

Identificación de cambios en la ciclogénesis del Atlántico Norte mediante un modelo de mezclas Gaussianas

• Julio César Buendía-Espinoza •
Universidad Autónoma Chapingo, México

• Paulino Pérez-Rodríguez* • Juan Manuel González-Camacho •
• Sergio Pérez-Elizalde • Adolfo Exebio-García •
Colegio de Postgraduados, México

*Autor para correspondencia

• Michel Rosengaus Moshinsky •
Consultor independiente

• Malaquías Peña •
National Centers for Environmental Prediction, NOAA, USA

DOI: 10.24850/j-tyca-2017-04-01

Resumen

Buendía-Espinoza, J. C., Pérez-Rodríguez, P., González-Camacho, J. M., Pérez-Elizalde, S., Exebio-García, A., Rosengaus-Moshinsky, M., & Peña, M. (julio-agosto, 2017). Identificación de cambios en la ciclogénesis del Atlántico Norte mediante un modelo de mezclas Gaussianas. *Tecnología y Ciencias del Agua*, 8(4), 5-18.

Varios modelos climáticos sugieren que la frecuencia e intensidad de los ciclones tropicales cambiarán a finales del siglo XXI, como consecuencia del calentamiento global. Existen diferentes reportes científicos sobre el calentamiento global y su relación con la intensidad de los ciclones tropicales. Sin embargo, poco se ha investigado sobre el impacto del calentamiento global sobre la ciclogénesis en las diferentes cuencas oceánicas. En este trabajo se estima el número de regiones ciclogénicas y sus centroides en la cuenca oceánica del Atlántico Norte entre los periodos 1951-1975 versus 1976-2013 y 1951-1989 versus 1990-2013 mediante un modelo estadístico de mezclas finitas Gaussianas, para identificar si existen cambios significativos debidos al cambio climático. En el presente estudio, los cambios de un intervalo a otro se le atribuirán al cambio climático. La estimación de los parámetros de la función de densidad de probabilidades (fdp) de los componentes de las mezclas se hizo a través del algoritmo esperanza-maximización (EM). Las fdp se compararon mediante la distancia de Bhattacharyya y se estimó el percentil 95 a través de la técnica de remuestreo paramétrico. Los resultados obtenidos muestran que solamente hay dos regiones ciclogénicas en los dos intervalos de estudio, es decir, no hubo cambio en el número de regiones de un periodo a otro. Un segundo resultado es que existen diferencias estadísticamente significativas en la ubicación de los centroides de las regiones de génesis estimadas, sugiriendo un impacto debido al cambio climático de acuerdo con los datos analizados.

Palabras clave: modelos de mezclas, verosimilitud, regiones ciclogénicas.

Abstract

Buendía-Espinoza, J. C., Pérez-Rodríguez, P., González-Camacho, J. M., Pérez-Elizalde, S., Exebio-García, A., Rosengaus-Moshinsky, M., & Peña, M. (July-August, 2017). Identification of changes in the North Atlantic cyclogenesis using a Gaussian mixture model. *Water Technology and Sciences (in Spanish)*, 8(4), 5-18.

Several climate models suggest that the frequency and intensity of tropical cyclones will change at the end of the 21st century, as consequence of global warming. Several scientific reports have described how global warming affects the intensity of tropical cyclones. However, little research has been done on the impact of global warming on cyclogenesis in different ocean basins. In this work, the number of cyclogenetic regions and their corresponding centroids are estimated in the North Atlantic Ocean basin for the intervals 1951-1975 versus 1976-2013 and 1951-1989 versus 1990-2013 through a finite Gaussian mixture model, to determine whether significant changes exist. In this study, the change from one interval to another will be attributed to climate change. The parameter estimation of the probability density function (fdp) of the mixture components was done using the Expectation-Maximization algorithm (EM). The fdp were compared using the Bhattacharyya's distance and the 95th percentile was estimated using the technique of parametric bootstrap. The results show that there are only two cyclogenetic regions in each of the intervals that is no increase or decrease in the number of regions. A second result indicates that there are significant differences in the centroid's locations from one interval to another, suggestion an impact due to climate change according to the analyzed data.

Keywords: Mixture models, likelihood, cyclogenetic regions.

