

Modelación espacial bayesiana de la marginación en localidades de Oaxaca, México a través de INLA

Cuauhtémoc Calderón Villarreal*
Gerardo Núñez Medina**

(Recibido: octubre, 2018/Aceptado: marzo, 2019)

Resumen

El objetivo principal de este trabajo es modelar el comportamiento espacial de la marginación en localidades de Oaxaca empleando modelos gaussianos latentes. La propuesta busca representar un campo gaussiano (GF) continuo a través de un campo aleatorio gaussiano markoviano (GMRF) discreto. Los datos provienen de estimaciones generadas por el Consejo Nacional de Población y del Censo General de Población 2010 levantado por el Instituto Nacional de Estadística y Geografía (INEGI). El ajuste se realizó empleando el paquete R-INLA. Los resultados parecen indicar que la marginación presenta un comportamiento espacial homogéneo en prácticamente la totalidad del territorio de Oaxaca, con algunas zonas de baja marginación y otras de muy alta marginación, a la vez que la incorporación de covariables permite mejorar el ajuste del campo gaussiano.

Palabras claves: análisis bayesiano, bienestar, marginación.

Clasificación JEL: C11, I31.

* Investigador de El Colegio de la Frontera Norte, A.C., investigador Nacional nivel 3 y miembro de la Academia Mexicana de Ciencias, A.C. Correo: <calderon@colef.mx>.

** Profesor en el Centro Mesoamericano de Salud Pública y Desastres de la Universidad Autónoma de Chiapas.

Bayesian space modeling of marginalization in localities of Oaxaca, Mexico through INLA

Abstract

The main objective of this work is to model the spatial behavior of marginalization in localities of Oaxaca using latent Gaussian models. The proposal seeks to represent a continuous Gaussian field (GF) through a discrete Gaussian Markovian (GMRF) random Field. The data comes from estimates generated by the National Population Council and the general population census 2010 taken by the National Institute of Statistics and Geography (INEGI). The adjustment was made using the R-INLA package. The results seem to indicate that the marginalization presents a homogeneous spatial behavior in practically the entire territory of Oaxaca, with some areas of low marginalization and others of very high marginalization, while the incorporation of covariates allows to improve the adjustment of the Gaussian countryside.

Keywords: bayesian analysis, welfare, marginalization.

JEL classification: C11, I31.

1. Introducción

La marginación es un concepto que busca dar cuenta de la capacidad de acceso de las comunidades a benefactores como caminos, servicios e infraestructuras, sin embargo, la marginación no es un atributo individual sino de nivel agregado, por lo que, es importante distinguir entre marginación y marginalidad. Mientras la marginación es un concepto válido y aplicable a poblaciones, la marginalidad se construye y es válida en el nivel individual (Cortés, 2006).

La marginación puede considerarse un atributo propio de las poblaciones y dado que en la actualidad la mayor parte de las poblaciones humanas son sedentarias, tiene sentido suponer que la marginación presenta un importante nivel de correlación con el espacio. En general, es de esperarse que el nivel medio de marginación de una población responda en gran medida

a la posición geográfica que ocupa, lo que a su vez define su relación con las poblaciones vecinas y su capacidad para acceder a una importante cantidad de bienes y servicios (Bustos, 2009).

En México, el Consejo Nacional de Población (CONAPO) define la marginación como una medida que busca dar cuenta de la capacidad de acceso a benefactores, por parte de las comunidades que habitan un espacio geográfico determinado: *...el índice de marginación es una medida-resumen que permite diferenciar entidades federativas y municipios según el impacto global de las carencias que padece la población, como resultado de la falta de acceso a la educación, la residencia en viviendas inadecuadas, la percepción de ingresos monetarios insuficientes y las relacionadas con la residencia en localidades pequeñas* (CONAPO, 2012).

Con base en la definición anterior, CONAPO, genera periódicamente un índice de marginación que busca cuantificar el impacto de las carencias provocadas por la falta de acceso a servicios de educación, servicios en la vivienda, bajos ingresos y residencia en localidades pequeñas (Cortés, 2002; Camberos, 2007). El índice se calcula para distintos niveles de división territorial, entre las que se distinguen las entidades federativas, municipios, localidades y zonas urbanas (Gutiérrez y Gama, 2010).

Una de las principales razones para analizar el comportamiento espacial de la marginación en Oaxaca en 2010, deriva del hecho de que Oaxaca ocupó el tercer lugar en marginación a nivel nacional. El porcentaje de Municipios con un alto y muy alto grado de marginación fue del 63.2%, lo que representó 360 de los 540 municipios que le integran, mientras que a nivel de localidad, de las poco más de ocho mil localidades habitadas reportadas, 94.2% presentaron un grado de marginación alto o muy alto (7 652 localidades).

