Use of density (number of individuals per unit area) and abundance (total number in a population) similarly

Title:

Authors: Brandon Chasco, Kara Anlauf-Dunn, Julie Firman, Courtney Zambory, Melanie Davis, Eric Ward, Katherine Kennedy

Journal: MEPS

Brandon Chasco, Washington Department of Fish and Wildlife,

Abstract

Being able to make fine scale predictions of abundance in space and time is a goal of many monitoring efforts, such as evaluating the effectiveness of conservation efforts for protected species. Along the west coast of North America, billions of dollars have been directed towards the recovery of Pacific salmon, whose status varies across regions, species, populations, and life stages. Past efforts to model juvenile and adult salmon abundance have typically relied on single parametric or nonparametric frameworks, often extrapolating future trends from historical data. In this study, we used long-term survey data from juvenile and adult coho salmon in Oregon (USA) to compare the performance of parametric (generalized linear mixed-effect model; GLMM), semi-parametric (generalized additive model; GAM), and nonparametric (random forest; RF) models. Each model was employed to elucidate spatial and temporal variability, as well as the influence of environmental covariates on abundance. Furthermore, we assessed these models' utility for monitoring and forecasting the recovery of juvenile and adult coho salmon populations in Oregon's coastal tributaries. Our comparative analysis revealed that models incorporating spatiotemporal variability yielded lower Root Mean Square Error (RMSE) values than those that did not. Regarding life history stages, RF outperformed in forecasting temporal and spatial out-of-sample densities of juveniles, while for adults, RF excelled in spatial forecasts and GLMM in temporal forecasts. We regard this study as a valuable decision support tool for salmon managers to evaluate population health and efficiently allocate resources for future conservation efforts.

*Introduction*

Worldwide, the majority of marine and freshwater fishes are expected to face the impacts of climate change (Comte et al., 2013; Hollowed et al., 2013; Morrongiello et al., 2011; Perry et al., 2005), which are predicted to alter community diversity (Buisson et al., 2013), affect physiology and growth (Genner et al., 2010), and disrupt global food security (Rice and Garcia, 2011). While both marine and freshwater fishes are vulnerable to climate stressors, anadromous fish populations may be particularly at risk due to their exposure to variable conditions in both freshwater and marine environments (Beamish et al., 2010; Jonsson and Jonsson, 2009; Miles et al., 2000). In certain regions, such as the west coast of North America, anadromous fishes have experienced negative impacts from human activities for over a century, including the effects of hydropower systems, habitat loss through development, harvest, and hatcheries. Introducing additional stressors through a warming climate has the potential to further jeopardize many of these ecologically, economically, and culturally important species.

Monitoring population changes and understanding responses to environmental change require long-term monitoring efforts (Lindenmayer et al., 2022; Magurran et al., 2010). In addition to sampling biological variables of interest, survey efforts often gather data on habitats or local environments to elucidate mechanistic drivers of change. However, increasing costs of monitoring programs, the need to reduce uncertainty, understand the status and trends of other native species populations, or shifting agency priorities have led to pressures on organizations to reduce sampling efforts. For instance, in the marine environment, many trawl survey programs that provide data for fisheries stock assessments are reducing sampling efforts (ICES, 2020; Oyafuso et al., 2021). Similar trends are observed in freshwater ecosystems, where sampling designs are being re-evaluated to maximize cost-effectiveness (Nieman et al., 2021). Consequently, there is a greater need for effective, quantitative methods to assess population trends, evaluate species-habitat relationships, and forecast commercially important fisheries, particularly for anadromous fishes like salmonids due to their complex life history strategies and their significant cultural, ecological, and economic value.

One of the most extensive and long-standing fish and aquatic habitat surveys on the west coast of North America has been conducted by the Oregon Department of Fish and Wildlife (ODFW) to monitor Oregon Coast coho salmon (Oncorhynchus kisutch). Oregon Coast coho were listed under the US Endangered Species Act (ESA) in 1998, prompting the initiation of ODFW’s survey efforts to monitor the status and trends of adult and juvenile coho salmon populations and their aquatic habitats. Surveys employed a spatially-balanced, rotating panel design (GRTS; Stevens, 2002). For this survey design, there were sites that were sampled annually, tri-annually, and every nine-years, as well as the indexes sites that were maintained across all years. Like many long term monitoring programs, the sampling effort associated with the ODFW coho salmon survey has been variable over its 25-year duration. From 1998 to 2012, approximately 550 sites were surveyed annually, but budget reductions after 2012 reduced the number of sites surveyed to 350 (Sounhein et al., 2015). Sampling during the COVID-19 pandemic further reduced sampling efforts to approximately 150 sites during 2020 and 2021. Naturally, pertinent questions persist about the effect of these reductions and future sampling effort on the stated goals of the research design.

