Appendix B. VAST Model of the Dover Sole in the Gulf of Alaska

Andrea Havron

Summary

Biomass observations from individual survey trawls were run through the VAST (Vector Autoregressive Spatio-Temporal) model to obtain model-based population indices (Thorson and Barnett, 2017). This approach can be used to fill in missing area-depth gaps using spatial information and has the potential to benefit the Dover sole assessment as surveys inconsistently extend out to 1000m where this species can be detected. VAST results were compared to design-based estimates from stratified surveys and results from a random effects (RE) model used to fill in area-depth gaps. VAST models run with default settings led to biomass estimates that were consistently above those from the RE model, while VAST models run with the Gamma observation model led to biomass estimates that were, for most years, similar to those from the RE model. Years in which VAST estimates either higher or lower than estimates from the RE model did not always correspond to years in which surveys were truncated to 500m. Further research is recommended before VAST estimates are used in the Dover Sole stock assessment. In addition to presenting results on a VAST GOA Dover sole model, this appendix also details model validation steps that are useful when evaluating VAST models.

Introduction

Spatial correlation has long been known to present a source of extra variability in ecological observations due to a fundamental property in geography: observations in close proximity tend to be more similar than observations distant from each other (Figure 1). Spatial correlation arises from different ecological processes. With respect to species distributions, the relationship between a species and its environment leads to spatial correlation that directly results from species organizing themselves around the spatial pattern of the environment. A second notable source of spatial correlation is the spatial structure of a species that is independent of the environment, such as aggregation patterns and patterns resulting from competition, predation, etc. Regardless of its source, spatial correlation leads to extra variability in observations. This, in turn, invalidates a key statistical assumption required for most basic analyses, that observations are independent.

The development of design-based population assessment methods were in part due to the need to minimize the impact of spatial correlation and environmental stochasticity on population estimates. Historically, model-based approaches that explicitly incorporated spatial correlation into model design were available, yet computational and statistical challenges made these approaches inferior to design-based methods. Recent advancements in computer technology and spatial modeling software tools, such as VAST, are now available to apply model-based methods to assessing population indices both precisely and efficiently.

VAST models used for stock assessment estimate two levels of spatial correlation for a species: an overall effect that is estimated by aggregating annual data; and a spatio-temporal effect, that estimates an annual spatial effect. The former can be thought of an ‘average’ spatial pattern for a species and the latter represents additional annual spatial variability. The ‘average’ spatial pattern can be explained, for

example, by static properties of an environment, such as depth, or consistent areas of species aggregation, such as in the case of stationary species, coral mounds. The spatio-temporal pattern could reflect associations with temporally variable environmental characteristics, such as temperature, or other complex biological patterns, such as migration or events related to high recruitment. With respect to GOA Dover sole, spatial patterns in this species could arise from its ontogenetic movement patterns and its association with depth.

VAST uses spatial correlation information to estimate annual biomass at unobserved locations. The prediction region, for example in GOA: Western, Central, and Eastern GOA, is covered by a grid. For each year, the 'average' spatial and spatio-temporal effect is added to the annual mean biomass value to derive an annual, locally unique biomass prediction for each grid cell. These values are then summed to obtain annual biomass estimates. As a distinct advantage over design-based methods, information can be borrowed from nearby locations, supplementing data when survey coverage is not complete.

Design-based methods are known to provide unbiased estimates of biomass, however, this is only the case when strata design is optimal to the species. Biases, therefore, can be occurring in the design-based estimates of Dover sole. Due to Dover sole's ontogenetic movement patterns, this species additionally occupies depths that are not consistently surveyed. Within the sixteen years of survey data, eleven years extended out to 700m, of which 6 years additionally extended out to 1000m. Random effects (RE) models are used in the Dover sole stock assessment to fill in depth-area gaps in surveys. These models do not take spatial correlation explicitly into account. Research has shown that model-based methods result in more precise biomass estimates (Thorson, 2019, Thorson, Shelton, Ward, & Skaug, 2015). While they are not guaranteed to be unbiased, estimates from VAST models can be optimally defined for a given species, potentially leading to less bias than a sub-optimal design-based approach.

Analytic Approach

The VAST model was run using settings for the purpose of index standardization. Under this framework, the underlying observation model is a 'Poisson-link' delta-lognormal model that specifies an expected encounter probability that is correlated with an expected positive catch rate (Thorson, 2017). Each expectation is a function of an annual intercept value plus additional latent effects (spatial, spatio-temporal, etc.).

A spatial random effects model is specified to handle extra variation in the observations as spatial correlation. The spatial likelihood is defined as a Gaussian Markov Random Field (GMRF). This likelihood is a multivariate normal likelihood that defines a precision matrix, which is the inverse of the multivariate normal covariance matrix. In the GMRF model, the precision matrix has a unique definition such that its inverse represents a covariance matrix defined using a Matérn covariance function, a common covariance structure used when correlations between locations decay with distance (Besag, 1974).

The GMRF likelihood results in efficient computation times as calculations only involve the precision matrix, skipping the computationally expensive task of inverting the covariance matrix. The precision matrix also has the Markov property of sparseness, that is, correlations are only non-zero for observations that are in close proximity. The model can efficiently learn the spatial structure by only considering the nearest neighbors of a given location.

The GMRF was originally developed for lattice, or spatially discrete data. The extension of the GMRF to continuous space relies on discretizing this space into a triangulated grid, or 'mesh'. This step creates

discrete spatial areas by overlaying a grid of triangles over the area of interest (Figure 2). The precision matrix of the GMRF is estimated using distances along the triangle edges and applying a mathematical function derived from engineering called the finite element method to construct the solution to a stochastic partial differential equation (SPDE), the solution being a continuous Gaussian Field with a Matérn covariance matrix (Lindgren, Rue, & Lindstrom, 2011). This method, therefore, provides an approximation to the precision matrix for a continuous Gaussian Field using a discrete GMRF, and thereby carries along all the computational benefits associated with its optimization. Correlation at observation locations are weighted according to the distance between the observation and the nearest triangle vertex. Markov properties hold such that correlations between two observations are only non-zero if locations are within two triangle edges of each other.

The resolution of the estimated spatial correlation structure, or 'spatial field' can be affected by the resolution of the discretization step. The vertices of the mesh triangles are the latent observations of the spatial field. If these are spaced at a distance larger than the spatial range (maximum distance of detectable correlation), then the latent observations are essentially independent and will not be able to detect any spatial correlation in observed data. While minimal research has been conducted on the relationship between maximum edge distance (maximum distance between two triangle vertices on the mesh), recommendations for optimal mesh properties have been made based on the underlying finite element method theory. Recommendations include a boundary area that extends beyond the area used for inference/prediction to about 1/5-1 times the spatial range of the field, and a maximum triangle edge in the interior mesh that is approximately 1/10-1/5th the spatial range (Bakka, 2018).

The 'mesh' method is the preferred approach in VAST and is used when calculating model-based indices. Initial recommendations are to use 250 knots in the GOA as this results in a mesh with an adequate number of triangles spread throughout the area. More research is needed to determine an optimal number of knots, and whether or not this optimal number needs to be species dependent. Each knot is a vertex of the inner triangulated mesh, its placement determined from a k-nearest-neighbor algorithm on observation locations. For k-defined knots (eg. k=250), the algorithm searches for placement locations given the clusters of observations to optimally minimize distance between knots and observations.

A boundary region is drawn around the inner mesh using a nonconvex hull algorithm. A convex hull boundary can be conceptualized as a boundary defined using the outermost points of an area, much like using a rubber band around a group of pegs. The nonconvex hull algorithm is used to draw a smoother boundary than the convex hull around a set of points. While locations of outermost points still define the general shape, observations do not fall directly on boundary but are rather buffered at a specified distance; boundary lines are not drawn with hard angles. Larger triangles are drawn between the inner mesh and the outer boundary; this region is not used for prediction, and therefore, does not require as fine a scale resolution as the inner mesh.

Increasing the number of knots decreases the maximum triangle edge (distances between triangles), improving the approximation of the Matérn covariance, however, at a computational cost. The mesh can also be drawn by specifying an inner and outer boundary and setting a maximum triangle edge that defines the size of triangles within the mesh (Figure 3). Increasing the maximum triangle edge lowers the spatial resolution of the mesh, decreases the accuracy of the Matérn covariance approximation, and increases computational run times.

Weight matrices are defined in VAST that link the locations of the mesh to those of the observations as well as locations on the extrapolation grid. In this step, the spatial correlation value at the location of the

observation or grid cell center is calculated using the value of the nearest mesh vertex but weighted by the distance between the two.

The default settings of the VAST model for survey indices does not specify any temporal correlation structure. The model instead estimates a spatial effect and a spatio-temporal effect. The end result is a spatial random effect that is consistent across years (approximating the 'average' spatial correlation structure) and a spatio-temporal effect which specifies a unique spatial correlation structure at each time step (approximating temporally unique spatial structure). The average spatial and annual spatio-temporal effects are realizations from a spatial process unique to the species and governed by two parameters: kappa, which is the rate of spatial correlation decay with distance, and tau, which is the scaled inverse of the spatial variance. Default settings in VAST also include the specification of a unique vessel random effect which introduces extra variability attributed to vessel-level variation.

Default settings in VAST also include an estimation of anisotropy. In spatial statistics, the property of isotropy defines spatial correlation as a process that decreases with distance away from a location at the same rate regardless of direction (Figure 4a). Geometric anisotropy occurs when this pattern is stretched (Figure 4b) and/or rotated (Figure 4c) along an axis. Correlation follows an elliptical, rather than circular pattern of decay away from the origin. In VAST, specifying anisotropy adds an additional two parameters to estimate the rotation and scale on kappa (spatial correlation decay parameter). Anisotropy likely occurs in the spatial pattern of depth and current-driven environmental variables along the GOA coastline due to the narrow shelf. Spatial patterns in fish resulting from these spatial environmental effects, as seen in Dover, translate into anisotropic effects in species biomass. The anisotropy of the GOA bathymetry, however, is not caused by a simple stretching and rotation; it is also affected by a second-order curvilinear pattern (Figure 4d). This type of anisotropy is more challenging to model and is not currently implemented in VAST.

GOA bottom trawl survey biomass data for Dover sole were fit using VAST settings for the purpose of index standardization. Random effects in this model included a spatial effect and a spatio-temporal effect. A second set of models were fit where the observation model was changed from the default lognormal distribution to a gamma distribution.

The lognormal and gamma VAST models were each run using a series of knots from 50 to 1000 knots at two hundred knot intervals. In addition, three user-defined meshes were generated using maximum triangle edges specified at 100, 50, and 30 meters. Models were evaluated for convergence and validated using QQ-plots output from VAST. Model-based indices and their standard errors were compared to results from design-based index values and results from the RE model results used in the Dover sole stock assessment and the movement models presented in Appendix A.

New data were simulated from the model using estimated parameters. Two types of simulations were created. Unconditional simulations were generated by first simulating data from spatio-temporal models after which biomass data were generated from both the simulated random effects and model parameter estimates. Conditional simulations generated data from the observation model conditioned on estimated spatio-temporal and random effects. Conditional simulations, as they rely on random effects values estimated by the model, contain less variability than unconditional simulations where data is generated from both the random effects and observation models.

Simulations were used in an additional model validation step using a new approach detailed in the DHARMa R package (Hartig, 2019). The empirical cumulative distribution function (ECDF) is used to compare an observation in the sample to a series of simulations of the observation given the model. The ECDF is an empirical measure of the sample and therefore does not rely on a specific distribution. Using

this approach, the simulated data can be used to calculate quantile residuals which are uniformly distributed when observations are similar to simulations from the model. Uniform patterns, therefore, are evidence for model validity as the observed data are considered a reasonable realization from the model. The Kolmogorov-Smirnov (KS) Test, which measures the difference between two distributions, was performed on the scaled quantile residuals to evaluate the null hypothesis that the simulations and observations were drawn from the same data generating process. Quantile residuals from the conditional simulations were also used to perform a Moran's I test for additional spatial correlation after fitting a spatial model.

Results

Model Validation

QQ-plots from VAST output (Figure 5, Figure 6, Figure 7, and Figure 8) and results from a KS Test on the quantile residuals (Table 1) led to similar findings. The KS Test on unconditional residuals, calculated from simulations generated from both the random and observation model, rejected the validity of all models. Conditional residual, or those calculated when data were only simulated from the observation model, failed to reject the Gamma model based on a 250-knot mesh (Table 1).

QQ-plots from VAST output also supported a lack of fit in Lognormal models (Figure 5 and Figure 6) and Gamma models with either low or high spatial resolution in the mesh (Figure 7 and Figure 8). All model validation tests rejected models based on the set of user-defined mesh. Visual interpretation of QQ-plots suggested the Lognormal models were underdispersed, that is, there was less variability in the data than what was estimated by the model. Similarly, Gamma models with a large number of knots (i.e. greater than 250) were also underdispersed. In contrast, Gamma models with a small number of knots (i.e. less than 250) were overdispersed, that is, there was greater variability in the data than what was estimated by the model.

Residual Spatial Correlation Results

The Moran's I test of quantile conditional residuals demonstrated that residuals from models based on a low resolution mesh still retained spatial correlation. The spatial correlation retained in residuals declined as the spatial resolution of the mesh increased (Table 2, Table 3).

Comparison of indices

Lognormal Models

VAST estimates from Lognormal model runs were either comparable or greater than stratified/RE model results (Figure 9, Figure 10). Differences between these VAST estimates and the design-based and RE-based estimates decreased as the number of knots increased (Figure 13, Figure 14). The decrease in differences leveled off as more knots were added, yet stabilized VAST estimates were still greater than the design-based and RE-based estimates for a majority of the time series. While survey years that only sampled 0-500m coincided with greater differences between VAST and non-VAST approaches (1990-1993, 2001, 2009), greater differences between approaches also occurred when surveys extended depths to 1000m (2005, 2007).

Gamma Models

VAST estimates from Gamma model runs resulted in model-based estimates with fewer differences between VAST indices and stratified/RE model indices than for the Lognormal model runs (Figure 11

and Figure 12). Differences decreased as the number of knots increased (Figure 15 and Figure 16). The decrease in differences leveled off as more knots were added, and unlike the Lognormal model runs, stabilized at zero difference around 750 knots. Excluding years 1990-1999, 2001, 2011, 2015, and 2017, results between VAST and non-VAST estimates were within model standard errors. There was no clear pattern between years with big vs small differences between VAST and non-VAST approaches and years with high vs. low survey coverage.

Index comparisons between VAST and design-based estimates

All VAST models consistently estimated a higher biomass in survey years 1993 and 1996 compared with non-VAST methods (Figure 9-Figure 12). Surveys extended out to 500m during these years, therefore, biomass estimates of Dover sole in the deep were more informed by the spatial and spatio-temporal models than by survey observations. As the spatial resolution of the models increased (i.e. VAST knots increased), the biomass indices decreased. Models using the 1000-knot mesh, however, still resulted in estimates significantly higher than non-VAST estimates. The RE model indices were also higher than those from design-based methods during this time frame. In contrast, other survey years with coverage out to 500m did not result in large differences between VAST and non-VAST methods. These results support a higher trend in Dover sole biomass during this time period than what was estimated using the design-based method. The VAST approach may be better suited to approximate the additional biomass than the RE model, given its ability to project spatial distributions in the deep strata from the 'average' spatial effect aggregated over time.

