

Online Resource 1: Supplementary Material

Bayesian Sparse Mixed-Response Species Distribution Models with Spatial Dependence

Environmental and Ecological Statistics submission

Hsin-Hsiung Huang and Osamu Komori

Correspondence: Hsin-Hsiung.Huang@ucf.edu

This file is linked to the main manuscript.

1 Notation used in the proofs

For site $i = 1, \dots, n$ and response $k = 1, \dots, q$, the main manuscript uses the linear predictor

$$\theta_{ik} = X_i^\top \beta_k + h_i^\top \gamma_k. \quad (1.1)$$

Here X is the $n \times p$ transformed environmental design matrix and H is the $n \times m$ spatial or spatio-temporal basis matrix. The full parameter is $\Theta = (B, \Gamma)$, with $B = (\beta_1, \dots, \beta_q)$ and $\Gamma = (\gamma_1, \dots, \gamma_q)$. The true value is $\Theta_0 = (B_0, \Gamma_0)$. The empirical prediction distance is

$$d_n^2(\Theta, \Theta_0) = \frac{1}{nq} \sum_{i=1}^n \sum_{k=1}^q \{X_i^\top (\beta_k - \beta_{0k}) + h_i^\top (\gamma_k - \gamma_{0k})\}^2. \quad (1.2)$$

All formal results in the main manuscript are restated and proved below. The arguments are written at a level intended to make the algebra and probability bounds checkable without interrupting the main article.

2 Proof of the Pólya–Gamma representation

Lemma 1 (Pólya–Gamma weighted-Gaussian representation). *For every Bernoulli or negative-binomial observation with likelihood proportional to $\exp(a\theta)/\{1 + \exp(\theta)\}^b$, $b > 0$, define $\kappa = a - b/2$. If $\omega \sim \text{PG}(b, 0)$, then*

$$\frac{\exp(a\theta)}{\{1 + \exp(\theta)\}^b} = 2^{-b} \exp(\kappa\theta) \int_0^\infty \exp(-\omega\theta^2/2) p_{PG}(\omega \mid b, 0) d\omega. \quad (2.1)$$

Conditioning on ω gives the kernel of a Gaussian observation with precision ω and working response $z = \kappa/\omega$.

Proof. The identity follows from a simple re-expression of the logistic likelihood and the Laplace transform of the Pólya–Gamma distribution (Polson et al., 2013). Since

$$1 + e^\theta = e^{\theta/2} \{e^{-\theta/2} + e^{\theta/2}\} = 2e^{\theta/2} \cosh(\theta/2), \quad (2.2)$$

we have

$$\frac{e^{a\theta}}{(1 + e^\theta)^b} = 2^{-b} \exp\{(a - b/2)\theta\} \cosh^{-b}(\theta/2) = 2^{-b} \exp(\kappa\theta) \cosh^{-b}(\theta/2). \quad (2.3)$$

For $\omega \sim \text{PG}(b, 0)$, the defining Laplace transform gives

$$\mathbb{E}\{\exp(-\omega\theta^2/2)\} = \cosh^{-b}(\theta/2). \quad (2.4)$$

Writing this expectation as an integral and substituting it into the preceding display proves the desired mixture representation.

It remains only to identify the conditional Gaussian kernel. After conditioning on ω , the part of the joint density that depends on θ is

$$\exp\{\kappa\theta - \omega\theta^2/2\}. \quad (2.5)$$

Completing the square gives

$$\kappa\theta - \omega\theta^2/2 = -\frac{\omega}{2} \left(\theta^2 - 2\frac{\kappa}{\omega}\theta \right) \quad (2.6)$$

$$= -\frac{\omega}{2} \left(\theta - \frac{\kappa}{\omega} \right)^2 + \frac{\kappa^2}{2\omega}. \quad (2.7)$$

The final term does not involve θ . Therefore, conditional on ω , the contribution is the kernel of a Gaussian observation with mean parameter θ , working response $z = \kappa/\omega$, and precision ω .

For Bernoulli observations, $a = y$ and $b = 1$, so $\kappa = y - 1/2$. For the negative-binomial parameterization used in the main manuscript,

$$L(\theta) \propto \frac{e^{y\theta}}{(1 + e^\theta)^{y+r}}, \quad (2.8)$$

so $a = y$, $b = y + r$ and $\kappa = (y - r)/2$. Both cases are therefore covered by the same identity. $\square$

3 Proof of the Gaussian full conditional blocks

Proposition 1 (Gaussian full conditional blocks). *Fix a response k and suppose the working model*

$$Z_k = X\beta_k + H\gamma_k + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}_n(0, W_k^{-1}) \quad (3.1)$$

holds. If $\beta_k \sim \mathcal{N}_p(0, D_k)$ and γ_k is fixed, then

$$\beta_k \mid - \sim \mathcal{N}_p(m_{\beta k}, V_{\beta k}), \quad V_{\beta k} = (X^\top W_k X + D_k^{-1})^{-1}, \quad m_{\beta k} = V_{\beta k} X^\top W_k (Z_k - H\gamma_k). \quad (3.2)$$

If $\gamma_k \sim \mathcal{N}_m(0, G_k)$ and β_k is fixed, then

$$\gamma_k \mid - \sim \mathcal{N}_m(m_{\gamma k}, V_{\gamma k}), \quad V_{\gamma k} = (H^\top W_k H + G_k^{-1})^{-1}, \quad m_{\gamma k} = V_{\gamma k} H^\top W_k (Z_k - X\beta_k). \quad (3.3)$$

Proof. With γ_k fixed, write $\tilde{Z}_k = Z_k - H\gamma_k$. The working model becomes $\tilde{Z}_k = X\beta_k + \varepsilon_k$, with $\varepsilon_k \sim \mathcal{N}_n(0, W_k^{-1})$. Up to constants not depending on β_k , the log likelihood is

$$-\frac{1}{2}(\tilde{Z}_k - X\beta_k)^\top W_k(\tilde{Z}_k - X\beta_k). \quad (3.4)$$

