Supplementary Information

Occupancy Model

We assume that the probability that species i is detected at site j in occupancy interval k, $x[i, j, k]$, is drawn from a Bernoulli distribution (0 or 1) with probability ($y[i, j, k]$),

$$x[i, j, k] \sim \text{Bernoulli}(y[i, j, k]) \tag{S1}$$

where $y[i, j, k]$ is the product of detection probability ($p[i, j, k]$) and the unknown, but true occupancy state, $z[i, j, k]$,

$$y[i, j, k] = p[i, j, k] * z[i, j, k] \tag{S2}$$

The true but unknown site occupancy for species i at site j, $z[i, j, k]$ is equal to 1 if that site is occupied and 0 if it is not. We assume that this true site occupancy is drawn from a Bernoulli distribution with mean equal to the species' occupancy probability at that site,

$$z[i, j, k] \sim \text{Bernoulli}(\psi[i, j, k]) \tag{S3}$$

Both occupancy probability, ψ, and detection probability, p, are formulated as functions time.

We model occupancy as:

$$\begin{aligned}\text{logit}(\psi[i,j,k]) =& \psi_0 + \\ & \psi_{\text{sp}}[i] + \\ & \psi_{\text{OI}}[i] \times k\end{aligned} \tag{S4}$$

where ψ_0 denotes the baseline occupancy (on the linear scale), $\psi_{\text{sp}}[i]$ is a random species-specific intercept, and $\psi_{\text{OI}}[i]$ is a random species-specific effect of occupancy interval (a positive value of $\psi_{\text{OI}}[i]$ would indicate that species i is increasing in occupancy through time). We assume that $\psi_{\text{sp}}[i]$ and $\psi_{\text{OI}}[i]$ are normally distributed, such that:

$$\begin{aligned}\psi_{\text{sp}}[i] &\sim \mathcal{N}(0, \sigma_{\psi,\text{sp}}) \\ \psi_{\text{OI}}[i] &\sim \mathcal{N}(\mu_{\psi,\text{OI}}, \sigma_{\psi,\text{OI}})\end{aligned} \tag{S5}$$

Then, we model detection probability as:

$$\begin{aligned}\text{logit}(p[i,j,k]) =& \; p_0 + \\ & p_{\text{sp}}[i] + \\ & p_{\text{site}}[j,k] + \\ & p_{\text{OI}} \times k\end{aligned} \tag{S6}$$

where p_0 is the baseline detection probability (on the linear scale), $p_{\text{sp}}[i]$ is a random species-specific intercept, $p_{\text{site}}[j,k]$ is a random site-specific intercept that varies by occupancy interval, and p_{OI} is an overall effect of occupancy interval. $p_{\text{site}}[j,k]$ allows spatiotemporal variability in detection probability (changing across sites and occupancy intervals), and helps account for the variation that is inherent in sample effort across

space and time in unstructured historical datasets. p_{OI} allows detection to change systematically through time, as has likely occurred in many groups as sampling techniques have improved, for example. p_{sp} and p_{site} are assumed to be normally distributed, such that:

$$\begin{aligned} p_{\text{sp}}[i] &\sim \mathcal{N}(0, \sigma_{p,\text{sp}}) \\ p_{\text{site}}[j,k] &\sim \mathcal{N}(0, \sigma_{p,\text{site}}), \end{aligned} \tag{S7}$$

We used uninformative priors for all parameter values with normal distributions for ψ_0, $\mu_{\psi,\text{OI}}$, p_0, p_{OI}, and uniform distributions for $\sigma_{\psi,\text{sp}}$, $\sigma_{\psi,\text{OI}}$, $\sigma_{p,\text{sp}}$, $\sigma_{p,site}$.

We ran all models for $100,000$ iterations, with a burn in period of 1000 iterations, after burn in iterations were thinned every 100 iterations, and we ran 3 chains of the MCMC. This ensured that across all of our simulations, most parameters converged (r-hat < 1.1), and this convergence was independent of parameter space and model choice.