In 2001, the Eastern GOA was not surveyed. Therefore, VAST estimates were only informed by the annual mean estimated biomass from observations in the Western and Central GOA, the 'average' spatial effect from the Eastern GOA aggregated over years in which this region was surveyed, and the spatio-temporal effect extrapolated from observation in Western and Central GOA. The RE model estimate was informed by Eastern GOA data in 1999 and 2003 as well as data from Central GOA. The design-based index from 2001 is not considered to be reliable and therefore excluded from stock assessments. As expected, both the VAST and RE model estimates were significantly higher than zero. Unlike 1993 and 1996, the RE model was estimated higher than the VAST estimate. Due to issues stated below, the VAST model may not be well approximating spatial correlation in the Eastern GOA, and therefore, may not be producing as reliable an estimate as the RE model.

Index comparison by area

The 250-knot VAST model was used to make comparisons to area apportioned design-based/RE model results (Figure 17 and Figure 18). Little to no differences between VAST and stratified/RE model indices were detected in the Western GOA. VAST indices consistently led to higher estimates than non-VAST indices in the Central GOA while a range of magnitude in differences between VAST and non-VAST estimates occurred in the Eastern GOA. The differences between estimates from different methodologies occurred in the Eastern GOA. While largest differences between VAST and non-VAST methods occurred during survey years with 500m survey coverage, this trend was not consistent (e.g. 2009) and significant differences also occurred in 700m and 1000m coverage years. this A comparison of proportion of the GOA area surveyed by area, depth, and year (Figure 19) failed to suggest any clear pattern between survey coverage and differences in approaches between VAST estimates and stratified/RE model estimates.

Discussion

The spatial and spatio-temporal model within the VAST model was likely mis-specified. Unconditional simulations generated from the both the random effects and observation models failed to generate data similar to observations. This level of misspecification is not uncommon in spatio-temporal models as the random effects tend to soak up variability better explained by other processes, such as missing covariates. In other words, when a species is correlated with a spatial or spatio-temporal covariate, a model designed to estimate the relationship with the covariate, a fixed effect, would better explain the data. It is not possible, however, to model or measure all possible covariates associated with species density distributions. One of the benefits of spatio-temporal models is in their ability to explain extra sources of variability due to missing covariates. Aim should be taken, however, to identify significant covariates that are readily available and include them in spatio-temporal models whenever possible. In the case of Dover sole, the inclusion of depth would likely improve model performance when generating unconditional simulations.

Additional mis-specification in the spatial model may be occurring due to issues in estimating anisotropy. Parameter estimation of anisotropy (Figure 20 and Figure 21) in the Dover sole VAST model was likely being driven by the north-eastern trending patterns in the Central/Western GOA rather than northern-western trending patterns that would be expected from the Eastern GOA. This is apparent in Figure 18 where differences between VAST and design-based estimates are highest in the Eastern GOA. Additionally, biomass density patterns show western-eastern trending 'strips' in the Eastern GOA which contrast the expectation for a more natural north-western pattern (Figure 28 and Figure 29). This striping pattern is most apparent in the low resolution 50-knot mesh. Mis-specification of anisotropy is known to lead to decreased accuracy in predictions (Yongtao Guan, 2004), and therefore, could be driving differences between VAST and non-VAST estimates in this region.

In general, models with the best performance, where performance was evaluated by comparing model fits to observations, had lower resolution than models with better predictive performance (ie. that is, resulted in biomass indices closer to what was estimated using non-VAST methods). Higher spatial resolution in the mesh better approximates the spatial structure throughout the predictive region. Consequently, models with higher spatial resolution tend to have better predictive performance. Notably, biomass densities from 50-knot mesh models oversimplified density patterns throughout the GOA (Figure 28 and Figure 29). As knots increased, the spatial pattern of density became more refined. The higher resolution models better approximated the absence of Dover sole, suggesting that an over-estimation of absence is a source in the discrepancies seen between VAST and non-VAST methods. Additional evidence of the relationship between spatial resolution and predictive performance can be seen in model test results. With increased spatial resolution, residuals were less likely to retain spatial correlation (Table 2 and Table 3) and differences between VAST and non-VAST survey indices decreased (Figure 13 and Figure 15).

Models with better fits to observations resulted in larger spatial ranges (Figure 20 and Figure 21), resulting in larger spatial and spatio-temporal offsets from the annual mean. Extrapolating these larger spatial and spatio-temporal effects to unobserved locations, however, resulted in larger differences between VAST and non-VAST biomass estimates. In contrast, higher spatial resolution models were estimating smaller spatial ranges resulting in smaller local offsets to the annual mean biomass. Inspection of QQplots suggested these models were likely not explaining enough variability present in the data (Table 1, Figure 5, and Figure 7), and therefore, were likely underestimating the spatial range of the true correlation structure.

The Dover sole spatial range was sensitive to both the observation model and the spatial resolution of the mesh (Figure 20 and Figure 21). The spatial range of both the positive catch rates and encounter probability extended from about 50-500m. Given a spatial range of 50m, observations further apart than this distance are independent, and spatial correlation between survey hauls may be difficult to detect. As noted above, a more realistic spatial range of the Dover sole is likely to be larger. This species, however, may still be particularly sensitive to changes in mesh structure and may require finer mesh resolution than other species in the GOA. Additionally, Dover is found in deep water and can therefore occur on the boundary of observations. Spatial effects along the boundary can be biased high when mesh boundaries do not extend far enough past the region of observations, resulting in biased high estimates of annual biomass.

Future Research Directions

Overall, a VAST model of the Dover sole has the potential to improve annual biomass estimates in the GOA. VAST models presented here in this appendix have been shown to detect higher abundance than design-based methods in years when missing depth-area gaps were present. Further work is needed, however, to identify a model that optimizes both performance in estimation and prediction. This author recommends exploring improvements to both the spatial and observation models.

VAST models with finer spatial resolution than presented in this appendix should be tested. User-specified meshes presented here did not result in a high enough spatial resolution to explain all of the spatial correlation structure in the data (Table 2 and Table 3). At higher resolutions, however, a user-specified mesh with larger boundaries may better approximate spatio-temporal correlation patterns and ameliorate any potential boundary effects (Figure 22, Figure 23, and Figure 24). It is recommended to rerun user-specific mesh models with inner maximum edges defined around 20 and a larger boundary area than what was presented in this appendix. This recommendation aligns with best practices that have been defined by statistical developers of the mesh spatial modeling methods (Bakka, 2018).

Current spatial-temporal modeling approaches cannot account for the 2nd order level of anisotropy (Figure 4d) that may be affecting results in the GOA. Research on this topic is ongoing, yet a definitive solution has yet to be determined. In the meantime, one model exploration step could estimate the Eastern GOA in a separate spatial model than the Western GOA and Central GOA to determine the degree to which misspecification of anisotropy in the Eastern GOA is occurring. Such sensitivity analyses could benefit other species as all GOA VAST models would be similarly affected if anisotropy is being mis-specified. The inclusion of significant covariates, such as depth, could minimize anisotropic effects. If anisotropy is being largely driven by the narrow, curved shape of the shelf, and Dover sole is correlated with depth, this anisotropic effect could be mollified by including depth as a covariate.

Additional recommendations by this author include exploring the underlying observation model and investigating the inclusion/exclusion of spatio-temporal effects and vessel random effects in both the probability of encounter and positive biomass components of the expectation.

Literature Cited

Bakka, H. (2018). How to solve the stochastic partial differential equation that gives a Matern random field using the finite element method. *arXiv preprint arXiv:1803.03765*.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society, Series B*, 192-236.

Hartig, F. (2019). DHARMa: Residual Diagnostics for Hierarchical. R package version 0.2.4.

Lindgren, F., Rue, H., & Lindstrom, E. (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*, 423-498.

Methot, R., & Wetzel, C. (2013). Stock Synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, 142:86-99.

Methot, R., Wetzel, C., & Taylor, I. (2019). *Stock Synthesis User Manual Version 3.30.13.* Seattle, WA: NOAA Fisheries.

Thorson, J. T. (2017). Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. *Canadian Journal of Fisheries and Aquatic Sciences*, 1369-1382.

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*, 210:143-161.

Thorson, J. T., Shelton, A. O., Ward, E. J., & Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science*, 1-14.

Yongtao Guan, M. S. (2004). A Nonparametric Test for Spatial Isotropy using Subsampling. *Journal of the American Statistical Association*, 99: 810-821.

Table 1. Model validation p-values from a Kolmogorov-Smirnov Test on the uniformity of simulation residuals from conditional simulations. Results listed for residuals from Lognormal versus Gamma observation models. Fits are evaluated at incrementally increasing spatial knots. A p-value less than 0.05 indicates strong evidence against the null hypothesis that the observed and simulated data are derived from the same distribution.

Knots	Max Edge	Lognormal	Gamma
50		0.013	0.006
250		0.01	0.348
450		0.003	0.039
650		0	0.003
850		0	0
1000		0	0
	100	0	0.036
	50	0	0.021
	30	0	0

Table 2. P-values from Moran's I test on residual correlation in **Lognormal** observation model residuals by year and spatial mesh complexity. Quantile residuals were calculated from conditional simulations. As VAST knots increase, spatial resolution of the mesh increases. When using a user-defined mesh, spatial resolution increases as the maximum triangle edge decreases. P-values in bold red indicate evidence against the null of no spatial correlation. Years from models that failed the Moran's I test all indicated residual spatial correlation was positive.

Year	VAST Knots					
	50	250	450	650	850	1000
1984	**0**	**0**	**0**	**0.001**	**0.005**	**0.017**
1987	**0**	0.25	0.158	0.335	0.596	0.504
1990	**0.011**	0.983	0.925	0.559	0.589	0.876
1993	**0.001**	0.395	0.686	0.39	0.992	0.889
1996	**0**	**0.004**	0.171	**0.041**	0.229	0.34
1999	**0**	0.103	0.705	0.585	0.796	0.119
2001	**0**	**0.013**	0.083	**0.041**	0.05	0.173
2003	**0**	0.054	0.167	0.627	0.594	0.55
2005	**0**	**0.007**	0.163	0.705	0.861	0.945
2007	**0**	0.778	0.622	0.906	0.642	0.407
2009	**0**	**0.04**	0.118	0.588	0.142	0.988
2011	**0.008**	0.086	0.352	0.599	0.339	0.736
2013	**0.012**	0.134	0.186	0.096	0.167	**0.026**
2015	**0**	**0.004**	**0.001**	**0.002**	**0.006**	0.053
2017	**0.001**	0.874	0.37	0.896	0.92	0.633
2019	**0**	0.051	0.082	0.577	0.218	0.172
	User Defined Mesh - max edge					
	100	50	30			
1984	**0**	**0**	**0**			
1987	0.511	0.784	0.409			
1990	0.332	0.852	0.539			
1993	0.831	0.147	0.888			
1996	0.09	**0.015**	0.073			
1999	0.69	0.136	0.219			
2001	0.108	0.23	0.164			
2003	0.882	0.146	0.502			
2005	0.877	**0.041**	0.933			
2007	0.849	0.814	0.813			
2009	0.708	0.055	0.638			
2011	0.298	0.8	0.811			
2013	0.117	0.122	0.057			
2015	0.245	**0.005**	**0.028**			
2017	0.22	0.772	0.977			
2019	**0.006**	0.298	0.12			

Table 3. P-values from Moran's I test on residual correlation in **Gamma** observation model residuals by year and spatial mesh complexity. Quantile residuals were calculated from conditional simulations. As VAST knots increase, spatial resolution of the mesh increases. When using a user-defined mesh, spatial resolution increases as the maximum triangle edge decreases. P-values in bold red indicate strong evidence against the null of no spatial correlation. Years from models that failed the Moran's I test all indicated residual spatial correlation was positive.

Year	VAST Knots					
	50	250	450	650	850	1000
1984	**0**	**0**	**0.003**	**0**	**0.001**	**0.015**
1987	**0**	**0.038**	0.324	0.71	0.903	0.331
1990	**0.007**	0.713	0.735	0.929	0.959	0.755
1993	**0.001**	0.238	0.736	0.874	0.92	0.543
1996	**0**	**0**	**0.022**	**0.025**	0.068	**0.031**
1999	**0**	**0.015**	0.24	0.796	0.689	0.494
2001	**0**	**0.044**	**0.033**	0.206	0.086	0.221
2003	**0**	**0.013**	0.114	0.356	0.261	0.635
2005	**0**	**0.003**	0.233	0.797	0.878	0.667
2007	**0**	0.805	0.905	0.975	0.924	0.831
2009	**0**	**0.01**	0.068	0.921	0.817	0.712
2011	**0.002**	**0.006**	0.458	0.485	0.96	0.236
2013	**0.001**	0.1	0.236	0.32	0.106	0.18
2015	**0**	**0**	**0.005**	**0.009**	0.292	0.123
2017	**0.002**	0.675	0.718	0.867	0.92	0.969
2019	**0.013**	**0.029**	0.126	0.307	0.065	0.6
	User Defined Mesh - max edge					
	100	50	30			
1984	**0**	**0**	**0.017**			
1987	0.125	0.292	0.44			
1990	**0.016**	0.175	0.362			
1993	**0.001**	0.446	0.573			
1996	**0.002**	**0.009**	**0.006**			
1999	**0.001**	0.086	0.555			
2001	**0.001**	**0.004**	0.149			
2003	**0.006**	**0.018**	0.298			
2005	**0**	**0.016**	0.418			
2007	**0.01**	0.392	0.659			
2009	**0.002**	**0.011**	0.744			
2011	**0.071**	**0.035**	0.155			
2013	**0.026**	**0.049**	**0.026**			
2015	**0**	0.08	0.188			
2017	**0.071**	0.517	0.694			
2019	**0.02**	0.152	**0.031**			

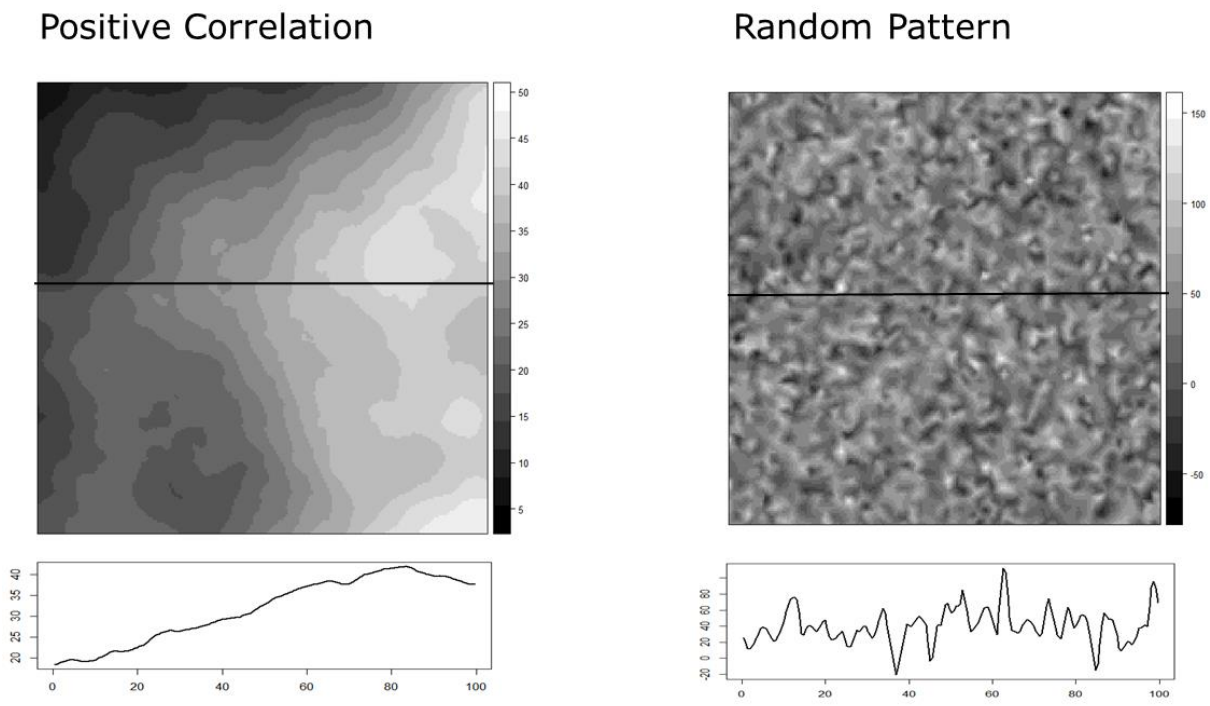

Figure 1. Left: Positive spatial correlation. Cross section plot on bottom shows smooth change in correlation value along the x-gradient. Observations sampled at close proximity would not be independent. Right: Random spatial pattern. Cross section plot on bottom shows no correlation pattern along the x-gradient. Observations are guaranteed to be independent, regardless of distance.