The Gaussian prior for β_k contributes $-\beta_k^\top D_k^{-1}\beta_k/2$. Combining likelihood and prior yields the conditional log posterior

$$-\frac{1}{2}(\tilde{Z}_k - X\beta_k)^\top W_k(\tilde{Z}_k - X\beta_k) - \frac{1}{2}\beta_k^\top D_k^{-1}\beta_k + C \quad (3.5)$$

$$= -\frac{1}{2}\beta_k^\top (X^\top W_k X + D_k^{-1})\beta_k + \beta_k^\top X^\top W_k \tilde{Z}_k + C', \quad (3.6)$$

where C and C' are constants with respect to β_k . Let

$$V_{\beta k} = (X^\top W_k X + D_k^{-1})^{-1}, \quad m_{\beta k} = V_{\beta k} X^\top W_k \tilde{Z}_k. \quad (3.7)$$

Completing the square gives

$$-\frac{1}{2}(\beta_k - m_{\beta k})^\top V_{\beta k}^{-1}(\beta_k - m_{\beta k}) + C'', \quad (3.8)$$

which is the kernel of $\mathcal{N}_p(m_{\beta k}, V_{\beta k})$. Substituting $\tilde{Z}_k = Z_k - H\gamma_k$ gives the displayed mean.

The update for γ_k is identical after exchanging $X\beta_k$ and $H\gamma_k$. With β_k fixed, set $\bar{Z}_k = Z_k - X\beta_k$. The likelihood becomes Gaussian with mean $H\gamma_k$, and the prior contributes $-\gamma_k^\top G_k^{-1}\gamma_k/2$. The conditional log posterior is

$$-\frac{1}{2}\gamma_k^\top (H^\top W_k H + G_k^{-1})\gamma_k + \gamma_k^\top H^\top W_k \bar{Z}_k + C. \quad (3.9)$$

Thus

$$V_{\gamma k} = (H^\top W_k H + G_k^{-1})^{-1}, \quad m_{\gamma k} = V_{\gamma k} H^\top W_k \bar{Z}_k, \quad (3.10)$$

and the conditional distribution is $\mathcal{N}_m(m_{\gamma k}, V_{\gamma k})$. Since $\bar{Z}_k = Z_k - X\beta_k$, the stated formula follows. For Gaussian responses, $Z_k = Y_k$ and $W_k = \sigma_k^{-2}I_n$; for Bernoulli and negative-binomial responses, Lemma 1 supplies the augmented $Z_{ik} = \kappa_{ik}/\omega_{ik}$ and $W_{ik} = \omega_{ik}$. The same weighted-Gaussian algebra therefore applies to all implemented families. $\square$

4 Proof of posterior contraction

This section proves Theorem 1 from the main manuscript. The proof is written for independent sites conditional on the spatial basis. The basis coefficients are finite-dimensional parameters; their dimension $m_n q$ is allowed to increase slowly with n .

Condition 1 (Design, likelihood and prior regularity). *The following hold as n increases.*

(i) *The transformed design and basis rows are uniformly bounded, and the active-column restricted eigenvalues of X are bounded away from zero and infinity.* (ii) *The true linear predictors are uniformly bounded.* (iii) *The Gaussian variances and negative-binomial size parameters are bounded away from zero and infinity.* (iv) *The row-sparse prior assigns at least $\exp\{-Cs_n q \log p\}$ mass to an $O(\varepsilon_n)$ Frobenius ball around B_0 and has exponentially small mass outside a sieve with at most a constant multiple of s_n active rows.* (v) *The Gaussian basis prior assigns at least $\exp\{-Cm_n q \log n\}$ mass to an $O(\varepsilon_n)$ Euclidean ball around Γ_0 and has exponentially small mass outside a Euclidean ball of radius polynomial in n .*

Theorem 1 (Posterior prediction contraction). *Under the preceding condition, set*

$$\varepsilon_n^2 = \frac{s_n q \log p + m_n q \log n}{nq}. \quad (4.1)$$

If $nq\varepsilon_n^2 \rightarrow \infty$ and $\varepsilon_n \rightarrow 0$, then there exists a constant $M < \infty$ such that

$$\Pi\{\Theta : d_n(\Theta, \Theta_0) > M\varepsilon_n \mid Y\} \rightarrow 0 \quad (4.2)$$

in P_{Θ_0} -probability.

Proof. Let $\eta_{ik}(\Theta) = X_i^\top \beta_k + h_i^\top \gamma_k$. For each response family, write the density or mass function as $p_k(y \mid \eta)$, with variance or count-size nuisance parameters restricted to compact sets. The bounded-predictor assumption places the true and sieve-restricted predictors in a compact interval. On that interval, the first two derivatives of $\log p_k(y \mid \eta)$ with respect to η are uniformly controlled for the Gaussian, Bernoulli and negative-binomial families considered here.

A Taylor expansion of the log likelihood around $\eta_{0ik} = \eta_{ik}(\Theta_0)$ gives the local Kullback–Leibler comparison. The score has mean zero at the truth, and the curvature is uniformly bounded above on the compact predictor set. Hence, for a constant C independent of n , p , q and m_n ,

$$\text{KL}(P_{\Theta_0}^{(n)}, P_{\Theta}^{(n)}) \leq C \sum_{i=1}^n \sum_{k=1}^q (\eta_{ik} - \eta_{0ik})^2 = Cnq d_n^2(\Theta, \Theta_0). \quad (4.3)$$

The same bounded-curvature argument, applied to the centered log likelihood ratio, gives

$$\mathbb{E}_0 \left[\left\{ \log \frac{p(Y \mid \Theta_0)}{p(Y \mid \Theta)} - \text{KL}(P_{\Theta_0}^{(n)}, P_{\Theta}^{(n)}) \right\}^2 \right] \leq Cnq d_n^2(\Theta, \Theta_0). \quad (4.4)$$

Consider the prediction ball $\mathcal{K}_n = \{\Theta : d_n(\Theta, \Theta_0) \leq c\varepsilon_n\}$. Uniform boundedness of the design and basis rows implies that, for suitable constants, a Frobenius ball around B_0 and a Euclidean ball around Γ_0 are contained in $\mathcal{K}_n$. The prior mass assumptions therefore imply

$$\Pi(\mathcal{K}_n) \geq \exp\{-C_1 s_n q \log p\} \exp\{-C_2 m_n q \log n\} \quad (4.5)$$

$$= \exp\{-C_3 (s_n q \log p + m_n q \log n)\} \quad (4.6)$$

$$= \exp\{-C_3 n q \varepsilon_n^2\}. \quad (4.7)$$

Together with (4.3) and (4.4), this gives a Kullback–Leibler neighborhood of the truth with sufficient posterior denominator mass.

