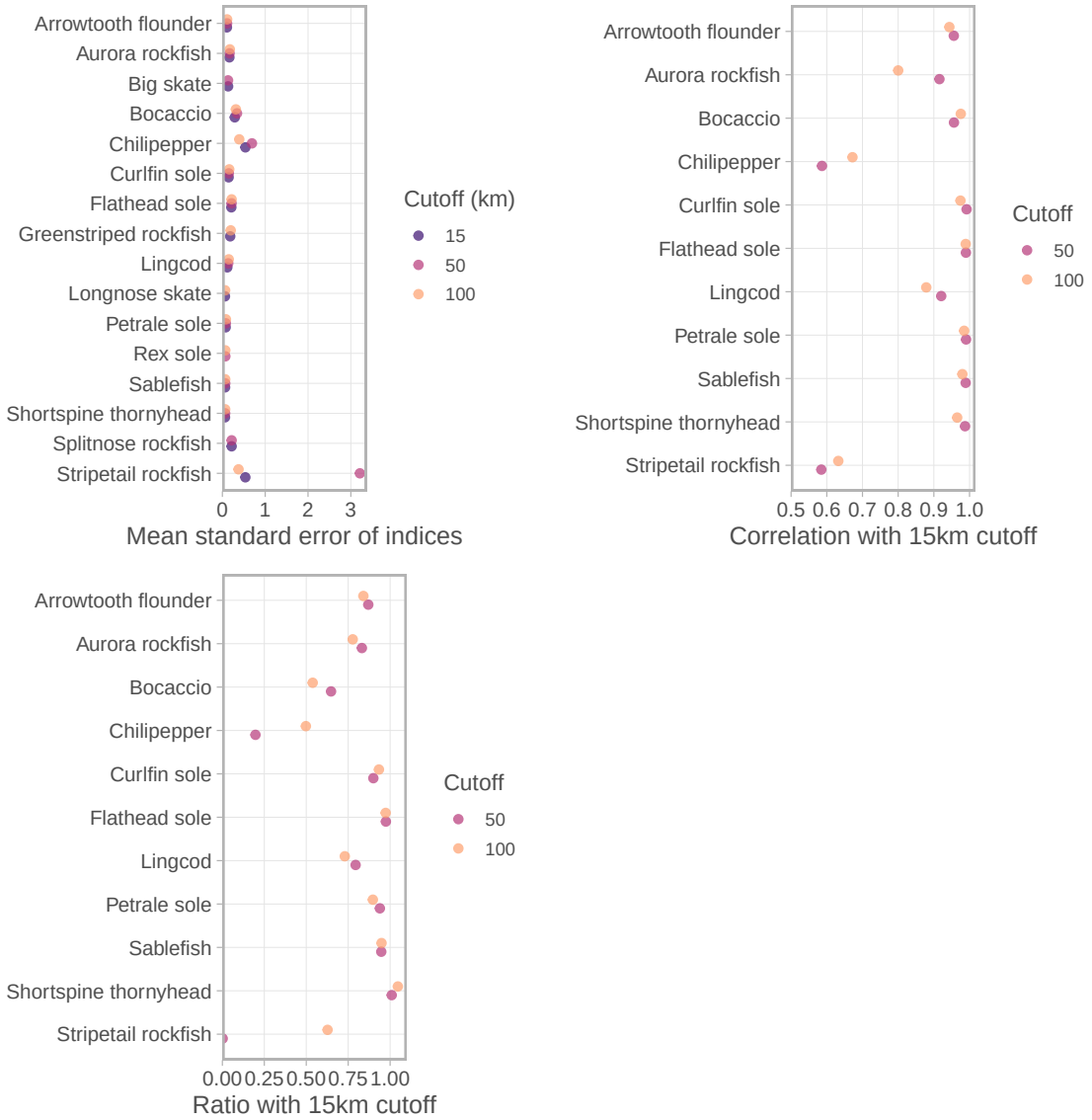

Figure S4: Relationships between groundfish biomass indices produced with different cutoff distances (larger distances correspond to coarser meshes). Panels represent the average standard error of biomass indices over the time series, the correlation between each biomass time series and that produced with the finest mesh, and the average ratio of each biomass time series to that produced using the coarsest mesh.