Recibido: 08/05/2016

Aceptado: 03/03/2017

Introducción

Existe un interés creciente en conocer el impacto del cambio climático en la actividad de los ciclones tropicales, pues son unos de los fenómenos climáticos más destructivos entre los fenómenos naturales que provocan mayores desastres. A medida que la temperatura global promedio de la superficie del planeta se incrementa, se espera que la intensidad, frecuencia, trayectorias, ubicación de ocurrencia y llegada a tierra de estos fenómenos se alteren por el clima actual. Walsh (2004) menciona que aunque no hay en este momento cambios perceptibles en las características de los ciclones tropicales que razonablemente podrían atribuirse al calentamiento global, las predicciones de los modelos de circulación general muestran incrementos en su intensidad máxima entre 5 y 10% para el año 2050. También indica que las regiones de formación de huracanes probablemente no cambien, y que ha habido poco consenso respecto a la modificación en el número de ciclones o trayectorias, además de la incertidumbre en las predicciones climáticas por algunas deficiencias en los modelos de circulación general; por lo tanto, si las predicciones de intensidades son correctas, sus cambios se detectarán en el Atlántico después de 2050. Con base en los registros de las mejores trayectorias del Centro de Advertencia de Tifones de la Marina Estadounidense (Joint Typhoon Warning Center, JTWC) y el Centro Nacional de Huracanes (National Hurricane Center, NHC) de la Agencia Atmosférica y Oceanográfica Nacional (National Oceanographic and Atmospheric Administration, NOAA), Emanuel (2005), Webster, Holland, Curry y Chang (2005), y más recientemente Elsner, Kossin y Jagger (2008) demostraron que la intensidad histórica de las tormentas ha aumentado tanto en el Pacífico Noroccidental (WNP) como en el Atlántico Norte. Webster *et al.* (2005) observaron un aumento considerable en los ciclones tropicales en todas las cuencas oceánicas durante los últimos 30 años para las categorías más fuertes (4 y 5 de acuerdo con la escala de Saffir-Simpson); en particular, en las regiones ciclogénicas del Pacífico Noroccidental y del

Atlántico Norte; 25 y 20%, respectivamente, de estas tormentas se presentaron en el intervalo de 1975-1989, y 41 y 25%, respectivamente, en el intervalo 1990-2004; esto significa que hubo aumentos de 16 y 5%, respectivamente. Sin embargo, los resultados obtenidos para la cuenca oceánica del Pacífico Noroccidental han sido cuestionados, pues la aparente tendencia que se observa en la intensidad de los ciclones es parte de un gran oscilación interdecadal (Chan, 2006) o de posibles errores de medición en el conjunto de datos (Knaff & Zehr, 2007). Además, Klotzbach (2006), con base en el análisis de los registros de las mejores trayectorias del periodo 1986-2006, encontró que la tendencia de la intensidad de los ciclones tropicales para la cuenca del Atlántico Norte no muestra evidencia de que haya cambiado, y la tendencia para el Pacífico Noroccidental presenta una baja considerable. También Kossin, Knapp, Vimont y Harper (2007), con base en el análisis de los registros de las mejores trayectorias, encontraron que no hay un aumento en la intensidad de los huracanes en cualquier cuenca distinta a la del Atlántico Norte en las dos décadas (1985-2005).

Respecto a la ubicación de ocurrencia (génesis) y la lisis de los ciclones tropicales, también han experimentado variaciones. Knutson *et al.* (2010) mencionan que no hay ninguna prueba concluyente de que los cambios observados en la génesis, trayectorias, duración e inundaciones de oleaje de los ciclones tropicales superen la variabilidad esperada a partir de causas naturales en los últimos 50 años; sin embargo, mencionan que hay sugerencias de cambios en la ubicación de génesis y en las trayectorias de las tormentas observadas en el océano Atlántico ofrecidas como una explicación de la falta de tendencias crecientes de la llegada a tierra de ciclones en los Estados Unidos y la Costa del Golfo. Asimismo, Kossin, Emanuel y Vecchi (2014) encontraron que los cambios observados en el cizallamiento vertical del viento y la intensidad potencial proporcionan evidencias o pruebas de que el desplazamiento global de los ciclones tropicales hacia los polos en ambos hemisferios, es decir, fuera de los trópicos, está siendo modificado por cambios ambientales sistemáticos en los últimos