Parameters varied in the simulations	**Values**
Number of simulated occupancy intervals K	10
Number of occupancy intervals aggregated in the data processing workflow K	2, 5, 10
Probability of community visits ρ_{com}	0, 0.25, 0.5, 0.75, 1
Visitation probability $\mu_{\nu,\text{OI}}$/ η_{OI}	-0.1, 0, 0.1
Species ranges	no species ranges (all species can be found in all sites), species ranges
Parameters values used in the simulations	**Values**
Number of species	50
Number of sites	100
Number of visits intervals per occupancy interval	3
Detection parameters	
p_0	-0.5
$\sigma_{p,\text{site}}$	1.5
$\sigma_{p,\text{sp}}$	0.5
p_{OI}	0
Occupancy parameters	
ψ_0	0
$\sigma_{\psi,\text{sp}}$	0.5
$\sigma_{\psi,\text{OI}}$	0.2
$\mu_{\psi,\text{OI}}$	0
Visitation parameters	
$\mu_{\nu,0}$, η_0	0

Table S1: All of the parameter values used in the simulations of occurrence data.

Supplementary Results

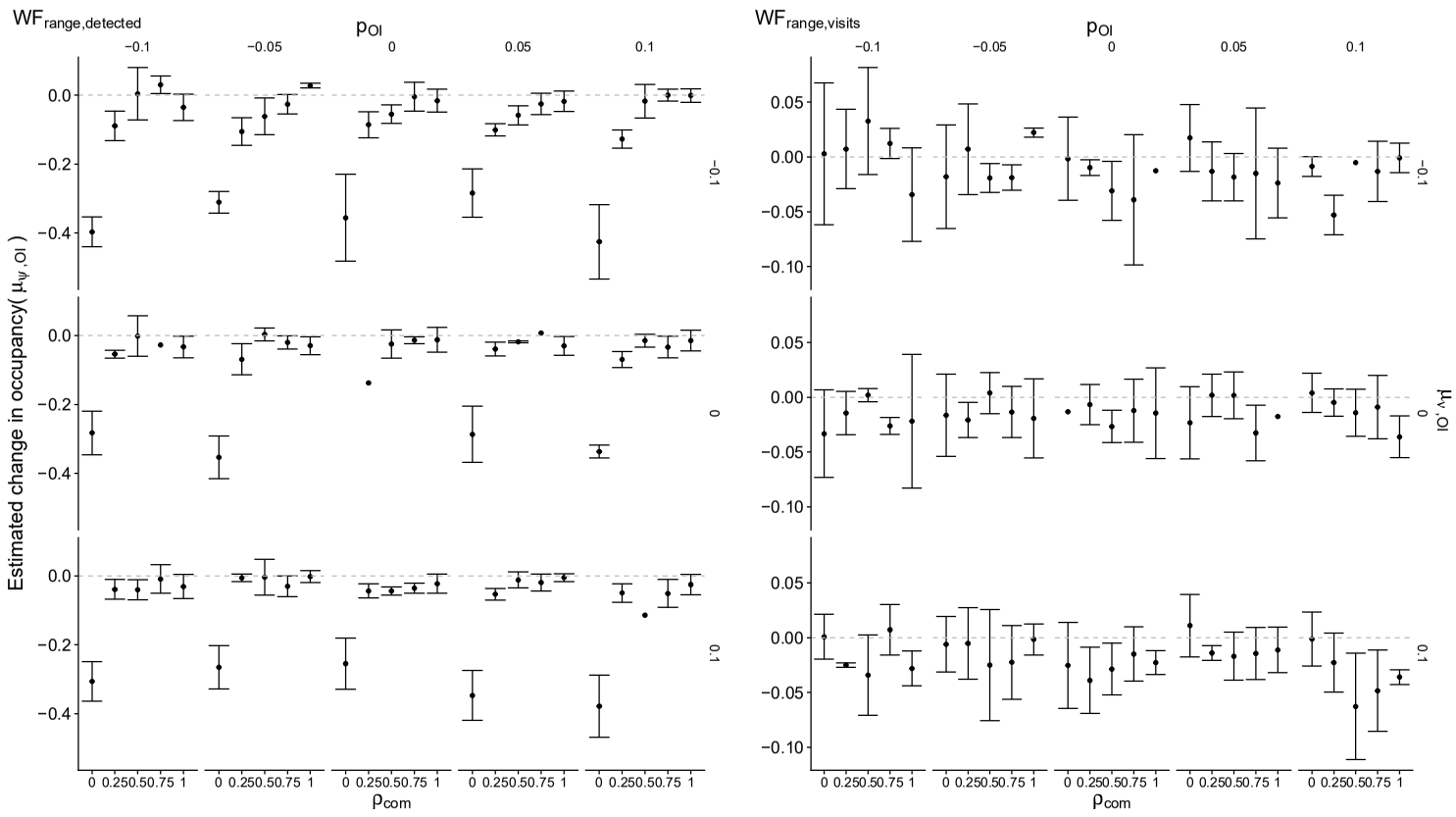

Figure S1: Estimated change of occupancy through time $\mu_{\psi,\mathrm{OI}}$, as a function of changes in detection through time (p_{OI}), visit history through time ($\mu_{\nu,\mathrm{OI}}$), and the probability of community visits (ρ_{com}). Points are means across ten replicates and error bars are the standard deviation across the replicates. The true value of $\mu_{\psi,\mathrm{OI}}$ is represented by the dashed grey line and is equal to zero.

