

Supplementary Methods

General Overview

Our overall analytic goal was to estimate potential changes in Red Snapper reproductive parameters (fecundity, spawning interval, and spawning period) over a 26 year period. Our data came from multiple studies conducted by different researchers at varying times throughout the northern Gulf of Mexico (Gulf). It would be naive to assume that the error associated with these factors is negligible. Therefore, each of our models contain two components: a time series, which estimates the expected value of the reproductive parameter for each year, and a meta-analysis, which estimates variation among data sources around this expected value.

Notation

A vector of data or parameters will be noted with an arrow (e.g., $\vec{\theta}$ is a group of parameters and its third element is θ_3). Raw data is denoted \vec{y} ; summarized data is $\vec{\bar{y}}$, with standard error $\vec{\sigma}_{\bar{y}}$. When discussing individual data points, raw data will be indexed with i and summarized data with j (e.g., y_i and \bar{y}_j). The expected values of the models will be denoted $\vec{\eta}$ (for the meta-analysis) or $\vec{\theta}$ (for the time series). Variances will be indicated with σ^2 (for residual or likelihood variances) or τ^2 (for the variance of a group of parameters). Probabilistic relations are indicated with “ \sim ” and deterministic relations with “ $=$ ”.

Meta-analysis details

Estimating effects of interest (such as the change in a parameter over time) can be complicated when the data have a group structure (such as belonging to multiple studies, being collected during multiple months, etc.). Even in a well controlled experiment, there would be some variation among the groups; observational data that were not collected with meta-analysis in mind would reasonably be expected to have effects that vary non-negligibly. A hierarchical modeling (also called random effects) approach accommodates this issue by treating each group as a sample from a distribution of possible parameter values that share a common variance term (τ^2). This partially pools the information from different groups while allowing for some flexibility (McElreath, 2016). If τ^2 is small, there are few differences among the groups, while larger values can allow for nearly independent parameters. In general, we have given τ^2 a lightly regularizing prior distribution, which puts a higher probability on low values but can allow for greater variation if the data require it. Conceptually, this is like placing a rubber band around the group of parameters: τ^2 keeps their values close together unless they are “pulled” by the data. A commonly used prior distribution is $\tau \sim t_{\nu}^{+}(0, s)$, where t_{ν}^{+} is the positive half of a Student’s t distribution with ν degrees of freedom and scale s . We used $\nu = 3$ for residual (σ^2) and parameter (τ^2) variances unless otherwise noted, as this provides moderate regularization without excluding large values that are warranted by the data. Note that we placed the prior on the standard deviation instead of the variance, as the distribution is more interpretable on this scale.

A slightly different method is necessary when variance parameters may differ among groups. Here, we define a hierarchical variance prior, $\mathcal{HV}(\bar{\sigma}^2, c)$. If $\sigma_g^2 \sim \mathcal{HV}(\bar{\sigma}^2, c)$ is a variance parameter associated with group g , then $\sigma_g^2 = \bar{\sigma}^2 G \phi_g$, where $\bar{\sigma}^2$ is the average variance of all groups, G is the number of groups, and θ_g is the proportion of the total variance ($\bar{\sigma}^2 G$) that belongs to group g . $\vec{\phi}$ is a G -length vector of probabilities that sums to 1; it is given a symmetric Dirichlet prior distribution with concentration c . The greater the value of c , the greater the probability that each element of $\vec{\phi}$ is near $1/G$, meaning that σ_g^2 converges towards $\bar{\sigma}^2$. We generally use $c = 3$, which limits the variation among groups while allowing it if needed by the data.

