

Modelado y estimación de ocupación para poblaciones y comunidades de especies bajo enfoque Bayesiano

Docentes: Vanesa Serafini, Facundo Contreras y Andrea Goijman.

Autor: Rodrigo J. Alonso

1- Describe brevemente tu sistema de estudio, indicando:

- **Objeto de estudio, N° de especies, cantidad de temporadas y años de muestreos**

Se estudió la comunidad de pequeños mamíferos en sistema de producción vacuna de la provincia de Buenos Aires. Estos sistemas corresponden a producción de leche (tambos) o de engorde de ganado a corral (*feedlots*). En total, el ensamble de pequeños mamíferos contó con siete especies (sp_1, ..., sp_7). Los muestreos se realizaron de forma estacional (verano, otoño, invierno y primavera) en cada establecimiento productivo, durante dos años. En cada año se muestrearon establecimientos distintos, siendo cuatro *feedlots* y cuatro tambos muestreados el primer año y otros cuatro *feedlots* y cuatro tambos el segundo año.

- **Metodología de obtención de datos (observación/cámaras trampas/trampas/otra)**

Para la obtención de los datos se utilizaron trampas de captura viva tipo Sherman y tipo jaula, dispuestas de forma pareada a lo largo de transectas. En cada transecta hubo entre 5 y 10 trampas de cada tipo, siendo la mayoría de ellas de 10 trampas.

- **Hipótesis de trabajo y objetivos**

El objetivo del trabajo es evaluar cómo el tipo de sistema (*feedlot* o tambo), la estación del año y el tipo de ambiente se asocian con el uso dado por las distintas especies. A su vez, se busca evaluar cómo la temperatura media y la remoción de individuos afecta la detectabilidad de las distintas especies

Las hipótesis para la *ocupación*(uso) son:

Las especies comensales usan sitios asociados a mayor actividad antrópica.

Las especies comensales no difieren en el uso de los sitios de acuerdo con las estaciones del año.

Las especies nativas usan sitios asociados a ambientes naturales.

Las especies nativas difieren en el uso de los sitios de acuerdo con las estaciones del año.

Las hipótesis para la *detectabilidad* son:

La detectabilidad de las especies es afectada negativamente por la disminución de la temperatura media diaria.

La detectabilidad de las especies es afectada negativamente por la remoción de las especies (capturas) realizadas los días previos.

- **¿Cuál es el sitio? (Unidad experimental/observacional) ¿Cuántos sitios hay? ¿Cómo se encuentran distribuidos?**

El sitio es cada transecta en cada campo. Hay 25 sitios por año. Los sitios están distribuidos dentro de cada establecimiento de forma tal de capturar la mayor cantidad de ambientes presentes.

- **¿Cuáles son las repeticiones? ¿Cuántas repeticiones hay?**

Cada noche de capturas es una repetición. Hay tres repeticiones por muestreo, es decir, cada transecta es visitada durante tres días consecutivos.

- **Indicar qué supuestos del modelo de ocupación se cumplen y cuáles no se cumplen. En caso de no cumplirse algún supuesto, ¿Qué modificaciones le realizarías a tu diseño para que los supuestos se cumplan?**

¡Zaran...! diría Jorge Suspenso (referencia: <https://www.youtube.com/watch?v=jM713XxPGIE>)

Supuesto de independencia entre sitios: de acuerdo con los estudios realizados por Montes de Oca et al. (2017) sobre el movimiento de los individuos de la especie dominante en estos sistemas (sp_1), no suelen moverse más allá de 50m del sitio de captura. Las transectas utilizadas en el estudio están separadas *al menos* 50m, llegando a encontrarse a varios cientos de metros de distancia en varios casos. Desde este punto de vista, podríamos suponer que el supuesto se cumple. Por otra parte, entiendo que el supuesto se establece para evitar que un individuo capturado en un sitio pueda ser capturado en otro sitio, y que, por lo tanto, ocupe dos sitios a la vez. En este estudio los individuos son removidos, es decir, la probabilidad de que eso ocurra es cero. Sin embargo, decidimos utilizar la palabra *uso* en lugar de *ocupación* de forma tal de evitar conflictos en la interpretación de los resultados y ser más conservadores con las conclusiones.