30 años. También mencionan que si estos cambios ambientales continúan, un desplazamiento simultáneo concomitante hacia los polos de la latitud donde los ciclones tropicales alcanzan su intensidad máxima tendría consecuencias potencialmente profundas sobre los residentes costeros e infraestructura. Los ciclones tropicales también tienen un papel importante en el sostenimiento de los recursos hídricos regionales; de igual forma, un desplazamiento hacia los polos de las tormentas podría amenazar el abastecimiento de agua potable en algunas regiones, al tiempo que aumentan los eventos de inundación, entre otras consecuencias. Por su parte, Mori, Kuniyoshi, Nakajo, Yasuda y Mase (2013) simularon con un modelo estadístico el impacto del calentamiento global sobre los centroides de la ciclogénesis y de la ciclólisis de las diferentes regiones ciclogénicas. Concluyeron que éstos se desplazarán hacia el centro de las cuencas oceánicas y que los cambios futuros en las condiciones dinámicas y termodinámicas en los océanos influirán en la frecuencia de la génesis de los ciclones tropicales y en el cambio de localización hacia el centro de las diferentes cuencas. Cabe mencionar que las temperaturas de la superficie del mar son más cálidas en el centro del océano que en las orillas (Chan, 2007; Yokoi & Takayabu, 2009), y que un ciclón que se desarrolla en la parte central dura más tiempo. Bajo tal contexto, estos autores mencionan que los cambios en la intensidad de los ciclones tropicales están más relacionados con el desplazamiento de los centroides que con la modificación de la temperatura de la superficie del mar. Holland y Bruyere (2013) desarrollaron un Índice de Cambio Climático Antropogénico (ACCI, por sus siglas en inglés, Anthropogenic Climate Change Index) para investigar la contribución del calentamiento global a la actividad de los ciclones tropicales. Dichos autores encontraron que en todas la cuencas oceánicas la frecuencia anual de los ciclones o huracanes tropicales no se ve afectada por el cambio antropogénico. Sin embargo, los huracanes categoría 4 y 5 han aumentado a tasas de 25 a 30% por grado Celsius (°C) de calentamiento global; en

contraste, los huracanes de categorías 1 y 2 han disminuido a tasas similares.

Además de la variación en la intensidad y la génesis, las precipitaciones inducidas por los ciclones tropicales también han experimentado cambios a largo plazo con la oscilación interdecadal. Kim, Ho, Lee, Jeong y Chen (2006), y Lau, Zhou y Wu (2008) encontraron un cambio sustancial en la precipitación durante la estación de ciclones, manifestado como un aumento en eventos y cantidad de lluvia extrema tanto en el Pacífico Noroccidental como en el Atlántico Norte. Sin embargo, en lo que respecta a las relaciones entre los ciclones tropicales y las lluvias extremas, descubrieron diferencias centrales entre los dos dominios oceánicos: en el Atlántico Norte, las contribuciones de los ciclones tropicales a la precipitación extrema son mayores que en el Pacífico Noroccidental.

En este trabajo de investigación se aplicó un modelo estadístico de mezclas Gaussianas (sección 2). Por una parte, para determinar el número de regiones ciclogénicas; por otra, para determinar los cambios temporales y espaciales de los centroides de dichas regiones en la cuenca oceánica del Atlántico Norte para dos periodos (1951-1975 *versus* 1976-2013 y 1951-1989 *versus* 1990-2013). Se utilizaron los puntos de inicio de los datos (latitud y longitud) de las “mejores trayectorias” o IBTrACS (Ncdc.noaa.gov, 2017). Se determinaron las funciones de densidad de probabilidades (fdp) de las regiones de génesis a través del algoritmo esperanza-maximización (EM), las cuales se evaluaron y compararon para verificar los cambios espacio-temporales. En la sección 3 se muestran los resultados, y su relación e interpretación con el fenómeno natural en estudio. Finalmente, en la sección 4, se concluye con una breve discusión.