La hipótesis central de la investigación sostiene que la marginación puede ser modelada a través de relaciones espaciales, establecidas por medio de criterios de vecindad entre localidades de Oaxaca. Esto implica que el grado de marginación estaría asociado a la estructura de vecindades, donde la posición geográfica de la localidad y aspectos como la proximidad a la cabecera municipal o el número de vecinos juega un papel fundamental en la definición del nivel medio de marginación observado en una localidad.

Adicionalmente se intentará modelar el comportamiento espacial de la marginación a partir de covariables como: el porcentaje de servicios disponibles, la escolaridad promedio y el porcentaje de población desocupada, variables que de manera directa o indirecta pueden modificar el nivel medio de marginación observado en las localidades.

2. Fuentes de datos y Variables

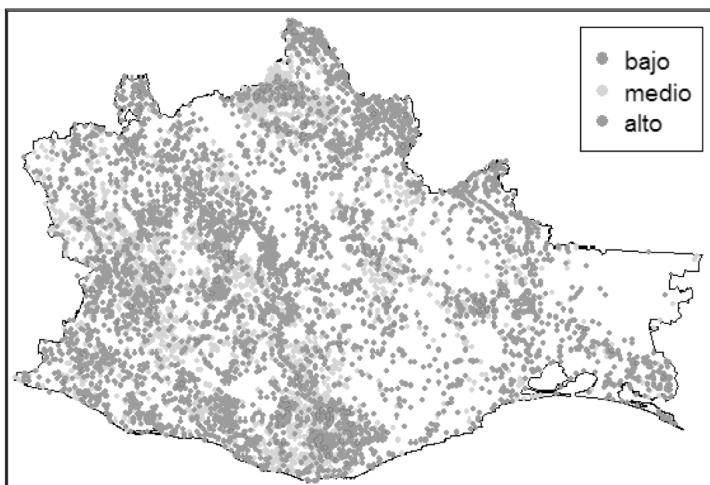
Los datos sobre marginación para las localidades de Oaxaca al 2010 fueron obtenidos directamente de las estimaciones realizadas por el CONAPO, mismos que se encuentran disponibles en línea (CONAPO, 2012). Los datos recabados comprenden un total de 8 127 localidades habitadas para el estado de Oaxaca al año señalado.

Es importante señalar que el CONAPO utilizó para la estimación del índice de marginación 2010, e información proveniente del Censo General de Población y Vivienda 2010. En este sentido fue posible referenciar espacialmente los datos de las covariables colectadas por el Censo de 2010, agregadas a nivel de localidad para Oaxaca.

Según la clasificación de CONAPO al 2010, un total de 7 652 localidades presentan un grado de marginación alto o muy alto (2 367 muy alta marginación y 5 285 de alta marginación). Esto significa que 94.1% de las localidades de Oaxaca carecen de servicios básicos como: drenaje, energía eléctrica, agua entubada y que la población que habita en ellas presentan altos niveles de analfabetismo (no saben leer y escribir), no concluyó la primaria, habita en viviendas con piso de tierra y no perciben ingresos por arriba de dos salarios mínimos. Esta situación afecta a 2.46 millones de personas que habita en localidades de alta o muy alta marginación (330 mil personas habitan localidades de muy alta marginación y 2.13 millones habitan localidades de alta marginación) lo que afecta al 65% de la población total de Oaxaca.

La distribución espacial de las localidades y su nivel de marginación puede verse con mayor claridad en el mapa 1, donde sólo 191 localidades tienen un grado de marginación bajo o muy bajo, lo que equivale al 2.3% de las localidades.

Mapa 1
Distribución espacial de los niveles de marginación
por localidad, Oaxaca 2010



Fuente: elaboración propia con datos de CONAPO, 2010.

Adicional al índice de marginación se incluyeron tres variables explicativas:

- x_1 : porcentaje de servicios disponibles en la localidad,
- x_2 : escolaridad promedio de la localidad y
- x_3 : porcentaje de población desocupada por localidad.

La variable x_3 tiene por objeto describir las condiciones de empleo presentes en las localidades. Es de esperar que las variaciones espaciales registradas en el nivel de servicios disponibles en cada localidad así como el nivel de educación ejerzan un efecto sobre los niveles observados de marginación, en las localidades de Oaxaca.