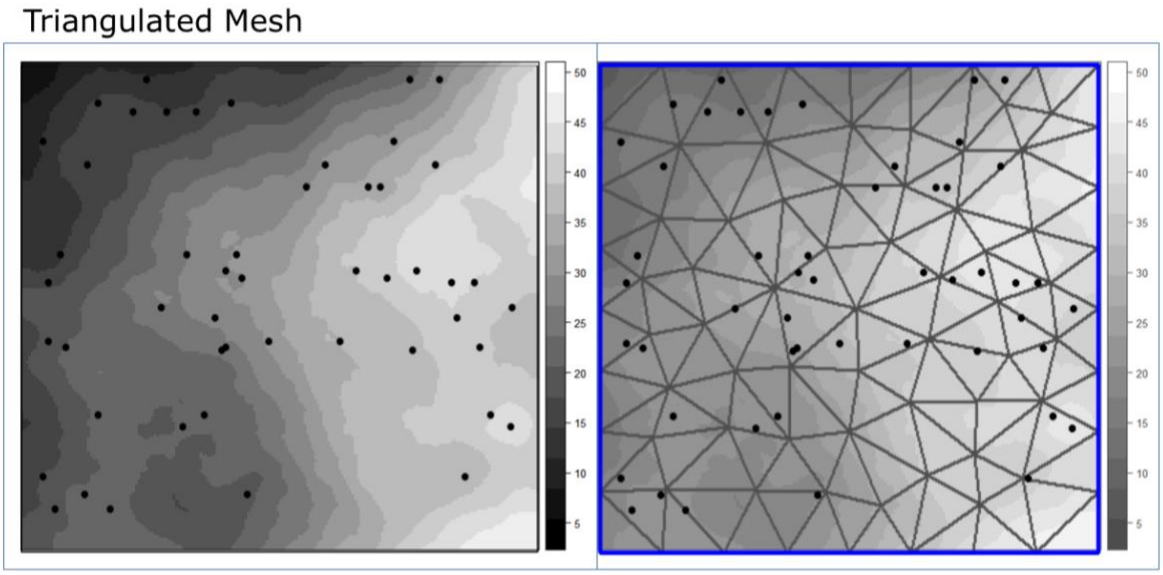

Figure 2. Left: Sample locations across the surface with spatial correlation. Right: Triangulated "Mesh" converts the continuous spatial surface to areas of discrete triangles. Spatial correlation is estimated along the triangle edges. Correlation at observation locations are weighted according to the distance between the observation and the nearest triangle vertex. Gaussian Markov properties hold such that correlations between two observations are only non-zero if they are within two triangle edges of each other.

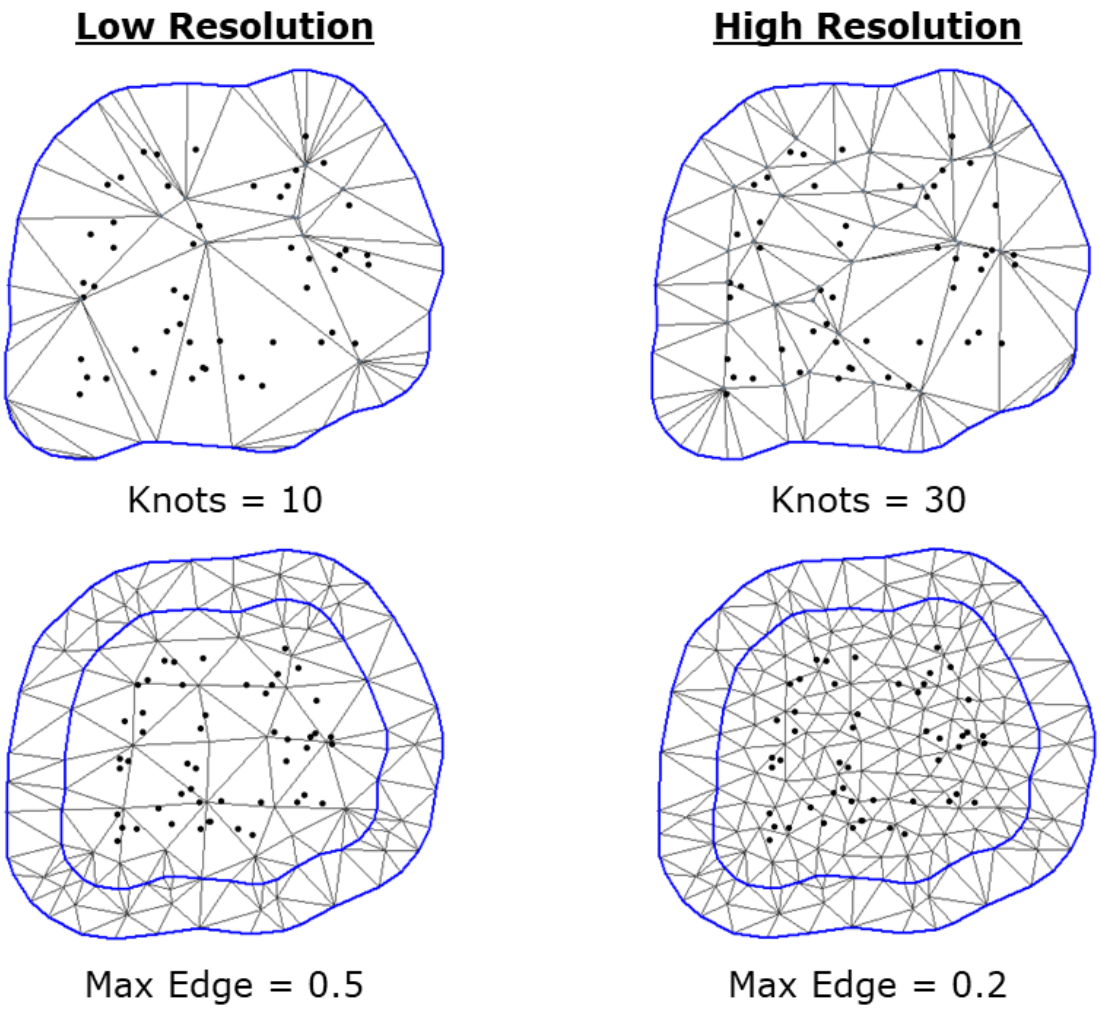

Figure 3. Examples of different Mesh structures for a single set of observations. Resolution increases from left to right. Boundaries were drawn using convex hulls. Top: Meshes were generated using knots similar to VAST. The number of knots are placed within the data region and triangles drawn between the knots and from the knots to the outer edge. Bottom: Meshes were generated using a maximum edge, that is a distance used to define the maximum edge of any triangle length. The number of knots are not specified.

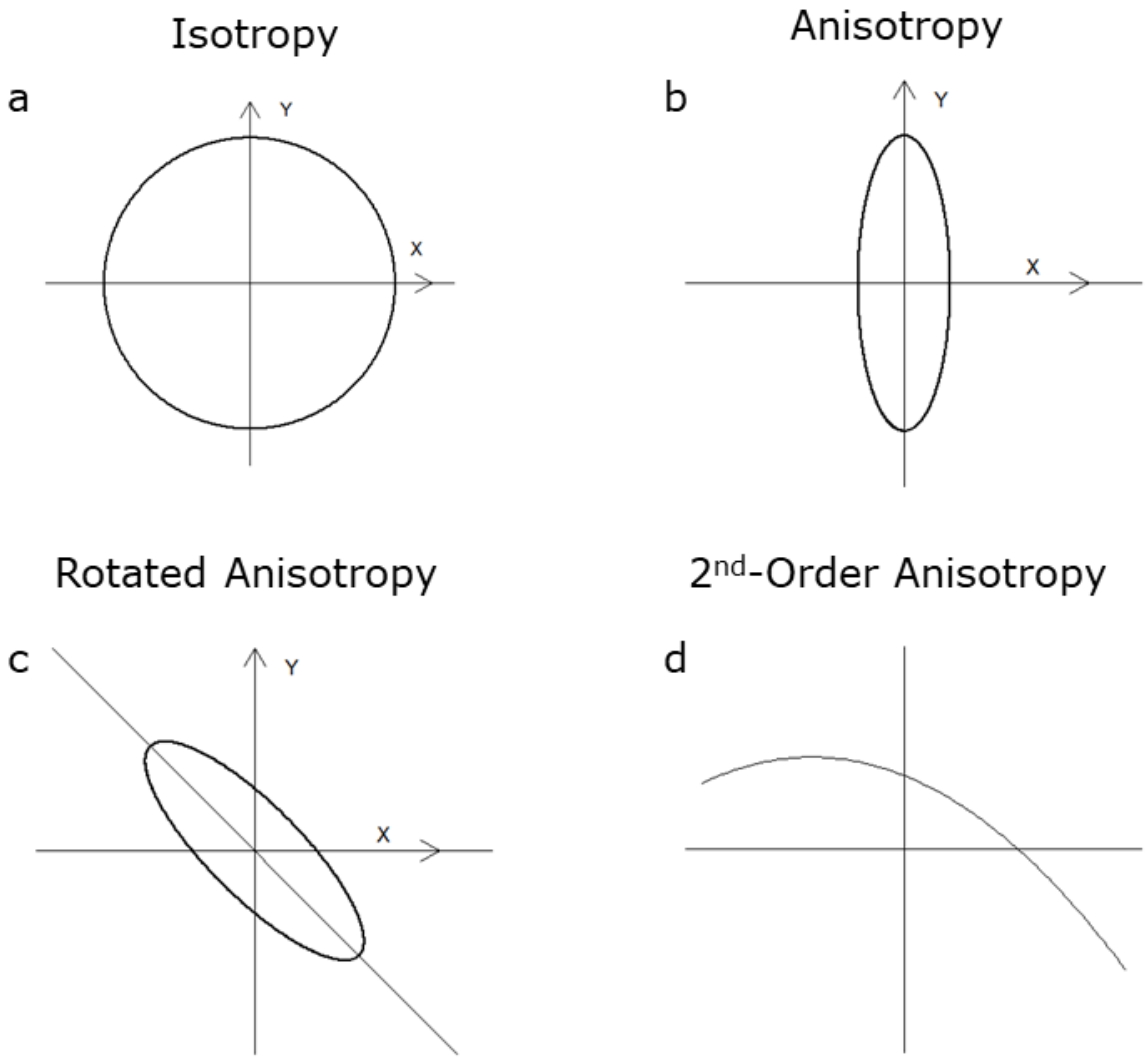

Figure 4. Isotropy (a) is a stationary property of spatial correlation where correlation decays at the same rate with distance regardless of direction. Geometric anisotropy occurs when correlation decay is stronger over one axes and weaker over a perpendicular one (b) and when this elliptical pattern of decay is rotated (c). Second-order anisotropy (d) occurs when there is a curvilinear pattern to the direction of stronger decay; this type of pattern is not well approximated given current spatial modeling methods.

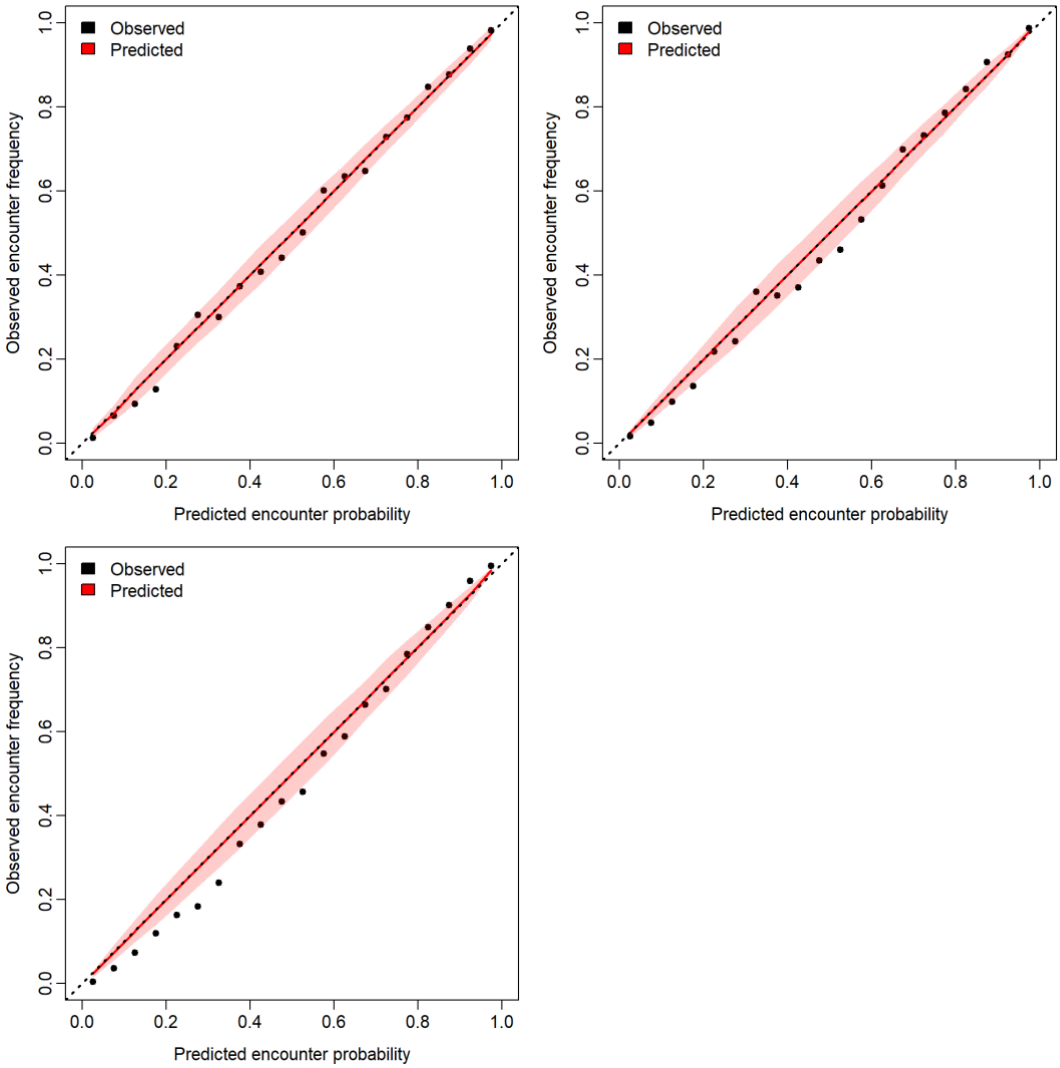

Figure 5. Diagnostics comparing observed versus predicted probability of encounter from **Lognormal** VAST models. From left to right: 50-knot Mesh, 250-Knot Mesh, 1000-knot Mesh