Next define a sieve $\mathcal{F}_n$ consisting of parameters with at most $\bar{s}_n = Cs_n$ nonzero rows in B , coefficient entries bounded by a polynomial in n , and $\|\Gamma\|_F$ bounded by a polynomial in

n . The assumed prior tail probabilities give

$$\Pi(\mathcal{F}_n^c) \leq \exp\{-(C_3 + 4)nq\varepsilon_n^2\} \quad (4.8)$$

for a sufficiently large sieve constant.

The metric entropy of $\mathcal{F}_n$ is of the same order as the numerator in the rate. For a fixed active row set S with $|S| \leq \bar{s}_n$, the coefficient matrix restricted to S has $|S|q$ free entries, and a Euclidean grid over a polynomially bounded ball contributes at most $C|S|q \log n$ to the log covering number. The number of possible active row sets satisfies

$$\sum_{j=0}^{\bar{s}_n} \binom{p}{j} \leq \left(\frac{ep}{\bar{s}_n}\right)^{\bar{s}_n+1}, \quad (4.9)$$

so row selection contributes at most $C\bar{s}_n \log p$. The spatial basis matrix has $m_n q$ entries and contributes at most $Cm_n q \log n$. Enlarging constants gives

$$\log N(c\varepsilon_n, \mathcal{F}_n, d_n) \leq C(s_n q \log p + m_n q \log n) = Cnq\varepsilon_n^2. \quad (4.10)$$

It remains to control posterior mass outside a multiple of the contraction radius. For any fixed Θ_1 with $d_n(\Theta_1, \Theta_0) > M\varepsilon_n$, the regularity of the Gaussian and exponential-family likelihoods on the bounded predictor set implies that the Hellinger distance between $P_{\Theta_1}^{(n)}$ and $P_{\Theta_0}^{(n)}$ is bounded below by a constant multiple of $\sqrt{nq} d_n(\Theta_1, \Theta_0)$. The Neyman–Pearson test between these two simple hypotheses has both errors bounded by

$$\exp\{-cnqd_n^2(\Theta_1, \Theta_0)\}. \quad (4.11)$$

Covering $\{\Theta \in \mathcal{F}_n : d_n(\Theta, \Theta_0) > M\varepsilon_n\}$ by d_n -balls of radius $M\varepsilon_n/4$ and combining the simple tests over the cover gives tests ϕ_n satisfying

$$\mathbb{E}_0 \phi_n \leq \exp\{-c_1 M^2 nq\varepsilon_n^2 + C_1 nq\varepsilon_n^2\}, \quad (4.12)$$

$$\sup_{\Theta \in \mathcal{F}_n: d_n(\Theta, \Theta_0) > M\varepsilon_n} E_{\Theta}(1 - \phi_n) \leq \exp\{-c_2 M^2 n q \varepsilon_n^2\}. \quad (4.13)$$

Choosing M large enough makes the exponent in the first bound negative with margin exceeding the denominator and sieve constants.

The posterior numerator of the bad set is decomposed into its part inside $\mathcal{F}_n$ and its tail outside $\mathcal{F}_n$. The tests above control the inside part, while the prior tail bound controls the outside part. The marginal likelihood denominator over $\mathcal{K}_n$ is at least $\exp\{-(C_3+1)nq\varepsilon_n^2\}$ with probability tending to one under P_{Θ_0} . Dividing the numerator bounds by this denominator shows that both contributions vanish in P_{Θ_0} -probability. Therefore

$$\Pi\{\Theta : d_n(\Theta, \Theta_0) > M\varepsilon_n \mid Y\} \rightarrow 0, \quad (4.14)$$

which proves the theorem. $\square$

5 Proof of posterior sure screening

Corollary 1 (Posterior sure screening). *Let $r_j(B) = \|B_{j\cdot}\|_2$ and suppose the minimum active row signal satisfies*

$$\min_{j \in \mathcal{S}_0} r_j(B_0) > 2M\varepsilon_n, \quad (5.1)$$

where $\mathcal{S}_0 = \{j : r_j(B_0) > 0\}$ and M is the contraction constant. If the active-block restricted eigenvalue converts prediction contraction into row-wise coefficient contraction, and rows are selected by $r_j(B) > M\varepsilon_n$, then the posterior probability of selecting every truly active row tends to one.

Proof. Theorem 1 implies that the posterior concentrates on a prediction ball of radius $M\varepsilon_n$. Under the active-block restricted eigenvalue condition, this prediction contraction implies coefficient contraction on the truly active rows: for a constant that can be absorbed into the

threshold,

$$\|B_{\mathcal{S}_0} - B_{0,\mathcal{S}_0}\|_F \leq M\varepsilon_n \quad (5.2)$$

with posterior probability tending to one. Consequently, for every active row $j \in \mathcal{S}_0$,

$$\|B_{j\cdot} - B_{0,j\cdot}\|_2 \leq M\varepsilon_n \quad (5.3)$$

on the same event. The reverse triangle inequality then yields

$$r_j(B) = \|B_{j\cdot}\|_2 \geq \|B_{0,j\cdot}\|_2 - \|B_{j\cdot} - B_{0,j\cdot}\|_2 > 2M\varepsilon_n - M\varepsilon_n = M\varepsilon_n. \quad (5.4)$$

Thus every truly active row exceeds the selection threshold on an event whose posterior probability tends to one. The result is a sure-screening statement: it guarantees inclusion of all active rows. It does not, by itself, rule out all false positives unless an additional selection-consistency condition is imposed on inactive rows. $\square$

6 Full conditional details for nuisance parameters

The main sampler uses standard conjugate updates for Gaussian variance components and scale-mixture updates for row-wise shrinkage. These are not stated as main-text theorems, but the algebra is included for completeness.