Previous studies have examined long-term trends in habitat characteristics and established connections between habitat variables and coho salmon densities (Anlauf-Dunn et al., 2014; Steel et al., 2017, 2012). These studies primarily focused on understanding the spatial variability in coho salmon distributions relative to both proximal (e.g., local, fine-scale) and landscape (e.g., broad-scale) habitat conditions. Advances in model-based approaches incorporating spatial or spatiotemporal variability (Bachl et al., 2019; Evans et al., 2011; Guisan et al., 2002) offer the potential to leverage spatial information to mitigate reductions in the precision of trends that would be expected under reduced effort scenarios (Ward et al., 2012). Another research avenue involves using this long-term monitoring effort to identify linkages between biological responses (such as juvenile densities) and new data sources. Derived products such as the NorWest stream temperature model (Isaak et al., 2017) enable the incorporation of potentially important climate covariates that were previously unavailable. Incorporating such covariates has the potential to increase both the precision of estimates and our mechanistic understanding of environmental factors that may limit coho salmon populations. Moreover, integrating climate variables with an extensive long-term monitoring dataset can aid in identifying variables and modeling approaches with reasonable forecast skill, providing insights into the long-term effects of climate change on Oregon coast coho salmon, as well as other species with similar environmental tolerances.

The primary objective of the present study is to evaluate a range of novel statistical methods incorporating spatiotemporal covariates to fit existing juvenile and adult data and identify those that yield good short-term (1–2 year) forecasts. We focus on three classes of models: generalized linear mixed-models (GLMMs), generalized additive models (GAMs), and random forests (RFs). GLMMs assume that the relationship between the response and predictor variables follows a specific mathematical form, such as a linear or logistic relationship. While not as flexible as other approaches in modeling nonlinear relationships, GLMMs are often easier to interpret and more computationally efficient than alternatives. GAMs allow for the inclusion of nonlinear relationships between the response and predictor variables by using smoothing functions (e.g., splines or cubic polynomials) to model the nonlinear effects of predictors. Like GLMMs, these functions can be added one at a time, allowing the model to capture complex relationships between the variables. However, smooths may also be multidimensional, permitting complex, nonlinear interactions between variables. Both GAMs and GLMMs benefit from the ability of random effects in the mixed model to account for the unbalanced nature of the data. In the case of coho salmon densities, accounting for biases introduced by the unbalanced temporal and spatial collection of data (e.g., landowner denials, weather, or disturbances impacting site access) is essential. Finally, we consider RFs as a machine learning approach to include spatiotemporal variation in salmon densities. RF models excel in handling high-dimensional and complex data and are often used for tasks such as feature selection and prediction. However, they can be more challenging to interpret than GAMs and GLMMs because they are based on an ensemble of decision trees rather than a single mathematical function. Unlike GAMs and GLMMs, which are based on a single outcome, an RF model is an ensemble model built using multiple decision trees. Individual decision trees in an RF model are trained on different subsets of data, and the outcomes of the individual trees are combined to make an ensembled prediction.

Given the expected differential effects of covariates on adult and juvenile densities, we developed separate models for both responses using all three spatiotemporal modeling approaches. Our secondary objective was to better understand the spatiotemporal distribution of coho salmon densities in the future could inform future abundances indices under novel and unbalanced survey designs. Such analyses are valuable for identifying high and low-risk areas, where densities are anticipated to shift.

*Methods*

*Study Area*

The Oregon Coast Coho Evolutionarily Significant Unit (OCC ESU) encompasses 21 independent populations stretching from the Necanicum River in the north to the Sixes River in the south (Figure x; 20,305 km2; Lawson, 2007). The region features a temperate maritime climate with wet, mild winters and warm, dry summers. Geologically, the area is primarily composed of marine sandstones, shales, or basalts, resulting in flashy coastal streams with stream flows closely correlated to recent precipitation. The OCC ESU exhibits a rain-dominated pattern, with peak flows occurring in winter and base flows in late summer and early fall (Harr, 1976).

Approximately one-third of the land in the area is publicly managed (Spies et al., 2007). While mountains predominate, large river valleys that support critical habitats for coho salmon have been cleared for agriculture (Ohmann and Gregory, 2002), with about 90% of streams with high coho salmon potential located on private lands (State of Oregon, 2005; Burnett et al., 2007). Despite efforts toward conservation of stream-side buffers, approximately half of the riparian areas adjacent to streams with high potential for coho salmon were non-forested as of 2007 (Burnett et al., 2007). Outside of the river valleys, the region is characterized by heavy forestation and significant logging activity.

*Data*

The Oregon Department of Fish and Wildlife (ODFW) has conducted a large-scale, spatially balanced, rotating panel sampling design throughout the OCC ESU since 1998, coinciding with the ESA listing of these coho salmon populations (Stevens and Olsen, 2004). This sampling design encompasses adult spawners, juvenile parr, and aquatic habitat, aiming to estimate trends and conduct spatial sampling across various sites (Stevens, 2002). Methods for deriving estimates for each subset of the data can be found in Jacobs and Nickelson (1998; adults), Rodgers (2000; juveniles), and Jones and Moore (1999; habitat). Data on adult and juvenile coho from 1998–2021 were compiled for this study, with exclusion of data from four basins dominated by coastal lakes due to substantial differences in life history patterns and high fish densities potentially obscuring relationships with landscape characteristics in the rest of the region (Jacobs et al., 2002).

