Appendix A: sdmTMB Model Selection to Estimate PIBKC mature male biomass from the NMFS EBS survey

William T. Stockhausen

2025-08-28

0.1 `sdmTMB` Models for PIBKC mature male biomass

The NMFS EBS bottom trawl survey captured no mature male PIBKC in 2023 and 2024, introducing observed zeros into the design-based survey time series for Pribilof Islands blue king crab mature male biomass (PIBKC MMB). The `rema` R package (R Core Team 2022; Sullivan 2022) has been used in previous assessments (e.g., Stockhausen (2023)) to fit a random effects first-order autoregressive (AR1) time series model to this time series to provide model-based estimates of the time series and terminal year survey MMB, but because `rema` fits the time series progression on the log-scale, the observed zeros are problematic because they become negative infinities when transformed to that scale. Several *ad hoc* approaches to dealing with the observed zeros while retaining the `rema` framework were presented to the CPT and SSC at their May and June (respectively) 2025 meetings (Stockhausen (2025)), but all were rejected in favor of deriving the model-based survey MMB time series based on spatiotemporal species distribution models (SDMs) using the `sdmTMB` `R` package (R Core Team 2022; Anderson et al. 2024).

`sdmTMB` is an `R` package that fits spatial and spatiotemporal Generalized Linear Mixed Effects Models (GLMMs) using Template Model Builder [TMB; Kristensen et al. (2016)], R-INLA (Lindgren and Rue 2015), and Gaussian Markov random fields. One common application of this framework is to create spatially-explicit species distribution models (SDMs) from species abundance data collected across space and (possibly) through time, although its principal advantage here is that it can also estimate stock-level time series by integrating estimated spatially-explicit patterns of abundance or biomass across space. A simple way to think about an `sdmTMB` model is that it provides a way to interpolate observations across space and time, taking into account both stochastic population variability across space and time as well as variability inherent to the sampling process itself. While several preliminary `sdmTMB` model configurations involving different statistical assumptions were presented to the CPT in May, it subsequently recommended (which the SSC supported) proceeding with models based on a Tweedie distribution, with spatiotemporal random effects having an AR1 error structure. The Tweedie distribution (Tweedie 1984) is a compound Poisson-gamma distribution that allows observed absences (zeros) in the data, incorporating a single linear predictor (rather than two in the case of two-stage models like the delta-gamma) while estimating a parameter that reflects the degree of mixing between the underlying Poisson and gamma distributions. (Thorson et al. 2021) recommend using the Tweedie distribution for spatiotemporal models

by default, particularly when it is not possible to compare the scale of model-derived indices with those from design-based estimators (although comparison is possible in this case).

0.2 Methods

Here, 18 different `sdmTMB` model configurations were examined (Table 1). Based on the CPT recommendation, only models with ln-scale spatiotemporal random effects modeled as Tweedie AR1 distributions were considered. Other "axes" of model construction included whether strictly spatial random effects were estimated, whether strictly temporal random effects were estimated, whether a smoothly-varying function of haul depth was estimated, the grid used to predict survey catch per unit effort (CPUE in [biomass]/[unit area]) across the management area for PIBKC, the spatial "mesh" (e.g., grid) used to estimate the model spatiotemporal and spatial random effects, and whether barriers (e.g., the Pribilof Islands) were included the estimation mesh.

0.2.1 Statistical models

All of the models examined here can be expressed as

$$\begin{aligned}\mathbb{E}[o_{s,t}] &= \mu_{s,t},\\ \mu_{s,t} &= f^{-1}\left(\alpha \cdot s[ln(d_{s,t})] + \beta \cdot \gamma_t + \delta \cdot \omega_s + \epsilon_{s,t}\right)\end{aligned}$$

where $o_{s,t}$ represents the observed survey CPUE at position s in year t; $\mathbb{E}[\cdot]$ is the expectation operator; $\mu_{s,t}$ is the mean at s, t; f represents the link function (the natural log in all cases here); α, β, and δ are 1 or 0 to turn on/off, respectively, a smooth function of depth ($s[ln(d_{s,t})]$), temporally-varying random intercepts (γ_t), or spatially-varying random effects (ω_s). $\epsilon_{s,t}$ represents the spatiotemporal random effects terms included in all the models.

Depth is included as a covariate in models with the "covariate" column value listed as "ln(depth)", with the effect estimated as a smooth, main effects function of $ln(d_s)$ using the "s" function from the `mgcv` R package. Spatial random fields are included in models with the "spatial" column value listed as "on". The random field, ω_s, is a Gaussian Markov random field with

$$\omega_s \sim \text{MVNormal}(0, \sigma_\omega^2 Q_\omega^{-1})$$

where Q_ω is a sparse precision matrix and σ_ω^2 is the marginal variance. Temporal random effects (random intercepts) are included in models with the "temporal" column value listed as "AR1". The random temporal intercepts, γ_t, follow a first-order autoregressive process, with

$$\begin{aligned}\gamma_{t=1} &\sim \text{Normal}\left(0, \sigma_\gamma^2\right),\\ \gamma_{t>1} &\sim \left(\rho_\gamma \gamma_{t-1}, \sqrt{1-\rho_\gamma^2}\sigma_\gamma^2\right)\end{aligned}$$

where the first time step is given a mean zero prior and ρ_γ is the correlation between subsequent time steps. The spatiotemporal random effects included in all the models were estimated using an AR1 temporal covariance structure such that

$$\begin{aligned}\delta_{t=1} &\sim \text{MVNormal}(0, \Sigma_\epsilon),\\ \delta_{t>1} &= \rho\delta_{t-1} + \sqrt{1-\rho^2}\epsilon_t,\ \epsilon_t \sim \text{MVNormal}\left(0, \Sigma_\epsilon\right)\end{aligned}$$

where Σ_ϵ is a spatial covariance matrix, ρ is the correlation coefficient, and the ϵ_t are temporally iid spatial random effects with the same spatial covariance structure as $\delta_{t=1}$.

0.2.2 Evaluation meshes

Six triangular stochastic partial differential equation (SPDE) meshes, based on two mesh construction strategies ("kmeans" and "cutoff") and three sizes per strategy, were originally considered for model evaluation (Figure 1). However, preliminary tests with the "cutoff" meshes (labelled "CT" in the figure) proved unsatisfactory, with sdmTMB models either failing to converge or taking a prohibitive amount of time. The majority of models evaluated here used the "kmeans" mesh with 80 knots, although the "best" 80-knot model was also evaluated at 40 and 60 knots to partially explore the influence of mesh size on results.

0.2.3 Prediction grids

The evaluation mesh is used to fit a `sdmTMB` model to CPUE data, with spatiotemporal random effects estimated at each mesh node for each time step. However, the evaluation mesh is generally at a coarser scale than that of the data to improve model stability and convergence times. Following model estimation, a "prediction grid" can be used to predict spatiotemporal patterns of species abundance/biomass at much finer spatial resolutions than the evaluation grid. Subsequently, area-integrated indices of species abundance or biomass can be obtained by integrating the predicted densities across the area of interest. Here, three prediction grids were evaluated as the basis for area-integrated indices of PIBKC mature male survey biomass (MMSB) to determine an estimate for terminal year MMSB (Figure 2). The spatial resolution was the same (5 km) for all three grids, but the coverage relative to the NMFS survey grid and the continental shelf differed: grid 0 covered the Pribilof District as defined by NMFS survey strata, extending across the Pribilof Islands and beyond the shelf edge in some areas; grid 1 extended to the shelf edge (500 m depth) and excluded the Pribilof Islands; and grid 2 was limited to depths < 500 m within the NMFS survey strata and excluded the Pribilof Islands. All three grids extend slightly beyond the kmeans 40 evaluation mesh, while prediction grid 0 extends slightly beyond the kmeans 60 mesh (Figure 2), which proved problematic when estimating the MMSB indices using these combinations of evaluation mesh and prediction grid.

0.2.4 Model evaluation

All models were estimated using `sdmTMB`'s 'sdmTMB' function by fitting the 1975-2024 NMFS EBS haul-level MMSB CPUE dataset for PIBKC using restricted maximum likelihood and adding 2020, the year the EBS survey was not conducted due to Covid-19, as an "extra" year. Models were first evaluated using `sdmTMB`'s 'sanity' function, which checks for issues with model convergence, hessian inversion, and the estimated size of the variances related to the various random effects components.

Models which passed `sdmTMB`'s sanity check were then subjected to a k-fold out-of-sample cross-validation procedure and subsequently ranked using the average out-of-sample predictive skill, with "k" chosen as 10. For each model, the cross-validation procedure randomly split the haul-level CPUE data into 10 "folds" of similar size, with approximately 10% of the data selected as the out-of-sample evaluation set and the remaining 90% of the data used as the training set. The evaluation data were selected across the 10 folds without replacement, so each datum featured in an evaluation set once. For each fold, the model was fit to the training data and used to predict the out-of-sample evaluation data. The out-of-sample log-likelihood score, root mean square error

(RMSE), and mean absolute error (MAE) were then calculated for each fold for a given model. The models were then ranked by the mean (negative) log-likelihood score (smaller being better) and the five "best" models were selected for further examination.

For the five "best" models, residuals were checked without regard to spatial pattern for uniformity, outliers, dispersion, and quantile behavior using the `DHARMa` `R` package's (Hartig 2022). They were then checked for spatial correlation by examining annual maps of the residuals for patterns and using the Moran's I test function ("moran.mc") in the `spdep` `R` package (Pebesma and Bivand 2023) to test for spatial autocorrelation in the residuals each year. Finally, aggregated biomass indices were calculated for each of these models using the three prediction grids and `sdmTMB`'s "get_index" function and the resulting time series compared.

0.3 Results

0.3.1 Model checks

Three models either failed to finish or failed the `sdmTMB`'s sanity checks Table 2. The remaining models passed all of the checks.

0.3.2 Cross-validation

Based on the 10-fold cross-validation, the best model in terms of out-of-sample predictive performance was `ar1S_nullT_logDepth_km080b`, based on all three cross-validation statistics (Table 3, Figures 3-5). Model `ar1s_nullT_logDepth_km040b` failed to converge during the cross-validation tests.

0.3.3 DHARMa residuals

All of the remaining models were tested for deviations from uniformity, significant dispersion, presence of outliers, and quantile bias and nonlinearity using the `DHARMa` R package. All of the models passed the outliers test at the $p > 0.05$ level of significance, but none passed the other tests at the same level (Table 4). Standard `DHARMa` residual plots are illustrated in Figures 6-10 for the top five models ranked on the average cross-validation score (lowest score is best).

0.3.4 Spatial residuals and statistics

Spatial plots of residuals scaled to range from 0 to 1 are illustrated in 5-year time blocks for the five "best" models ranked on the average cross-validation score (lowest score is best) are illustrated in Figures 11-20. The distributions associated with the null hypothesis of Moran's I test for spatial autocorrelation were very similar across models and years while the values of the statistics were different (Figure 21). The distributions were determined using 1000 randomly-drawn permutations of the residuals for each model for each year; the statistic for each model and year was determined from the "observed" spatial pattern of residuals. Two of the models failed the Moran's I test for spatial correlation in a couple of years (Table 5), but this was not unexpected given the number of years for which the statistic was calculated.

0.3.5 Fixed effects (smoothed covariates)

Three of the "best" models contained a smooth function of log-scale haul depth as a fixed effects covariate (Figures 22-24). The estimated smooth function has an amplifying effect at depths shallower than ~100 m and a diminishing effect at greater depths, with the uncertainty increasing dramatically at depths greater than ~150 m due to a lack of observations at greater depths.

0.3.6 Biomass indices

The biomass indices calculate using the evaluation grid and the three prediction grids are illustrated for the five "best" models in Figures 25-29. The results are generally similar across models and grids, with the notable exceptions that 1) the estimates for 1979 differs substantially between the evaluation grid and the three prediction grids for all five models, and 2) the results using prediction grid 1 for models (except the "best" model) that include log depth as a covariate. Using the evaluation grid to generate the biomass index results in estimates almost equivalent to the design-based values (not shown) and the spatial pattern of haul locations in 1979 was unique among survey years, so differences are not surprising. As noted previously, prediction grid 1 extended to the shelf edge (500 m depth) and excluded the Pribilof Islands. The inclusion of depths up to 500 m led to obviously poor estimates for the biomass indices when using this grid with models "r1S_nullT_logDepth_km060b" and "r1S_nullT_logDepth_km080", but not with the "best" model r1S_nullT_logDepth_km080b.

0.3.7 Model and biomass index selection

Based on the analysis presented, Model "ar1S_nullT_logDepth_km080b" appears to be the "best" of the 18 models considered. Although the indices based on prediction grid 1 were poor for two of the models that included a smooth function of log depth as a fixed effect, this was not the case for "ar1S_nullT_logDepth_km080b". Because prediction grid 1 best represents the shelf area within the Pribilof District, the mature male survey biomass index based on Model "ar1S_nullT_logDepth_km080b" and prediction grid 1 is considered the "best" index to use to calculate MMB-at-mating in the Tier 4 status determination for PIBKC.

References

Anderson, S.C., Ward, E.J., English, P.A., Barnett, L.A.K., and 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. bioRxiv **2022.03.24.485545**. doi:10.1101/2022.03.24.485545.

Hartig, F. 2022. DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models. Available from http://florianhartig.github.io/DHARMa/.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software **70**(5): 1–21. doi:10.18637/jss.v070.i05.

Lindgren, F., and Rue, H. 2015. Bayesian spatial modelling with R-INLA. Journal of Statistical Software **63**(19): 1–25. Available from http://www.jstatsoft.org/v63/i19/.

Pebesma, E., and Bivand, R.S. 2023. Spatial data science with applications in R. Chapman & Hall. Available from https://r-spatial.org/book/.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Stockhausen, W.T. 2023. 2023 Stock Assessment and Fishery Evaluation Report for the Pribilof Islands blue king crab fisheries of the Bering Sea and Aleutian Islands regions. *In* Stock Assessment and Fishery Evaluation Report for the KING AND TANNER CRAB FISHERIES of the Bering Sea and Aleutian Islands regions 2023 Final Crab SAFE. North Pacific Fishery Management Council, Anchorage, AK. p. 112. Available from https://meetings.npfmc.org/CommentReview/DownloadFile?p=05baad65-e92f-4998-8655-a076b17b7af3.pdf&fileName=Priblof%20Island%20Blue%20King%20Crab%20SAFE.pdf.

Stockhausen, W.T. 2025. 2025 PIBKC assessment considerations. North Pacific Fishery Management Council, Juneau, AK. Available from https://meetings.npfmc.org/CommentReview/DownloadFile?p=753c6ba1-2567-468a-b6eb-faa77ec3e4a3.pdf&fileName=PIBKC2025-05.pdf.

Sullivan, J. 2022. rema: A generalized framework to fit the random effects (RE) model, a state-space random walk model developed at the Alaska Fisheries Science Center (AFSC) for apportionment and biomass estimation of groundfish and crab stocks. Available from https://github.com/JaneSullivan-NOAA/rema, https://afsc-assessments.github.io/rema/.

Thorson, J.T., Cunningham, C.J., Jorgensen, E., Havron, A., Hulson, P.-J.F., Monnahan, C.C., and 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:https://doi.org/10.1016/j.fishres.2020.105745.

Tweedie, M.C.K. 1984. An index which distinguishes between some important exponential families. *In* Statistics: Applications and new directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. *Edited by* J.K. Ghosh and J. Roy. Indian Statistical Institute, Calcutta. pp. 579--604.