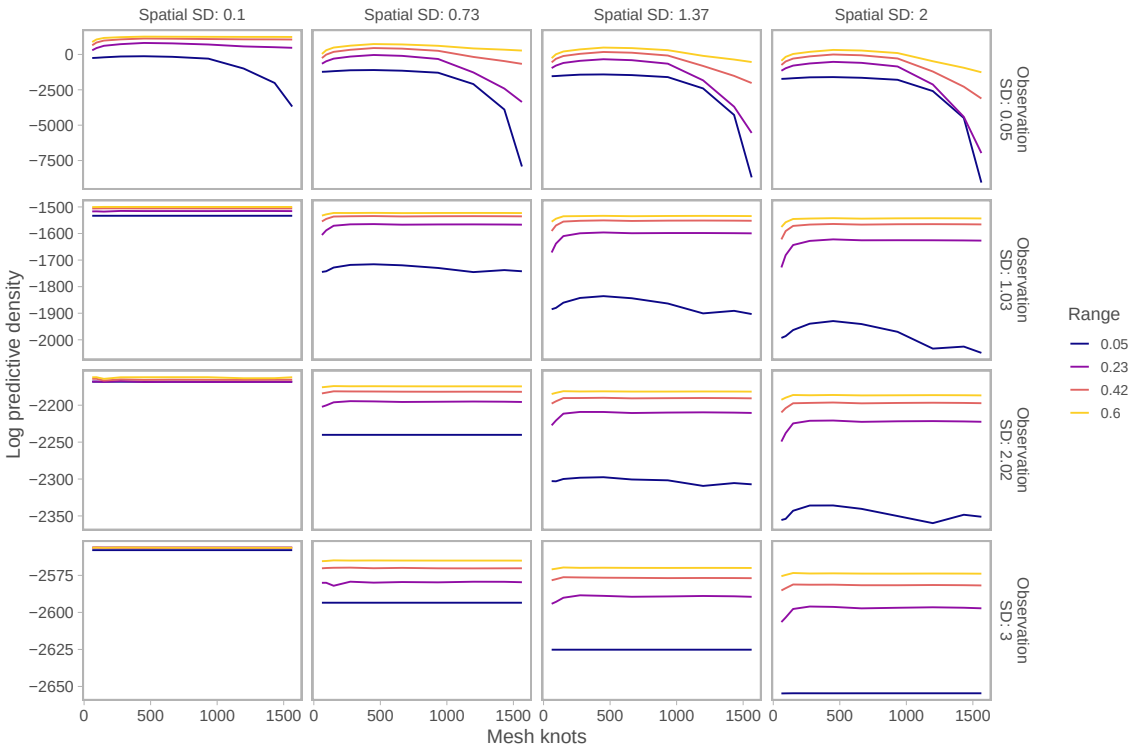

Figure S5: Same as Figure 6 but at four levels of spatial standard deviation, observation standard deviation, and spatial range. Log predictive density of models fit to simulated Gaussian random fields across a range of mesh vertices (x-axis), spatial standard deviations (columns), observation error standard deviations (rows), and spatial ranges (colours). The scale of the spatial coordinates ranged from zero to one.

