



Modelado espacial bayesiano de niveles de retorno de precipitaciones diarias extremas en Honduras

Bayesian spatial modeling of extreme daily rainfall return levels in Honduras

S. MÉNDEZ¹, C. CRUZ²

Recibido: 9 de septiembre de 2022 / Aceptado: 25 de abril de 2023

RESUMEN

Modelar valores extremos en precipitación es muy importante, una forma de cuantificarlos es mediante los niveles de retornos. El modelo completo utilizado para la estimación se desarrolla en una primera etapa asociada a los parámetros de la Distribución Pareto Generalizada (GPD) y una segunda etapa asociada a las tasas de excedencia. Los modelos son abordados mediante modelos jerárquicos. Se toma en cuenta la componente espacial del fenómeno usando covariables. El mejor modelo es escogido mediante en criterio de la log-verosimilitud marginal, las estimaciones de los modelos escogidos se usan de manera conjunta para poder construir con las mismas mapas de retorno. Este estudio está siendo desarrollado para Honduras tomando valores de precipitación diaria de 1972 hasta 2012 en 59 estaciones meteorológicas, evidenciando que la elevación es una covariable que influye en la estimación de los parámetros de la GPD y las tasas de excedencia tienen un comportamiento constante para todas las estaciones.

ABSTRACT

Modeling extreme values in precipitation is very important, and one way to quantify them is by means of return levels. The complete model used for the estimation is developed in a first stage associated with the parameters of the Generalized Pareto Distribution (GPD) and a second stage associated with the exceedance rates. The models are approached by means of hierarchical models. The spatial component of the phenomenon is taken into account using covariates. The best model is chosen using the marginal log-likelihood criterion, and the estimates of the chosen models are used jointly in order to construct return maps with the same estimates. This study is being developed for Honduras taking daily precipitation values from 1972 to 2012 in 59 meteorological stations, showing that elevation is a covariate that influences the estimation of the GPD parameters and the exceedance rates have a constant behavior for all stations.

PALABRAS CLAVES

Niveles de Retorno, Modelos Jerárquicos, Teoría de Valores Extremos, Inferencia Bayesiana.

KEYWORDS

Return Levels, Hierarchical Models, Extreme Value Theory, Bayesian Inference.

¹Departamento de Estadística, Escuela de Matemáticas y Ciencias de la Computación. Universidad Nacional Autónoma de Honduras. sonia.mendez@unah.edu.hn

²Maestría de Matemáticas, Escuela de Matemáticas y Ciencias de la Computación. Universidad Nacional Autónoma de Honduras. cristian.cruz@unah.edu.hn

I | INTRODUCCIÓN

La precipitación es uno de los fenómenos meteorológicos donde la ocurrencia de valores extremos es cada vez más frecuente, esto como consecuencia del cambio climático (IPCC, 2022). En Honduras, se han presentado fenómenos como ser los huracanes Fifi (1974), Mitch (1998), Eta e Iota (2020), estos últimos catalogados como el fenómeno natural más severo en los últimos 20 años. Todos afectaron fuertemente al país y dada la naturaleza de los mismos se consideran poco usuales. Al observar la ocurrencia de este tipo de fenómenos nos surgen varias interrogantes, como por ejemplo: ¿En qué momento se puede presentar nuevamente un fenómeno de esta magnitud?, ¿Existe un periodo de tiempo específico donde estos fenómenos se repiten?. Poder dar respuesta a estas preguntas es muy importante.

Como una de las posibles respuestas a las interrogantes planteadas, surge la necesidad de estimar periodos de retorno, es decir el número de años que en promedio un evento considerado como extremo se igualará o superará, y además poder determinar la frecuencia con la que se espera que ocurra. Si bien es cierto, no es posible detener la ocurrencia de fenómenos extremos, conocer su impacto brinda la oportunidad de crear planes para mitigar las consecuencias de las ocurrencias de los mismos y con esta información contribuir con los planes de gestión de riesgos.

El estudio de estos eventos extremos y hablar de la probabilidad de ocurrencia de los mismos se vuelve difícil, por la cantidad de datos que se poseen, además de que su comportamiento no es el mismo en todos los puntos, geográficamente hablando, ya que la precipitación como variable de interés tiene implícitos muchos factores entre ellos: latitud, altitud, la estación del año y el relieve.

La presencia de valores extremos en nuestros datos de estudio implica que realicemos un abordaje diferente. Este análisis se realiza mediante la teoría de valores extremos, para estudiarla puede consultar (Coles, 2001), (Gnedenko y Kolmogorov, 1954), (Mochales, 2013) y (Embrechts, Klüppelberg, y T., 1994). Esta teoría es muy importante por su aplicación en áreas como ingeniería, meteorología, economía, entre otras. El objetivo de este trabajo es poder estimar niveles de retorno para diferentes periodos de tiempo usando estos valores para generar mapas de retorno para precipitación en Honduras.

II | PRELIMINARES

Generalmente cuando estudiamos valores extremos buscamos relacionarlos con una aplicación, por ende queremos darle a estos valores una interpretación. Con la finalidad de relacionar la teoría de valores extremos con sus aplicaciones, es conveniente hablar de cuantiles o de manera análoga de niveles de retorno, más que de las estimaciones de parámetros como tal.

En el caso específico de los fenómenos extremos, es importante disponer de un valor que permita cuantificar la probabilidad de ocurrencia de los mismos, para un periodo de tiempo determinado. La teoría de valores extremos proporciona un enlace entre los datos tomados por hora o día y periodos de tiempo más largos, como lo son los niveles de retorno que pueden ser tomados por periodos de 5, 10, 15 o más años. Escogimos determinar los niveles de retorno de forma cerrada, pero como una función no lineal de los parámetros de la GPD (ξ, σ) y la probabilidad de excedencia (ζ_u).

Existen 2 tipos de modelos para el abordaje de valores extremos:

1. **Máximo por bloques:** En lugar de analizar todos los datos, se toman bloques de datos de igual longitud. Para cada bloque se calcula el máximo de acuerdo a lo que se desea analizar generando una serie de máximos, convirtiéndose estos en el objeto de interés.
2. **Método POT:** Método de picos sobre el umbral (Peaks Over Threshold) en este modelo se trabaja con valores de excedencias, se escoge un umbral u adecuado y se calculan las excedencias, el objeto de interés son todos aquellos valores que superan el mismo. El método de valores sobre un umbral fue propuesto inicialmente en el área de Hidrología y luego formalizado por R.L. Smith en base a resultados obtenidos previamente por Pickands.

