# Appendix C

Saint Matthew Island blue king crab spatiotemporal model-based biomass index

Caitlin Stern[1,2]
[1]Alaska Department of Fish and Game
[2]caitlin.stern@alaska.gov

May 2026

## Summary

The St. Matthew blue king crab (SMBKC) stock assessment uses estimated biomass from the National Oceanic and Atmospheric Administration (NOAA) Eastern Bering Sea (EBS) summer bottom trawl survey as one of two indices of abundance in the stock assessment model. Following the end of survey sampling at the extra "corner stations" in 2023, the EBS survey estimated biomass time series no longer consists of directly comparable annual estimates, since fewer survey stations were sampled in 2024 and 2025 than in the 1983 - 2023 surveys that included the corner stations. Index standardization is thus needed to derive a survey biomass index for inclusion in the stock assessment model. This document details the development of a spatiotemporal model-based survey biomass index that accounts for differences in survey station sampling among years. This index will be included in proposed SMBKC stock assessment model runs presented at the Crab Plan Team meeting in May 2026.

## Introduction

Indices of abundance are key inputs to stock assessment models, providing information crucial for estimating stock size (Patterson et al. 2001; Maunder and Punt 2004). However, design-based methods for producing these indices from fishery-independent surveys can yield biased results when practical constraints lead survey sampling to deviate from survey design. For example, financial resources may not be sufficient for surveys to occur as frequently as planned, or changes in survey priorities may limit vessels to sampling only a subset of survey stations. Spatiotemporal models including spatial and spatiotemporal random effects have the potential to mitigate the effects of inconsistent survey sampling by accounting for the spatiotemporal processes that underlie survey observations (Thorson et al. 2021; Yalcin et al. 2023). Simulation work suggests than index standardization using well-specified spatiotemporal models can reduce the impact of changes in survey effort by providing a consistent method of inferring abundance in missing areas (Yalcin et al. 2023).

In addition to suffering from bias due to changes in survey sampling, design-based indices of abundance can inflate temporal variability by failing to account for spatial dependence (Shelton et al. 2014; Thorson et al. 2015). Due to the importance of indices of abundance in stock assessment models, this can lead to increased uncertainty in stock assessment model outputs used in fishery management (Cao et al. 2017). By accounting for spatial variation in survey catch, as well as incorporating habitat variables, spatiotemporal model-based index standardization can produce abundance indices with reduced variability compared to design-based indices (Cao et al. 2017, Chen et al. 2024). Important outcomes for stock assessment models can include less extreme retrospective patterns when using model-based rather than design-based indices (Cao et al. 2017, Chen et al. 2024).

The stock assessment for blue king crab (*Paralithodes platypus*) around St. Matthew Island, Alaska, uses two fishery-independent survey time series: the National Oceanic and Atmospheric Administration (NOAA) Eastern Bering Sea trawl survey (1978-2025) and the Alaska Department of Fish and Game (ADF&G) pot survey (1995-2025) (Stern and Palof 2024). The 2024 and 2025 NOAA EBS surveys did not sample the high density "corner" stations that were added to the St. Matthew Island survey area in 1983 and sampled through 2023 (Figure 1); the planned exclusion of these corner stations was discussed at the May 2024 Crab Plan Team (CPT) meeting (NPFMC 2024). Corner stations were originally added to the survey in order to improve survey data quality for the blue king crab stocks near St. Matthew and the Pribilof Islands, but the declines in these stocks and the low probability that they will rebuild to harvestable levels in the near future motivated examination of the utility of continuing to allocate survey time and funds to sampling these stations (DePhilippo et al. 2023). Across the time series, NOAA EBS survey biomass estimates are consistently lower when corner stations are excluded than when they are included; size compositions are similar overall when corner stations are excluded versus included but can show large differences for individual years (Stern and Palof 2024). A previous analysis found that excluding corner stations reduced the scale of, but not the trends in, biomass estimates for the St. Matthew Island blue king crab stock assessment, and that spatiotemporal model-based biomass estimates were more robust to the removal of corner stations than design-based estimates (DePhilippo et al. 2023). This document details the development of a spatiotemporal model-based biomass index for the EBS trawl survey that will be evaluated for inclusion in the 2026 SMBKC stock assessment model.

# Methods

## Data sources

I used the R (R Core Team 2026) package **crabpack** (Hennessey 2025) to extract the NOAA EBS trawl survey data for 1978 - 2025. Three different visualizations of these data are shown in Figures 1 - 3. Figure 1 illustrates the changes across the time series in both survey station sampling and the set of survey stations at which at least one male blue king crab $\geq$ 90 mm in carapace length (CL) was recorded. The dropping of the higher-density corner stations in 2024 and 2025 is visually apparent, as is a possible contraction of the area occupied by blue king crab. Figure 2 shows crab biomass density at each survey station, illustrating patterns such as the large estimated densities of crab caught to the north of St. Matthew Island in some years (e.g., 2010, 2011, 2014, 2015). Figure 3 shows the same information as Figure 2, but with crab biomass density represented using point color rather than point size, making it easier to determine at which survey stations the highest crab biomass densities were recorded.

## Spatiotemporal models

I fit geostatistical GLMMs with spatiotemporally correlated random effects to survey data using the R package **sdmTMB** (Anderson et al. 2022). In this approach, spatial effects are modeled using the stochastic partial differential equation (SPDE) approximation to Gaussian random fields (Lindgren et al. 2011). The rate at which spatial covariance decays with distance is defined by the Matèrn covariance function, a derived parameter of which is the spatial (or Matèrn) range, the distance at which spatial correlation decays such that two points are effectively independent (Anderson et al. 2022). **sdmTMB** uses geostatistical time series data to estimate spatial and spatiotemporal generalized linear mixed effects models (GLMMs). This approach allows for index standardization when the set of stations surveyed is not consistent across years: one can generate a spatial grid that covers the area of interest, predict from the model onto that grid, and sum the predicted abundance to obtain an area-weighted index that is independent of sampling locations (Anderson et al. 2022).

Applying the SPDE approach to modeling random fields requires defining a triangulation mesh of the spatial domain (Anderson et al. 2022). I constructed a triangulation mesh for the EBS survey data set using the **sdmTMB** function make_mesh(), a wrapper for the triangulation mesh functions in the **fmesher** package (Lindgren 2023), and the **sdmTMBextra** (Anderson 2025) function add_barrier_mesh(). The add_barrier_mesh function permitted incorporation of correlation barriers based on the St. Matthew Island coastline. The resulting mesh is shown in Figure 4. Figure 5 shows the vertices included (black circles) versus excluded (red crosses) from the mesh, demonstrating that the barrier mesh approach effectively excluded vertices on land and outside of the designated spatial polygon from the mesh.

For all models, I estimated spatiotemporal random fields as independent and identically distributed (IID). The spatial random fields are estimated for each time slice, the period of which is specified in the model (e.g., time = year). Estimating spatiotemporal random fields as IID is likely most appropriate for the standardization of survey indices intended to be used in stock assessment models as this approach minimizes the estimation covariance among years, which is usually ignored in stock assessment models (Thorson et al. 2020, Chen et al. 2024). I compared models in which the only fixed effect was year, specified as a factor, with models that also included survey station depth (m) as a covariate, as variation in the depth of stations surveyed from year to year could influence estimated abundance if not taken into account. I compared two approaches for handling the depth covariate: 1) depth centered and scaled by its standard deviation, and 2) a smooth on depth. I ran models with the Tweedie, delta gamma, and delta lognormal distribution families.

## Model evaluation and diagnostics

After fitting models, I checked model convergence using the **sdmTMB** sanity() function. Models were considered converged if they met the following criteria evaluated by the sanity() function: the non-linear minimizer suggested successful convergence, the Hessian matrix was positive definite, no extreme or very small eigenvalues were detected, all gradients with respect to fixed effects were $< 0.001$, no fixed-effect standard errors were NA, no standard errors looked unreasonably large, no sigma parameters were $< 0.01$, no sigma parameters were $> 100$, and the range parameter was of a reasonable size. Models that did not converge were excluded from further consideration.

For models that passed all the sanity checks, I calculated residuals using the **DHARMa** package (Hartig 2024) and examined diagnostics included in the package. I plotted the DHARMa residuals over space and time to visualize potential patterns of autocorrelation.

I evaluated model predictive skill (the predictive ability of the model for new observations; Anderson *et al.* 2024) using the cross validation function sdmTMB_cv(). This function measures model predictive skill by holding out subsets of the data in turn and using each as a test set. These subsets of data are termed "folds". Following the practices of similar investigations (e.g., Vihtakari *et al.* 2026), I compared models by performing cross validation with 10 randomly arranged folds for each model before extracting the model's total predicted out-of-sample negative log-likelihood (NLL). The model with the smallest NLL was deemed the best-fitting model. I calculated Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) based on the predictive values; smaller RMSE and MAE values indicate better model fit.

## Predictions

To generate a prediction grid, I pulled the SMBKC survey strata from the **akgfmaps** package (AFSC 2026; Figure 6) using the get_crab_strata() function, then produced a grid with a resolution of 5 $km^2$ that matched the SMBKC survey area (Figure 7). I used the **sdmTMB** predict() function to generate predicted crab abundance from each model over the prediction grid.

## Indices of abundance

I used the **sdmTMB** get_index() function to calculate total biomass estimates and standard errors from each model, then plotted these model-based indices with the design-based index currently used in the stock assessment model for visual comparison.

# Results

## Spatiotemporal model diagnostics

Five models converged and passed sanity checks: the Tweedie distribution family models with year as a fixed effect, year and scaled depth as fixed effects, and year and a smooth on depth as fixed effects; and the delta gamma distribution family models with year as a fixed effect and with year and depth as fixed effects.

Comparing diagnostics using DHARMa residuals among the models showed that the models performed similarly: all five models showed no evidence of significant deviation in the Kolmogorov-Smirnov or outlier tests, but showed evidence of significant deviation in the dispersion test (Figures 8 - 12). The spatial distributions of residuals did not appear to show concerning patterning (Figures 13 - 17).

Cross-validation for the delta gamma distribution family model with year and depth as fixed effects did not converge, so this model was excluded from further consideration. Among the models for which cross-validation analysis was successful, the model using the delta gamma distribution family with year as a fixed effect had the smallest NLL value, indicating that this model had the best out-of-sample predictive skill of the models evaluated; this model also had the smallest RMSE and MAE values (Figure 18). Based on these model selection criteria, the model using the delta gamma distribution family with year as a fixed effect is the best-fitting model and I used this model to generate the biomass estimates for the model-based index in the stock assessment model.

The Matèrn range for the presence/absence component of the best-fitting model was 114 km, indicating spatial autocorrelation in crab presence/absence over a large distance, while the range was much smaller for the positive catch component of the model, indicating spatial autocorrelation in crab abundance over a shorter distance (Table 1). Standard deviations were higher for spatial fields ($\sigma_\omega$) than for spatiotemporal fields ($\sigma_\epsilon$) for both model components, indicating that spatial variability was larger than spatiotemporal variability for both crab presence/absence and crab abundance (Table 1).

## Spatiotemporal model predicted biomass

The spatial distribution of model-predicted crab presence/absence and abundance for the selected model is shown in Figures 19 and 20.

## Model-based biomass indices compared to the design-based biomass index

Biomass estimates from the design-based index are shown alongside estimates from the selected model, as well as estimates from the Tweedie distribution family model with year as a fixed effect for comparison (Figure 21). The same information with only years from 2000 to the present included is shown in Figure 22, providing a finer-scale visualization as early years with large uncertainties for the Tweedie model index are excluded. The spatiotemporal model-based biomass estimates tend to be smaller in scale than the design-based biomass estimates for the model using the delta gamma distribution family, though this is not the case for the model using the Tweedie distribution family. The model-based biomass estimates from the model using the delta gamma distribution family also show reduced interannual variability compared to the design-based biomass estimates, a finding consistent with previous work (Cao et al. 2017, Chen et al. 2024).