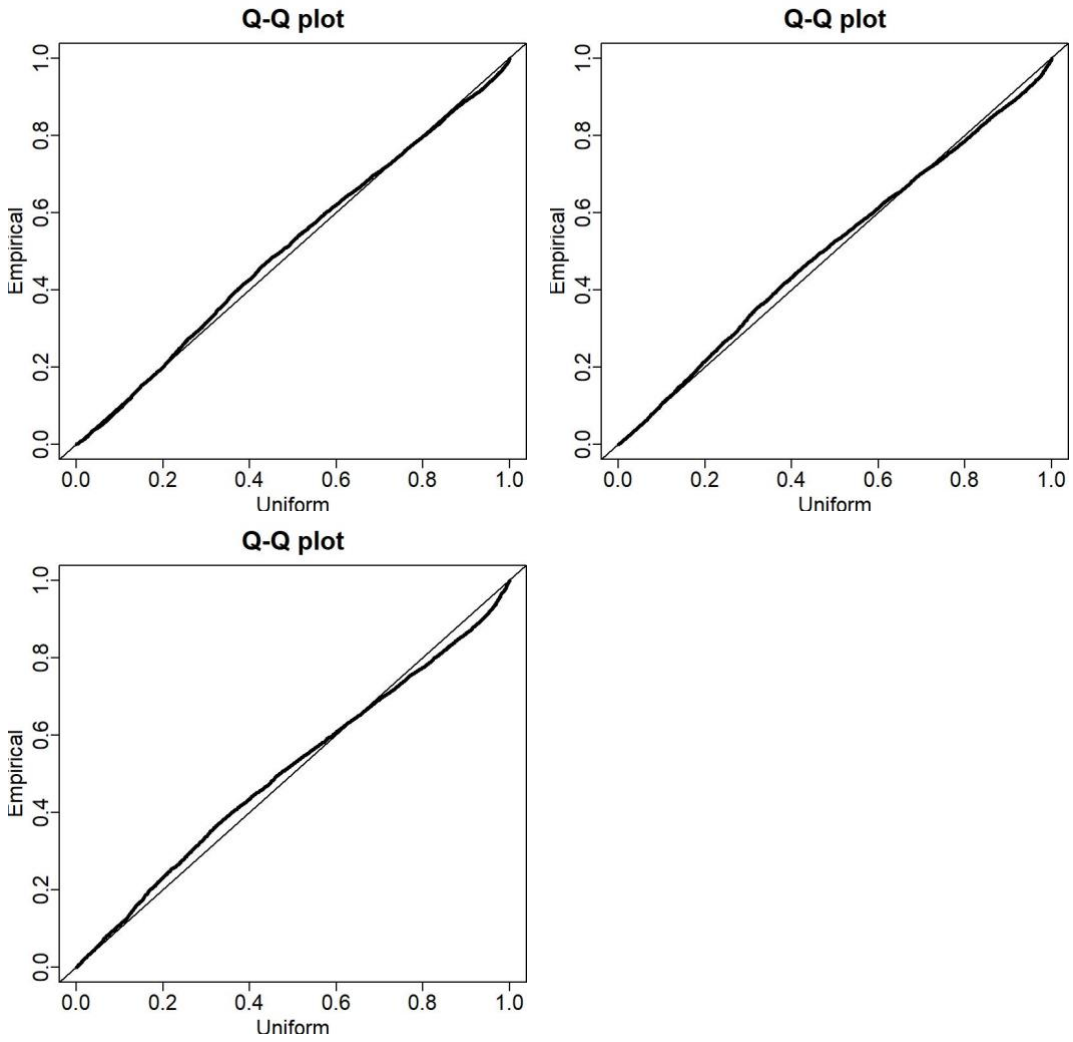

Figure 6. . Diagnostics comparing observed versus predicted probability of encounter from **Lognormal** VAST models. From left to right: 50-knot Mesh, 250-Knot Mesh, 1000-knot Mesh

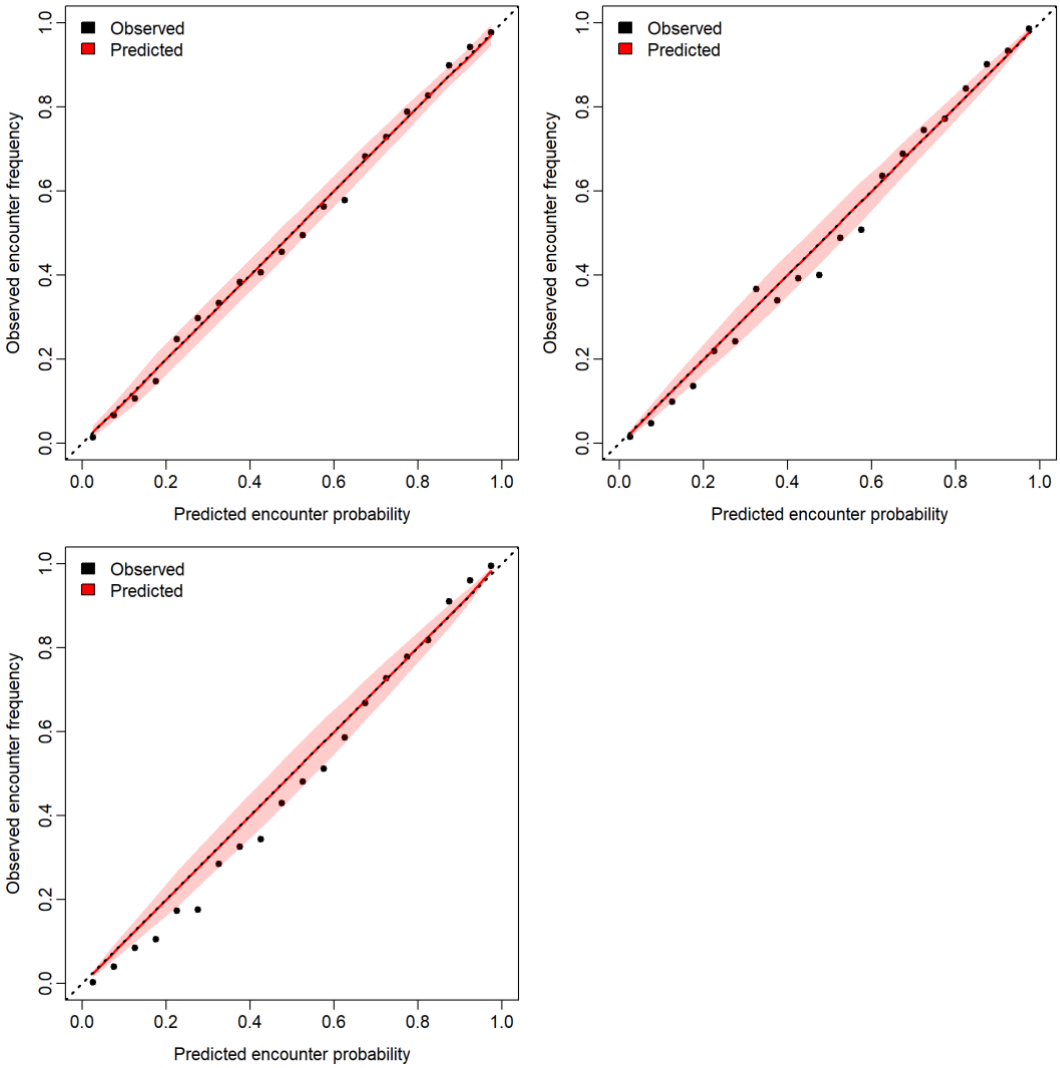

Figure 7. Diagnostics comparing observed versus predicted probability of encounter from **Gamma** VAST models. From left to right: 50-knot Mesh, 250-Knot Mesh, 1000-knot Mesh

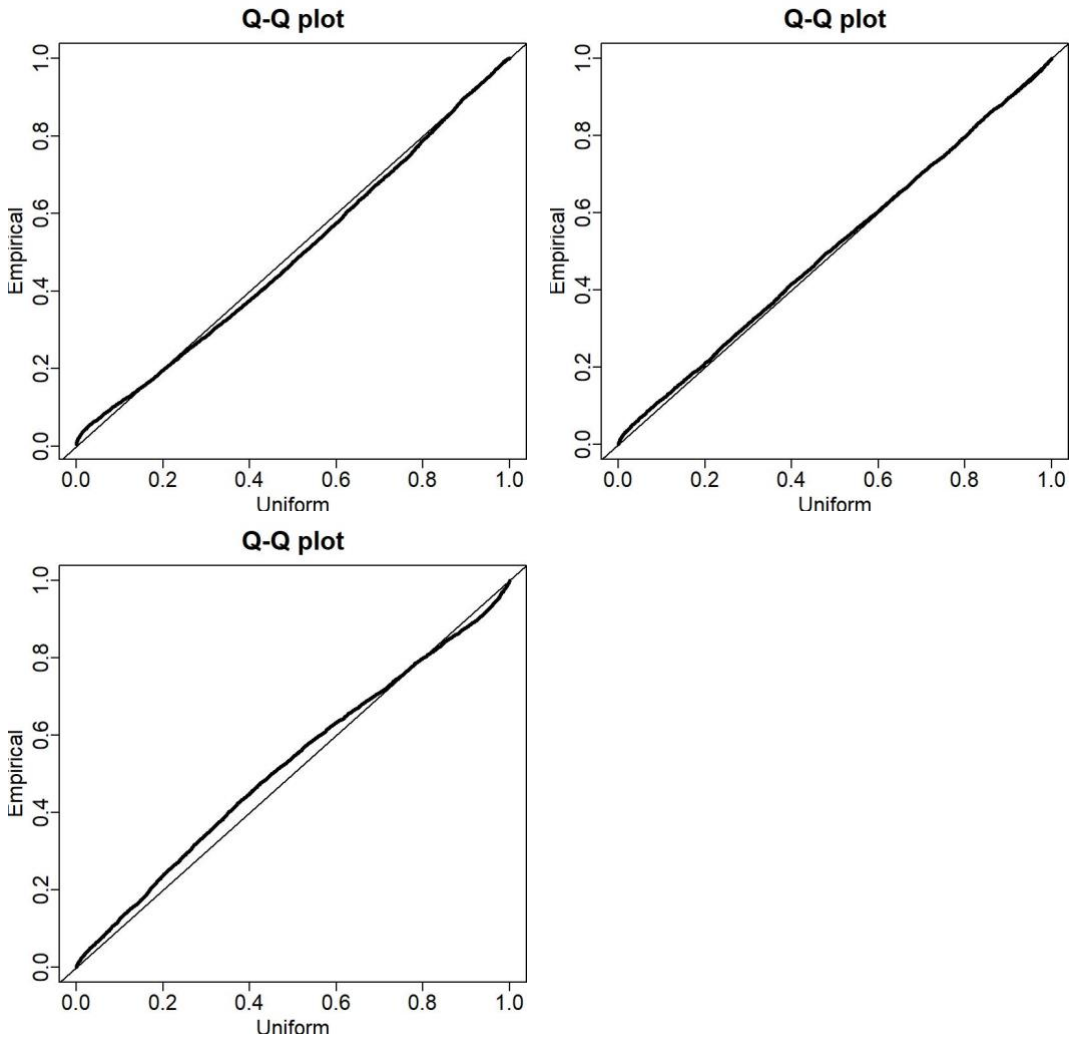

Figure 8. . Diagnostics comparing observed versus predicted probability of encounter from **Gamma** VAST models. From left to right: 50-knot Mesh, 250-Knot Mesh, 1000-knot Mesh

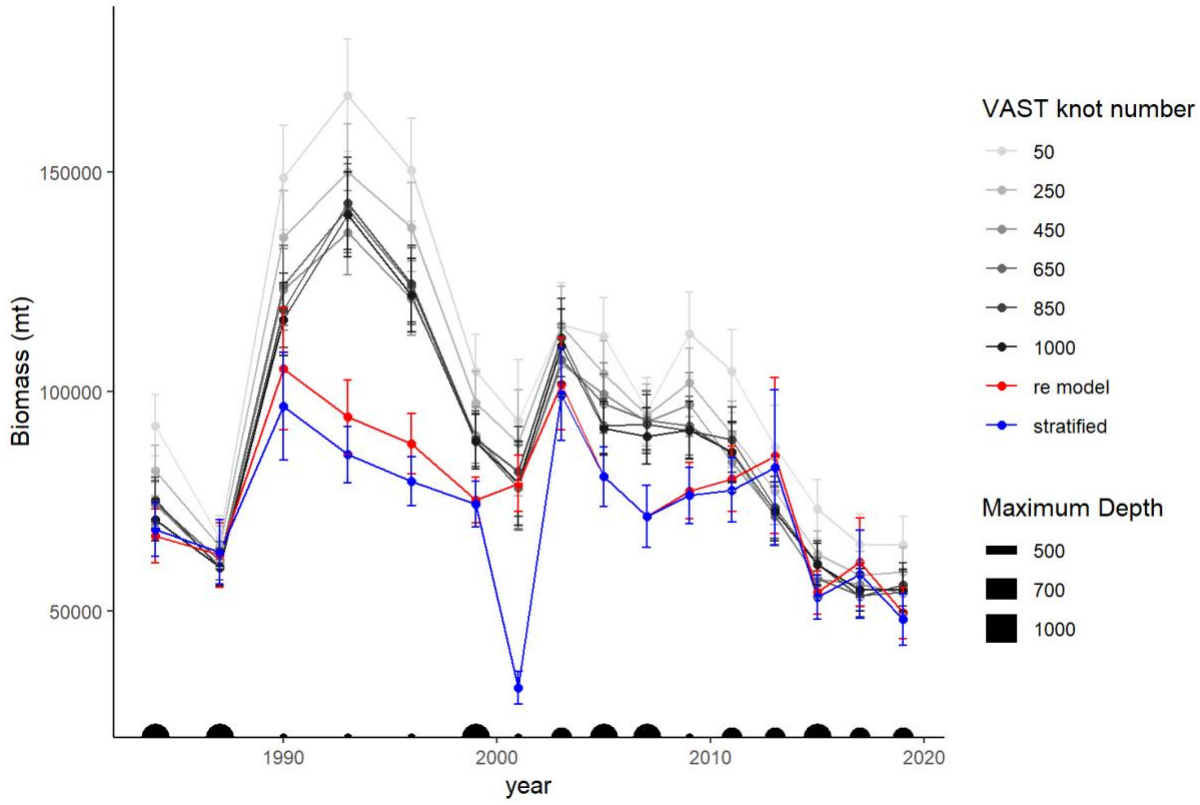

Figure 9. Biomass estimates from the **Lognormal** VAST model under different knot scenarios (colors light grey-black) compared with index results from the random effects model which filled in area-depth gaps (red) and original stratified estimates (blue). Years which surveyed out to 500m (small), 700m (medium), and 1000m (large) are displayed along the x-axis at the bottom of the plot. The survey did not sample the Eastern GOA in 2001.

