

Disease Mapping: Modelo Logístico, INLA y Reproducibilidad

Autores: Santiago Moncalero, Diego Fernández y Jorge Guitart

1. Introduccion

La reproducibilidad y la replicabilidad son pilares fundamentales de la investigación científica. Estos conceptos no solo permiten verificar y validar los análisis realizados por terceros, sino que también facilitan su reutilización en nuevos contextos, fortaleciendo la confianza en los resultados obtenidos y en su validez empírica.

Esta práctica se centra en el mapeo de enfermedades (disease mapping), un área clave de la bioestadística que combina datos geográficos y modelos estadísticos para identificar patrones espaciales de incidencia o mortalidad. Para ello, se emplean dos enfoques metodológicos: WinBUGS, basado en simulaciones de Monte Carlo, e INLA, que utiliza aproximaciones determinísticas. Ambos métodos se aplican para estimar tasas de mortalidad ajustadas y probabilidades espaciales en la región de Aragón.

El objetivo principal de este trabajo es comparar los resultados obtenidos con cada enfoque, analizando su consistencia y destacando las diferencias clave en sus predicciones.

2. Resultados: Mapas de RME y $RME > 1$

En este apartado, se presentan los mapas resultantes del análisis de tasas de mortalidad ajustadas, específicamente el Riesgo Relativo Estandarizado (RME) y las áreas donde $RME > 1$, utilizando dos metodologías: WinBUGS e INLA.

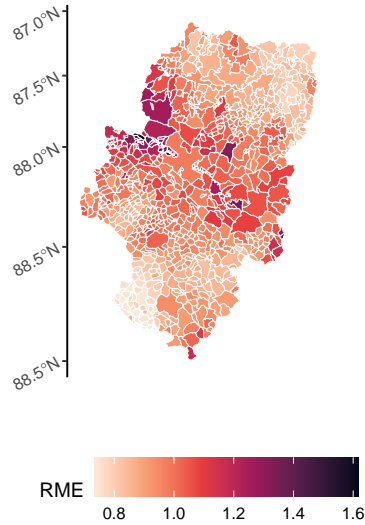
Los mapas generados para ambos enfoques muestran una alta similitud en la distribución espacial de los valores estimados. Esto se evidencia tanto en la intensidad de los colores como en los patrones espaciales observados, lo que refuerza la robustez de las estimaciones obtenidas. En particular, las áreas con $RME > 1$, que indican un riesgo relativo superior al promedio, son consistentes entre los modelos, lo cual es relevante para identificar zonas de mayor mortalidad ajustada.

La consistencia entre los resultados de WinBUGS e INLA se puede atribuir a la calidad del ajuste de los modelos y a la naturaleza de los datos. Ambos enfoques utilizan modelos jerárquicos bayesianos para incorporar efectos espaciales y ajustar las tasas de mortalidad observadas, lo que garantiza una adecuada captura de patrones subyacentes. Sin embargo, es importante destacar que, aunque los algoritmos utilizados son distintos (Monte Carlo en WinBUGS y aproximaciones determinísticas en INLA), los

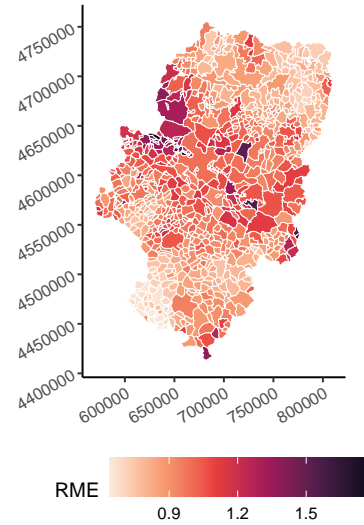
dos métodos logran resultados similares debido a su capacidad para modelar de manera efectiva las estructuras espaciales en los datos.

Esta convergencia en los resultados refuerza la validez de los modelos implementados y sugiere que ambos métodos son herramientas confiables para el análisis de datos espaciales en contextos de mapeo de enfermedades. La elección entre uno u otro dependerá, en última instancia, de las necesidades específicas del análisis, como la velocidad de procesamiento (ventaja de INLA) o la exploración de incertidumbre más detallada (fortaleza de WinBUGS).