La distribución de Pareto generalizada se utiliza como aproximación, para la distribución límite de las excedencias de umbrales.

Utilizamos en nuestro abordaje el Método POT, haciendo uso de gráficos de vida media residual para la elección del umbral. Una vez determinado el umbral lo que tenemos para modelar son excedencias (colas).

Dentro de la teoría formal de valores extremos se puede ubicar el primer teorema fundamental con Fisher-Tippet (Fisher y Tippett, 1928) y Gnedenko (Gnedenko, 1943), que caracteriza la distribución asintótica del máximo. El segundo teorema fundamental con Pickands (Pickands III, 1975) y Haan (Balkema y Haan, 1974) en el que se describe el comportamiento de las colas de una distribución asociándolo a la distribución de valores generalizados.

Segundo teorema fundamental de la teoría de valores extremos

Si se toma como objeto de estudio todos aquellos valores que exceden un umbral fijo (excedencias), estos valores quedan caracterizados por la Distribución de Pareto Generalizada (GPD).

Función de distribución de excesos: Sea X una v.a con distribución F y extremo derecho X_F . Para $u < X_F$ fijo decimos que ha ocurrido una excedencia de u si $X > u$.

La distribución de X_u esta dada por:

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad x \geq 0. \quad (1)$$

A la función $e(u) = E((X - u \leq x | X > u))$ se le conoce como la función media de exceso, es posible encontrar un valor apropiado del umbral haciendo un gráfico de la función media de excesos empírica.

Distribución de Pareto: Introducida por Pickands en 1975 es la familia bi-parámetrica de funciones de distribución dada por:

$$G(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(\frac{-x}{\sigma}\right), & \xi = 0 \end{cases} \quad (2)$$

La GPD queda descrita por:

- σ : Parámetro de escala
- ξ : Parámetro de forma

Con $\xi < 0$, para $0 < x < -\frac{\sigma}{\xi}$ y $x > 0$, para $\xi > 0$.

Los niveles de retorno son cantidades climatológicas asociados con precipitación, que se comportan de manera diferente para diferentes puntos geográficos, es decir hay una componente espacial implícita que debe ser tomada en cuenta. De las reglas de probabilidad básica, siguiendo la línea de trabajo de (Cooley, Nychka, y Naveau, 2007) para el abordaje de los niveles de retorno, tenemos que:

$$P(Z > z + u) = \zeta_u \left(1 + \xi \frac{z}{\sigma_u} \right)^{-\frac{1}{\xi}}, \text{ con } \zeta_u = P(Z > u). \quad (3)$$

Es conveniente trabajar los niveles de retorno en años, sea n_y el número de observaciones tomadas en un año, se obtiene el nivel de retorno a r años resolviendo la ecuación $P(Z > z_r) = \frac{1}{r n_y}$ para z_r . De manera que

$$z_r = u + \frac{\sigma_u}{\xi} \left[(r n_y \zeta_u)^\xi - 1 \right]. \quad (4)$$

Para lograr estimar los niveles de retorno es necesario obtener las estimaciones de los parámetros y sustituirlas en la ecuación 4. En resumen los niveles de retorno Z_r son valores de precipitación que en promedio serán alcanzados o excedidos con probabilidad $1/Z_r$ al menos una vez en r años.

I Precipitación en Honduras

Honduras es un país muy montañoso, debido a esto la distribución de los valores de precipitación es muy variable. Honduras es además influenciado por el fenómeno de El Niño y La Niña, el cual tiene un impacto muy importante en la distribución de las lluvias tanto espacial como temporal, presentando un fuerte déficit de lluvia en las regiones sur, suroccidente, oriente y central en su etapa cálida. Este comportamiento cambia con condiciones frías (La Niña) o neutras, donde la precipitación puede reflejarse entre el promedio normal o arriba de lo normal en la mayor parte del país así se enuncia en (Capel, 1994). Puede encontrar más detalles respecto a la climatología y gestión de riesgo en Honduras en (Navarro, Llano, Obando, y D. Córdoba, 2018), (Rodríguez y cols., 2012).

III | DESCRIPCIÓN DE LA METODOLOGÍA

Basándonos en el modelo usado en (Cooley y cols., 2007), artículo usado como base para el desarrollo de este trabajo. El modelo completo se aborda en 2 etapas:

- **Etapla 1:** asociada a los parámetros de la Distribución Pareto generalizada (GPD).
- **Etapla 2:** asociada a las tasas de excedencia.

A cada etapa le asociamos un modelo jerárquico, estos modelos se describen en (Banerjee, Carlin, y Gelfand, 2003). Nuestros modelos jerárquicos están compuesto de 3 niveles, los modelos asociados a cada etapa son independientes entre si, debido a la naturaleza de los parámetros.

- **Nivel 1:** Nivel de los datos.
- **Nivel 2:** Nivel del proceso latente que controla la precipitación climatológica extrema.
- **Nivel 3:** Nivel de las distribuciones a priori de los parámetros involucrados en el proceso latente.

La inferencia sobre los parámetros θ , dados los datos de las estaciones meteorológicas $\mathbf{Z}(\mathbf{x})$, se realiza haciendo uso de herramientas básicas de inferencia descritas en (Infante, A., y Urquiola, 2010), (Liu y Wasserman, 2014), (Casella y Berger, 2002) y (Gamerman, Migon, y Louzada, 2015). A partir del Teorema de Bayes tenemos que:

$$p(\theta|\mathbf{Z}(\mathbf{x})) \propto p(\mathbf{Z}(\mathbf{x})|\theta)p(\theta). \quad (5)$$

La distribución condicional de nuestros modelos jerárquicos queda descrita por:

$$p(\theta|\mathbf{Z}(\mathbf{x})) \propto p_1(\mathbf{Z}(\mathbf{x})|\theta_1)p_2(\theta_1|\theta_2)p_3(\theta_2), \quad (6)$$

donde p_j es la densidad asociada con el nivel j , en nuestro caso $j = 1, 2, 3$.

