

## *Ecological Archives* A020-051-A1

**J. R. Thomson, W. J. Kimmerer, L. R. Brown, K. B. Newman, R. Mac Nally, W. A. Bennett, F. Feyrer, and E. Fleishman. 2010. Bayesian change-point analysis of abundance trends for pelagic fishes in the upper San Francisco Estuary. *Ecological Applications* 20: 1431–1448.**

Appendix A. Details of prior distributions used in change-point and associated regression models.

*Priors for the number, timing and magnitude of change-points.*

We specified binomial prior distributions for the number of step changes  $k_\alpha$  and the number of trend changes  $k_\beta$ :  $k_\alpha \sim \text{Bin}(k_{\max}, \pi)$ ,  $k_\beta \sim \text{Bin}(k_{\max}, \pi)$ , where  $k_{\max}$  is the maximum possible number of each type of change, and  $\pi$  is the binomial probability. Under these priors, the prior probability that there will be  $k_\alpha$  step changes was

$$p(k_\alpha) = \binom{k_{\max}}{k_\alpha} \times \pi^{k_\alpha} (1 - \pi)^{(k_{\max} - k_\alpha)} = \frac{k_{\max}! \pi^{k_\alpha} (1 - \pi)^{(k_{\max} - k_\alpha)}}{k_\alpha! (k_{\max} - k_\alpha)!}.$$

The priors for the timing of change-points  $\delta$ (step) and  $\theta$ (slope) were conditional on  $k_\alpha$  and  $k_\beta$ , respectively. All models with a given number of change-points were equally probable a priori. That is, all combinations of  $k_\alpha$  step changes were treated as equally probable. For step changes, the prior probability of a specific combination of  $k_\alpha$  change-points (e.g. in 1973 and 1999, given that  $k_\alpha=2$ ) was

$$p(\delta | k_\alpha) = \binom{T}{k_\alpha}^{-1} = \frac{k_\alpha! (T - k_\alpha)!}{T!},$$

where  $T$  is the number of possible change points (= number of survey years – 1). Therefore, the prior probability for a particular combination of step changes was

$$p(\delta) = p(\delta | k_\alpha) \cdot p(k_\alpha) = \frac{k_\alpha!(T - k_\alpha)!}{T!} \cdot \frac{k_{\max}! \pi^{k_\alpha} (1 - \pi)^{(K_{\max} - k_\alpha)}}{k_\alpha!(k_{\max} - k)!} = \frac{k_{\max}!(T - k_\alpha)! \pi^{k_\alpha} (1 - \pi)^{(K_{\max} - k_\alpha)}}{T!(k_{\max} - k)!}$$

The probability that any specific year  $y$  is included in the vector  $\delta$  of  $k_\alpha$  change-points is  $k_\alpha / T$ . Accordingly, the prior probability for a change-point at any give year  $y$  was

$$p(y) = \int_{k=1}^{K_{\max}} p(y | k_\alpha) \cdot p(k_\alpha) \cdot dk_\alpha = \sum_{k=1}^{K_{\max}} \left( \frac{k_\alpha}{T} \cdot \frac{k_{\max}! \pi^{k_\alpha} (1 - \pi)^{(K_{\max} - k_\alpha)}}{k_\alpha!(k_{\max} - k)!} \right) = \frac{\pi K_{\max}}{T}$$

These model priors are uninformative about the timing of change-points, but are somewhat informative about the number of change-points. A maximum number of change-points is specified, and there is a prior expectation of  $\pi k_{\max}$  step and slope changes. We used  $\pi = 0.5$  and  $k_{\max} = 4$ . Importantly, the prior also allows for no change points. In fact, the model with no change-points has higher prior probability ( $= 0.5^4 \times 0!(T - 0)!/T! = 0.5^4$ ) than any other single model (i.e., any specific combination of  $\geq 1$  change-points).

The uninformative priors used for all other model parameters are shown in Tables A1 through A3.

TABLE A1. Parameters and their prior distributions for trend models.

