Developing a fishery-independent index of relative abundance for Gulf of Mexico White Shrimp using VAST

Lisa Ailloud, Molly Stevens, Brendan Turley, Adam Pollack, and David Hanisko

SEDAR87-AP-05

31 January 2025

Please cite this document as:

Ailloud, Lisa, Molly Stevens, Brendan Turley, Adam Pollack, and David Hanisko. 2025. Developing a fishery-independent index of relative abundance for Gulf of Mexico White Shrimp using VAST. SEDAR87-AP-05. SEDAR, North Charleston, SC. 24 pp.

Developing a fishery-independent index of relative abundance for Gulf of Mexico White Shrimp using VAST.

Lisa Ailloud[1], Molly Stevens[2], Brendan Turley[3], Adam Pollack[4], and David Hanisko[4]
[1]*Southeast Fisheries Science Center (SEFSC), National Marine Fisheries Service (NMFS), National Oceanographic and Atmospheric Administration (NOAA), Charleston SC, USA*
[2]*SEFSC, NMFS, NOAA, St. Petersburg FL, USA*
[3]*Cooperative Institute of Marine and Atmospheric Studies; University of Miami, Miami, FL*
[4]*SEFSC, NMFS, NOAA, Pascagoula, MS, USA*

1. Introduction

White shrimp (*Litopenaeus setiferus*) indices of relative abundance were developed for consideration in the SEDAR 87 Gulf of Mexico (GOM) white, pink, and brown shrimp stock assessment process for input into JABBA, a Bayesian state-space surplus production modelling tool (Winker, Carvalho, and Kapur 2018).

We used the Vector Autoregressive Spatio-Temporal (VAST) modeling platform to develop a single, continuous standardized index of abundance for white shrimp. VAST is an R package for implementing a spatial delta-generalized linear mixed model (delta-GLMM) when standardizing survey data. It allows for predicting population density based on both habitat covariates (i.e., covariates that affect population density) and spatial and spatiotemporal random effects. This flexible framework capable of accounting for changes in the spatiotemporal footprint of surveys through time. For white shrimp, the Louisiana Department of Wildlife and Fisheries (LDWF) trawl survey (SEDAR 2023) was used for developing a single, joint index of relative abundance for the stock. While the general design of the survey has not changed through time, the number of stations sampled has increased throughout the years (*Figure 1*) and relative monthly sampling intensity shifted through time (*Figure 2*). There was therefore an interest in using VAST to help correct for any bias that might arise from changes in the spatial footprint of the survey. In addition, since many studies point to the fact that environmental conditions (e.g., salinity and temperature) in the nursery grounds play an important role in driving the abundance and distribution of shrimp each year. We were interested in exploring whether the inclusion of environmental covariates could help explain the variability observed in the index through the used of habitat covariates in VAST.

This paper details the data, workflow, and results of applying the VAST modeling approach to Gulf of Mexico white shrimp data to produce an index of relative abundance for the stock.

2. Methods

2.1 Stock description

White shrimp in the GOM range from the Ochlochonee River, Florida, to Campeche, Mexico (SEDAR 2024a). They are commonly found in estuaries and coastal areas out to about 100 feet offshore. Adults spawn offshore in the spring through fall, and their larvae settle into nearshore estuarine or marsh nursery habitat in the spring. They then migrate back offshore in the fall as subadults when water temperature declines. White shrimp have a short life span of less than 2 years (SEDAR 2024a). The bulk of the US Gulf of Mexico white shrimp catches occur in the inshore and offshore waters of Louisiana (Williams et al. 2024).

2.2 Data

2.2.1 Fishery Independent Surveys

A number of fishery-independent survey data time series were made available for SEDAR 87, and vetted during the Data Workshop (see Data Workshop Report (SEDAR 2024b) for more detail). A single data source was retained for developing the index of relative abundance for the stock : the Louisiana Department of Wildlife and Fisheries (LDWF) survey (**ldwf2023?**). White shrimp are present in inshore waters throughout the year as overwintering adults in spring months and as new recruits in summer months, giving the best option for an abundance index that includes all size classes. The SEAMAP survey was deemed to operate too far offshore to accurately track the bulk of the stock (SEDAR 2024b). Other state surveys operating with similar survey designs as the LDWF (MS, AL) were considered during the Data Workshop but the analytical team ultimately decided not to use these datasets due to the shorter time scales available and changes in gear, protocol, and sampling timing that were not well documented.

The LDWF survey is a fishery-independent 16-foot inshore marine otter trawl survey that is conducted monthly at fixed sampling stations in coastal LA. The dataset consists of 10 minute tows and spans 1980-2023 (SEDAR 2023). In October 2010, additional fixed stations were added to this survey allowing more spatial coverage within each coastal area (*Figure 1*).

2.2.2 Time Series of Environmental Data

Indices of mean annual temperature and salinity in the nursery grounds were developed by Turley, Ailloud, and Stevens (2023), and used to test the impact of environmental conditions on stock abundance. These indices were constructed based on the timing and regions that are thought to supply the bulk of the commercial catches for white shrimp to capture the variability and trends of the environmental drivers (salinity and temperature) that are most likely to influence recruitment success and, as a result, exploitable biomass. These indices were input into VAST as habitat covariates to test whether their inclusion helped explain a portion of the year-to-year variability observed in the index of relative abundance. A set of lagged indices (one and two years) were also tested. The median of the time series was used as substitutes for the initial years that preceded lagged data availability (i.e. 1980 for the 1-year lag time series and 1980-1981 for the 2-year lag time series).

2.3 Modeling Approach

2.3.1 Model Description

We used the Vector Autoregressive Spatiotemporal (VAST) model v3.11.0 (Thorson and Barnett 2017) implemented in R v.4.4.0 (R Core Team, 2020) to develop indices of relative abundance for each stock. This spatial-delta generalized linear mixed effects model is defined by two linear predictors : p_1 to calculate encounter probabilities (or zero-inflation probabilities in a count-data model) and p_2 for positive catch rates in numbers of shrimp per tow (or the count-data intensity function in a count-data model) for each observation i at location s_i and time t_i.

$$p_1(i) = \beta_1(t_i) + \omega_1^*(s_i) + \epsilon_1^*(s_i, t_i) + \gamma_1(t_i) + \lambda_1(i)$$

$$p_2(i) = \beta_2(t_i) + \omega_2^*(s_i) + \epsilon_2^*(s_i, t_i) + \gamma_2(t_i) + \lambda_2(i)$$

Where β are the year intercepts for each linear predictor (temporal variation, fixed effect), ω^* are the spatial factors (random effects), ϵ^* are the spatio-temporal factors (random effects), γ represent the impacts of habitat covariates on each linear predictor and λ represent the impacts of catchability covariates on each linear predictor. Linear predictors are then transformed to predict encounter probabilities and positive catch rates. In the case of a conventional delta-model :

$$r_1 = logit^{-1}\big(p_1(i)\big)$$

$$r_2 = a_i * exp\big(p_2(i)\big)$$

where a_i is the effort offset (in minutes fished). For white shrimp, no effort offset was included since all tows were of equal duration. Then, VAST predicts population density $d(s,t)$ at extrapolation grids across a continuous spatial domain (pre-defined to cover the extent of the surveys) and discrete time intervals (years) from both transformed linear predictors as (in the case of a conventional delta-model):

$$d(s,t) = logit^{-1}\big(\beta_1(t_i) + \omega_1^*(s_i) + \epsilon_1^*(s_i,t_i) + \gamma_1(t_i)\big) * exp\big(\beta_2(t_i) + \omega_2^*(s_i) + \epsilon_2^*(s_i,t_i) + \gamma_2(t_i)\big)$$

and calculates the abundance index $I(t)$ in time t as :

$$I(t) = \sum_{s=1}^{n_s}\big(a(s) * d(s,t)\big)$$

where $a(s)$ is the area associated with knot s. Habitat covariates are processes that affect true underlying densities, they are used in both the fitting process and the prediction process. Catchability covariates are processes that affect sampling but do not reflect underlying densities, their estimated effect is removed when calculating the index. Estimates of "density" for white shrimp are not directly interpretable on an absolute scale as no meaningful estimate of area swept could be produced for these surveys.

Spatial and spatiotemporal variation are estimated using Gaussian Markov random fields (Thorson 2019a). The spatial correlation matrix is modeled using a Matérn covariance function (Lindgren, Rue, and Lindström 2011). The user can specify a Matérn covariance function that is either isotropic (where correlations decline at the same rate in any direction) or anisotropic (where the rate at which correlations decline depends upon the direction of movement – common in marine ecosystems, where correlations decline slowly when moving along a depth contour but rapidly when moving perpendicular to the depth contour (J. T. Thorson 2019a)). The number of prediction locations (i.e. knots) must be predefined. VAST then distributes the knots spatially using a k-means algorithm that minimizes the average distance between samples and knots. A detailed description of the model can be found in (Thorson 2019b) and in the VAST User Manual (Thorson 2024).