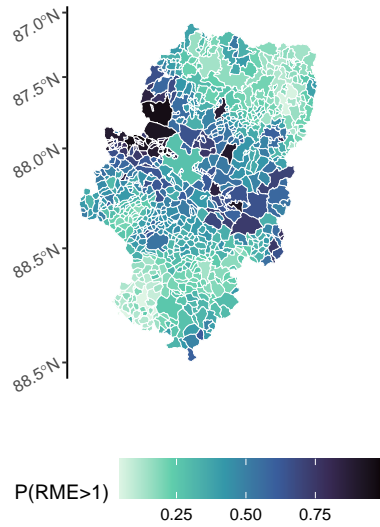
Aragón, Media posterior (INLA)



Aragón, Media posterior (WinBUGS)



Aragón, P(RME > 1) (INLA)



Aragón, P(RME > 1) (WinBUGS)

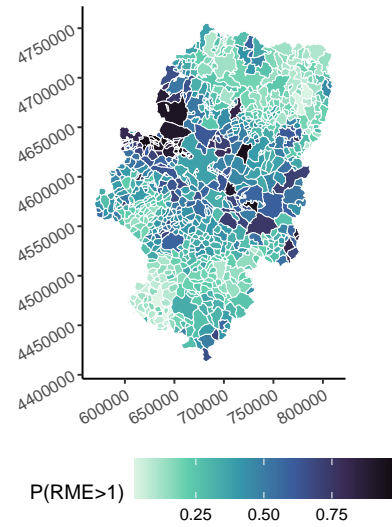


Figura 1: Mapeo de modelos realizados por INLA y WinBUGS.

3. Correlación entre Resultados de WinBUGS e INLA

Para evaluar cuantitativamente la similitud entre los resultados obtenidos mediante las metodologías WinBUGS e INLA, se calculó la correlación entre las estimaciones de RME y las probabilidades de que $RME > 1$.

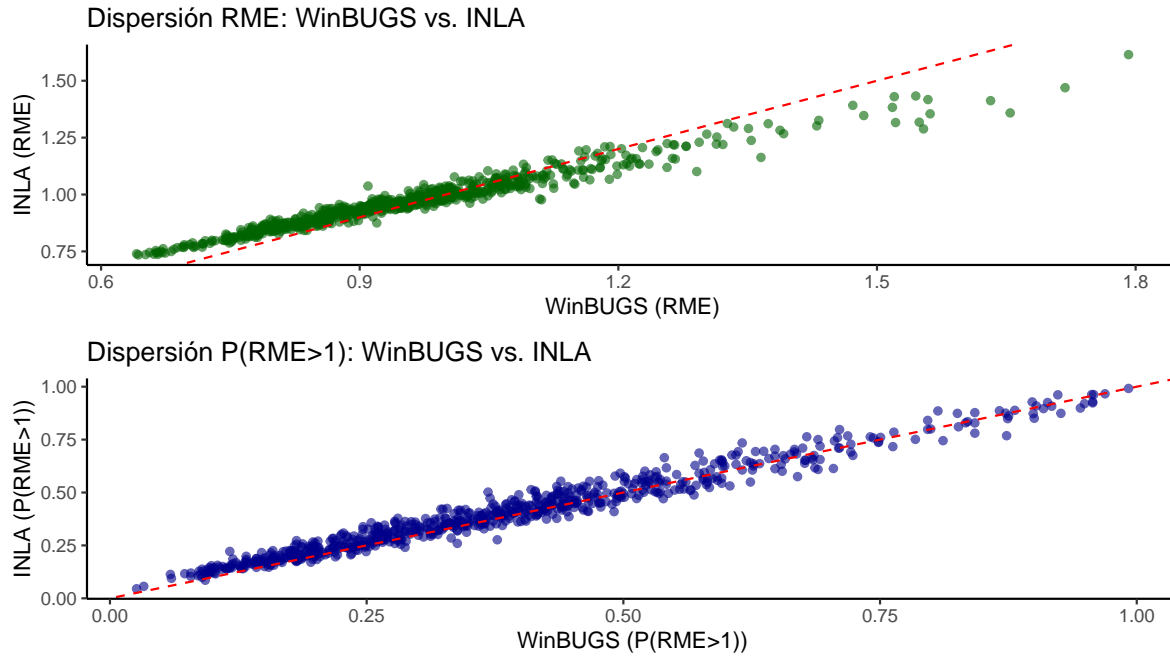


Figura 2: Dispersiones entre modelos.

Los coeficientes de correlación obtenidos son notablemente altos:

- Correlación entre RME: 0.9786469
- Correlación entre $P(RME > 1)$: 0.9830644

Estos valores reflejan una excelente concordancia entre los dos enfoques, lo que respalda la consistencia de las estimaciones generadas independientemente del método utilizado. Tal nivel de similitud indica que ambos modelos logran capturar de manera adecuada los patrones subyacentes en los datos, tanto en las tasas ajustadas como en las probabilidades asociadas a riesgos relativos superiores al promedio.

4. Comparación de las Densidades Esperadas

En este apartado, se comparan las distribuciones de las densidades esperadas estimadas por los modelos WinBUGS e INLA. Estas densidades reflejan la distribución a posteriori del RME en toda la región de estudio, permitiendo evaluar las características globales de las tasas ajustadas obtenidas con cada metodología.

