

# Proyecto 2 - Ubicaciones de Volcanes en Japón

*Andrés Arguedas Leiva - Natalia Díaz Ramírez*

*11 Octubre 2018*

## Introducción

El objetivo del presente trabajo consiste en determinar si existen agrupaciones de volcanes en Japón. Con base en el objetivo mencionado anteriormente, se pretende contestar las siguientes preguntas:

- ¿Los volcanes están distribuidos aleatoriamente o existe evidencia de agrupaciones en la región de estudio?
- ¿Existe algún modelo para predecir la intensidad de volcanes en ciertas áreas de la región de estudio?

## Metodología

Los datos utilizados en este trabajo se obtuvieron del Global Volcanism Program (GVP) del Instituto Smithsonian, los cuales hacen referencia a ubicaciones de volcanes que han estado activos en los últimos 10.000 años a nivel mundial. El GVP informa sobre las erupciones actuales de todo el mundo y mantiene un repositorio de bases de datos sobre volcanes activos y sus erupciones, además de ciertas variables de interés como su ubicación, año de la última erupción, tipo de volcán, comportamiento sísmico, entre otros. Para efectos de este trabajo, solamente se usaron las variables de la latitud y longitud en la cual está ubicado el volcán (para poder hacer una identificación del punto en el espacio). Además, se usaron los datos de la página GADM para obtener un mapa administrativo de Japón, para poder hacer gráficos sobre este.

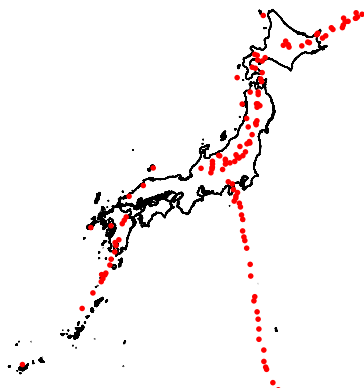
En términos del análisis, de la lista de todos los volcanes, se escogieron solamente aquellos que estuvieran en territorio de Japón, con el fin de hacer un análisis con resultados y conclusiones más precisas al tratarse de un área más reducida. Adicionalmente, por la historia geológica de Japón, en este país hay una gran cantidad de volcanes, por lo que, en términos de datos, se usaron los 128 volcanes de la serie de datos original que estuvieran cercanos al territorio japonés. Los datos se analizaron mediante métodos y modelos de procesos puntuales ya que, al tratarse de eventos geográficos en un solo punto, se pueden considerar como puntos en el espacio, además que se tiene la ubicación del volcán en latitud-longitud. Los datos fueron procesados mediante el software R, en la interfaz gráfica RStudio y se usaron las bibliotecas `spatstat`, para hacer los modelos de procesos puntuales, `dplyr`, para manipulación de datos y `sp`, para poder hacer algunos gráficos.

## Resultados

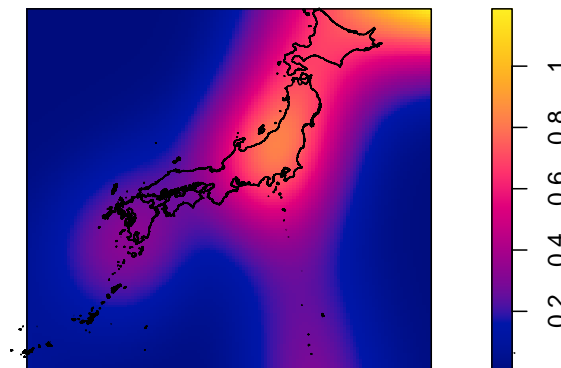
En esta sección se presenta, primeramente, un análisis descriptivo de los datos mediante mapas con las ubicaciones de los volcanes y gráficos de intensidad. Posteriormente, se hace un análisis sobre la posible existencia de aglomeraciones de datos mediante las funciones L y K, usando varios ajustes por borde y simulaciones. Por último, se estiman dos modelos de procesos puntuales para intentar explicar la ubicación de los volcanes, haciendo pronósticos con base en estos y seleccionando al mejor.

Para empezar, es importante tener una idea sobre la distribución de los datos y la intensidad de la cantidad de volcanes en Japón. Para esto, el Gráfico 1 presenta un mapa de Japón con puntos rojos que representan cada uno de los volcanes, mientras que el Gráfico 2 presenta la intensidad estimada según la locación de los eventos.