**The model**

$$y_t \sim \text{Normal}(n_t, \sigma_{0t}^2) \tag{A.1}$$

$$n_t \sim \text{LogNormal}(\alpha_t + f_1(t), \sigma_p^2) \tag{A.2}$$

$$\alpha_t = \alpha_1 + \sum_{j=1}^{k_\alpha} \chi_j I(t \geq \delta_j) \tag{A.3}$$

$$f_1(t) = \beta_1 t + \sum_{j=1}^{k_\beta} \beta_{[j+1]}(t - \theta_j)_+ \tag{A.4}$$

Parameters	Description	Prior	Comments
$\sigma_{0t}^2$	Variance of observation error at time $t$ :	Point estimate calculated from catch	uncertainty of estimated

		data	abundance $y_t$
$\sigma_p^2$	Variance of process error	IG(0.001,1000)	Uninformative
$\alpha_1$	Estimated initial abundance	N(0,10000)	Uninformative
$k_\alpha$	Number of step changes in abundance.	Bin(4, 0.5)	Maximum of 4 step changes, prior expectation of 2.
$\delta_j, j = 1, \dots, k_\alpha$	year when the $j^{\text{th}}$ step change occurred	$p(\delta_j   k_\alpha) = \frac{k_\alpha!(T - k_\alpha)!}{T!}$ $T = \text{survey years} - 1.$	Prior is conditional on $k_\alpha$ . All possible combinations of $k_\alpha$ Step changes are equally likely
$\chi_j, j = 2, \dots, k_\alpha$	Vector of step change sizes occurring at the $k_\alpha$ change-points	N(0, $\sigma_\alpha^2$ )	Uninformative, exchangeable prior
$\sigma_\alpha^2$	Variance of the normal distribution of step change sizes	$\frac{(\ln y_{\max} - \ln y_{\min})}{1.96}$	Point estimate derived from data range
$k_\beta$	Number of changes in the slope; number of times linear trend in abundance changes	Bin(5, 0.5)	Maximum of 5 changes-in-slope. Prior expectation of 2.5.
$\theta_j, j = 1, \dots, k_\beta$	year when the $j^{\text{th}}$ trend change (change in slope) occurred	$p(\theta_j   k_\beta) = \frac{k_\beta!(T - k_\beta)!}{T!}$	All possible combinations of $k_\beta$ changes-in-slope are equally likely
$\beta_j, j = 1, \dots, k_\beta$	Slope of linear trend	N(0, $\sigma_\beta^2$ )	Uninformative, exchangeable prior
$\sigma_\beta^2$	Variance of the normal distribution of linear trend parameters	$\frac{(\ln y_{\max} - \ln y_{\min})}{4 \times 1.96}$	Point estimate derived from data range

Distributions: N= Normal, Bin = Binomial, IG = inverse Gamma. In WinBUGS, Normal distributions are specified with precisions (1/variance) and Gamma distributions with inverse scale parameters, e.g., Gamma(0.001,1000) is specified as dgamma(0.001,0.001).

TABLE A2. Parameters and their prior distributions for variable selection models.

**The model**

$$y_t \sim \text{Normal}(n_t, \sigma_{0t}^2)$$

$$n_t \sim \text{Lognormal} \left( \alpha + \sum_{j=1}^Q \sum_{m=1}^{k_j} \beta_{jm} (x_{jt} - \phi_{jm})_+ + \rho \log n_{t-1}, \sigma_p^2 \right) \quad (\text{A.5})$$

Parameters	Description	Prior	Comments
$\sigma_{0t}^2$	Variance of observation error at time $t$ :	Point estimates calculated from catch data	uncertainty of estimated abundance $y_t$
$\sigma_p^2$	Variance of process error	IG(0.001,1000)	Uninformative
$\alpha$	Estimated initial abundance	N(0,10000)	Uninformative
$k_j, j = 1, \dots, Q$	Number of linear segments in piecewise linear spline for variable $j$	Cat( $p_0, p_1, p_2, p_3$ ) $p_0 = 0.5, p_1 = 0.3,$ $p_2 = 0.1, p_3 = 0.1$ $p_n$ is probability of $n$ segments.	$p_0 = 0.5$ is prior probability of no effect of variable $j$ Max. segments (knots) is 3
$\phi_{jm}$ $m=1, \dots, k_j$ $j = 1, \dots, Q$	Knot value for $m^{\text{th}}$ segment of linear spline for variable $j$	Cat( $p_1, \dots, p_{10}$ ) $p_n = 0.1$ is probability of knot at $n^{\text{th}}$ candidate value. There were 10 evenly spaced candidate knot values starting at $\min(x_j)$ .	Categorical prior with 10 discrete knots used to limit model space (hence increase speed of MCMC) in variable selection