Coho salmon density estimates are georeferenced, enabling linkage of observations to important habitat features (Table 1). Landscape habitat drivers are spatially explicit but do not vary with time. Examples include stream size (WidthM; m), stream power (StrmPow; watts/m), distance to spawning/rearing sites (OUT\_DIST; km), mean solar irradiance (SolMean; W-h/m2), stream slope (StrmSlope; rise/run), stream order (STRM\_ORDER; n), and intrinsic potential of coho at each site (Burnett et al., 2007) (IP\_COHO; n).

Additionally, coho salmon densities were related to climate-influenced variables that fluctuated temporally and spatially across the sampling domain. We used the NorWeST model of maximum weekly mean stream temperature estimated for the period of 1993-2011 (Isaak et al., 2017), indexed by annual PRISM July mean air temperature to provide an annual index of maximum stream temperature (MWMT\_Index, ℃), and two metrics based on the PRISM precipitation model: winter maximum precipitation (WD3ppt, mm, annual maximum of the 3-day rolling total), and annual total spring precipitation (SprPpt). An annual index of stream temperature was generated by calculating the mean temperature observed in the PRISM air temperature data for the period between 2002 and 2011, then dividing the PRISM value for each year between 1992 and 2021 by the mean value to generate a deviation from the mean and multiplying this annual deviation by the appropriate NorWeST modeled estimate to provide an index of stream temperature at the points of interest for each year (Isaak et al., 2017).

*Modeling*

Prior to fitting spatiotemporal models, we examined potential covariance among environmental variables to reduce overfitting and identify parsimonious models. Using a Random Forest model with 500 trees and three splits per branch, we examined the partial dependence to quantify the mean decrease in model accuracy for the full suite of model variables. Variables with the lowest decrease in model accuracy included total spring precipitation (SprPpt) and maximum winter precipitation (W3Dppt) for juveniles and stream width (WidthM) and stream order (STRM\_ORDER) for adults; however, we included all covariates listed in Table 1 in the GAM, GLMM, and RF models for juveniles and adults and focused our analysis on the information gained from including spatiotemporal complexity.

Generalized linear mixed-effect models (GLMM) describe the relationship between dependent and independent variables, where observations arise from various statistical distributions (e.g., binomial, Poisson, Gaussian, etc.), and effects of independent variables may be fixed and/or random. Mathematically, we represent the linear model for juvenile and adult coho salmon densities in coastal Oregon tributaries as:

We constructed separate models for juvenile and adult coho salmon densities, with fixed parameters (**b**) related to environmental covariates (**X**), including random effects for the latent spatial process (**ω**) and describing spatiotemporal processes with **ϵ\_y**. Spatial correlation for both spatial and spatiotemporal effects was defined by a Matérn covariance function with a parameter (*κ*) describing decorrelation distance between spatial locations *i* and *j* (|*si* - *sj*|) and two parameters for anisotropy (Thorson et al. 2016). Spatiotemporal deviates could be independent across years or temporally autocorrelated (i.e., i.i.d. or AR(1), respectively). The statistical distributions of random deviations were defined by Gaussian Markov Random Fields (GMRF) approximated using stochastic partial differentials equations (SPDE) to improve computational efficiency (Lindgren et al. 2011). Because both juvenile and adult densities were zero-inflated and skewed, we modeled each response with a Tweedie distribution (Shono 2008) to account for the zero and positive observations in our data. We used the sdmTMB package in R (Anderson et al. 2023), which combines SPDE approximations for spatial GMRF using the Integrated Nested Laplace Approximation package (INLA; Lindgren & Rue 2015) and optimization routines in the Template Model Builder package (TMB; Kristensen et al. 2016), which uses a Laplace approximation to optimize the marginal maximum likelihood of the fixed parameters of the model and the variances of the random effects given the data.

We made several transformations of the data. To increase interpretability of the spatial range parameter, the coordinate data were transformed from latitude and longitude to UTM (Universal Transverse Mercator) Zone 10 with distance measured in kilometers. All non-categorical covariates were Z-scored. Unlike GAM or RF models, the GLMM is a parametric model that does not require tuning. For both rearing juveniles and adult spawners, we included all covariates in the model to make forecasts under future climate conditions and then tested 12 different model permutations based on temporal (include or not include), spatial (include or not include), and spatiotemporal random deviates (include, not include, i.i.d.).

Generalized additive models (GAMs) were constructed and fit using the mgcv package (Wood and Wood 2015) in R. Spatial effects were modeled as two-dimensional thin-plate regression splines, considering models that only included these smoothing for the spatial effects (e.g., no interaction with year) and models that estimated the interactions with year as a three-dimensional spline. Stream order was treated as a factor, and all remaining covariates were modeled as one-dimensional thin-plate regression splines with a basis dimension of 4. Although the observation model for our GAMs is a Tweedie distribution, like the GLMMs, the non-parametric description of the temporal, spatial, and spatiotemporal does not require decisions about the covariance structure (i.e., i.i.d., random walk, or Matern). Thus, there are only four permutations of the GAM model: a covariate-only model or a covariate model with temporal, spatial, or spatiotemporal effects.