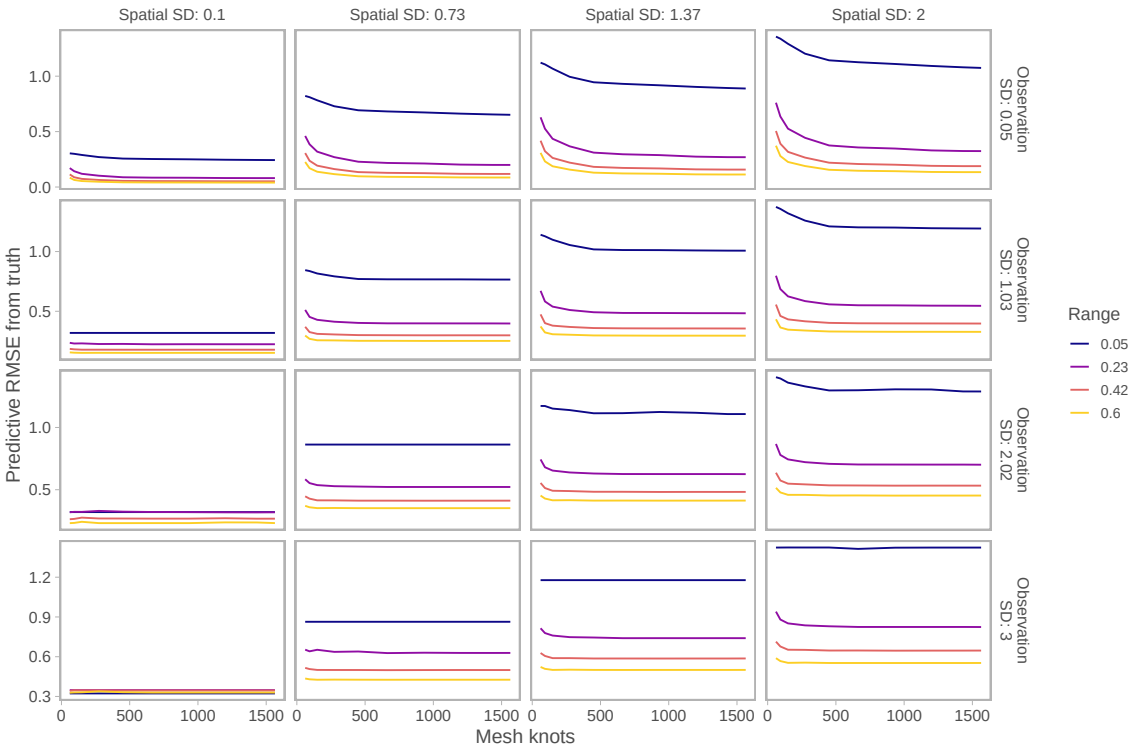

Figure S6: Predictive root mean square error (RMSE) calculated against the true known values from models shown in Figure S5.