Los resultados sobre infraestructura y servicios de las localidades fueron recopilados de la base de datos complementaria al Censo de Población y Vivienda 2010. La misma contiene variables que informan sobre la conexión a carretera, disponibilidad y tipo de transporte público a la cabecera municipal, así como el tiempo de traslado, frecuencia y costo, además de variables que indican la existencia de servicios públicos y equipamiento como: escuelas por nivel educativo, servicios de salud, telecomunicaciones,

establecimientos comerciales, servicios de abasto de alimentos y recintos culturales y deportivos. Todos los datos permitieron generar la variable x_1 que captura la proporción de infraestructuras y servicios presentes en una localidad en relación al número máximo de servicios posibles (62).

La variable x_2 captura los años promedio de escolaridad alcanzados por la población de 12 años y más que habita en la localidad, mientras que la variable x_3 concentra el porcentaje de población desocupada por localidad. Resulta al dividir la población desocupada de 15 a 65 años entre la población en edad laboral residente en la localidad.

3. Modelación geoestadística

La modelación espacial de los datos de marginación (de las localidades de Oaxaca) tiene por objeto identificar el proceso que especifica los mecanismos subyacentes que dan origen a dichos datos, además de estimar el valor de los componentes paramétricos que especifican la distribución que identifica el proceso.

La variable respuesta Y_i representa el nivel medio de marginación al 2010 para cada una de las localidades del estado de Oaxaca, ubicadas en el punto $i \in \mathbb{R}^2$ con coordenadas (latitud, longitud) lo que define el dominio espacial del proceso (Blangiardo & Cameletti, 2015). La distribución de probabilidad asociada a los datos observados Y_i está en función de la distribución asociada al vector Θ de parámetros y al vector ψ de hyperparámetros, y se define a partir de:

$$Y|\theta, \psi \sim \prod_i f(y_i|\theta, \psi, \eta_i)$$

donde los vectores Θ y ψ asociados a la distribución Y se definen a través de.

$$E(Y|\eta) = f^{-1}(\eta)$$

De manera que la respuesta media Y_i se mapea mediante una función de enlace al predictor lineal η que toma la forma de un modelo de regresión aditivo estructurado (Rue & Lindgren, 2015), también conocido como modelo gaussiano latente, cuya forma está dada por:

$$\eta_i = \beta_0 + \sum_{m=1}^M \beta_m X_{mi} + \xi_i u_i \quad (1)$$

donde la i-ésima observación y_i tiene una distribución:

$$y_i \sim Normal(\eta_i, \sigma_e^2)$$

σ_e^2 es la varianza del error e_i , que se supone independiente de e_j cuando $i \neq j$, y se conoce como efecto nugget (Ingebrigtsen *et al.*, 2014). El término

$$\beta_0 + \sum_{m=1}^M \beta_m X_{mi}$$

Es el predictor lineal de efectos fijos. Mientras que z_i es la i-ésima observación de un campo gaussiano (GF) aleatorio latente, cuyo objetivo es captar el efecto de los factores no incluidos explícitamente en el modelo, pero que impactan sobre el valor de la respuesta observada a través de variaciones espacialmente estructuradas (Rozanov, 1982). La función de distribución de z está dada por:

$$\zeta | \psi \sim MN(0, Q^{-1}(\psi))$$

don $Q(\psi)$ se modela a partir de la función de covarianza espacial de Matérn definida como:

$$C(\Delta_{ij}) = \frac{1}{\Gamma(a)} 2^{a-1} (k \Delta_{ij})^a K_a(k \Delta_{ij}) \quad (2)$$

la cual depende de la distancia euclíadiana medida entre dos observaciones cualesquiera:

$$\Delta_{ij} = \|S_i - S_j\|$$

el parámetro K_a denota la función de Bessel modificada de segundo orden (Rue & Held, 2005). El parámetro $a = r-d/2$, mide el grado de suavizamiento del proceso, su varianza marginal está dada por:

$$\sigma^2 = \frac{\Gamma(a)}{\Gamma(r)(4\pi)^{d/2} k^{2a} \tau^2}$$

$y, k > 0$ es un parámetro de escalamiento relacionado con la distancia a partir de la cual la correlación espacial tiende a anularse (Rue *et al.*, 2009).

La distribución de las observaciones está dada por la verosimilitud en términos de un vector de parámetros $q = \{ b_0, b_i, x \}$, y un vector de hiper-parámetros Ψ , de manera que:

Adicionalmente el modelo incluye un efecto aleatorio espacial no estructurado, u_i , que captura los efectos de las características no medidas que afectan el nivel medio de marginación de las localidades (Rue & Lindgren, 2015). Cada u_i presenta una distribución gaussiana independiente, con media cero y precisión τ_u .

