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

Méndez Ramos, Sonia Marisol[1];

Cruz Torres, Cristian Andrés[2];

Palabras Clave: Niveles de Retorno, Modelos Jerárquicos, Teoría de Valores Extremos, Inferencia Bayesiana, Precipitación

Resumen:

Producto del cambio climático es cada vez más frecuente la presencia de valores extremos en términos de precipitación. Modelar valores extremos en precipitación es importante por sus implicaciones, una forma de cuantificarlos es mediante los niveles de retornos. Para poder estimar estos niveles de retorno de precipitación, se desarrolla una primera etapa asociada a los parámetros de la Distribución Pareto generalizada y una segunda etapa asociada a las tasas de excedencia, existe independencia entre cada etapa por la naturaleza de los parámetros. Los modelos en ambas etapas son abordados mediante modelos jerárquicos desde un punto de vista bayesiano.

La precipitación es un fenómeno atmosférico fuertemente ligado a la ubicación geográfica, por esta razón se toma en cuenta esta componente espacial del fenómeno usando covariables climatológicas y geográficas como son la altitud y la precipitación media. Para cada etapa se simulan las distribuciones a posteriori de los parámetros de interés usando algoritmos de MCMC, el mejor modelo es escogido usando como criterio de comparación la log-verosimilitud marginal, las estimaciones de los modelos escogidos se usan de manera conjunta para poder construir con ellas mapas de niveles de retorno.

[1] Maestría en Matemática ,Universidad Nacional Autónoma de Honduras (UNAH); sonia.mendez@unah.edu.hn
[2] Maestría en Matemática ,Universidad Nacional Autónoma de Honduras (UNAH); cristian.cruz@unah.edu.hn

Este estudio está siendo desarrollado para Honduras tomando valores de precipitación diaria de 1972 hasta 2012 en 59 estaciones meteorológicas, los resultados evidencian que la altitud 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. Los valores de los niveles de retorno estimados tienen un comportamiento bastante similar para todas las estaciones escogidas en el estudio, las variaciones presentes en las estimaciones de los mismos están asociadas a la ubicación geográfica de las estaciones meteorológicas y a su elevación.

Introducción

Según resultados obtenidos en modelos desarrollados por el grupo de expertos internacional que estudia el cambio climático, llamado Panel Intergubernamental para el Cambio Climático (IPCC), presentados en su reporte técnico de 2022 se enuncia en términos de precipitación que debido al cambio climático:

- Se espera que a lo largo del siglo XXI, las sequías e inundaciones podrían ser más frecuentes y prolongadas.
- Se esperan incrementos en los actuales niveles de evaporación y precipitación.
- Un aumento muy probable de las precipitaciones intensas en regiones donde la precipitación es mayor.
- El suministro de agua potable podría peligrar en diversas zonas y algunas epidemias podrían extenderse más fácilmente

Estos resultados muestran la importancia de estudiar valores extremos en precipitación, ya que la precipitación es uno de los fenómenos meteorológicos donde la ocurrencia de valores extremos es cada vez más frecuente.

La presencia de valores muy altos o muy bajos en un estudio implica que realicemos un abordaje diferente. Este análisis se puede realizar mediante la teoría de valores extremos, esta teoría es muy importante por su aplicación en áreas como ingeniería, meteorología, economía, entre otras. En Honduras, si hablamos específicamente de precipitaciones extrema, 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.

Si bien es cierto no es posible detener la ocurrencia de fenómenos extremos, conocer su impacto brinda la oportunidad de crear planes de gestión de riesgos que permitan mitigar las consecuencias de las ocurrencias de los mismos y desarrollar a partir de ellos planes de prevención a nivel de país. De manera que es necesario 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.

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.