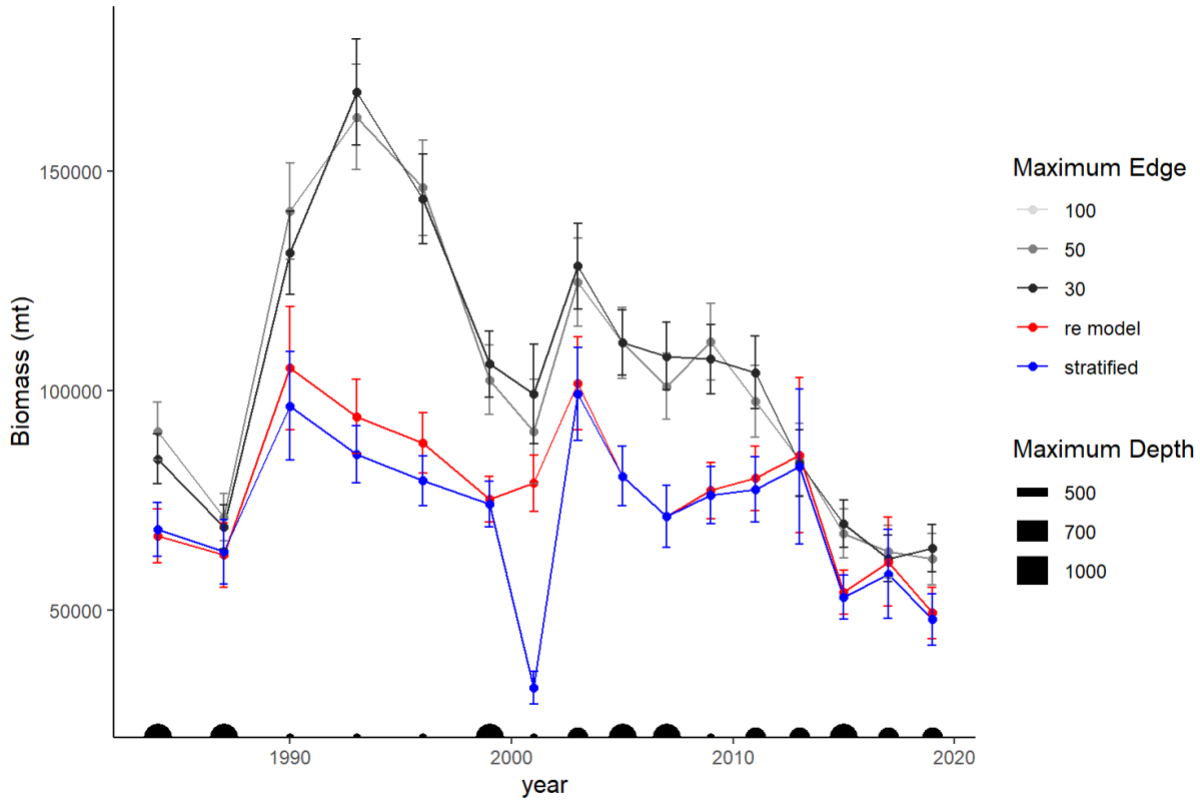

Figure 10. Biomass estimates from the **Lognormal** VAST model under different maximum edge scenarios from user-defined meshes (colors light grey-black) compared with index results from the random effects model which filled in area-depth gaps (red) and original stratified estimates (blue). Years which surveyed out to 500m (small), 700m (medium), and 1000m (large) are displayed along the x-axis at the bottom of the plot. The survey did not sample the Eastern GOA in 2001.

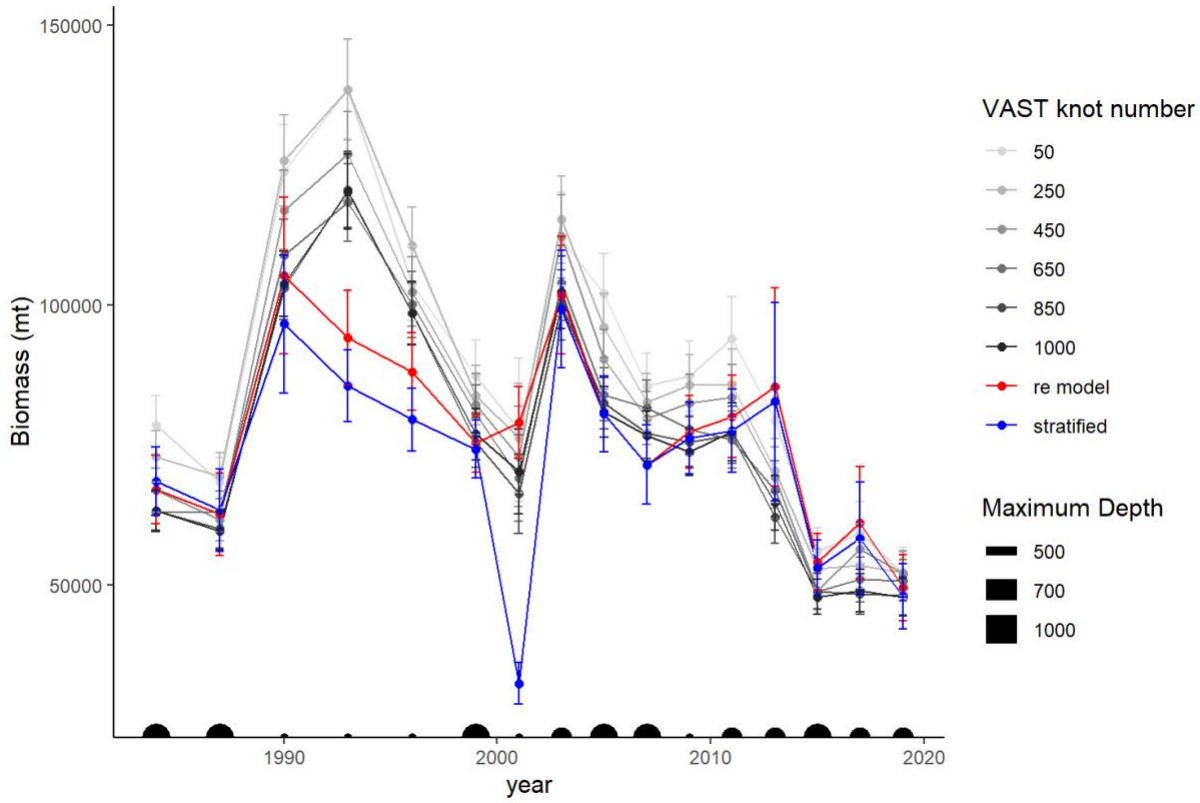

Figure 11. Biomass estimates from the **Gamma** VAST model under different knot scenarios (colors light grey-black) compared with index results from the random effects model which filled in area-depth gaps (red) and original stratified estimates (blue). Years which surveyed out to 500m (small), 700m (medium), and 1000m (large) are displayed along the x-axis at the bottom of the plot. The survey did not sample the Eastern GOA in 2001.

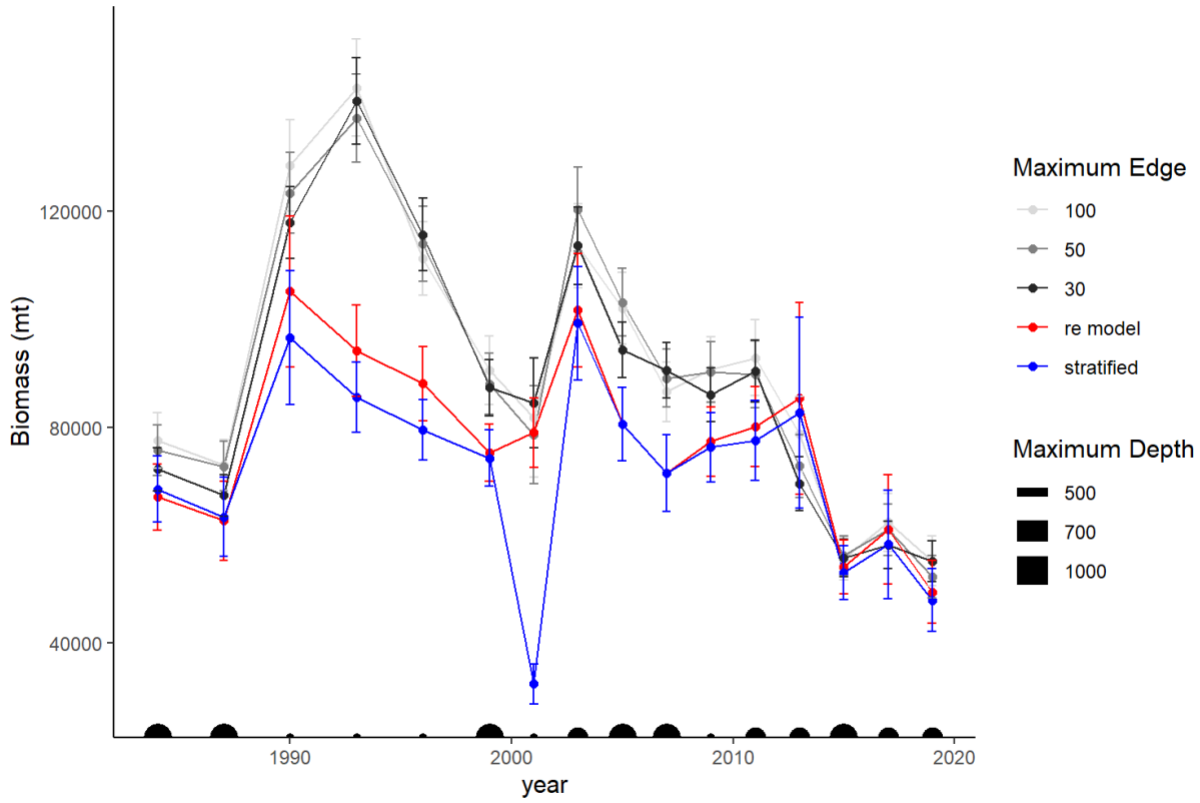

Figure 12. Biomass estimates from the **Gamma** VAST model under different maximum edge scenarios from user-defined meshes (colors light grey-black) compared with index results from the random effects model which filled in area-depth gaps (red) and original stratified estimates (blue). Years which surveyed out to 500m (small), 700m (medium), and 1000m (large) are displayed along the x-axis at the bottom of the plot. The survey did not sample the Eastern GOA in 2001.

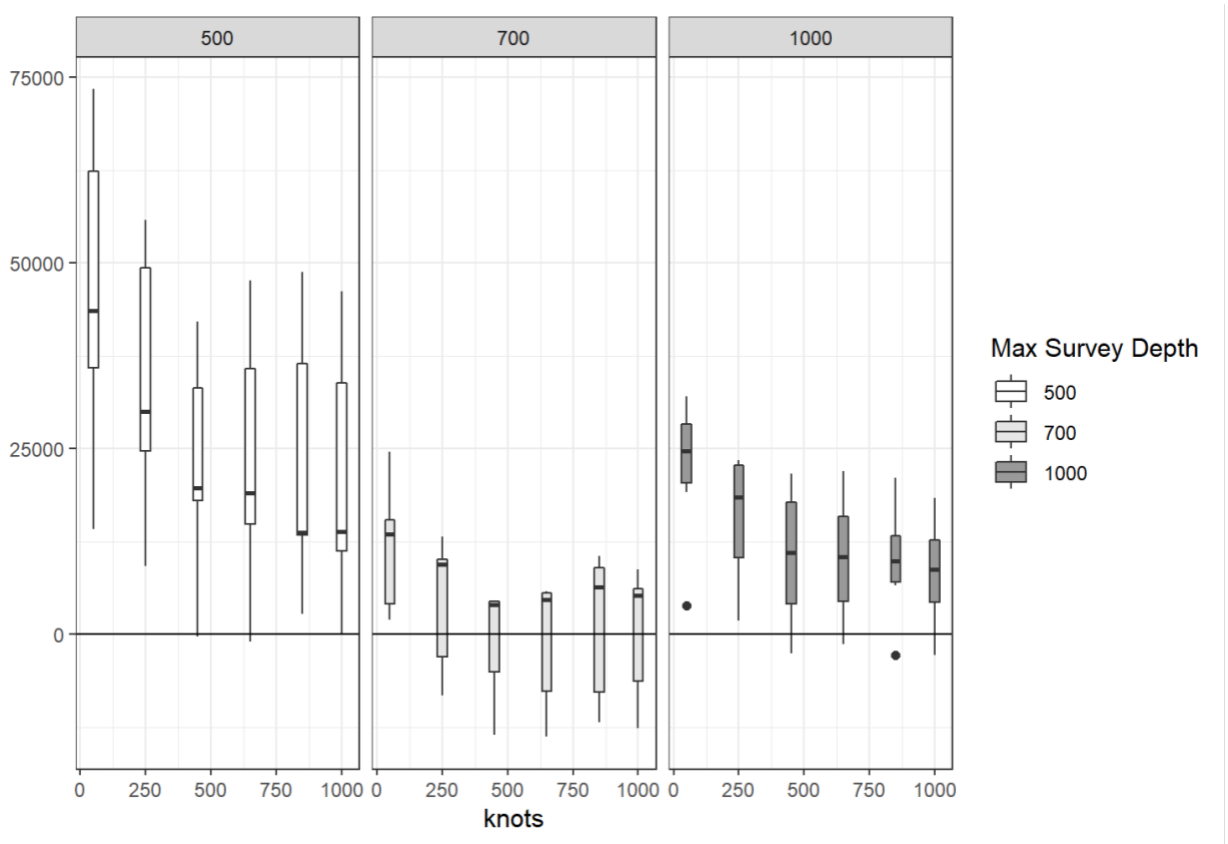

Figure 13. Differences between **Lognormal** VAST biomass estimates and the random effect model estimates (VAST – RE) compared against the number of knots used to construct the triangulated grid for spatial correlation estimation.