4. Implementación del modelo

La modelación de un campo aleatorio gaussiano continuo (GF) se complejiza al incrementarse el número de observaciones, para lo cual se emplea el enfoque de ecuaciones diferenciales parciales estocásticas (SPDE) en el que un campo gaussiano continuo (GF) se modela a través de un campo aleatorio gaussiano markoviano discreto (GMRF) mediante la definición de una triangulación del dominio espacial (Lindgren & Rue. 2011). De forma que el GMRF capture las dependencias espaciales sobre una malla triangular que permita identificar una estructura de vecindades.

La triangulación del dominio espacial de los datos, a partir del enfoque SPDE (Rue & Lindgren, 2015) utiliza combinaciones lineales de funciones tipo *spline* sobre las ubicaciones de los vértices de la triangulación para representar el campo GMRF a través de la ecuación:

$$\zeta(s) = \sum_{g=1}^G \phi_g(s) \xi_g \quad (4)$$

Donde G es el número total de vértices en la triangulación, $\{j_g\}$ es un conjunto de funciones lineales básicas y $\{\mathbf{x}_g\}$ representa ponderadores espaciales gaussianos con media cero. La función base $\{j_g\}$ es igual a 1 en el vértice g y 0 en los demás vértices.

Una vez que el GF se escribe en términos de la ecuación [4] se obtiene un mapeo de la función de covarianza [2] del campo gaussiano a la matriz de precisión Q , de modo que el modelo espacial puede reescribirse en términos de un GMRF (Rue *et al.*, 2009).

La matriz de precisión Q se define de forma que para cualquier localidad i distinta de j se tiene que:

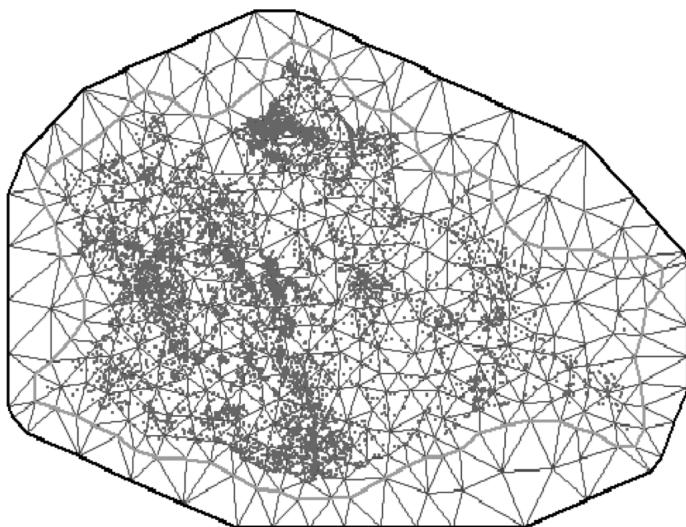
$$i \notin \text{vecindad}\{j\} \Leftrightarrow Q_{ij} = 0$$

Por lo que la dependencia condicional basada en esta estructura de vecindades genera una matriz Q dispersa, con importantes ventajas computacionales (Blangiardo & Cameletti, 2015; Lindgren & Rue, 2011).

Es importante señalar que para obtener una buena definición del modelo, durante el proceso de triangulación se deberán definir triángulos de tamaño y forma regular, además se deberá extender la triangulación más allá de los límites del dominio espacial con la finalidad de evitar el llamado efecto límite cuya consecuencia es una inflación de la varianza en los puntos cercanos a la frontera (Lindgren & Rue, 2011).

La triangulación del dominio espacial de las localidades de Oaxaca puede verse en el mapa 2. La misma comprende un total de $G = 3\,728$ vértices.

Mapa 2
Triangulación de los niveles de marginación para las localidades,
Oaxaca 2010



Fuente: elaboración propia con datos del Censo de Población y Vivienda 2010, INEGI.

5. Hiperparámetros

El último elemento para la definición del modelo consiste en fijar los valores iniciales del vector de parámetros $\mathbf{q} = \{ \mathbf{b}_0, \mathbf{b}_i, \mathbf{x} \}$, y del vector de

hiperparámetros $\Psi = \{\sigma_e^2, k, \sigma_{rf}^2\}$, donde k y σ_{rf}^2 son parámetros de la función de covarianza de Matérn definidos en la ecuación [2] (Fuglstad *et al.*, 2015). La representación interna en R-INLA para la aproximación SPDE de los parámetros establece las especificaciones previas de las distribuciones en términos de:

$$\sigma^2 \approx \frac{1}{4\pi k^2 \tau^2}$$

$$r \approx \frac{\sqrt{8}}{k}$$

los valores iniciales asignados fueron: $k = 0.75$ y $\alpha = 2$. Para el vector de parámetros, mientras que los valores asignados a la distribución inicial de del vector de parámetros fueron: $\beta_0 \sim N(0, \tau^{-1})$, y para los componentes de los efectos fijos $\beta_i \sim N(0, 1/0.001)$, para $i = 1, 2, 3$.