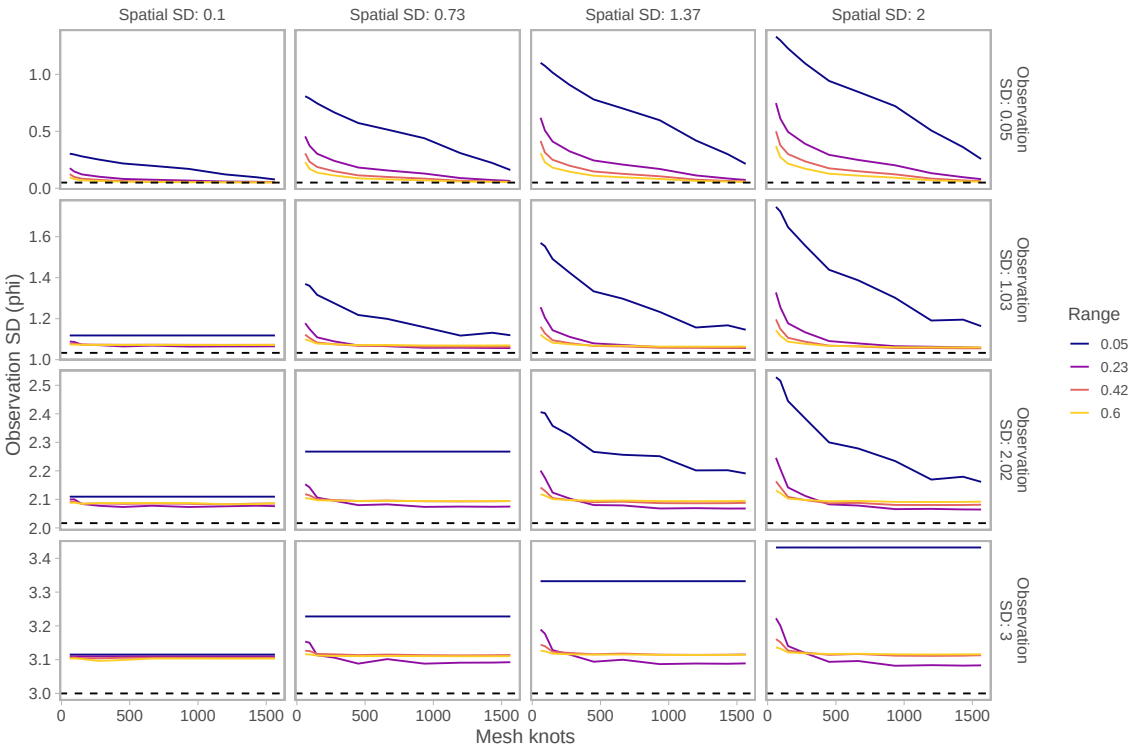

Figure S7: Estimated Gaussian observation standard deviation of models from models shown in Figure S5. True values are shown as horizontal dashed lines.

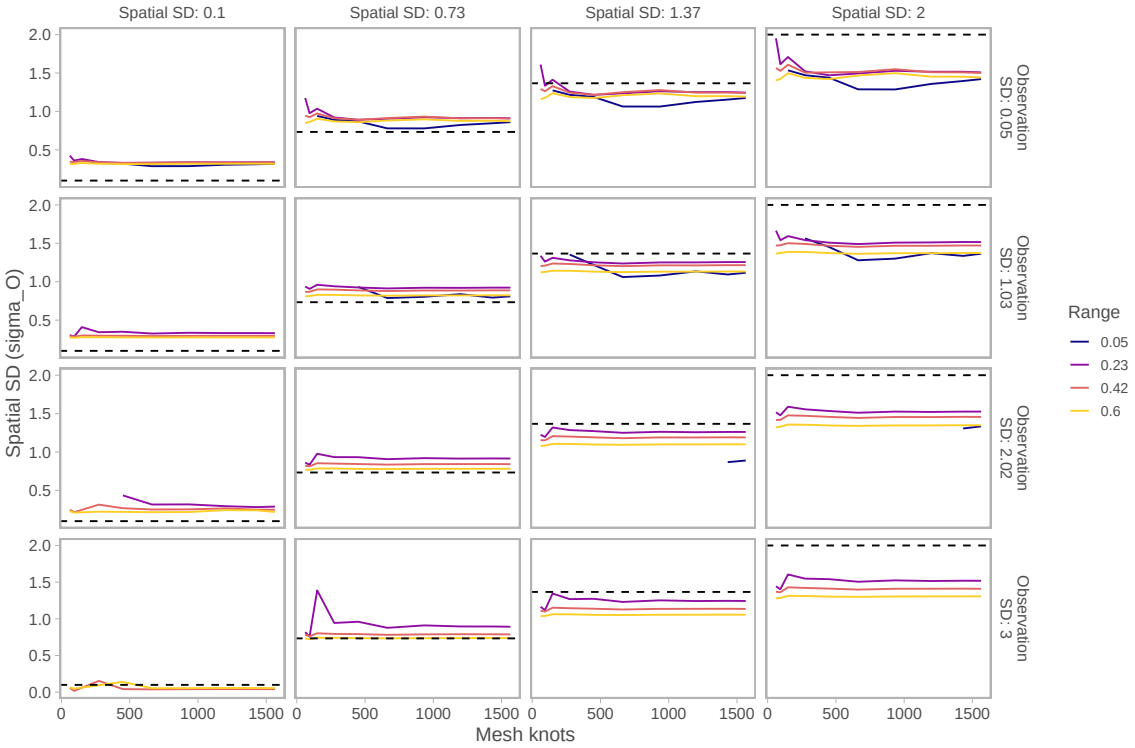

Figure S8: Estimated spatial GMRF standard deviation from models shown in Figure S5. True values are shown as horizontal dashed lines.