6.1 Gaussian variance update

For $k \in \mathcal{G}$, let $e_k = Y_k - X\beta_k - H\gamma_k$. With inverse-gamma prior $\sigma_k^2 \sim \text{IG}(a_\sigma, b_\sigma)$ under the density proportional to $(\sigma^2)^{-a_\sigma-1} \exp(-b_\sigma/\sigma^2)$,

$$\sigma_k^2 \mid - \sim \text{IG}\left(a_\sigma + \frac{n}{2}, b_\sigma + \frac{1}{2}e_k^\top e_k\right). \quad (6.1)$$

This follows by multiplying the Gaussian likelihood contribution $(\sigma^2)^{-n/2} \exp\{-e_k^\top e_k / (2\sigma^2)\}$ by the inverse-gamma prior and collecting powers of σ^2 .

6.2 Normal scale-mixture shrinkage update

For row j , suppose $\beta_j \mid \lambda_j, \tau, \Sigma_\beta \sim \mathcal{N}_q(0, \tau^2 \lambda_j^2 \Sigma_\beta)$ and the local variance $u_j = \lambda_j^2$ has a mixing prior $p(u_j)$. The conditional density is proportional to

$$u_j^{-q/2} \exp \left\{ -\frac{1}{2\tau^2 u_j} \beta_j \cdot \Sigma_\beta^{-1} \beta_j^\top \right\} p(u_j). \quad (6.2)$$

For inverse-gamma mixing, this remains inverse-gamma. For horseshoe or generalized double-Pareto mixing, the same expression is used in either Gibbs sub-updates or a one-dimensional slice or Metropolis update. The key point for Proposition 1 is that, conditional on u_j , the coefficient prior is Gaussian.

7 Additional simulation diagnostics

The simulation diagnostics complement the main-text heat maps by showing distributional variability across replications, row-selection behavior and residual spatial structure. They are intended to prevent overinterpretation of any single summary score. In particular, coefficient recovery and support recovery are assessed separately because a method can predict well while still selecting a correlated surrogate for the true feature.

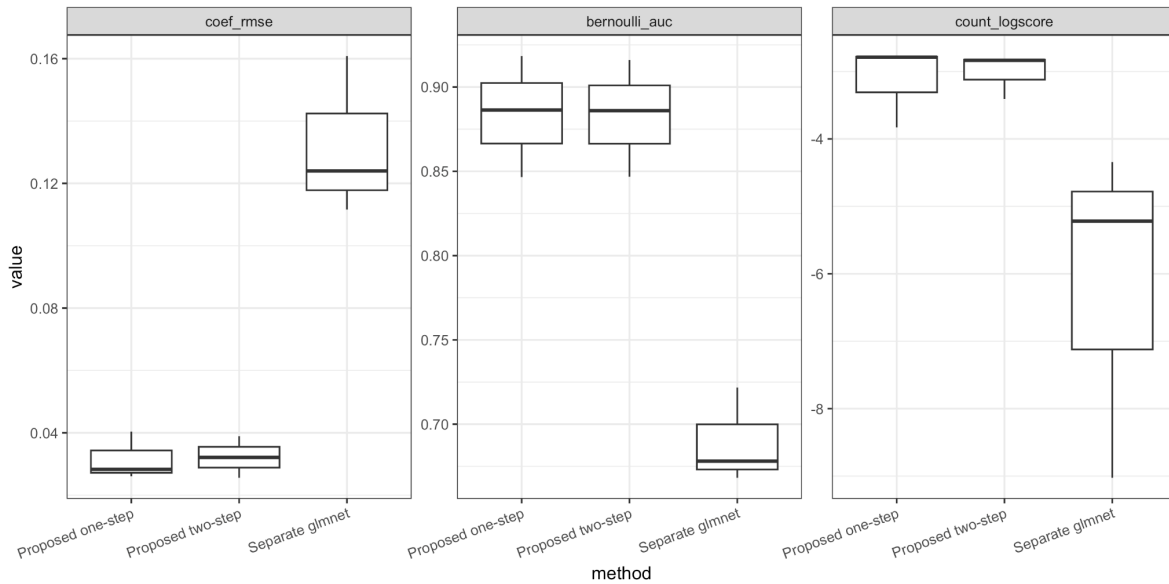


Figure 1 Simulation box plots for the repeated mixed-response benchmark.