La primera etapa del modelo completo, consta de 3 niveles que se describen a continuación:

Nivel de los datos

Nivel base del modelo jerárquico asociado a la excedencia de umbrales, descrito por la GPD definida en la ecuación 2. Dadas las ubicaciones de las s estaciones meteorológicas x_1, \dots, x_s , denotamos los parámetros de la GPD, asociados a cada estación por:

$$\sigma_u = [\sigma_u(x_1), \dots, \sigma_u(x_n)]^T, \quad (7)$$

$$\xi = [\xi(x_1), \dots, \xi(x_n)]^T. \quad (8)$$

Reparametrizamos la GPD haciendo $\phi = \log(\sigma_u)$, de esta manera permitimos que el parámetro ϕ tome valores positivos y negativos.

Sea $Z_k(\mathbf{x}_i)$ la k -ésima cantidad de precipitación diaria que excede el umbral u y fue registrada en la estación meteorológica \mathbf{x}_i con $i = 1, \dots, s$. Dado que $Z_k(\mathbf{x}_i)$ es una excedencia, asumimos que puede ser descrita mediante una GPD cuyos parámetros dependen de la ubicación de la estación de manera que:

$$p\{Z_k(\mathbf{x}_i) - u > z | Z_k(\mathbf{x}_i) > u\} = \left(1 + \frac{\xi(\mathbf{x}_i)z}{\exp\phi(\mathbf{x}_i)}\right)^{-1/\xi(\mathbf{x}_i)}, \quad (9)$$

derivamos la función de distribución asociada a la ecuación 9, para obtener la función de densidad:

$$p_1(Z_k(\mathbf{x})|\theta_1) = \prod_{i=1}^s \prod_{k=1}^{n_i} \frac{1}{\exp\phi(\mathbf{x}_i)} \left(1 + \frac{\xi(\mathbf{x}_i)z}{\exp\phi(\mathbf{x}_i)}\right)^{-1/\xi(\mathbf{x}_i)-1}, \quad (10)$$

donde $\theta_1 = [\phi, \xi]^T$.

La ecuación 10 nos da la primera parte de la ecuación 6 asociada al nivel 1 de la etapa 1.

I Nivel del proceso espacial latente

En este nivel buscamos describir la relación existente entre el nivel de los datos, con la orografía y climatología de las estaciones meteorológicas que son las que caracterizan el proceso espacial latente. Esperaríamos que de acuerdo a la ubicación de las estaciones meteorológicas, se presenten variaciones en los valores extremos de precipitación diaria, debido a factores como ser el clima, la altitud, la temperatura, la ubicación geográfica como tal, entre otros. Las variaciones que logremos identificar, no pueden ser explicadas completamente por las covariables y sus funciones. Con el fin de lograr capturar de forma eficiente ese efecto utilizamos métodos de la estadística espacial. Para conocer un poco acerca de Estadística Espacial y sus herramientas puede consultar (Giraldo, 2002), (Sierra, 2014), y (Armstrong, 1998).

Sea $\phi(x)$ el parámetro log-transformado de escala y $\xi(x)$ el parámetro de forma del proceso. Estos parámetros son tratados como variables aleatorias, cada uno tiene una distribución a priori asociada. Para describir el comportamiento de $\phi(x)$, tomando en cuenta su relación con el proceso espacial latente, hacemos uso de métodos estándar en geoestadística. Modelamos $\phi(x)$ mediante un proceso Gaussiano con:

$$\text{media: } E[\phi(x)] = \mu_{\phi}(x), \quad (11)$$

$$\text{covarianza: } \text{cov}(\phi(x), \phi(x')) = k_{\phi}(x, x'). \quad (12)$$

La media es función de los parámetros α_{ϕ} y las covariables:

$$\mu_{\phi}(x) = f_{\phi}(\alpha_{\phi}, \text{covariables}). \quad (13)$$

La función f podemos modificarla para lograr incorporar relaciones entre las diferentes covariables involucradas en el modelo.

La covarianza es función de la distancia entre las estaciones meteorológicas y los parámetros β_{ϕ} , está dada por:

$$k_{\phi}(x, x') = \beta_{\phi,0} \times \exp(-\beta_{\phi,1} \times \|x - x'\|), \quad (14)$$

que corresponde al modelo teórico de un variograma exponencial.

Tenemos que:

- $\beta_{\phi,0}$: meseta.
- $\beta_{\phi,1}$: rango.

Analizamos el variograma en distintas direcciones confirmando que el comportamiento de los parámetros espaciales de interés (meseta y rango) es similar en las distintas direcciones, esto como consecuencia de la distancia entre las estaciones escogidas y por el número de estaciones consideradas en nuestro estudio. De manera que similar a lo enunciado en (Cooley y cols., 2007), nuestro modelo de covarianza asume que el proceso es isótropo (comportamiento similar, independiente de la dirección) y estacionario; no detectamos no estacionariedad o anisotropía con sólo 59 estaciones. Elegimos trabajar con modelos exponenciales debido a su simplicidad, porque no asumen ningún nivel de suavidad en el proceso latente y se ajusta a las particularidades identificadas en nuestros datos. Más detalles pueden ser consultados en (Armstrong, 1998).

Con respecto a $\xi(x)$ de manera similar a lo que se describe en (Cooley y cols., 2007), no sabemos como se comporta este parámetro ni su sensibilidad a las variables regionalizadas. De manera que el parámetro de forma es modelado como un valor único para todas las estaciones, con distribución a priori $U(-\infty, \infty)$.

Tomando en cuenta la información que la ubicación de las estaciones nos proporciona acerca de los datos y del proceso espacial latente, tenemos:

$$p_2(\theta_1|\theta_2) = \frac{1}{\sqrt{(2\pi)^s |\Sigma|}} \exp \left[-\frac{1}{2} (\phi - \mu)^T \Sigma^{-1} (\phi - \mu) \right] \times p_\xi(\xi|\theta_\xi) \quad (15)$$

donde μ es un vector definido por la ecuación 13 evaluado en las covariables de la estación x_i , Σ es la matriz de covarianza generada por la ecuación 14 en la ubicación de las estaciones, p_ξ viene de la distribución a priori escogida para ξ con parámetro θ_ξ , $\theta_1 = [\phi, \xi]^T$ y $\theta_2 = [\alpha_\phi, \beta_\phi, \theta_\phi]^T$.

La ecuación 15 nos proporciona la segunda parte del modelo completo, descrito en la ecuación 6, asociada al nivel 2 de esta primera etapa.