Tables

List of Tables

Table 1. Brief description of the models examined here. All models fit to mature male survey CPUE in t/sq. km., assumed a Tweedie error distribution with log link, and estimated spatiotemporal random effects with a temporal AR1 structure.

name	spatial	temporal	covariate	evaluation mesh	barriers
onS_ar1T_noCovar_km080	on	AR1	none	kmeans 80	no
onS_offT_noCovar_km080	on	off	none	kmeans 80	no
offS_ar1T_noCovar_km080	off	AR1	none	kmeans 80	no
offS_offT_noCovar_km080	off	off	none	kmeans 80	no
onS_ar1T_logDepth_km080	on	AR1	ln(depth)	kmeans 80	no
onS_offT_logDepth_km080	on	off	ln(depth)	kmeans 80	no
offS_ar1T_logDepth_km080	off	AR1	ln(depth)	kmeans 80	no
offS_offT_logDepth_km080	off	off	ln(depth)	kmeans 80	no
onS_ar1T_noCovar_km080b	on	AR1	none	kmeans 80	yes
onS_offT_noCovar_km080b	on	off	none	kmeans 80	yes
offS_ar1T_noCovar_km080b	off	AR1	none	kmeans 80	yes
offS_offT_noCovar_km080b	off	off	none	kmeans 80	yes
onS_ar1T_logDepth_km080b	on	AR1	ln(depth)	kmeans 80	yes
onS_offT_logDepth_km080b	on	off	ln(depth)	kmeans 80	yes
offS_ar1T_logDepth_km080b	off	AR1	ln(depth)	kmeans 80	yes
offS_offT_logDepth_km080b	off	off	ln(depth)	kmeans 80	yes
onS_offT_logDepth_km040b	on	off	ln(depth)	kmeans 40	yes
onS_offT_logDepth_km060b	on	off	ln(depth)	kmeans 60	yes

Table 2. Summary table for model checks using sdmTMB's 'sanity'function. For models that finished running

name	finished	hessian	eigen values	minimizer	range	gradients	se magnitude	NAs	sigmas	all
onS_ar1T_noCovar_km080	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
onS_offT_noCovar_km080	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
offS_ar1T_noCovar_km080	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
offS_offT_noCovar_km080	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
onS_ar1T_logDepth_km080	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
onS_offT_logDepth_km080	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
offS_ar1T_logDepth_km080	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
offS_offT_logDepth_km080	false	–	–	–	–	–	–	–	–	–
onS_ar1T_noCovar_km080b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
onS_offT_noCovar_km080b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
offS_ar1T_noCovar_km080b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
offS_offT_noCovar_km080b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
onS_ar1T_logDepth_km080b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
onS_offT_logDepth_km080b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
offS_ar1T_logDepth_km080b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
offS_offT_logDepth_km080b	true	not ok	ok	ok	ok	ok	not ok	some NAs	not ok	not ok
onS_offT_logDepth_km040b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok
onS_offT_logDepth_km060b	true	ok	ok	ok	ok	ok	ok	no NAs	ok	ok

Table 3. Summary statistics from the 10-fold cross-validation tests, randked by the negative out-of-sample predictive log-likelihood score. Model ar1s_nullT_logDepth_km040b failed to converge during the cross-validation tests and is not listed in the table.

model	-mean(log-likelihood)	mean(RMSE)	mean(MAE)
ar1S_nullT_logDepth_km080b	95.76340	0.2756831	0.0451916
offS_nullT_noCovar_km080b	99.95572	0.3323269	0.0483967
ar1S_nullT_logDepth_km060b	100.11473	0.3822746	0.0521069
offS_ar1T_noCovar_km080	108.38197	0.3215885	0.0475334
ar1S_nullT_logDepth_km080	110.54570	0.3506269	0.0496702
ar1S_nullT_noCovar_km080b	110.88060	0.3387133	0.0466871
ar1S_ar1T_noCovar_km080b	112.51754	0.3095588	0.0455972
offS_nullT_noCovar_km080	113.31247	0.3446132	0.0491336
offS_ar1T_noCovar_km080b	114.30084	0.4230001	0.0539670
offS_ar1T_logDepth_km080b	119.00278	0.3546071	0.0479412
ar1S_nullT_noCovar_km080	119.50918	0.3557230	0.0505758
offS_ar1T_logDepth_km080	126.04203	0.4012171	0.0507293
ar1S_ar1T_logDepth_km080b	129.57397	0.3449482	0.0483890
ar1S_ar1T_logDepth_km080	129.73837	0.3411131	0.0453765

Table 4. DHARMa statistical checks on the `sdmTMB` models. Checks with p-values < 0.05 were regarded as failing the test.

model	uniformity	dispersion	outliers	quantiles
onS_ar1T_noCovar_km080	0.003	0.000	0.801	0.001
onS_offT_noCovar_km080	0.004	0.000	0.380	0.001
offS_ar1T_noCovar_km080	0.010	0.000	1.000	0.001
offS_offT_noCovar_km080	0.023	0.000	0.449	0.003
onS_ar1T_logDepth_km080	0.026	0.000	0.380	0.002
onS_offT_logDepth_km080	0.006	0.000	0.531	0.000
offS_ar1T_logDepth_km080	0.007	0.000	1.000	0.004
onS_ar1T_noCovar_km080b	0.008	0.000	0.900	0.001
onS_offT_noCovar_km080b	0.004	0.000	0.258	0.001
offS_ar1T_noCovar_km080b	0.010	0.000	1.000	0.001
offS_offT_noCovar_km080b	0.038	0.000	0.900	0.009
onS_ar1T_logDepth_km080b	0.019	0.000	0.801	0.001
onS_offT_logDepth_km080b	0.006	0.000	0.900	0.000
offS_ar1T_logDepth_km080b	0.006	0.000	0.707	0.004
onS_offT_logDepth_km040b	0.247	0.092	0.614	0.048
onS_offT_logDepth_km060b	0.060	0.000	0.380	0.000

Table 5. Models which failed the permutation-based Moran's I in at least one year ($p<0.05$), with the year, statistic, and p-value determined by permutation.

name	year	statistic	p_value
tw_ar1ST_offS_ar1T_noCovar_km080	1979	0.115	0.024
tw_ar1ST_offS_ar1T_noCovar_km080	1992	0.087	0.017
tw_ar1ST_offS_ar1T_noCovar_km080	2005	0.099	0.012
tw_ar1ST_ar1S_nullT_logDepth_km080	2024	0.071	0.041
tw_ar1ST_ar1S_nullT_logDepth_km080b	2024	0.074	0.040

Figures

List of Figures

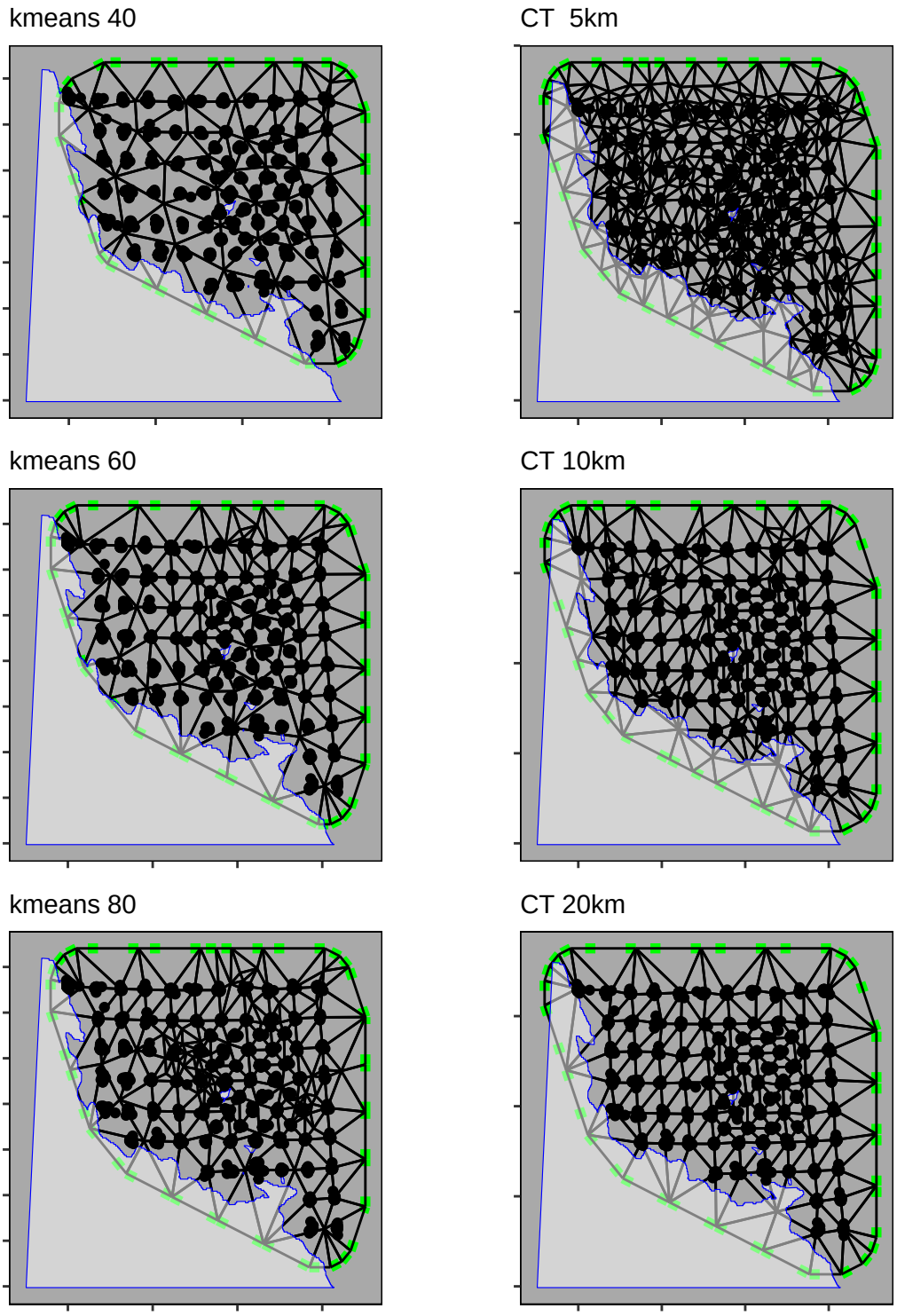

Figure 1. The triangular SPDE evaluation meshes considered while developing the sdmTMB models. For each, the sdmTMB function "make_mesh" was applied to the NMFS survey haul-level CPUE data for MMB (black dots). "kmeans" (left column) or "CT" (right-hand column) indicate the method used; the value indicates the number of allowd knots (kmeans) or the minimum distance allowed between nodes (CT). The green dots indicate the boundary vertices, the segments indicate triangle edges. The light gray polygons (with blue outlines) were used to create interpolation barriers representing the shelf edge and the Pribilof Islands.

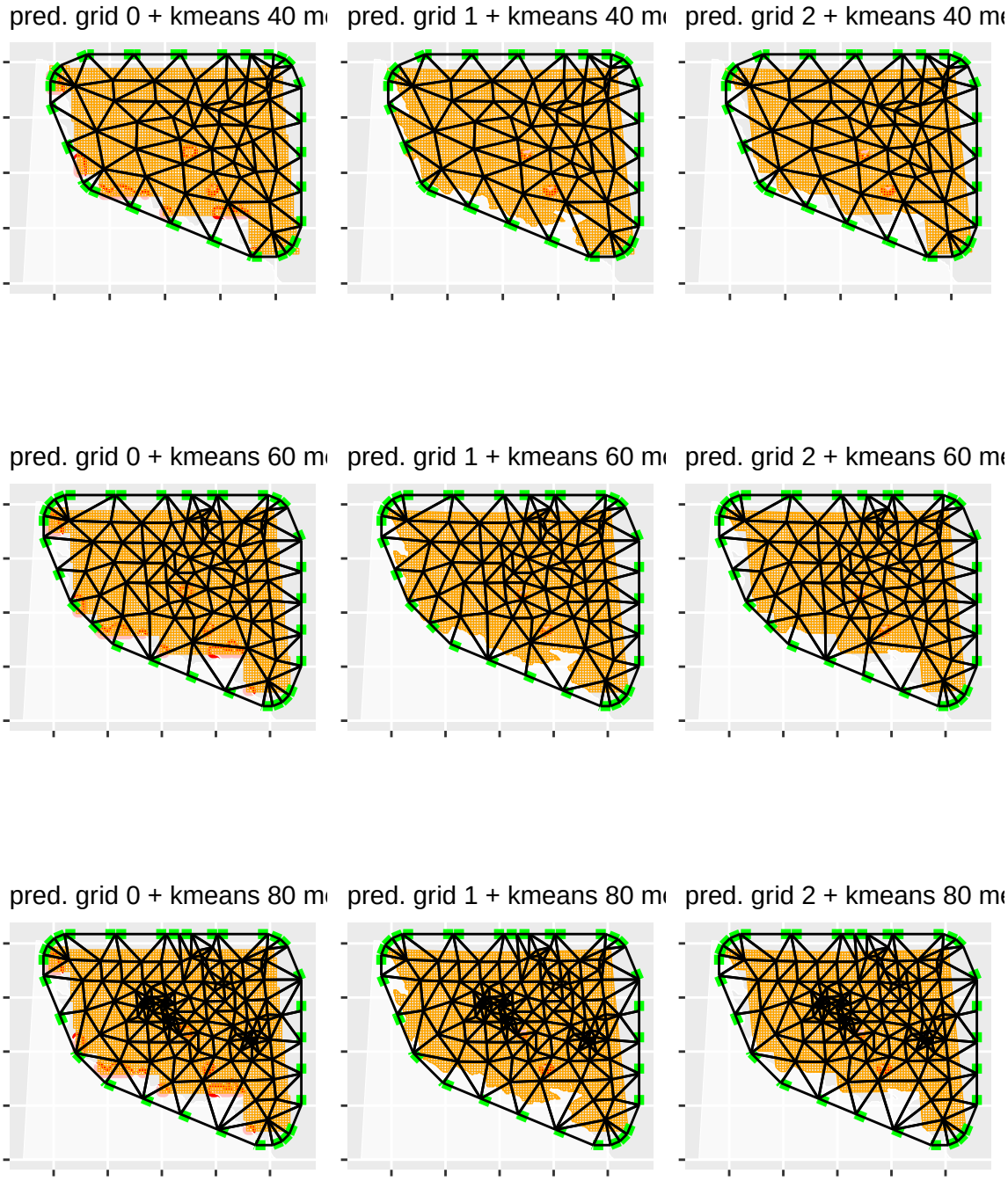

Figure 2. Three prediction grids evaluated (each compared with the "kmeans" evaluation grids). Each prediction grid (orange squares) uses a 5 km grid cell size but covers a slightly different area: grid 0 covers the Pribilof District as defined by NMFS survey strata, and extends across the Pribilof Islands and beyond the shelf edge in some areas (darker orange); grid 1 extends to the shelf edge (500 m depth) and excludes the Pribilof Islands; and grid 2 is limited to depths < 500 m within the Pribilof District and excludes the Pribilof Islands.