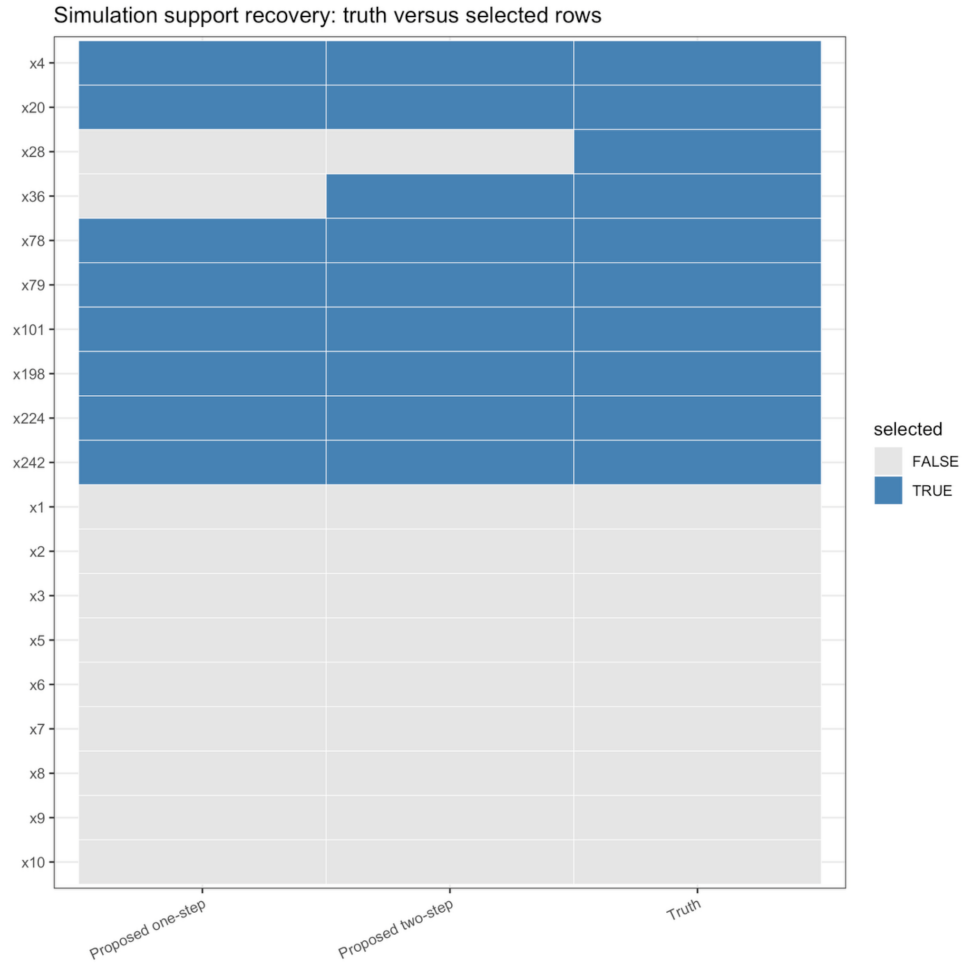


Figure 2 Simulation support recovery: truth versus selected feature rows.

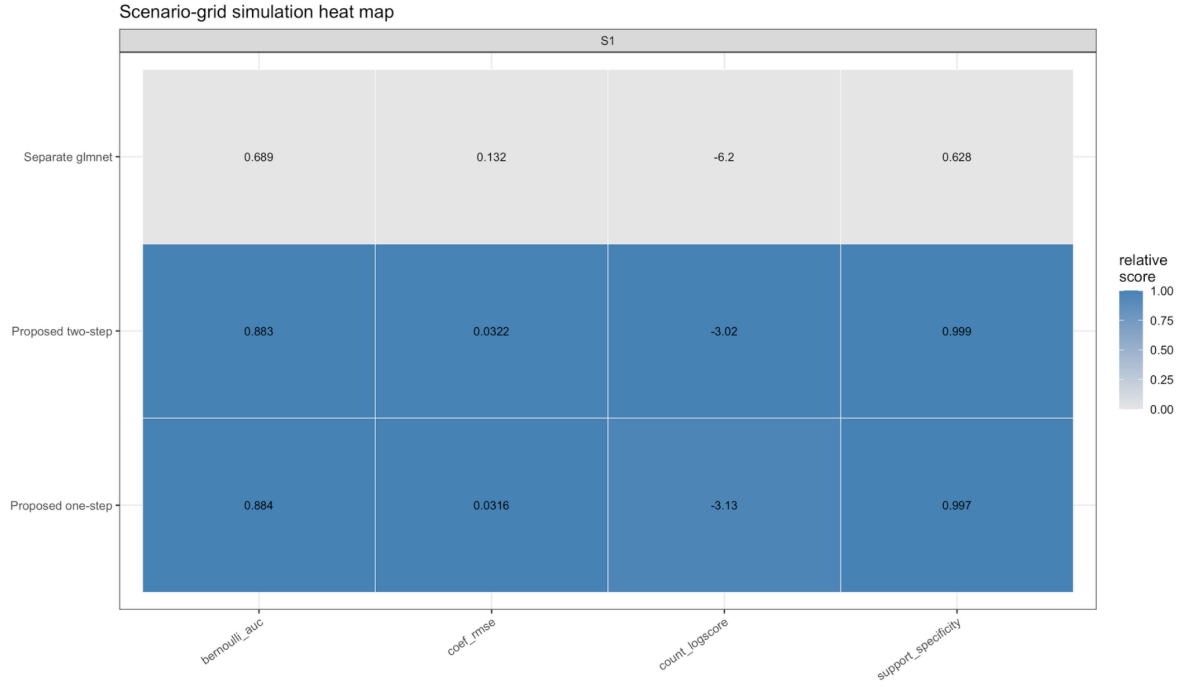


Figure 3 Scenario-grid simulation heat map summarizing representative metrics across the baseline scenario.

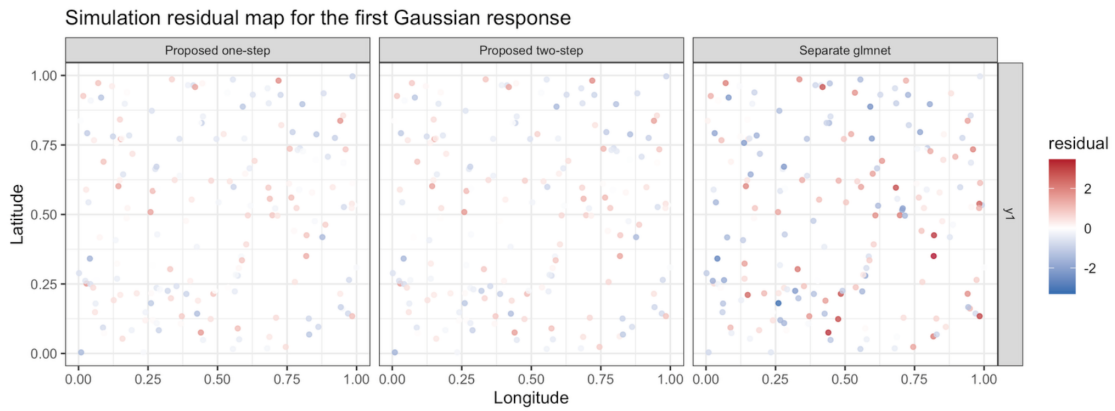


Figure 4 Simulation residual map for the first Gaussian response.

