Supporting Information: Alan E. Gelfand and Shinichiro Shirota. 2019. Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. *Ecological Monographs.*

Appendix S2. A technical issue

In "What does "probability of presence" mean?" in the main manuscript, we argued that, under point-level modeling, the realized presence/absence surface should be *locally* constant, i.e., there should be areas where it takes only the value 1 and areas where it takes only the value 0. The $Y(\mathbf{s})$ surface provides conditionally independent Bernoulli trials and, therefore, will be everywhere discontinuous. However, they will be marginally dependent and smoothness of $p(\mathbf{s})$ will encourage a gridded image of a realization to offer a locally constant (0 or 1) appearance.

An alternative presence/absence specification to remedy this discontinuity is a first stage or *direct* model which introduces a latent Gaussian process at the first modeling level, setting $Y(\mathbf{s}) = 1(Z(\mathbf{s}) > 0)$, a function of $Z(\mathbf{s})$. If $Z(\mathbf{s})$ is a realization of a Gaussian process which is smooth, then the realized $Y(\mathbf{s})$ surface will be locally constant. For instance, if $Z(\mathbf{s}) = \mathbf{w}^T(\mathbf{s})\boldsymbol{\beta} + \omega(\mathbf{s})$, with $w(\mathbf{s})$ and $\omega(\mathbf{s})$ almost everywhere continuous, we will have this behavior. This first level modeling approach can be attractive for joint species distribution modeling (as in Clark et al. (2017)) since it

allows direct modeling of dependence between species rather than deferring it to the second stage (as in Ovaskainen et al. (2016)).

Unfortunately, a technical problem arises in fitting the direct model. This concerns the difference between the probability of presence surface, $p(\mathbf{s})$, that is, $\Phi(\mathbf{w}^T(\mathbf{s})\boldsymbol{\beta}+\omega(\mathbf{s}))$ under the second stage model and the realized presence surface under the direct model, $1(\mathbf{w}^T(\mathbf{s})\boldsymbol{\beta}+\omega(\mathbf{s})\geq 0)$. The realized presence surface has to "agree" well with the observed presences and absences while the probability of presence surface does not. At a given location, we can observe the presence of a species which has small probability of occurring or an absence which has a small probability of occurring. As a result, the probability of presence surface does not have to work as hard to fit the data. Specifically, with $\omega(\mathbf{s})$ in the modeling, under the direct model, the GP has to react strongly to observed presences and absences. Under second stage modeling, it can react less so. Therefore, when fitting the direct model, the flexibility of the GP results in the $\omega(\mathbf{s})$ surface becoming spiky in the neighborhood of a presence in order to explain well the observed presence.

Can we achieve a locally constant realized presence/absence surface and a smoothed probability of presence surface? Suppose we let $Y(\mathbf{s}) = 1, 0$ according to $Z(\mathbf{s}) \geq 0, < 0$. However, we introduce a second GP in specifying $Z(\mathbf{s})$, i.e., $Z(\mathbf{s}) = \mathbf{w}^T(\mathbf{s})\boldsymbol{\beta} + \omega(\mathbf{s}) + \gamma(\mathbf{s})$. Here, $\omega(\mathbf{s})$ has a larger range, a smaller decay parameter while $\gamma(\mathbf{s})$ has a smaller range with a larger decay parameter. (We are capturing the frequently used interpretation of the "nugget" as microscale dependence (Banerjee et al., 2014)). Then, we define the probability of presence surface as $p(\mathbf{s}) = P(Z(\mathbf{s}) \geq 0|\boldsymbol{\beta}, \mathbf{w}(\mathbf{s}), \omega(\mathbf{s})) = \Phi(\mathbf{w}^T(\mathbf{s})\boldsymbol{\beta} + \omega(\mathbf{s}))$ while we define the realized presence/absence surface as $1(Z(\mathbf{s}) \geq 0)$. Since $\gamma(\mathbf{s})$ is smooth, we will have locally constant behavior in this surface. The γ's will be spiky but the ω's will be smoother.

Strong prior information will be needed to control the decay parameters in the GP's. Specifi-

cally, we would impose an order restriction on the ranges or decays, demanding more rapid decay for the $\gamma(\mathbf{s})$ process. Additionally, we would impose a range for $\omega(\mathbf{s})$ which is related to the maximum inter-point distance available in the study region and a range for $\gamma(\mathbf{s})$ which is related to the smallest inter-point distance among the observed presence/absence locations.

If we let $\gamma(\mathbf{s})$ be a pure error process, then we would return to the problem of $Z(\mathbf{s})$ being everywhere discontinuous so that the realized $Y(\mathbf{s})$ surface would be everywhere discontinuous. However, a pure error process with very small variance will provide results similar to that for a GP with very short range, with very rapid decay and the pure error process model will be easier to fit. In fact, in the applications, we adopt the first stage model with pure error for $\gamma(\mathbf{s})$, i.e., $\gamma(\mathbf{s}) \sim \mathcal{N}(0, \tau^2)$ where we set $\tau^2 = 1$ for the identifiability of other parameters.

References

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2014). *Hierarchical Modeling and Analysis for Spatial Data, 2nd ed.* Chapman and Hall/CRC.

Clark, J. S., D. Nemergut, B. Seyednasrollah, P. Turner, and S. Zhange (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs 87*, 34–56.

Ovaskainen, O., D. B. Roy, R. Fox, and B. J. Anderson (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution 7*, 428–436.