Random forest (RF) models represent a non-parametric alternative to GLMM and GAM approaches. As these models are non-parametric, there is no assumption about the distribution of the response variable, and they implicitly account for nonlinearities or complicated interactions between variables. We constructed RF models using the randomForest package in R. Because these models may be sensitive to the choice of tuning parameters, we performed a grid search over the number of covariates used at each split in the tree and the number of trees to grow. To maintain consistency among GLMM, GAM, and RF approaches, we included all the same dependent variables as well as iterations of spatial, temporal, or spatiotemporal effects similar to the GLMM and GAM models.

*Evaluating model performance for different spatial, temporal, and spatiotemporal survey designs*

In comparing the modeling approaches, which range from fully parametric (GLMM) to fully non-parametric (RF), we evaluated their performance using the average root mean-squared error (RMSE) for predicted out of sample juvenile and spawner fish densities from 2017 to 2021. To identify the optimal model configuration within each approach, an exploratory analysis was conducted, considering the configuration resulting in the lowest RMSE. This analysis can be interpreted as a zero-year-ahead approach, where predictions for the years 2017 to 2021 utilized all available data up to and including those years.

Following the determination of the optimal model configuration within each approach, we conducted a series of temporal and spatial examinations of predictive ability using Leave Future Out Cross Validation (LFOCV). Temporal assessments included forecasting one and two years ahead, where out-of-bag predictions for 2017 were based on data up to 2016 and 2015, respectively. Spatial evaluations for zero years ahead involved out-of-bag samples comprising annual, annual plus triannual, and index sample locations up to and including 2017. For each LFOCV analysis, we computed the average RMSE for the years 2017 to 2021.

*Effect of survey design on derived inidces*

In our final analysis, we focused on comparing the predictive performance of various survey designs using the GLMM approach exclusively. In this Leave Future Out Cross Validation (LFOCV) analysis, we systematically eliminated each population group and then computed the average root mean square error (RMSE) across all populations for the years 2017 to 2021, forecasting zero years ahead. Subsequently, by retaining only the five populations resulting in the highest RMSE (indicating a greater loss of predictive ability), we compared the estimated juvenile and spawner indices for each population group against maintaining the current Generalized Random Tessellation Stratified (GRTS) sampling design.

Results

*Exploratory analysis*

In our exploratory analysis, we observed that the random forest model consistently exhibited the lowest average root mean square error (RMSE) for predictions spanning 2017 to 2021, followed by the GLMM and GAM models for both juvenile and spawner life history stages (Fig. 1). Notably, the formulas for the models with the lowest RMSE differed between the life stages.

For the models predicting juvenile density, the GAM incorporated a three-dimensional spline to capture spatio-temporal variability (i.e., s(UTM\_E\_km, UTM\_N\_km, yr) ), while the GLMM accounted for included a spatial random across all years and spatio-temporal deviations that were assumed to be independent and identically distributed (i.i.d.). The most predictive random forest model for juvenile densities included all environmental covariates, spatial, and temporal data, utilizing 11 splits and 1000 trees.

Regarding spawner densities, the formula for the GAM model with the lowest RMSE mirrored that of the juvenile model. However, the GLMM model for spawner densities only included spatio-temporal random deviates and omitted average random spatial deviates constant across all years. The optimal random forest model for spawner densities incorporated data for the year but not the UTM coordinates for the survey location, consisting of 11 splits and 200 trees.

*Temporal and spatial out-of-bag predictability*

In our analysis, we investigated the predictive performance of all three approaches using various temporal and spatial out-of-bag samples, and then compared them based on the average root mean square error (RMSE) for predictions spanning from 2017 to 2021. Three notable patterns emerged: First, the random forest model consistently demonstrated the lowest RMSE for both juveniles and spawners across different spatial and temporal out-of-bag samples (Fig. 1). However, it's worth noting that the random forest models showed the largest increases in RMSE relative to its exploratory model. For the RF approach, the average percent increase for the different spatial and temporal out-of-bag sample was 118% for juvenile data and 120% for spawner data. Conversely, the relative percent increases for the GAMM approach were 9% and 7% for juvenile and spawner data, respectively, and 14% and 20% for the GLMM. Second, out-of-bag predictions derived from survey designs incorporating "annual" and/or "tri-annual" observations demonstrated lower RMSE compared to observations solely based on "index" sampling across all models and life stages. Third, the models displayed greater accuracy in predicting spatial out-of-bag samples compared to temporal out-of-bag samples one or two years into the future, used data from all survey types. This suggests that maintaining a consistent set of sites each year yields more precise model predictions than retaining all sites but not sampling every year.

*Prioritizing modeling approach*

In our study, we found that while the random forest approach consistently exhibited lower RMSE compared to the GAM or GLMM in both the exploratory and out-of-bag cross-validation analyses. However, the random forest model with the lowest RMSE had 11 variables sampled at each split, which is high considering the total of 13 variables in the model. This high number of variables sampled per split leads to increased independence among trees, potentially resulting in reduced generalizability to unseen data. This is evidenced by the larger relative increases in RMSE for out-of-bag sample estimates using the random forest model (Fig. 1). Because the GLMM approach has interpretable parameter estimates, and an intermediate level of overall and relative RMSE compared to the GAMM or RF approaches, we ultimately selected the GLMM as the preferred model for characterizing the effects of model covariates on coho salmon densities from 1998 to 2021.