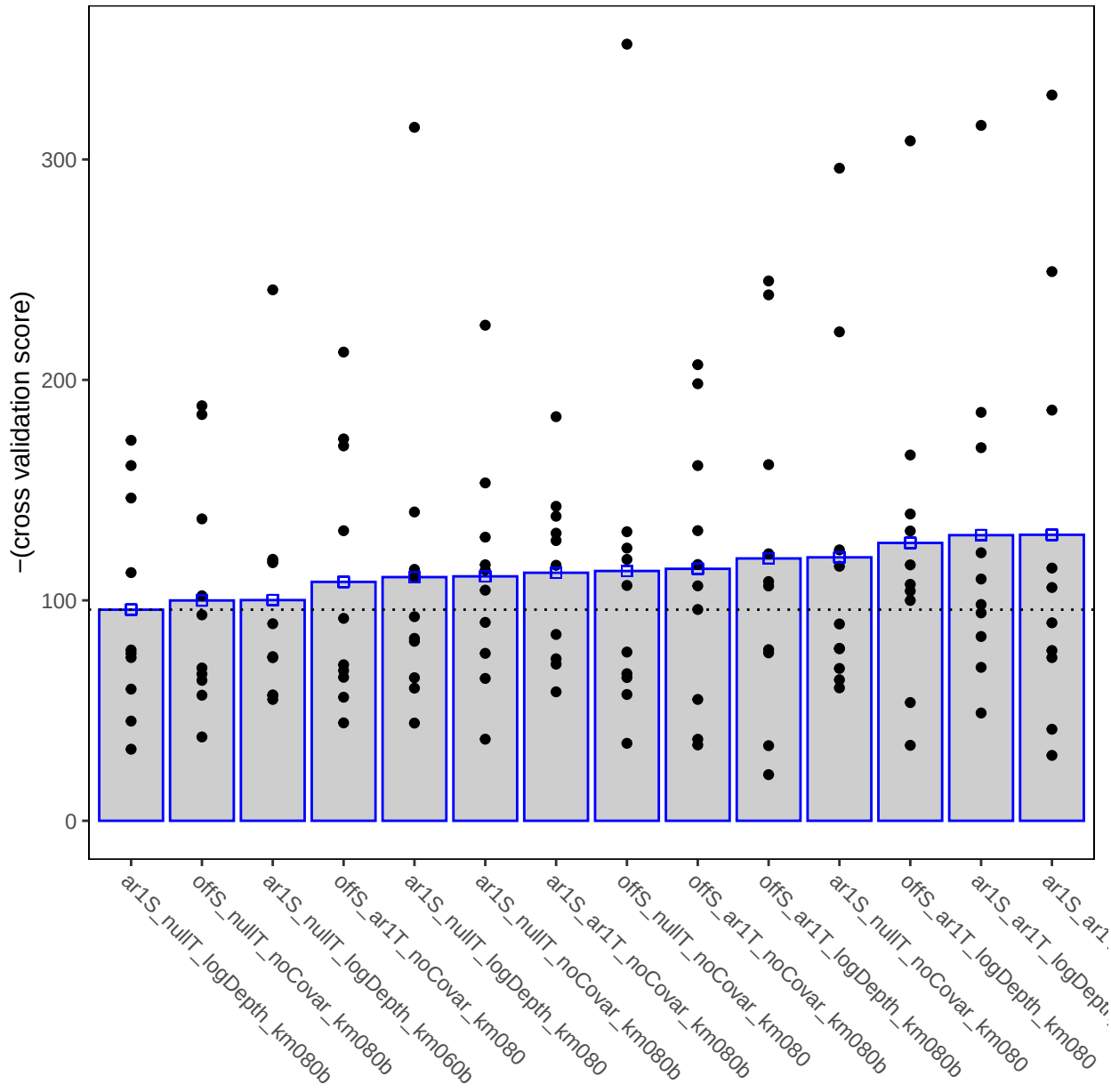

Figure 3. Results from cross-validation of the sdmTMB models, using 10 random folds/model. The cross-validation score for each run is the negative log-likelihood of the left out (out-of-sample) data. The model results are ordered by the average out-of-sample negative log likelihood, with the "best" model atken as the one with the lowest average negative log-likelihood.

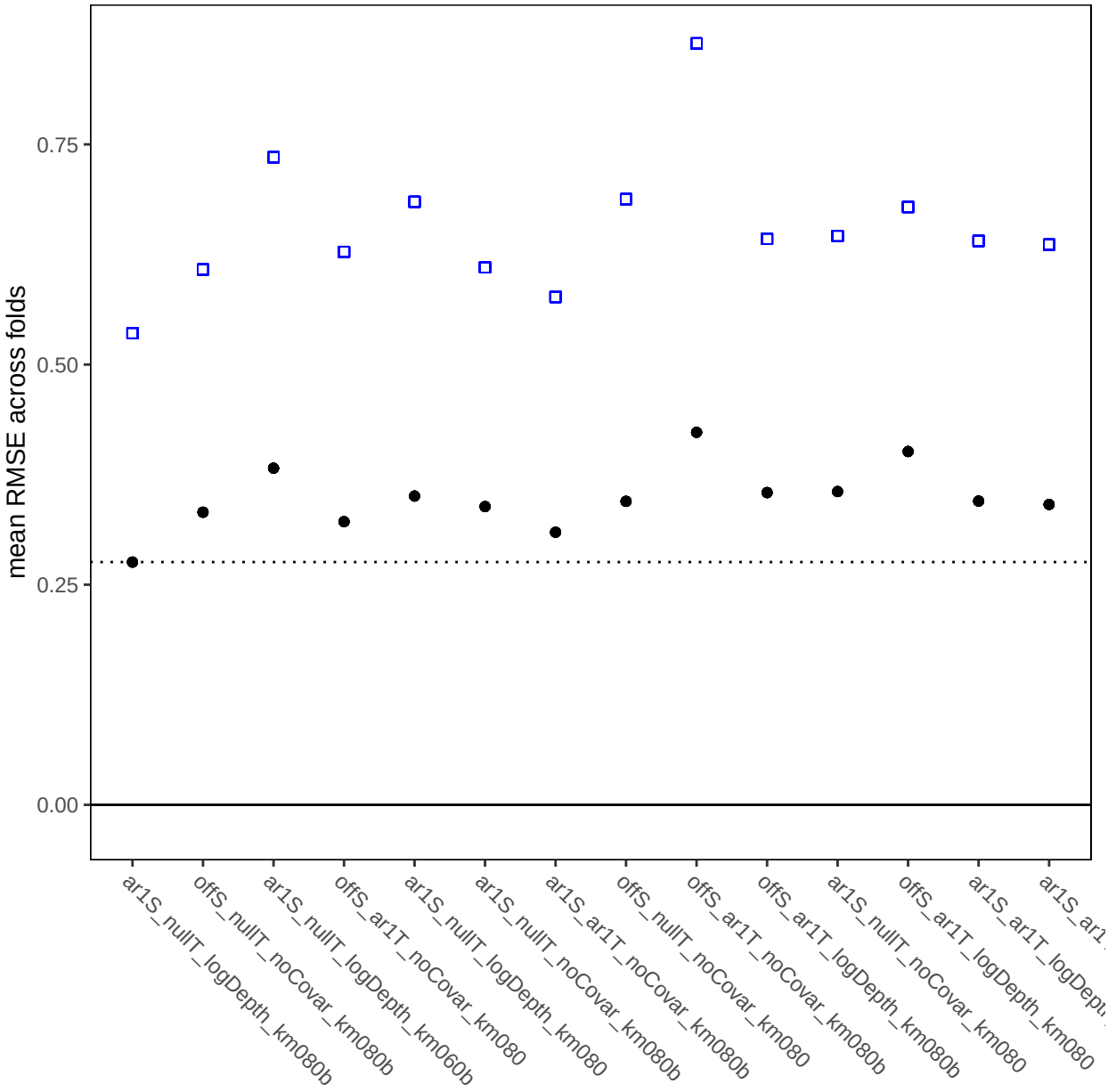

Figure 4. Mean (black symbols) and mean plus standard error (blue symbols) for root mean square errors (RMSEs) from the cross-validation of the sdmTMB models, using 10 random folds/model. The RMSE for each run was the RMSE for the out-of-sample data relative to the model predictions. The model results are ordered left-to-right as in Figure 3.

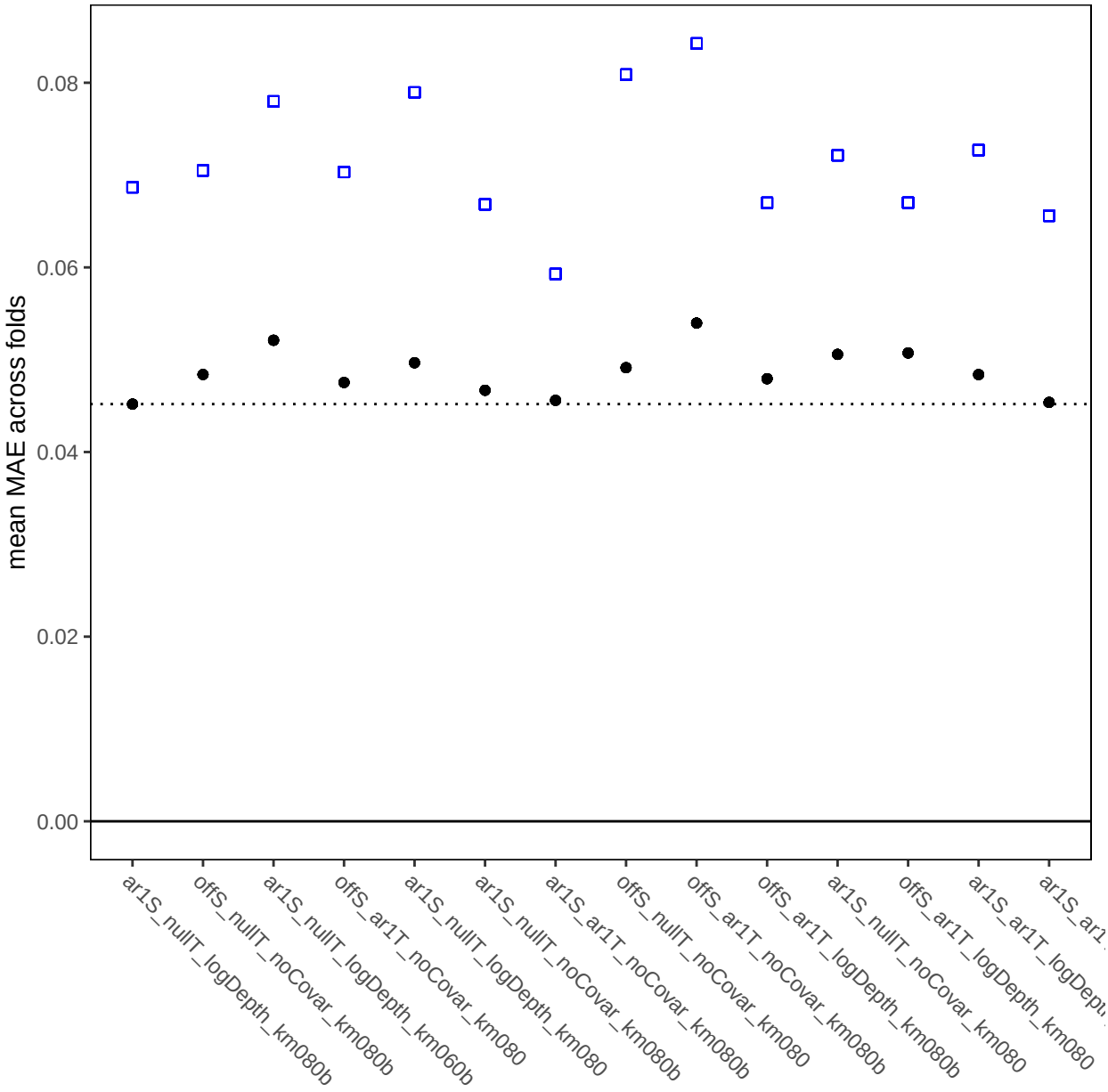

Figure 5. Mean (black symbols) and mean plus standard error (blue symbols) for mean absolute error (MAEs) from the cross-validation of the sdmTMB models, using 10 random folds/model. The MAE for each run was the MAE for the out-of-sample data relative to the model predictions. The model results are ordered left-to-right as in Figure 3.

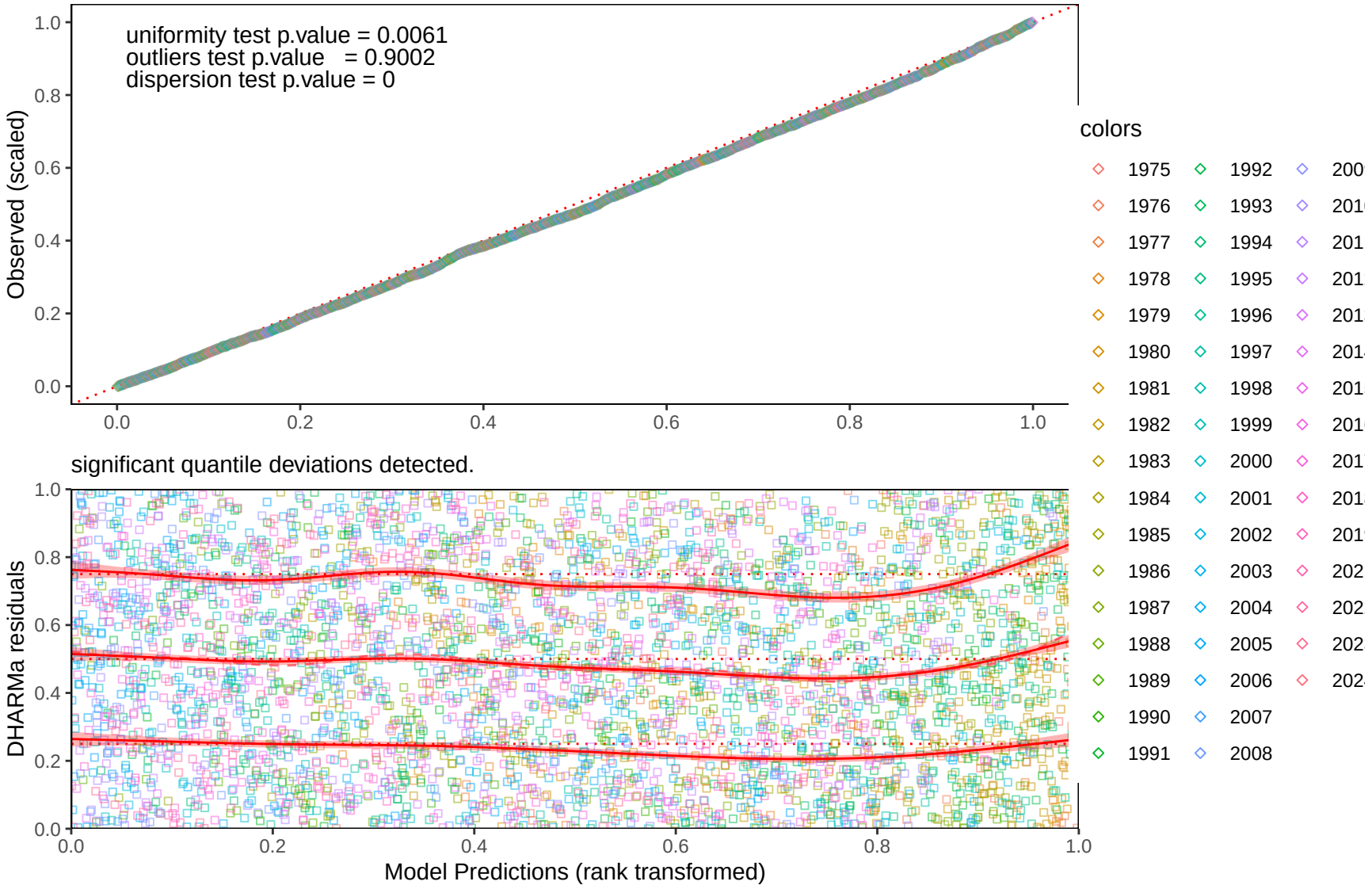

Figure 6. DHARMa residuals diagnostic plots for ar1S_nullT_logDepth_km080b.