**Gráfico 1:**  
**Ubicación de volcanes en Japón**

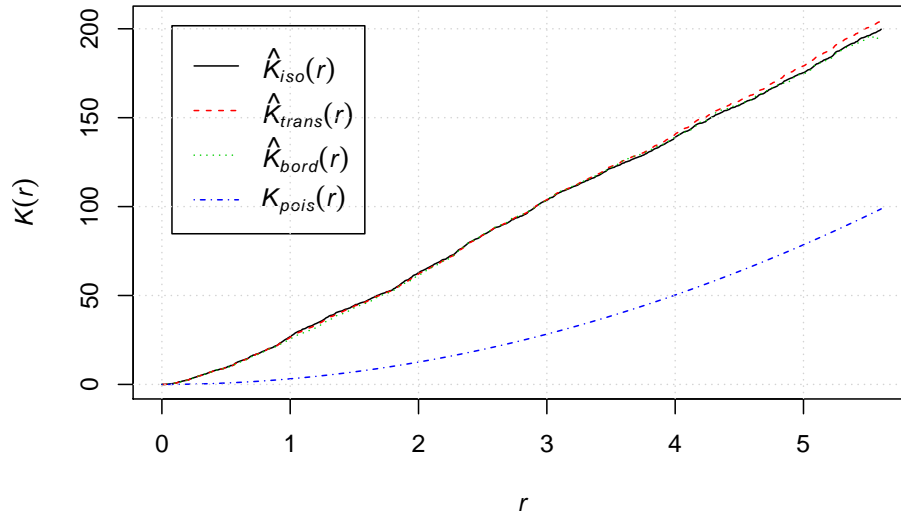


**Gráfico 2:**  
**Gráfico de densidad de volcanes en Japón**

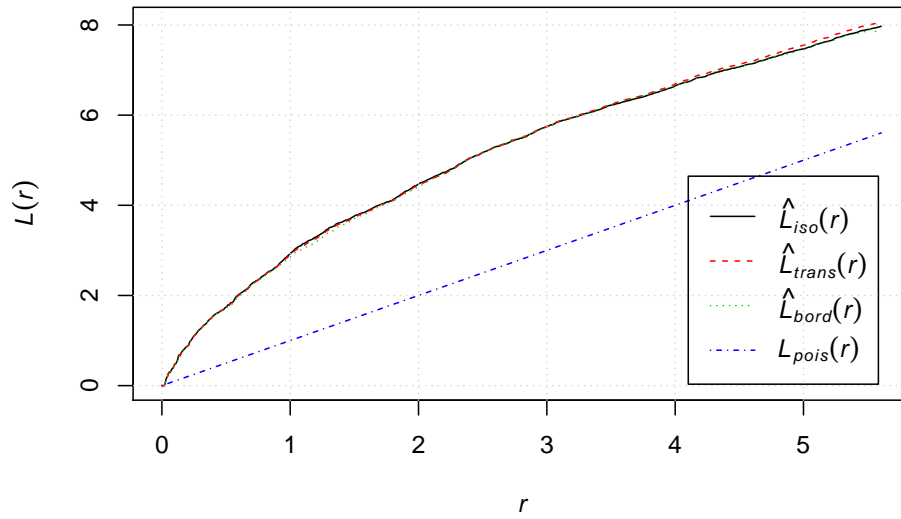


Con base en el Gráfico 1, se puede observar que hay concentraciones en ciertas áreas del territorio, de forma específica sobre las partes noreste de las islas, y un gran conjunto de volcanes que bajan hacia el sureste, de forma lineal. Complementando los resultados del Gráfico 1 con los del Gráfico 2, parece que las intensidades más grandes se encuentran localizadas sobre la parte noreste del mapa, con un foco en la isla al sur y una intensidad moderada sobre las islas al sureste. Por lo tanto, con base en estos dos gráficos, parecen haber indicios que realmente la distribución de los volcanes no es aleatoria, sino que hay ciertas aglomeraciones en secciones específicas. Para lograr ver este fenómeno de algomeración a mayor profundidad, a continuación se presentan las funciones K y L, basados en las propiedades de segundo orden, con el fin de determinar posibles aglomeraciones de datos. Por lo anterior, los Gráficos 3 y 4 presentan la función K y L, respectivamente, en términos teóricas y la obtenido de los datos usando tres ajustes por bordes.