al vector de pesos espaciales $\xi = (\xi_1, \dots, \xi_n)'$ se le asignó una distribución gaussiana $\xi \sim N(0, Q^{-1})$ donde Q es una matriz de precisión dispersa que depende de la función de covarianza Matérn dada por la ecuación [2].

5. Ajuste y selección de modelos

Uno de los principales objetivos del modelo es la predicción del nivel medio de marginación para las localidades donde no se dispone de información. A partir de las relaciones teóricas establecidas en (Bistrain, 2010; Camberos, 2007; Gutiérrez y Gama, 2010) entre el concepto de marginación y las covariables x_1 (porcentaje de servicios disponibles en la localidad), x_2 (escolaridad promedio de la localidad), x_3 (porcentaje de población desocupada por localidad) se construyeron ocho modelos surgidos de la totalidad de combinaciones de covariables en términos de los efectos fijos para la ecuación [1].

Los modelos comprenden: un modelo sin covariables, tres modelos con una covariable, tres modelos con dos covariables y un modelo con tres covariables.

Los modelos se numeraron del M0 al M7 de acuerdo con las siguientes definiciones: el primer modelo M0 se desarrolló sin ninguna variable de efectos fijos, lo que equivale a modelar la respuesta en función de sólo el efecto espacial; el modelo M1 se conformó con una variable servicios como

el efecto fijo, más la parte de efecto espacial; el modelo M2 se construyó con la variable educación más el efecto espacial; el modelo M3 consideró la variable de ocupación y el efecto espacial; el modelo M4 se construyó con las variables servicios y educación como efectos fijos más el efecto espacial; el modelo M5 al igual que el anterior, empleó dos variables, servicios y ocupación, más la parte de efecto espacial; el modelo M6 se construyó añadiendo al componente espacial, la parte de efectos fijo con las variables educación y desocupación; finalmente el modelo M7 incorporó las tres covariables: educación, servicios y desocupación más la parte de efecto espacial.

Los modelos propuestos fueron ajustados utilizando el método “Integrated Nested Laplace approximation” (INLA) por sus siglas en inglés, implementado en el paquete R-INLA desarrollado para r-project en su versión 3.15. La utilización de R-INLA permitió realizar una inferencia bayesiana rápida y eficiente de los modelos gaussianos latentes propuestos (Martins *et al.*, 2013).

El enfoque INLA resulta ser una alternativa computacionalmente menos costosa que el tradicional método MCMC (Gamerman & Lopes, 2006). El uso de INLA permitió aproximar los marginales posteriores del campo gaussiano latente en tres pasos: (1) ajusta la distribución posterior de θ utilizando la aproximación de Laplace para la ecuación [3]. (2) utiliza una versión simplificada de Laplace para estimar la distribución del GMRF a partir de los valores estimados de θ , lo que mejora la aproximación gaussiana dada por [4] y (3) estima los parámetros restantes a partir de la ecuación:

$$P(\theta, \psi | Y) \propto P(\psi) x |Q(\psi)|^{\frac{1}{2}} \cdot \exp\left(-\frac{\theta^T}{2} Q(\psi) \theta + \sum_{i=1}^n \log(P(y_i | \theta_i))\right) \quad (5)$$

que combina las ecuaciones [3] y [4] mediante el uso de métodos de integración numérica (Cressie, 1993). El modelo supone independencia condicional entre el campo gaussiano latente y el vector de parámetros θ .

Una vez estimados los modelos M0 a M7, la forma de identificar el modelo que presenta la combinación de covariables más adecuada para explicar el comportamiento de la marginación, se realiza través del criterio de información de Akaike (AIC). El modelo seleccionado será el que alcance el AIC más bajo en términos de:

$$AIC = -2l(\hat{\theta}) + 2k$$

donde $l(\hat{\theta})$ es la función de máxima de log-verosimilitud y k es el número de parámetros (Blangiardo & Cameletti, 2015).

6. Marginación en localidades de Oaxaca

Las mediciones de marginación por localidad permitieron generar mapas que muestran los niveles de marginación previos y modelar su incertidumbre a través de las distribuciones estimadas de forma posterior para cada una de las localidades de Oaxaca (mapa 1 y 2).