Materiales y métodos

Descripción y alcance geográfico de la base de datos de ciclones tropicales

El análisis de los datos fue de 1951-2013 (ver figura 1), dividido en dos periodos: 1951-1975

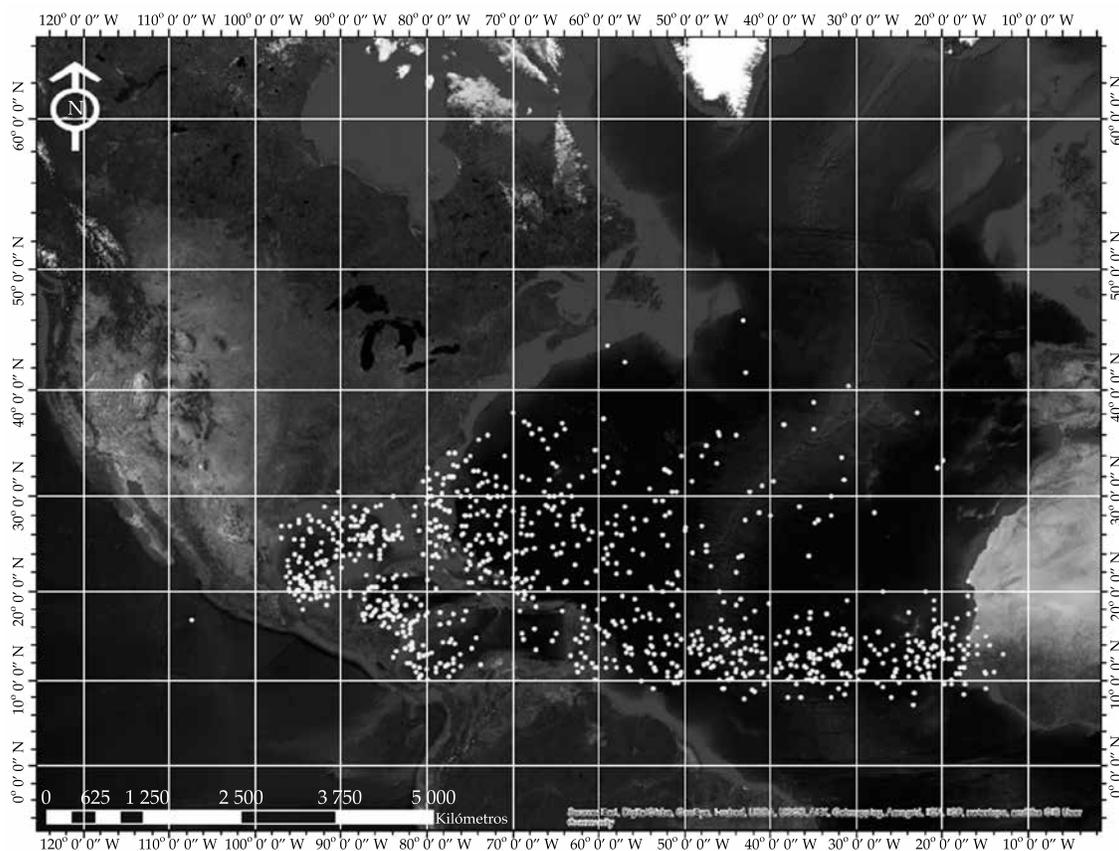


Figura 1. Distribución espacial de los puntos de ubicación de ocurrencia de los ciclones tropicales en la región del Atlántico Norte desde el año 1951 hasta 2013. Fuente: elaboración propia con base en los datos del IBTrACS.

versus 1976-2013 y 1951-1989 versus 1990-2013. El primero se determinó con base en el calentamiento ocurrido durante el siglo XX en los intervalos de 1928-1975 y 1976-2013 (ver figura 2). Cabe mencionar que se tomó como valor de inicio 1950, porque a partir de este año fue usado de manera rutinaria el reconocimiento aéreo para supervisar los ciclones tropicales. Esto significa que la información sobre ciclones medidos antes de esta fecha es menos confiable (Vecchi & Knutson, 2008, citados por Villarini, Vecchi, & Smith, 2011).

Respecto al segundo periodo, se determinó de acuerdo con Webster *et al.* (2005), quienes observaron un aumento considerable en los ciclones tropicales en todas las cuencas oceánicas durante los últimos 30 años para las

categorías más fuertes (4 y 5 con base en la escala de Saffir-Simpson) debido al incremento de la temperatura de la superficie marina (TSM). Cabe mencionar que dicho intervalo de tiempo lo dividieron en dos periodos: 1975-1989 versus 1990-2004.