**Gráfico 3:**  
**Función k teórica y con base en los datos**

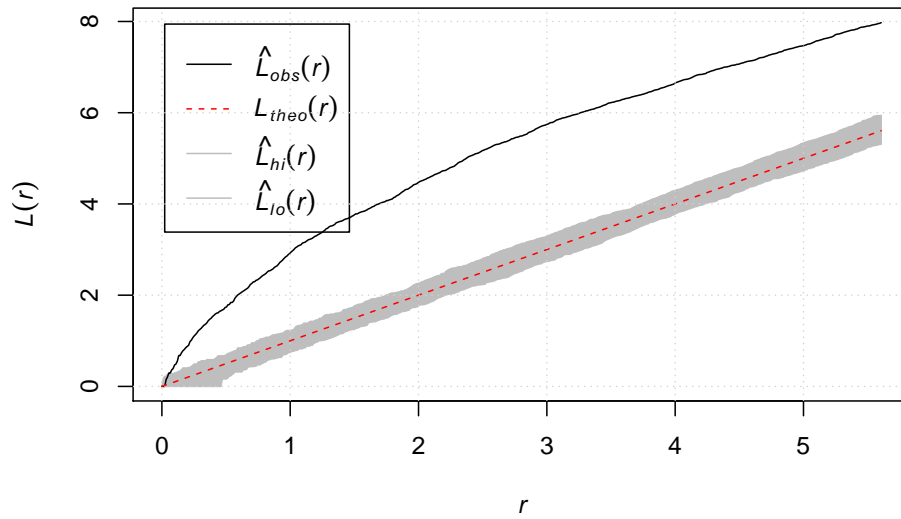


**Gráfico 4:**  
**Función L teórica y con base en los datos**



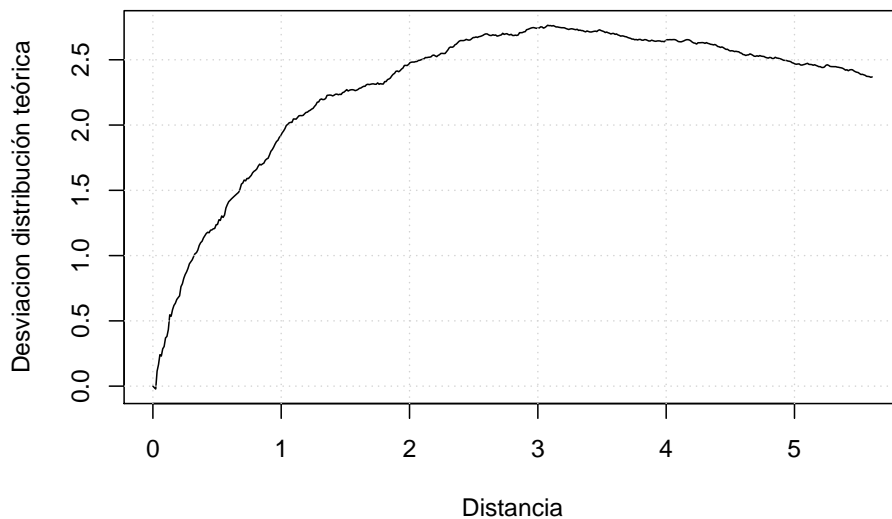
Con base en los Gráficos 3 y 4, se puede notar que, como las funciones basadas en los datos de la muestra están por encima de la teórica, entonces hay aglomeraciones en la posición de los volcanes. Para lograr tener una estimación mediante intervalo, en lugar de una puntual, en Gráfico 5 se presenta la función L teórica mediante 999 simulaciones de Monte-Carlo de datos provenientes de una Poisson.

**Gráfico 5:**  
**999 Monte Carlo envelopes con la función L**



Al igual que en los Gráficos 3 y 4, el Gráfico 5 nos ayuda a confirmar que realmente hay aglomeraciones en los datos, ya que las funciones estimadas con base en la muestra están fuera de los intervalos de las simulaciones teóricas. En términos de la forma en la cual se dan las agrupaciones, el Gráfico 6 presenta la distancia entre puntos con respecto a la desviación de estos de la distribución teórica. De este gráfico, se evidencia que a distancias cortas hay mayor agrupación y, como el crecimiento es rápido, se puede suponer que hay agrupaciones.