# Conclusions

The spatiotemporal model-based biomass index for St. Matthew Island blue king crab accounts for differences in survey stations sampled among years, producing a consistent time series that includes both years in which corner stations were sampled (1983-2023) and years in which corner stations were not sampled (1978-1982; 2024-2025). This model-based biomass index will be evaluated for inclusion in the 2026 SMBKC stock assessment model.

## Acknowledgements

I thank the members of the North Pacific Fishery Management Council Crab Plan Team, as well as Chris Siddon and Alex Reich of ADF&G, for their helpful feedback on earlier versions of this work. I thank Lewis Barnett and Sean Hardison for providing invaluable advice on the development of the spatiotemporal models.


## References


Anderson SC (2025) sdmTMBextra: Extra Functions for Working with 'sdmTMB' Models. R package version 0.0.4, commit 63f236912e12ce78b5a0529eedf1e11cb93d0a10, https://github.com/pbs-assess/sdmTMBextra.

Anderson SC, Ward EJ, English PA, Barnett LAK (2022) sdmTMB: An R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. bioRxiv, *2022.03.24.485545*. doi:10.1101/2022.03.24.485545.

Anderson SC, Ward EJ, English PA, Barnett LAK, Thorson JT (2024) Cross-validation for model evaluation and comparison. Retrieved from https://pbs-assess.github.io/sdmTMB/articles/cross-validation.html.

Bivand RS. 2022. spdep: Spatial Dependence: Weighting Schemes, Statistics. R Package Version 1.2-1, https://CRAN.R-project.org/package=spdep.

Cao J, Thorson JT, Richards RA, Chen Y (2017) Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine. Canadian Journal of Fisheries and Aquatic Sciences 74: 1781–1793 doi:10.1139/cjfas-2016-0137.

Cacciapaglia C, Brooks EN, Adams CF, Legault CM, Perretti CT, Hart D (2024) Developing workflow and diagnostics for model selection of a vector autoregressive spatiotemporal (VAST) model in comparison to design-based indices. Fisheries Research 275:107009 doi:10.1016/j.fishres.2024.107009.

Chen J, Gao J, Zhang F (2024) Spatiotemporal model improves survey indices for witch flounder stock assessment in the Grand Banks. Canadian Journal of Fisheries and Aquatic Sciences 81:459–487 doi:10.1139/cjfas-2023-0101.

DeFilippo L, Kotwicki S, Barnett L, Richar J, Litzow MA, Stockhausen WT, Palof K (2023) Evaluating the impacts of reduced sampling density in a systematic fisheries-independent survey design. Frontiers in Marine Science 10:1219283 doi:10.3389/fmars.2023.1219283.

Hartig F (2024) DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.7, https://CRAN.R-project.org/package=DHARMa.

Hennessey S (2025) crabpack: Calculate Bering Sea Crab CPUE, Abundance, and Biomass. R package version 1.0.0, commit 65322a20b78cf157b69587eb3039e30ec8ee114d, https://github.com/AFSC-Shellfish-Assessment-Program/crabpack.

Moran PA (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–23. doi:10.1093/biomet/37.1-2.17.

R Core Team (2026) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Shelton AO, Thorson JT, Ward EJ, Feist BE (2014) Spatial semiparametric models improve estimates of species abundance and distribution. Canadian Journal of Fisheries and Aquatic Sciences 71:1655–1666 doi:10.1139/cjfas-2013-0508.

Stern CA, Palof K (2024) Saint Matthew Island blue king crab stock assessment 2024. In: Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands: 2024 Final Crab SAFE. North Pacific Fishery Management Council, Anchorage AK.

Thorson JT, Cunningham CJ, Jorgensen E, Havron A, Hulson P-JF, Monnahan CC, von Szalay P (2021) The surprising sensitivity of index scale to delta-model assumptions: recommendations for model-based index standardization. Fisheries Research 233:105745. doi:10.1016/j.fishres.2020.105745.

Vihtakari M, Aune M, Assmann KM, Howell D, Hallfredsson EH, Ward EJ (2026) Demersal distribution and fisheries independent trends of beaked and golden redfish in the Barents and Norwegian Seas. ICES Journal of Marine Science 83:fsag009 doi:10.1093/icesjms/fsag009.

Yalcin S, Anderson SC, Regular PM, English PA (2023) Exploring the limits of spatiotemporal and design-based index standardization under reduced survey coverage. ICES Journal of Marine Sciences 80:2368-2379. doi:10.1093/icesjms/fsad155.

## Tables

Table 1: Spatial parameters for the best-fitting model, which used the delta-gamma distribution family and included year as a fixed effect. Delta-gamma models include one component that models presence/absence and one component that models positive catch (abundance); results from both model components are shown. Terms include the Matèrn range (the distance at which spatial correlation decays such that two points are effectively independent), the standard deviation for spatial fields ($\sigma_\omega$), and the standard deviation for spatiotemporal fields ($\sigma_\epsilon$).

| Component | Fixed effects | Term | Estimate | SE |
|---|---|---|---|---|
| presence/absence | year | range | 114.39 | 23.25 |
| presence/absence | year | sigma_O | 2.72 | 0.38 |
| presence/absence | year | sigma_E | 0.65 | 0.23 |
| positive catch | year | range | 7.09 | 4.45 |
| positive catch | year | phi | 2.26 | 0.17 |
| positive catch | year | sigma_O | 1.94 | 1.01 |
| positive catch | year | sigma_E | 1.12 | 0.52 |

# Figures

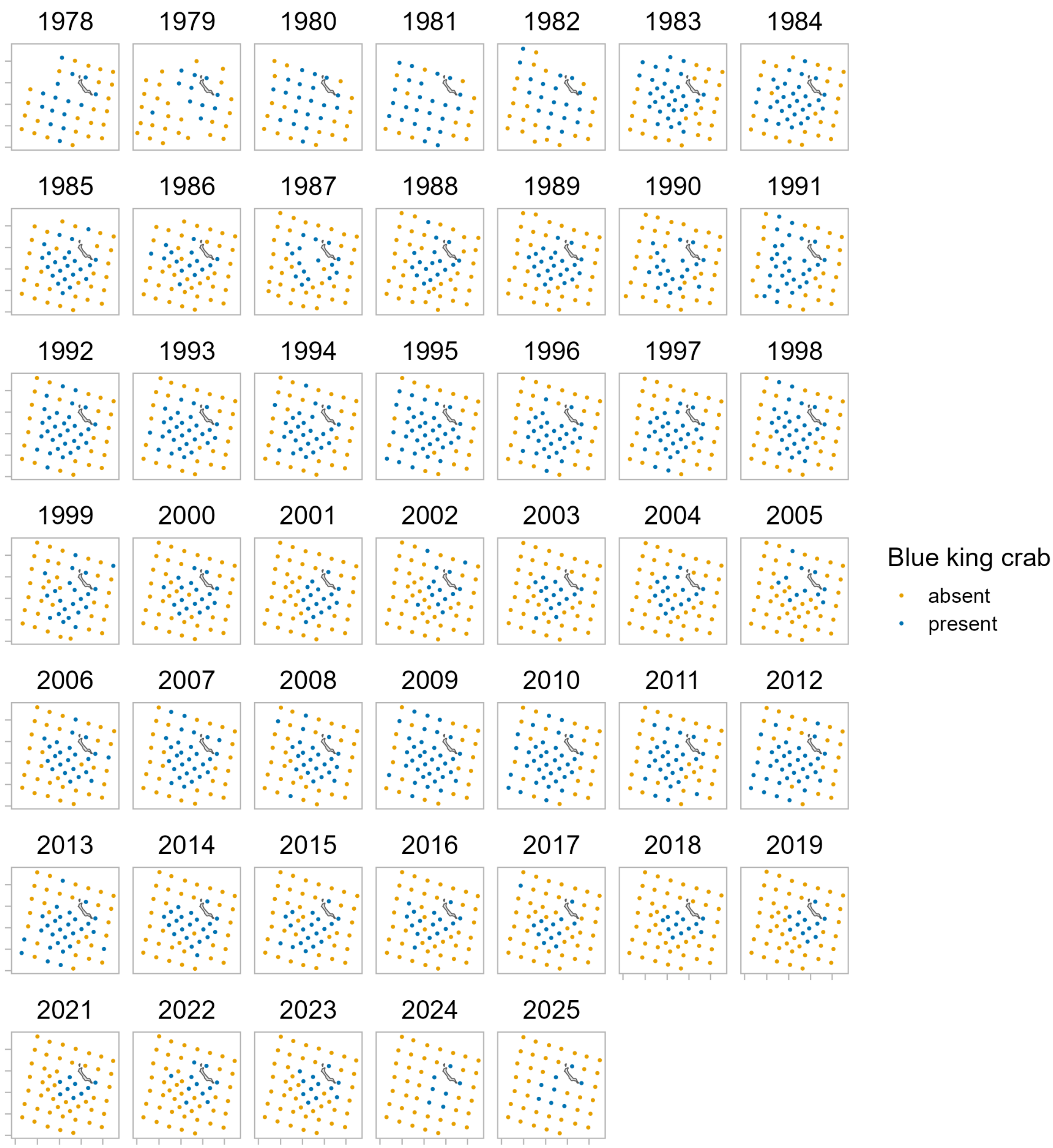


Figure 1: St Matthew Island blue king crab survey stations sampled in each year of the Eastern Bering Sea bottom trawl survey time series. Note that corner stations were not sampled in 2024 and 2025. Stations colored orange recorded no male blue king crab $\geq$ 90 mm in carapace length, while stations colored blue recorded at least one male blue king crab $\geq$ 90 mm in carapace length.