Supuesto de población cerrada: este supuesto definitivamente *no* lo cumplimos porque, como se mencionó anteriormente, los individuos son removidos. La tasa de supervivencia/mortalidad está siendo afectada por el propio muestreo. Para ello, lo que hicimos fue poner como una covariable de detección la noche de muestreo (noche 1, 2 o 3) y ver si el transcurso de las noches afectaba la detectabilidad de las distintas especies. En caso de no ver un efecto de la noche, podríamos suponer que la población es mucho más grande que el número de individuos capturado y que la remoción no estaría afectando las tasas vitales de las poblaciones (el supuesto, por lo tanto, se cumpliría).

Supuesto de *falsos positivos*: Suponiendo que el observador es experto, la identificación de las especies del estudio no debería producir errores en la determinación de esta. Por lo tanto, este supuesto se cumpliría.

Supuesto de *detección homogénea*: Probablemente este supuesto no sea cierto, pero es modelado con las distintas covariables. Cada especie, así, debería tener su propia probabilidad de detección, relacionada a su vez, con la covariable propuesta para su modelado.

Pensando en qué deberíamos modificar del diseño experimental, lo primero que debería ser modificado es el incumplimiento del supuesto de población cerrada, es decir, no debería realizarse una remoción. Este estudio no fue pensado para analizar ocupación, sino con otros fines. La dificultad reside en explicarle a un productor que vamos a capturar estos individuos y los vamos a liberar en el mismo lugar... En segundo lugar, correspondería chequear o pensar en el supuesto de independencia. Si bien el estudio sobre la especie dominante de Montes de Oca et al. (2017) es una base, aún resta estudiar el movimiento del resto de las especies, sobre las cuales hay nula o poca información.

2- Indica tu modelo de detección y ocupación incluyendo las covariables (en caso de tenerlas, y sus subíndices) que consideres relevantes para tu estudio en cada uno de los modelos.

Siendo: i= sitio, j=repeticion, k = especie, l = estación, m = año,

Modelo de proceso (teórico): $Z_{ik} \sim \text{Bernoulli}(\psi_{ik})$

Ajustado a los datos: $Z_{i,k,l,m} \sim \text{Bernoulli}(\psi_{i,l,k,m})$

Modelos para la heterogeneidad de especies ajustado a los datos (covariables):

$$\begin{aligned} \text{logit}(\psi_{i,l,k,m}) = & \text{lpsi}_k + \text{betalpsi1}_k * \text{summer}_{i,m,l} + \text{betalpsi2}_k * \text{fall}_{i,m,l} + \\ & \text{betalpsi3}_k * \text{winter}_{i,m,l} + \text{betalpsi4}_k * \text{spring}_{i,m,l} + \text{betalpsi5}_k * \text{tipo}_{i,m} + \text{betalpsi6}_k * \text{zanja}_{i,m} + \\ & \text{betalpsi7}_k * \text{galpon}_{i,m} + \text{betalpsi8}_k * \text{tambo}_{i,m} + \text{betalpsi9}_k * \text{vegetacion}_{i,m} + \\ & \text{betalpsi10}_k * \text{comedero}_{i,m} + \text{betalpsi11}_k * \text{silo}_{i,m} + \text{betalpsi12}_k * \text{foza}_{i,m} + \text{betalpsi13}_k * \text{tipo}_{i,m} \end{aligned}$$

Modelo de observación (teórico): $Y_{ik} \sim \text{Bernoulli}(Z_{ik} * P_{ijk})$

Ajustado a los datos: $y_{i,j,k,l,m} \sim \text{Bernoulli}(\mu_{p_{i,j,k,l,m}})$

$$\text{Donde } \mu_{p_{i,j,k,l,m}} = p_{i,j,k,l,m} * Z_{i,k,l,m}$$

Probabilidad de detección afectada por las covariables: $\text{logit}(p_{i,j,k,l,m}) = \text{lp}_k + \text{betalp1}_k * \text{temp}_{i,j,l,m} + \text{betalp2}_k * \text{noches}_{i,j}$

3- Adjunta tu base de datos y tu script de R en la carpeta de Drive (Enlace). Dentro de la carpeta de Drive, crea una carpeta con los apellidos y nombres de los integrantes.