2.3.2 Prediction grid

Before fitting VAST models, we constructed prediction grids for the study species. A 3km x 3km extrapolation grid was constructed to encompass the total area covered by the survey, excluding land and rivers (*Figure 3*).

2.3.3 Modeling Workflow
Model development involved the following steps:

1. *Selecting the optimal number of knots :* a delta-lognormal model with both spatial and spatiotemporal random effects and no habitat or catchability covariates was run using a series of knots from 250 to 1500 knots to determine the optimal number of knots to use while balancing computational speed and spatial resolution. The root mean square error (RMSE) between raw observations and fitted values of the index was used to measure differences between runs and find the number of knots beyond which results appeared to stabilize.

2. *Identifying the optimal distribution model :* Candidate models included :

   – a Poisson-link delta-model (*ObsModel=c(2,1)*) : log-linked linear predictor for encounter probability (i.e., Poisson), and a gamma error distribution for positive catch rates;

   – a Zero-inflated negative binomial model (*ObsModel=c(5,0)*) : 1st linear predictor for logit-linked zero-inflation, 2nd linear predictor for log-linked conditional mean of negative binomial;

   – a Conventional lognormal delta-model (*ObsModel=c(4,0)*) : logit-link for encounter probability and log-link for positive catch rates.

   Standard model selection tools (i.e., AIC, qq plots) were used to identify the best fitting model.

3. *Testing the inclusion of spatial and spatiotemporal random effects :* Using the optimal number of knots and distribution function determined in steps 1. and 2., a series of models were built with increasing complexity with respect to including spatial and spatiotemporal random effects. If the estimated variances of the spatial and spatiotemporal terms were greater than zero, the random effects were retained. Random effects, when present, were included in both linear predictors. AIC was used to select the Matérn covariance functions (anisotropic vs. isotropic).

4. *Selecting catchability covariates :* Using the optimal model determined in step 3., we explored the impact of including the following catchability covariates: month. Catchability covariates were included in both linear predictors. AIC was used to assess model fit.

5. *Selecting habitat covariates :* Using the optimal model determined in step 4., we explored the impact of including the following habitat covariates: indices of mean annual temperature and salinity in the nursery grounds and these same indices lagged by 1 and 2 years. Indices were normalized to a mean of zero and standard deviation of 1 prior to input into the model and set up as spatially explicit annual zero-centered covariates to predict changes in density across space and time (i.e., single value across space within a year but varied from year to year). Habitat covariates were included in both linear predictors. AIC was used to assess model (m) fit. Following (Cao et al. 2017), a pseudo-

R^2 metric was calculated to determine the proportion of variance (σ^2) from the null model ($null$; i.e., habitat covariates) that was explained by including habitat variables :

$$pseudo - R^2 = 1 - \frac{\sigma^2_{\omega,m} + \sigma^2_{\epsilon,m}}{\sigma^2_{\omega,null} + \sigma^2_{\epsilon,null}}$$

3. Results

1. *Selecting the optimal number of knots :* The RMSE reached an asymptote around 500 knots (*Figure 4*) and there was very little difference in the estimated index using 500 (runtime : 0.4hrs) vs. 1500 knots (run time : 1.6hrs) (*Figure 5*). A model with 700 knots was therefore retained for subsequent model building steps.

2. *Identifying the optimal distribution model :* The zero-inflated negative binomial model did not converge. Comparing the poisson-link delta-model (VAST default option) with the conventional lognormal delta-model, AIC and qqplots indicated that the delta-lognormal model was preferred (*Figure 6*).

3. *Testing the inclusion of spatial and spatiotemporal random effects :* The marginal standard deviation of both spatial (ω) and spatiotemporal (ϵ) random effects were significantly greater than zero and thus included in the final model (ω_1 = 1.44; ω_2 = 0.97; ϵ_1 = 0.53; ϵ_2 = 0.33). The model that assumed anisotropic covariance was favored in terms of AIC (*Table 1*, *Figure 7*). Including spatial and spatio-temporal random effects brought significant improvements to the model fit.

4. *Selecting catchability covariates :* The model with the lowest AIC was the model that included time of day, survey and month as catchability covariates (*Table 2*). Month had the largest effect, followed by time of day (*Figure 8*). The fixed effect for survey (Q) had only a minor impact on the overall fit. Effects plots for month are shown in *Figure 9*.

5. *Selecting habitat covariates :* The model with the lowest AIC was the model that included the average nursery temperature index with a 1 year lag as habitat covariate (*Table 3*). However, it resulted in a nearly identical fit as the null model with no habitat covariates ($\Delta AIC = 21$; *Figure 10*) and only explained a trivial fraction (<4%) of the spatial and spatiotemporal variation observed. None of the habitat covariates explained any considerable amount of spatial or spatiotemporal variation in the model.

The final model included spatial and spatiotemporal random effects using the anisotropic estimation of correlation (*Figure 11*), and a catchability covariate for month, in both the 1st and 2nd linear predictors. The use of spatial REs had the largest effect in the standardization process. None of the habitat covariates had an appreciable impact on the index trend and each explained only a very small portion of the variance from the null model with no habitat covariates. Because of that and the fact that this index is meant for assessment purposes and benefits from being simple to update, the final model did not include any habitat covariates. *Figure 12* shows the final standardized index plotted against the non-standardized index. The distribution of quantile residuals did not show any substantial spatial pattern (*Figure 13*). The model fit the data reasonably well (*Figure 14*).

4. Discussion

The standardization process resulted in an index with an, overall, steeper increase across the time series (*Figure 12*). Index values prior to the survey expansion (pre 2020) generally fell below the nominal values and index values post expansion generally fell above.

While many studies point to the relationship between shrimp production and temperature and salinity on the nursery grounds (see SEDAR (2024c)), these variables were not found to have an appreciable impact on spatial variation in shrimp density. It is also likely that the relationship between CPUE and environmental conditions is not linear and would be better described using a spline where shrimp production is maximized under certain conditions but hampered under extremes. At the time that this work was conducted, it was not possible in VAST to specify a more complex relationship for the environmental index so that option was not explored.

The final model including spatial and spatiotemporal random effects and a catchability covariate for month for both the 1st and 2nd linear predictors but no habitat covariate is recommended as the index for input into JABBA.

References

Cao, Jie, James T. Thorson, R. Anne Richards, and Yong Chen. 2017. "Spatiotemporal Index Standardization Improves the Stock Assessment of Northern Shrimp in the Gulf of Maine." *Canadian Journal of Fisheries and Aquatic Sciences* 74 (11): 1781–93. *https://doi.org/10.1139/cjfas-2016-0137*.

Lindgren, Finn, Håvard Rue, and Johan Lindström. 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 (4): 423–98. *https://doi.org/10.1111/j.1467-9868.2011.00777.x*.

SEDAR. 2023. "SEAMAP Trawl Shrimp Data and Index Estimation Work Group Report." North Charleston, SC. *https://sedarweb.org/documents/sedar-rd01-seamap-trawl-shrimp-data-and-index-estimation-work-group-report/*.

SEDAR. 2024a. "SEDAR 87 Gulf of Mexico White, Pink, and Brown Shrimp SECTION II: Data Workshop Report." SEDAR. *https://sedarweb.org/documents/sedar-87-gulf-of-mexico-white-pink-and-brown-shrimp-data-workshop-report/*.

SEDAR. 2024b. "SEDAR 87 Gulf of Mexico White, Pink, and Brown Shrimp SECTION II: Data Workshop Report." *https://sedarweb.org/documents/sedar-87-gulf-of-mexico-white-pink-and-brown-shrimp-data-workshop-report/*.

SEDAR. 2024c. "SEDAR 87 Gulf of Mexico White, Pink, and Brown Shrimp SECTION II: Data Workshop Report." *https://sedarweb.org/documents/sedar-87-gulf-of-mexico-white-pink-and-brown-shrimp-data-workshop-report/*.

Thorson, James. 2024. "VAST Manual." *https://github.com/James-Thorson-NOAA/VAST/blob/main/manual/VAST_model_structure.pdf*.

Thorson, James T. 2019a. "Guidance for Decisions Using the Vector Autoregressive Spatio-Temporal (VAST) Package in Stock, Ecosystem, Habitat and Climate Assessments." *Fisheries Research* 210 (February): 143–61. *https://doi.org/10.1016/j.fishres.2018.10.013*.

Thorson, James T. 2019b. "Guidance for Decisions Using the Vector Autoregressive Spatio-Temporal (VAST) Package in Stock, Ecosystem, Habitat and Climate Assessments." *Fisheries Research* 210 (February): 143–61. *https://doi.org/10.1016/j.fishres.2018.10.013*.