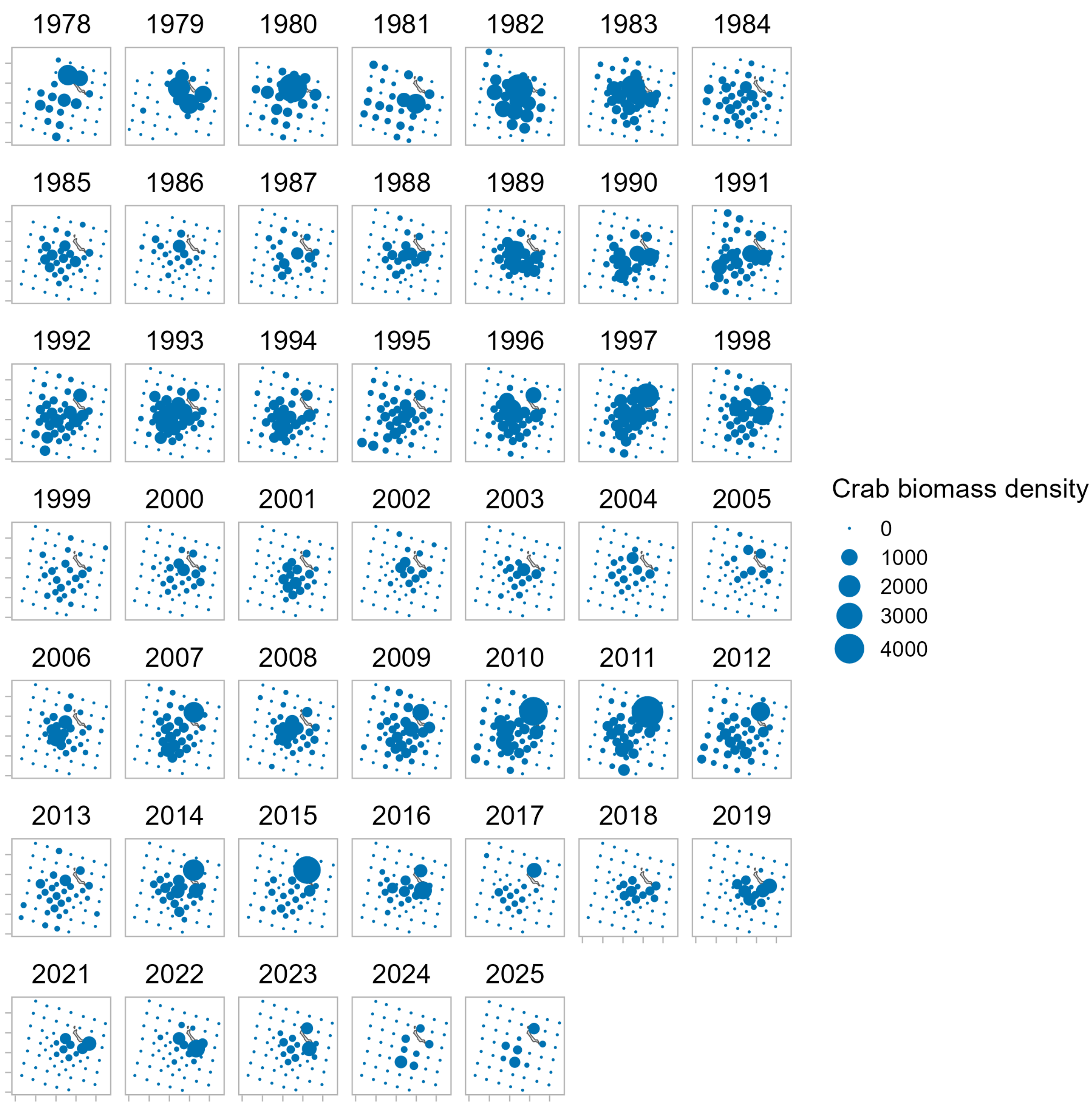


Figure 2: St Matthew Island blue king crab biomass density at survey stations sampled in each year of the Eastern Bering Sea bottom trawl survey time series. Point size is proportional to the biomass density of male blue king crab $\geq$ 90 mm in carapace length at the station.

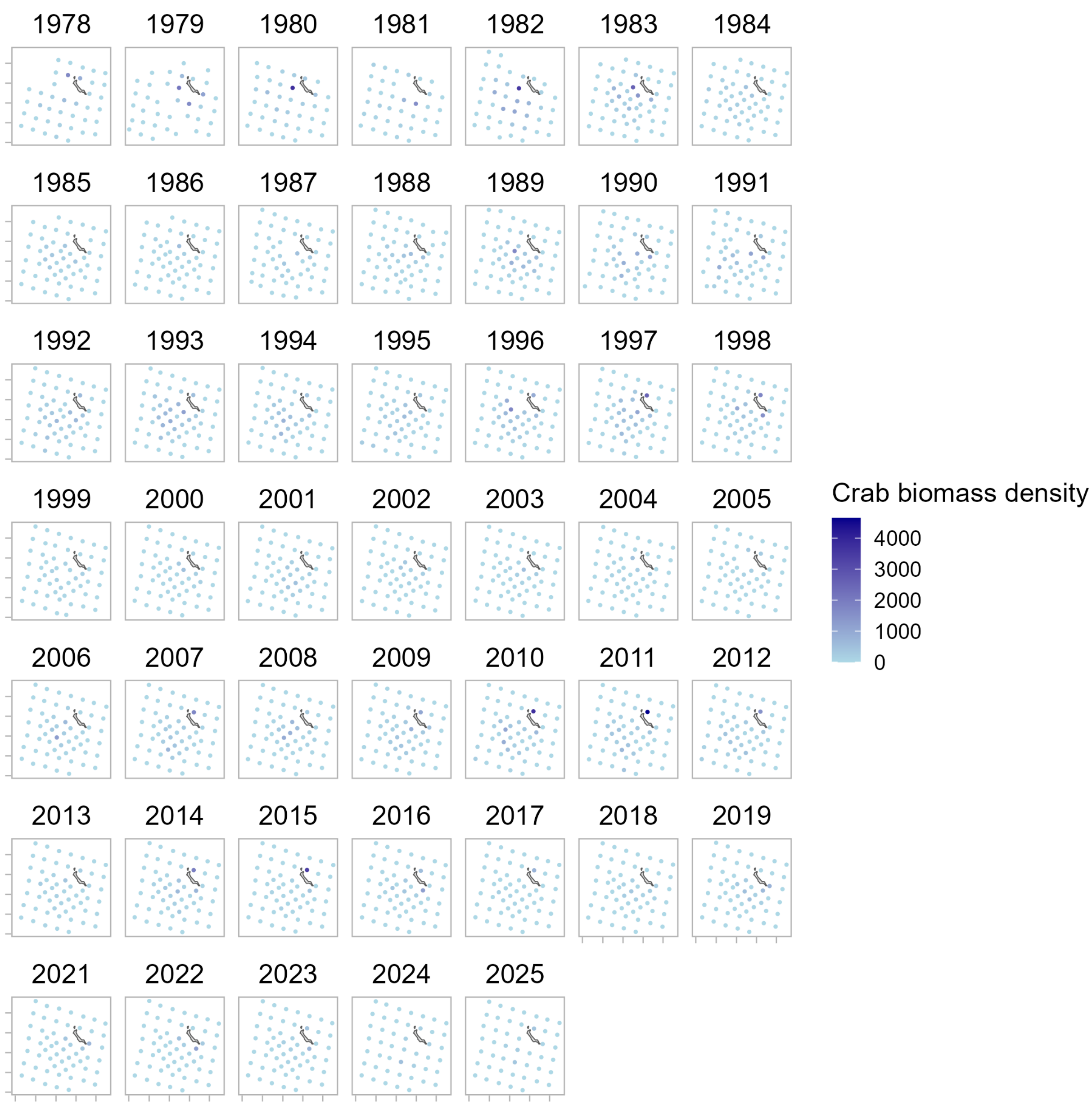


Figure 3: St Matthew Island blue king crab biomass density at survey stations sampled in each year of the Eastern Bering Sea bottom trawl survey time series. Point color is proportional to the biomass density of male blue king crab $\geq$ 90 mm in carapace length at the station.

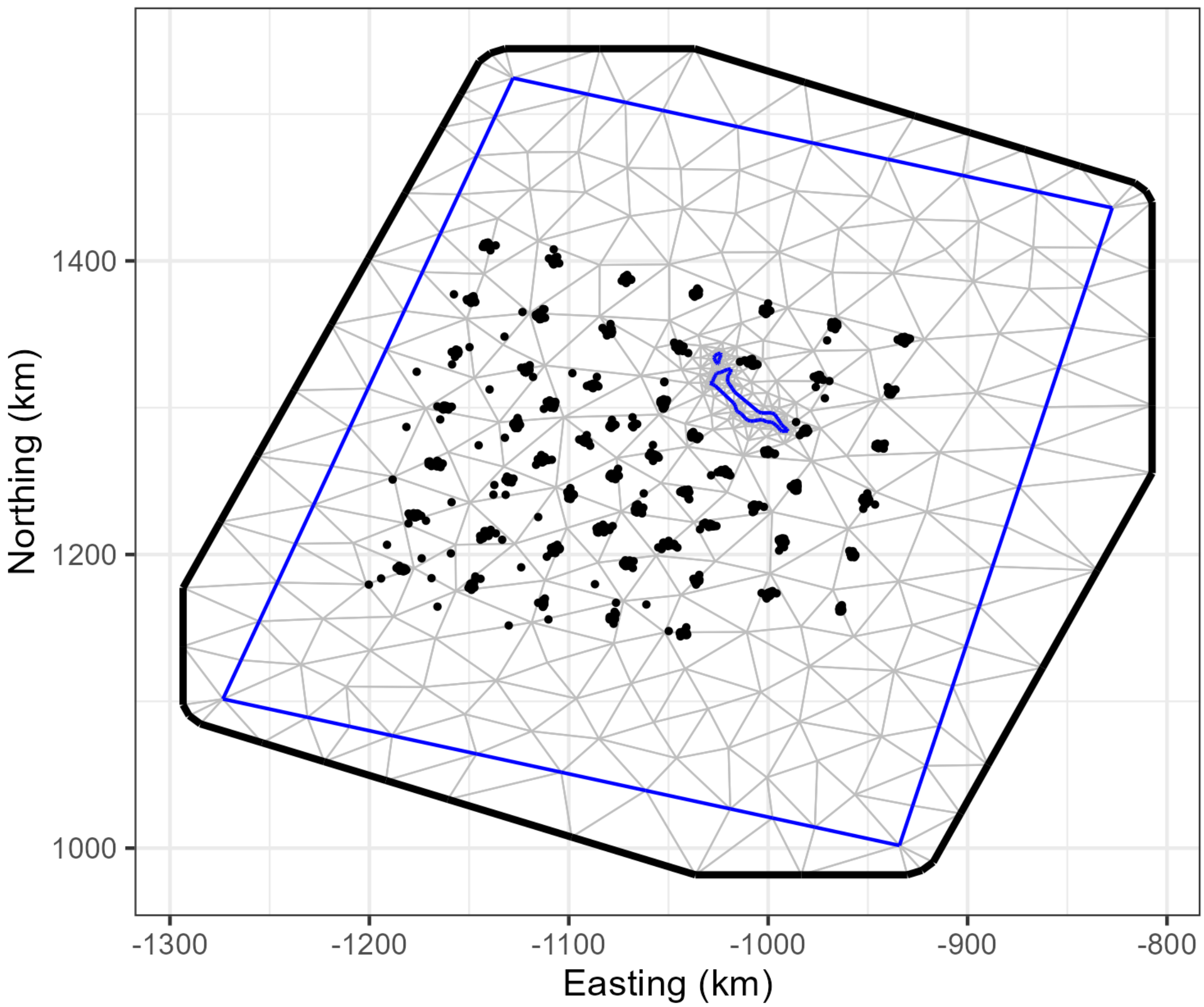


Figure 4: Spatial mesh used in spatiotemporal models, with NOAA EBS survey data observations (black points) overlaid.

