

INTRODUCTION

Ecological studies involve understanding and identifying the underlying drivers of the distribution of species. There is a high volume of species distribution data obtained from citizen science records. Environmental data is used together with citizen science data to construct species distribution models. However, there are biases in citizen science data due to their presence-only nature and also imperfect detection; another source of bias is measurement error inherent in environmental data across sites, which results in bias estimates of the effect sizes of covariates in our species distribution models. We propose a Bayesian approach that accounts for these biases. We do Simulation-Based Calibration (SBC) to show the effect of measurement error in environmental data (i.e., species distribution data and covariates).

BAYESIAN MODEL

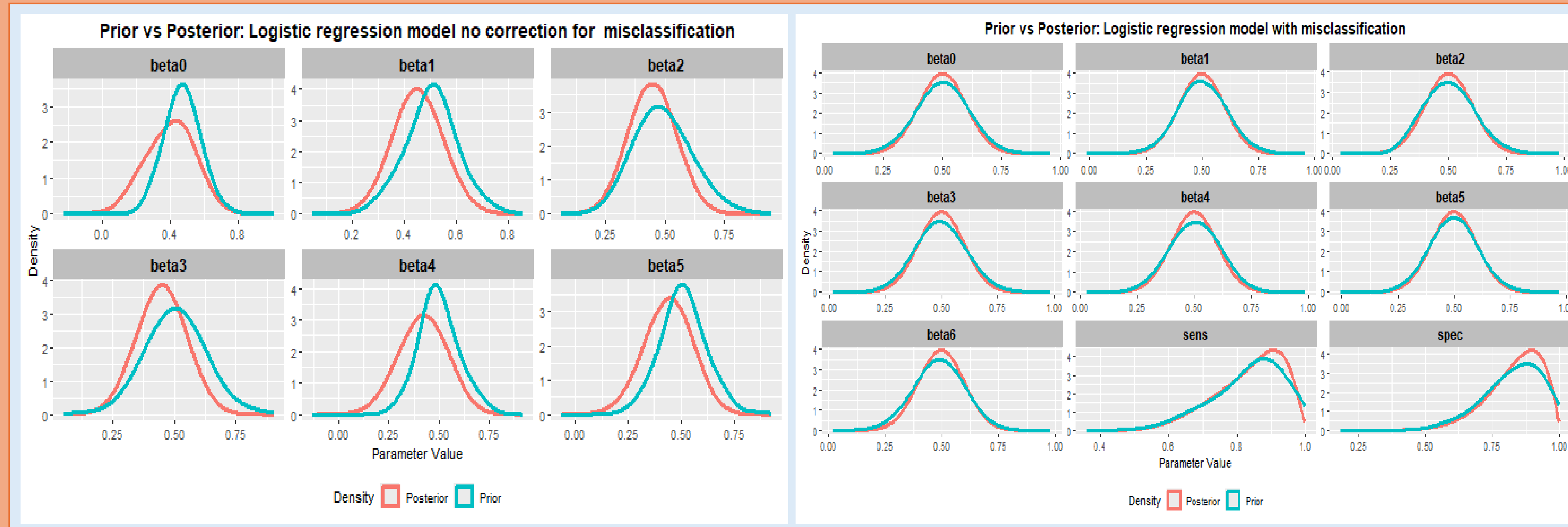
$$\mathbb{P}(y(A_i) = 1 | \mathbf{x}(A_i)) = \text{logit} \left[\frac{\mathbb{P}(y^*(A_i) = 1 | \mathbf{x}(A_i)) - (1 - \rho_{00})}{\rho_{11} - (1 - \rho_{00})} \right] = \beta_0 + \beta_k \mathbf{x}_k(A_i) + \phi(A_i)$$

where A_i is the spatial location or site and $i = 1, 2, \dots, n$.