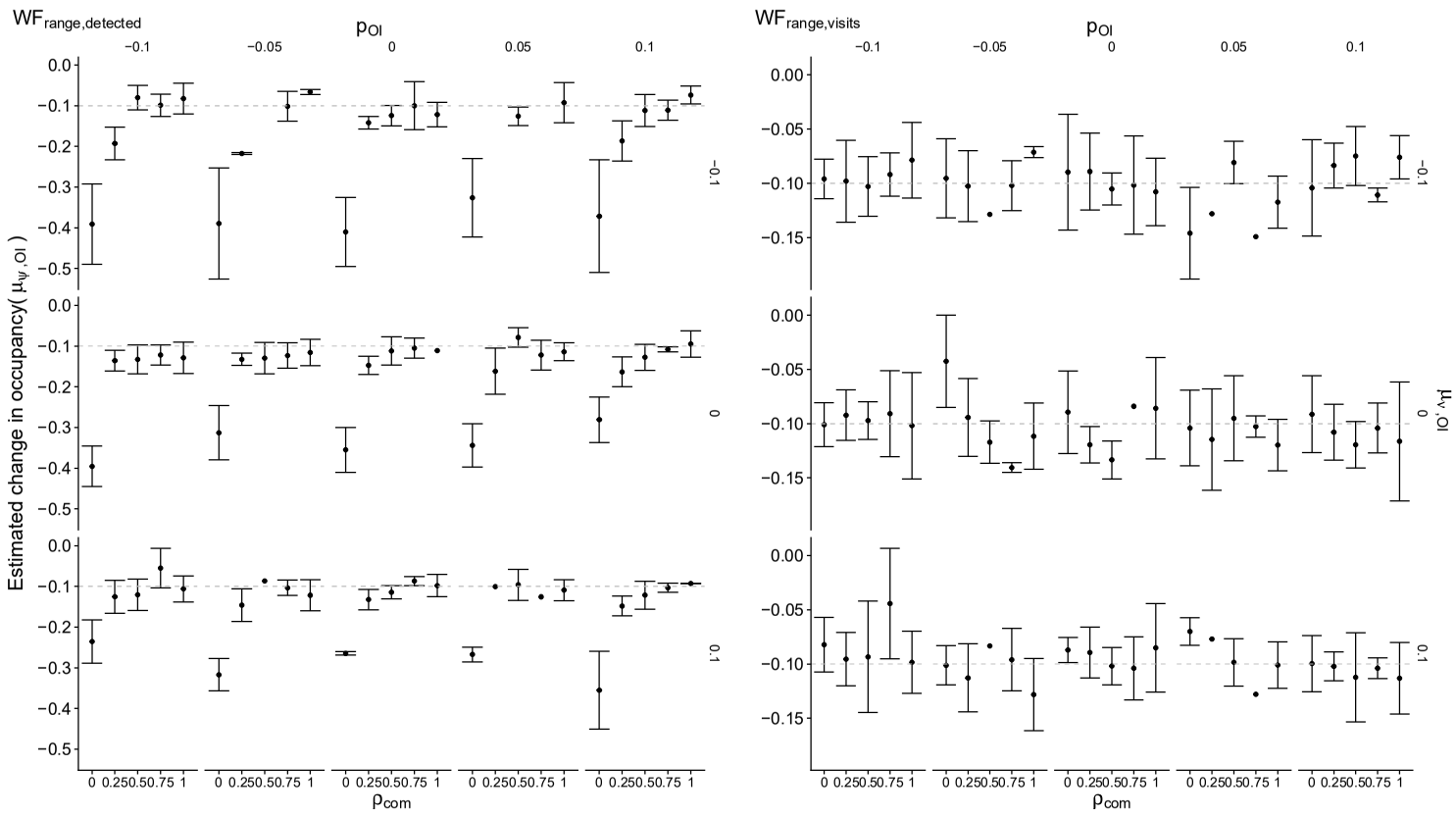

Figure S2: Estimated change of occupancy through time $\mu_{\psi,\mathrm{OI}}$, as a function of changes in detection through time (p_{OI}), visit history through time ($\mu_{\nu,\mathrm{OI}}$), and the probability of community visits (ρ_{com}). Points are means across ten replicates and error bars are the standard deviation across the replicates. The true value of $\mu_{\psi,\mathrm{OI}}$ is represented by the dashed grey line and is equal to -0.1.