Thorson, James T., and Lewis A. K. Barnett. 2017. "Comparing Estimates of Abundance Trends and Distribution Shifts Using Single- and Multispecies Models of Fishes and Biogenic Habitat." Edited by Emory Anderson. *ICES Journal of Marine Science* 74 (5): 1311–21. *https://doi.org/10.1093/icesjms/fsw193*.

Turley, Brendan, Lisa Ailloud, and Molly Stevens. 2023. "Development of Estuarine Environmental Indices for SEDAR 87 Gulf of Mexico White, Pink, and Brown Shrimp Stock Assessment." North Charleston, SC. *https://sedarweb.org/documents/sedar-87-ap-01-development-of-estuarine-environmental-indices-for-sedar-87-gulf-of-mexico-white-pink-and-brown-shrimp-stock-assessment/*.

Williams, Jo, Kimberly Johnson, Kyle Detloff, and Alan Lowther. 2024. "SEDAR 87 Commercial Fishery Landings and Effort Figures for White, Pink, and Brown Shrimp in the US Gulf of Mexico, 1960-2022." North Charleston, SC. *https://sedarweb.org/documents/sedar-84-dw-16-sedar-84-commercial-fishery-landings-and-effort-figures-for-white-pink-and-brown-shrimp-in-the-us-gulf-of-mexico-1960-2021/*.

Winker, Henning, Felipe Carvalho, and Maia Kapur. 2018. "JABBA: Just Another Bayesian Biomass Assessment." *Fisheries Research* 204 (August): 275–88. *https://doi.org/10.1016/j.fishres.2018.03.010*.

Tables

Run Description	$\sigma_{\omega 1}$	$\sigma_{\omega 2}$	$\sigma_{\varepsilon 1}$	$\sigma_{\varepsilon 2}$	AIC	ΔAIC
No RE					381,862	18,446
Spatial RE (iso)	1.43	0.88			365,112	1,695
Spatial (iso) & spatiotemporal REs	1.43	0.97	0.52	0.33	363,423	6
Spatial (aniso) & spatiotemporal REs	**1.44**	**0.97**	**0.53**	**0.33**	**363,416**	**0**

Table 1. Marginal standard deviation of spatial (ω) and spatiotemporal (ϵ) terms and AIC across runs with different specifications for the spatial and spatiotemporal random effects and associated Matérn covariance function. The run with the lowest AIC is bolded. *RE: random effects; iso : isotropic Matérn covariance function; aniso: anisotropic Matérn covariance function.*

Run Description	$\sigma_{\omega 1}$	$\sigma_{\omega 2}$	$\sigma_{\varepsilon 1}$	$\sigma_{\varepsilon 2}$	AIC	ΔAIC
Null model	1.44	0.97	0.53	0.33	363,416	7,809
Null model + month	**1.56**	**1.01**	**0.58**	**0.36**	**355,607**	**0**

Table 2. Marginal standard deviation of spatial (ω) and spatiotemporal (ϵ) terms and AIC across runs with different specifications for the catchability covariates. The run with the lowest AIC is bolded.

Run Description	$\sigma_{\omega 1}$	$\sigma_{\omega 2}$	$\sigma_{\varepsilon 1}$	$\sigma_{\varepsilon 2}$	pseudoR^2_1	pseudoR^2_2	AIC	ΔAIC
Null model	1.56	1.01	0.58	0.36	0.000	0.000	355,607	21
Null model + temperature	1.56	1.00	0.58	0.35	0.003	0.008	355,603	17
Null model + temperature (lag=1)	**1.55**	**0.99**	**0.57**	**0.35**	**0.013**	**0.038**	**355,586**	**0**
Null model + salinity (lag=1)	1.56	1.01	0.57	0.35	0.002	0.006	355,596	10
Null model + temperature (lag=2)	1.56	1.01	0.57	0.36	0.008	0.003	355,605	18
Null model + salinity (lag=2)	1.57	1.01	0.57	0.36	-0.005	0.002	355,598	12

Table 3. Marginal standard deviation of spatial (ω) and spatiotemporal (ϵ) terms and pseudo-R2 showing the proportion of variance from the null model (i.e., the model with no habitat covariates included) that is explained by including habitat covariate(s) in the model. The run with the lowest AIC is bolded.

Figures

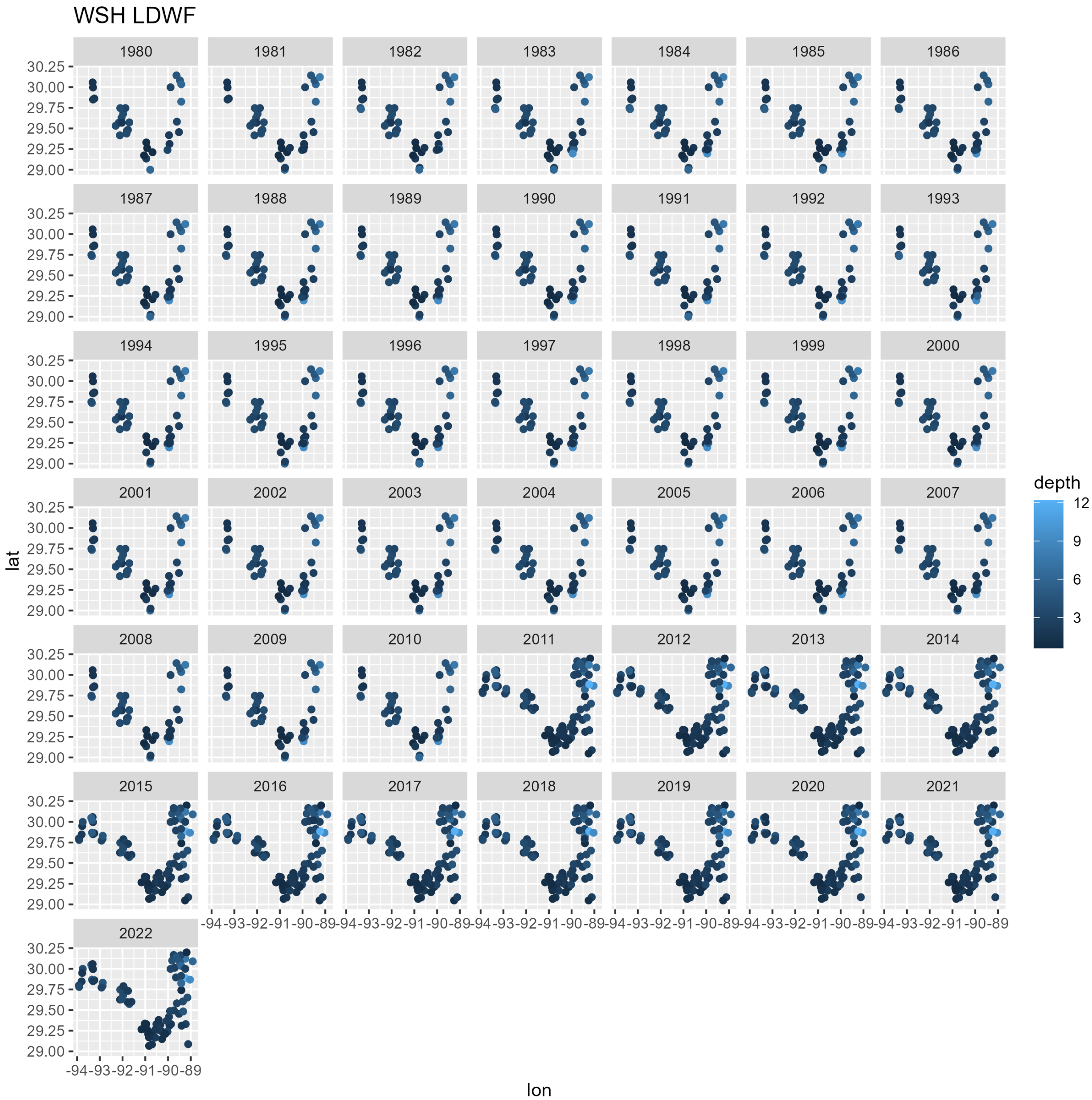

Figure 1: Change in the spatial footprint of the LDWF survey through time.