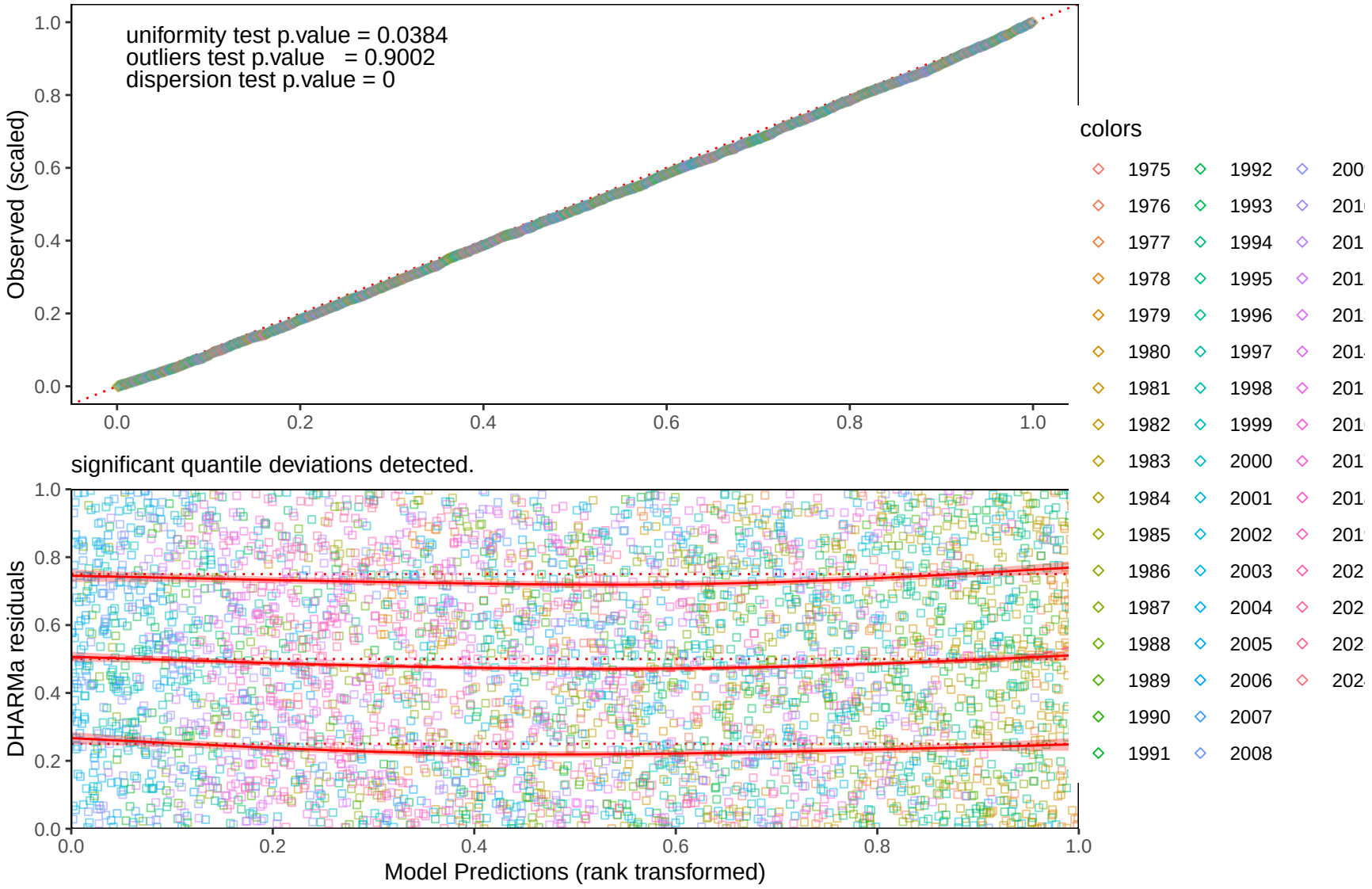

Figure 7. DHARMa residuals diagnostic plots for offS_nullT_noCovar_km080b.

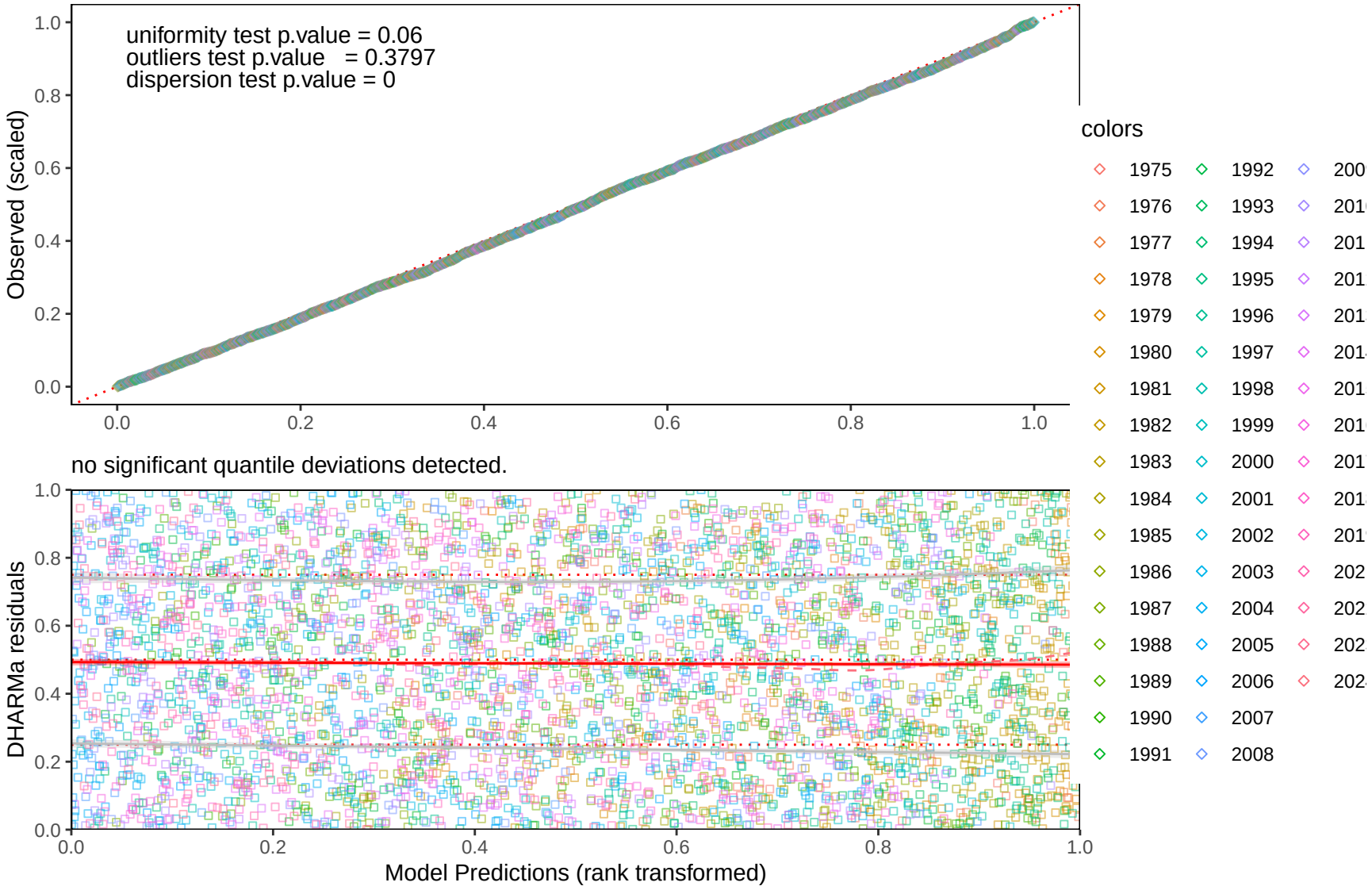

Figure 8. DHARMa residuals diagnostic plots for ar1S_nullT_logDepth_km060b.

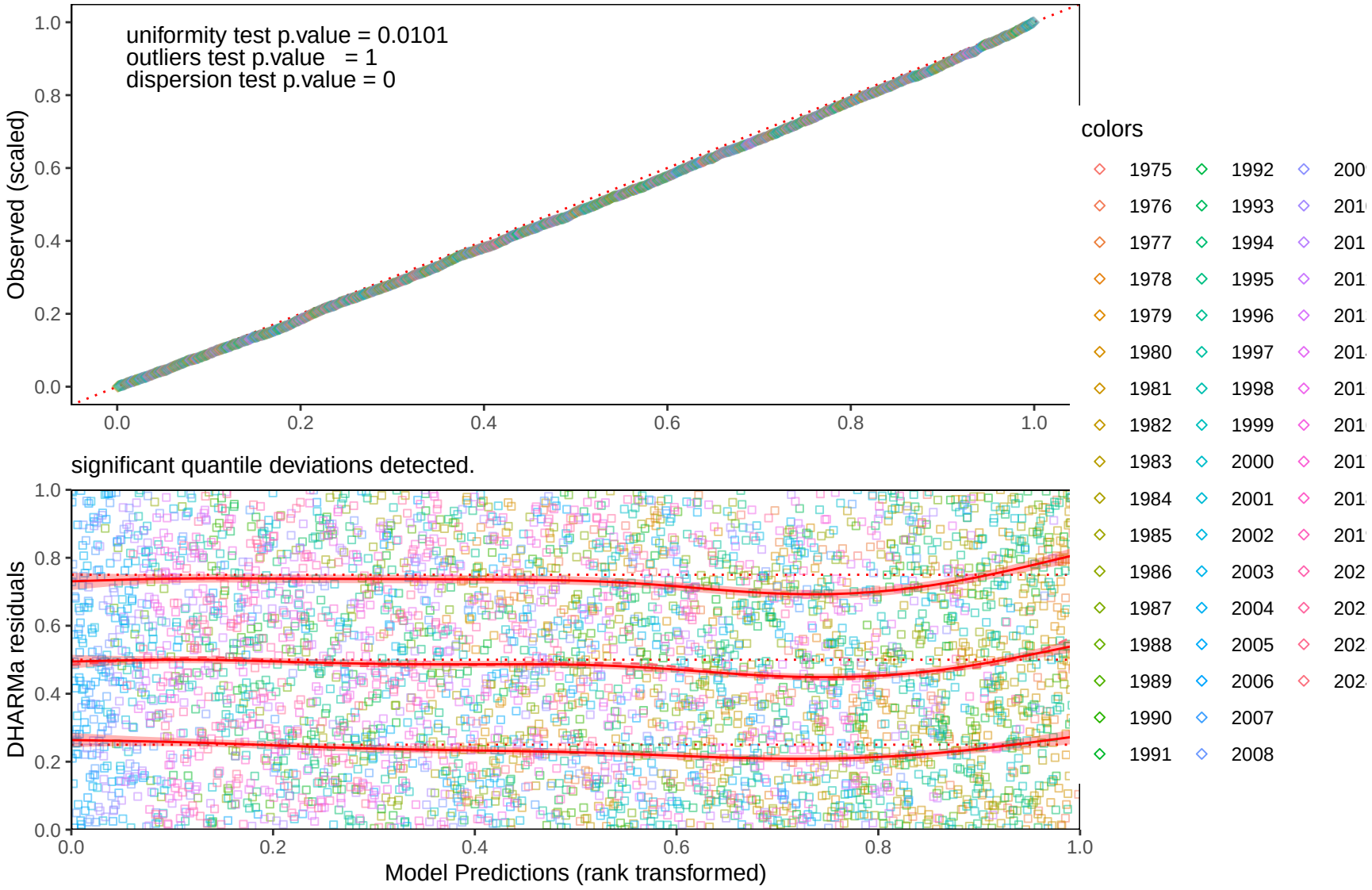

Figure 9. DHARMa residuals diagnostic plots for offS_ar1T_noCovar_km080.

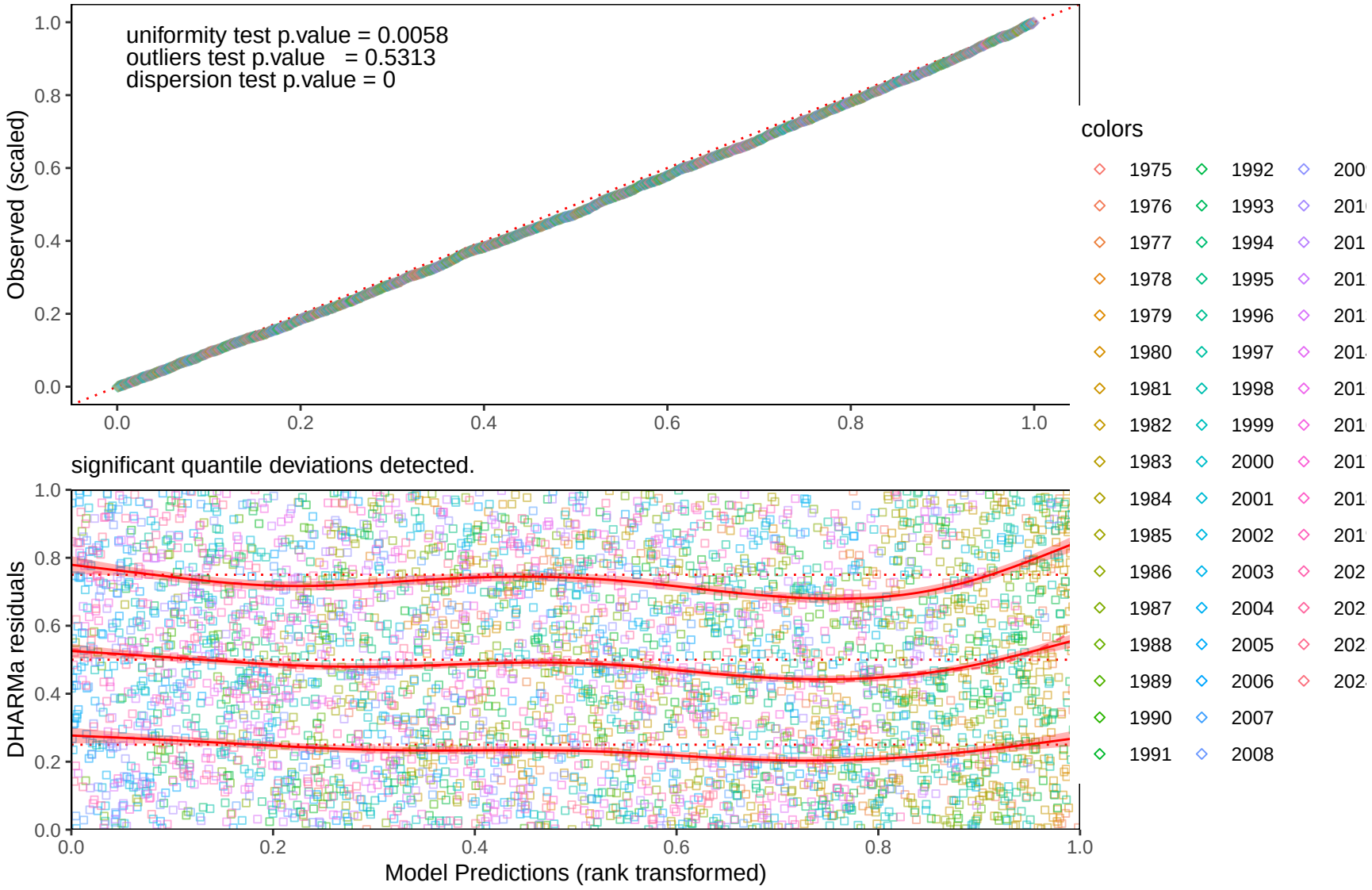

Figure 10. DHARMa residuals diagnostic plots for ar1S_nullT_logDepth_km080.