*Spatio-temporal variability*

Spatial and spatiotemporal variability in juvenile and spawner densities are characterized by the standard deviation and anisotropy of the random effects, along with the overdispersion of the Tweedie distribution for the observed process. The median and 90% quantiles for observed juvenile densities, 164 fish/km (0, 1 421), exceed those of observed spawner densities, 10 fish/mi (0, 130). The estimated overdispersion for juvenile densities is 12.95, and for spawner densities, it is 4.92, indicating more variability than expected based on the mean-variance relationship for the Tweedie observation process, even after considering environmental covariates. In the case of juvenile densities, a greater proportion of the overdispersion is attributed to spatial variability (standard deviation equal to 4.89) (Supplemental Fig. 1) compared to spatiotemporal variability (standard deviation equal to 0.36) (Supplemental Fig. 2). For spawner densities, the best-fit model incorporates only spatiotemporal variability, with a value of 1.27 (Supplemental Fig. 3).

The anisotropy of fish densities characterizes the directional dependence of these densities across latitudinal and longitudinal space. For spawner densities, which includes only spatiotemporal deviates, the estimated distance where decorrelation decreases to 10% along the latitudinal and longitudinal axes is 37.9 and 41.5 kilometers, respectively, at 94 degrees. In contrast, for rearing juveniles, where the spatial and spatiotemporal deviates are assumed to have the same anisotropic properties, the decorrelation distances are 136.1 and 249.7 kilometers, respectively, at 73 degrees. This suggests that the spatial correlation for spawners is lower than for juveniles, exhibiting an equal gradient both north and south (Supplemental Fig. 4). Conversely, the gradient for juveniles is elongated along a southwest to northeast axis (Supplemental Fig. 5).

*Covariate effects for the GLMM model*

In our GLMM approach, we included six spatially static and three spatiotemporal environmental covariates without interactions in the different models, and we observed minimal evidence of multicollinearity. The maximum absolute correlation between any two variables was 0.44 for juvenile models and 0.41 for spawner models. All continuous covariates were Z-scored to make effects size comparable.

Analyzing the effects on juvenile coho salmon densities, we found that stream slope (-0.39 ± 0.04) exerted the most significant negative influence after the stream order, whereas outmigration distance (0.32 ± 0.14) had the most substantial positive impact (Fig. 2). For densities of adult spawning coho salmon, both stream slope (-018 ± 0.02) and the maximum weekly mean temperature index demonstrated notable negative effects on densities, while solar intensity (0.11 ± 0.02) exhibited a positive effect (Fig. 3). Notably, only the maximum weekly mean temperature index and spring precipitation could be forecasted into the future under various climate scenarios. Although both variables showed negative effects (-0.04 ± 0.04 and -0.01 ± 0.04, respectively) for juveniles were relatively weak compared to other covariates, and a mix of positive and negative effects that were comparatively larger for spawner densities (-0.17 ± 0.03 and 0.03 ± 0.04) (Table 3).

*Annual population abundance based on observed and model-estimated densities*

At the population level, we assumed each observed or model-predicted juvenile and spawner density account for one kilometer or mile, respectively, of tributary. Summing the observed and model-estimated densities for each year, we found these derived variables of the aggregated juvenile and spawning Coho salmon abundance in Oregon coastal tributaries closely aligned with the observed data (Fig. 4 and Fig. 5, respectively). However, we observed variations in the precision and accuracy of the abundance estimates between juvenile and spawner populations. Specifically, the higher average standard errors for the log-transformed estimates of abundance for juveniles (average s.e. 0.28) relative to spawner (average s.e. = 0.41) reveals lower precision. The percent of observed abundances within the 95% CI for the estimated spawner abundance is 98%; conversely, only 73% of the observed juvenile abundances are within the 95% CI suggesting lower accuracy of the GLMM model for this life history stage. Beaver Creek and Sixes River have the least accurate predictions for juvenile and spawners (Fig. 4 and 5).

*Population monitoring strategy evaluation*

To prioritize which populations should be considered in future monitoring efforts, we initially assessed the relative increase in RMSE across all populations after removing each individual population. For spawner densities, we observed that the removal of Coos River, Siletz River, Lower Umpqua, Siuslaw River, and Beaver Creek resulted in the largest relative increases in the average RMSE for the period spanning 1998 to 2021 (Fig. 6). Similarly, in the case of juvenile densities, eliminating Nestucca River, Siuslaw River, Lower Umpqua, Coos River, and South Umpqua led to the most significant increases in the average RMSE (Fig. 7). Notably, for both life history stages, the gradient of relative RMSE across the populations becomes more pronounced when the survey encompasses all GRTS observations, gradually diminishing as the temporal GRTS coverage is reduced to annual sites only.