Adjunto con este documento un archivo comprimido con la base de datos usada, junto con el environment, el modelo, las salidas del modelo (para que no lo tengan que correr).

4- Indique la convergencia y el ajuste del modelo (Indicando los valores que observó o gráficamente).

El modelo se corrió con los siguientes parámetros para las MCMC: 500 000 iteraciones, thin: 10, burn-in: 50 000 y el número de cadenas fue de 3.

Para todas las salidas (conjuntos de parámetros monitoreados), se observó el valor R-hat (Gelman & Rubin, 1992), y se consideró una buena convergencia si dichos valores se encontraron entre 1.00 y 1.10. Ningún estimador tuvo valores por fuera del intervalo (**Tabla 1**). Además, se analizó la convergencia gráficamente, utilizando las trazas de cada una de las cadenas (**Figura 1**) y la distribución posterior obtenida por cada cadena (**Figura 2**), como evidencia de “buena mezcla”. Por razones de espacio, en la tabla y en las figuras solo mostramos algunas de las salidas.

Tabla 1. Valores de Rhat para cada uno de los parámetros de la primera salida (out1). Los valores se encuentran entre 1 y 1.10, dando muestras de una buena convergencia de las cadenas.

<i>Parámetro</i>	<i>Especie</i>						
	sp_1	sp_2	sp_3	sp_4	sp_5	sp_6	sp_7
lp	1.00	1.00	1.00	1.00	1.00	1.00	1.00
lpsi	1.04	1.02	1.05	1.05	1.07	1.02	1.05
betalp1	1.00	1.00	1.00	1.00	1.00	1.00	1.00
betalp2	1.00	1.00	1.00	1.00	1.00	1.00	1.00
betalpsi1	1.05	1.05	1.04	1.06	1.05	1.03	1.03
betalpsi2	1.05	1.05	1.05	1.06	1.05	1.03	1.04
betalpsi3	1.05	1.05	1.05	1.06	1.05	1.04	1.03
betalpsi4	1.05	1.05	1.05	1.06	1.05	1.04	1.03
betalpsi5	1.00	1.00	1.00	1.00	1.00	1.00	1.00
betalpsi6	1.00	1.00	1.00	1.00	1.01	1.01	1.01
betalpsi7	1.00	1.00	1.00	1.00	1.00	1.00	1.01