Figure 11. Scaled spatial residuals (0 to 1) in the 5-year time interval 1975-1979 by model for the five "best" models as ranked in the cross-validation exercise.

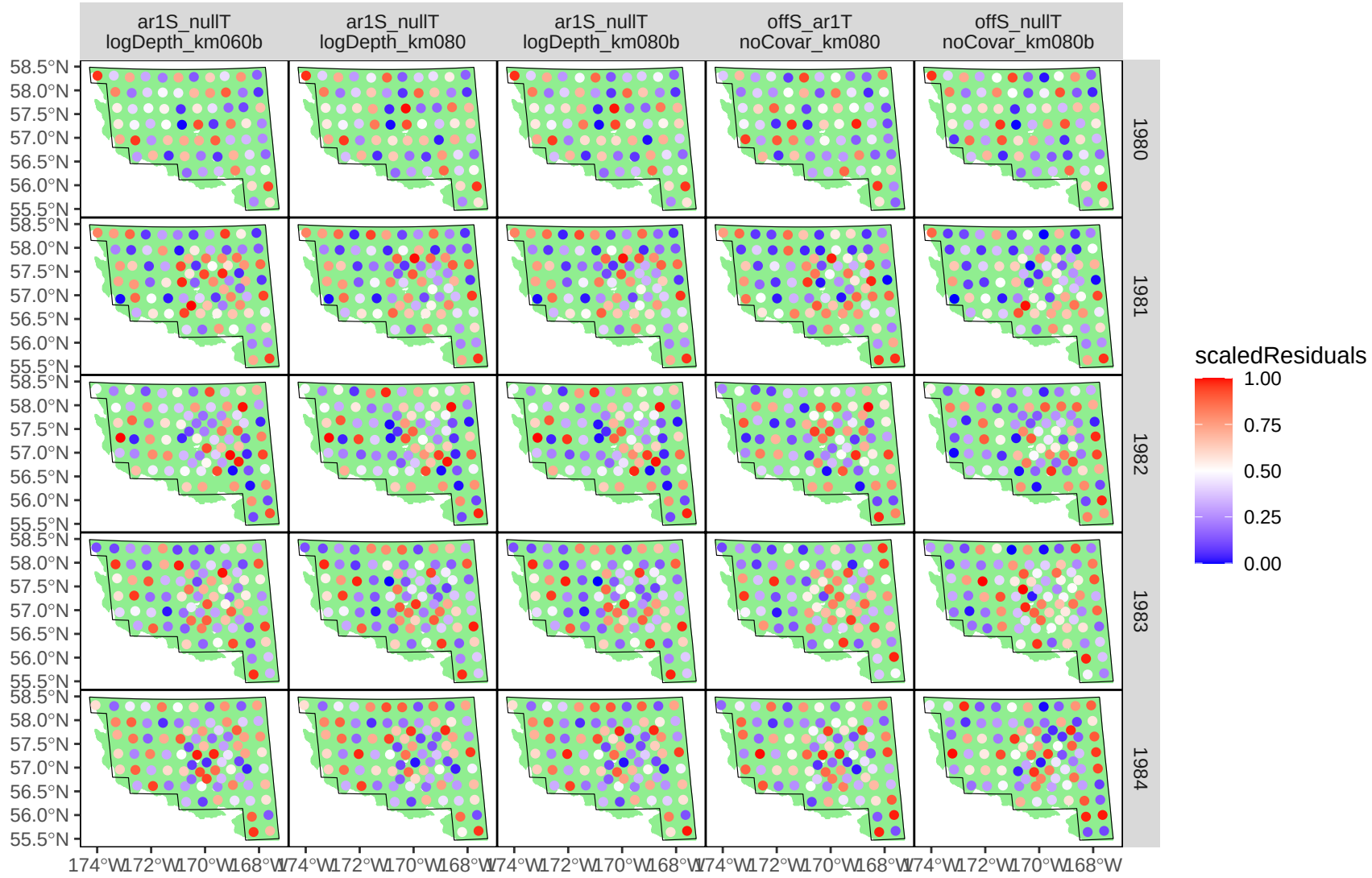

Figure 12. Scaled spatial residuals (0 to 1) in the 5-year time interval 1980-1984 by model for the five "best" models as ranked in the cross-validation exercise.

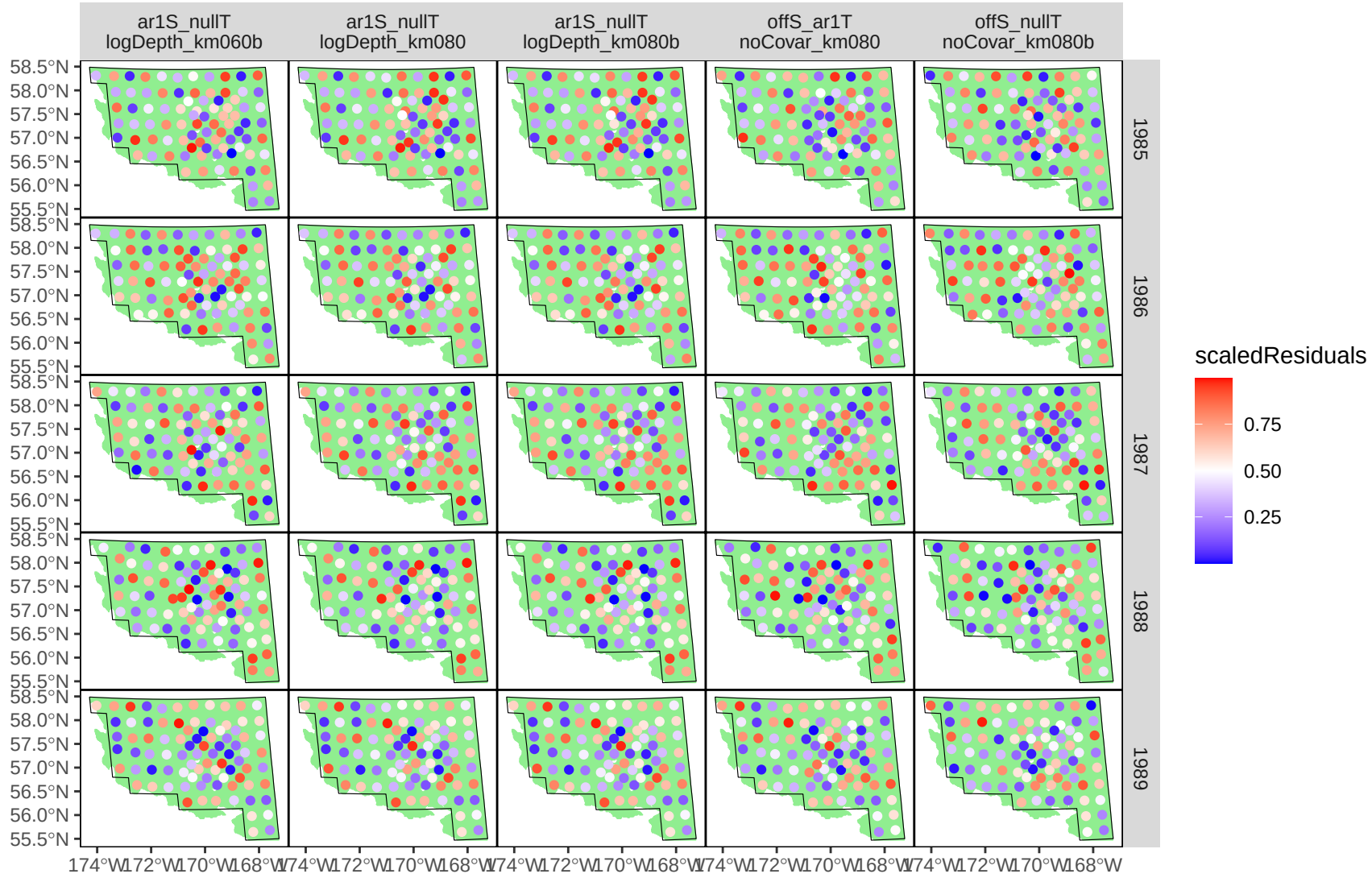

Figure 13. Scaled spatial residuals (0 to 1) in the 5-year time interval 1985-1989 by model for the five "best" models as ranked in the cross-validation exercise.

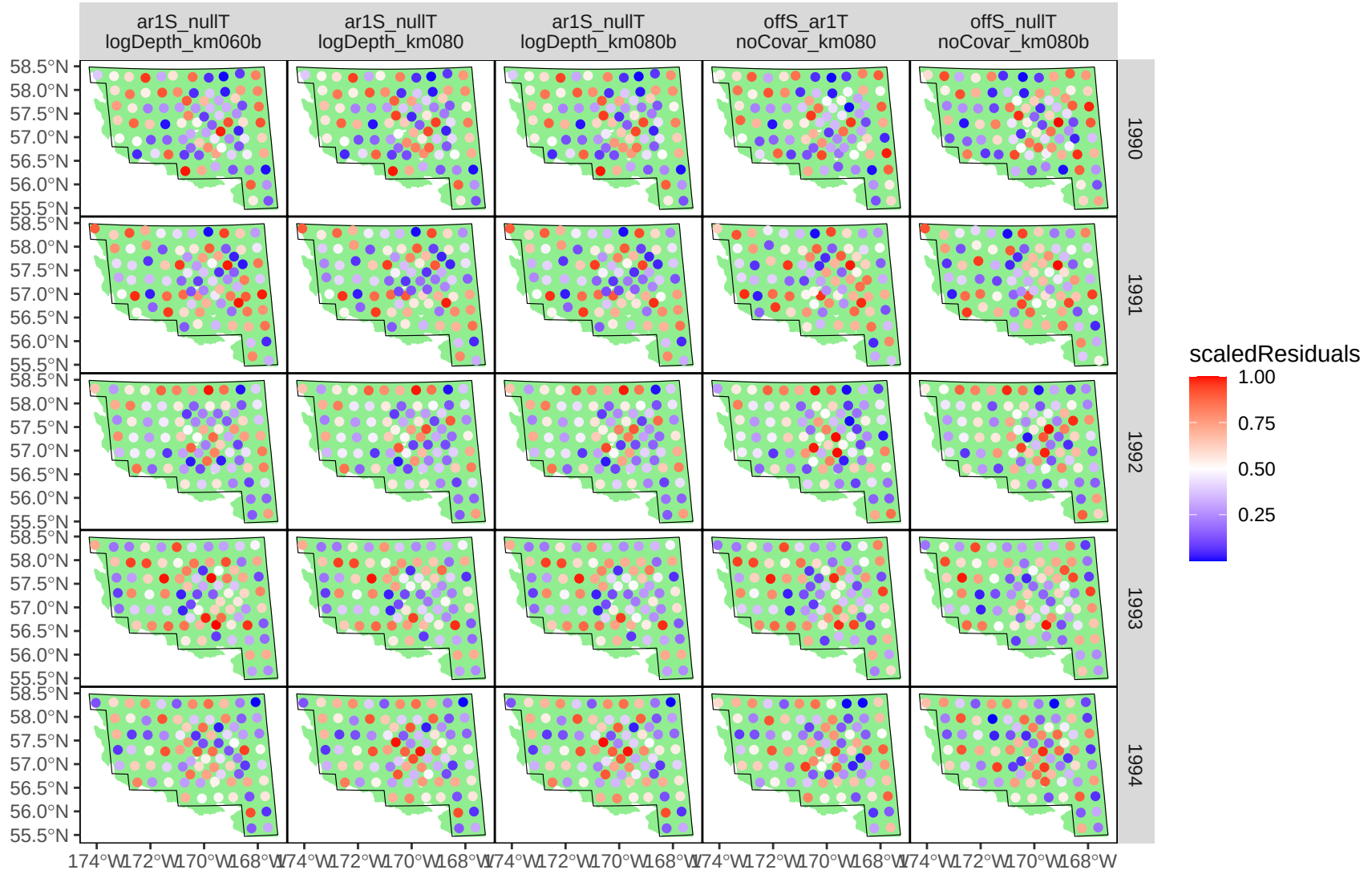

Figure 14. Scaled spatial residuals (0 to 1) in the 5-year time interval 1990-1994 by model for the five "best" models as ranked in the cross-validation exercise.

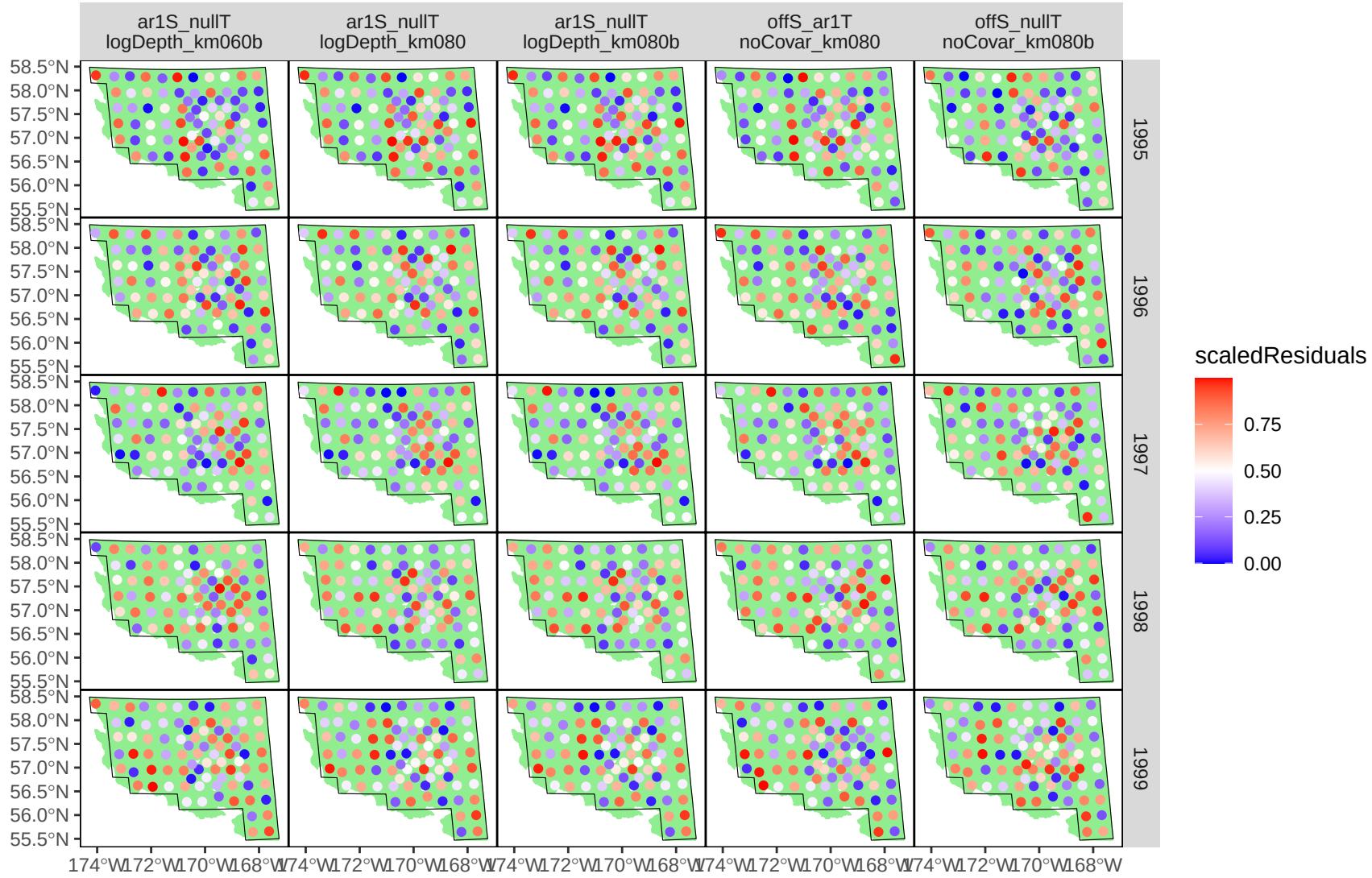

Figure 15. Scaled spatial residuals (0 to 1) in the 5-year time interval 1995-1999 by model for the five "best" models as ranked in the cross-validation exercise.