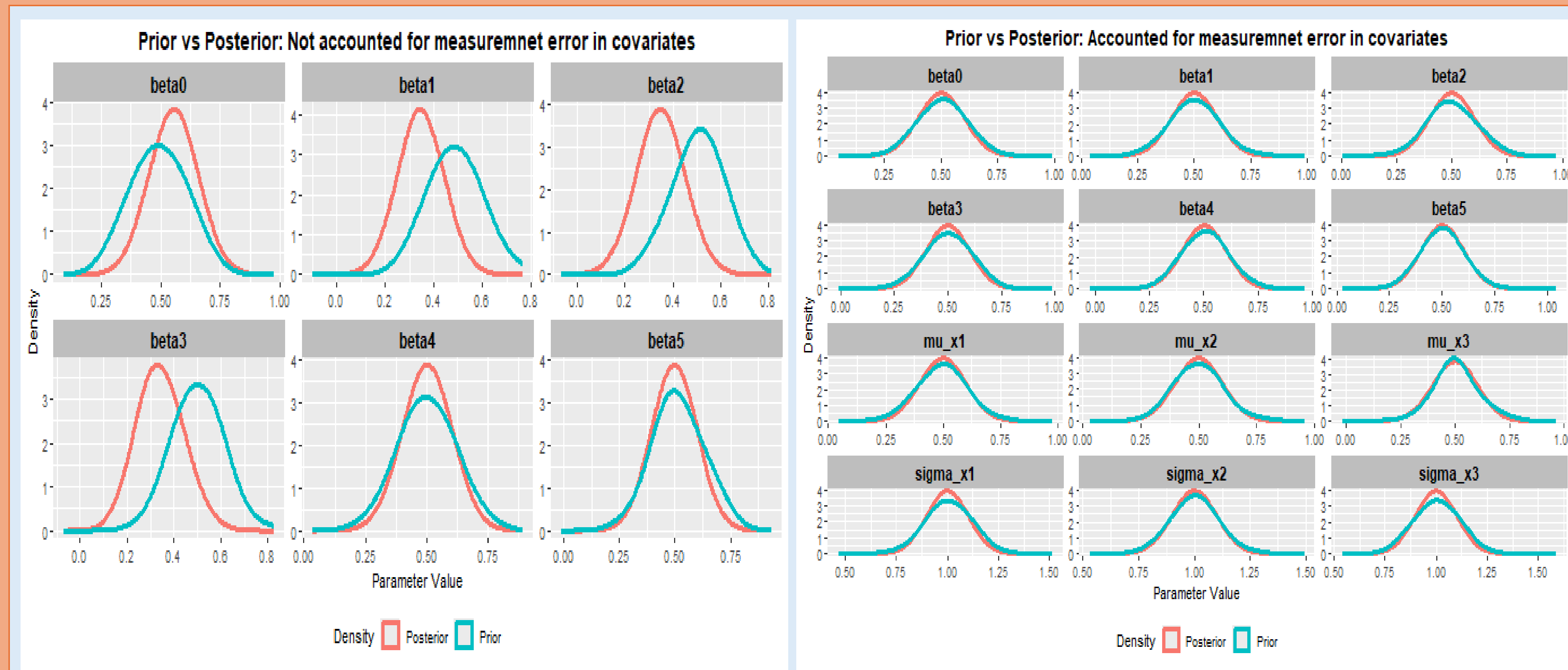
$$\beta_0 \sim N(0, \sigma_0^2); \beta_k \sim N(0, \sigma_k^2); \phi_i \sim N(0, \tau^{-1}(\text{diag}(\mathbf{W}\mathbf{1}) - \rho\mathbf{W})^{-1});$$

$$\mathbf{x}_{ik}^* \sim N(\mathbf{x}_{ik}, \sigma_{*k}^2); \mathbf{x}_{ik} \sim N(\mu_x, \sigma_x^2); \mu_x \sim N(0, \sigma_\mu^2); \sigma_x^2 \sim \text{Exp}(1); \rho_{00} \sim \text{Beta}(\alpha_0, \beta_0); \rho_{11} \sim \text{Beta}(\alpha_1, \beta_1)$$

SBC: EFFECT OF MISCLASSIFICATION



SBC: MEASUREMENT ERROR IN COVARIATES



PRESENCE-ONLY DATA

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
L1	1	1	1	1	NA	1	NA	1	NA	1
L2	1	NA	1	NA	NA	NA	NA	1	1	1
L3	1	NA	1	1	1	NA	NA	1	NA	NA
L4	NA	NA	NA	1	NA	1	1	1	NA	NA
L5	1	1	1	NA	1	1	NA	NA	NA	1
L6	1	1	1	1	1	NA	NA	NA	1	NA
L7	NA	NA	NA	1	1	NA	NA	NA	NA	1
L8	NA	NA	NA	NA	NA	NA	NA	1	NA	NA
L9	1	NA	NA	NA	NA	1	NA	1	1	NA
L10	NA	NA	NA	NA	1	NA	NA	1	NA	1