I Nivel de las distribuciones a priori

Nivel de las distribuciones a priori, en este le asignamos las distribuciones a priori a todos los parámetros involucrados en el modelo. Tenemos que α_ϕ , β_ϕ y θ_ϕ son los parámetros caracterizan el proceso latente, asumimos que cada parámetro en esta etapa es independiente de los otros.

No poseemos información a priori de como el parámetro ϕ se relaciona con las covariables, por esa razón escogemos prioris no informativas para los parámetros de regresión α_ϕ , asumimos que $\alpha_{\phi,i} \sim U(-\infty, \infty)$ para todos los modelos.

Los parámetros β_ϕ son los que describen la estructura espacial del parámetro de escala log-transformado de la GPD, para estos es aconsejable usar prioris informativas. Queremos estimar estos parámetros en nuestros datos, para ese fin utilizamos el paquete GeoR que nos proporciona herramientas para el análisis de datos geoestadísticos en R. A nuestro variograma empírico le asociamos un modelo teórico, con el fin de poder estimar los parámetros de covarianza, ajustamos el variograma experimental mediante un modelo exponencial, tomando en consideración el efecto nugget.

En la Figura 1, las líneas punteadas representan las envolventes correspondientes al modelo exponencial y los puntos corresponden al variograma experimental omnidireccional, que es además la franja de confianza para la estimación de los parámetros del variograma, lo que lo hace significativo a los valores que utilizaremos. Con la información obtenida del variograma exponencial definimos las distribuciones a priori de la siguiente manera:

- $\beta_{\phi,0} \sim U(0.01, 0.4)$,
- $\beta_{\phi,1} \sim U(0.01, 1.99)$.

Para las prioris de α_ξ tomamos $U(-\infty, \infty)$ y usamos información empírica para determinar prioris apropiadas para β_ξ .

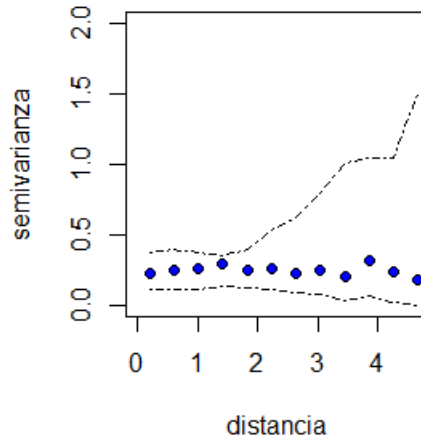


Figura 1: Ajuste del variograma experimental de los datos mediante el modelo exponencial

I Modelos de tasas de excedencias

Segunda etapa de nuestro modelo completo, esta es independiente de la primera etapa. Para estimar completamente los niveles de retorno, debemos incluir un parámetro adicional que denotaremos como ζ_u . Donde ζ_u representa la tasa a la que un grupo de observaciones supera el umbral u . Las tasas de excedencia asociadas a las estaciones meteorológicas las denotamos por:

$$\zeta_u = [\zeta_u(x_1), \dots, \zeta_u(x_n)]^T. \quad (16)$$

Inicialmente nuestros datos están desagrupados, no tenemos la probabilidad de que una observación exceda el umbral u . En nuestros datos tenemos de momento que ζ_u es la probabilidad que una observación sea un máximo de grupo. Por la forma en como están divididos nuestros datos y dada la elección previa del umbral, ζ_u varía para cada estación y puede ser modelada espacialmente.

Para la escogencia del umbral hay diferentes métodos, el valor que el umbral tome determina la cantidad de excedencias a considerar en el estudio. A menor valor de umbral mayor cantidad de excedencias, por otro lado cuanto más grande el umbral tendríamos muy pocas excedencias. En nuestro caso buscamos que el número de excedencias represente un porcentaje considerable del total de datos disponibles, de manera que tomar el 3% de las observaciones más altas resulto ser adecuado para nuestro objetivo. Tomamos el 3% de las observaciones mas altas para cada una de las estaciones meteorológicas incluidas en el estudio y con estos datos calculamos el promedio. El valor de umbral que escogimos fue de 45mm. (4.5 cm) para todas las estaciones, este valor también lo comparamos con el valor obtenido mediante el gráfico de vida media residual que es uno de los métodos clásicos mas usados para la elección del umbral (Coles, 2001).

Dado que la precipitación en Honduras no es homogénea para todo el país, la tasa de excedencia ζ_u cambia de acuerdo a la estación que tomemos. La cantidad de excedencias está directamente relacionada con la ubicación de la estación, razón por la cual la tasa de excedencia debe ser modelada espacialmente. Denotamos al parámetro de la tasa de excedencia en la estación \mathbf{x} por $\zeta(\mathbf{x})$.

De manera análoga a lo desarrollado en la etapa 1, para poder obtener inferencias sobre ζ_u utilizamos

un modelo jerárquico con 3 niveles.

I Nivel de los datos

Primer nivel, el nivel base de esta etapa, para cada estación calculamos el número de excedencias y lo denotamos como n_i . El número de excedencias total para cada estación sigue una distribución binomial $\text{Bin}(m_i, \zeta(\mathbf{x}_i))$, con:

- m_i : número total de observaciones por estación.
- $\zeta(\mathbf{x}_i)$: probabilidad de ser un grupo máximo.

I Nivel del proceso

Segundo nivel del modelo jerárquico, en este es donde modelamos el proceso espacial latente, lo trabajamos de manera similar a como se trabajó el parámetro $\phi(\mathbf{x})$ en la GPD del nivel 1 en la etapa 1. Siguiendo la línea de (Diggle y Tawn, 1998) transformamos nuestro parámetro $\zeta(\mathbf{x})$ usando una transformación logit, para modelar este parámetro transformado como un proceso Gaussiano con $E[\zeta(\mathbf{x})] = \mu_\zeta(\mathbf{x})$ y $\text{cov}(\zeta(\mathbf{x}), \zeta(\mathbf{x}')) = k_\zeta(\mathbf{x}, \mathbf{x}')$.

Donde $\mu_\zeta(\mathbf{x})$ es función de α_ζ y la información de las covariables, la covarianza es función de los β_ζ obtenidos a partir del variograma exponencial.