Time series details

The time series components of our processes are derived from Gaussian process (GP) regressions, which model how the similarity of parameters decreases as their temporal distance increases. This is a flexible method that allows the estimation of a temporal trend where there is similarity among nearby years, but not a linear increase or decrease. The expected values for each year are modeled as $\vec{\theta} \sim \text{MultiNormal}(\mu, \Sigma)$, where μ is the overall average parameter value. The covariance (Σ) is given by function of the amount of time between each pair of years, \mathbf{D} , where more distant years have less covariance. We used an exponential

covariance function, $\Sigma = \tau_{\Sigma, T}^2 \exp(\frac{-\mathbf{D}_T}{\ell_T})$, where $\tau_{\Sigma, T}^2$ is the scale of among-year variance, \mathbf{D}_T is the number of decades between time points and ℓ_T is the length-scale, the amount of time where the correlation between years equals $e^{-1} \approx 0.368$. Larger values of ℓ_T result in more consistent patterns of similarity over time. The parameters of a GP are only weakly identified without the use of moderately informative priors on either $\tau_{\Sigma, T}^2$ or ℓ_T .

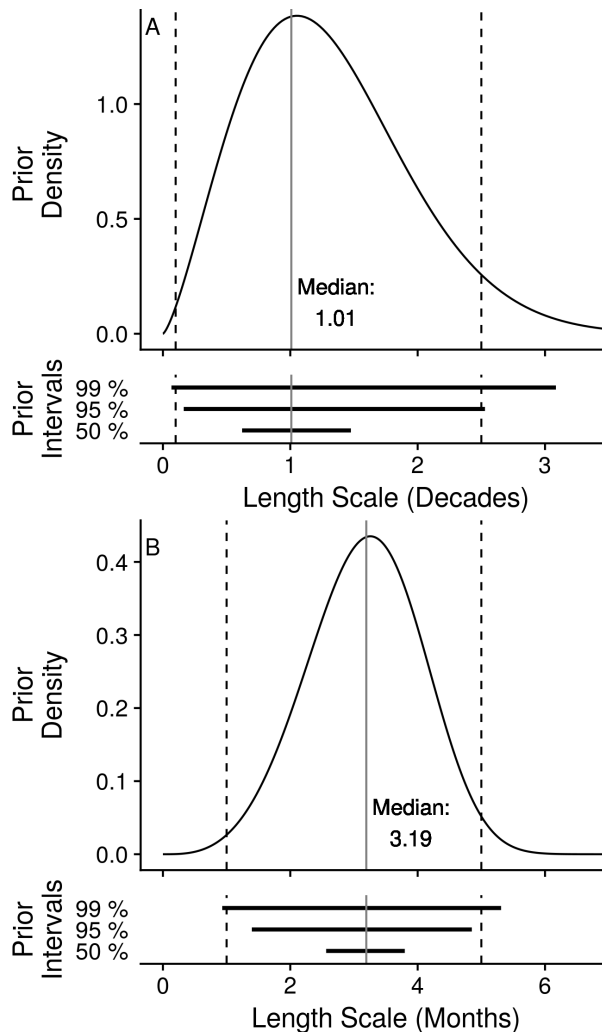
For our models, we placed a Weibull prior on the transformation of length-scale parameter: ${}^{2.2}\sqrt{\ell_T} \sim \text{Weibull}(4, 1.1)$. We selected this prior so that most of the probability of ℓ_T was concentrated in a region between the smallest distance (1 year) and the largest distance (26 years) while still allowing for a small amount of probability outside this region (Suppl. Figure, panel A). For the spawning seasonality model, we also used a GP for variation among months. We used a different length-scale prior for this time series, which concentrates probability between one and five months: $\ell_M \sim \text{Weibull}(4, 3.5)$. The distance matrix \mathbf{D}_M is the pairwise distance in months, which ranges from 1 to 6. This is illustrated in the Suppl. Figure, panel B. Gaussian processes are indicated by $\theta \sim \mathcal{GP}(\mu, \ell, \tau_{\Sigma}^2, \mathbf{D})$.