Figure 16. Scaled spatial residuals (0 to 1) in the 5-year time interval 2000-2004 by model for the five "best" models as ranked in the cross-validation exercise.

Figure 17. Scaled spatial residuals (0 to 1) in the 5-year time interval 2005-2009 by model for the five "best" models as ranked in the cross-validation exercise.

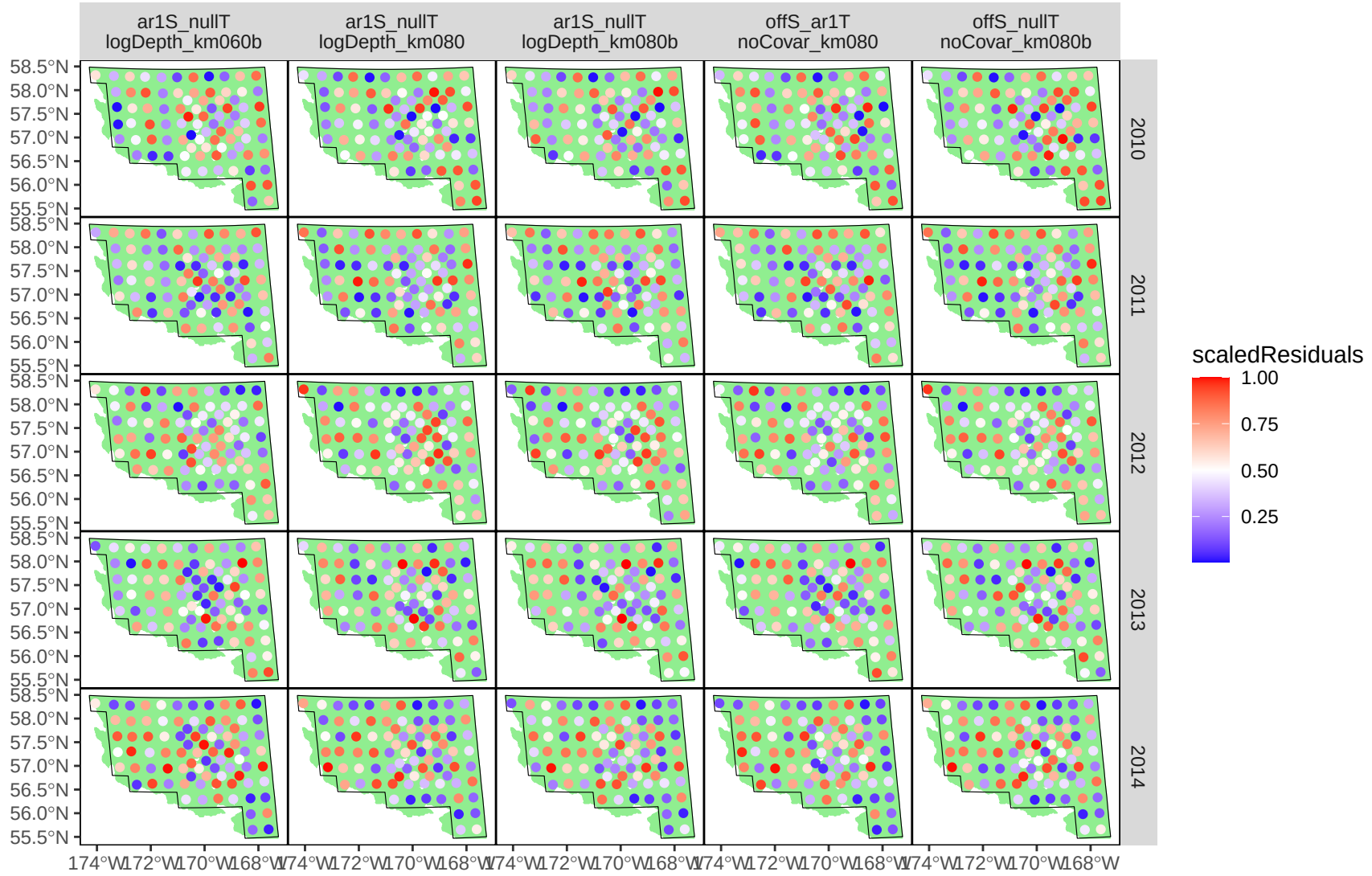

Figure 18. Scaled spatial residuals (0 to 1) in the 5-year time interval 2010-2014 by model for the five "best" models as ranked in the cross-validation exercise.

Figure 19. Scaled spatial residuals (0 to 1) in the 5-year time interval 2015-2019 by model for the five "best" models as ranked in the cross-validation exercise.

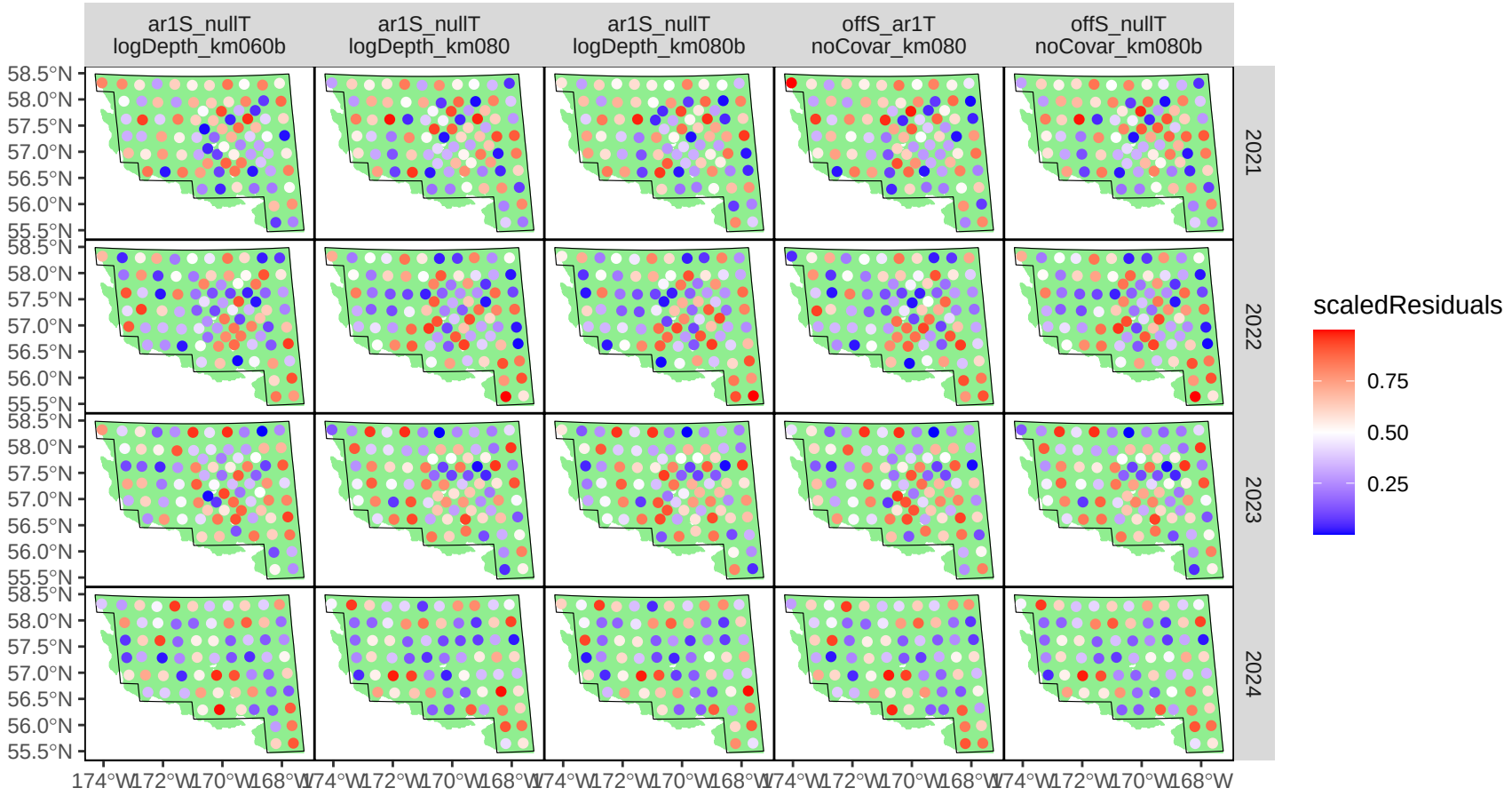

Figure 20. Scaled spatial residuals (0 to 1) in the 5-year time interval 2020-2024 by model for the five "best" models as ranked in the cross-validation exercise.

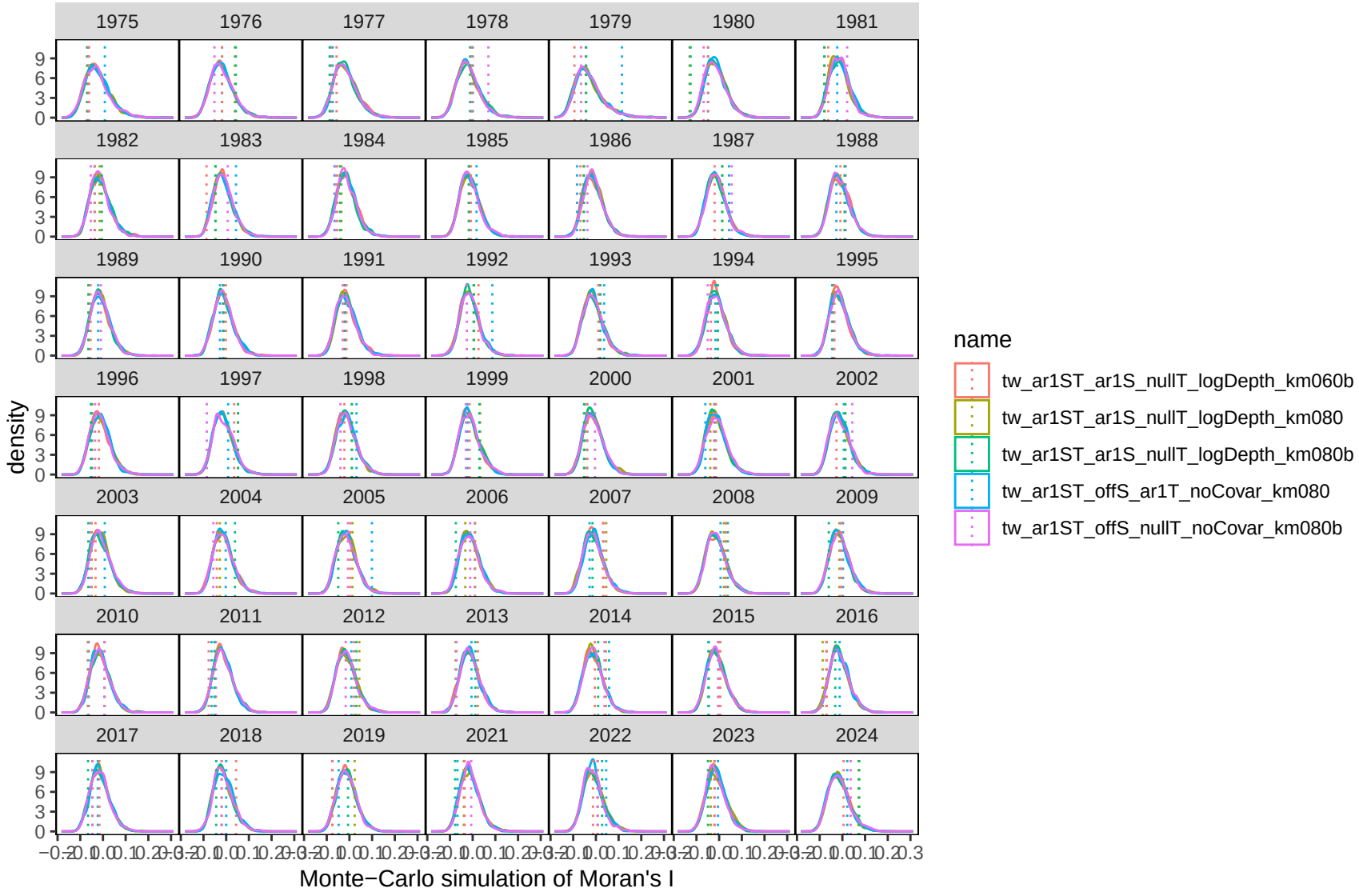

Figure 21. Null distributions (from permutation tests) and observed values of Moran's I statistic for spatial correlation.

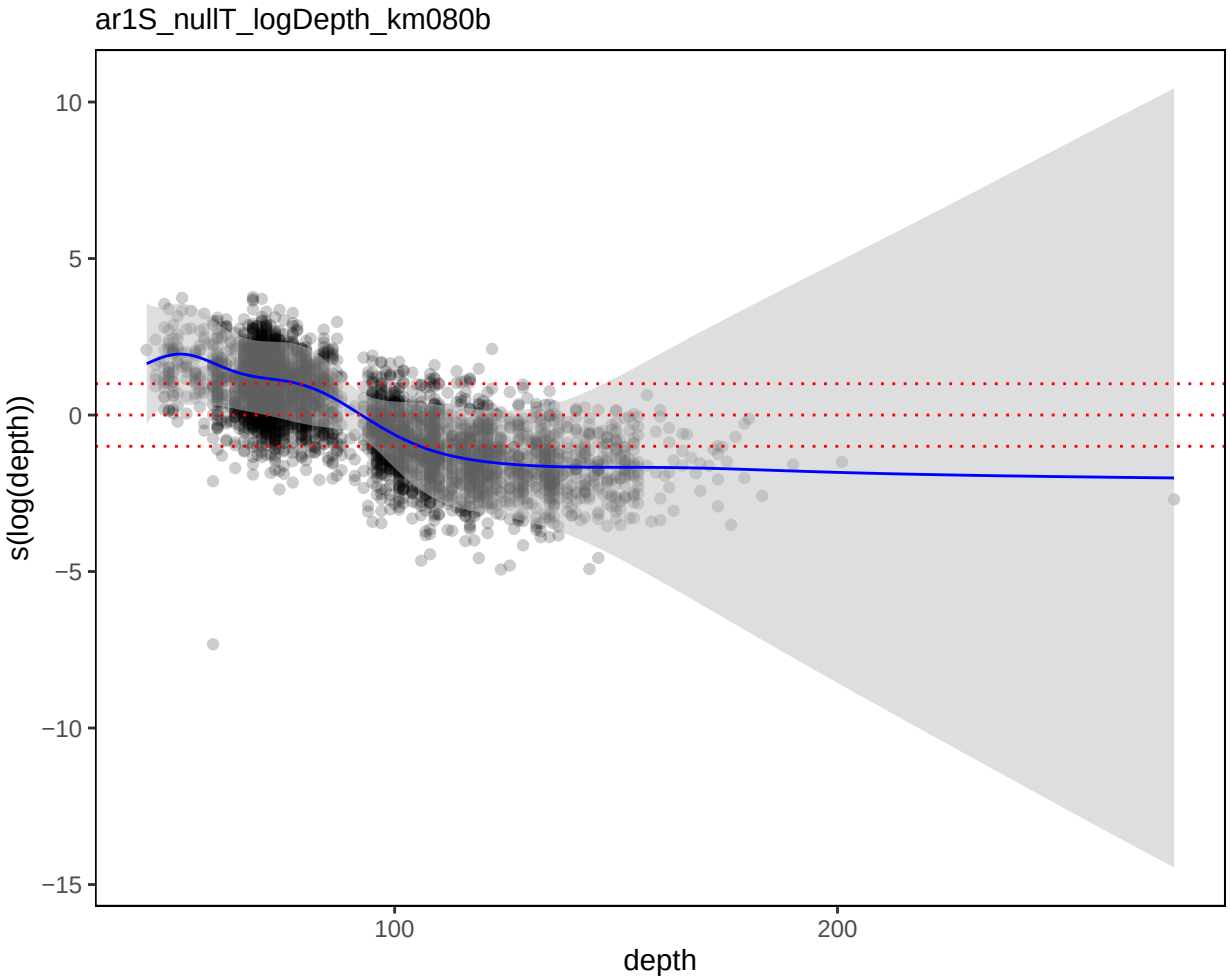

Figure 22. Estimated smooth function of ln(bottom depth) as a main effect covariate, on the link (ln) scale, for Model ar1S_nullT_logDepth_km080b. Line: estimated effect; shading: 95% confidence intervals; points: residuals; horiontal dotted lines are plotted at link-scale values of -1, 0, and 1.