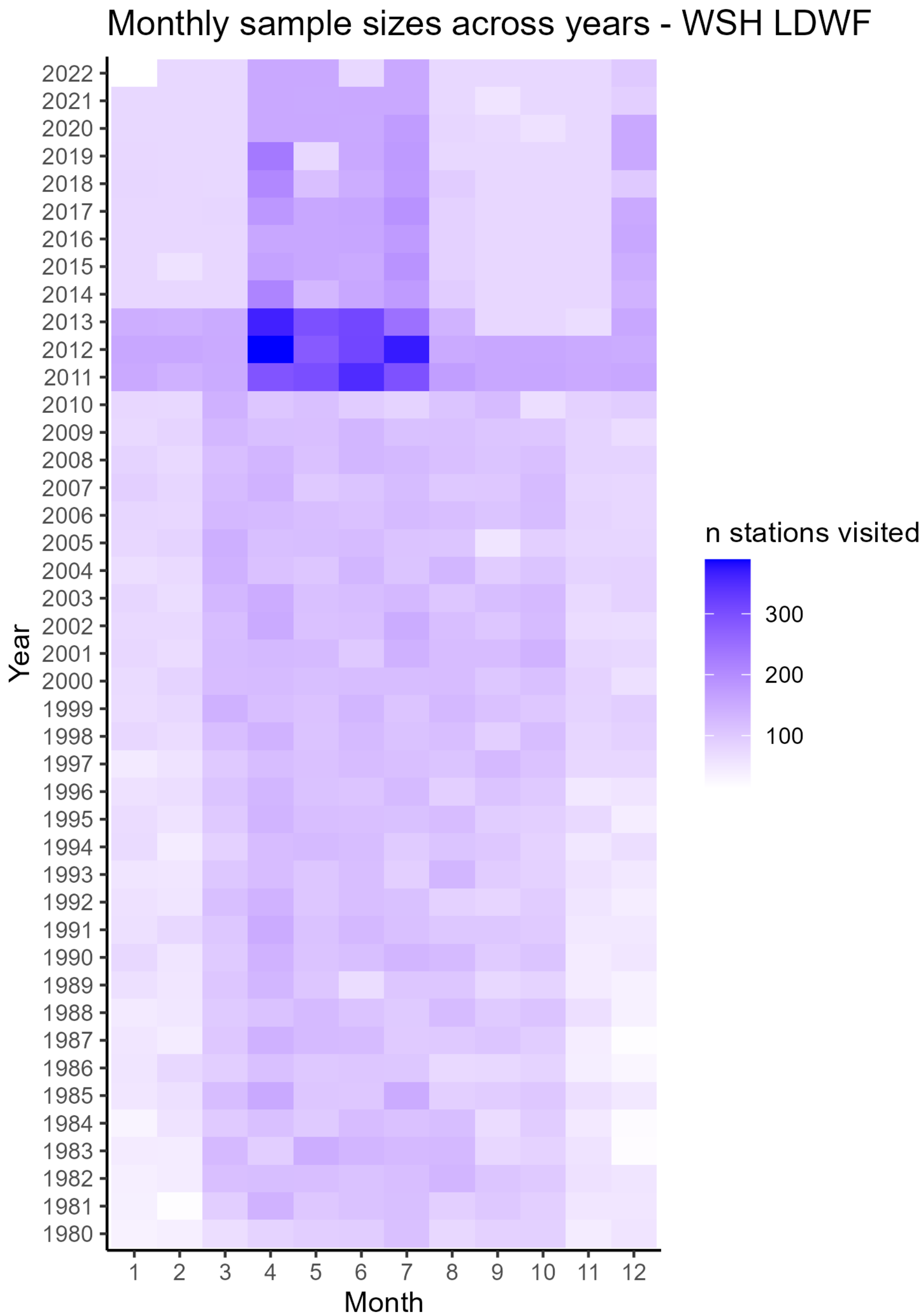

Figure 2: Sampling intensity of the LDWF survey through time.

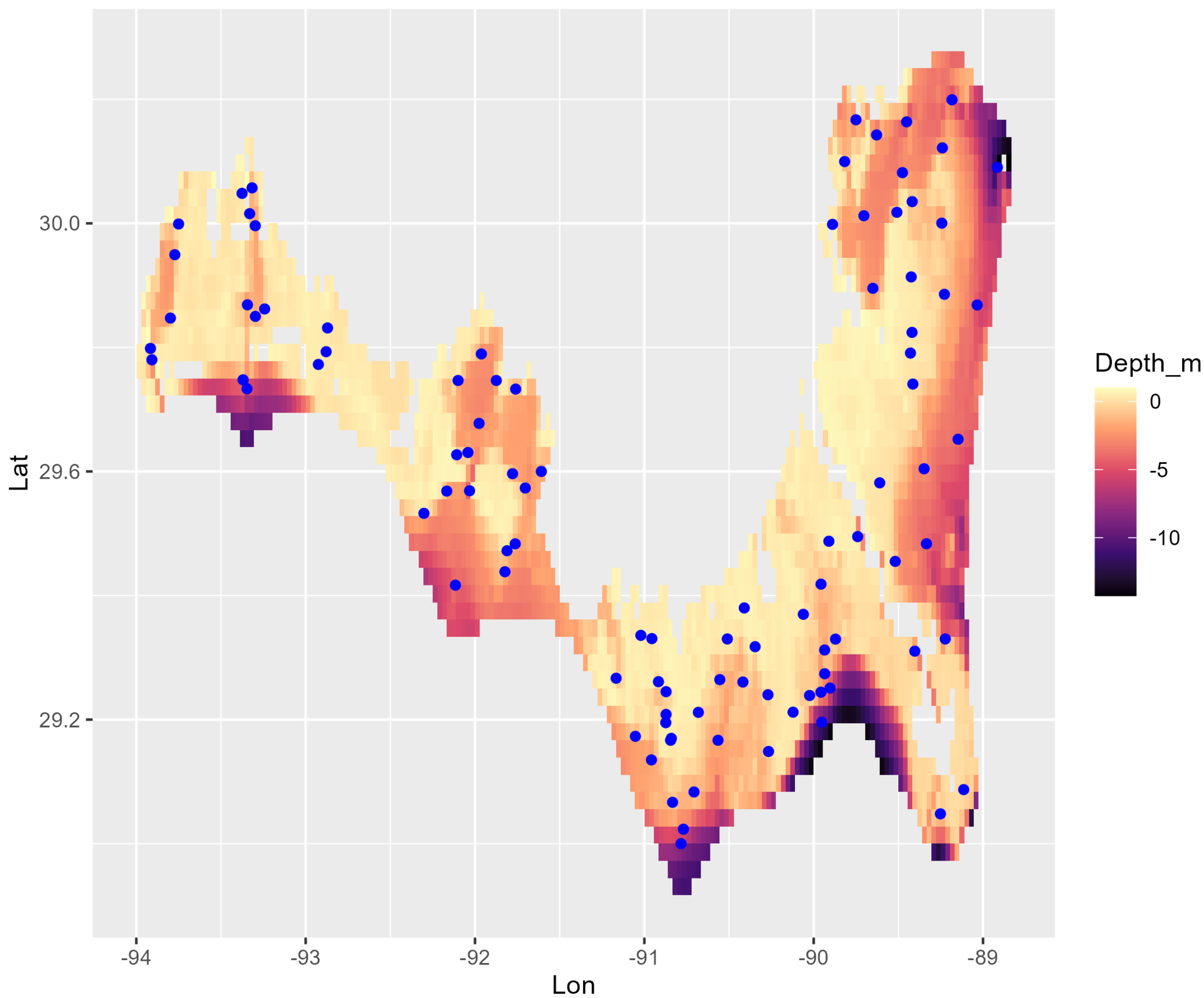

Figure 3: WSH index extrapolation region. Survey stations are shown in blue.