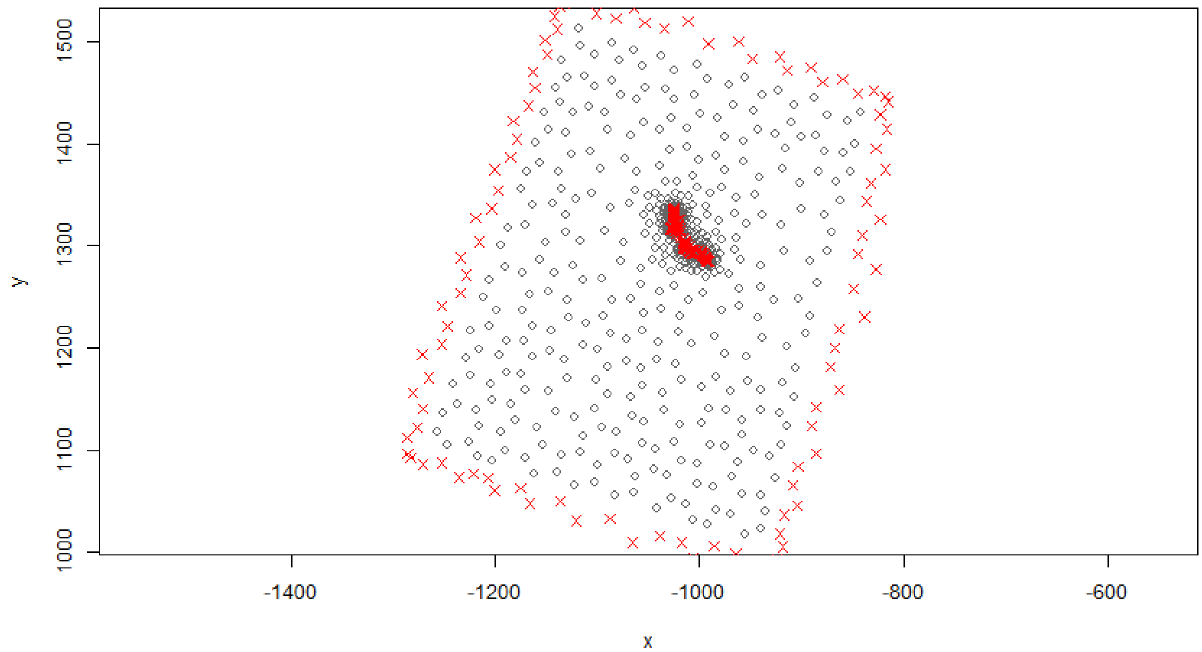


Figure 5: Vertices of the spatial barrier mesh used in spatiotemporal models, demonstrating that land as well as the area outside the designated polygon are excluded from the mesh (red crosses).

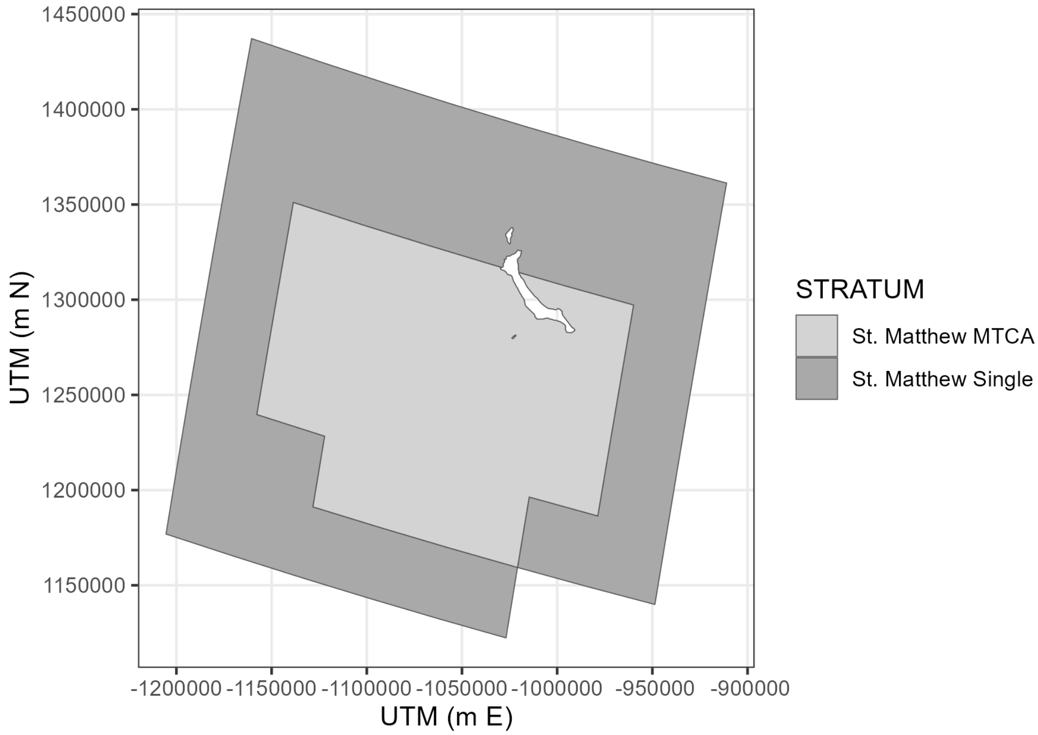


Figure 6: NOAA Eastern Bering Sea summer bottom trawl survey strata for the St. Matthew Island blue king crab stock.

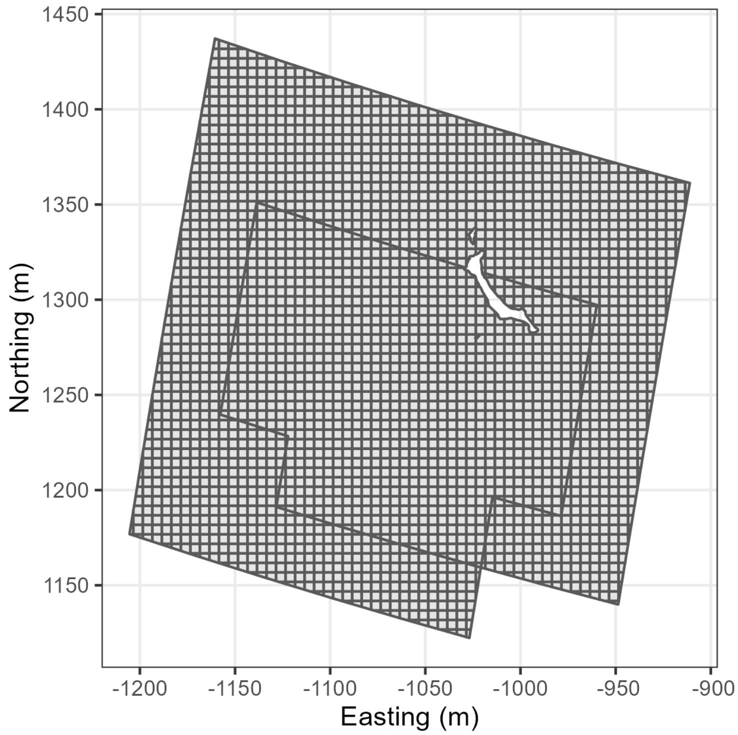


Figure 7: Prediction grid used to generate St. Matthew Island blue king crab biomass estimates.

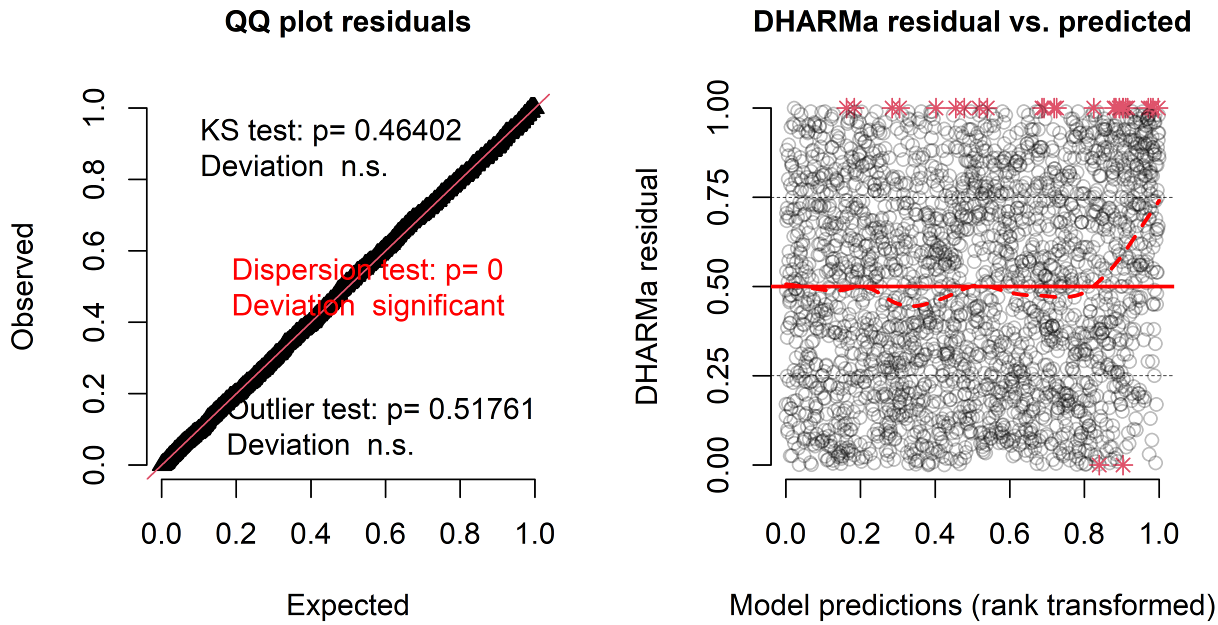


Figure 8: Diagnostics for the delta gamma distribution family model with year as a fixed effect, using DHARMa residuals.

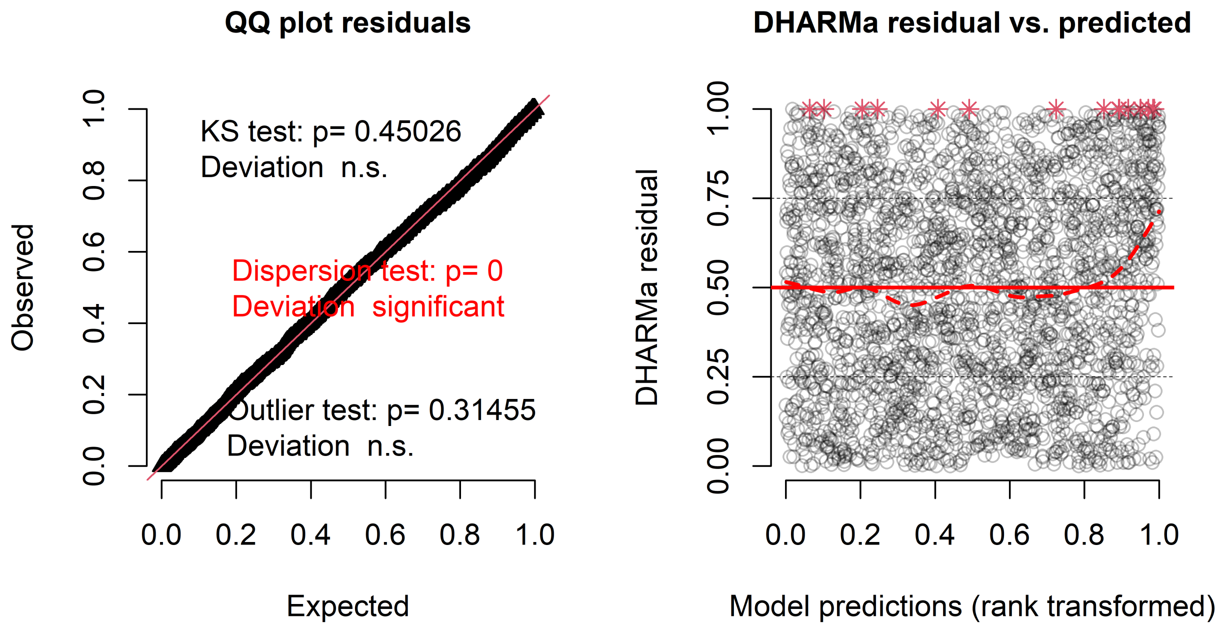


Figure 9: Diagnostics for the delta gamma distribution family model with year and depth as fixed effects, using DHARMa residuals.