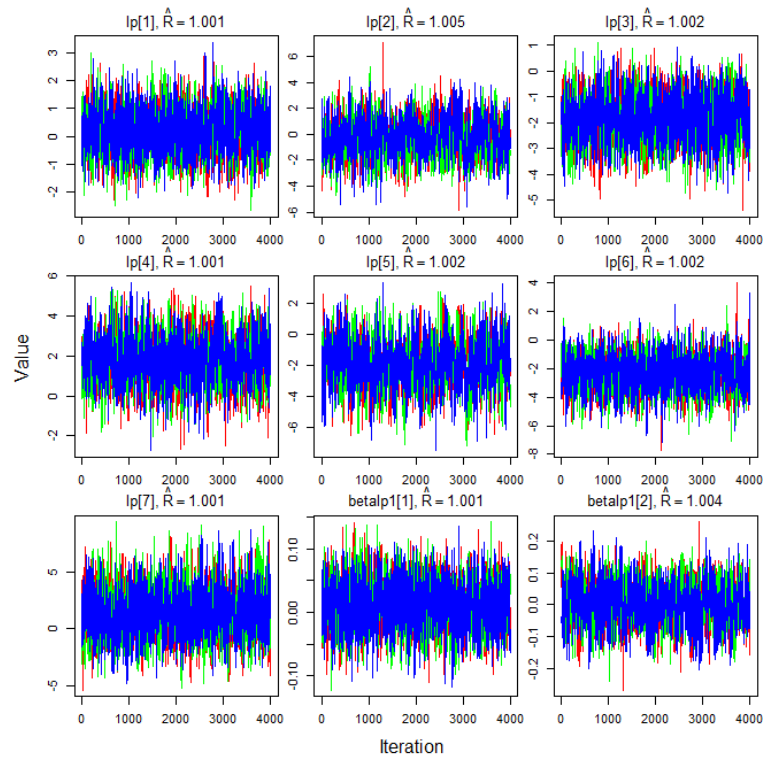


Figura 1. Gráfico de trazas de las MCMC. Se observa una “buena mezcla” de los valores tomados por las tres cadenas.

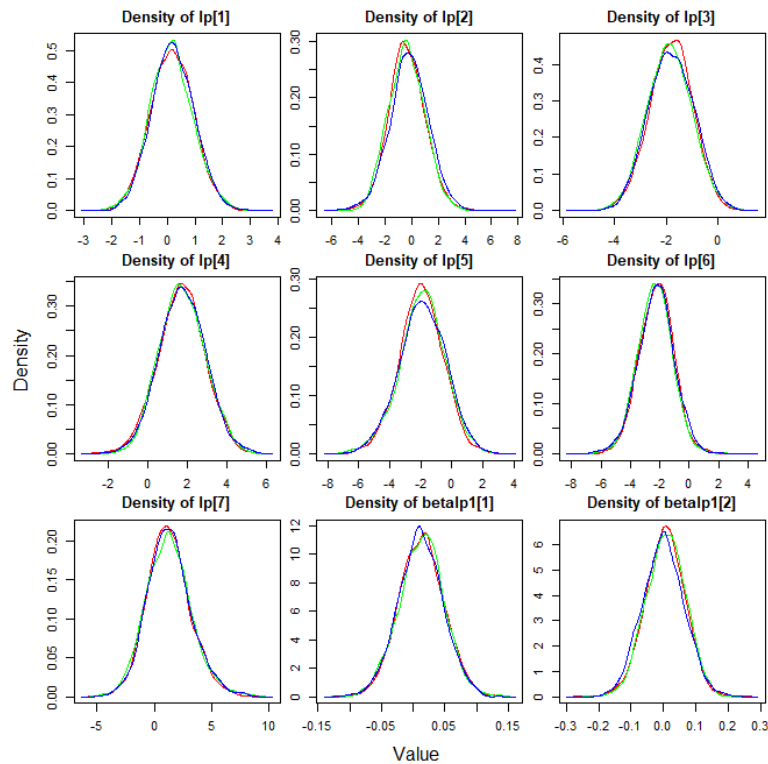


Figura 2. Gráfico de las distribuciones posteriores estacionarias de cada MCMC. Se observa que las tres cadenas obtuvieron distribuciones posteriores estacionarias similares.

5- Presente sus resultados en forma escrita y gráfica.

Resultados sobre la *detectabilidad*:

Nosotros evaluamos el efecto de la temperatura media diaria de la repetición y de la noche de muestreo sobre la probabilidad de detección. De acuerdo con el modelo, la temperatura media diaria no tuvo ningún efecto sobre la detección. En cambio, la noche de muestreo tuvo un efecto negativo sobre algunas de las especies, particularmente sobre la sp_1, la sp_4 y levemente sobre la sp_7 (**Figura 3 y 4**). A partir de este resultado, podríamos concluir que no se cumple uno de los supuestos del modelo (población cerrada), por lo tanto, queda planteado pensar cómo incorporar en el modelo el efecto de la remoción (una suerte de corrección sobre la detección en el evento $n+1$ en caso de tener captura en el evento n).