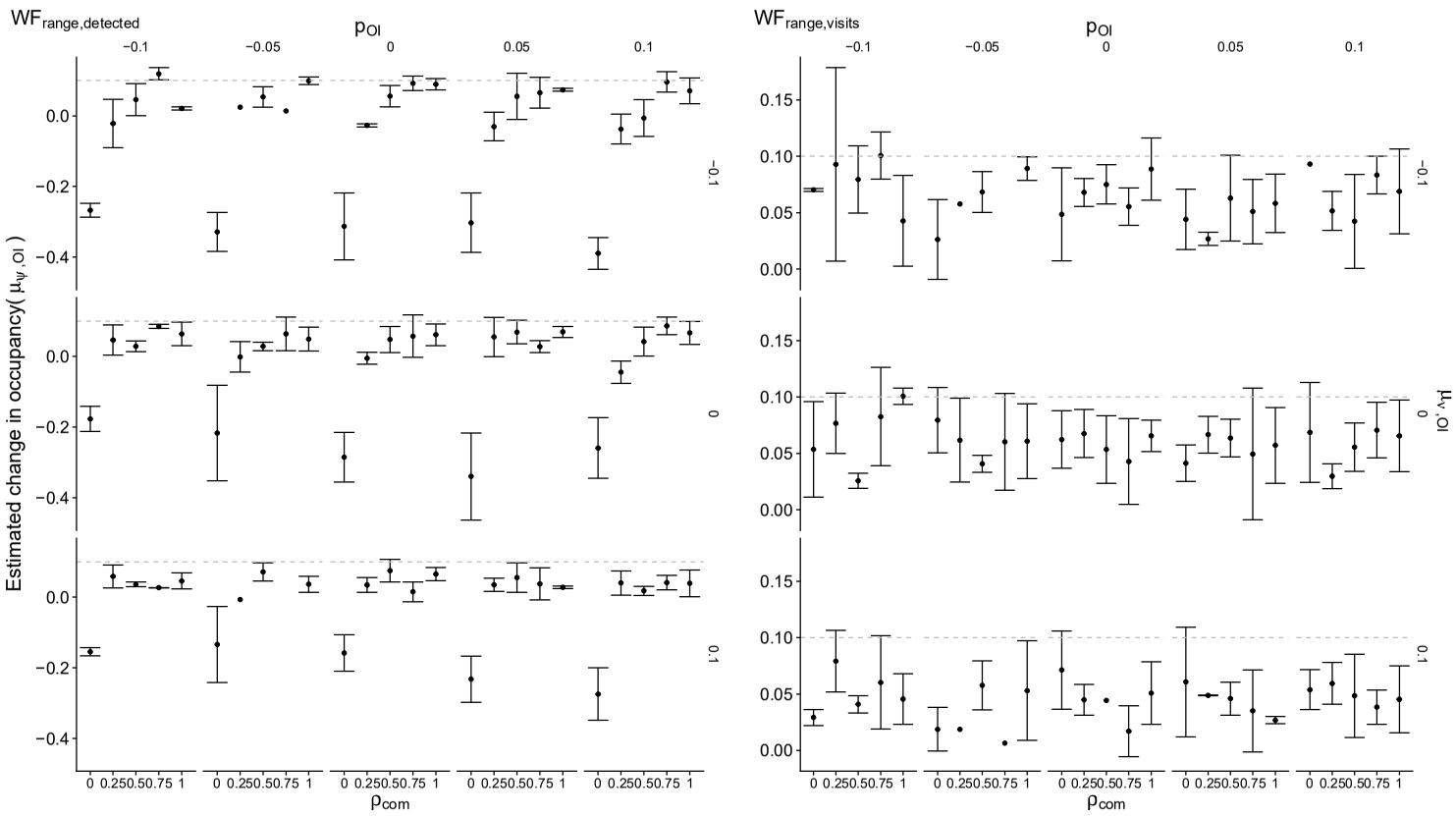

Figure S3: Estimated change of occupancy through time $\mu_{\psi,\mathrm{OI}}$, as a function of changes in detection through time (p_{OI}), visit history through time ($\mu_{\nu,\mathrm{OI}}$), and the probability of community visits (ρ_{com}). Points are means across ten replicates and error bars are the standard deviation across the replicates. The true value of $\mu_{\psi,\mathrm{OI}}$ is represented by the dashed grey line and is equal to 0.1.