I Nivel de las prioris

Tercer nivel de nuestra segunda etapa, nivel donde asociamos las distribuciones a priori asociadas a los parámetros involucrados. Usaremos prioris no informativas para los parámetros de regresión, $\alpha_\zeta \sim U(-\infty, \infty)$. Basándonos en el variograma mostrado en la Figura 1, tomamos $\beta_{\zeta,0} \sim U[0.01, 0.4]$ y $\beta_{\zeta,1} \sim U[0.01, 1.99]$.

I MCMC en los modelos

Para obtener muestras en ambas etapas simulamos la distribución posterior haciendo uso de los Métodos de Monte Carlo por Cadenas de Markov (MCMC): Metrópolis Hastings y Muestreador de Gibbs. Estos métodos se describen (Infante y cols., 2010).

Para cada parámetro involucrado calculamos las distribuciones condicionales completas, ya que son las que nos permitirán simular muestras para los mismos, sin perder de vista que estamos usando modelos jerárquicos y los mismos nos marcan el orden en el cual simulamos los parámetros. Generamos las propuestas usando caminata aleatoria y propuestas independientes con el objetivo de alcanzar tasas de aceptación adecuadas.

Para acelerar la convergencia, cuando actualizamos ϕ , usamos el estimador de máxima verosimilitud de los parámetros de la GPD de cada estación para obtener un candidato adecuado. Sea $\hat{\phi}$ el estimador de máxima verosimilitud del parámetro ϕ de la GPD. Por las propiedades asintóticas del estimador de

máxima verosimilitud,

$$\hat{\phi} \approx \phi + \varepsilon, \quad (17)$$

donde $\varepsilon \sim MVN(\mu, \Sigma)$ y μ es definida por la ecuación 13 y Σ por la ecuación 14. Podemos denotar la distribución conjunta de $\hat{\phi}$ y ϕ por:

$$\begin{pmatrix} \hat{\phi} \\ \phi \end{pmatrix} = MVN \left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{bmatrix} \Sigma + I^{-1} & \Sigma \\ \Sigma & \Sigma \end{bmatrix} \right) \quad (18)$$

la distribución condicional es:

$$\phi | \hat{\phi} \sim MVN(\mu + \Sigma(I^{-1} + \Sigma)^{-1}(\hat{\phi} - \mu), \Sigma - \Sigma(I^{-1} + \Sigma)^{-1}\Sigma). \quad (19)$$

Usamos la ecuación 19 como la densidad del candidato en nuestro paso de Metropolis Hastings. Después del parámetro ϕ , actualizamos la media α_ϕ y β_ϕ . La densidad candidata de Metropolis Hastings para α_ϕ y β_ϕ son implementadas como caminatas aleatorias.

Para el parámetro de forma, repetimos el proceso, actualizamos ξ como una caminata aleatoria cuando lo modelamos, ya sea como un valor único o como un par de valores y de la misma manera que con ϕ cuando lo modelamos espacialmente. La tasa de excedencia se maneja de manera análoga.

Las simulaciones las realizamos usando el programa estadístico R. En cada simulación realizamos 100,000 iteraciones, posteriormente realizamos un *burn-in* de 1,000 iteraciones. Las cadenas resultantes mostraban dependencia, por lo que para reducir la misma aplicamos un *thinning* promedio de 30. Después de haber realizado *burn-in* y el *thinning* obtuvimos muestras finales de 3,300 valores aproximadamente.

I Estimador Gamma desplazado

Este método se basa en el hecho de que la distribución a posteriori de la log-verosimilitud es aproximadamente una distribución gamma. Esto conduce a un estimador de la máxima verosimilitud alcanzable, y también un estimador del número efectivo de parámetros que es sencillo de calcular a partir de las log-verosimilitudes, independientemente de la parametrización del modelo.

Este criterio se describe a detalle en (Raftery, Newton, Satagopan, y Krivitsky, 2007). El enfoque es utilizado para estabilizar la estimación de la media armónica. La log-verosimilitud marginal esta dada por:

$$\log\pi(y) = \log E[e^{-t}|y] = l_{max} + \alpha \log(1 - \alpha).$$

Los mejores modelos son aquellos con el mayor valor de log-verosimilitud marginal, esto permite seleccionar el modelo que realiza el mejor ajuste de los datos escogiendo así el modelo que hace a los datos observados los más probables.

IV | ANÁLISIS DE RESULTADOS

Este estudio está siendo desarrollado para Honduras tomando valores de precipitación diaria de 59 estaciones meteorológicas. Agradecemos a la Unidad de Meteorología del Instituto Hondureño de Ciencias de la Tierra (IHCIT), por su valiosa colaboración al compartimos los datos para poder desarrollar este

Precipitación mensual acumulada de referencia

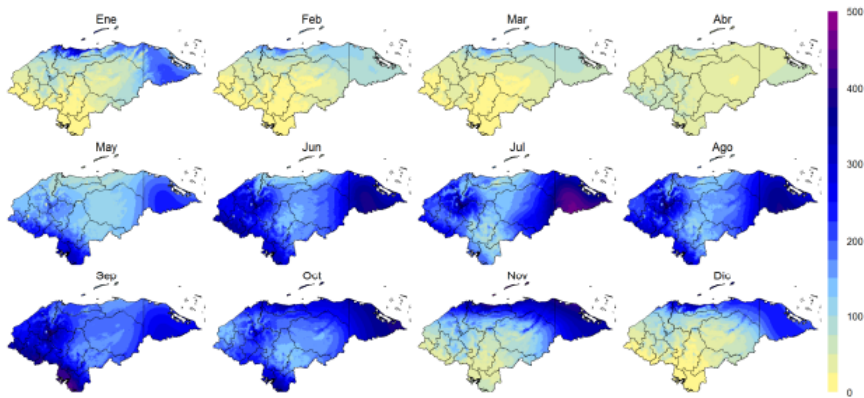


Figura 2: Mapas derivados de la interpolación de datos de estaciones meteorológicas de diversas fuentes. Los valores están dados en mm/mes. Fuente: (Navarro, Llano, Obando, y D. Córdoba, 2018)

trabajo.

De la data completa de valores de precipitación diaria se seleccionaron las estaciones que tienen datos desde 1972 hasta 2012, para cada estación seleccionada tenemos registros diarios de precipitación diarias de 41 años.