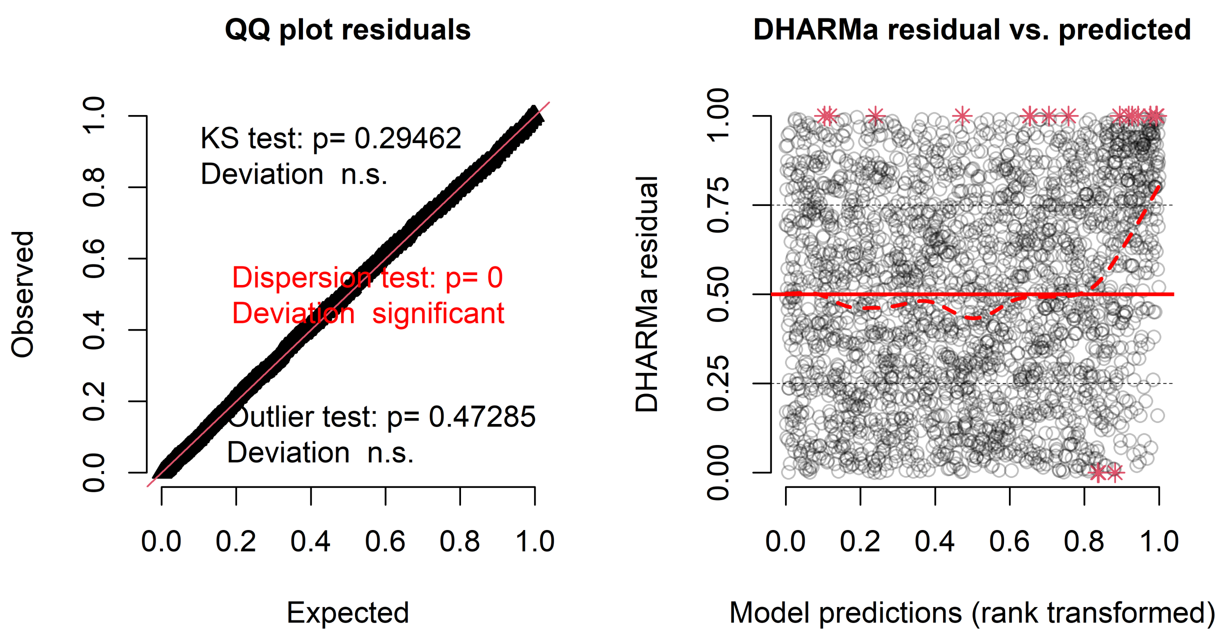


Figure 10: Diagnostics for the Tweedie distribution family model with year as a fixed effect, using DHARMa residuals.

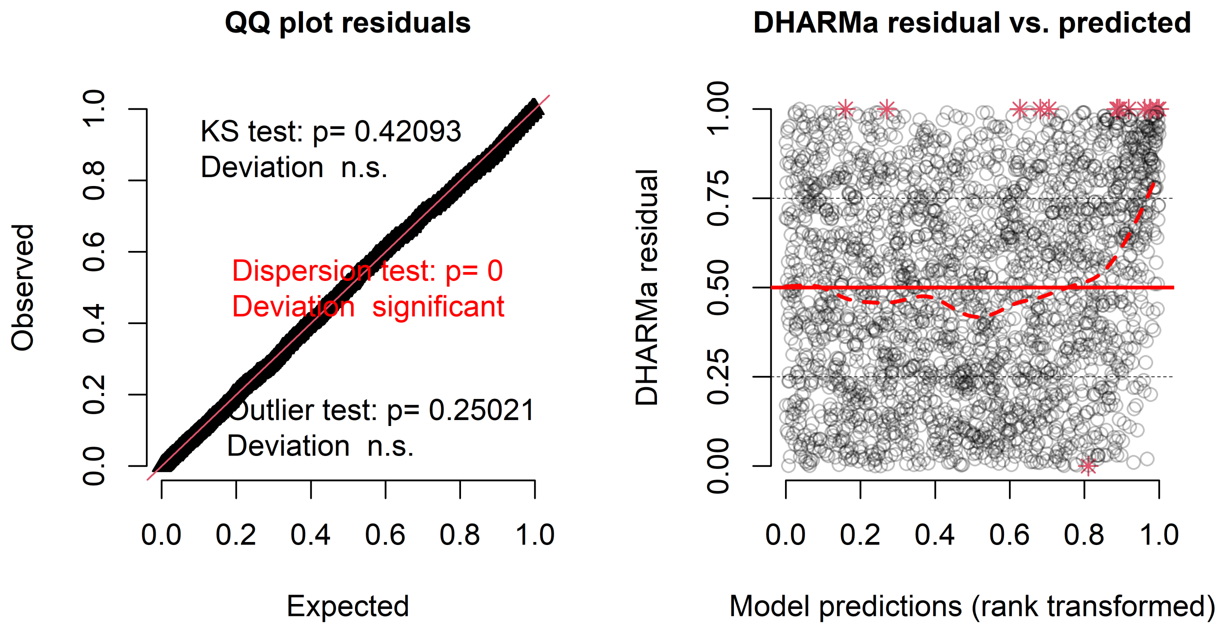


Figure 11: Diagnostics for the Tweedie distribution family model with year and depth as fixed effects, using DHARMa residuals.

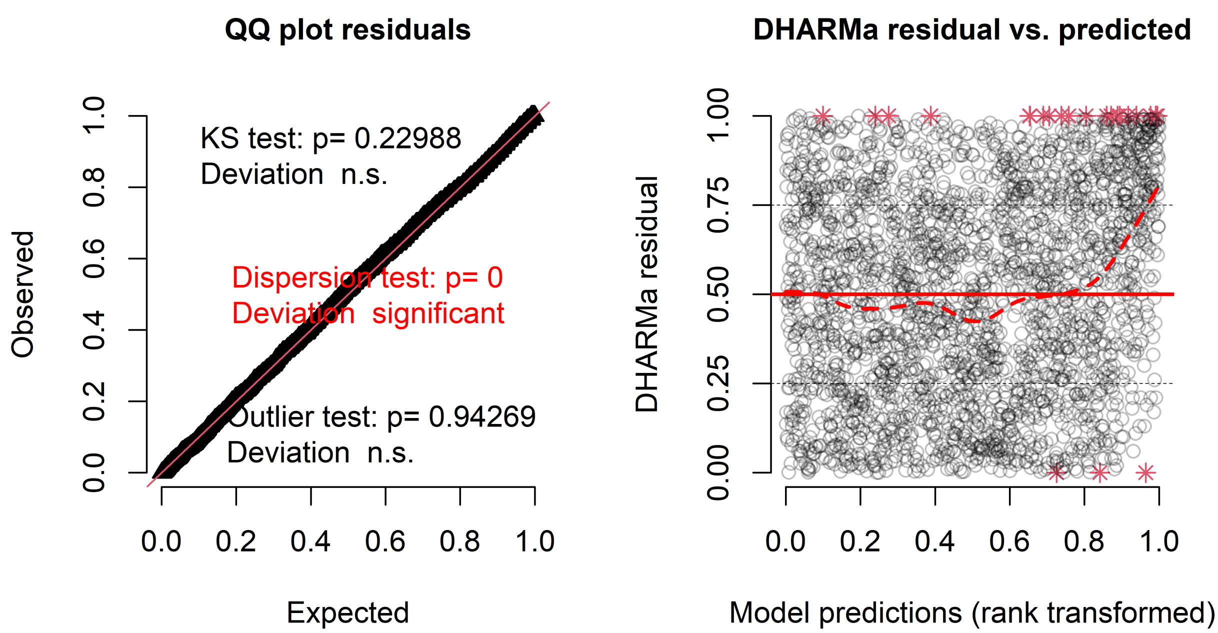


Figure 12: Diagnostics for the Tweedie distribution family model with year and a smooth on depth as fixed effects, using DHARMa residuals.

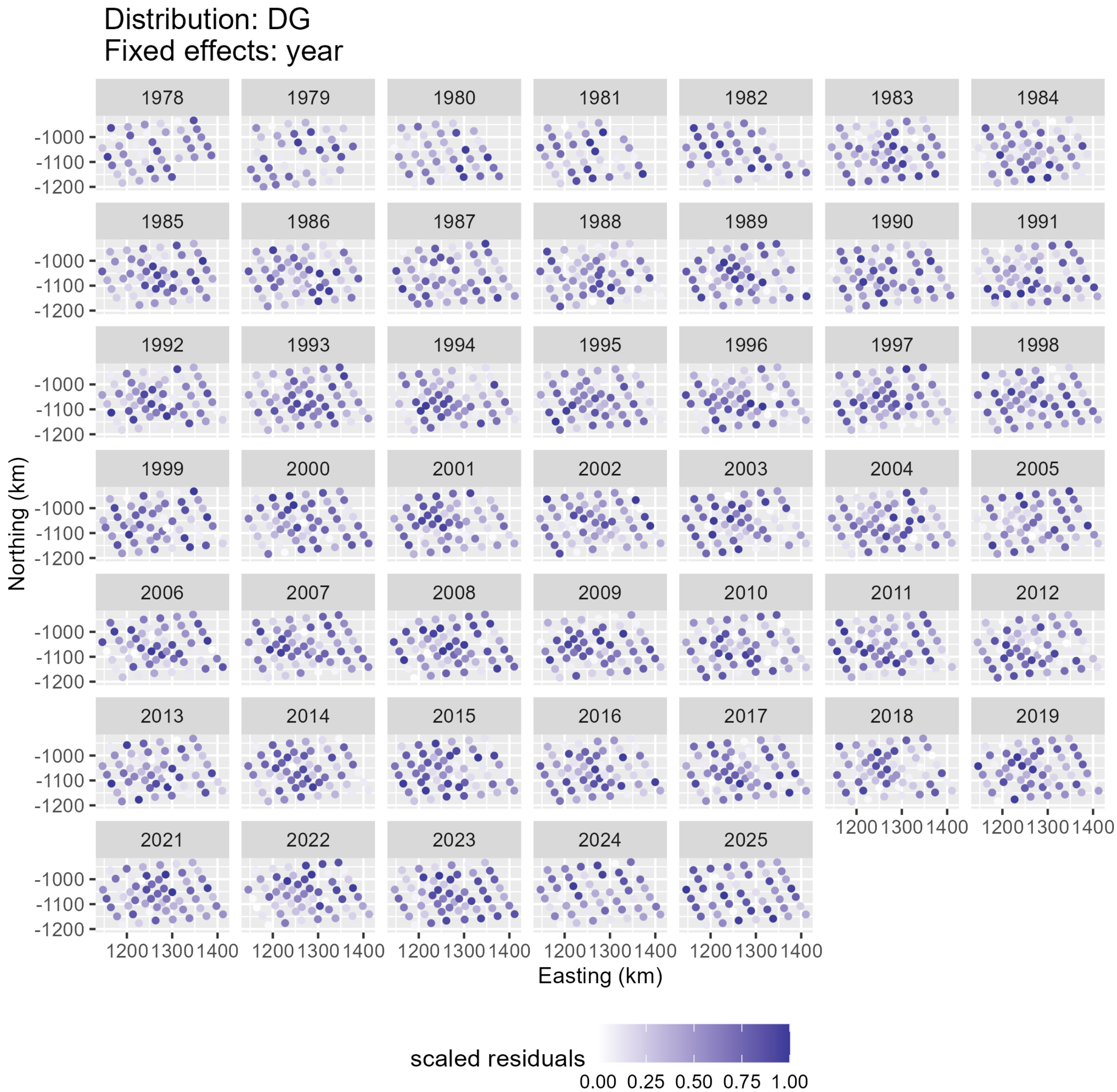


Figure 13: Spatial distribution of DHARMa residuals for the model using the delta gamma distribution with year as a fixed effect.