Específicamente, las estimaciones obtenidas para la parte de efectos fijos de los modelos M0 a M7 pueden observarse en la tabla 1. Adicionalmente la tabla 1 presenta el AIC alcanzado para cada modelo, lo que permite identificar el modelo que, bajo este criterio, ajusta de mejor manera los niveles de marginación municipales. En este caso resultó ser el modelo 7, no obstante, se optó por elegir el modelo 4 debido a tres razones, (1) la variable porcentaje de población desocupada por localidad ejerce un efecto muy pequeño sobre el nivel de marginación, (2) la diferencia de AIC entre los modelos 4 y 7 es baja y (3) el modelo 4 es más parsimonioso.

El modelo M0, presentado en la tabla 1, tiene como finalidad modelar el comportamiento espacial del campo aleatorio (GF) mismo que se muestra en el panel superior izquierdo del mapa 3. Recuerde que la media posterior del predictor lineal presenta una distribución normal multivariada a partir de la cual es posible estimar la respuesta media que genera el gráfico espacial en escala natural (Krainski *et al.*, 2016; Nychka *et al.*, 2013).

Tabla 1
Modelos espaciales bayesianos M0 a M7

Modelo	B0	x1	x2	x3	AIC
M0	0.3621				-11704.49
M1	0.3952	-0.1938			-12851.22
M2	0.5954		-0.8612		-14724.84
M3	0.3672			-0.3449	-11854.97
M4	0.6219	-0.1834	-0.8427		-16223.72
M5	0.5948		-0.8485	-0.1920	-14790.33
M6	0.3949	-0.1916		-0.0310	-12844.01
M7	0.6238	-0.1930	-0.8507	0.1250	-16250.09

Fuente: elaboración propia.

La parte de efectos fijos del modelo M4, tabla 2, busca evaluar el efecto que las covariables (x_1) el porcentaje de servicios disponibles en la localidad y (x_2) la escolaridad promedio ejercen sobre los niveles de marginación locales. La estimación de los efectos fijos del modelo M4, pueden observarse con mayor detalle en la tabla 2. Los resultados indican que un incremento relativo de un punto porcentual en el acceso a servicios disponibles en la localidad implicaría una reducción del 18.3 puntos porcentuales de los niveles de marginación locales, mientras que el incremento de un año en los niveles de escolaridad promedio de la localidad provocarían una reducción de 84.2 puntos porcentuales en los niveles de marginación locales, una vez considerado el efecto espacial.

Tabla 2
Modelo espacial bayesiano M4

M4.	mean	sd	0.025quant	mode	0.975quant
b0	0.6219	0.0076	0.6071	0.6219	0.6369
x_1	-0.1833	0.0045	-0.1921	-0.1833	-0.1745
x_2	-0.8428	0.0129	-0.8682	-0.8428	-0.8175

Fuente: elaboración propia

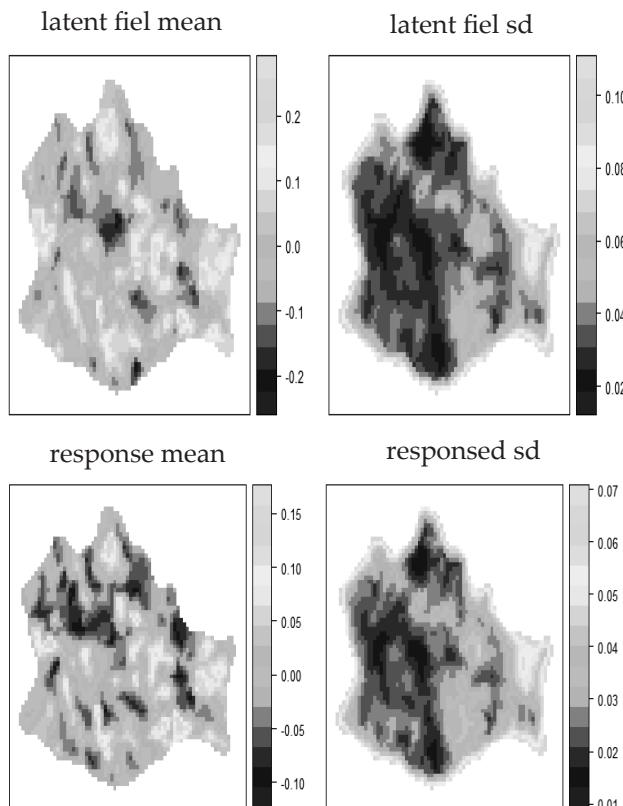
La combinación de los efectos fijos (de las covariables) y las variaciones generadas por los efectos aleatorios a partir de la salida del campo aleatorio discreto, generan la media posterior del predictor lineal que es la estimación del nivel de marginación esperado en cada localidad.