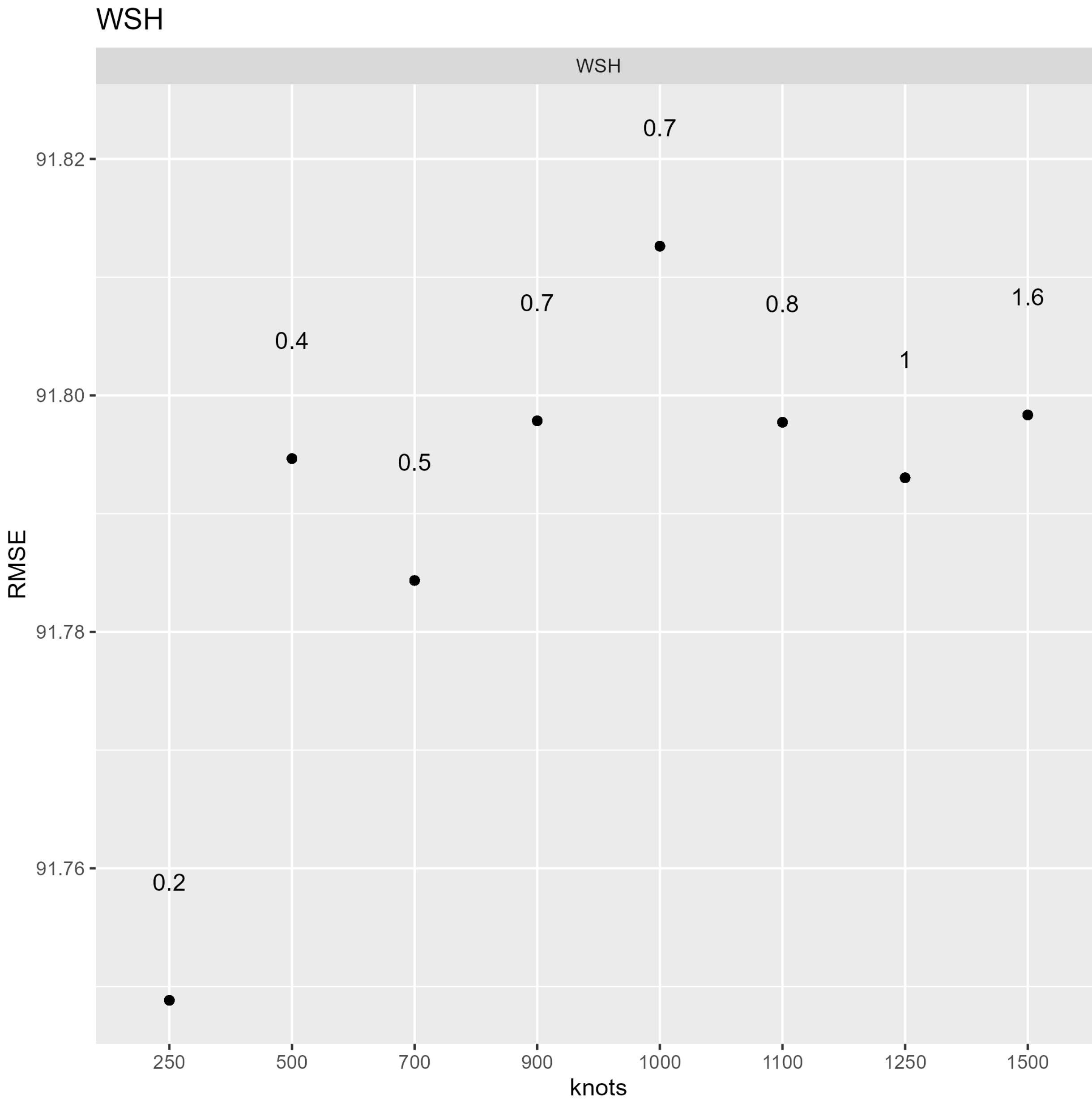

Figure 4: Index RMSE associated with each model run for the different numbers of knots attempted. Run time in hours is printed above each data point. Increasing knot size leads to an asymptote in the RMSE using starting at around 500 knots.

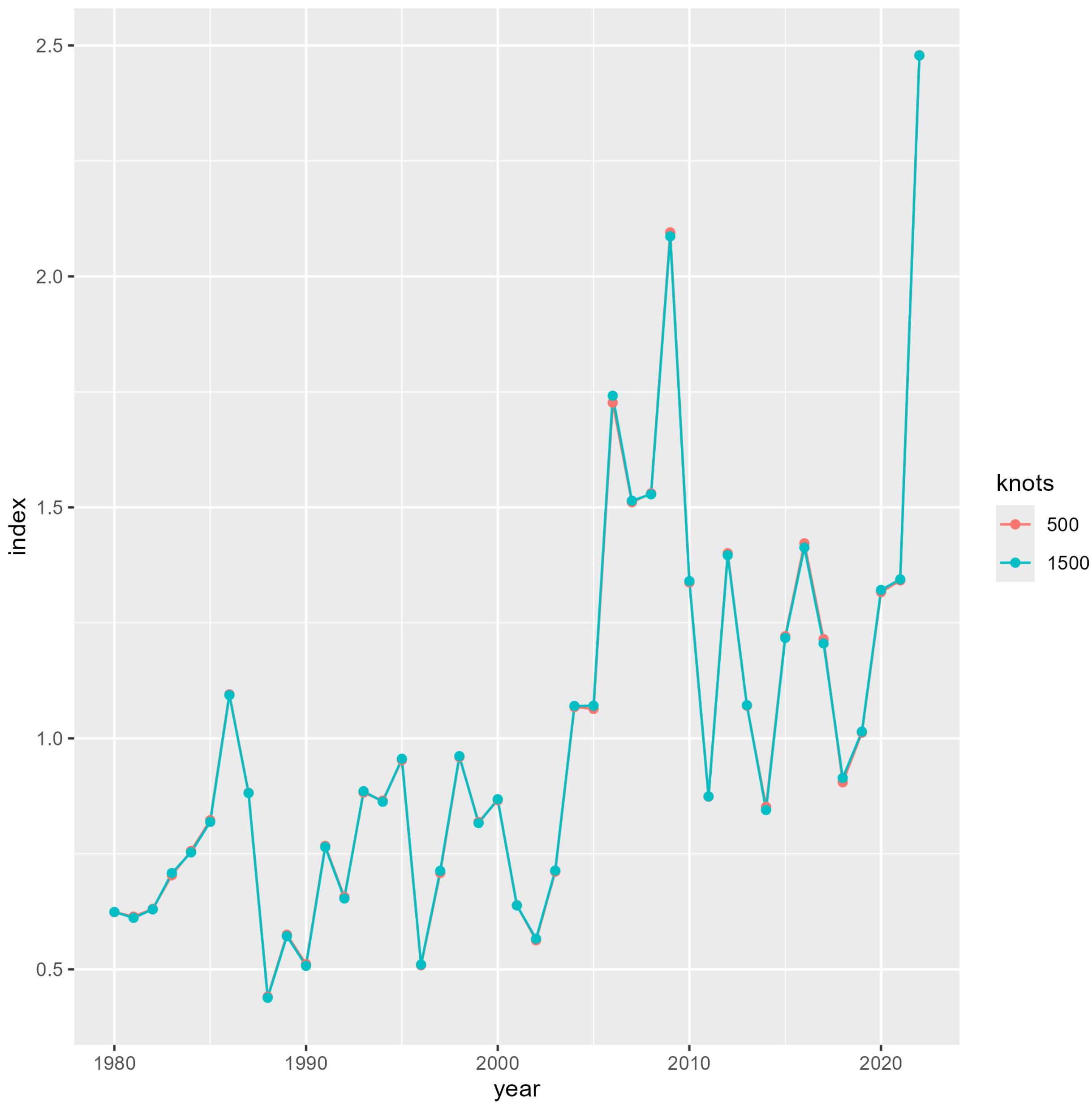

Figure 5: Index estimates for the model run with the highest number of knots (1500) compared with those for the selected model (500).