Efecto de la noche de muestreo sobre la detectabilidad (betalp2)

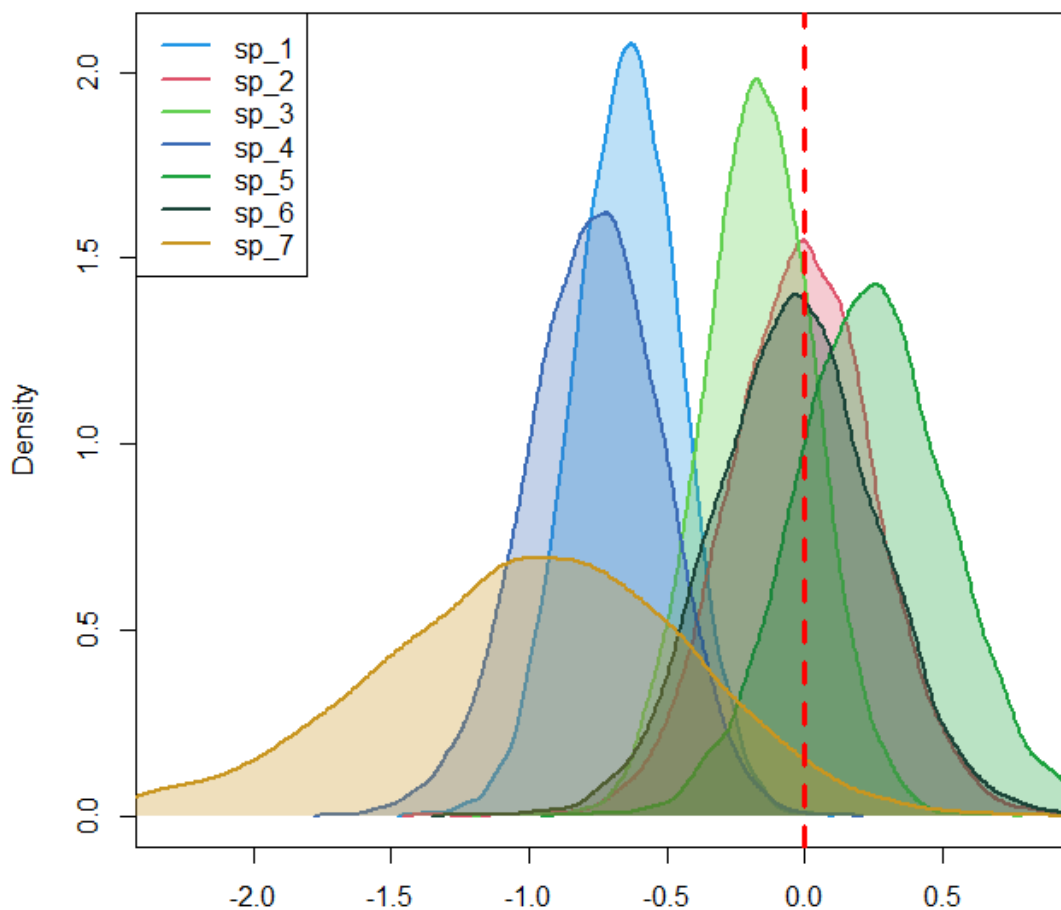


Figura 3. Distribución posterior del $\text{logit}(p)$ debido al efecto de la remoción de individuos de cada especie. La línea punteada marca el cero. Las curvas de las distribuciones posteriores están limitadas entre el CRI del 95%.