Al graficar las densidades esperadas de ambos métodos, se observa una similitud destacable en las formas de las distribuciones. Ambas presentan una distribución unimodal con patrones que indican consistencia en las estimaciones. Las diferencias, aunque mínimas, pueden atribuirse a la forma en que cada metodología aborda el ajuste:

- WinBUGS: Utiliza simulaciones MCMC, lo que permite explorar con mayor profundidad la incertidumbre en las estimaciones, pudiendo generar ligeras variaciones en las colas de la distribución.
- INLA: Recurre a aproximaciones determinísticas, lo que se traduce en estimaciones más rápidas, aunque puede suavizar ligeramente las colas de la distribución. A pesar de estas diferencias técnicas, las distribuciones esperadas de ambos modelos muestran un alto grado de solapamiento, lo que refuerza la confianza en los resultados obtenidos y la consistencia de las metodologías aplicadas.

La similitud en las densidades esperadas refleja que ambos enfoques capturan de manera adecuada los patrones subyacentes en los datos y que las diferencias metodológicas no impactan significativamente en la interpretación global de los resultados. Esto demuestra que ambos métodos son igualmente válidos y confiables para realizar análisis de mapeo de enfermedades en contextos similares.

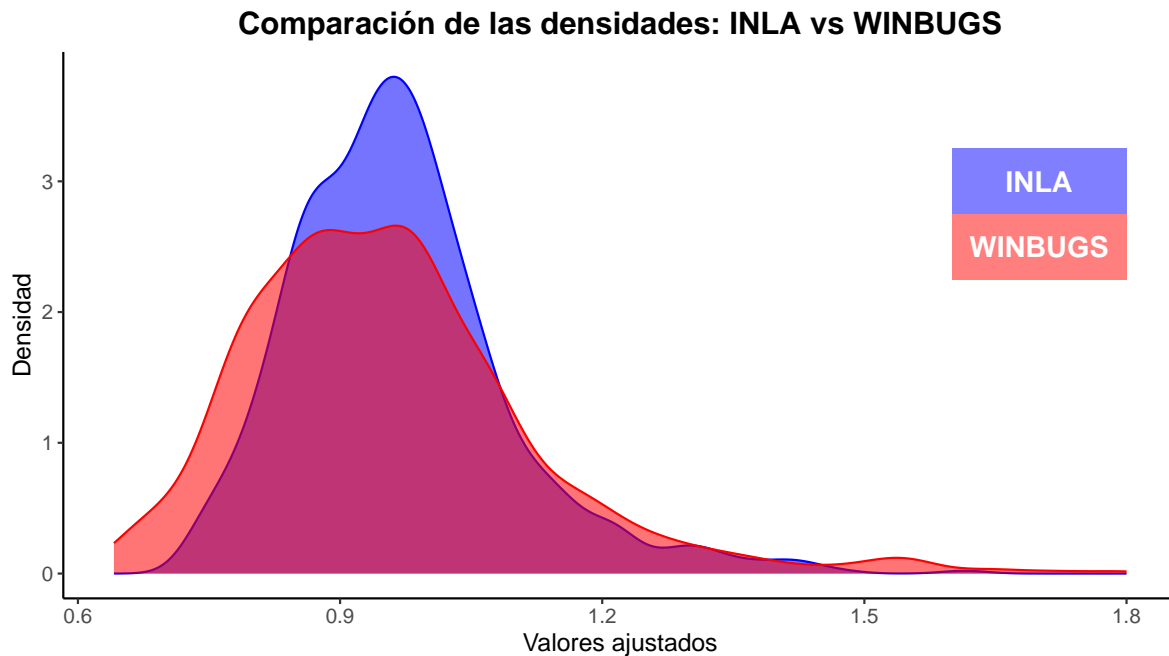


Figura 3: Gráficos de densidad de la media.

5. Conclusiones

En esta práctica se ha llevado a cabo un análisis de mapeo de enfermedades empleando el modelo de Besag, York, y Mollié con dos metodologías diferentes: WinBUGS e INLA. El objetivo era estimar las tasas de mortalidad ajustadas y las probabilidades de que dichas tasas superen el valor 1 en la región de Aragón. Los resultados se han visualizado mediante mapas de la media posterior y probabilidades ajustadas, que han permitido identificar de manera clara las áreas de mayor y menor riesgo relativo.

Por otro lado, la comparación entre los dos métodos ha demostrado una alta consistencia en las predicciones, como lo reflejan tanto los gráficos como los coeficientes de correlación elevados entre las estimaciones de RME y $P(\text{RME} > 1)$. Estas similitudes destacan la robustez de ambos enfoques para capturar los patrones espaciales subyacentes en los datos, a pesar de las diferencias metodológicas en su implementación.

Estos ensayos son plenamente reproducibles, ya que se han establecido semillas para garantizar la replicación de los resultados, y el código completo del proyecto está disponible en GitHub, acompañado de un archivo README que facilita su descarga y ejecución.