Descripción del método

Para la estimación de los parámetros involucrados en el modelo final utilizado para la construcción de los mapas de los niveles de retorno se desarrollan 2 etapas:

1. Asociada a los parámetros de la Distribución Pareto generalizada.
2. Asociada a las tasas de excedencia.

Las etapas son independientes entre sí por la naturaleza de los parámetros, cada una de las etapas se aborda mediante modelos jerárquicos de 3 niveles descritos a continuación:

1. Nivel de los datos
2. Nivel del proceso latente que controla la precipitación climatológica extrema

3. Nivel de las distribuciones a priori de los parámetros involucrados en el proceso latente.

En la etapa 1 la finalidad es modelar las excedencias, siendo una excedencia aquel valor que supera un umbral (**u**) escogido. De la teoría de valores extremos tenemos que las excedencias obtenidas luego de la escogencia de un umbral óptimo deben aproximarse a una Distribución de Pareto Generalizada (GPD, por sus siglas en inglés) a medida que el umbral aumenta y el tamaño de muestra crece, esto por la teoría asintótica.

La Distribución de Pareto Generalizada introducida por Pickands en 1975 es la familia bi-paramétrica de funciones de distribución dada por:

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

Donde:

σ: $parámetro\ de\ escala$

ξ: $parámetro\ de\ escala$

El parámetro de forma nos permite también clasificar las colas de una distribución de la siguiente manera:

- Cola ligera: $\xi < 0$
- Colas exponenciales o normales: $\xi = 0$
- Cola pesada: $\xi > 0$

En la etapa 2 se busca modelar las tasas de excedencia: ζ_u. Para hacer las inferencias sobre ζ_u utilizamos también un modelo jerárquico con 3 niveles, teniendo siempre en cuenta que hay un proceso espacial latente que influye en las probabilidades de excedencia y cuyo efecto debe ser incluido y modelado.

En 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 y aunque el nivel de retorno tiene una forma cerrada, es una función no lineal de los parámetros GPD y las tasas de excedencia.

Para relacionar la GPD y los niveles de retorno se tiene que:

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

Si tomamos a n_y como el número de observaciones en un año con $P(Z > z_r) = \frac{1}{rn_y}$, despejando tendríamos que:

$$z_r = u + \frac{\sigma_u}{\xi}\left[(rn_y\zeta_u)^{\xi} - 1\right] \quad (2)$$

con esta ecuación obtenemos el r-anual nivel de retorno z_r. El objetivo es poder estimarlos para cada estación meteorológica incluida en el estudio, ya que los niveles de retorno son cantidades climatológicas y la precipitación es un fenómeno que por su naturaleza se comporta de manera diferente para diferentes puntos geográficos, es decir hay una componente espacial que debe ser incluida en los modelos.

Sea $Z(x)$ la precipitación diaria para la ubicación x con sus coordenadas dadas por la longitud y latitud, las excedencias son todos aquellos valores de precipitación diaria que superen el umbral escogido para todas las estaciones, el componente espacial está implícito ya que σ_u ,ξ y ζ_u son funciones de x.

Para la etapa 1, el modelo jerárquico se describe a continuación:

1. Nivel de los datos

Sean $x_1, x_2, \ldots, x_s$ las ubicaciones de las estaciones denotamos el parámetro de escala de la GPD como $\sigma_u = [\sigma_u(x_1), \ldots, \sigma_u(x_s)]^T$ de manera equivalente se definen ξ y ζ_u.

La inferencia sobre θ, dados los datos de las estaciones está dada por la regla de Bayes

$p(\theta|Z(x)) \propto p(Z(x)|\theta)p(\theta),$ p denota una densidad de probabilidad. A partir de la distribución condicional del modelo jerárquico se tiene

$$p(\theta|Z(x)) \propto p_1(Z(x)|\theta_1)p_2(\theta_1|\theta_2)p_3(\theta_2) \qquad (3)$$

donde p_j es la densidad asociada al nivel j del modelo jerárquico. Respecto a los modelos jerárquicos para excedencias de umbrales, se reparametriza $\phi = log\sigma_u$ para permitir que ϕ tome tanto valores positivos como negativos.