$\beta_{jm}$ , $m=1, \dots, k_j$ , $j = 1, \dots, Q$	Linear coefficient for $m^{\text{th}}$ segment of linear spline for variable $j$	$N(0, \sigma_\beta^2)$	Uninformative, exchangeable prior for non-zero coefficients
$k_{Q+1}$	Binary indicator for inclusion of autocorrelation term $\rho$	$\text{Bin}(1, 0.5)$	
$\rho$	Autocorrelation coefficient	$N(0, \sigma_\beta^2)$	
$\sigma_\beta$	Standard deviation of the non-zero coefficients.	$\sigma_\alpha =  \zeta  \times \sigma_z^{-0.5}$ $\zeta \sim N(0, A)$ $\sigma_z \sim G(0.5, 2)$ $A=0.5$	Half-Cauchy prior

Distributions: N= Normal, Bin = Binomial, IG = inverse Gamma, Cat = Categorical (equivalent to Multinomial with  $n = 1$ ).

TABLE A3. Parameters and their prior distributions for covariate-conditioned change-point models.

**The model**

$$y_t \sim \text{Normal}(n_t, \sigma_{O_t}^2)$$

$$n_t \sim \text{Lognormal} \left( \alpha_t + \sum_{j=1}^q \sum_{m=1}^{k_j} \beta_{jm} (x_{jt} - \phi_{jm})_+ + \rho \log n_{t-1}, \sigma_p^2 \right) \quad (\text{A.6})$$

In the single species model,

$$\alpha_t = \alpha_1 + \sum_{j=1}^{k_\alpha} \chi_j I(t \geq \delta_j)$$

In the multi-species model,  $\alpha_t$  for species  $s$  (denoted  $\alpha_{st}$  in text), was

$$\alpha_t = \alpha_{s1} + \sum_{j=1}^{k_{s\alpha}} \chi_{sj} I(t \geq \delta_{sj}) + \sum_{l=1}^{k_{c\alpha}} \psi_l I(t \geq \zeta_l) \quad (\text{A.7})$$

Parameters	Description	Prior	Comments
$\sigma_{0t}^2$	Variance of observation error at time $t$ : relative uncertainty of estimate, $y_{\text{obs}_t}$	Point estimates calculated from catch data	
$\sigma_p^2$	Variance of process error	IG(0.001,1000)	Uninformative
$\alpha_1$	Estimated initial abundance	N(0,10000)	Uninformative
$k_\alpha$ ( $k_{s\alpha}$ in multi-species model)	Number of step changes in abundance.	$k_\alpha \sim \text{Bin}(4, 0.5)$ $k_{s\alpha} \sim \text{Bin}(2, 0.5)$	Maximum of 4 (2) step changes, prior expectation of 2 (1)
$\delta_j, j = 1, \dots, k_\alpha$	year when the $j^{\text{th}}$ step change occurred	$p(\delta_j   k_\alpha) = \frac{k_\alpha!(T - k_\alpha)!}{T!}$  $T = \text{survey years} - 1.$	Prior is conditional on $k_\alpha$ . All possible combinations of $k_\alpha$ Step changes are equally likely
$\chi_j, j = 2, \dots, k_\alpha$	Size of $j^{\text{th}}$ step change	N(0, $\sigma_\alpha^2$ )	Uninformative, exchangeable prior
$\sigma_\alpha^2$	Variance of the normal distribution of step change sizes	$\frac{(\ln y_{\text{max}} - \ln y_{\text{min}})}{1.96}$	Point estimate derived from data range
$k_j, j = 1, \dots, Q$	Number of linear segments in piecewise linear spline for variable $j$	$1 + \kappa_j$ $\kappa_j \sim \text{Bin}(3, 0.3)$	At least 1 segment (linear effect), up to 3 changes in slope
$\phi_{jm}$ $m = 1, \dots, k_j.$ $j = 1, \dots, Q$	Knot value for $m^{\text{th}}$ segment of linear spline for variable $j$	U(min( $x_j$ ), max( $x_j$ ))	Uniform prior for continuous knots