No se tomo en cuenta a todas las estaciones meteorológicas de Honduras porque no poseían la cantidad de registros completos para estos años. Conforme al diseño del estudio necesitamos un rango de meses en los cuales se observe mayor cantidad de precipitación para definir lo que denotamos como temporada lluviosa. Los valores de precipitación diaria correspondientes a los meses de diciembre a abril no son incluidos en los modelos. Los valores que se incluyen son los que corresponden a los meses de mayo a noviembre, como podemos observar en la Figura 2 la mayor precipitación acumulada se presenta en estos meses.

Definimos un rango de fechas como temporada lluviosa, tomamos los valores de precipitación diaria para las fechas del 1 de mayo al 30 de noviembre, por ser estos los meses en los que se identifica la mayor cantidad de precipitación en Honduras. Para cada una de las estaciones meteorológicas se cuenta con 214 registros tomados por año, es decir 8774 registros para cada una de las 59 estaciones meteorológicas escogidas. Lo que deseamos estimar son los extremos en precipitación, en nuestro caso valores máximos, si usáramos todos los valores del año estaríamos incluyendo días con poca precipitación que influirían en los resultados, llevándonos a conclusiones erróneas. Por otra parte no se introduce sesgo en las estimaciones ya que buscamos estudiar extremos, más aún no trabajamos con valores de precipitación diaria sino con valores de excedencia. Nuestro umbral es de 45mm., dentro de las estaciones seleccionadas usamos aquellas que igualen o superen este valor, estas excedencias se convierten entonces en nuestro objeto de interés. Estos datos conforman la muestra con la que se desarrollo el estudio, previo a la modelización los datos de nuestra muestra se realizo una limpieza de

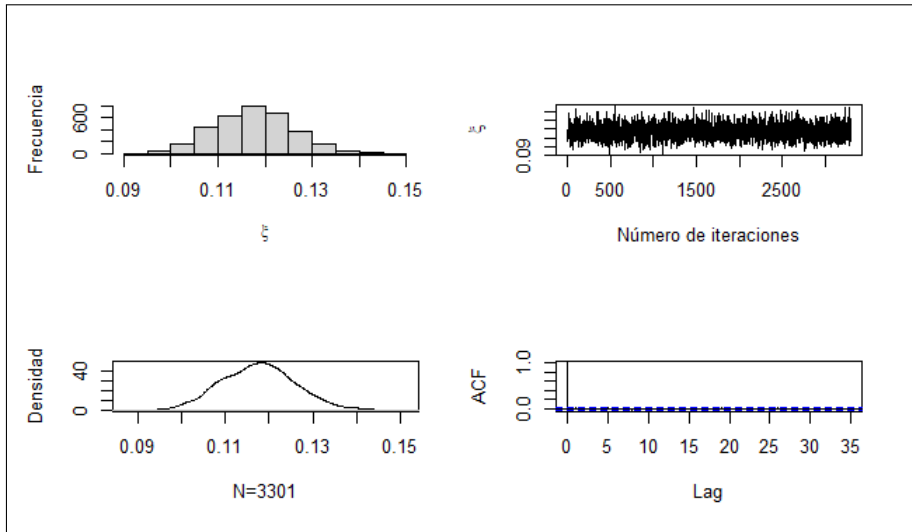


Figura 3: Valores de ξ modelo 3

los datos, revisando inconsistencias para verificar la calidad de los mismos y que cumplan también con los requerimientos de homogeneidad, independendencia y estacionalidad.

En la Tabla 1 presentamos los valores de precipitación promedio, así como los valores de precipitación máximo y mínimo para algunas de las estaciones de la muestra por regiones. Se pueden observar también los valores de desviación estándar y percentil 95 que nos da una idea de que valores podríamos considerar como umbrales si decidiéramos hacer el análisis por región.

En las Tablas 2 y 3 se muestran todos los modelos considerados en las etapas 1 y 2 respectivamente. Para cada modelo se indica el valor de log-verosimilitud asociado, los modelos resaltados son los modelos escogidos en cada etapa.

Las prioris que utilizamos son uniformes, por lo tanto no informativas, en las Figuras 3, 4a, 4b y 5 se muestran las gráficas de las estimaciones a posteriori. Por la naturaleza de las distribuciones a priori tomadas no se introduce sesgo en el análisis a posteriori.

En la Figura 3, se presentan los gráficos asociados a ξ de la muestra final para el modelo 3. Los valores que toma ξ evidencian la presencia de colas pesadas, estos valores se mantienen para todas las estaciones. Se puede observar la convergencia del parámetro y como después de realizar el refinamiento a la muestra no se observa dependencia entre las observaciones.

En las Figuras 4a y 4b, se muestran los gráficos resultantes de ϕ , en las muestras obtenidas para el modelo 3, asociado a las estaciones meteorológicas 30 y 59 respectivamente, los resultados obtenidos en el resto de las estaciones son bastantes similares. En los gráficos se puede observar la convergencia del parámetro en las estaciones después de realizar el refinamiento, a no se observa dependencia entre las observaciones. El parámetro de forma presenta variaciones en sus valores, esta variación se debe a que la covariable altitud es significativa en este modelo. Podemos decir que las variaciones en la

	# de estación	Media	Máximo	Percentil 95	Desviación típica
Región Occidental	7	61.1	104.0	92.2	14.1
	23	65.4	220.5	100.0	20.7
	24	59.0	117.2	90.9	14.6
	58	62.4	172.7	91.4	18.2
Región Nor-occidental	2	60.8	159.3	99.7	19.1
	4	59.4	122.5	88.7	15.1
	13	67.7	198.4	115.8	24.1
	25	62.3	130.0	97.2	17.3
	51	66.8	236.0	113.4	28.3
Región Nor-oriental	26	81.2	263.4	165.7	41.8
	49	102.1	556.4	266.1	73.3
	50	88.3	360.0	195.0	49.6
	52	75.6	271.1	147.7	33.9
Región Centro-occidental	5	61.4	138.8	90.4	15.5
	8	74.1	226.9	129.6	30.9
	12	60.6	123.2	90.2	15.4
	14	59.3	134.5	84.6	13.9
	57	62.0	173.0	91.8	19.1
Región Centro-occidental	6	58.4	140.0	93.1	15.7
	20	66.0	204.3	153.3	31.5
	37	75.1	390.1	130.8	33.1
	38	61.7	200.2	90.0	19.0
	56	57.2	92.5	82.0	10.8
Región Sur	48	71.6	260.1	128.0	28.5
	55	70.7	285.0	136.6	30.2