Sea $Z_k(x_i)$ el k-ésimo registro de precipitación diaria en la ubicación(x_i), $i = 1, \ldots s$ dado que $Z_k(x_i)$ supera el umbral queda descrito su comportamiento por la GDP*:*

$$P\{Z_k(x_i) - u > z|Z_k(x_i) > u\} = \left(1 + \frac{\xi(x_i)z}{exp\ \phi(x_i)}\right)^{-1/\xi(x_i)} \qquad (4)$$

Con $\phi(x_i)$ y $\xi(x_i)$ los parámetros medidos en la ubicación x_i.

Derivando se obtiene la función de distribución asociada:

$$p_1(Z(x)|\theta_1) = \prod_{i=1}^{s}\prod_{k=1}^{n_i}\frac{1}{exp\ \phi(x_i)}\left(1 + \frac{\xi(x_i)z}{exp\phi(x_i)}\right)^{-1/(x_i)-1} \qquad (5)$$

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

2. Nivel del proceso especial latente

Cada estación escogida tiene una ubicación geográfica diferente dada por su longitud y latitud, dentro del marco bayesiano $\phi(x)$ y $\xi(x)$ son variables aleatorias por lo tanto tienen distribuciones a priori asociadas.

Se hace uso de métodos de geoestadística para modelar $\phi(x)$ como un proceso Gaussiano con $E[\phi(x)] = \mu_\phi(x)$ y $cov(\phi(x), \phi(x') = k_\phi(x, x')$, la media $\mu_\phi(x)$ es función de los parámetros α_ϕ y las covariables:

$$\mu_\phi(x) = f_\phi(\alpha_\phi, covariables) \qquad (6)$$

La función f es fácilmente adaptable para permitir que la media sea función de combinaciones entre las covariables .La función de covarianza es función de la distancia entre las estaciones y los parámetros β_ϕ esta dada por:

$$k_\phi(x, x') = \beta_{\phi,0} \times \exp\left(-\beta_{\phi,1} \times \|x - x'\|\right) \qquad (7)$$

que está asociado a un variograma exponencial, los parámetros $\beta_{\phi,0}$ representan la meseta y $1/\beta_{\phi,1}$ el rango, en el contexto geoestadístico estos capturan el componente espacial de los modelos y son calculado empíricamente a partir de las ubicaciones geográficas de las estaciones.

3. Nivel de las prioris

Los parámetros de interés tienen distribuciones a priori asociadas, tenemos que $\alpha_{\phi,i} \sim Unif(-\infty, \infty)$. Después de hacer un análisis geoestadístico exploratorio basándonos en el variograma exponencial determinamos las prioris adecuadas para $\beta_{\phi,0}$ y $\beta_{\phi,1}$.

La etapa 2 se desarrolla de manera análoga a la etapa 1, el parámetro ζ_u se aborda de manera similar a como se trabaja ϕ_u.

Discusión de los resultados

Este estudio está siendo desarrollado para Honduras, tomando valores de precipitación diaria de 59 estaciones meteorológicas. De la data completa se seleccionaron las estaciones que tienen datos desde 1972 hasta 2012, para cada estación seleccionada tenemos registros de precipitación diaria de 41 años para las fechas del 1 de mayo al 30 de noviembre por ser las fechas en las que se identifica mayor cantidad de precipitación incluyendo la temporada lluviosa. El umbral se escogió haciendo uso de gráficos de vida media residual, una vez escogido el umbral se calcularon las excedencias por estación.

Decidimos incluir a la precipitación media (mp) y la altitud como covariables. Para poder hacer las estimaciones realizamos simulaciones a partir de las distribuciones condicionales completas usando Métodos de Monte Carlo vía Cadenas de Markov como ser el algoritmo de Metrópolis Hastings y muestreador de Gibbs para ambas etapas. Generamos las propuestas usando caminata aleatoria y propuestas independientes con el objetivo de alcanzar tasas de aceptación adecuadas. Usamos el criterio de log-verosimilitud marginal mostrado en (Raftery, Satagopan, Newton, y Krivitsky, 2007) se escogió el mejor modelo para cada etapa y con las estimaciones obtenidas en las etapas 1 y 2 se estima el nivel de retorno.