Implementation details

All models were written in Stan v. 2.17 (Carpenter et al. 2017), an analysis engine that uses highly efficient Markov chain Monte Carlo (MCMC) methods to estimate a posterior distributions. Data pre- and post-processing was handled in R v 3.4.1 (R Core Team 2017). Unless otherwise mentioned, models were run on 4 independent chains for 1000 warmup and 1000 sampling iterations each. Results were checked for non-convergence as recommended in the Stan user manual (Stan Development Team 2017), and model fit was checked with posterior predictive checks.

Spawning Seasonality

Our raw data (y_i) is individual gonadosomatic index (GSI). For summarized data, \bar{y}_j is the average GSI of a certain month and year, with $\sigma_{\bar{y}, j}$ either taken directly from the paper or calculated from standard deviation and sample size. The month, year, and study were recorded for each raw or summarized data point. Data were assigned into groups based on the combination of their month, year, and study. Because we were estimating GSI for each month and year, we used a compound Gaussian Process composed of three components: $\vec{\alpha}$ (the annual trend), $\vec{\beta}$ (the monthly trend), and γ (the month-by-year interaction). The interaction component is a matrix in which the rows have monthly autocorrelation and the columns have annual autocorrelation; this requires a more complex covariance matrix than a typical GP. We considered spawning activity to be occurring in



Suppl. Figure: Prior distributions for Gaussian process length-scale parameters (ℓ), for yearly (A) and monthly (B) time series. Each figure includes the prior's probability density, credible intervals, and median. Dashed horizontal lines indicate the range over which the prior was calibrated to be informative. Note that the monthly prior is only used for the spawning seasonality model.

month m and year t ($S_{m,t}$) if the expected mean GSI $\theta_{m,t} > 1$. For the meta-analysis, variation among groups was modeled hierarchically with ε_g .

Integrating raw and summarized GSI data into a single model required a bias correction. Under central limit theorem, \bar{y} are normally distributed around the expected value ($\bar{\eta}$) regardless of the distribution of y . However, the heavily skewed raw GSI data were best described by a log-normal distribution, which centers around the geometric mean (noted here as λ). Fortunately, the expected value of a log-normal distribution is $\bar{\eta} = \exp(\lambda + \frac{\sigma^2}{2})$, where σ^2 is an estimated parameter of the log-normal distribution. We were thus able link both data types together with this bias correction and model the expected value of GSI. Since both normal and log-normal likelihoods contribute information to the posterior distribution in an amount that is proportional to standard error (Gelman et al, 2013), this approach ensured that raw and summarized data were appropriately weighted and combined. For data in month m , year t , and group g :

$$y_i \sim \text{logNormal}\left(\ln \eta_i - \frac{\sigma_m^2}{2}, \sigma_m^2\right) \quad (1)$$

$$\bar{y}_j \sim \text{Normal}(\eta_j, \sigma_{\bar{y},j}^2) \quad (2)$$

$$\sigma_m^2 \sim \mathcal{HV}_1(\bar{\sigma}^2, 3) \quad (3)$$

$$\bar{\sigma} \sim t_3(0, 3) \quad (4)$$

$$\ln \eta_{\{i,j\}} = \ln \theta_{m,t} + \varepsilon_g \quad (5)$$

$$\varepsilon_g \sim \text{Normal}(0, \tau_\varepsilon^2) \quad (6)$$

$$\tau_\varepsilon \sim t_3(0, 2) \quad (7)$$

$$\ln \theta_{m,t} = \mu + \alpha_t + \beta_m + \gamma_{m,t} \quad (8)$$

$$\mu \sim \text{Normal}(0, 2) \quad (9)$$

$$\vec{\alpha} \sim \mathcal{GP}(0, \ell_{T,1}, \tau_\alpha^2, \mathbf{D}_T) \quad (10)$$

$$\vec{\beta} \sim \mathcal{GP}(0, \ell_{M,1}, \tau_\beta^2, \mathbf{D}_M) \quad (11)$$