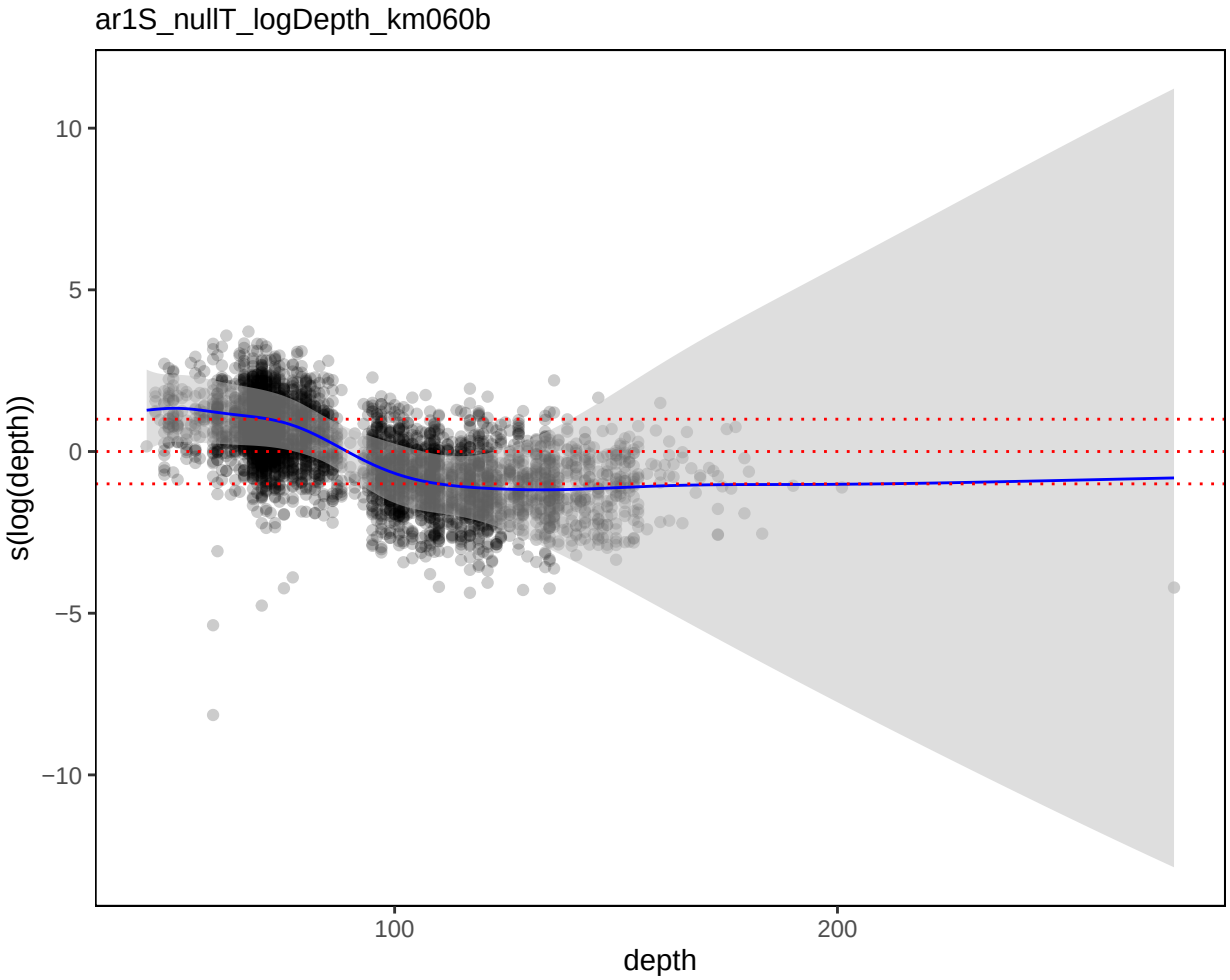

Figure 23. Estimated smooth function of ln(bottom depth) as a main effect covariate, on the link (ln) scale, for Model ar1S_nullT_logDepth_km060b. Line: estimated effect; shading: 95% confidence intervals; points: residuals; horiontal dotted lines are plotted at link-scale values of -1, 0, and 1.

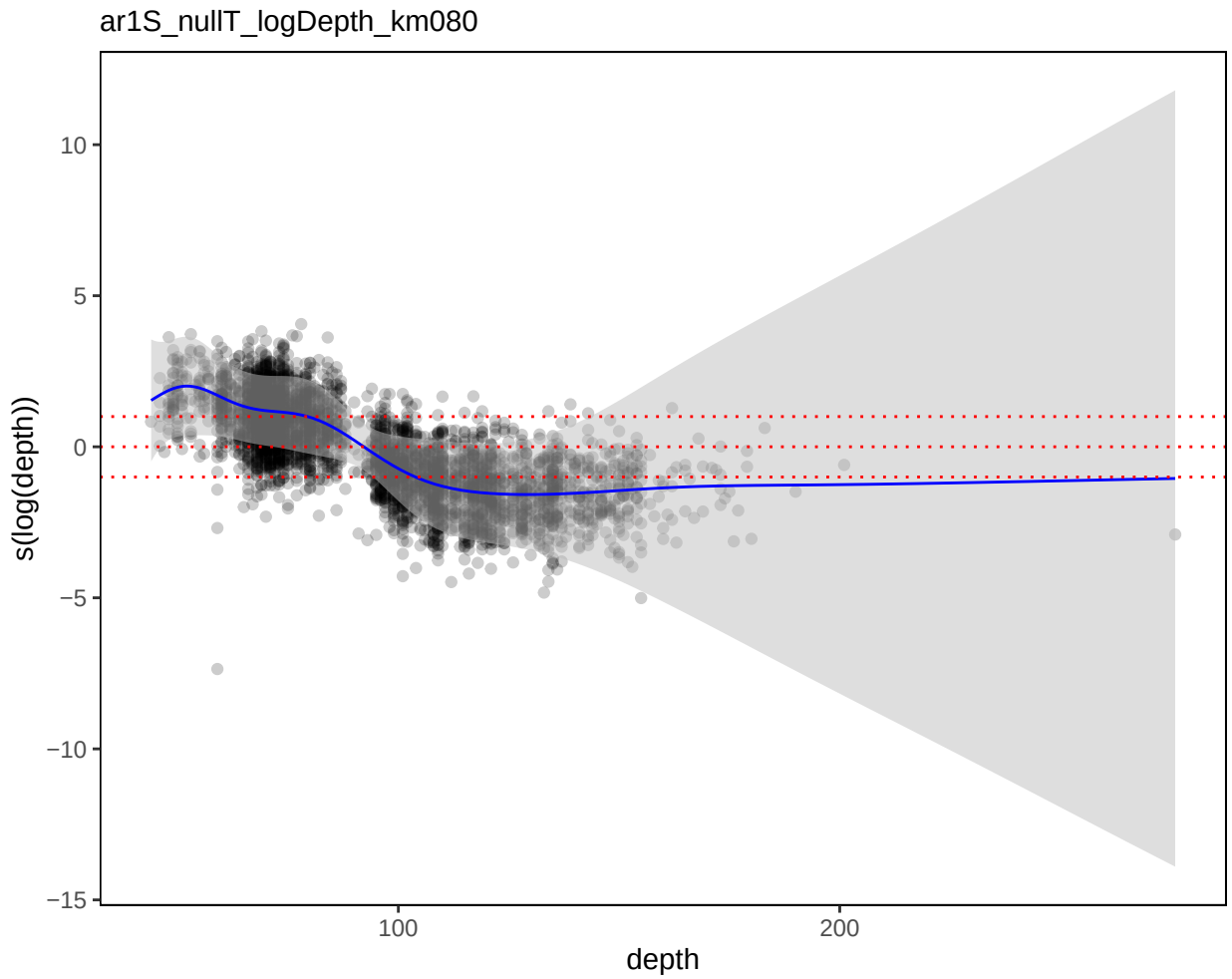

Figure 24. Estimated smooth function of ln(bottom depth) as a main effect covariate, on the link (ln) scale, for Model ar1S_nullT_logDepth_km080. Line: estimated effect; shading: 95% confidence intervals; points: residuals; horiontal dotted lines are plotted at link-scale values of -1, 0, and 1.

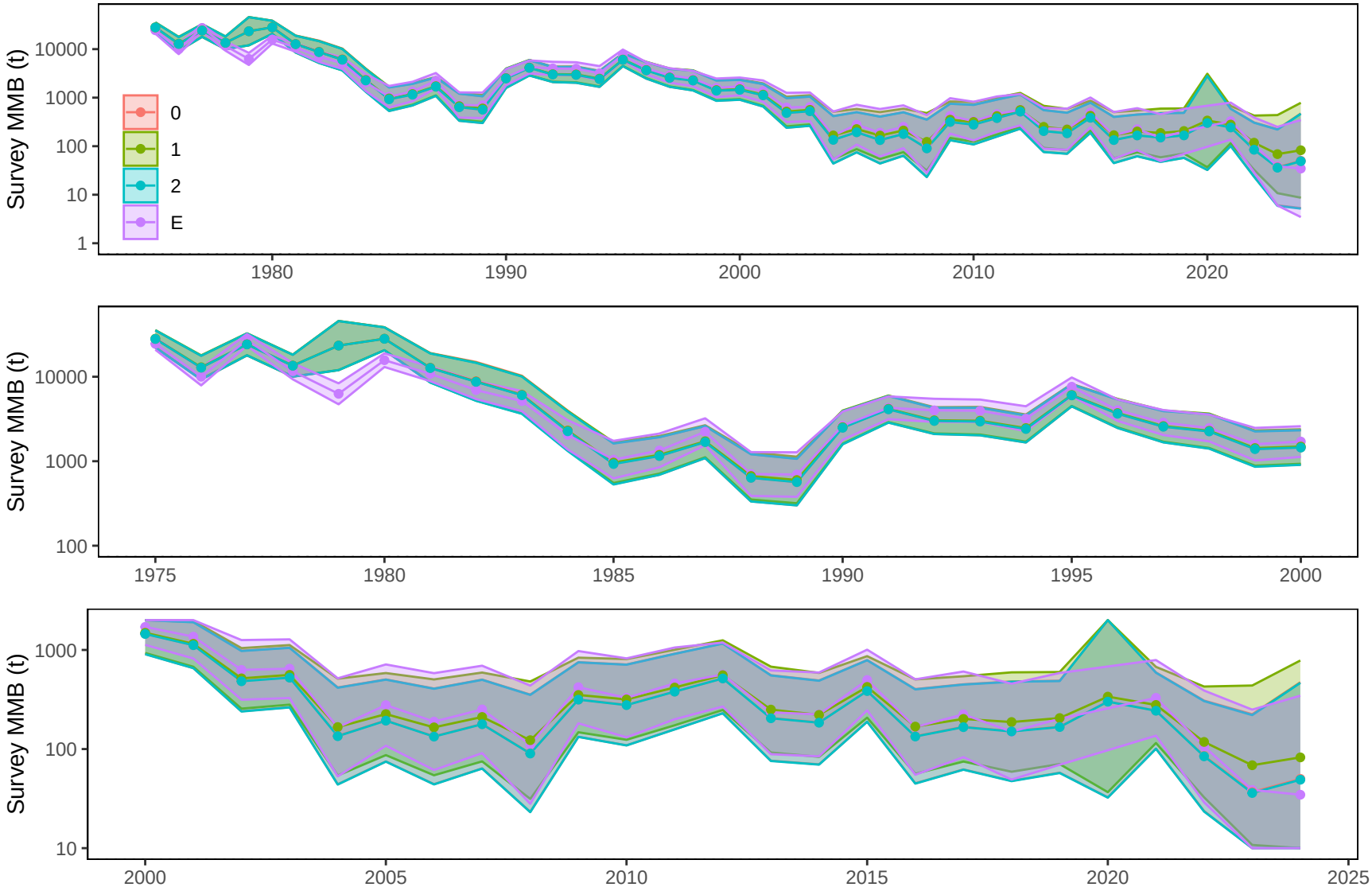

Figure 25. Comparison of `sdmTMB` model-based indices (colored points, lines, and regions) using different prediction grids for Model tw_ar1ST_ar1S_nullT_logDepth_km080b. The legend indcates the fine scale prediction grid (0, 1, or 2) or the original haul locations ("E") used to obtain the area-expanded survey MMB. The shaded areas represent 80% confidence intervals. Upper plot: full time series; middle plot: 1975-2000; lower plot: 2000-current year.