La distribución espacial de la marginación de las localidades de Oaxaca modelada como un campo gaussiano, se presenta en el mapa 3 panel superior izquierdo. En la misma se muestra un comportamiento homogéneo, donde en prácticamente la totalidad del territorio se presentan niveles de marginación altos (verde) con algunas manchas (amarillas) que dibujan zonas de muy alta marginación y algunas zonas (azules) que indican la presencia de clusters de localidades con bajos niveles de marginación. Es importante señalar que los clusters amarillos de mayor tamaño (que identifican zonas de muy alta marginación en Oaxaca) se corresponden con áreas prácticamente despobladas que coincide en su mayor parte con la zona de cerro azul y con la reserva ecológica de los Chimalapas, en contra parte

las zonas azules (de menor marginación) se corresponden a regiones mayormente pobladas ubicadas en la zona de valles centrales, de la Mixteca, zona de la cañada y algunas ciudades costeras como Huatulco y Puerto Escondido.

Mapa 3

presenta mapa de la distribución espacial de los niveles de marginación por localidad, Oaxaca 2010.



Fuente: elaboración propia

El panel superior derecho del mapa 3 muestra la varianza de la respuesta del modelo M0. En el mismo, es posible observar que la varianza tiene una magnitud media de 0.052, similar a la varianza del GMRF para la mayor

parte del dominio, lo que indica que la incertidumbre que da origen a los datos tiene una escala equivalente. Por otra parte, el panel inferior derecho del mapa 3 muestra la varianza del modelo M4, cuya magnitud media es 0.04, variación que puede atribuirse a la magnitud de las covariables incorporadas al modelo.

Es importante señalar que a pesar del cambio en la magnitud de la varianza, el modelo M4 parece ajustar de mejor manera los datos sobre marginación por localidad que el modelo M0, debido a la notable reducción del AIC, lo que implica también que las covariables mejoran el nivel de explicación respecto del ajuste producido por el campo gaussiano markoviano.

El mapa 3 en el panel inferior izquierdo, muestra la distribución espacial de la marginación considerando tanto los efectos fijos como espaciales. Es importante señalar que en el mapa se observa un patrón de localidades con baja marginación que dibujan una serie de caminos (color azul) que coinciden con el trazo de las principales carreteras, una de Arriaga con destino a Puebla y otra de Puerto Escondido dirección valles centrales de Oaxaca.

6. Discusión

El uso de modelos espaciales bayesianos proporciona un enfoque robusto y flexible especialmente útil para analizar el comportamiento de fenómenos socioeconómicos considerando el efecto de la estructura de dependencia espacial y proporcionando la posibilidad de realizar inferencias sobre el comportamiento de la marginación, no sólo en términos de su nivel sino también de su distribución espacial.

Una de las principales ventajas de implementar las SPDE es que permiten pasar de un GF a un GMRF con lo cual se obtienen importantes propiedades, a partir del uso de métodos numéricos computacionalmente más eficientes que hacen viable la implementación del enfoque INLA, a través del paquete en *R*.

La implementación del enfoque INLA permitió la estimación eficiente y por tanto la comparación de ocho modelos a través del AIC, lo que proporcionó evidencia estadística para establecer la existencia de una relación negativa entre los niveles de marginación y las covariables: porcentaje de servicios disponibles en la localidad y escolaridad promedio, donde altos niveles de marginación son indicativos de bajos niveles de escolaridad y escasa presencia de servicios públicos. Adicionalmente se encontró la

presencia de algunos clusters de localidades con muy altos niveles de marginación los cuales presumiblemente corresponden a zonas de baja o muy baja densidad poblacional, mientras que por el contrario, se detectó la presencia de algunos clusters de baja marginación cuyos niveles pueden explicarse a partir de su ubicación en importantes zonas urbanas de Oaxaca o por estar ubicadas sobre alguna ruta carretera importante.

Los mapas de distribución de la marginación basados en la modelación espacial bayesiana representan un primer intento por analizar la variación geográfica que refleja la marginación por localidad en Oaxaca. Los mapas sugieren una distribución homogénea de la marginación, que refleja la presencia generalizada de niveles altos de marginación a lo largo de todo el territorio de Oaxaca, con las excepciones señaladas.

Si bien, la inferencia bayesiana se ha implementado comúnmente a través de métodos MCMC, estos presentan problemas en términos de costo computacional y eventualmente de convergencia. Estos problemas se subsanaron a través del uso de INLA que resultó ser un método cuyas aproximaciones rápidas y precisas resultaron en una alternativa efectiva al tradicional MCMC.