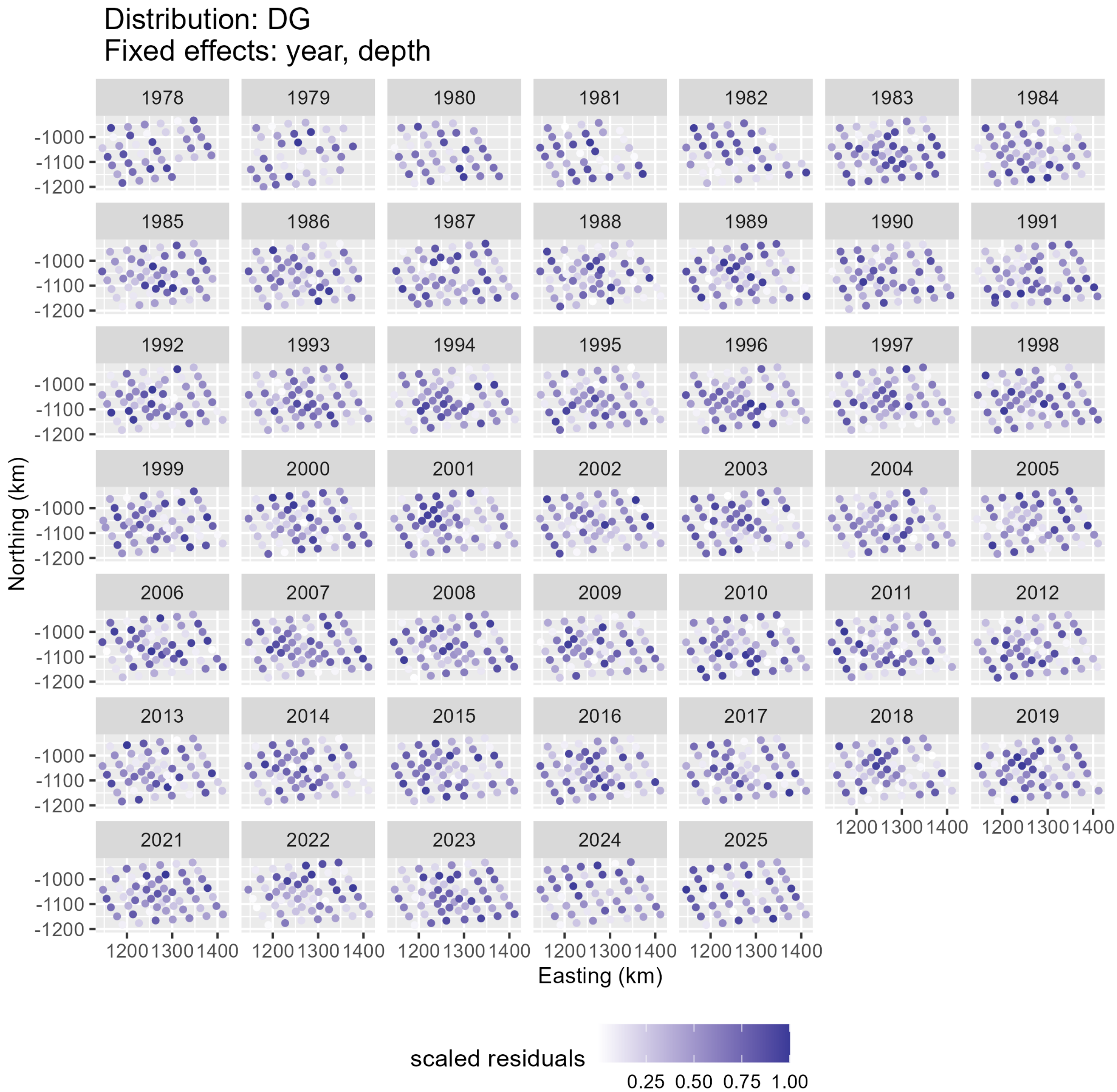


Figure 14: Spatial distribution of DHARMa residuals for the model using the delta gamma distribution with year and depth as fixed effects.

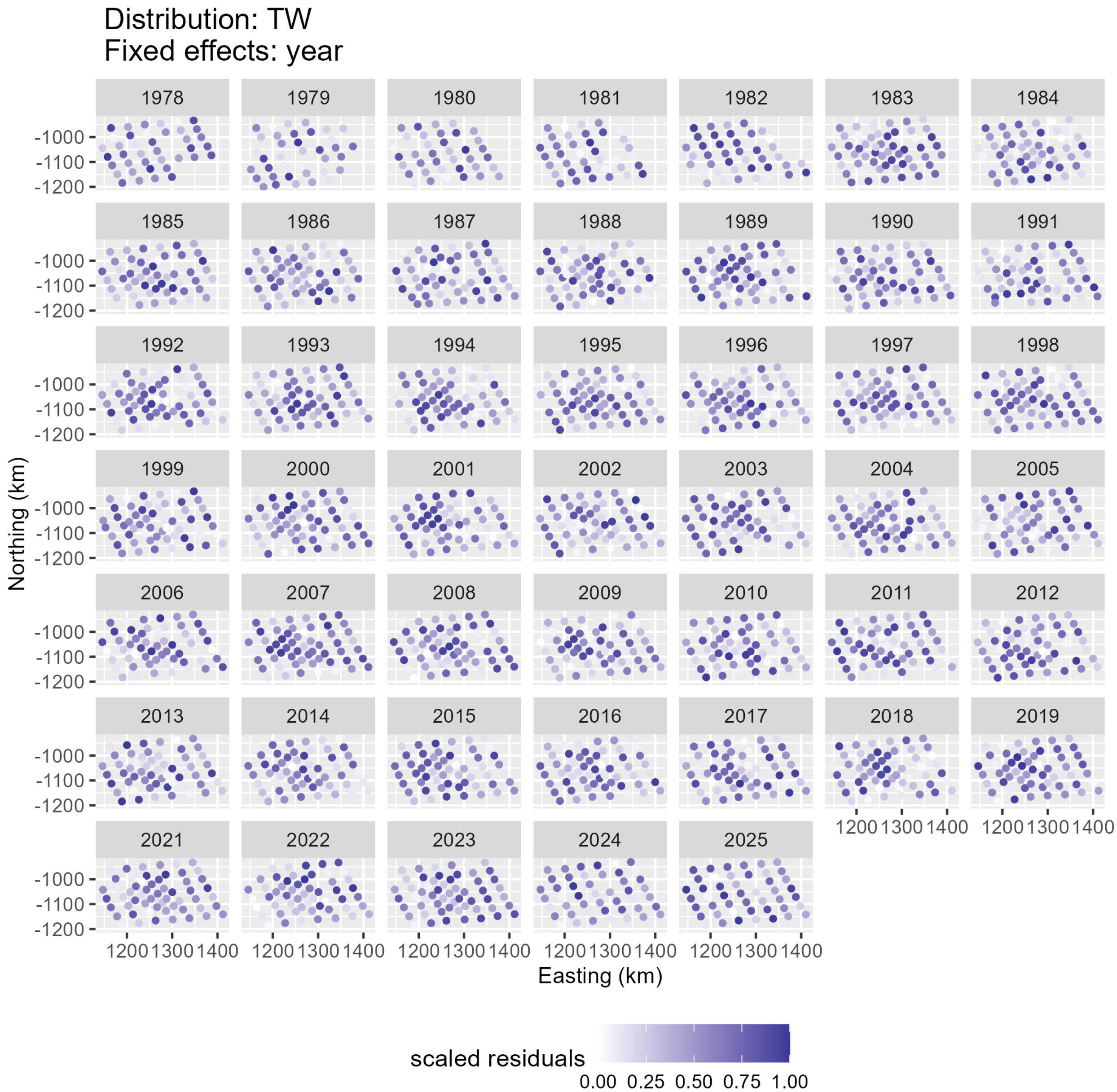


Figure 15: Spatial distribution of DHARMa residuals for the model using the Tweedie distribution with year as a fixed effect.

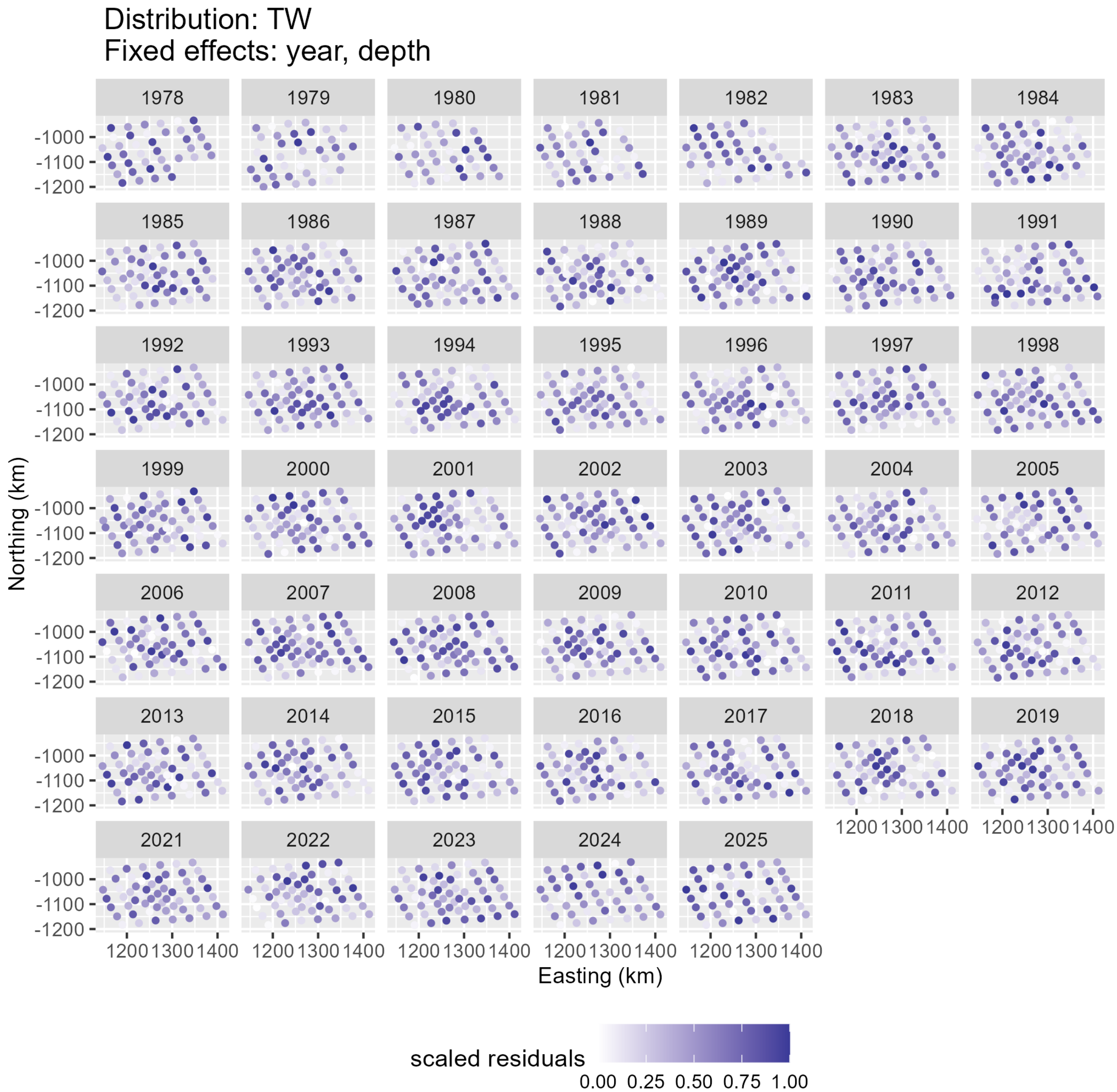


Figure 16: Spatial distribution of DHARMa residuals for the model using the Tweedie distribution with year and scaled depth as fixed effects.