**Gráfico 6:**  
**Cercanía de puntos según distancia**



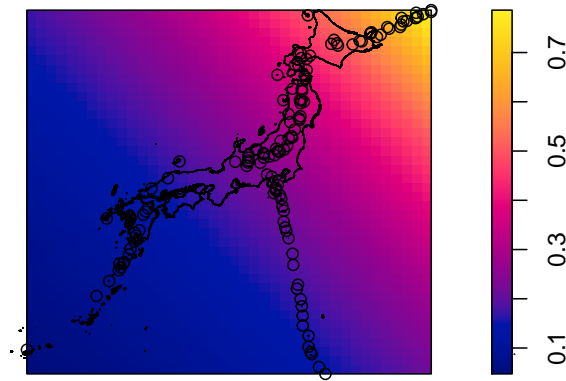
Habiendo determinado que hay agrupaciones en los datos, se desea modelar la intensidad, para lo cual se utilizó un proceso de Poisson no homogéneo (IPP) con una estimación no paramétrica usando kernels. El modelo teórico es el siguiente:

**Modelo:**

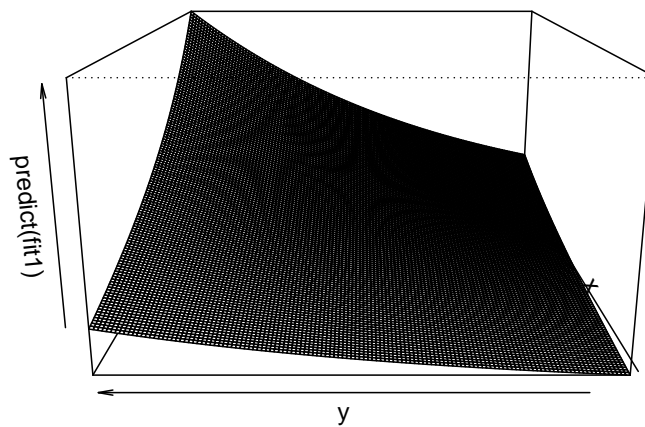
$$\lambda((x, y)) = \exp(-12.15 + 0.062x + 0.059y)$$

	Estimate	S.E.	CI95.lo	CI95.hi	Ztest	Zval
(Intercept)	-12.1475700	1.8917330	-15.8552986	-8.4398413	***	-6.421398
x	0.0620714	0.0130190	0.0365547	0.0875880	***	4.767769
y	0.0594193	0.0142656	0.0314592	0.0873794	***	4.165209

**Gráfico 7:**  
**Estimación de la intensidad del PPM lineal**



**Gráfico 8:**  
**Estimación de la intensidad del PPM lineal en 3D**



Ajustando el modelo de Poisson con intensidad log-lineal en las coordenadas cartesianas se obtuvo que no predice tan bien la intensidad, ya que, comparandolo con los resultados del Gráfico 2, solo está tomando las

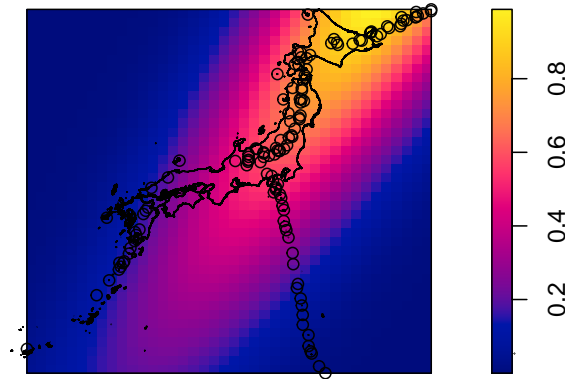
agrupaciones en la parte noreste. Ahora, ajustando el modelo con una intensidad que es log-cuadrática en las coordenadas cartesianas se obtiene que la predicción de la intensidad mejora.