7. Conclusiones

La estimación de distintos modelos gaussianos latentes permitió estimar la variación espacial de la distribución de la marginación en las localidades de Oaxaca, a través de un campo aleatorio gaussiano markoviano discreto representado por el modelo M0, a la par se ajustaron siete diferentes modelos M1 a M7, que además de una parte de efectos aleatorios incluyeron una parte de efectos fijos con hasta tres covariables.

Los niveles de marginación presentaron un comportamiento espacialmente homogéneo a lo largo de todo el territorio oaxaqueño, con excepción de algunos clusters, lo cual resulta consistente con los datos observados. Asimismo, se encontró evidencia de la existencia de una relación causal indirecta entre los niveles de marginación observados en las localidades y las covariables porcentaje de servicios y niveles de escolaridad promedio.

El análisis de los efectos espaciales del modelo M4 permite tener una idea de las condiciones del comportamiento de la marginación, una vez controladas las covariables, de forma que los resultados pueden constituirse en

una herramienta útil para la formulación políticas sociales destinadas a combatir los altos niveles de marginación al identificar áreas de alta marginación en zonas relativamente pobladas o susceptibles de requerir intervención.

Finalmente, se debe señalar que la intención del estudio no es generar conclusiones definitivas sobre el comportamiento de la marginación en Oaxaca, sino más bien evidenciar la capacidad de los métodos espaciales bayesianos para modelar datos a nivel de localidad, además de poner de relieve la importancia del uso de covariables en este tipo de modelos.

Referencias

- Bistrain Coronado, César (2010). "Revisión de los índices de marginación elaborados por el Conapo", en Estudios Demográficos y Urbanos, vol. 25, núm. 1 (73).
- Bustos, Víctor A. (2009). "Indicadores sintéticos para seguir la evolución en el tiempo de fenómenos multidimensionales: una propuesta metodológica", en Boletín del Sistema Nacional de Información Estadística y Geografía, vol. 2, núm. 3.
- Blangiardo M. and M. Cameletti (2015). Spatial and Spatio-temporal Bayesian Models with R-INLA. Wiley, 1st edition.
- Camberos, Mario y Joaquín Bracamontes (2007). "Marginación y políticas de desarrollo social: un análisis regional para sonora", en Problemas del Desarrollo, vol. 38, núm. 149.
- Consejo Nacional de Población (2012). "Índice de Marginación por Localidad 2010". México, DF.

- Cortés F. (2002). "Consideraciones sobre la marginalidad, marginación, pobreza y desigualdad en la distribución del ingreso". *Papeles de población*, vol. 8 núm. 31.
- Cortés F. (2006). "Consideraciones sobre la marginación, la marginalidad, marginalidad económica y exclusión social", *Papeles de Población*, vol. 12. núm. 47.
- Cressie N. (1993). *Statistics for spatial data*. J. Wiley & Sons.
- DE SAL (1969). "La marginalidad en América Latina: un ensayo de diagnóstico", Herder, Barcelona.
- Gamerman D. and H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC.
- Geir-Arne Fuglstad, Daniel Simpson, Finn Lindgren, and Håvard Rue (2015). Interpretable priors for hyperparameters for gaussian random fields.
- Gutiérrez Pulido y Gama Hernández (2010). "Limitantes de los índices de marginación de Conapo y propuesta para evaluar la marginación municipal en México", *Papeles de población*, núm. 66, pp 227-257.
- Ingebrigtsen E., F. Lindgren, and I. Steinsland (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, vol 8, pp. 20-38.
- Krainski E.; F. Lindgren, D. Simpson and H. Rue (2016). The r-inla tutorial on spde models. <http://www.math.ntnu.no/inla/r-inla.org/tutorials/spde/spde-tutorial.pdf>.
- Lindgren Finn and Håvard Rue (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.
- Martins TG.; D. Simpson; F. Lindgren y H. Rue (2013). "Bayesian Computing with INLA: New Features." *Computational Statistics and Data Analysis*, núm. 67, 68-83.
- Nychka D; D. Hammerling; S. Sain y N. Lenssen (2013). LatticeKrig: Multiresolution Kriging based on Markov Random Fields. URL <http://cran.r-project.org/web/packages/LatticeKrig/>.
- Rozanov A. (1982). *Markov Random Fields*. Springer-Verlag, New York.
- Rue H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- Rue H.; S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B.*, 71, 319-392.
- Rue H. and F. Lindgren (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, vol., v, Issue II.