To gain insights into how Root Mean Square Error (RMSE) can guide future population monitoring efforts, our study employed a model trained on data available up to 2016 for all populations. Subsequently, we focused on adding to the training data only the five spawner or juvenile population groups exhibiting the greatest RMSE increases when years 2017 to 2021, as detailed in the preceding paragraph. We refer to the populations as the surveyed populations since they were added to the training data, and then remaining out-of-bag populations as unsurveyed. Utilizing the get\_index function within sdmTMB, we derived annual "abundance" estimates for both surveyed and unsurveyed populations by aggregating densities within each year.

For annual spawner abundance, relative to the exploratory analysis encompassing all populations, the RMSE for years 2017 to 2021 decreased by 18% for surveyed populations but increased by 298% for unsurveyed populations. Notably, the average standard errors rose for both surveyed and unsurveyed populations (8% and 117%, respectively). This suggests that the inclusion of out-of-bag population groups in the exploratory analysis diminished the predictability of the training populations used in this monitoring strategy evaluation.

In the case of juveniles, the RMSE in the monitoring strategy evaluation rose by 48% for surveyed populations and by 117% for unsurveyed populations compared to the exploratory analysis. Concurrently, the average standard error for annual abundance estimates exhibited increases of 63% and 103% for surveyed and unsurveyed populations, respectively.

**Figures**

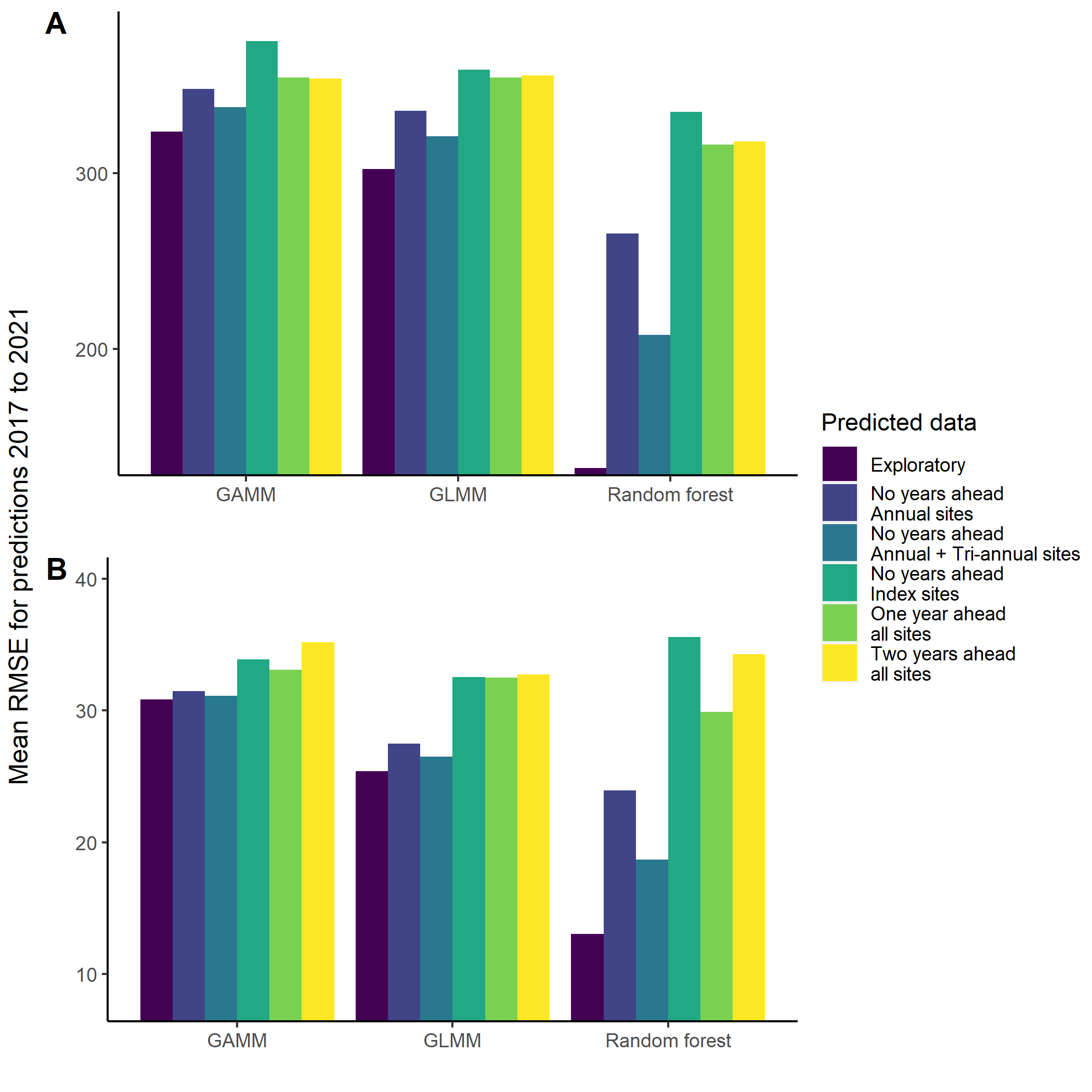


Figure 1. Average root mean square error (RMSE) from 2017 to 2021 for each model with the lowest RMSE in the exploratory analysis, and five out-of-bag samples of juvenile (panel A) and spawner (panel B) densities with different levels of survey effort. (ggplot\_predictive\_RMSE.r)