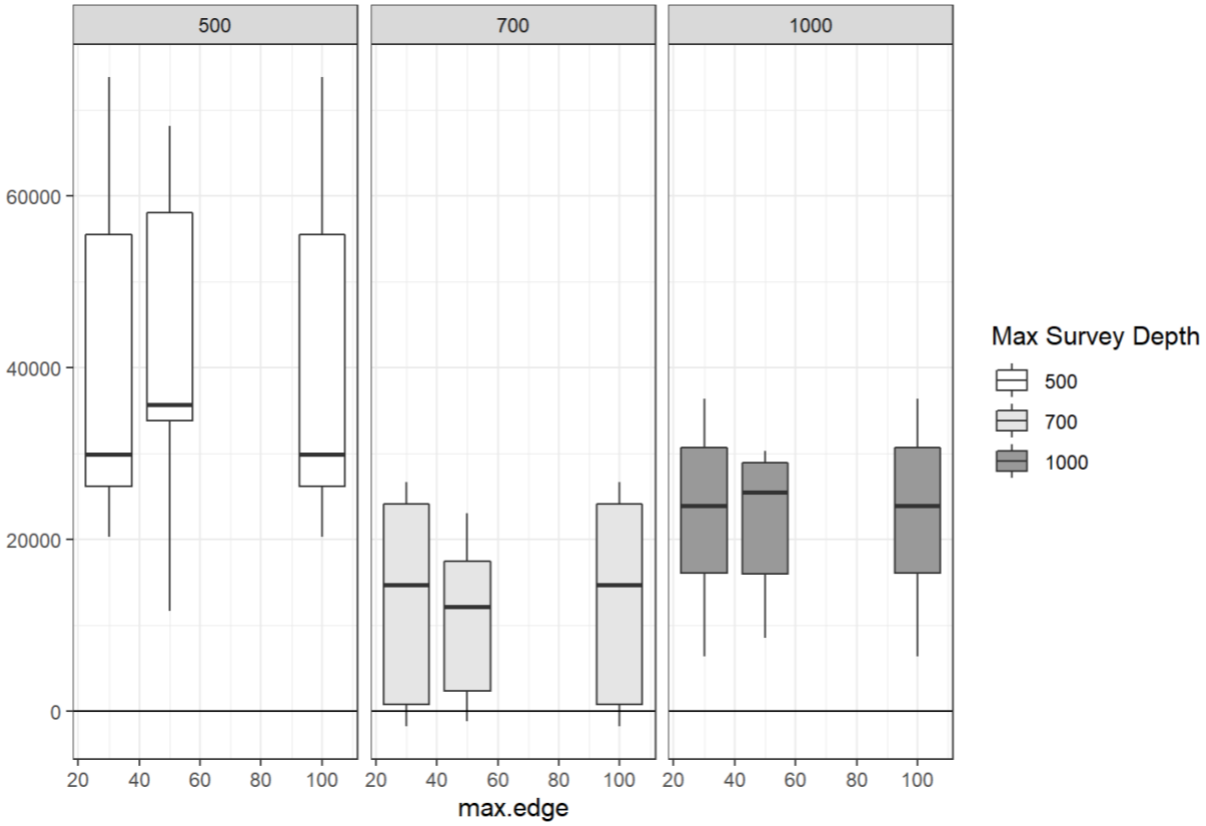

Figure 14. Differences between **Lognormal** VAST biomass estimates and the random effect model estimates (VAST-RE) compared against the maximum edge used to construct the used-defined triangulated grid for spatial correlation estimation.

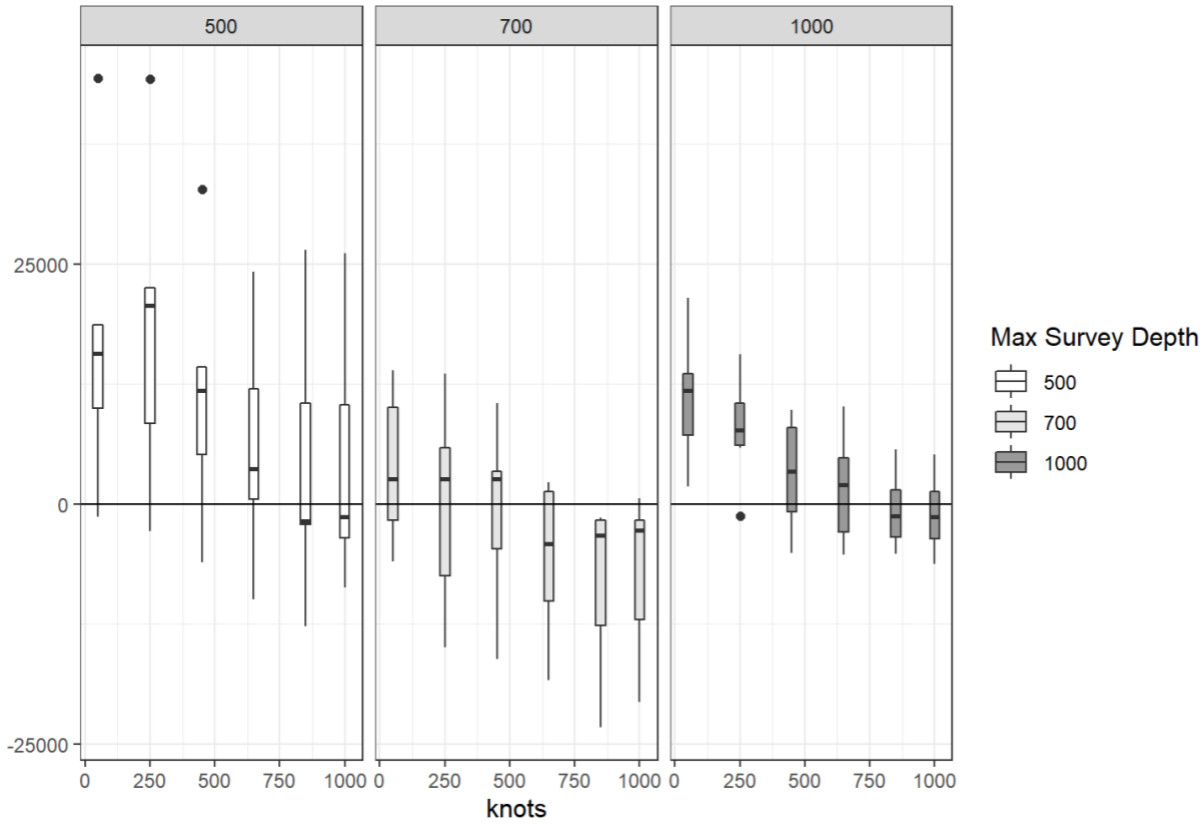

Figure 15. Differences between **Gamma** VAST biomass estimates and the random effect model estimates (VAST-RE) compared against the number of knots used to construct the triangulated grid for spatial correlation estimation.

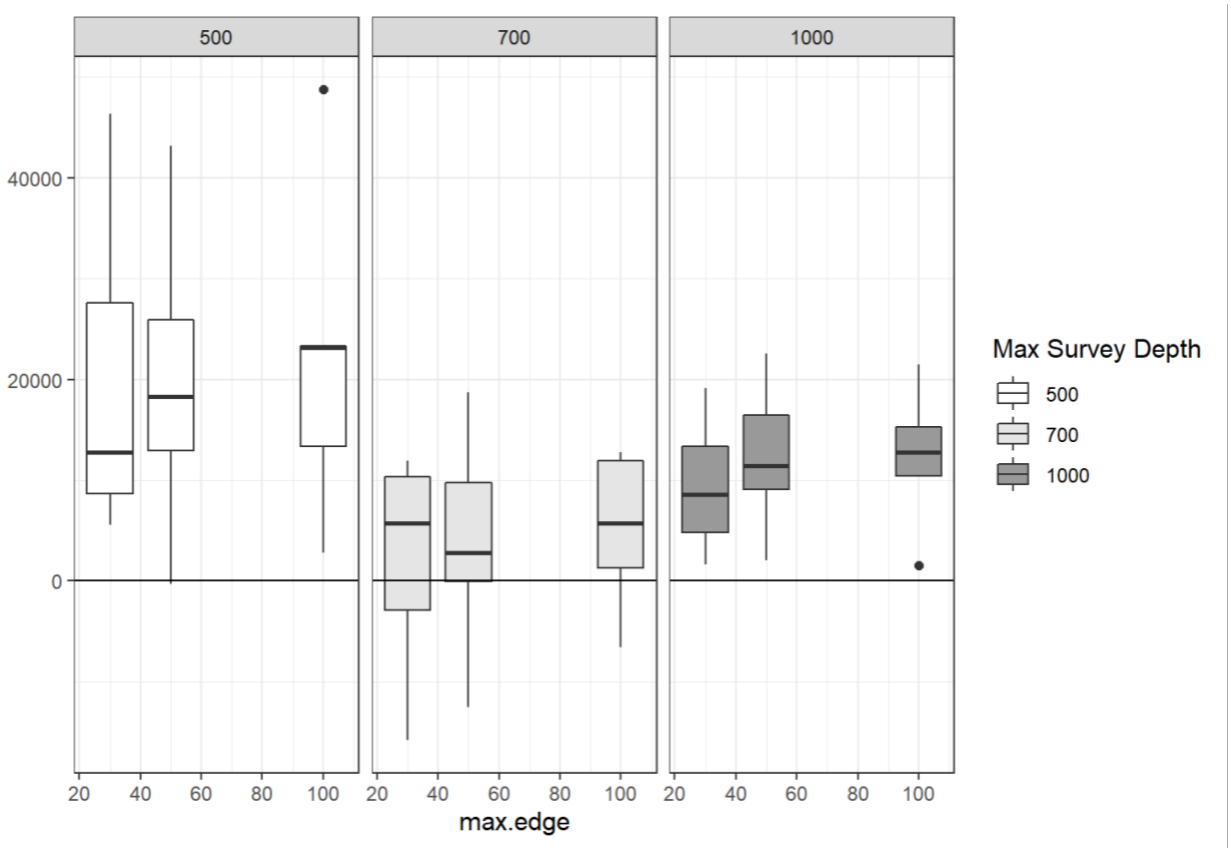

Figure 16. Differences between **Gamma** VAST biomass estimates and the random effect model estimates (VAST-RE) compared against the maximum edge used to construct the used-defined triangulated grid for spatial correlation estimation.

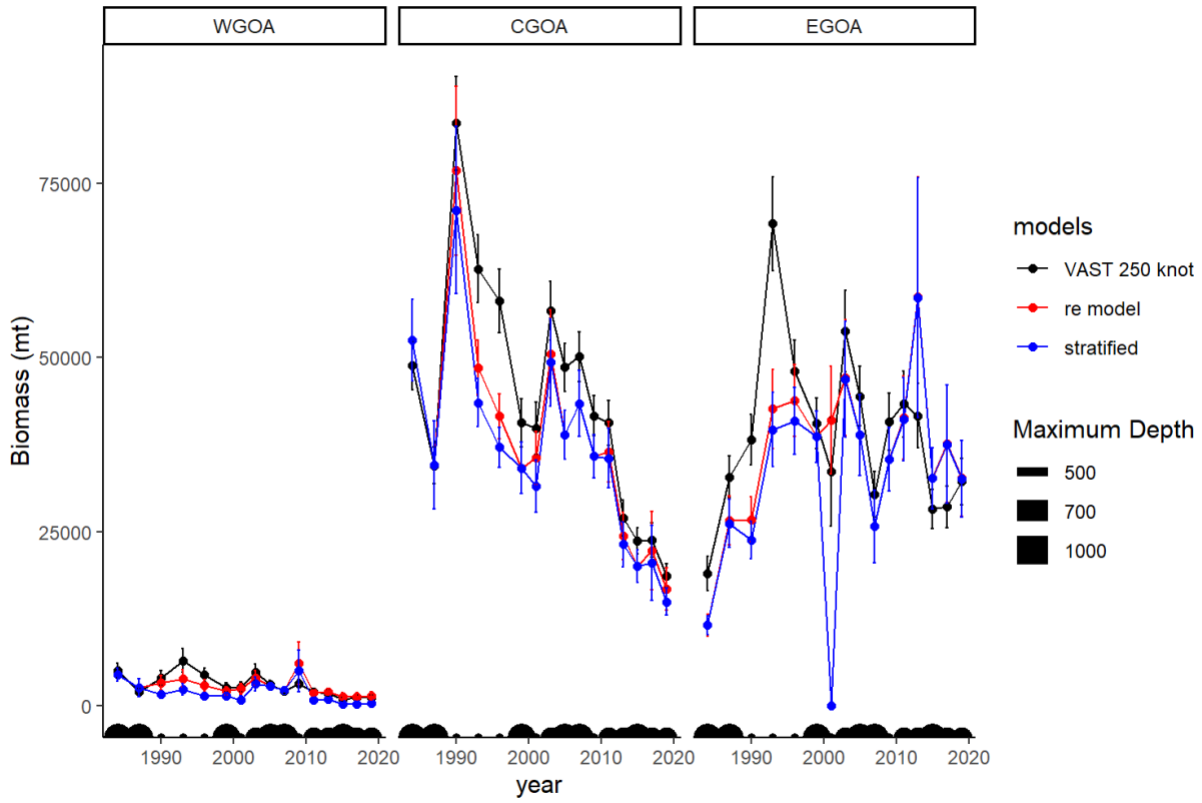

Figure 17. Estimates from the **Gamma** 250 knot Mesh VAST model (black) compared with index results from the random effects model which filled in area-depth gaps (red) and original stratified estimates (blue). Years which surveyed out to 500m (small), 700m (medium), and 1000m (large) are displayed along the x-axis at the bottom of the plot. Data is partitioned by NMFS Areas: Western GOA, Central GOA, and Eastern GOA

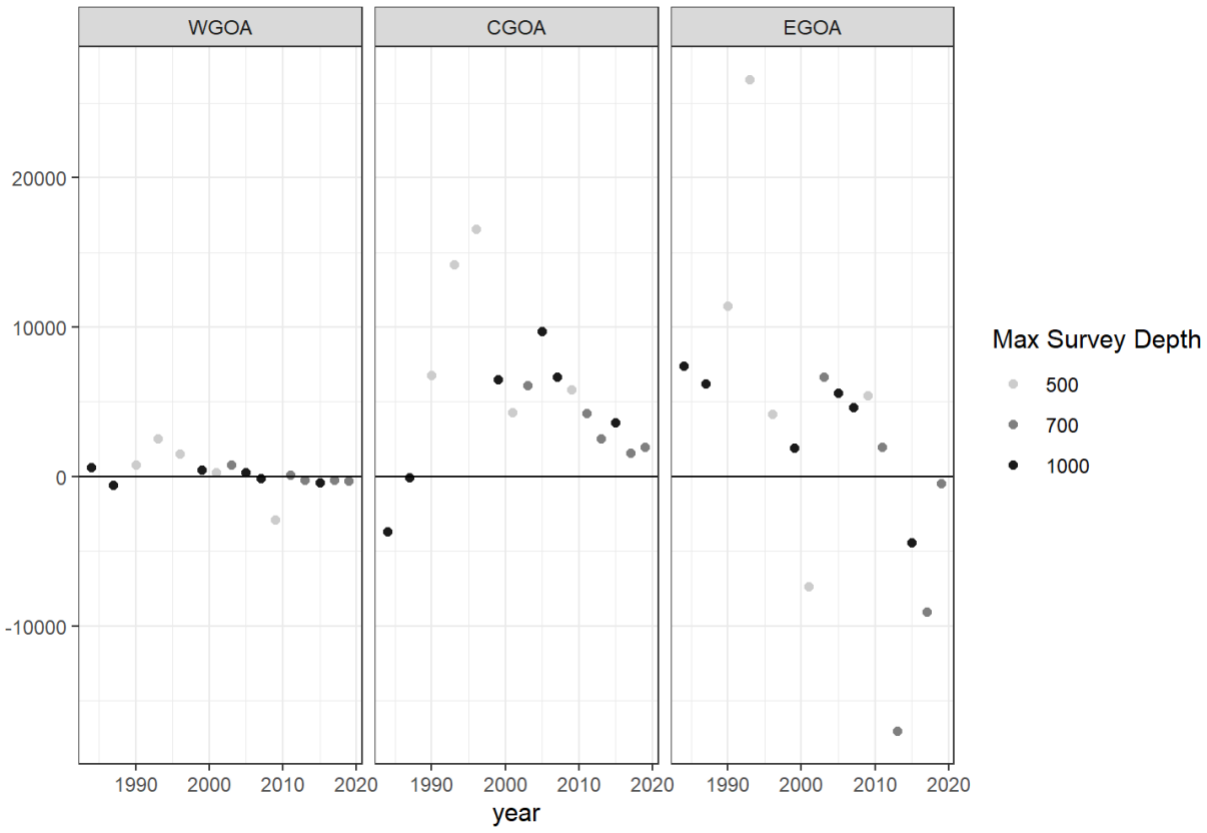

Figure 18. Differences between the Gamma 250-knot Mesh Vast model and the random effect model estimates (VAST-RE). Bias is partitioned by NMFS Areas: Western GOA, Central GOA, and Eastern GOA.