$$\text{vec}(\gamma) \sim \text{MultiNormal}(0, \Sigma_\gamma) \quad (12)$$

$$\Sigma_\gamma = \tau_\gamma^2 \left(\exp\left(-\frac{\mathbf{D}_T}{\ell_{T,2}}\right) \otimes \exp\left(-\frac{\mathbf{D}_M}{\ell_{M,2}}\right) \right) \quad (13)$$

$$\vec{\ell}_M \sim \text{Weibull}(4, 3.5) \quad (14)$$

$$\sqrt[2.2]{\vec{\ell}_T} \sim \text{Weibull}(4, 1.1) \quad (15)$$

$$\vec{\tau} \sim t_3(0, 2) \quad (16)$$

$$S_{m,t} = \begin{cases} 1 & \theta_{m,t} \geq 1 \\ 0 & \theta_{m,t} < 1 \end{cases} \quad (17)$$

$$\hat{p}_{m,t} = \frac{1}{K} \sum_{k=1}^K S_{m,t,k} \quad (18)$$

$$\hat{n}_t = \frac{1}{K} \sum_{k=1}^K \sum_{m=1}^{12} S_{m,t,k} \quad (19)$$

From the posterior distribution of the model, we estimated the probability of spawning in a given month and year ($\hat{p}_{m,t}$) and the width of the spawning period (\hat{n}_t). K is the number of MCMC iterations that the model was run and $S_{m,t,k}$ indicates spawning in month m , year t , iteration k . Due to the complexity of this model, additional sampling was necessary. We ran 6 independent chains with 2000 warmup and 1500 sampling iterations, for a total of 9000 posterior draws.

Spawning Interval

Spawning interval (SI) is the reciprocal of the proportion of fish spawning (p). Assuming that whether a given fish is spawning is independent of the spawning status of the fish around it, then spawning status can be considered a Bernoulli trial and the number of spawning fish can be described with a binomial distribution. Under these conditions, standard errors can be estimated from published values of p as long as n is known, so the data can be analyzed with a modified logistic regression. For each method of assessing spawning interval (OM or POF), the raw data (y_i) was the number of fish with a spawning marker (OM or POF); the total number captured was n_i . Our summarized data was the logit-transformed proportion of fish spawning; means and standard errors were calculated as:

$$y_j = \text{logit}(p_j) = \ln \left(\frac{p_j}{1 - p_j} \right) \quad (20)$$

$$\sigma_{\bar{y},j} = \sqrt{\frac{1}{n_j p_j} + \frac{1}{n_j(1 - p_j)}} \quad (21)$$

While it should have been theoretically possible to recover the raw data from the summarized data, this often resulted in fractional numbers of spawning fish, so the logit-transformed approach was taken to avoid making unnecessary assumptions. Integrating the raw and summarized data for the spawning interval analysis was fairly straightforward, since the transformation used on the summarized data is a widely-used normal approximation that gives it the same scale and weight as the equivalent raw data (see Gelman et al, 2013, section 5.6 for a good example). Data were divided into groups that represented the data collected during one year in one region of the Gulf that was published in a single study. Each summarized data point was in its own group, but raw data could share the same group.

As with spawning seasonality, variation among groups was modeled with ε_g . Some groups contained multiple samples (e.g., spawning interval was assessed monthly). Variation within these groups was modeled with the partially pooled $\epsilon_{g,w}$, where w is the sub-group identifier. This parameter was set to 0 for groups with only a single data point. Variance within each group was partially pooled with $\tau_{\epsilon,w}^2$ parameters, which were themselves partially pooled with a hierarchical variance prior. Raw data were modeled with logistic regression, while summarized data were modeled with a normal regression.