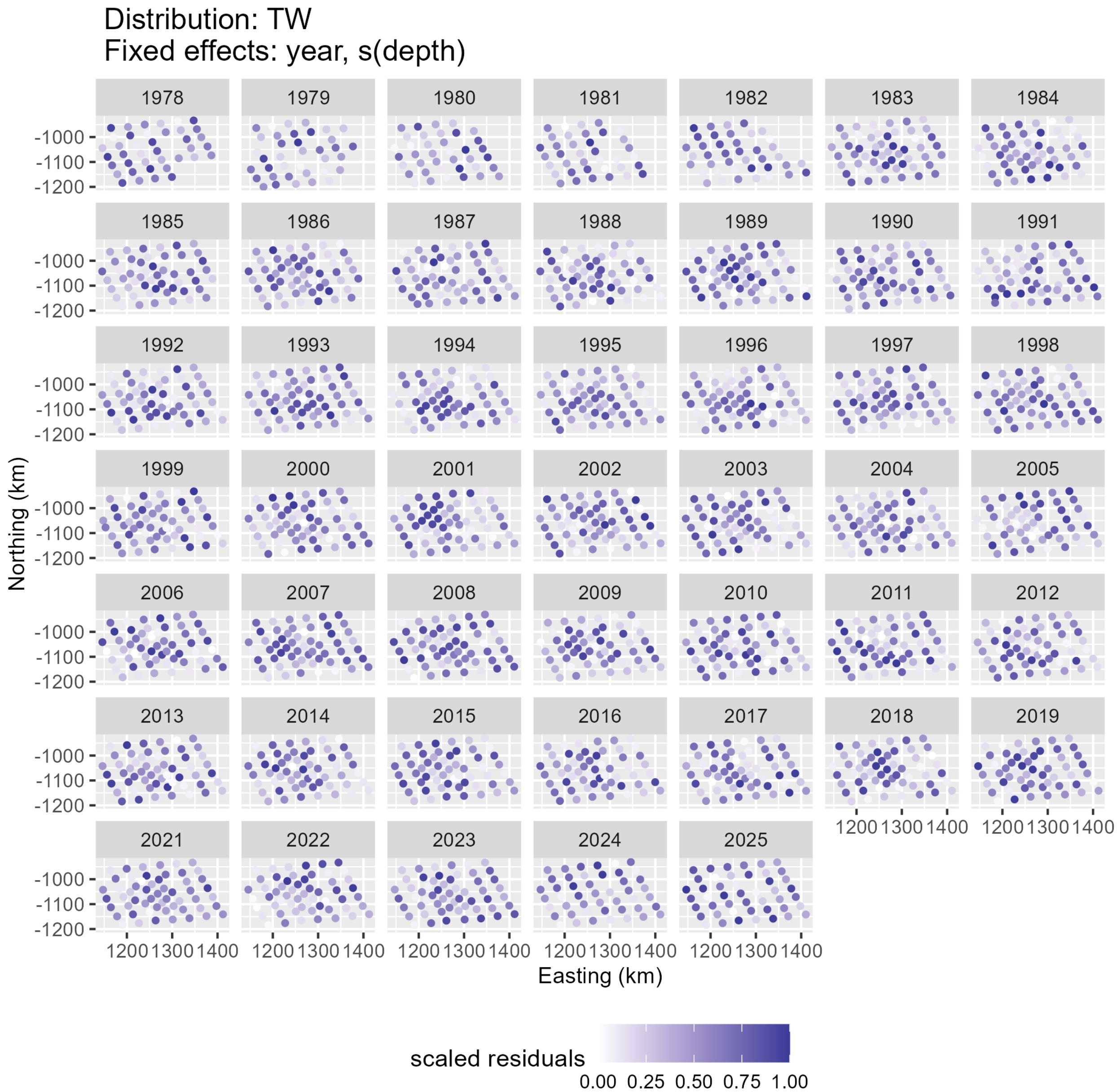


Figure 17: Spatial distribution of DHARMa residuals for the model using the Tweedie distribution with year and a smooth on depth as fixed effects.

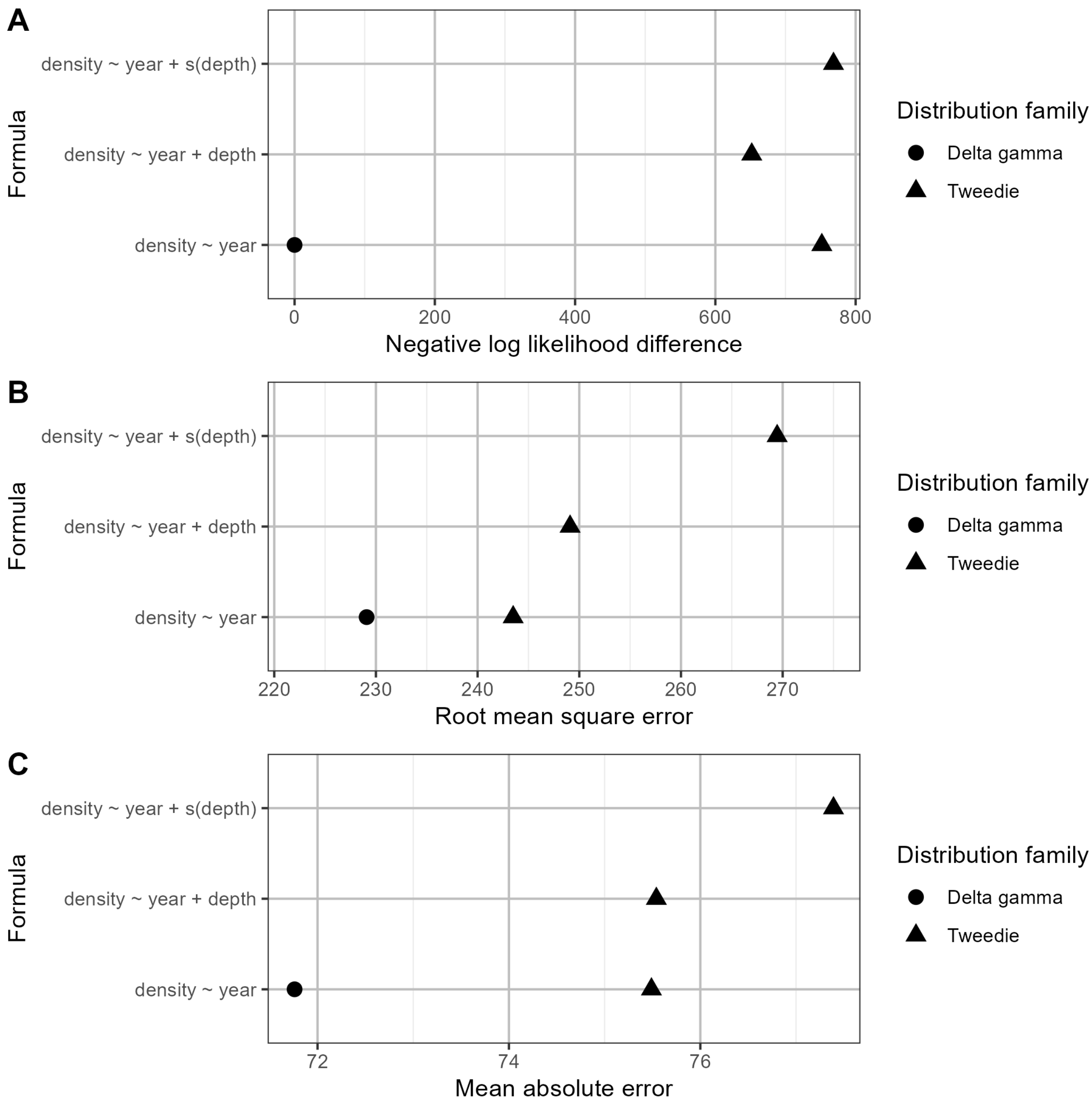


Figure 18: Metrics used in model selection, based on 10-fold cross validation. Note that one model using the delta gamma distribution family is shown and three using the Tweedie distribution family, because the delta gamma distribution family model with a smooth on depth failed to converge, and the cross-validation for the delta gamma distribution family model with scaled depth failed to converge. A) Negative log likelihood difference compared to the best-fitting model, which has NLL difference = 0, where smaller values indicate better model fit. B) Root mean square error, a measure of the average difference between observed and predicted values, where a lower value indicates better model fit. C) Mean absolute error, a measure of the average absolute value of the difference between observed and predicted values, where a lower value indicates better model fit.

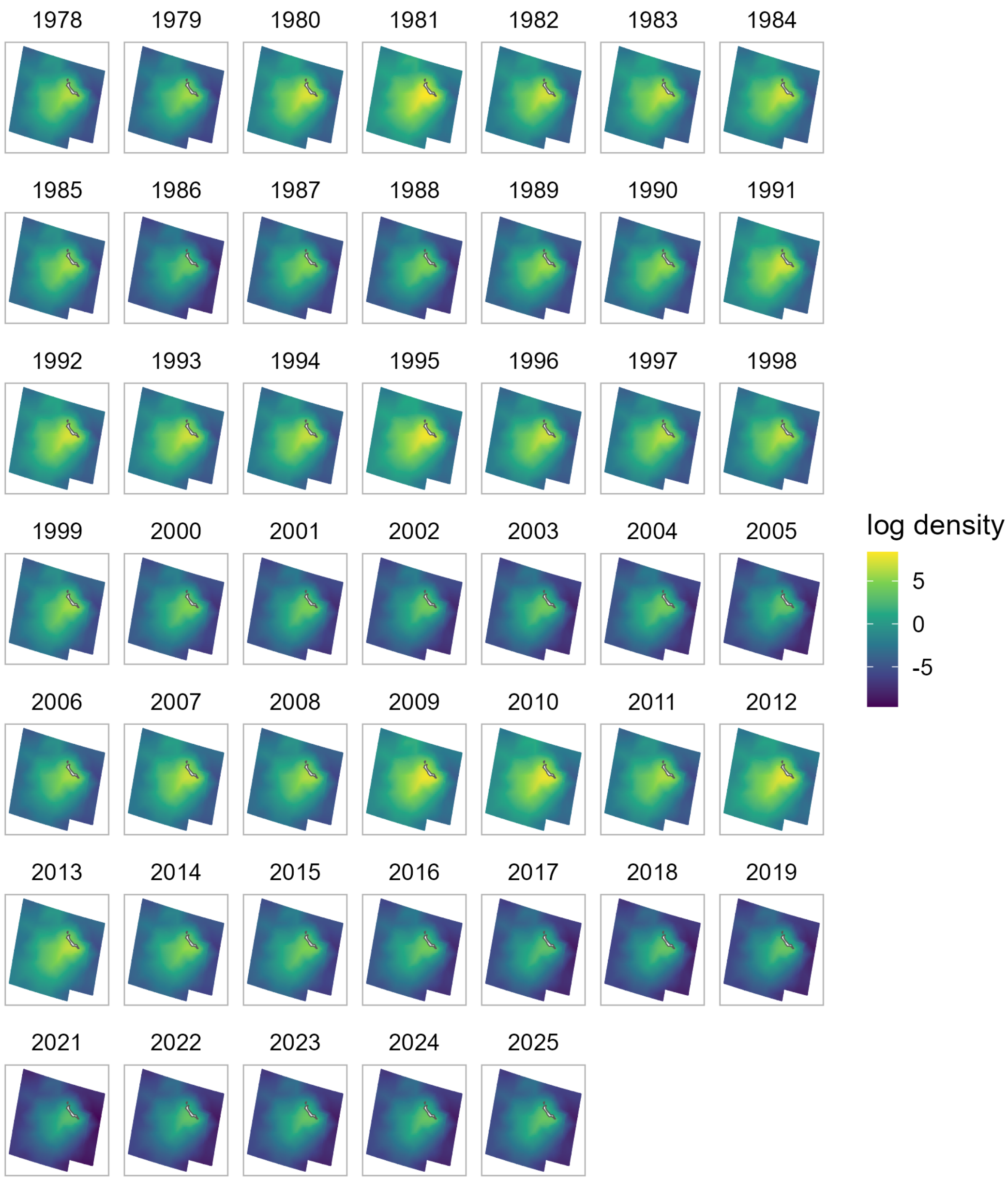


Figure 19: Spatial predictions of St. Matthew Island blue king crab males $\geq$ 90 mm in carapace length by year from the model using the delta gamma distribution with year as a fixed effect. These predictions are from the presence/absence component of the hurdle model.