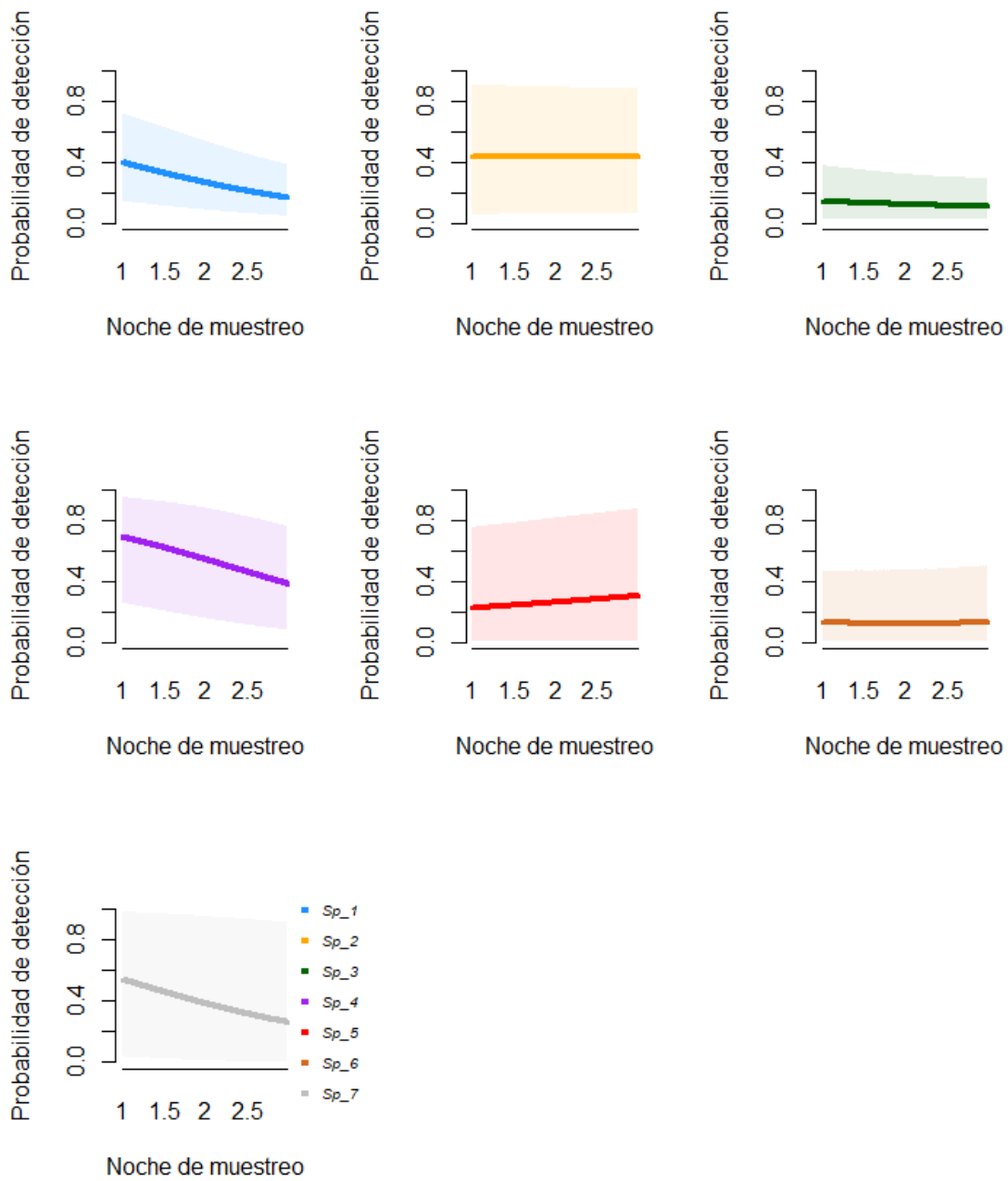


Figura 3. Probabilidad de detección (p) debido al efecto de la remoción de individuos de cada especie. Las bandas claras muestran el CRI del 95%.

Resultados sobre el uso (ocupación):

De acuerdo con nuestras hipótesis, la estacionalidad tendría un efecto en el uso de los distintos sitios de acuerdo con el carácter comensal o nativo de las especies. Sin embargo, el modelo no arrojó diferencias en la probabilidad de uso atribuibles a la estación para ninguna especie (**Figura 5**).