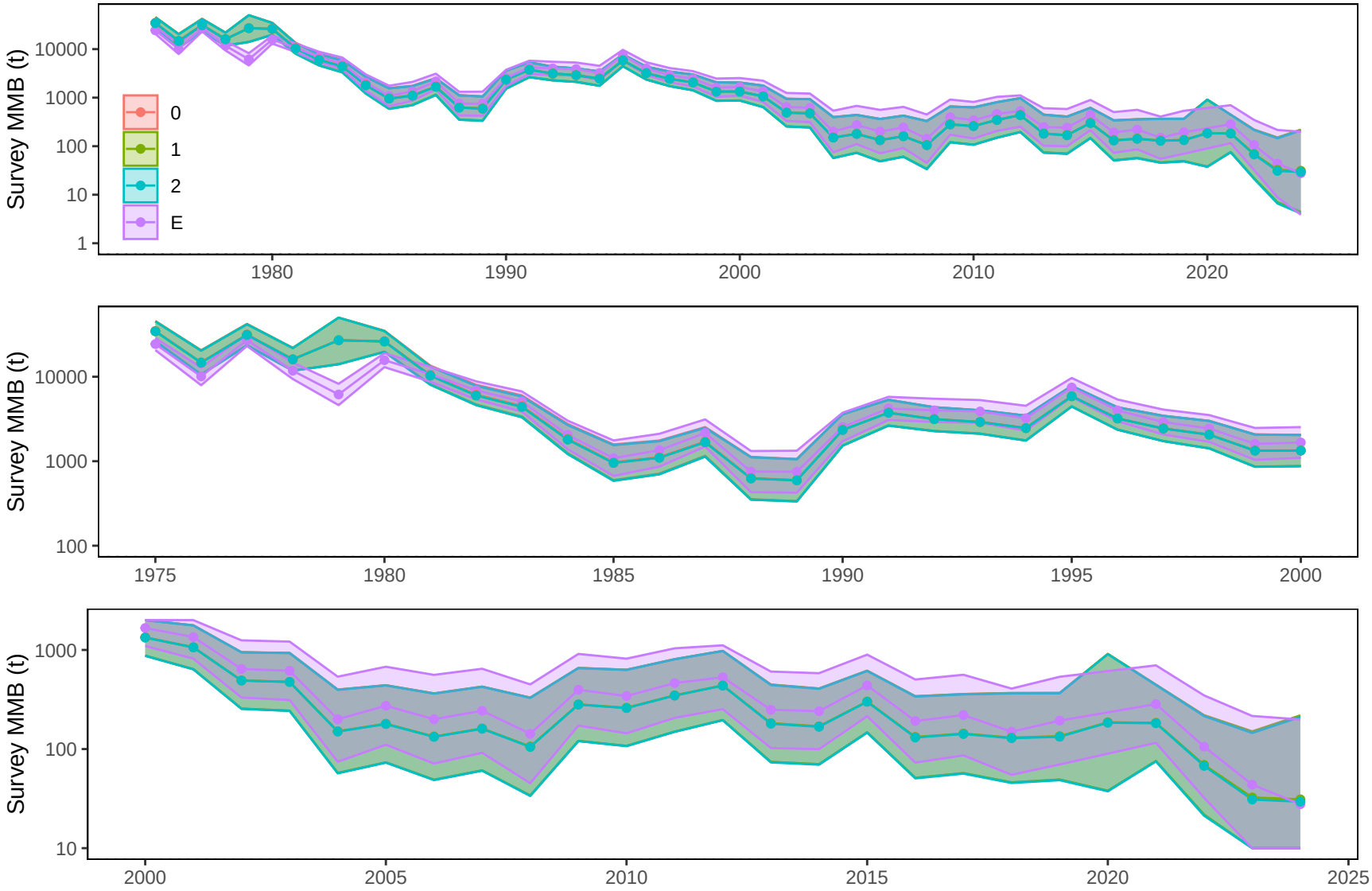

Figure 26. Comparison of `sdmTMB` model-based indices (colored points, lines, and regions) using different prediction grids for Model tw_ar1ST_offS_nullT_noCovar_km080b. The legend indcates the fine scale prediction grid (0, 1, or 2) or the original haul locations ("E") used to obtain the area-expanded survey MMB. The shaded areas represent 80% confidence intervals. Upper plot: full time series; middle plot: 1975-2000; lower plot: 2000-current year.

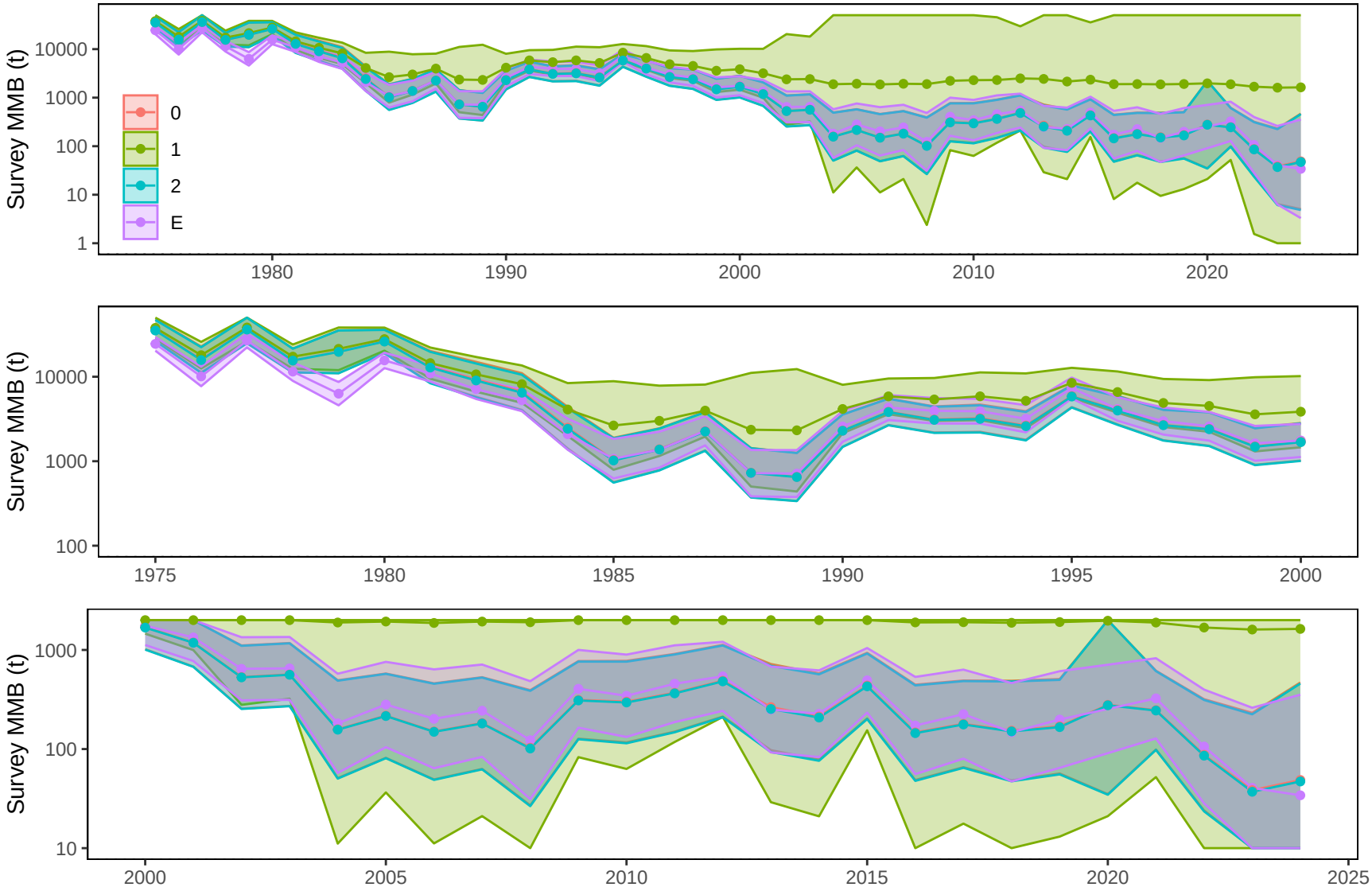

Figure 27. Comparison of `sdmTMB` model-based indices (colored points, lines, and regions) using different prediction grids for Model tw_ar1ST_ar1S_nullT_logDepth_km060b. The legend indcates the fine scale prediction grid (0, 1, or 2) or the original haul locations ("E") used to obtain the area-expanded survey MMB. The shaded areas represent 80% confidence intervals. Upper plot: full time series; middle plot: 1975-2000; lower plot: 2000-current year.

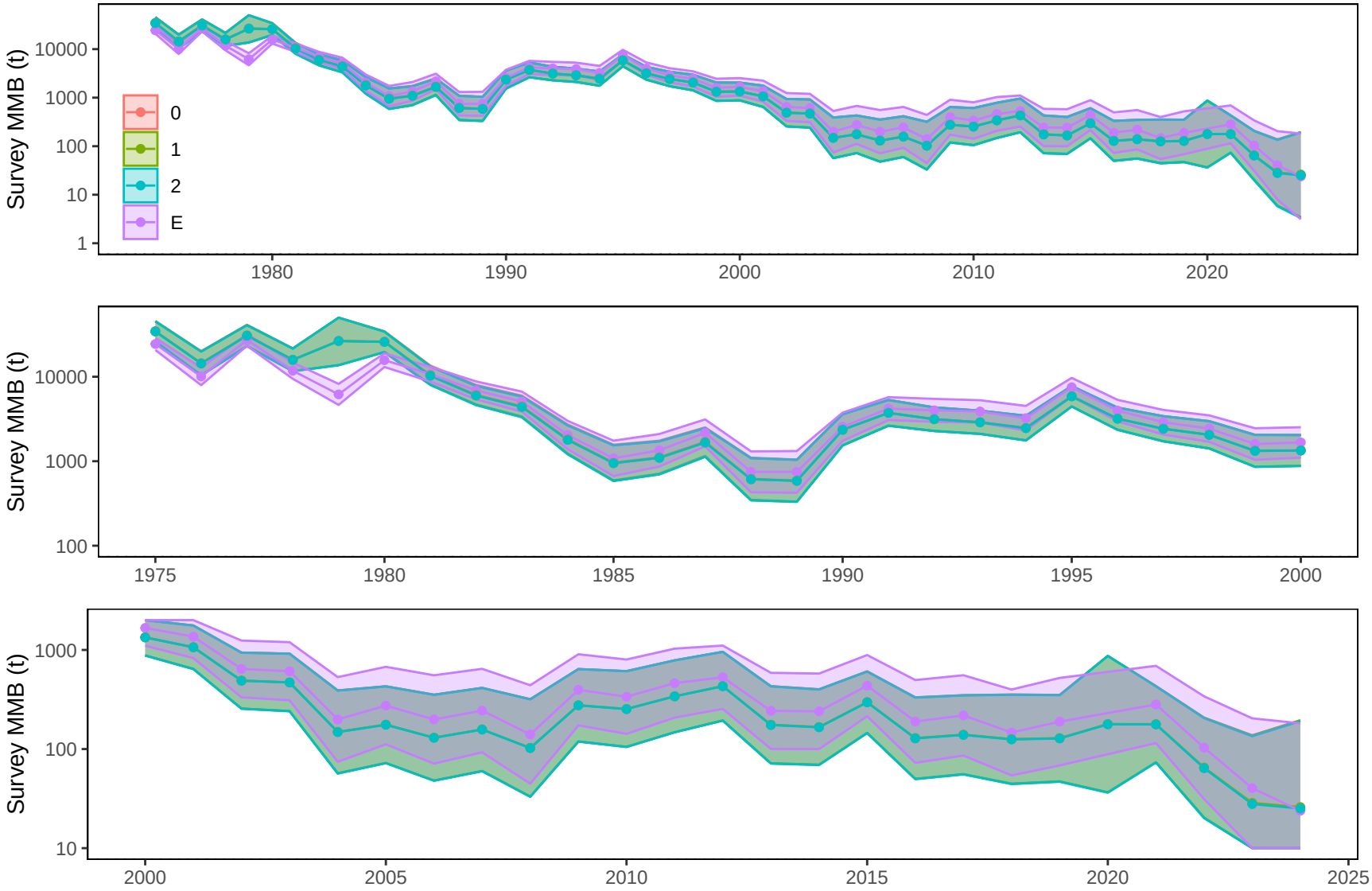

Figure 28. Comparison of `sdmTMB` model-based indices (colored points, lines, and regions) using different prediction grids for Model tw_ar1ST_offS_ar1T_noCovar_km080. The legend indcates the fine scale prediction grid (0, 1, or 2) or the original haul locations ("E") used to obtain the area-expanded survey MMB. The shaded areas represent 80% confidence intervals. Upper plot: full time series; middle plot: 1975-2000; lower plot: 2000-current year.

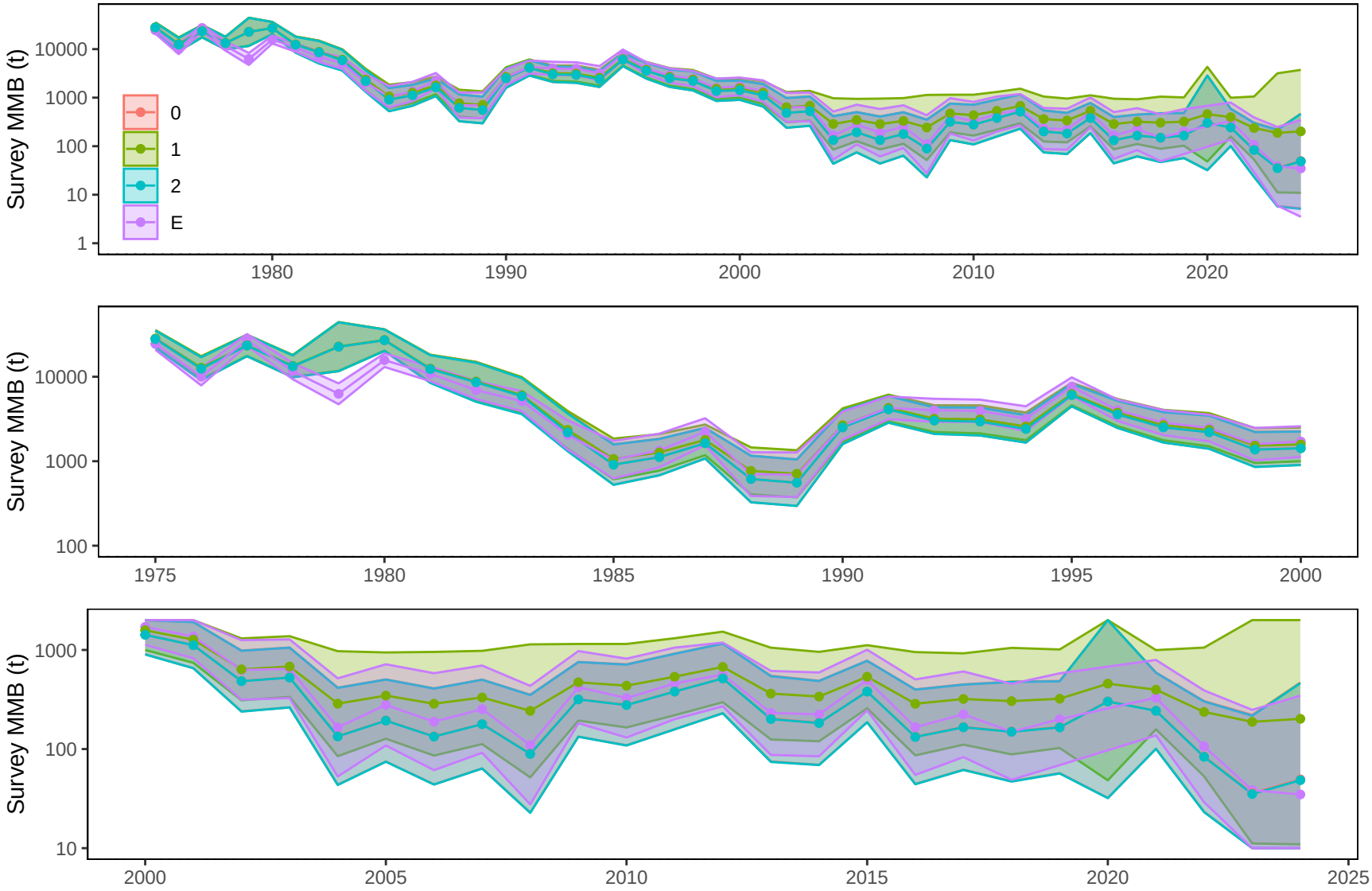

Figure 29. Comparison of `sdmTMB` model-based indices (colored points, lines, and regions) using different prediction grids for Model tw_ar1ST_ar1S_nullT_logDepth_km080. The legend indcates the fine scale prediction grid (0, 1, or 2) or the original haul locations ("E") used to obtain the area-expanded survey MMB. The shaded areas represent 80% confidence intervals. Upper plot: full time series; middle plot: 1975-2000; lower plot: 2000-current year.