8 Spatial-sparse Bernoulli implementation for the elephant benchmark

The main manuscript’s linear predictor contains both an environmental dictionary term and a spatial-basis term,

$$\theta_i = X_i^\top \beta + h_i^\top \gamma. \quad (8.1)$$

The EcoData elephant benchmark showed that the environment-only sparse ablation, obtained by setting $h_i^\top \gamma = 0$, is too restrictive for this data set. The observed occurrence-background surface has a strong broad geographic gradient, and a sparse set of environmental hinge and threshold features cannot reliably emulate that smooth spatial trend. The revised empirical workflow therefore fits the proposed Bernoulli model in its full spatial-sparse form. The environmental design X is the screened MaxEnt-style dictionary, while $H = (h_1^\top, \dots, h_n^\top)^\top$ is a low-rank radial spatial basis built from the observation coordinates. In each spatial training fold, the basis knots are chosen from the training sites, the Bernoulli likelihood is fit by the same Pólya–Gamma update as in the main model, and a one-dimensional probability calibration is estimated from training-fold predictions. Test-fold probabilities are then obtained by applying the fitted environmental coefficients, spatial basis coefficients and calibration map to the held-out spatial block.

This revision changes the interpretation of the EcoData comparison. The method labelled as the proposed Bernoulli fit is the spatial-sparse model above, whereas the environment-only sparse fit is an ablation used to diagnose whether the sparse environmental dictionary alone is adequate. In the updated spatially blocked benchmark, the full spatial-sparse Bernoulli fit is the empirical winner for all three reported scores: it has the largest AUC and the smallest Brier score and log score. The ablation remains scientifically useful because it shows that the sparse environmental dictionary alone is not adequate when the dominant signal is smooth geography. Its purpose is to justify the $H\gamma$ component in the main model, not to serve as the final proposed map.

The elephant reference diagnostics in Fig. 5 display the same-grid empirical occurrence reference surface together with the fitted habitat-suitability maps. These are the appropriate empirical references because continuous habitat suitability is unobserved; the continent-wide distribution map of Pečnerová et al. (2026) is cited in the main text as external range context, but is not reproduced here to avoid third-party artwork permissions. The metric table below records the same blocked-CV values shown in the main text heat map. The prediction-correlation heat map in Fig. 6 shows that the competing models are not merely rescaled versions of one another. The residual maps in Fig. 7 identify locations where fitted probabilities remain too low for observed occurrences or too high for background points. Together, these diagnostics support the revised main-text conclusion: the full proposed Bernoulli implementation must include the spatial basis, while the environment-only sparse fit is an ablation that exposes the consequence of omitting smooth spatial structure.

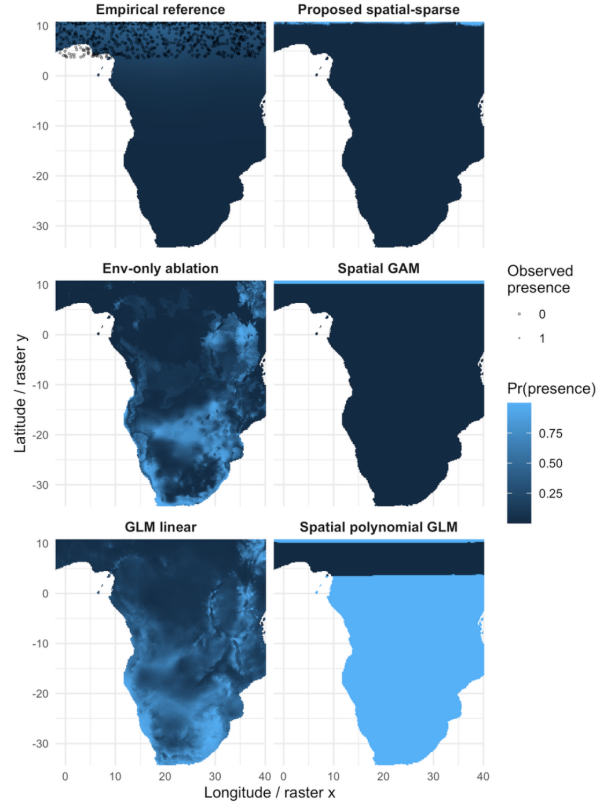


Figure 5 Kernel-smoothed African elephant occurrence-background reference surface and fitted habitat-suitability maps on the same EcoData prediction grid. The proposed panel corresponds to the spatial-sparse Bernoulli implementation; the environment-only sparse fit is an ablation.

Table 1 African elephant spatially blocked cross-validation summary. AUC is better when larger; Brier score and log score are better when smaller. The asterisks identify the best method for each criterion.

Method	AUC	Brier score	Log score
Proposed spatial-sparse Bernoulli	0.997*	0.012*	0.084*
Proposed sparse Bernoulli (environment only)	0.924	0.158	0.474
GAM environmental plus spatial smooth	0.991	0.014	0.158
Spatial polynomial GLM	0.949	0.077	0.949
GLM linear	0.807	0.250	0.703