Tabla 1: Estadísticos descriptivos de algunas estaciones por región

Modelo	Log-verosimilitud marginal
Modelo 0	
$\phi = \phi$	-42623.61
$\xi = \xi$	
Modelo 1	
$\phi = \alpha_0 + \varepsilon_\phi$	-32001.55
$\xi = \xi$	
Modelo 2	
$\phi = \alpha_0 + \alpha_1(mp) + \varepsilon_\phi$	-31921.52
$\xi = \xi$	
Modelo 3	
$\phi = \alpha_0 + \alpha_1(altitud) + \varepsilon_\phi$	-31901.19
$\xi = \xi$	

Tabla 2: Descripción de los modelos utilizados en la etapa 1 y sus valores de log-verosimilitud marginal asociada

altitud de las estaciones meteorológicas se ven reflejadas en el parámetro de escala de la GPD que ajusta nuestros valores de excedencias.

La tasa de excedencia es modelada como un parámetro que se comporta de manera uniforme para todas las estaciones. En la Figura 5, podemos observar que bajo este modelo ζ toma valores entre 0.0206 y 0.0218, se evidencia la convergencia del parámetro. En esta muestra final que se obtiene al realizar el refinamiento a la muestra inicial no hay dependencia en las observaciones.

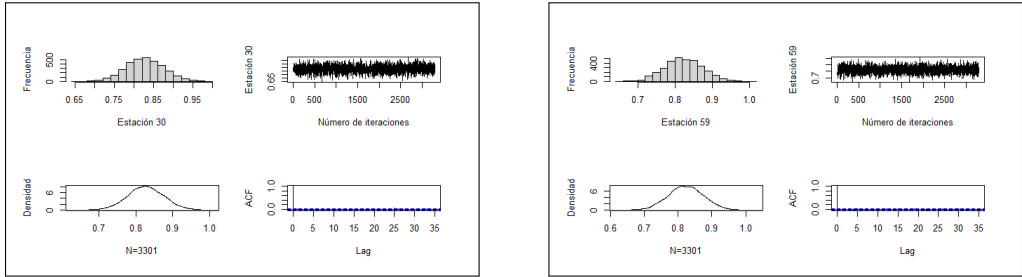
1 Estimaciones de los parámetros

Para estimar los niveles de retorno tomamos los resultados de los modelos escogidos en las 2 etapas, para las estimaciones puntuales calculamos la media a posteriori de los parámetros. Los resultados obtenidos para ξ y ζ se muestran en la Tabla 4.

Las estimaciones para el parámetro de forma se hacen para cada estación, recordemos que $\phi = \log(\sigma)$, de manera que $\sigma = e^\phi$. Los valores de las estimaciones puntuales para cada una de las estaciones se muestran en la Figura 6.

1 Estimaciones de los niveles de retornos

Se muestran en la Figura 7, los valores de los niveles de retorno estimados a 5 años (Figura 7a) y 50 años (Figura 7b) mediante mapas de niveles de retornos. A medida que tomamos periodos de tiempo más largos los valores asociados a los niveles de retorno van aumentando, de manera que la probabilidad de ser alcanzados o excedidos va disminuyendo.



(a) Valores de ϕ para el modelo 3 obtenidos en la estación 30 (b) Valores de ϕ para el modelo 3 obtenidos en la estación 59

Figura 4: Estimaciones de ϕ modelo 3

Modelo	Log-verosimilitud marginal
Modelo A $\zeta = \zeta$	-24.54897
Modelo B $\zeta = \alpha_0 + \epsilon_\zeta$	-30.39126
Modelo C $\zeta = \alpha_0 + \alpha_1(mp) + \epsilon_\zeta$	-30.37409
Modelo D $\zeta = \alpha_0 + \alpha_1(altitud) + \epsilon_\zeta$	-30.68142
Modelo E $\zeta = \alpha_0 + \alpha_1(mp) + \alpha_2(altitud) + \epsilon_\zeta$	-32.36031

Tabla 3: Descripción de los modelos utilizados en la etapa 2 y sus valores de log-verosimilitud marginal asociados

Al tener presencia de valores de niveles de retorno tan altos, es menos probable que ese valor sea alcanzado o excedido. Los valores de niveles de retorno asociados a cada una de las estaciones involucradas en este estudio y para las diferentes ubicaciones geográficas en general, presentan variaciones porque como hemos mencionado a lo largo de este trabajo, la precipitación no es un fenómeno con un comportamiento constante. Los valores de precipitación cambian, de igual manera los niveles de retorno, debido a la ubicación de la estación meteorológica de interés y mas aún como lo describen nuestros modelos, debido a la altitud.

V | CONCLUSIONES

- El parámetro de forma ξ es un parámetro que es altamente sensible.
- El modelo escogido evidencia que la covariable de elevación es significativa, lo que nos indica que los valores de precipitación diaria en las estaciones escogidas presentan variaciones entre si, producto de su elevación. Por otra parte para las tasas de excedencia no se encontro evidencia de que las

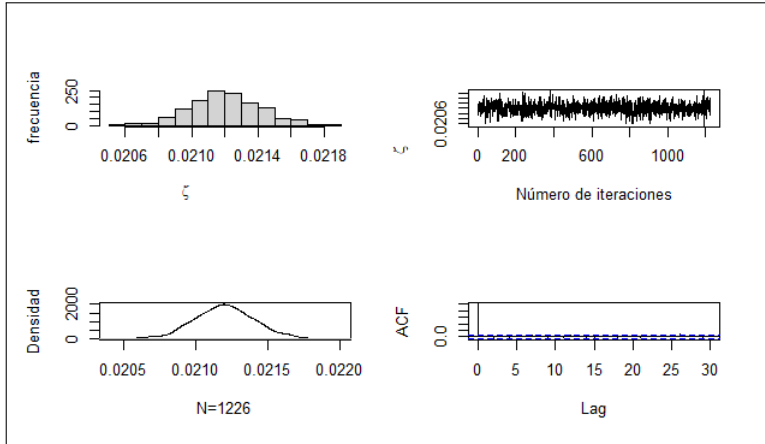


Figura 5: Valores de ζ para el modelo A

Parámetro	Estimación	IC _{95%}
ξ	0.1176	[0.1017, 0.1341]
ζ	0.0212	[0.0208, 0.0216]

Tabla 4: Valores estimados de los parámetros para el modelo completo

covariables tengan incidencia en el cálculo de las mismas, ya que estas se mantienen constantes para todas las estaciones.