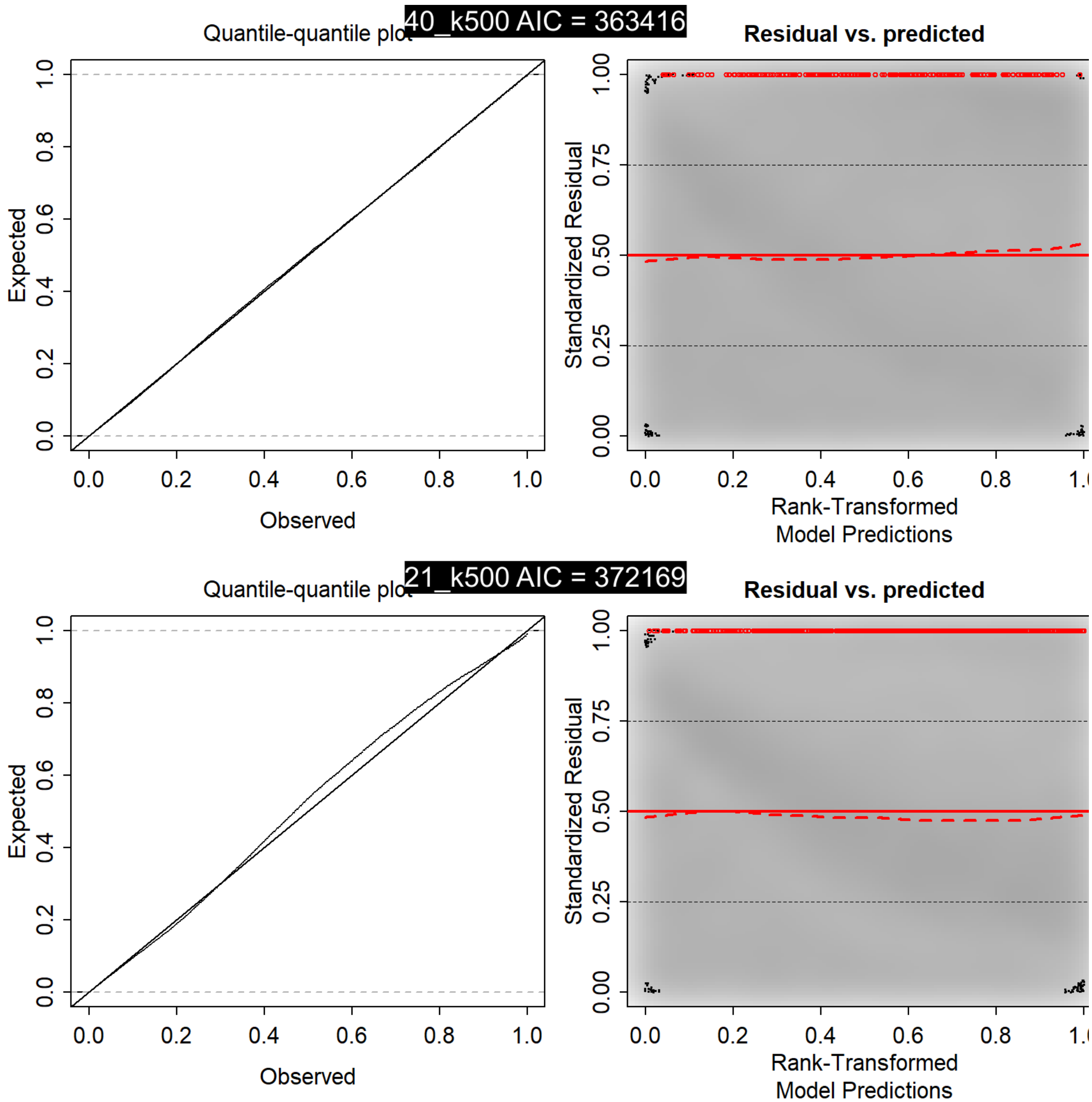

Figure 6: Dharma residual diagnostic plots for the delta-lognormal (top; AIC=363,416) and poisson-link (bottom; AIC=372,169) model.

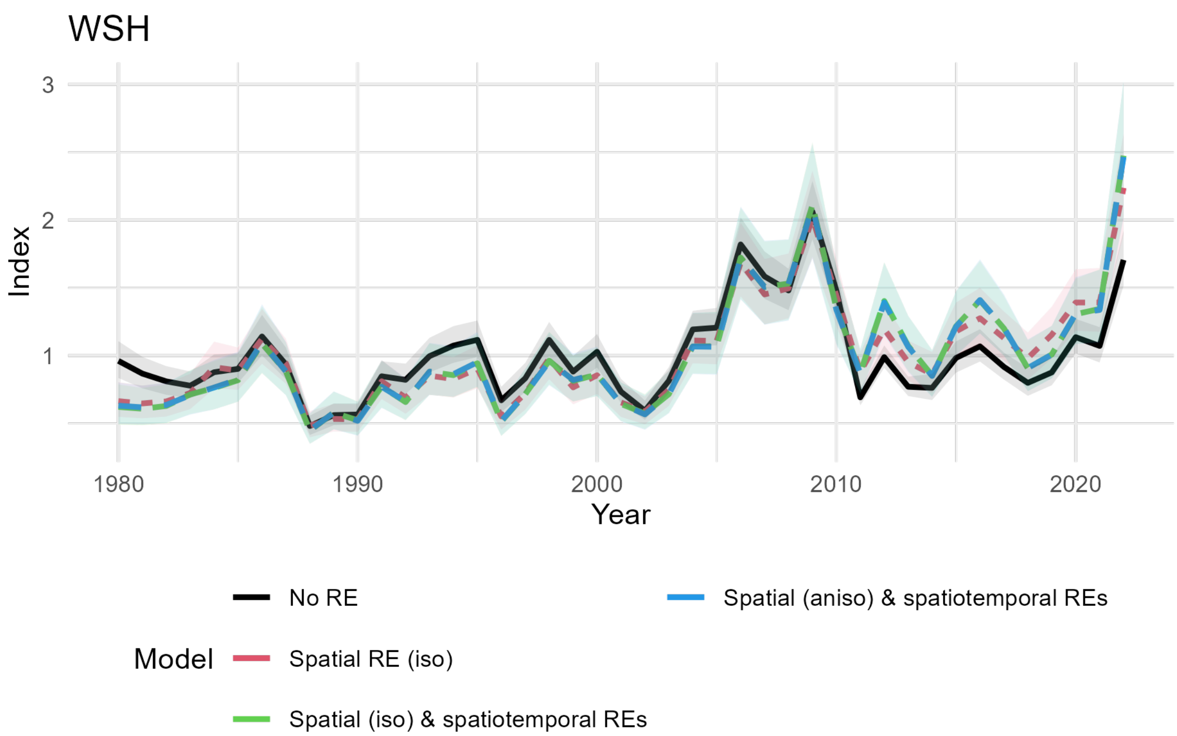

Figure 7: Comparing index estimates and associated confidence intervals across runs with different spatial and spatiotemporal random effects (RE) specifications. iso: isotropic Matérn covariance function; aniso: anisotropic Matérn covariance function.

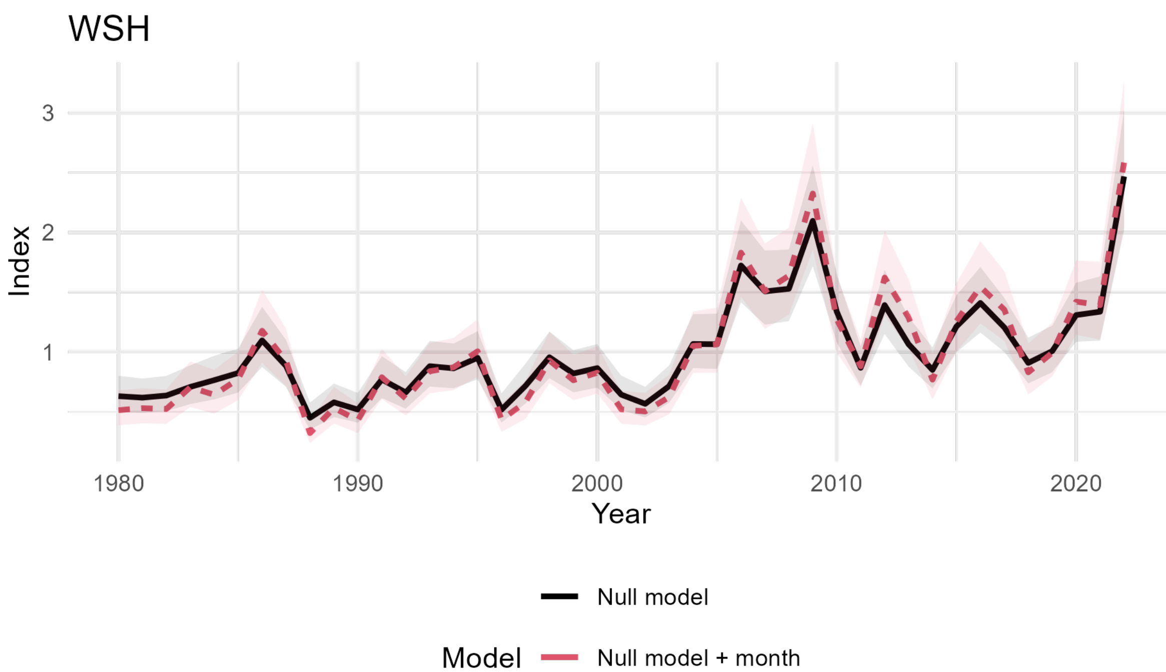

Figure 8: Comparing index estimates and associated confidence intervals across runs with different catchability covariates. The Null model is a model with both spatial and spatiotemporal random effects included in both the 1st and 2nd linear predictors and assuming geometric anisotropy but with no catchability covariates included.