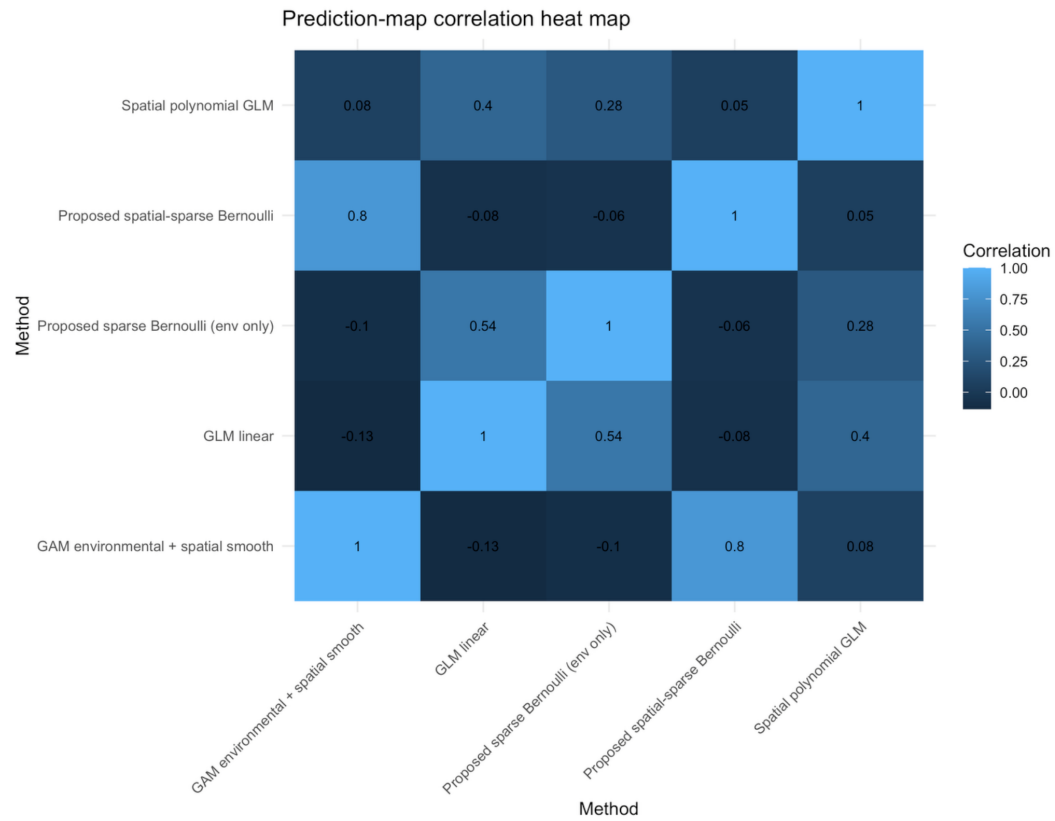


Figure 6 Prediction-map correlation heat map among competing Bernoulli models.

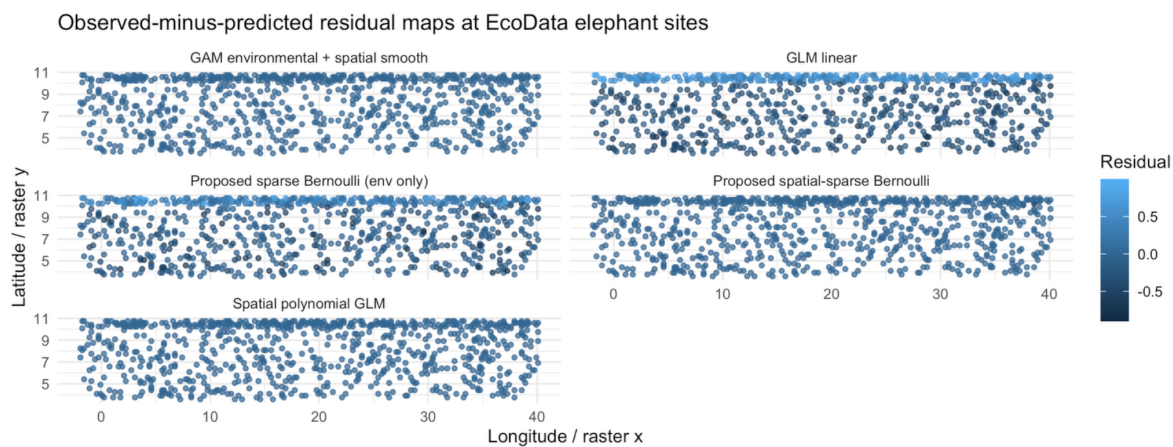


Figure 7 Observed-minus-predicted residual maps at elephant occurrence-background sites.

9 Additional empirical diagnostics

The FishGlob panels in Figs. 8 and 9 are included to emphasize the mixed-response target of the paper. The predictive panel should be read family by family: occurrence metrics evaluate discrimination and calibration, continuous metrics evaluate regression accuracy, and count log scores are sensitive to overdispersion and rare large catches. The residual-correlation heat map complements those predictive summaries by showing which response components remain associated after environmental feature expansion and spatial adjustment. This is why the main manuscript recommends survey-stratified diagnostics, residual-dependence checks and sensitivity to negative-binomial size parameters before attaching a single ecological interpretation to the count components.

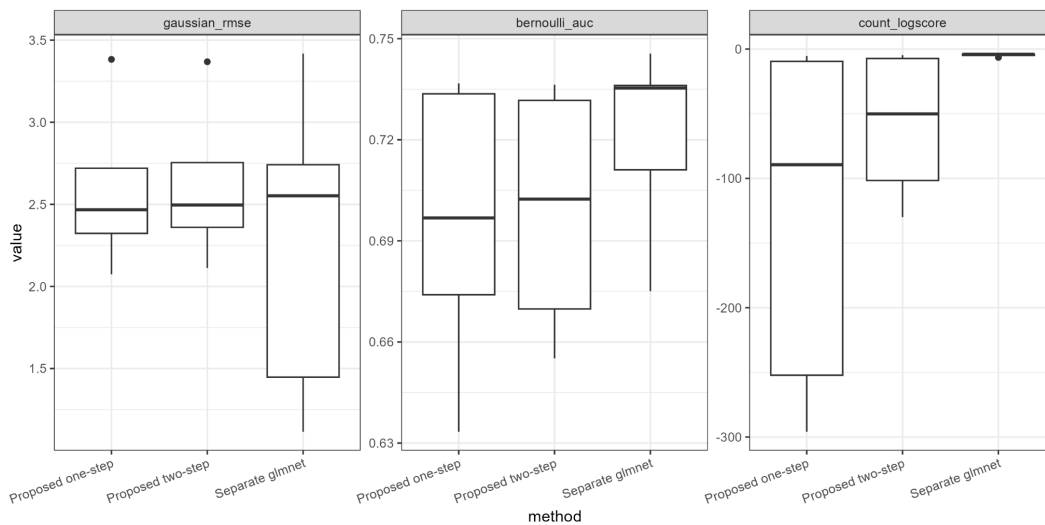


Figure 8 FishGlob cross-validation metric panel for the survey-scale mixed-response target.