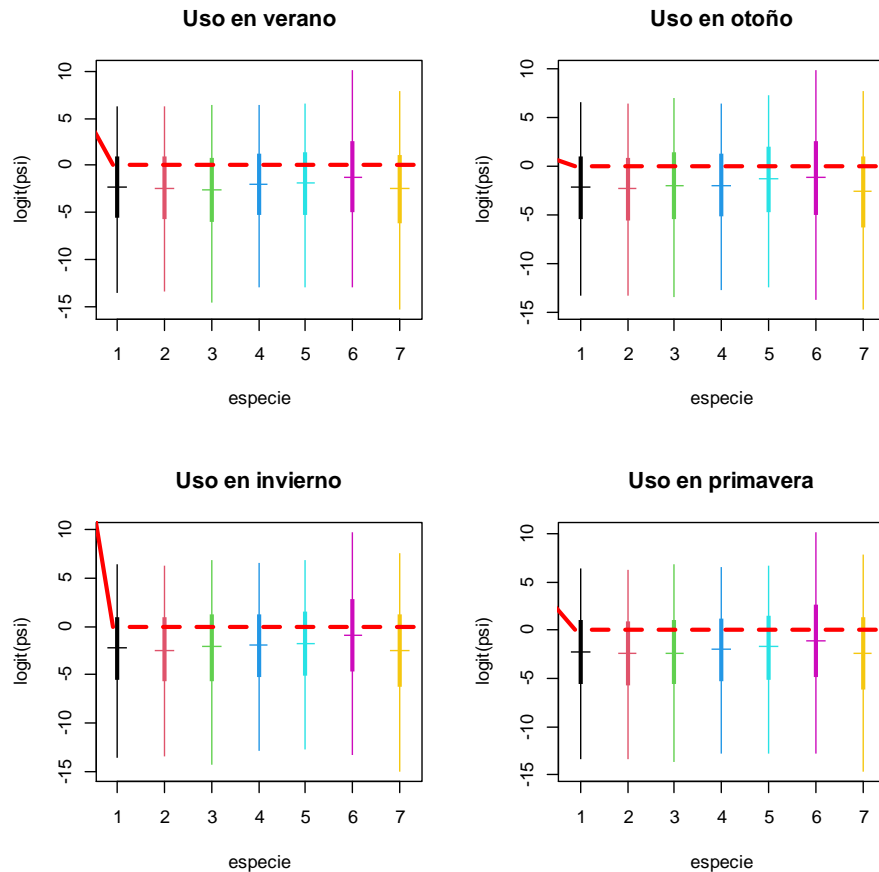


Figura 5. Distribución posterior del $\text{logit}(\psi_i)$ debido al efecto de la estación en cada especie. La línea punteada marca el cero. Las barras horizontales señalan la mediana de la distribución posterior. Barras verticales anchas: CRI del 50%. Barras verticales finas: CRI del 95%.

Por otra parte, se propuso que las especies comensales (sp_1, sp_2 y sp_4) usarían más los sitios en ambientes antropizados que las especies nativas. Los resultados apoyan esta hipótesis parcialmente, dado que los tipos de ambientes que tuvieron efectos sobre las especies corresponden a ambientes antrópicos. En el caso del ambiente zanja (que corresponde a desagotes y lagunas con desechos, ambiente por lo general patogénico), este fue usado por las especies comensales sp_1, sp_2 y sp_4, aunque también por la especie nativa sp_3 (**Figura 6A**). El ambiente galpón (guardado de alimentos, maquinaria y herramientas) fue utilizado por casi todas las especies, tanto nativas como comensales, a excepción de la especie sp_7 (**Figura 6B**). En el caso del ambiente tambo (lugar de ordeño), este propició el uso de los sitios por las especies comensales sp_2 y sp_4 (**Figura 6C**). En el caso de los silos, las transectas ubicadas en estos ambientes fueron usadas por la especie nativa sp_7 (**Figura 6D**) y, por último, las transectas ubicadas en el ambiente foza (cadáveres de vacas), fueron utilizadas por la especie comensal sp_4 (**Figura 6E**).