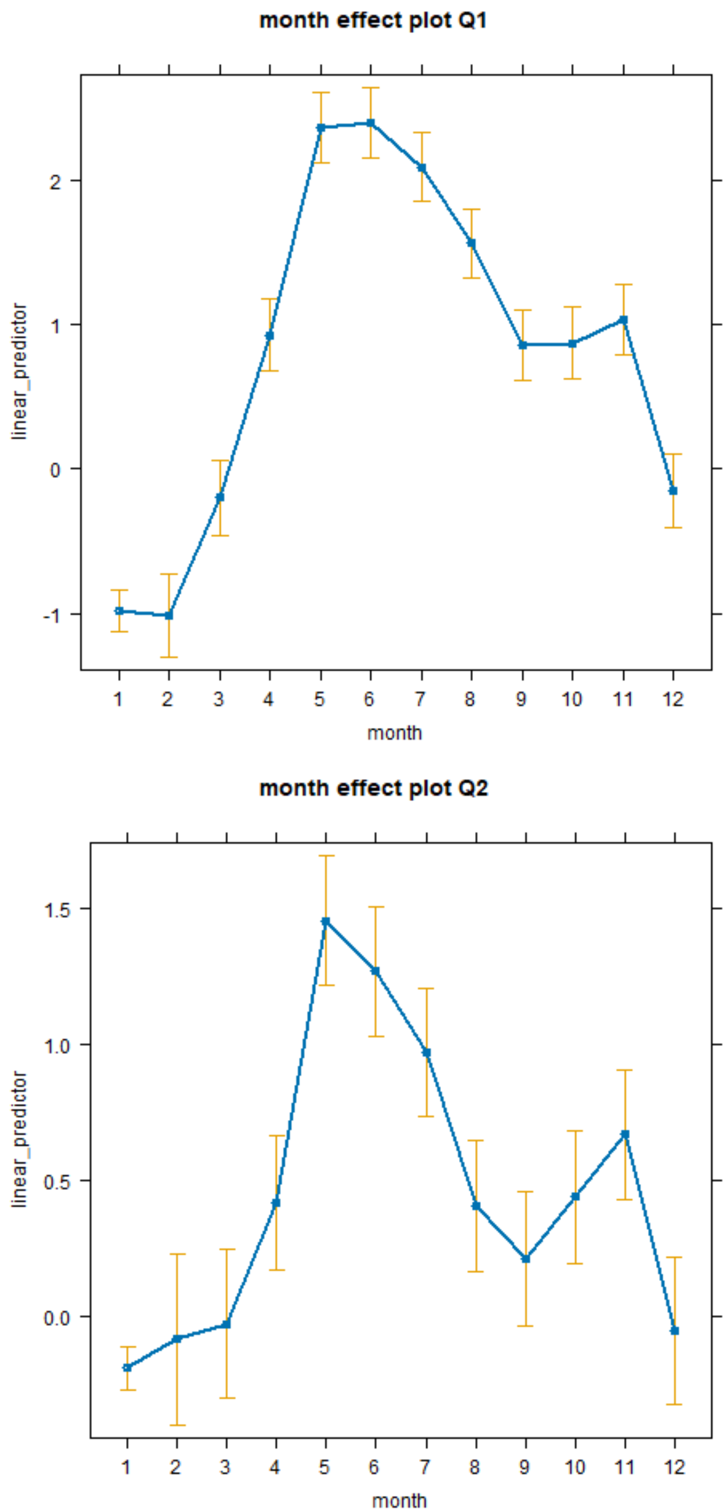

Figure 9: Effect plots for the month catchability covariate. Top panel shows the impact on the first linear predictor, bottom panel shows the impact on the second linear predictor.

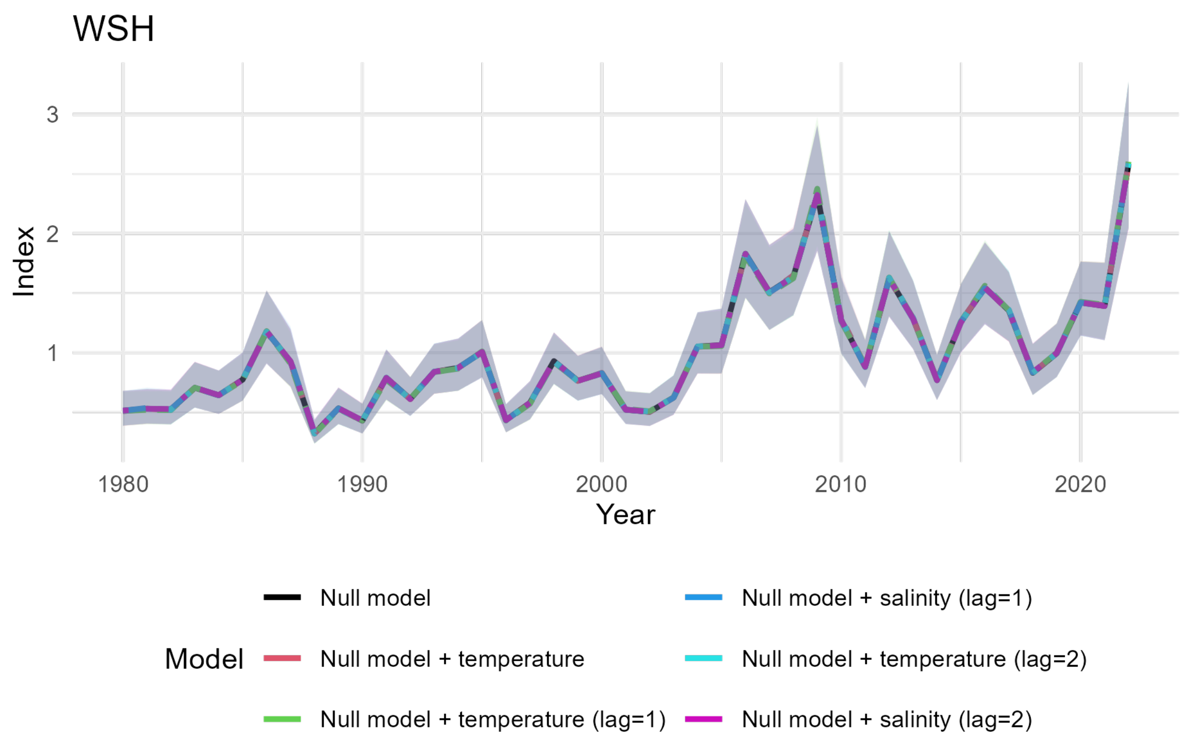

Figure 10: Comparing index estimates and associated confidence intervals across runs with different habitat covariates included. The Null model includes both spatial and spatiotemporal random effects (assuming geometric anisotropy) and a month catchability covariate. temperature = mean annual temperature on the nursery grounds, salinity = mean annual salinity on the nursery grounds, environmental indices are lagged by 0, 1 or 2 years compared with the cpue observations.

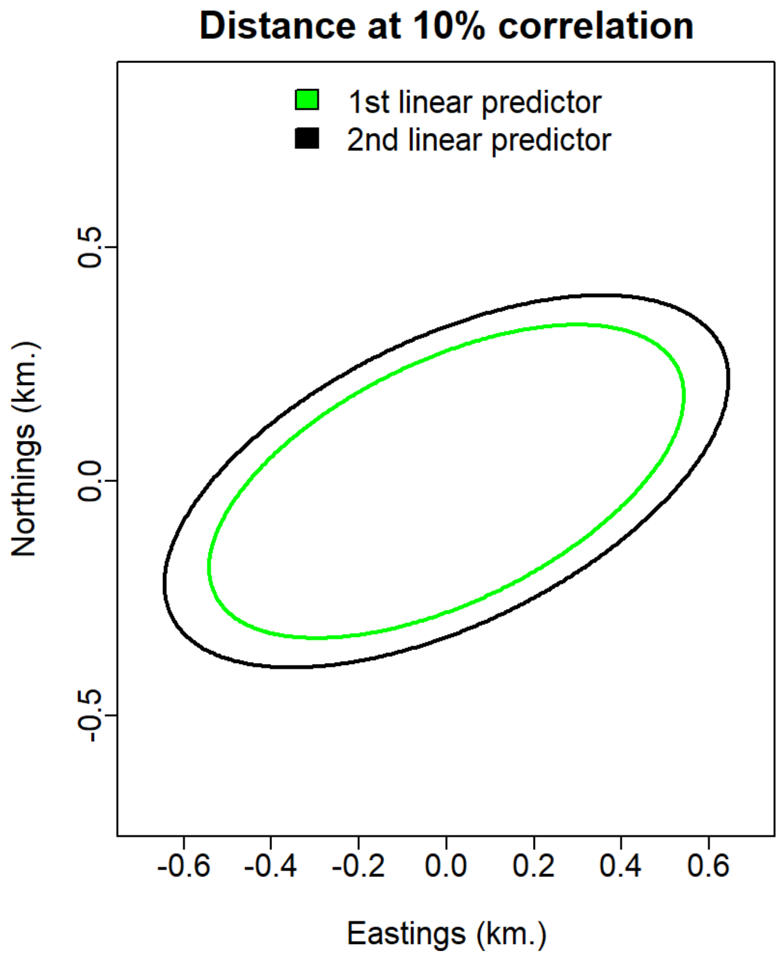

Figure 11: Distance needed to achieve a correlation of approximately 10% from a location centered at coordinates (0,0). Correlations decline slower along the southwest-northeast axis than along the northwest-southeast axis.