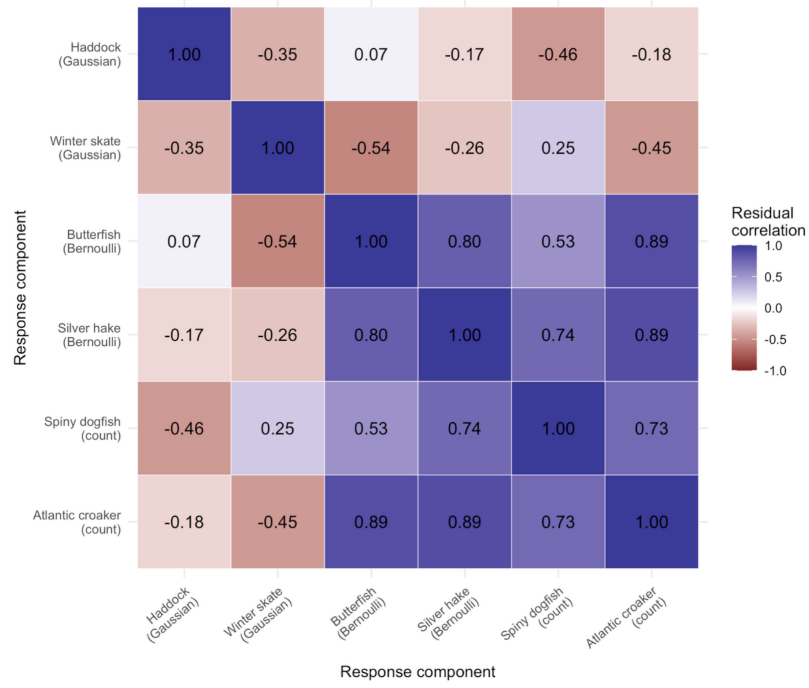


Figure 9 FishGlob residual-correlation heat map for the six-response mixed panel. This diagnostic visualizes the residual dependence that motivates joint modeling after adjustment for the expanded environmental design and low-rank spatial basis.

10 Supplementary numerical tables

The following tables collect the numerical settings and summaries underlying the figures in the main manuscript and supplement. They are included to make the simulation design, FishGlob response construction, and elephant occurrence-background diagnostics transparent without overloading the main article.

Table 2 Core simulation design used for the mixed-response benchmark. G, B and C denote Gaussian, Bernoulli and count response blocks.

n	p	G	B	C	active rows	basis rank
180	250	2	2	2	10	10

Table 3 Reviewer-facing simulation scenario grid. The baseline scenario is shown in the main figures; the remaining scenarios define stress tests for larger feature dictionaries, weaker sparsity, spatial-signal strength, and screening difficulty.

Scenario	Name		n	p	active rows	basis rank	Stress test
S1	Well-specified mixed response	row-sparse	180	250	10	10	Baseline target regime with shared sparsity and cross-response dependence
S2	Higher-dimensional $p > n$ design		180	500	10	10	Feature dictionary much larger than sample size
S3	Small- n high- p screening stress test		120	500	10	10	Sure-screening under limited sample size
S4	Denser active support		180	250	25	10	Weaker sparsity and harder variable selection
S5	Weak spatial structure		180	250	10	1	Little benefit expected from the spatial basis
S6	Strong spatial structure		180	250	10	20	Spatial confounding and stronger low-rank nuisance
S7	Larger sample with moderate p		300	300	15	12	Stability as sample size increases
S8	High-dimensional sparse support with larger basis		240	600	12	20	Large feature dictionary plus stronger spatial component

Table 4 Representative baseline simulation summary. For RMSE, Brier score and log loss, smaller values are better; for AUC, sensitivity and specificity, larger values are better. Count log score is shown on the original reported scale.

Method	coef. RMSE	sensitivity	specificity	Gaussian RMSE	Bernoulli log loss	Brier	AUC	count RMSE	count log score
Proposed one-step	0.032	0.933	0.997	0.696	0.428	0.138	0.884	415.047	-3.130
Proposed two-step	0.032	0.933	0.999	0.744	0.429	0.139	0.883	236.485	-3.019
Separate glmnet	0.132	0.967	0.628	1.548	0.632	0.221	0.689	175.814	-6.196

Table 5 Low-rank basis approximation diagnostic used to select a computationally stable spatial-basis dimension.

Rank	relative Frobenius error	variance explained
5	0.324	0.422
10	0.221	0.544
20	0.145	0.666
40	0.090	0.783

Table 6 FishGlob mixed-response panel used to define the survey-scale ecological target. The labels G , B and C refer to Gaussian, Bernoulli and count likelihood blocks.

Component	Common name	Likelihood family	Role in the mixed panel
G_1	Haddock	Gaussian	Demersal shelf groundfish with expected depth and bottom-temperature response
G_2	Winter skate	Gaussian	Benthic skate representing bottom-habitat gradients
B_1	Butterfish	Bernoulli	Shelf-associated occurrence contrast in the same trawl footprint
B_2	Silver hake	Bernoulli	Shelf predator with seasonal and temperature-associated movement
C_1	Atlantic spiny dogfish	negative binomial	Overdispersed near-bottom count response
C_2	Atlantic croaker	negative binomial	Coastal and estuarine bottom-dwelling count response

Table 7 Selected expanded FishGlob features from the final sparse fit. These terms illustrate the environmental dictionary used for joint row selection.

Selected feature
hinge_r__depth__50.000
hinge_r__depth__72.500
hinge_r__year__1977.000
lin__month_cos
hinge_r__log_duration__0.288
hinge_f__log_duration__0.405
thr__temperature__16.728
hinge_f__temperature__16.728
prod__depth__log_duration
prod__depth__temperature

References

- Patricia Pečnerová, Yasuko Ishida, Genís Garcia-Erill, Laura D. Bertola, Cindy G. Santander, Xiaodong Liu, Anna Brüniche-Olsen, et al. The genomic impact of population connectivity and decline in africa’s elephants. *Nature Communications*, 17:3223, 2026. doi: <https://doi.org/10.1038/s41467-026-71262-w>.
- Nicholas G. Polson, James G. Scott, and Jesse Windle. Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013. doi: <https://doi.org/10.1080/01621459.2013.829001>.