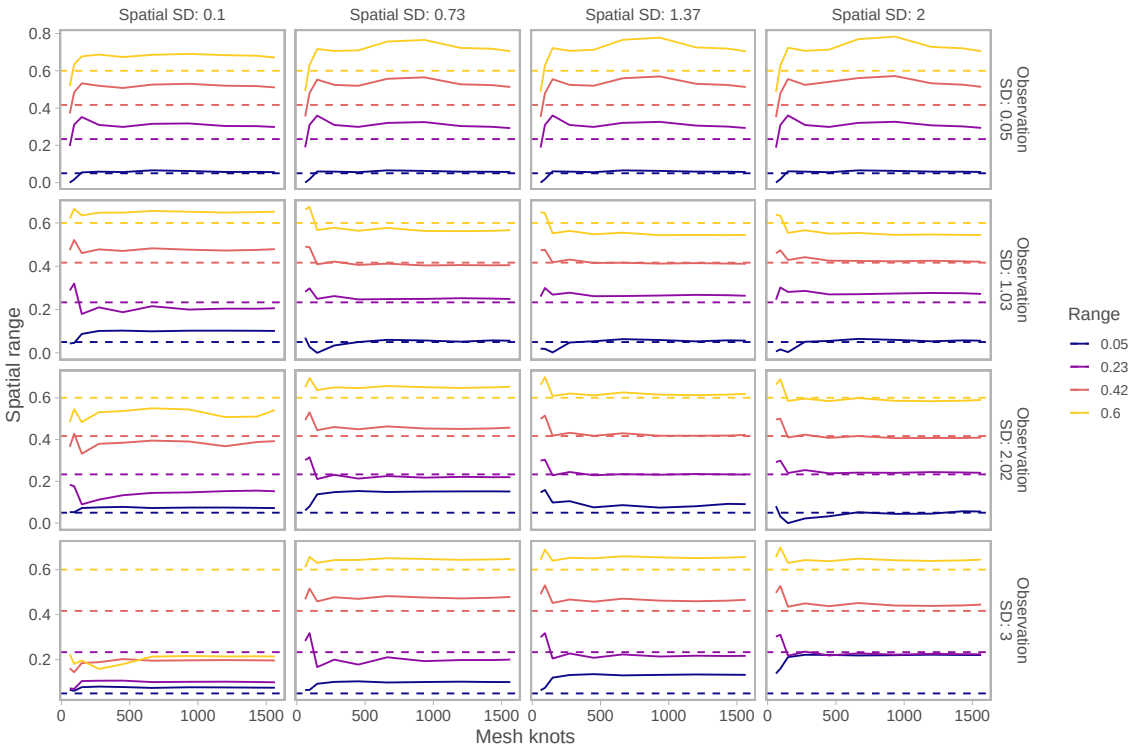

Figure S9: Estimated spatial GMRF Matérn range from models shown in Figure S5. True values are shown as horizontal dashed lines.