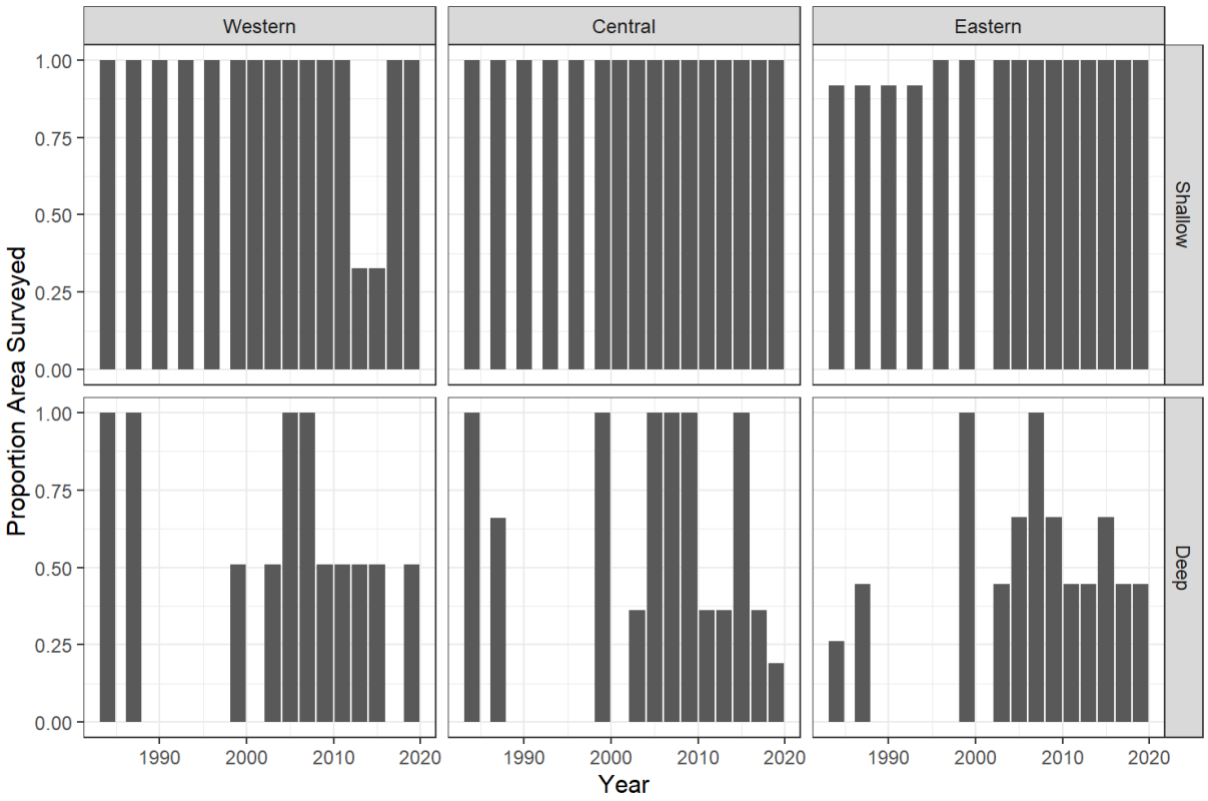

Figure 19. Proportions of areas surveyed partitioned by NMFS area (Western, Central ,Eastern) and shallow (0-500m) versus deep (501-1000m) strata. Proportions were calculated by comparing GOA strata areas from strata surveyed to total area for each NMFS area – depth combination.

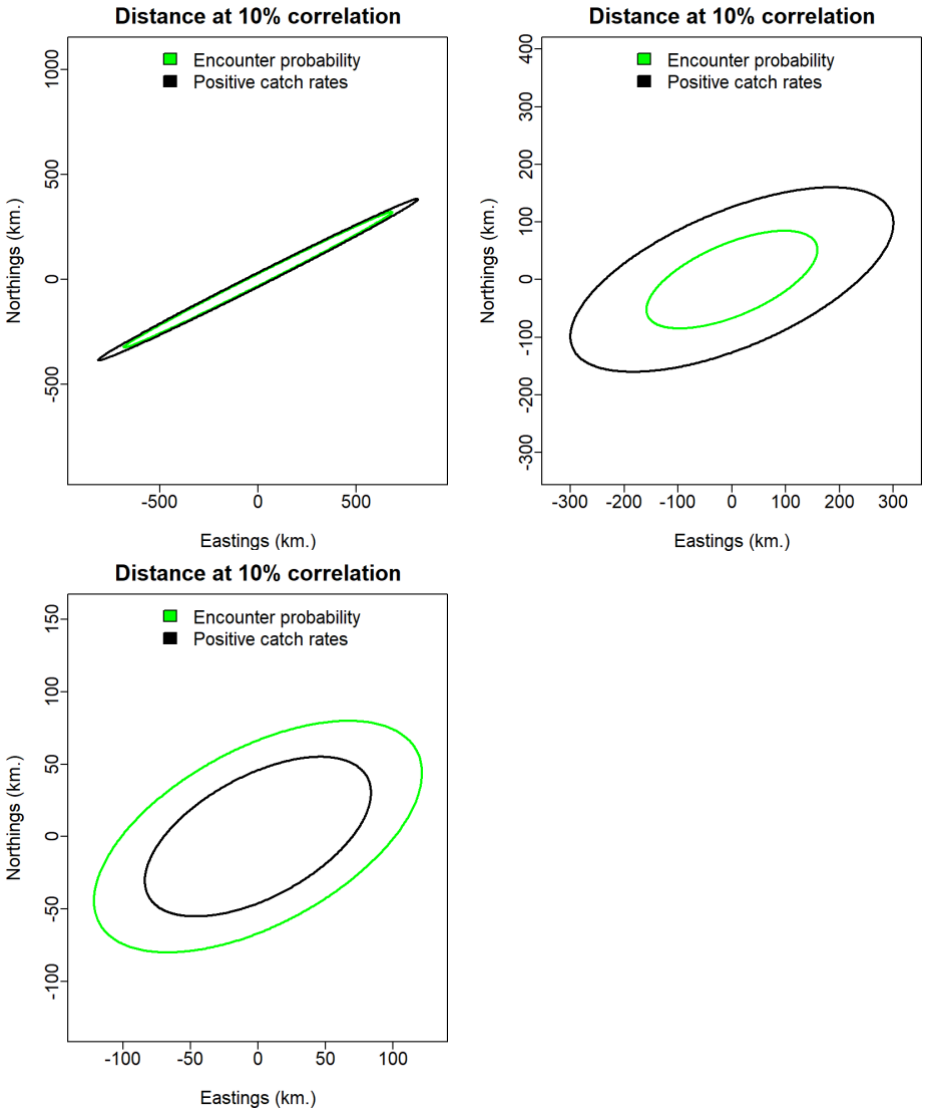

Figure 20. Anisotropy results from **Lognormal** models. Top 50-knots (left), 250-knots (right); Bottom: 1000-knots

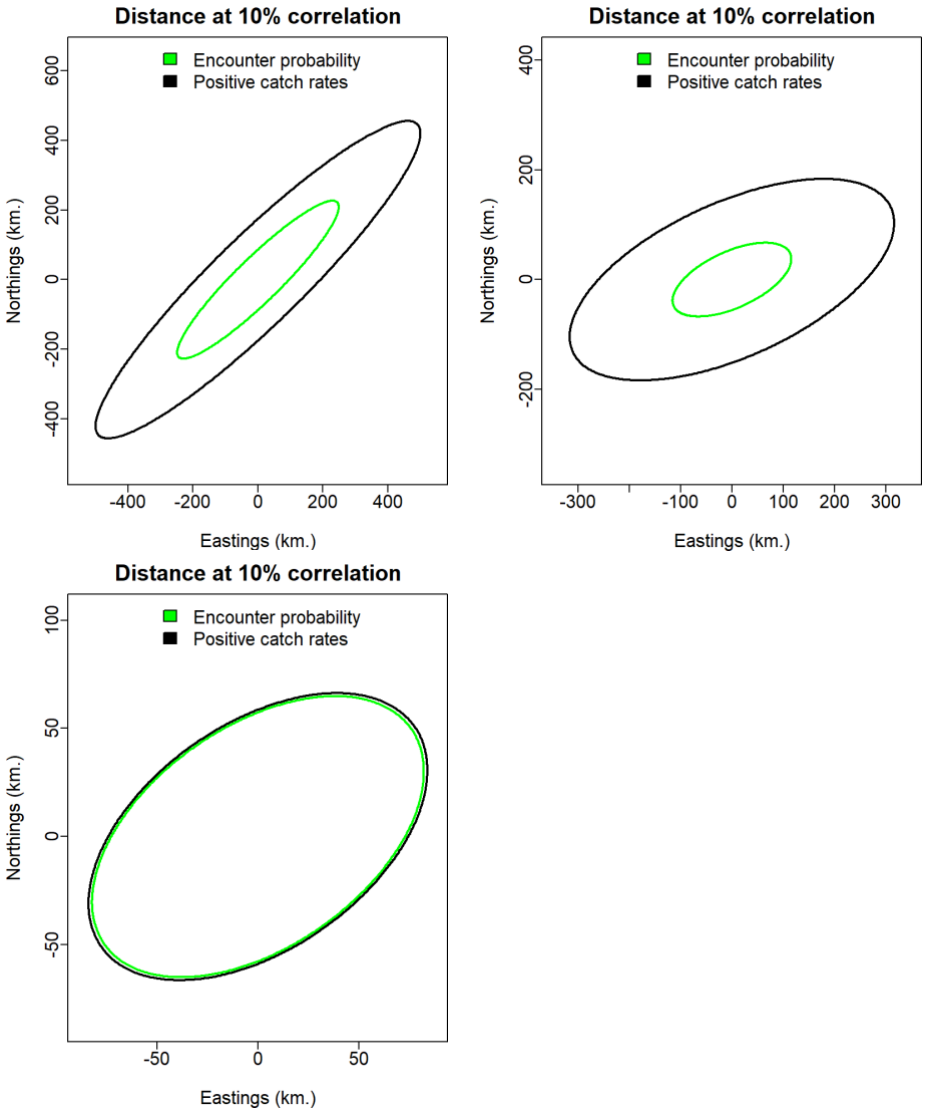

Figure 21. Anisotropy results from **Gamma** models. Top: 50-knots (left), 250-knots (right); Bottom: 1000-knots.

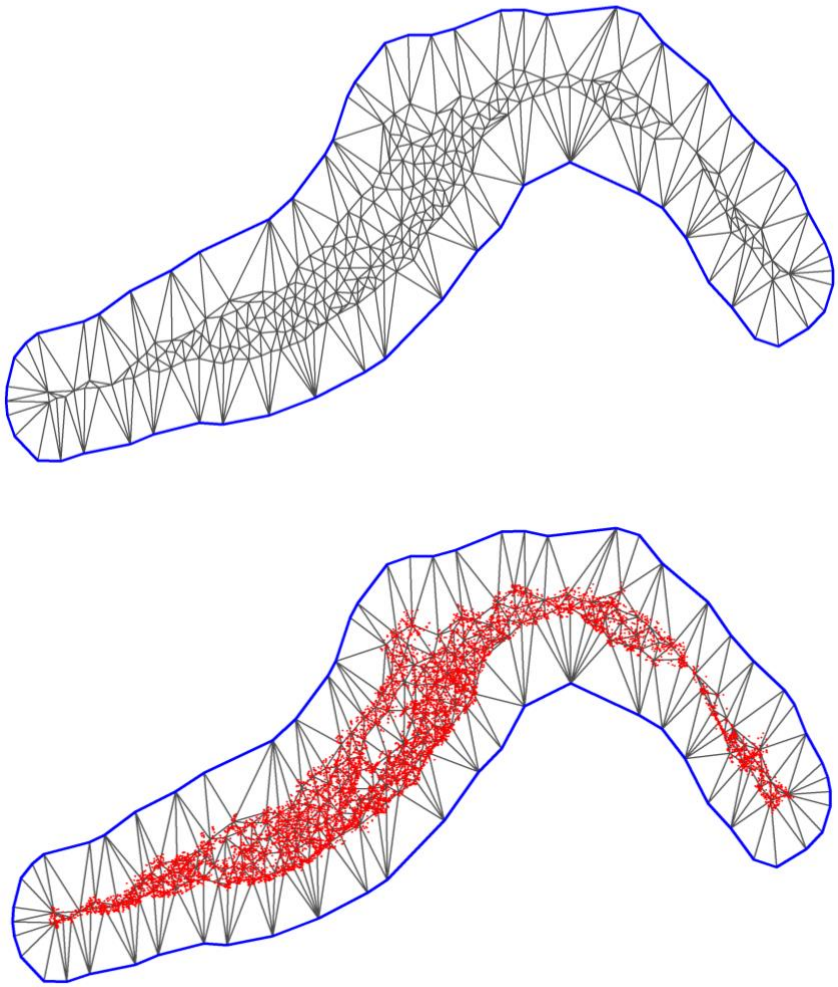

Figure 22. Left: Mesh generated with 250 knots. Right: 250-knot Mesh with survey haul locations in red.

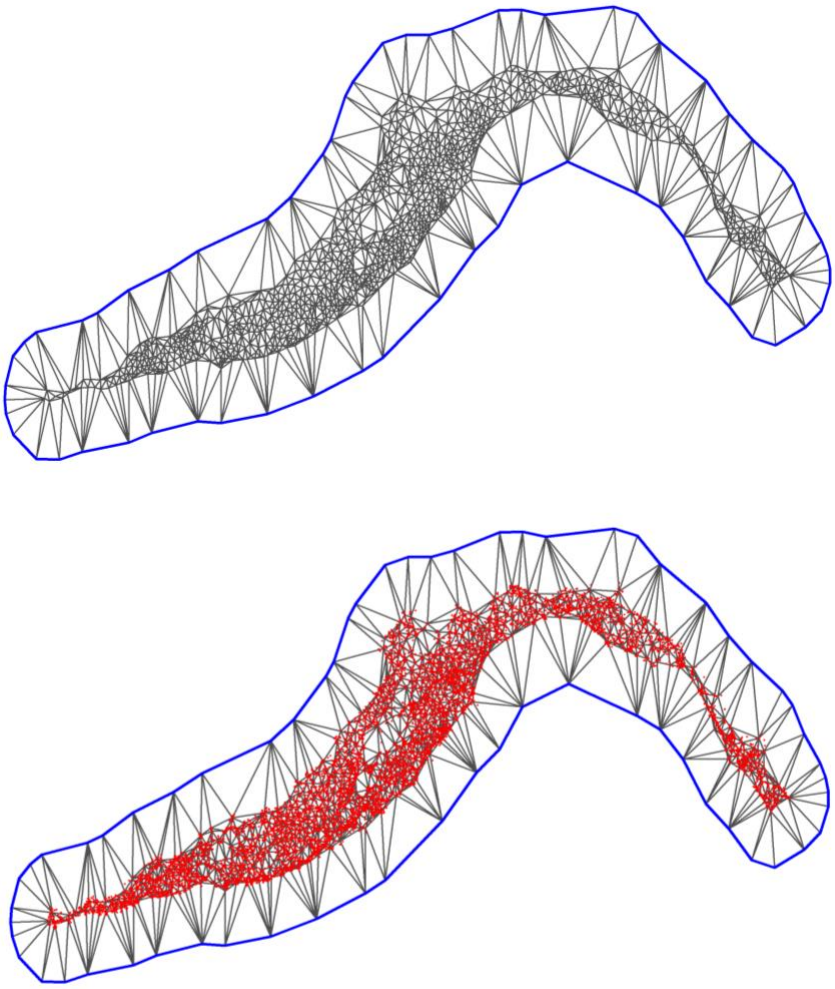

Figure 23. Left: Mesh generated with 1000 knots. Right: 1000-knot with survey haul locations in red.