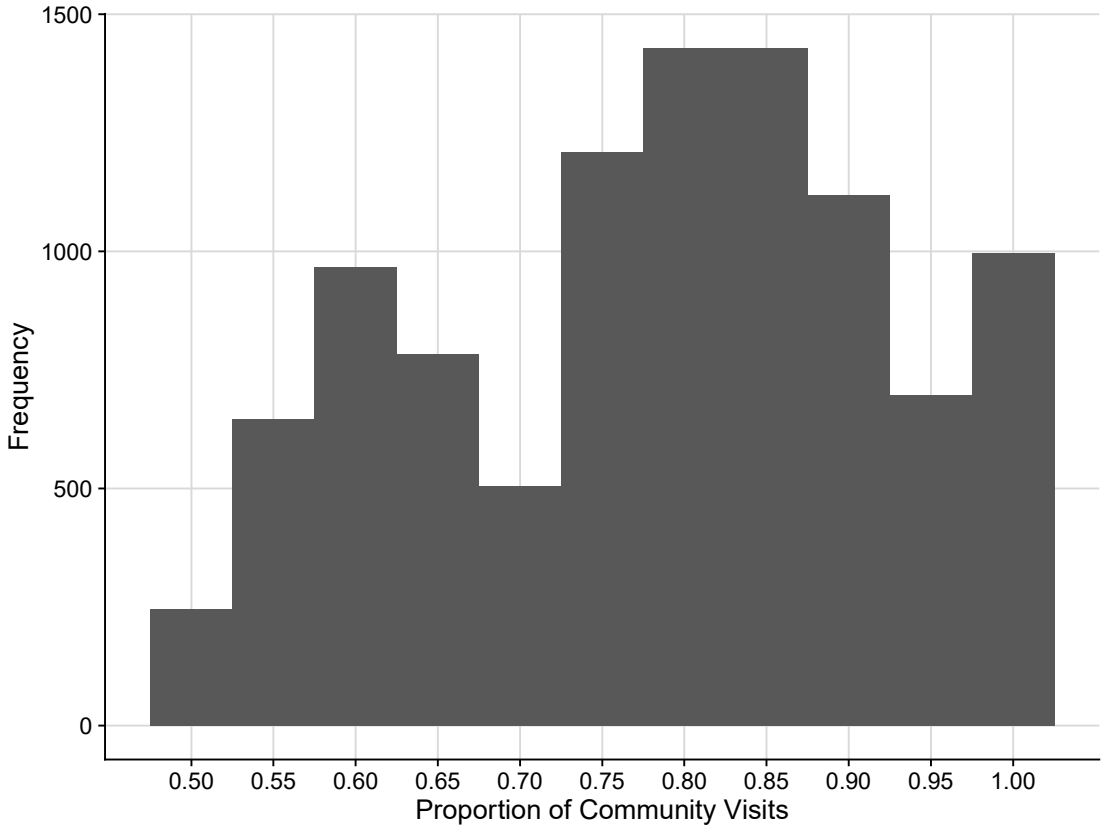

Figure S4: Distribution of the proportion of community visits by site/occupancy interval/time interval combinations for a real-world odonate dataset. Only site/occupancy interval/time interval with over half of visits as ″community sampling events″ (greater than 2 species detected based on a spatiotemporal clustering algorithm) are shown.