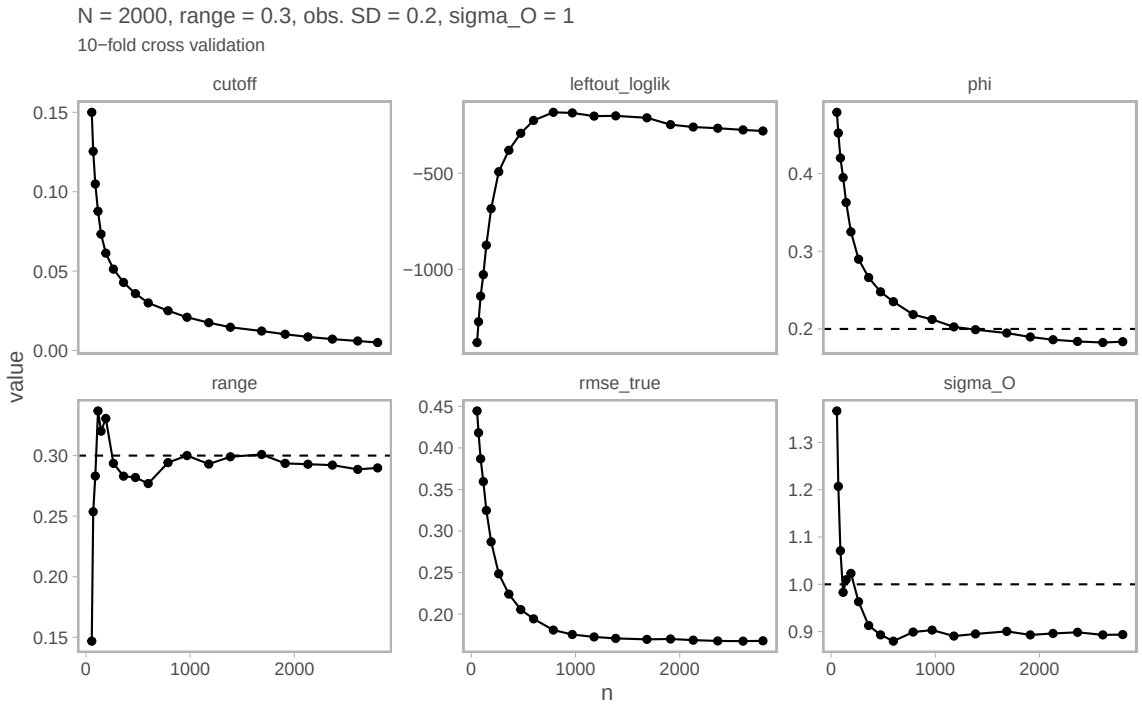

Figure S10: An example from simulated data, similar to as in Figure 6 and S5–S9. Dots and lines show estimated values or calculated statistics. Horizontal dashed lines show true values.

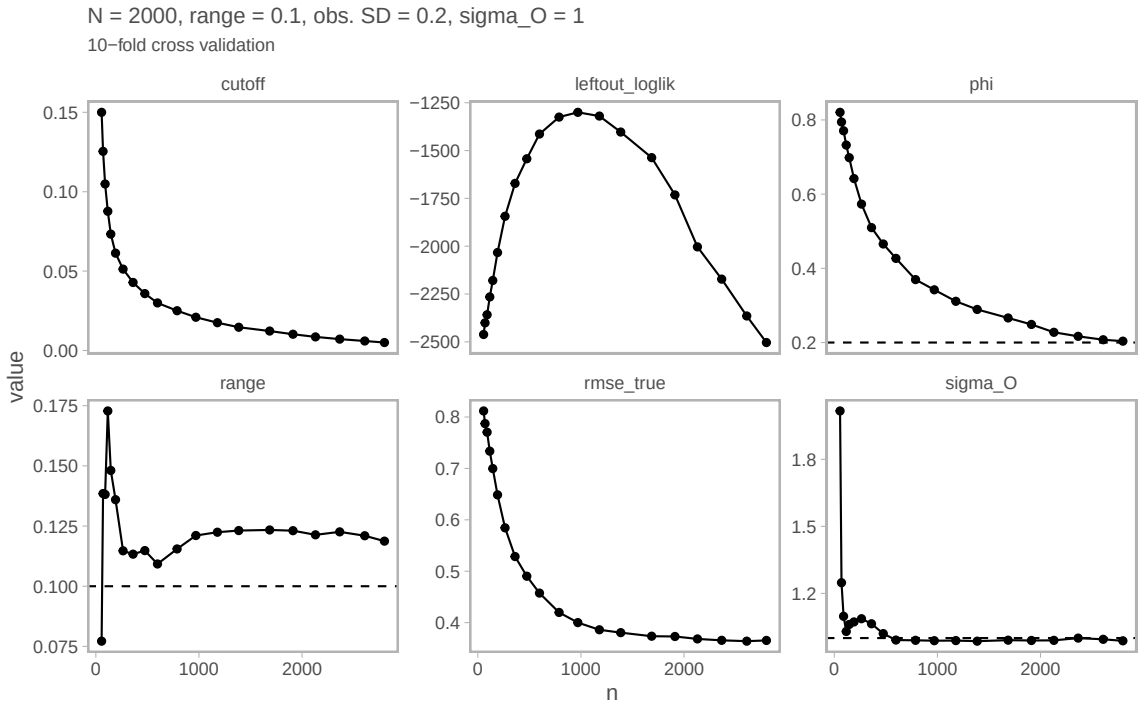

Figure S11: Same as Figure S10 but with a lower true Matérn spatial range.