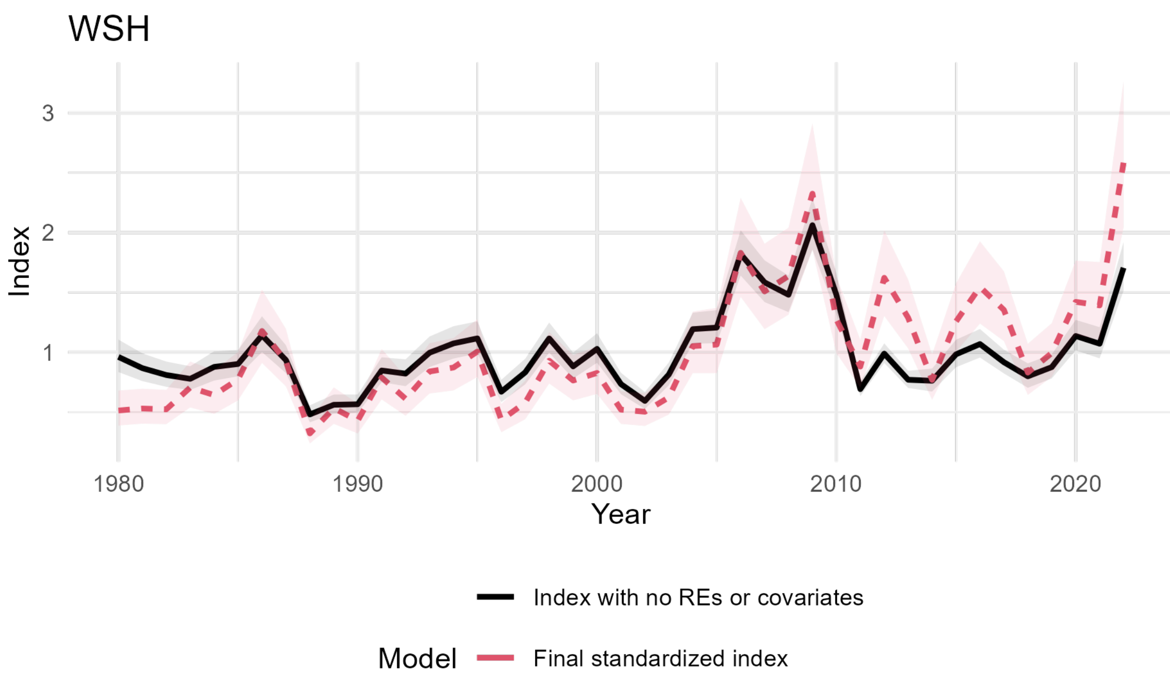

Figure 12: Comparing index estimates and associated confidence intervals for the final model (red) with the preliminary unstandardized index (black). The final index includes both spatial and spatiotemporal random effects assuming geometric anisotropy and a month catchability covariate.

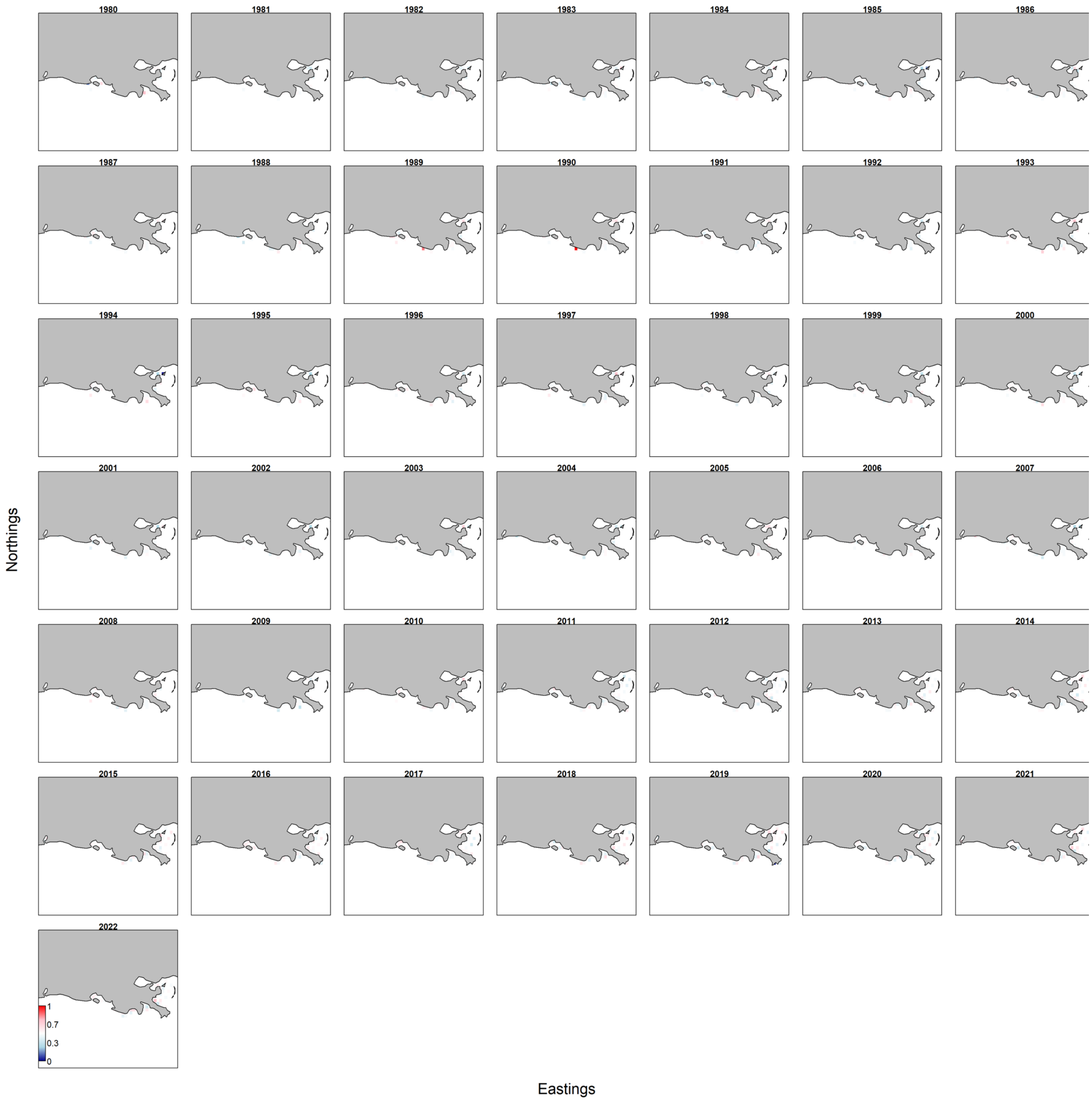

Figure 13: Spatial distribution of quantile residuals for the final model, red color indicating overestimation and blue indicating underestimation.

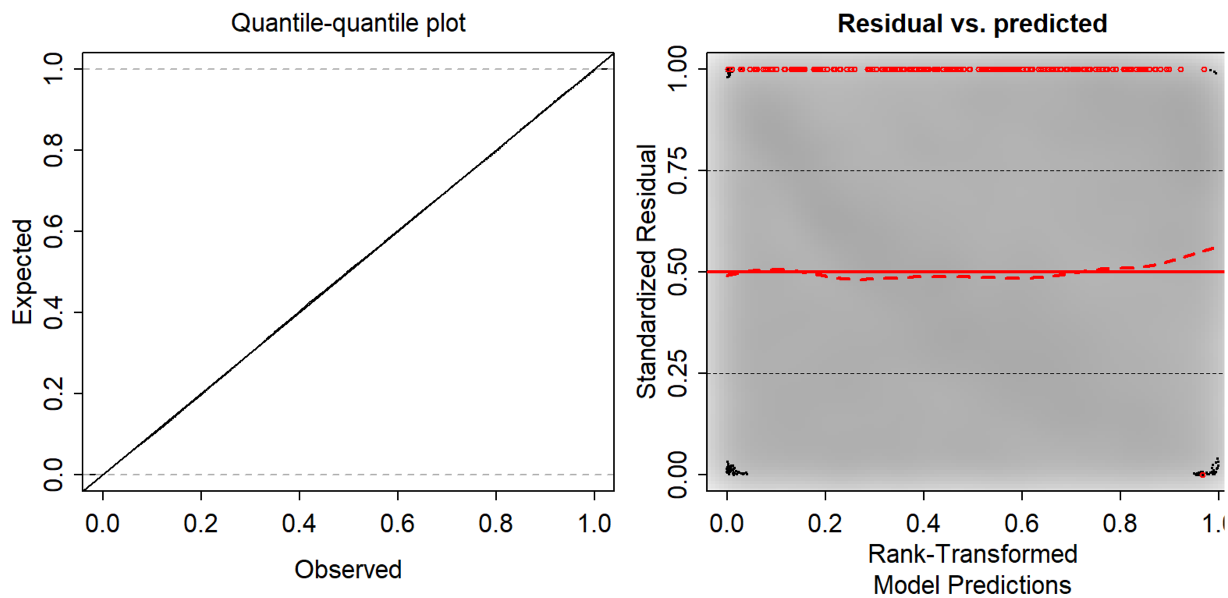

Figure 14: Dharma residual diagnostic plots for the final model.

Figure 15: Predicted densities in each year for the final model. The absolute scale is not directly interpretable since no meaningful area swept estimates were available for the development of the index.