**Modelo:**

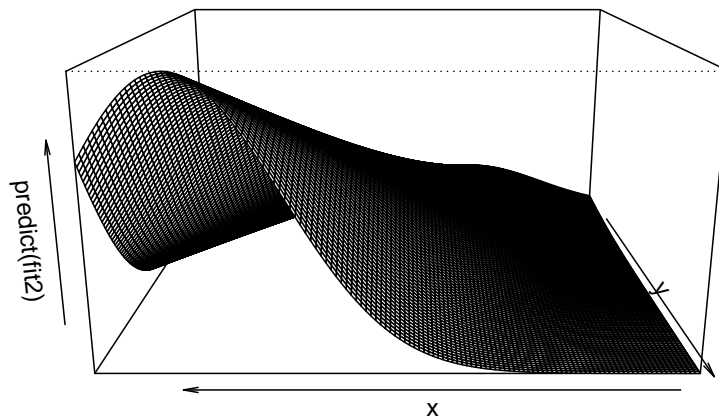
$$\lambda((x, y)) = \exp(-369.09 + 6.21x - 3.63y - 0.026x^2 + 0.032xy - 0.01y^2)$$

	Estimate	S.E.	CI95.lo	CI95.hi	Ztest	Zval
(Intercept)	-369.0873315	65.6662686	-497.7908529	-240.3838100	***	-5.620653
x	6.2064495	1.0480953	4.1522205	8.2606785	***	5.921646
y	-3.6345108	0.6346774	-4.8784556	-2.3905660	***	-5.726549
I(x^2)	-0.0264077	0.0042556	-0.0347485	-0.0180668	***	-6.205368
I(x * y)	0.0322102	0.0054783	0.0214730	0.0429474	***	5.879634
I(y^2)	-0.0109948	0.0034089	-0.0176761	-0.0043135	**	-3.225314

**Gráfico 9:**  
**Estimación de la intensidad del PPM cuadrático**



**Gráfico 10:**  
**Estimación de la intensidad del PPM cuadrático en 3D**



Para el caso del modelo Poisson con intensidad log-cuadrática con base en las coordenadas, nos da un resultado parecido al obtenido en la intensidad original, donde en la parte noreste y central del mapa estima mayor intensidad, aunque todavía no estima del todo bien las islas al sureste. Por lo mencionado anteriormente, se decidió usar el modelo de Poisson con intensidad log-cuadrática para modelar la intensidad. A continuación se presenta la matriz de covariancia de los parametros del modelo log-cuadrático:

	(Intercept)	x	y	$I(x^2)$	$I(x * y)$	$I(y^2)$
(Intercept)	4312.0588315	-68.2259961	23.1964641	0.2700430	-0.1877351	0.0410731
x	-68.2259962	1.0985037	-0.4389709	-0.0044210	0.0035744	-0.0008130
y	23.1964641	-0.4389709	0.4028154	0.0020113	-0.0033204	0.0008145
$I(x^2)$	0.2700430	-0.0044210	0.0020113	0.0000181	-0.0000168	0.0000045
$I(x * y)$	-0.1877351	0.0035744	-0.0033204	-0.0000168	0.0000300	-0.0000118
$I(y^2)$	0.0410731	-0.0008130	0.0008145	0.0000045	-0.0000118	0.0000116

Por último, al comparar el modelo log-cuadrático contra el modelo que sólo contiene el intercepto, se obtuvo que el modelo con una intensidad que es log-cuadrático es el mejor.

```
## Analysis of Deviance Table
##
## Model 1: ~1    Poisson
## Model 2: ~x + y + I(x^2) + I(x * y) + I(y^2)    Poisson
##   Npar Df Deviance   Pr(>Chi)
## 1     1
## 2     6   5    136.55 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Conclusiones

Como primera conclusión, es importante resaltar que se probó que los volcanes con erupciones durante los últimos 10.000 años en Japón no presentan una distribución espacial aleatoria y se observa una tendencia de agrupaciones de dichos puntos. Esto significa que hay un efecto de atracción entre los volcanes, donde se espera encontrar más volcanes cerca de volcanes que ya existen. Por otro lado, se logró modelar la intensidad de los volcanes mediante un modelo de Poisson no homogéneo con intensidad log-cuadrática, que tuvo mejores predicciones que el modelo con solo las coordenadas y logró predecir la intensidad de forma bastante buena.

## Referencias

- <https://www.kaggle.com/smithsonian/volcanic-eruptions>
- <https://gadm.org>
- <https://volcano.si.edu>