We fit two versions of this model: one in which all data shared the same temporal trend, and one in which the time series were allowed to differ between the northeastern and northwestern Gulf. For data in region r , year t , group g , and sub-group w :

$$y_i \sim \text{Binomial}(n_i, \text{logit}^{-1}\eta_i) \quad (22)$$

$$\bar{y}_j \sim \text{Normal}(\eta_j, \sigma_{\bar{y},j}^2) \quad (23)$$

$$\eta_{\{i,j\}} = \theta_{r,t} + \varepsilon_g + \epsilon_{g,w} \quad (24)$$

$$\varepsilon_g \sim \text{Normal}(0, \tau_\varepsilon^2) \quad (25)$$

$$\epsilon_{g,w} \sim \text{Normal}(0, \tau_{\varepsilon,g}^2) \quad (26)$$

$$\tau_\varepsilon \sim t_3(0, 2) \quad (27)$$

$$\tau_{\varepsilon,g}^2 \sim \mathcal{HV}_3(\bar{\tau}_\varepsilon^2, 3) \quad (28)$$

$$\bar{\tau}_\varepsilon^2 \sim t_3(0, 2) \quad (29)$$

$$\vec{\theta}_r \sim \mathcal{GP}(\mu_r, \ell_r \tau_r^2, \mathbf{D}_T) \quad (30)$$

$$\mu_r \sim \text{Normal}(0, 3) \quad (31)$$

$${}^{2.2}\sqrt{\vec{\ell}_r} \sim \text{Weibull}(4, 1.1) \quad (32)$$

$$\tau_r \sim t_3(0, 2) \quad (33)$$

$$r = \begin{cases} 1 & \text{northeast region or in no-region model} \\ 2 & \text{northwest region when region effects are being estimated} \end{cases} \quad (34)$$

$$\hat{S}I_{r,t} = \frac{1 + \exp \theta_{r,t}}{\exp \theta_{r,t}} \quad (35)$$

$$\hat{S}I_g = \frac{1 + \exp(\theta_{r,t} + \varepsilon_g)}{\exp(\theta_{r,t} + \varepsilon_g)} \quad (36)$$

For the model in which the two regions shared a common trend, the r subscripts were treated as identical for all data. We calculated estimated annual spawning interval in each region and year ($\hat{S}I_{r,t}$) from the reciprocal of the inverse-logit of the posterior distribution of $\vec{\theta}$. We estimated the group-level spawning interval ($\hat{S}I_g$) in a similar manner. These correspond with the horizontal and vertical lines in Figure 4, respectively.

Batch Fecundity

Changes in relative batch fecundity (RBF) were analyzed entirely with raw data. The RBF was square-root transformed, then scaled to approximate normality with a mean of 0 and variance of 1. Data were grouped as indicated in the year column of Table 1. As with spawning interval, variation between the northeastern and northwestern regions was assessed by running two versions of the model: one with separate time series, one with shared time series. The meta-analysis included a partially pooled effect for among-group variability (ε_g). Due to a small number of abnormally large RBF values (even after the transformation), the likelihood was modeled with a Student's t distribution, which allows for a more robust estimate that minimizes the influence of the outliers on the parameter estimates. For data in region r , year t , and group g :

$$y_i \sim t_6(\eta_i, \sigma_g^2) \quad (37)$$

$$\sigma_g^2 \sim \mathcal{HV}_1(\bar{\sigma}^2, 3) \quad (38)$$

$$\bar{\sigma} \sim t_3(0, 3) \quad (39)$$

$$\eta_i = \theta_{t,r} + \varepsilon_g \quad (40)$$

$$\varepsilon_g \sim \text{Normal}(0, \tau_\varepsilon^2) \quad (41)$$

$$\tau_\varepsilon \sim t_3(0, 2) \quad (42)$$

$$\vec{\theta}_r \sim \mathcal{GP}(\mu_r, \ell_r \tau_r^2, \mathbf{D}_T) \quad (43)$$

$$\mu_r \sim \text{Normal}(0, 3) \quad (44)$$

$$\sqrt[2.2]{\vec{\ell}_r} \sim \text{Weibull}(4, 1.1) \quad (45)$$

$$\tau_r \sim t_3(0, 2) \quad (46)$$

$$r = \begin{cases} 1 & \text{northeast region or in no-region model} \\ 2 & \text{northwest region when region effects are being estimated} \end{cases} \quad (47)$$