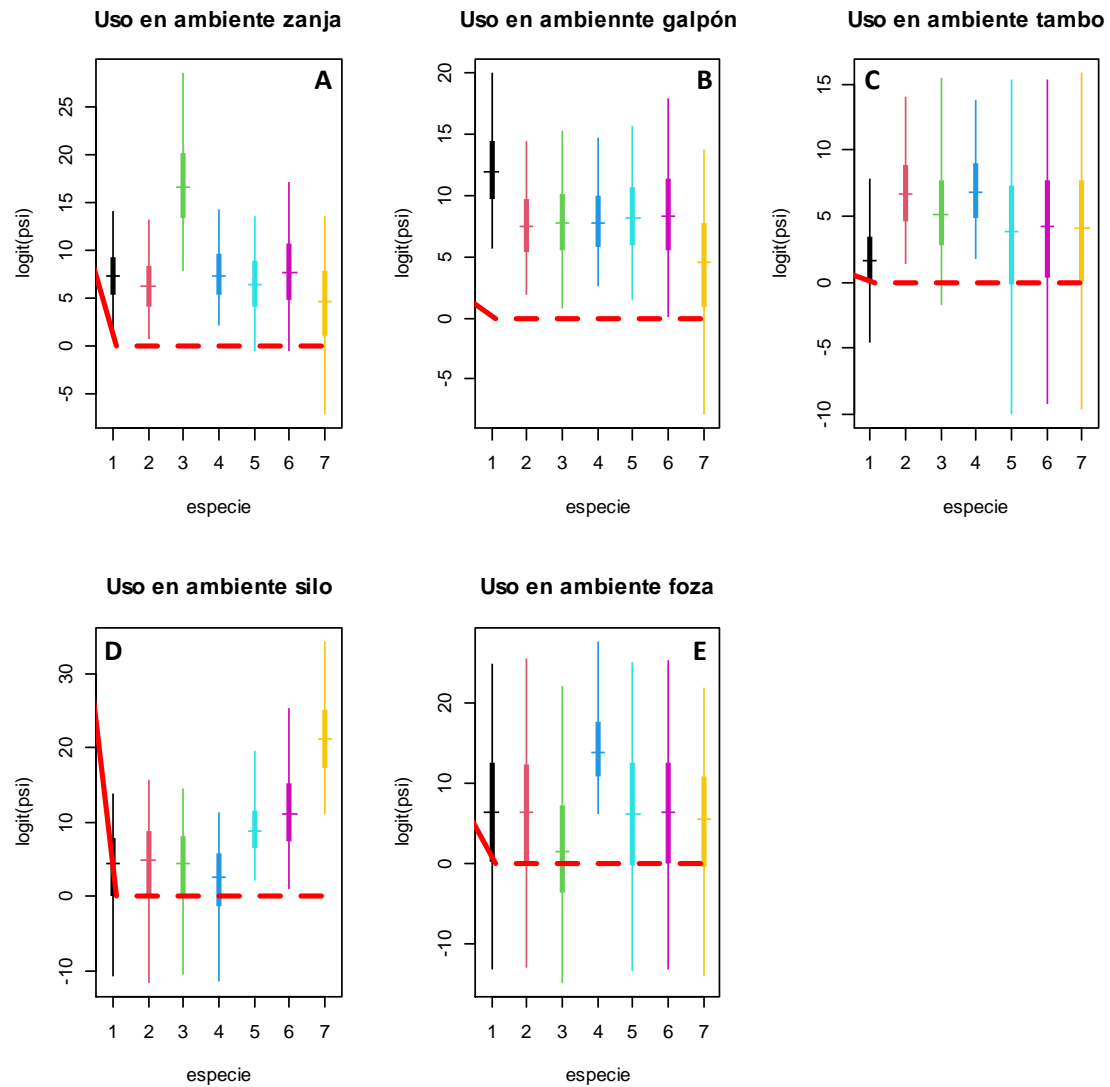


Figura 6. Distribución posterior del $\text{logit}(\text{psi})$ debido al efecto del ambiente sobre cada especie. La línea punteada marca el cero. Las barras horizontales señalan la mediana de la distribución posterior. Barras verticales anchas: CRI del 50%. Barras verticales finas: CRI del 95%.

Finalmente, ningún ambiente natural (vegetación) tuvo efecto sobre las especies del ensamble. Tampoco se observaron diferencias en el uso de los sitios debido al tipo de sistema productivo donde se encontraban inmersos, sean estos *tambos* o *feedlots*.