Chart, diagram

Description automatically generated

Figure 2. Marginal effect of the nine environmental covariates for the GLMM model predicting juvenile coho salmon densities in coastal Oregon tributaries from 1998 to 2021. The range of the covariates have been Z-scored to standardize the estimated effects. Shaded areas represent the 90% confidence intervals and the solid represents the mean effect. (ggplot\_marginalPlot.r; Life\_stage = ‘rear’)

Diagram, engineering drawing

Description automatically generated

Figure 3. Marginal effect of the nine environmental covariates for the GLMM model predicting spawning coho salmon densities in coastal Oregon tributaries from 1998 to 2021. The range of the covariates have been Z-scored to standard the estimated effects. Shaded areas represent the 90% confidence intervals and the solid represents the mean effect. (ggplot\_marginalPlot.r; Life\_stage = ‘Spwn’)

Calendar

Description automatically generatedFigure 4. GLMM model-estimated aggregate abundance for all surveyed locations within a year and population (black points) with 90% confidence intervals (lines) and observed aggregated abundance of (red points) juvenile Coho salmon abundance for 18 Oregon coastal tributaries from 1998 to 2021 based on survey locations in each year. Aggregated abundances assume each survey location accounts for one kilometer of tributary (Figure\_4.r).

Calendar

Description automatically generated

Figure 5. GLMM model-estimated aggregate abundance for all surveyed locations within a year and population (black points) with 90% confidence intervals (lines) and observed aggregated abundance of (red points) spawning Coho salmon abundance for 18 Oregon coastal tributaries from 1998 to 2021 based on survey locations in each year. Aggregated abundances assume each survey location accounts for one kilometer of tributary (Figure\_5.r).

Chart, bar chart

Description automatically generated

Figure 6. Sensitivity analysis of the Generalized Linear Mixed Model (GLMM) to forecast spawner densities. The analysis evaluates the mean root-mean-squared error across all populations from 2017 to 2021 when individual populations are excluded from the analysis starting in 2017. Three levels of sampling intensity are considered: all Generalized Random Tessellation Stratified (GRTS) locations, only annual and triannual locations, and only annual locations .

Chart, bar chart

Description automatically generated

Figure 7. Sensitivity analysis of the Generalized Linear Mixed Model (GLMM) to forecast spawner densities. The analysis evaluates the mean root-mean-squared error across all populations from 2017 to 2021 when individual populations are excluded from the analysis starting in 2017. Three levels of sampling intensity are considered: all Generalized Random Tessellation Stratified (GRTS) locations, only annual and triannual locations, and only annual locations.

Chart

Description automatically generated

Figure 8. The annual abundance of spawning Coho salmon across survey locations within a population group. Observed abundances are denoted by black circles, while estimated abundances, represented by points and lines indicating mean and 90% confidence intervals, are superimposed. Abundance values are derived from the summation of observed and predicted densities, with each observation presumed to signify a mile of stream (Figure\_8.r).

A picture containing calendar

Description automatically generated

Figure 9. The annual abundance of juvenile Coho salmon across survey locations within a population group. Observed abundances are denoted by black circles, while estimated abundances, represented by points and lines indicating mean and 90% confidence intervals, are superimposed. Abundance values are derived from the summation of observed and predicted densities, with each observation presumed to signify a kilometer of stream (Figure\_9.r).

Table 1. Covariate effects for juvenile and spawner densities of coho salmon in Oregon coastal tributaries for the spatially explicit models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Abbrev | Covariate | Units | Source | Resolution | Interval |
| WidthM | Active channel width | Meters | TerrainWorks | Reach | Static |
| StrmSlope | Channel slope multiplied by drainage area | rise / run | TerrainWorks | Reach | Static |
| OUT\_DIST | Distance upstream from outlet | Kilometers | TerrainWorks | Reach | Static |
| SolMean | Solar shading/irradience | W-h/m2 | TerrainWorks | Reach | Static |
| IP\_COHO | Coho salmon intrinsic potential | Numbers | TerrainWorks | Reach | Static |
| MWMT Index | Mean weekly maximum temperature index | °C | NorWest | Reach | Annual |
| W3Dppt | Winter greatest 3-day sum precipitation | Millimeters | PRISM | 30 km pixel | Annual |
| SprPpt | Spring seasonal sum precipitation | Millimeters | PRISM | 30 km pixel | Annual |
| StrmPow | Stream power | watts / m |  |  |  |
| Stream order | Stream order | Integer |  |  |  |

*PRISM* Daley et al. (1994)