$$(48)$$

The estimated mean fecundity ($R\hat{B}F_{r,t}$) required a bias correction to estimate. Let x_i be the untransformed RBF, y_i be the transformed RBF value used in the model, and \bar{y} and s_y represent the sample mean and standard deviation of \sqrt{x} . Since y_i is modeled with a t_ν -distribution, then $y_i = \mu + \varepsilon_i \sigma$, where μ is the expected value of y_i , σ is the scale of the residuals, and $\varepsilon_i \sim t_\nu(0, 1)$. With a little algebra, this can be solved for x_i :

$$\sqrt{x_i} = y'_i \quad (49)$$

$$y_i = \frac{y'_i - \bar{y}}{s_y} \quad (50)$$

$$y_i = \mu + \varepsilon_i \sigma \quad (51)$$

$$y'_i = s_y(\mu + \varepsilon_i \sigma) + \bar{y} \quad (52)$$

$$x_i = (y'_i)^2 = s_y^2(\mu + \varepsilon_i \sigma)^2 + 2\bar{y}s_y(\mu + \varepsilon_i \sigma) + \bar{y}^2 \quad (53)$$

$$x_i = s_y^2(\mu^2 + 2\mu\varepsilon_i\sigma + \sigma^2\varepsilon_i^2) + 2\bar{y}s_y\mu + 2\bar{y}s_y\varepsilon_i\sigma + \bar{y}^2 \quad (54)$$

We can now take the expectation of both sides. Because constants can be factored out of expectations, the only variables that need to be considered are the ones which contain ε_i . Since ε_i is a random variable with a standard Student's t -distribution and $\nu \geq 3$, $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \frac{\nu}{\nu-2}$. Finally, we can substitute these values for our model parameters to determine mean RBF. For a data point in year t and region r and an unspecified group, $E(x_i)$ is $R\hat{B}F_{r,t}$, μ is $\theta_{r,t}$, σ is $\bar{\sigma}$, s_y and \bar{y} are properties of the data, and ν is 6.

$$E(x_i) = E(s_y^2(\mu^2 + 2\mu\varepsilon_i\sigma + \sigma^2\varepsilon_i^2)) + E(2\bar{y}s_y\mu) + E(2\bar{y}s_y\varepsilon_i\sigma) + E(\bar{y}^2) \quad (55)$$

$$E(x_i) = s_y^2(\mu^2 + 2\mu\sigma E(\varepsilon_i) + \sigma^2 E(\varepsilon_i^2)) + \bar{y}^2 + 2\bar{y}s_y\mu + 2\bar{y}s_y\sigma E(\varepsilon_i) \quad (56)$$

$$E(x_i) = s_y^2\left(\mu^2 + \sigma^2 \frac{\nu}{\nu-2}\right) + \bar{y}^2 + 2\bar{y}s_y\mu \quad (57)$$

$$R\hat{B}F_{r,t} = s_y^2\left(\theta_{r,t}^2 + \frac{3}{2}\bar{\sigma}^2\right) + \bar{y}^2 + 2\bar{y}s_y\theta_{r,t} \quad (58)$$

Model Validation

To verify that the models converged to a stationary posterior distribution, we calculated the split Gelman-Rubin diagnostic \hat{R} and effective sample size n_{eff} for each parameter. We also examined the trace plot of

the log posterior probability.