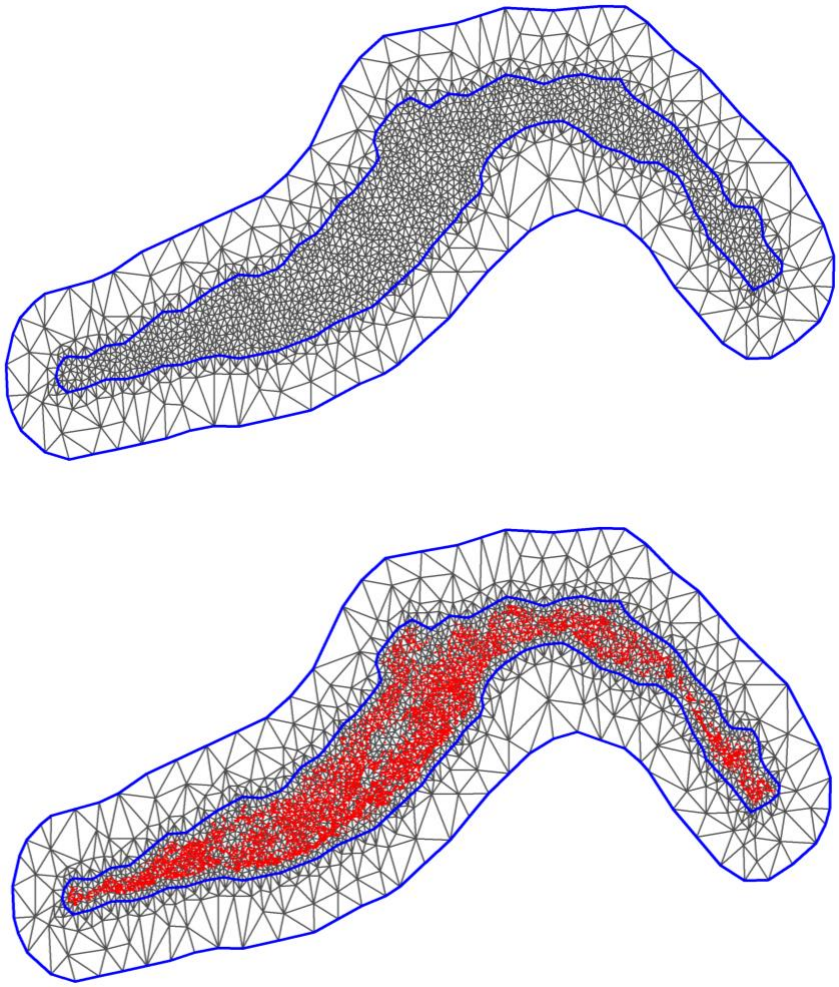

Figure 24. Left: User specified Mesh. Maximum triangle edge is 30m in the inner and 200m in the outer mesh. Right: User specified mesh with haul locations in red

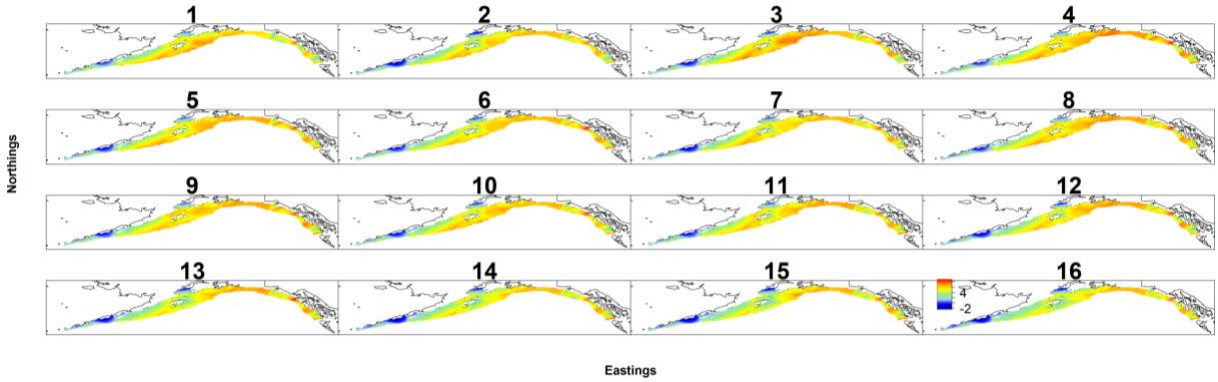

Figure 25. Biomass density plots from the 50-knot VAST model. Number represent survey years: 1984-1999 (triennial), 2001-2019 (biennial). In contrast to annual biomass indices, biomass densities were not bias corrected.

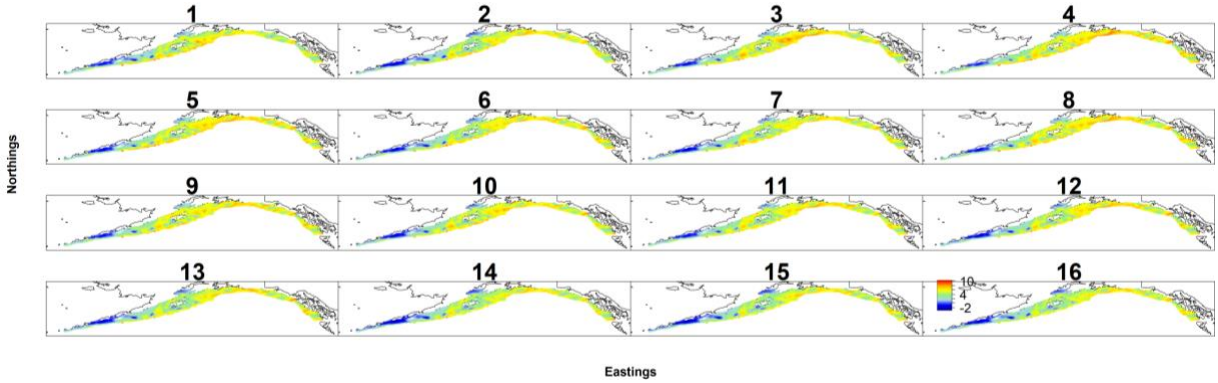

Figure 26. Biomass density plots from the 250-knot VAST model. Number represent survey years: 1984-1999 (triennial), 2001-2019 (biennial). In contrast to annual biomass indices, biomass densities were not bias corrected.

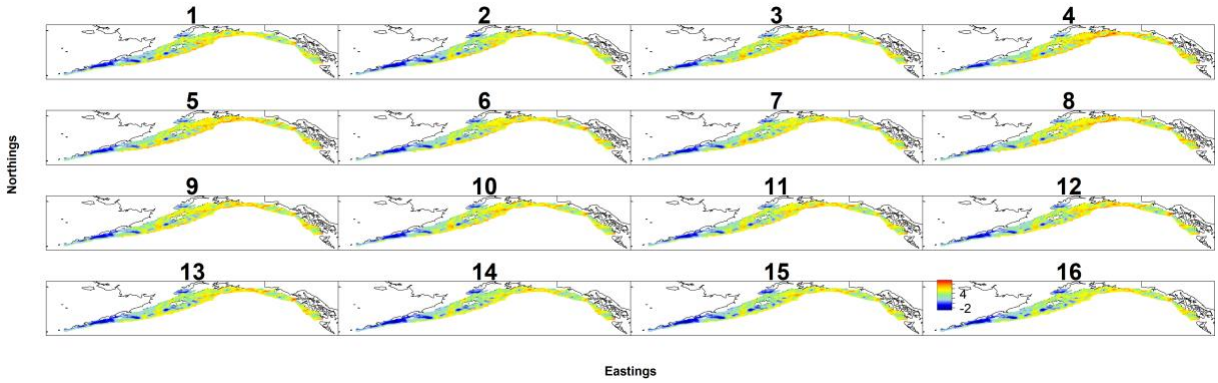

Figure 27. Biomass density plots from the 1000-knot VAST model. Number represent survey years: 1984-1999 (triennial), 2001-2019 (biennial). In contrast to annual biomass indices, biomass densities were not bias corrected.

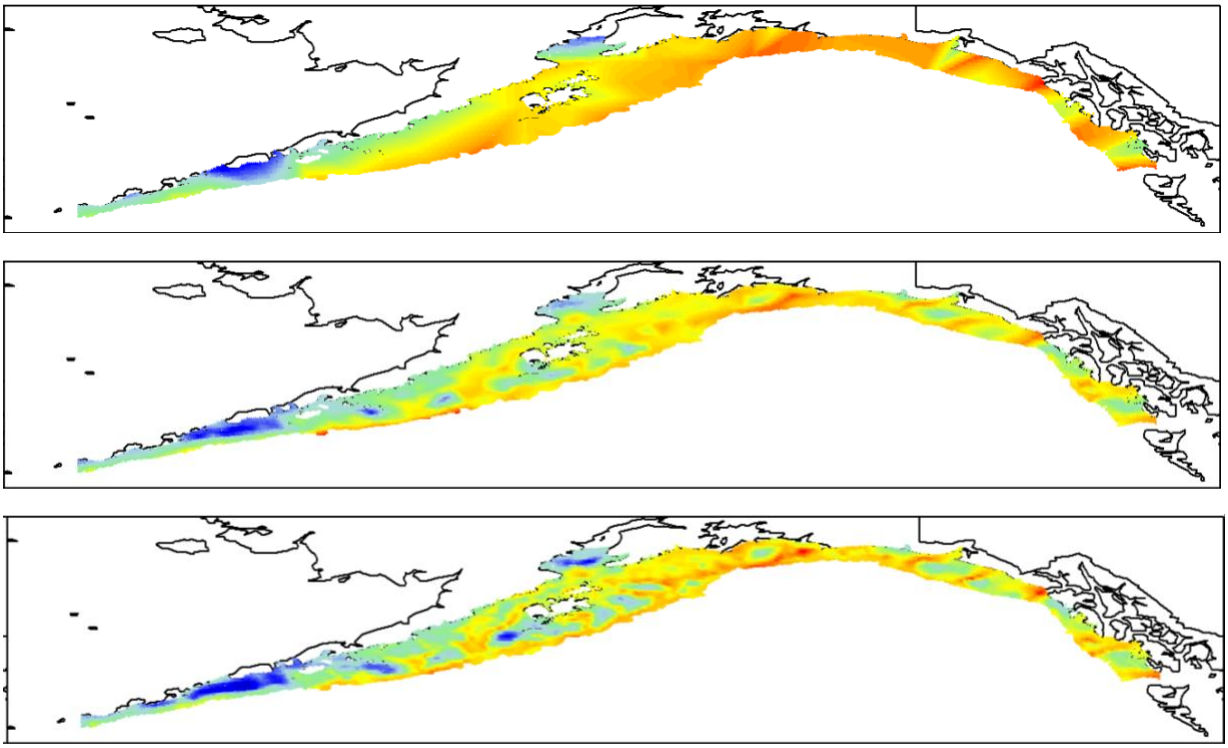

Figure 28. Comparison of VAST biomass density predictions for 1993. This survey year extended out to 500m, therefore, deep regions were informed more from the spatial and spatio-temporal models than observations. In contrast to annual biomass indices, biomass densities were not bias corrected. Top to bottom: 50-knot mesh, 250-knot mesh, 1000-knot mesh.

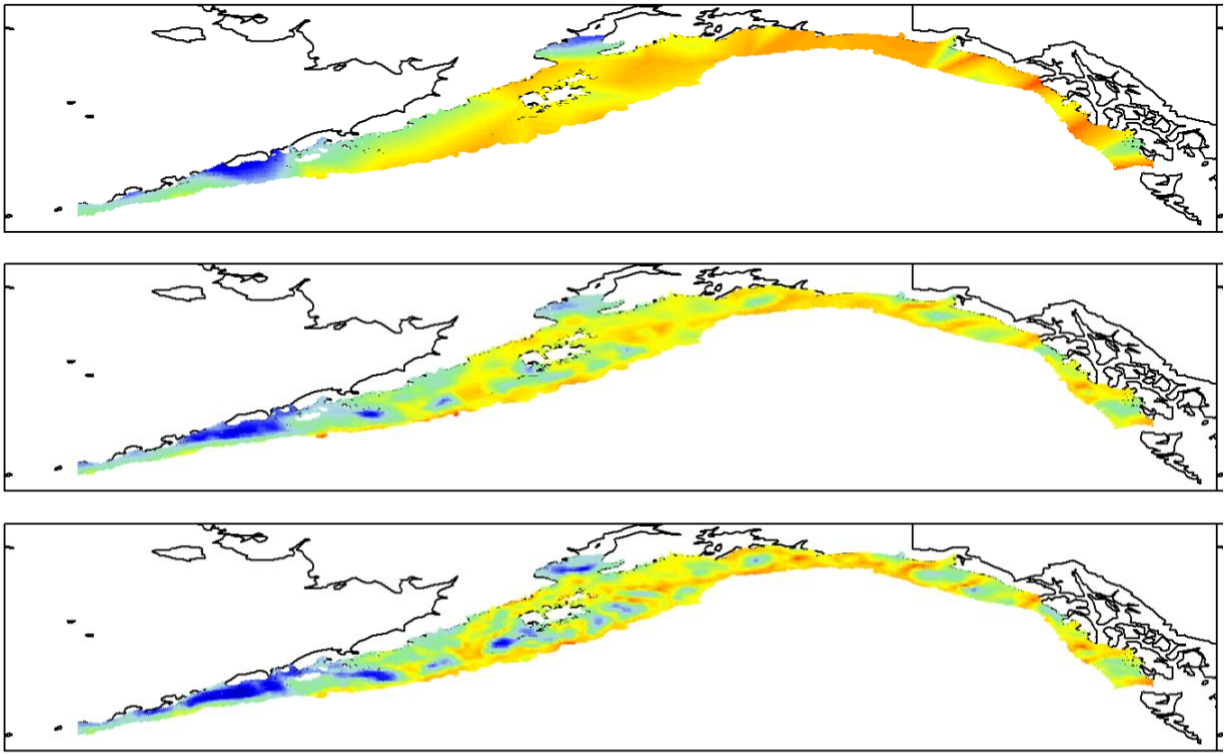

Figure 29. Comparison of VAST biomass density predictions for 2005. This survey year extended out to 1000m, therefore, deep regions were well informed by observations. Biomass densities were not bias corrected and therefore may be different to bias-corrected biomass indices. Top to bottom: 50-knot mesh, 250-knot mesh, 1000-knot mesh.