$\beta_{jm}$ , $m=1, \dots, kj$ , $j = 1, \dots, Q$	Linear coefficient for $m^{\text{th}}$ segment of linear spline for variable $j$	$N(0, \sigma_\beta^2)$	Uninformative, exchangeable prior for non-zero coefficients
$\rho$	Autocorrelation coefficient	$N(0, 0.001)$ for striped bass 0 for all other species	Uninformative. Included only for striped bass
$\sigma_\beta$	Standard deviation of the non-zero coefficients.	$\sigma_\beta =  \zeta  \times \sigma_z^{-0.5}$ $\zeta \sim N(0, A)$ $\sigma_z \sim G(0.5, 2)$ $A=0.04$	Half-Cauchy prior
$k_{C\alpha}$	Number of step changes common to all species	$\text{Bin}(2, 0.5)$	Multi-species model only
$\zeta_j, j = 1, \dots, k_{C\alpha}$	year when the $j^{\text{th}}$ common step change occurred	$p(\psi_j   k_{C\alpha}) = \frac{k_{C\alpha}!(T - k_{C\alpha})!}{T!}$  $T = \text{survey years} - 1.$	Multi-species model only
$\psi_j, j = 1, \dots, k_{C\alpha}$	Size of $j^{\text{th}}$ common step change	$\frac{(\text{mean}(\ln y_{\max}) - \text{mean}(\ln y_{\min}))}{1.96}$	Multi-species model only

Distributions: N= Normal, Bin = Binomial, IG = inverse Gamma, U = Uniform, G= Gamma (note G(0.5,2) is equivalent to a  $\chi^2$  distribution with 1 d.f.)

### *Sensitivity of change-point detection to prior distributions*

Absolute posterior probabilities of change-points obviously will be sensitive to the prior distributions on the numbers of change-points  $k_\alpha$  and  $k_\beta$  (in trend models). Posterior probabilities for change-points in particular years will generally increase with the prior expectation for the number of change-points. Clearly it is important to use prior distributions that reflect appropriate definitions of change-points and plausible expectations about the numbers of such change points (e.g., priors that allowed up to 40 change-points would not be sensible). Across a range of sensible priors for  $k_\alpha$  and  $k_\beta$  (e.g.  $k_{\max} = 4$  vs.  $k_{\max} = 2$ ) the relative probabilities of change-points and odds ratios (which mostly remove the influence of prior

model probabilities) generally should be consistent. Therefore, inferences about the timing of change-points will rarely be sensitive to the exact choice of priors for  $k_\alpha$  or  $k_\beta$  (within reasonable limits). This was certainly true in sensitivity tests for our trend models in which we fitted models with  $k_{max} = 1, 2, 4$  and  $6$  (for both  $k_\alpha$  and  $k_\beta$ ).

The prior variances  $\sigma_\alpha^2$  and  $\sigma_\beta^2$  control the possible magnitudes of any change-points in trend models (and covariate-conditioned change-point models for  $\sigma_\alpha^2$ ). Posterior model probabilities can also be sensitive to these parameters, because the degree to which integrated likelihoods penalize complexity largely depends on the prior variance for model parameters. Larger prior variances will tend to favor less complex models, and vice versa. In regression models, the prior variance essentially specifies the expected magnitudes of effects. Thus, large prior variances will favor models with few large effects, whereas small prior variances will favor models that include a greater number of variables with relatively small effects. For change-point models, this equates to a choice between favoring few large change-points, or relatively many (up to  $k_{max}$ ) smaller changes.

We tested the sensitivity of posterior probabilities for change-points to prior variances by fitting models with point priors set at 0.5, 1, and 2 times the data-range values described in the main text (and table A1). We also fitted models with hyper-priors on the variances or standard deviations  $\sigma_\alpha$  and  $\sigma_\beta$ . This approach reflects prior uncertainty (ignorance) about the expected magnitudes of any effects (e.g., change-points, covariate effects). We fitted models using three different hyper-prior specifications discussed by Gelman 2006 (inverse uniform on standard deviations, inverse Gamma on variances, and Half-Cauchy priors), each with 3 different scale parameters that define the credible effect sizes (Table A4). Results generally were consistent in relative probabilities and odds ratios for change-points in particular years, and invariably led to consistent inferences about the most probable change-points. The absolute probabilities of change-points were generally lower with the hyper-priors because these placed relatively more prior weight on large effect sizes, including some extreme values.

TABLE A4. Priors used in sensitivity analysis for change-point parameters.