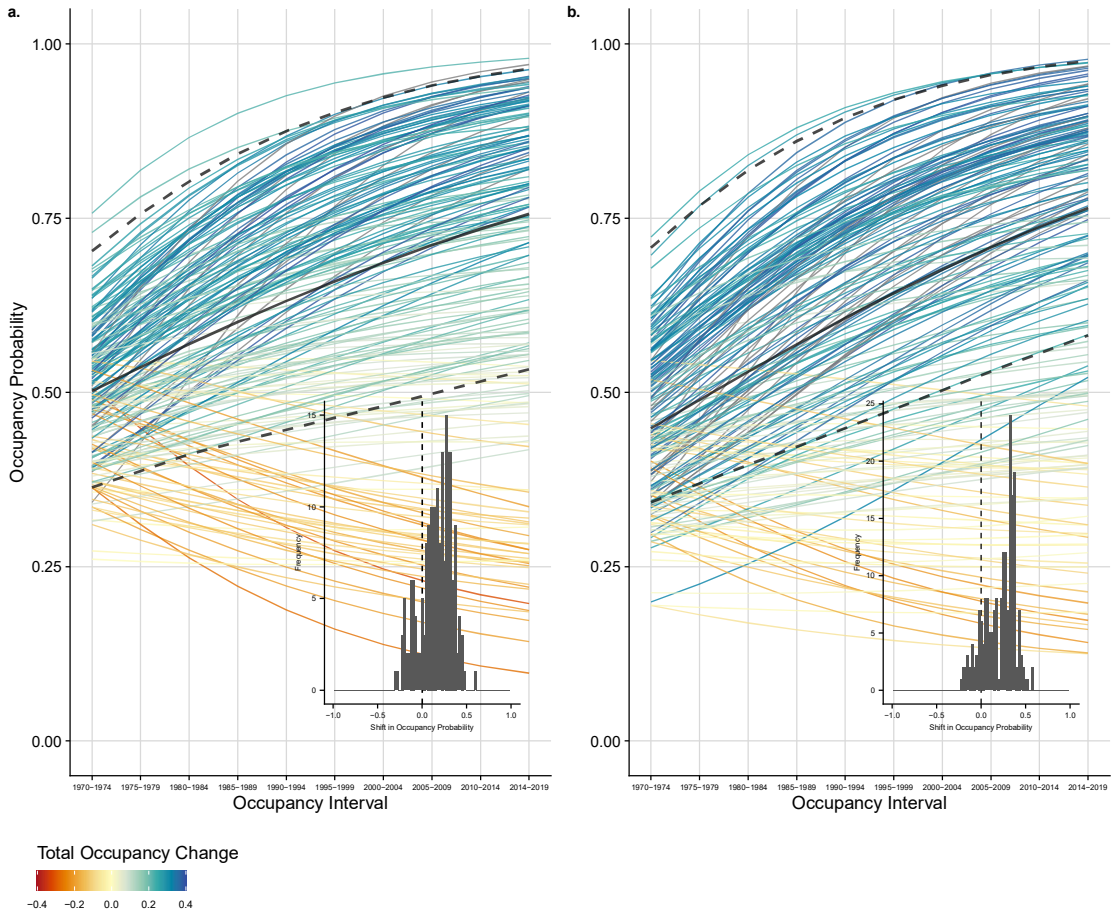

Figure S5: Species-specific (colored lines) and community (black line) trends for eastern North American odonates under three data processing workflows: (a-b) $WF_{\text{range,detected}}$. In (b), we further restrict analyses to only site $\times$ occupancy interval $\times$ time interval combinations where the proportion of clustered collection events containing 2 or more species exceeds 0.5. Lines are coloured to indicate occupancy trend through time (difference from Occupancy Interval 1 to Occupancy Interval 10). Inset histograms show these occupancy differences. Black dashed lines denote 95% Bayesian Credible Intervals.