Modelo de mezclas Gaussianas (MMG)

Bajo el modelo de mezclas finitas de distribuciones normales, ajustado en este estudio, cada punto de ubicación de ocurrencia $x_i = (\text{latitud}_i, \text{longitud}_i)$ de los ciclones tropicales se considera que proviene de una súper población, la cual es una mezcla de un número finito K de poblaciones con proporciones o pesos π_1, \dots, π_K respectivamente, donde:

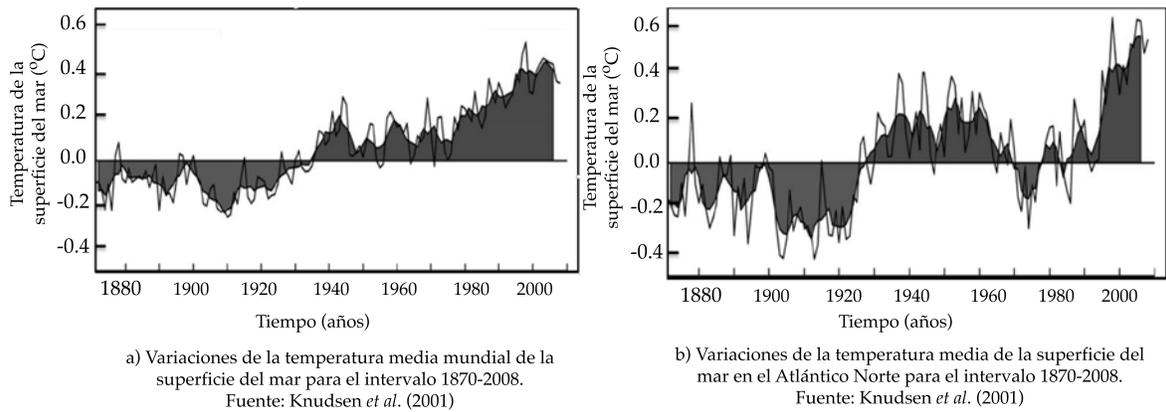


Figura 2. Variaciones de la temperatura media de la superficie del mar global y del Atlántico Norte.

$$\sum_{j=1}^K \pi_j = 1; \quad \pi_j \geq 0 \quad (1)$$

La función de densidad probabilidad (fdp) de la variable aleatoria X (de dimensión d) en la forma de mezclas finitas es:

$$p(x, \phi) = \sum_{j=1}^K \pi_j \cdot p(x; \theta_j) = \sum_{j=1}^K \pi_j \cdot \frac{1}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} e^{-\frac{1}{2}[(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)]} \quad (2)$$

donde ϕ denota el vector de parámetros de la mezcla; $p(x; \theta_j)$ es la función de densidad de probabilidades correspondientes al K -ésimo componente; π_j es el peso de la mezcla; y θ_j denota el vector de todos los parámetros desconocidos de los elementos de los vectores de las medias μ_j y elementos distintos de matrices de covarianzas Σ_j para $j = 1, \dots, K$, que pertenecen a algún espacio de parámetros Θ .

a) Estimación de los parámetros utilizando el algoritmo EM

Existen varios procedimientos para determinar los parámetros de un modelo de mezclas Gaussianas (MMG) de un conjunto de datos (McLachlan & Basford, 1988). Sin embargo, el

método más popular es el de máxima verosimilitud (ML, por sus siglas en inglés). En este trabajo, la estimación de los parámetros se hizo utilizando técnicas de máxima verosimilitud vía el algoritmo EM (Dempster, Laird, & Rubin, 1977; Redner & Walker, 1984; Maitra & Melnykov, 2010).

El algoritmo EM para mezclas Gaussianas es un proceso iterativo que consiste de dos etapas: a) esperanza (E) y b) maximización (M). En el caso de componentes Gaussianos, la densidad de la mezcla contiene los siguientes parámetros: π_j , μ_j y Σ_j donde $j = 1, \dots, K$. La log-verosimilitud condicionada completa esperada para el conjunto de datos, conocida como función Q , es igual a:

$$Q(\phi, x_1, \dots, x_n) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^K \pi_{ij} \left\{ \ln |\Sigma_j| + (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right\} + \sum_{i=1}^n \sum_{j=1}^K \pi_{ij} \ln(\pi_j) - \frac{dn}{2} \ln(2\pi) \quad (3)$$

El paso-E consiste en la actualización de las probabilidades de pertenencia $\pi_{ij}^{(t+1)}$, dada la estimación de los parámetros actuales ϕ^t :

$$\pi_{ij}^{(t+1)} = \frac{\pi_{ij}^{(t)} p(x_i; \mu_j^{(t)}, \Sigma_j^{(t)})}{\sum_{j=1}^K \pi_{ij}^{(t)} p(x_i; \mu_j^{(t)}, \Sigma_j^{(t)})} \quad (4)$$