Prior name	Details	Scale parameters for $\sigma_\alpha$	$\sigma_\beta$ scales
point	$\sigma_\alpha^2 = (scale / 1.96)^2$	$scale=range/2, range, 2\times range$	$scale/4$
Gamma	$\sigma_\alpha^2 \sim \text{InverseGamma}(a, 1/a)$	$a = 0.1, 0.01, 0.001$	$a$
Uniform	$\sigma_\alpha \sim \text{Uniform}(0, 0.8\times scale)$	$scale=range/2, range, 2\times range$	$scale/4$

Half-Cauchy	$\sigma_\alpha =  \zeta  \times \sigma_z^{-0.5}$ $\zeta \sim N(0, 100/scale^2)$ $\sigma_z \sim \text{Gamma}(0.5, 2)$	$scale = range/2, range, 2 \times range$	$scale/4$
-------------	--	--	-----------

$$range = \ln y_{\max} - \ln y_{\min}$$

### *Sensitivity of variable selection model to prior distributions*

In variable selection models, posterior model probabilities can be sensitive to the prior on regression coefficients  $\beta$ . We used a Half-Cauchy prior (see Table B2) with scale parameter chosen so that ca. 90% and 95% of the prior probability mass was in the interval (-1,1) and (-2,2) respectively. This prior puts most weight on more plausible coefficients while still allowing larger effects. We tested sensitivity of model posterior probabilities to the prior on  $\beta$  by fitting models with a range specifications for the prior variance  $\sigma_\beta^2$  of the regression coefficients (the jump interface in WinBUGS allows only exchangeable normal priors for the vector of coefficients  $\beta$ ). We varied the scale parameter of the Half-Cauchy prior and fitted models with a range of different priors on  $\sigma_\beta^2$ , including point estimates (0.25, 0.5, 1, 2), uniform on  $\sigma_\beta$  with upper limits (0.5, 1, 2, and 5). and inverse Gamma (0.01,0.01). We also implemented an approximation to the unit information prior (corresponding to Bayesian Information Criterion penalty when all models are equally probable, George and Foster 2000). Posterior model probabilities (hence probabilities of variable inclusion,  $\Pr(k_j > 0)$ ) varied predictably with the prior (more diffuse priors yielded lower probabilities), but the relative values among variables were consistent.  $\Pr(k_j > 0)$  values for variables with strongest effects (e.g., spring X2 for longfin smelt, water clarity for striped bass) always were high ( $> 0.9$ ) regardless of the prior used, and the set of variables with  $\Pr(k_j > 0) > 0.75$  was generally consistent among different prior specifications (though  $\Pr(k_j > 0)$  for some variables varied between 0.7 and 0.85).  $\Pr(k_j > 0)$  values for winter exports in the delta smelt and winter and spring exports in threadfin shad models were the most sensitive to prior specifications. This sensitivity to priors suggests that only relatively small effects of winter exports on abundances of fishes are supported by the data.

We also tested the sensitivity of odds ratios to prior probabilities of inclusion (i.e., to prior  $\Pr(k_j > 0)$ ) by increasing the probability of 0 in the categorical prior for the number of linear segments in nonlinear variable selection models. A consistent set of variables with odds ratio  $> 3$  emerged from each analyses.

*Note on change-point detection in autoregressive models*

The inclusion of an autoregressive term,  $\rho n_{t-1}$ , in change-point models alters the interpretation of parameters and therefore complicates the detection and interpretation of change-points. In the covariate condition change-point model (Eq. A.6), if  $\rho = 0$ , then  $e^\alpha$  is the initial abundance, and a step change in year  $y$  is modelled well by a new intercept value for year  $y$  and all subsequent years (as in Eq. A.3). But if  $\rho = 1$ , then  $e^\alpha$  is the proportional change in abundance from year  $y-1$  to year  $y$ , and a sustained change in  $\alpha$  (Eq. A.3) would model a trend change (a change in the annual rate of change in  $y$ ). With  $\rho = 1$  a step change in year  $y$  is better modelled by a change in  $\alpha$  at year  $y$  only, which can be achieved by modifying the  $\alpha$  submodel:

$$\alpha_t = \alpha_1 + \sum_{j=1}^{k_\alpha} \chi_j I(t = \delta_j) \quad (\text{A.8})$$

Either or both types of change-point (Eqns. A.3 or A.8) can be included in change-point models. But when  $0 < \rho < 1$  it is not clear which model is most appropriate because the interpretation of  $\alpha$ , and any change in it, is difficult. This difficulty of interpretation makes the specification of appropriately bounded priors (i.e., credible effect sizes) difficult, which in turn may affect the probability of detecting change-points.

LITERATURE CITED

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1:515–533.

George, E. I., and D. P. Foster. 2000. Calibration and empirical Bayes variable selection. *Biometrika* 87:731–747.