Table 2. Combinations of fixed and random effects for the random forest, generalized additive and generalized linear mixed models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model components |  | Random forest | GAM | GLMM |
| Fixed effects |  | Included or not included | Included or not included | Included or not included |
| Random effects | Temporal | Year | s(year, k) | (1 | year) |
|  | Spatial | UTM\_E + UTM\_N | s(UTM\_E,UTM\_N, k) | sp = c(TRUE, FALSE) |
|  | Spatiotemporal | UTM\_E + UTM\_N + year | s(UTM\_E,UTM\_N,year, k) or  s(UTM\_E,UTM\_N) + s(year,k) | st = c('i.i.d.', FALSE) |
| Tuning parameters |  | mtry = {3,5,7,9,11} | k = 4 |  |
|  |  | ntree = {200, 400, 600, 800, 1000} |  |  |
| Model combinations |  | 500 | 4 | 8 |

Table 3. Estimated covariate effects for the juvenile and spawners densities of Coho salmon in the Coastal Oregon ESU (Figure\_2\_and\_3\_results\_for\_text.r).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Juveniles | |  | Spawners | |
|  | Estimate | Standard error |  | Estimate | Standard error |
| (Intercept) | 3.66 | 3.85 |  | 3.18 | 0.13 |
| fSTRM\_ORDER2 | -1.75 | 0.28 |  | -0.84 | 0.13 |
| fSTRM\_ORDER3 | -0.89 | 0.24 |  | -0.68 | 0.12 |
| fSTRM\_ORDER4 | -0.10 | 0.23 |  | -0.49 | 0.12 |
| fSTRM\_ORDER5 | -0.10 | 0.24 |  | -0.65 | 0.12 |
| fSTRM\_ORDER6 | -0.55 | 0.26 |  | -0.57 | 0.16 |
| fSTRM\_ORDER7 | -1.74 | 0.39 |  | 0.21 | 0.33 |
| fSTRM\_ORDER8 | -4.61 | 1.87 |  |  |  |
| StrmSlope | -0.39 | 0.04 |  | -0.18 | 0.02 |
| WidthM | -0.02 | 0.05 |  | -0.02 | 0.03 |
| SolMean | 0.03 | 0.03 |  | 0.11 | 0.02 |
| IP\_COHO | 0.15 | 0.03 |  | 0.06 | 0.02 |
| OUT\_DIST | 0.32 | 0.14 |  | 0.07 | 0.04 |
| StrmPow | -0.04 | 0.03 |  | -0.03 | 0.02 |
| MWMT\_Index | -0.08 | 0.05 |  | -0.17 | 0.03 |
| W3Dppt | -0.04 | 0.04 |  | 0.05 | 0.04 |
| SprPpt | -0.01 | 0.04 |  | 0.03 | 0.04 |

**Supplemental Information**



Supplemental Figure 1. Spatial random deviates of juvenile rearing densities (number / km) of Coho salmon in the Oregon Coastal ESU across all years (Supplemental\_Fig1\_Fig2\_Fig3.r).

Diagram

Description automatically generated

Supplemental Figure 2. Spatiotemporal random deviates of juvenile rearing densities (number / km) of Coho salmon in the Oregon Coastal ESU from 1998 to 2021 (Supplemental\_Fig1\_Fig2\_Fig3.r).



Supplemental Figure 3. Spatiotemporal distribution of spawner densities (number / mi) of Coho salmon in the Oregon Coastal ESU from 1998 to 2021 (Supplemental\_Fig1\_Fig2\_Fig3.r).

Diagram

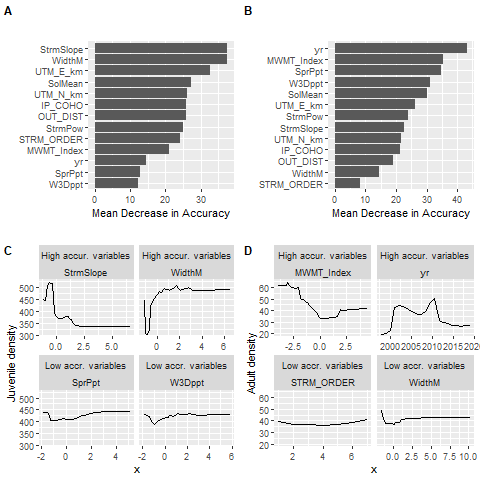
Description automatically generated

Supplemental Figure 4. Estimated distance and rotation where the decorrelation between adjacent spawning densities (number / mi) for Coho salmon in the Oregon Coastal ESU reach 10% (Supplemental\_Fig4\_and\_Fig5.r).

Diagram

Description automatically generated

Supplemental Figure 4. Estimated distance and rotation where the decorrelation between adjacent juvenile rearing densities (number / km) for Coho salmon in the Oregon Coastal ESU reach 10% (Supplemental\_Fig4\_and\_Fig5.r).



PROBABLY GOING TO LEAVE THIS ONE OUT. Supplemental Figure 1. Mean decrease in accuracy for the full suite of variables (see Table 1) in random forest models (ntree = 500) for rearing juveniles (panel A) and adult spawner (panel B) densities of coho salmon in western Oregon tributaries. Partial dependence plots for rearing juveniles (panel C) and adult spawners (panel D) for variables with the highest (upper row; panels C and D) and lowest mean decrease in model accuracy (low row; panels C and D).