- Tenemos pocas estaciones que cumplen con los requerimientos para este estudio, ya sea porque la cantidad de datos no es suficiente o por la calidad de los mismos. Hay zonas bastantes grandes del país que no se alcanzan a cubrir, hacer interpolaciones en esta zonas nos llevaría a tomar conclusiones erróneas por eso se presentan únicamente las estimaciones puntuales asociadas a las estaciones escogidas.
- La mayoría de las estaciones incluidas en el estudio tienen un comportamiento similar en cuanto a los niveles de retorno se refiere, se ubican en su mayoría en la escala más baja esto tomando en cuenta sus ubicaciones y la altitud como covariable influyendo en los valores de precipitación diaria. Esto tiene sentido ya que ocurren 2 cosas con las ubicaciones de las estaciones escogidas para este estudio, la mayoría se ubican geográficamente cerca y solo estamos tomando en cuenta la altitud de las estaciones y segundo la mayoría de las estaciones se encuentran en zonas con valores de precipitación bajos.

REFERENCIAS

Armstrong, M. (1998). *Basic linear geostatistics*. Springer Springer Berlin Heidelberg.
 Balkema, A., y Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2, 792-804.
 Banerjee, S., Carlin, B., y Gelfand, A. (2003). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.
 Capel, J. (1994). Aspectos geográficos de la pluviometría de honduras. *Papeles de geografía*, 20, 9-26.

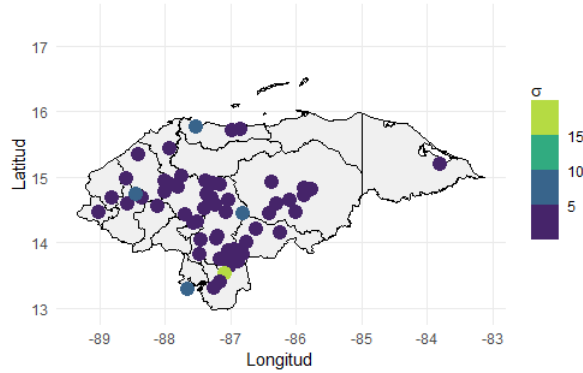
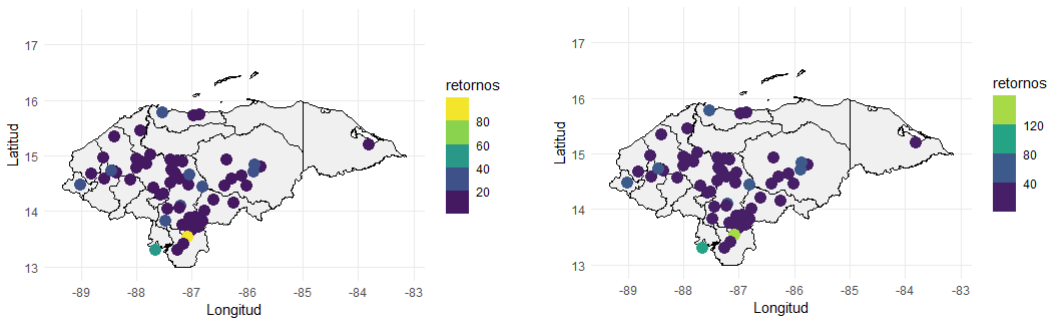


Figura 6: Mapa de estimaciones de σ para cada estación meteorológica



(a) Mapa de niveles de retorno estimados a 5 años

(b) Mapa de niveles de retorno estimados a 50 años

Figura 7: Estimaciones de los niveles de retornos

- Casella, G., y Berger, R. (2002). *Statistical inference*. Duxbury.
- Coles, S. (2001). *An introduction to statistical modeling of extrem values*. London: Springer series in statistics.
- Cooley, D., Nychka, D., y Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Associations*, 102, 824-840.
- Diggle, P., y Tawn, J. (1998). Model-based geostatistics. *The Applied Statistics*, 47, 299-350.
- Embrechts, P., Klüppelberg, C., y T., M. (1994). *Modelling extremal events for insurance and finance*. Stuttgart: UTB.
- Fisher, R., y Tippett, (1928). On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceeding of the Cambridge Philosophical Society*, 24, 180-190.
- Gamerman, D., Migon, H., y Louzada, F. (2015). *Statistical inference an integrated approach*. Boca Raton: Chapman and Hall Texts in Statistical Science Series.
- Giraldo, R. (2002). *Introducción a la geoestadística-teoría y aplicación*.
- Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44, 423-453.
- Gnedenko, B. V., y Kolmogorov, A. N. (1954). *Limit distributions for sums of independent random variables*. Cambridge: Addison-Wesley.
- Infante, S., A., H., y Urquiola, A. (2010). *Introducción a la inferencia bayesiana*.
- IPCC. (2022). *Climate change 2022: Impacts, adaptation, and vulnerability*.
- Liu, H., y Wasserman, L. (2014). *Statistical machine learning*.

- Mochales, I. (2013). *Modelos estadísticos para valores extremos y sus aplicaciones* (Tesis de Doctorado). Universidad Autónoma de Barcelona, Barcelona.
- Navarro, F., C. and Monserrate, Llano, L., Obando, y D. Córdoba, J. (2018). *Desarrollo de los escenarios climáticos de Honduras y módulo académico de capacitación*. (Inf. Téc.). Honduras: Centro Internacional de Agricultura Tropical (CIAT); Programa de las Naciones Unidas para el Desarrollo (PNUD); Dirección Nacional de Cambio Climático de MiAmbiente.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3, 119-131.
- Raftery, A., Newton, M. A., Satagopan, J. M., y Krivitsky, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian statistics*, 8, 1-45.
- Rodríguez, M., Kawas, N., Elvir, O., Torres, L., Wiese, K., Cardona, A., ... Castillo, P. (2012). *Atlas climático y de gestión de riesgo de honduras*.
- Sierra, E. (2014). *Herramientas geoestadísticas en la construcción de modelos digitales de elevación*.