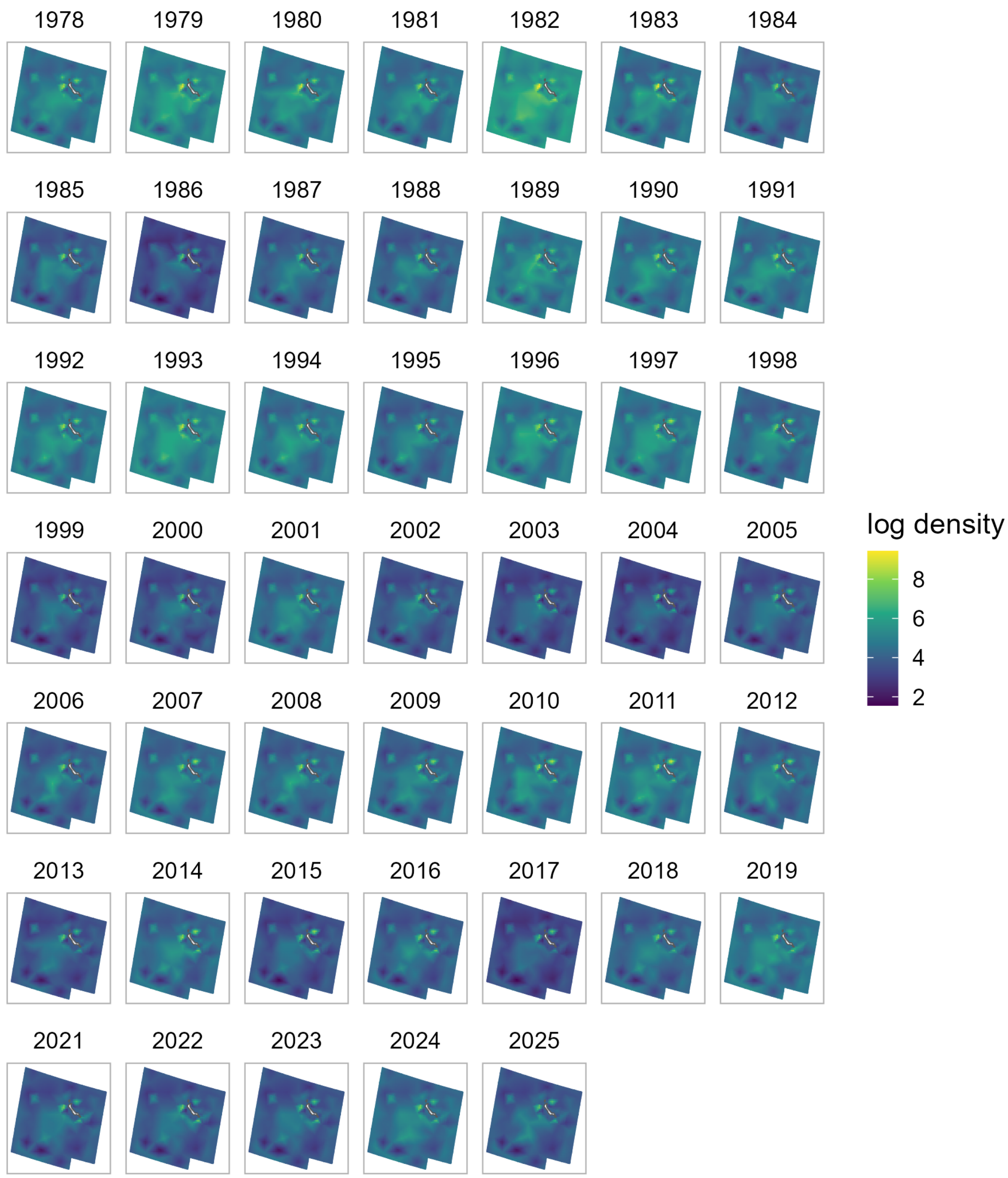


Figure 20: Spatial predictions of St. Matthew Island blue king crab males $\geq$ 90 mm in carapace length by year from the model using the delta gamma distribution with year as a fixed effect. These predictions are from the positive catch (abundance) component of the hurdle model.

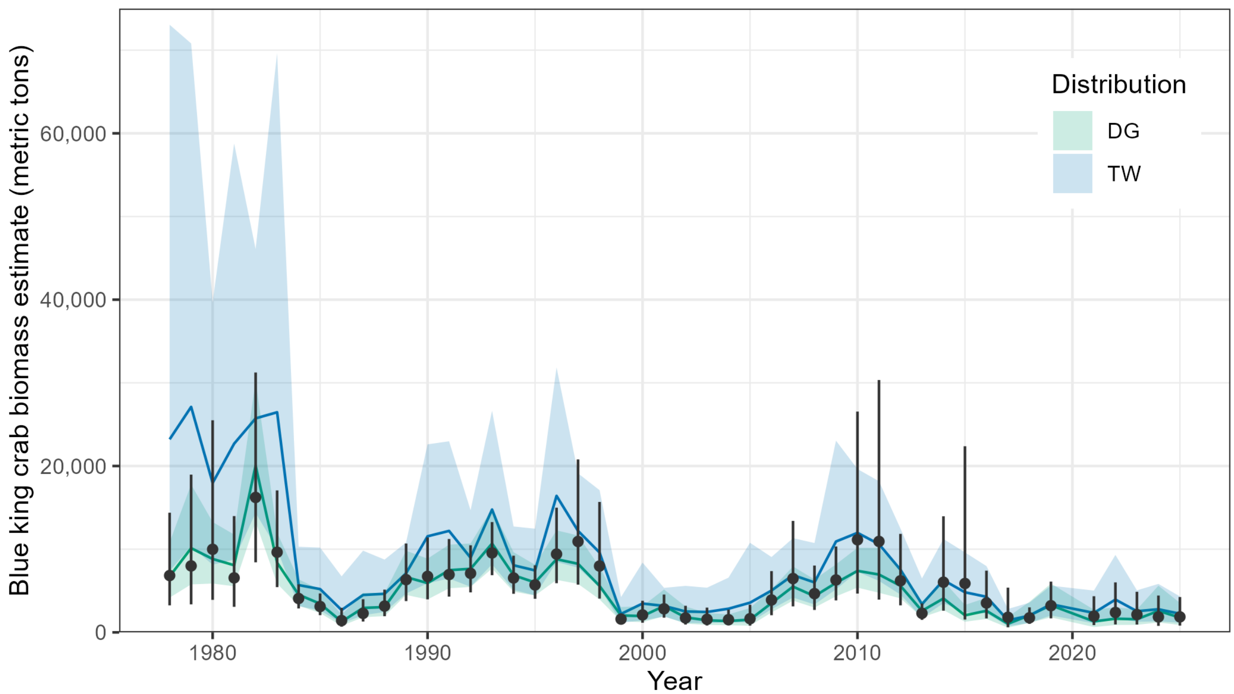


Figure 21: Estimated St. Matthew Island blue king crab biomass for males with carapace length $\geq$ 90 mm. Colored lines represent abundance ($\pm$95% CI) estimated using spatiotemporal models with either the delta gamma distribution family (green) or Tweedie distribution family (blue). Black points represent design-based biomass ($\pm$95% CI) estimates currently used in the stock assessment model.

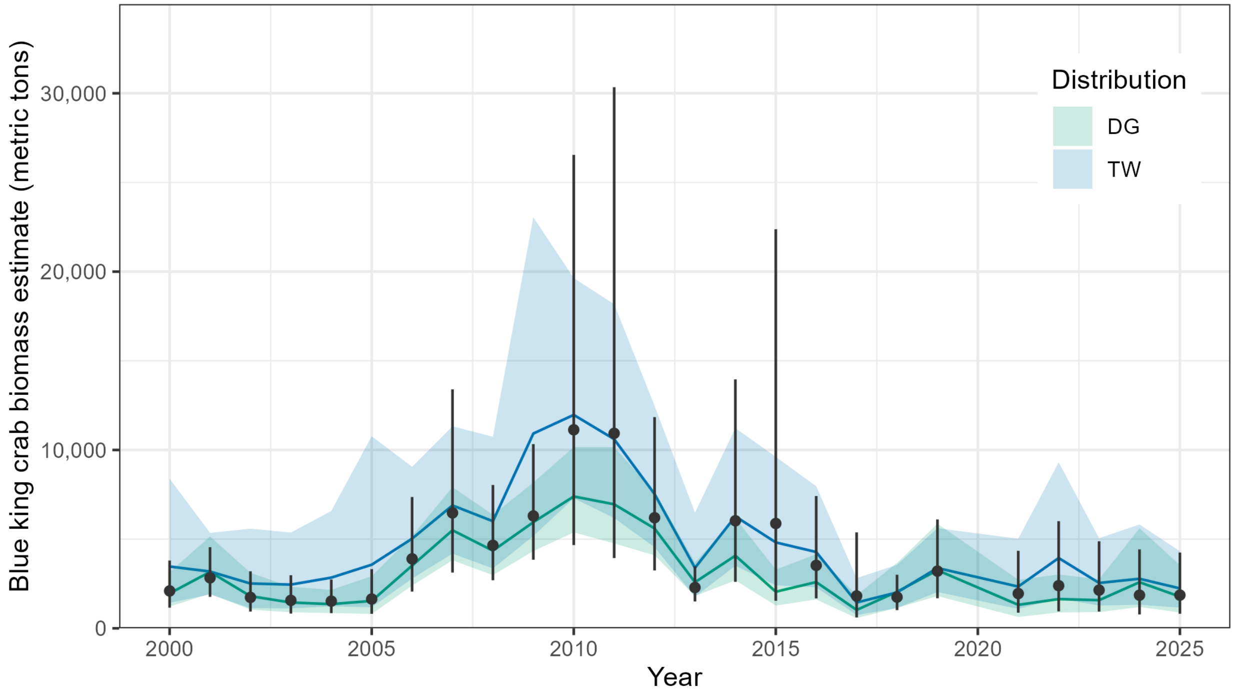


Figure 22: Estimated St. Matthew Island blue king crab biomass for males with carapace length $\geq$ 90 mm for the years 2000 - 2025 only. Colored lines represent abundance ($\pm$95% CI) estimated using spatiotemporal models with either the delta gamma distribution family (green) or Tweedie distribution family (blue). Black points represent design-based biomass ($\pm$95% CI) estimates currently used in the stock assessment model.