Los modelos son trabajados en el espacio Latitud/longitud, las tasas de excedencia son también modeladas espacialmente es decir para cada estación se tiene una tasa de excedencia, los modelos propuestos para cada etapa y los resultados de los valores de la log-verosimilitud Marginal asociados a cada modelo se resumen en la Tabla 1.

*Tabla 1: Resumen de los modelos en cada etapa con sus valores de L*og-verosimilitud marginal

Modelo		**Log-verosimilitud Marginal**
	Etapa 1	
Modelo 1	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	16763657
Modelo 2:	$\phi = \alpha_0 + \alpha_1(mp) + \epsilon_\phi$ $\xi = \xi$	27973473
Modelo 3:	$\phi = \alpha_0 + \alpha_1(altitud) + \epsilon_\phi$ $\xi = \xi$	80700983
	Etapa 2	
Modelo A:	$\zeta = \zeta$	-24.54897
Modelo B:	$\zeta = \alpha_0 + \epsilon_\zeta$	-30.39126
Modelo C:	$\zeta = \alpha_0 + \alpha_1(mp) + \epsilon_\phi$	-30.37409
Modelo D:	$\zeta = \alpha_0 + \alpha_1(altitud) + \epsilon_\phi$	-30.68142
Modelo E:	$\zeta = \alpha_0 + \alpha_1(mp) + \alpha_2(altitud) + \epsilon_\phi$	-32.36031

Los mejores modelos para cada etapa son aquellos con el mayor valor de log-verosimilitud marginal, estos se resaltan en la tabla 1, esto permite seleccionar el modelo que realiza el mejor ajuste de los datos de manera que se escoja el modelo que hace a los datos observados los más probable.

Para la etapa 1 se escogió el modelo 3, en este tenemos que:

$$\boldsymbol{\phi = \alpha_0 + \alpha_1(altitud) + \epsilon_\phi}$$

$$\boldsymbol{\xi = \xi}$$

Cada estación tiene un valor ϕ asociado diferente y depende del valor de α_0, α_1 y de la altitud, que fueron los parámetros significativos para este modelo en cambio para ζ se mantiene constante en todas las estaciones, al escoger este modelo se observó que para nuestros datos la altitud es significativa.

En la etapa 2 escogimos el modelo A, en el que tenemos:

$$\boldsymbol{\zeta = \zeta}$$

esto nos muestra que las tasas de excedencia se mantienen constantes para todas las estaciones y que no hay evidencia con estos datos para tomar en cuenta la influencia de la precipitación media y la altitud en la estimación de las tasas de excedencia.

Finalmente en nuestro trabajo en proceso está el poder determinar las estimaciones de los parámetros de interés para poder determinar periodos de retorno adecuados que nos permitan construir los mapas para los niveles de retorno y poder tener las conclusiones finales.

Los modelos escogidos en cada etapa son aquellos con mayor valor de log-verosimilitud marginal. Para los datos trabajados la covariable de altitud resulto ser significativa, es decir tiene una influencia notable sobre los valores de precipitación usados, lo que nos dice que los valores de precipitación para estas estaciones de Honduras escogidas presentan variaciones entre si producto de la elevación de las ubicaciones de las mismas. Por otra parte con las tasas de excedencia no se evidencia que las covariables tengan incidencia en el cálculo de las mismas. Se construirán los mapas de retorno usando el Modelo 3 y el modelo A para así poder presentar las conclusiones finales que son el punto focal de este trabajo.

Referencias

Banerjee, S., Carlin, B., y Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* Boca Raton: Chapman and Hall/CRC.

Coles, S. (2001). *An Introduction to statistical modeling of extreme values.* London,U.K.: Springer.

Cooley, D., Nychka, D., y Naveau, P. (2007). Bayesian Spatial Modelling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association*.

Diggle, P., y Tawn, J. (1998). Model- based geoestatistics. *The applied Statistics*, 61-111.

Gelman, A., Carlin, J., Stern, H., y Rubin, D. (2003). *Bayesian Data Analysis* (Vol. 2nd ed.). Boca Raton: Chapman & Hall.