We assessed the fit of each model with posterior predictive checks. This involves using the expected value of each data point $E(y_i)$ and the posterior distribution to simulate new data, \tilde{y}_i . Each posterior sample produces a different value for \tilde{y}_i . If the model fits well, the real data will lie well within the distribution of \tilde{y} . This correspondence is best evaluated with graphical checks comparing y to \tilde{y} (Gelman et al, 2013). Additionally, we compared summary statistics calculated from each posterior sample's \tilde{y} with the real sample statistics. Because we had a mixture of individual and summarized data for spawning seasonality and spawning interval, individual data was averaged by group for the posterior predictive checks. For fecundity, we presented both individual and group-level checks. Individual spawning interval data was also inverted and logit-transformed to be on the same scale as the summarized data.

To assess the robustness of the models to prior assumptions, we re-ran each model multiple times with different hyperparameters. For each of the three model classes (seasonality, spawning interval, and fecundity), we varied all of the τ , σ , ν , μ , and c (hierarchical variance concentration) parameters between 2^{-3} and 2^7 of their original scale. As it would be computationally prohibitive to fully explore all combinations of these parameters, we used latin hypercube sampling to select 240 sensitivity trails with hyperparameter combinations that approximated the entire grid of possibilities. The models were then re-run with these hyperparameters for 4 chains with 1000 post-warmup iterations each. Prior sensitivity was assessed by comparing the posterior predictive summary statistics.

Validation Results

Convergence checks

Models are considered to be convergent if $\hat{R} < 1.1$, n_{eff} is reasonably large, and the log posterior is well-mixed. The models for spawning seasonality (Suppl. Figure 1), spawning interval (Suppl. Figures 2 - 3) and fecundity (Suppl. Figure 4) all converged.

Posterior Predictive Checks

The spawning seasonality model did a good job estimating average GSI (Suppl. Figures 5 - 6), although it had a tendency to overestimate large values (GSI > 2.5). This is not particularly concerning, since these values are already well over the spawning threshold (GSI = 1) and the model performs well in that region. The OM spawning interval model slightly underestimated the average transformed SI for values above one (Suppl. Figures 7 - 8); this corresponds to a slight underestimation of spawning intervals less than 2. However, summary statistics matched well with the real data (Suppl. Figures 9 - 10). This concern was negligible with the POF methods (Suppl. Figures 11 - 14).

Both fecundity models produced similar summary statistics to the data (Suppl. Figures 15 - 16). While the minimum value is slightly underestimated, the actual difference (around 0.5) is biologically insignificant. These models did not do a particularly good job of predicting individual fecundities (Suppl. Figures 17 - 18), since there was a large amount of variation among individuals (ranging over two orders of magnitude) and we did not have consistent data on individual predictor variables. However, both models predicted group-level fecundity (Suppl. Figures 19 - 20) quite well.

Study-level effects showed no indication of bias over time or region (Suppl. Figures 21 - 24).

Sensitivity Analyses

Our sensitivity analyses showed qualitative agreement across the different sensitivity trials for seasonality (Suppl. Figure 25), spawning interval (Suppl. Figures 26 - 29), and fecundity (Suppl. Figures 30 - 31). This indicates that the models are relatively robust to prior assumptions. With these figures, each panel indicates a statistic calculated from the posterior predictive distribution of the repeated run. Each gray line is the density estimate of that statistic for one of the sensitivity trials. Some hyperparameters resulted in non-convergent models and were not included in the figures. Spawning interval minimum and maximum values were somewhat sensitive to prior assumptions. This has little practical importance, as the values presented in Suppl. Figures 28 - 29 are on a logit-transformed scale, where changes in magnitudes that are greater than 2 result in minimal changes in the real scale.

Literature Cited

- Carpenter, B., A. Gelman, M.D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software* 76(1), 32 p. <https://doi.org/10.18637/jss.v076.i01>
- Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. 2013. *Bayesian Data Analysis, 3rd edition*. Boca Raton: CRC Press/Taylor & Francis, 639 p.
- McElreath, R. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton: CRC Press/Taylor & Francis, 469 p.
- R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Stan Development Team. 2017. *Stan Modeling Language Users Guide and Reference Manual*, Version 2.17.0. <http://mc-stan.org>