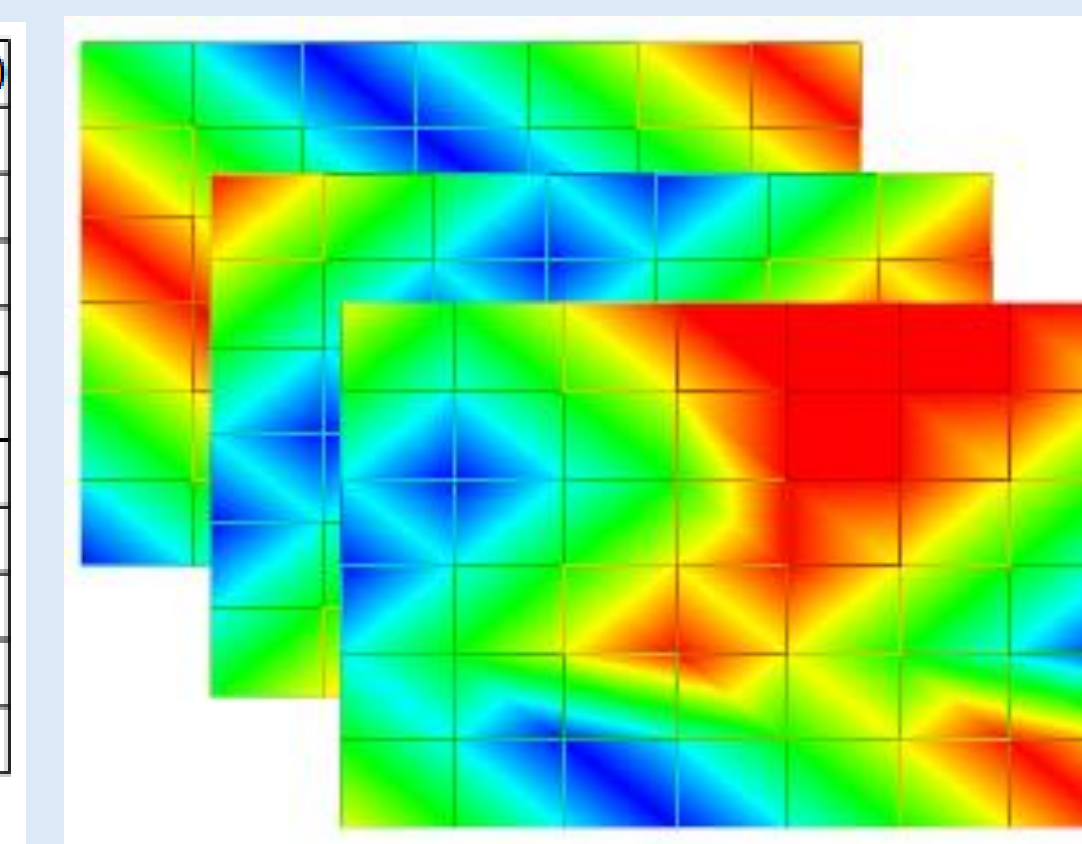
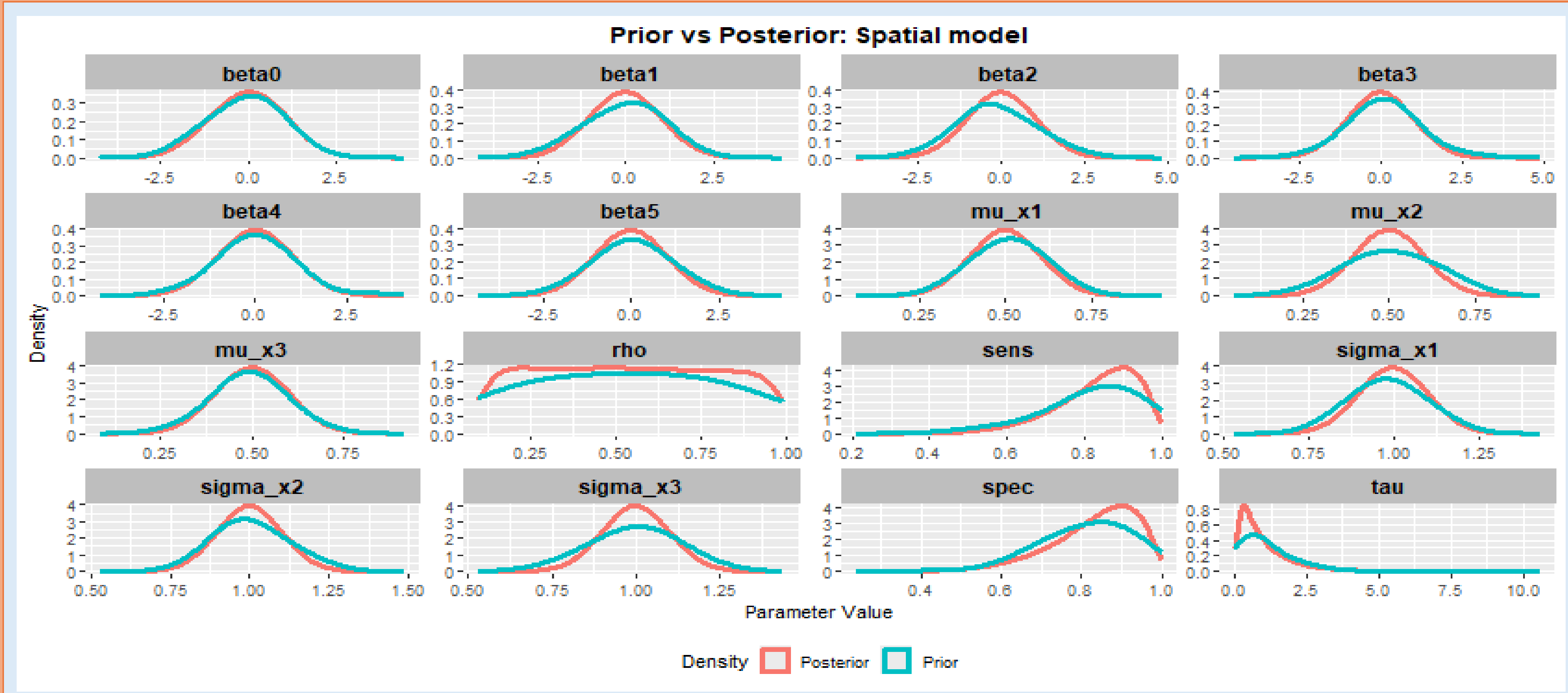
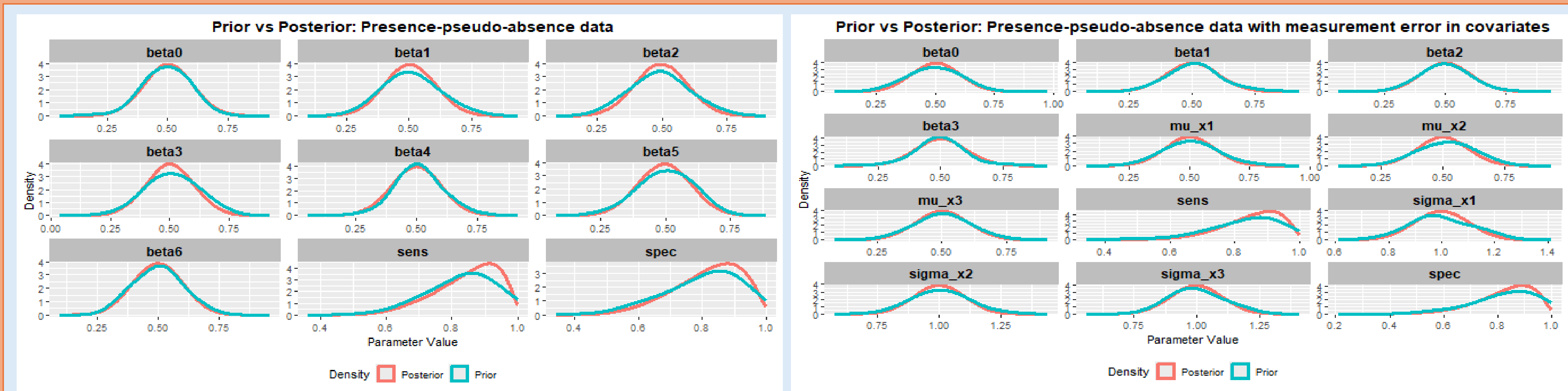


Figure 1: Example: Records for presence-only data

Consider A to be a finite space containing n grid cells; A_1, A_2, \dots, A_n and for each of the grid cells, there is a corresponding information of environmental covariates $\mathbf{x}(A_i)$. There is only records for presence in some of the grid cells. For example, in Figure 1, L_1 to L_{10} denote different locations (A_1 to A_{10}), S_1 to S_{10} are different species, 1 indicate species was recorded as present and "NA" means there is no records for species.

SBC: USING PRESENCE-PSEUDO-ABSENCE DATA



POSTERIOR PREDICTIVE CHECKS