El paso-M tiene una solución analítica:

$$\pi_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \pi_{ij}^{(t+1)}$$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^n \pi_{ij}^{(t+1)} x_i}{\sum_{i=1}^n \pi_{ij}^{(t+1)}}$$

y

$$\sum_j^{(t+1)} = \frac{\sum_{i=1}^n \pi_{ij}^{(t+1)} (x_i - \mu_j^{(t+1)}) (x_i - \mu_j^{(t+1)})^T}{\sum_{i=1}^n \pi_{ij}^{(t+1)}} \quad (5)$$

El algoritmo EM comienza con una estimación inicial del valor de los parámetros ϕ , llamada $\phi^{(t)}$. Luego, mediante las ecuaciones (4) y (5), se estiman los valores de los nuevos parámetros, llamado $\phi^{(t+1)}$. El proceso se repite hasta que la diferencia entre dos evaluaciones sucesivas de la log-verosimilitud sea menor que una ϵ dada ϵ , es decir:

$$\left| l(\phi^{(t+1)} | x) - l(\phi^{(t)} | x) \right| < \epsilon \quad (6)$$

en este trabajo se fijó $\epsilon = 1 \times 10^{-3}$. Este resultado depende de la selección de los parámetros iniciales (Seidel, Mosler, & Alker, 2000).

b) Inicialización en el algoritmo EM

El algoritmo EM es un procedimiento iterativo de maximización que depende del valor inicial de los parámetros, pues la función de verosimilitud puede tener máximos locales (McLachlan & Peel, 2000). Por lo tanto, una buena inicialización es crucial para encontrar los estimadores de máxima verosimilitud.

Se han sugerido diferentes procedimientos de inicialización en la literatura (Figueiredo & Jain, 2000; Maitra, 2009); sin embargo, ningún método supera a los demás. En este trabajo se utilizó el procedimiento de Fraley, Raftery, Murphy y Scrucca (2012), implementado en la biblioteca de funciones de R (R Core Team, 2016)

mclust, para encontrar los valores iniciales de los parámetros que permiten obtener el valor máximo en el marco de mezclas Gaussianas multivariadas.

Identificación del número óptimo de componentes o grupos

Hay una vasta lista de literatura dedicada al tema de la elección de K (número de componentes). McLachlan y Peel (2000) proporcionan una interpretación detallada de los diferentes enfoques disponibles para abordar este problema. La mayoría de los métodos destinados a la estimación de K se divide por lo general en dos categorías: modelos basados en el principio de la parsimonia y modelos basados en procedimientos de prueba, ambos sustentados en la función de log-verosimilitud. Sin embargo, en este estudio, K se determinó mediante un método heurístico, conocido como partición alrededor de los medoides (PAM, por sus siglas en inglés, Partitioning Around Medoids).

El algoritmo de la PAM se basa en la formación de K particiones u objetos representativos (medoides) de n observaciones de un conjunto de datos. Un medoide se define como la observación de un agrupamiento, cuya diferencia promedio, con respecto a todas las observaciones en el grupo, es mínima. Se eligen aleatoriamente K medoides de un conjunto de datos. El medoide, que representa un grupo, se ubica en el centro del grupo. Los objetos restantes se agrupan con el medoide al que son más similares, basándose en la distancia entre el objeto y el medoide. La estrategia, entonces, es reemplazar uno de los medoides por los no medoides, siempre y cuando la calidad del agrupamiento mejore. Esta calidad es estimada usando una función de costo que mide el promedio de disimilaridad o diferencia entre un objeto, y el medoide de su grupo (Kaufman & Rousseeuw, 2005).

El método PAM genera una gráfica, conocida como gráfica de "siluetas". Para cada observación se muestra una medida que indica la calidad de la clasificación. Valores cercanos a 1 indican que la observación está bien clasificada

en su grupo; valores cercanos a 0 significan que la observación podría pertenecer a otro grupo; en valores cercanos a -1, la observación está pobremente clasificada. En la gráfica hay una medida resumen denominada ancho promedio de silueta o coeficiente de silueta, que se interpreta de acuerdo con el cuadro 1.

Medida de la distancia entre MMG

De acuerdo con Sfikas, Constantinopoulos, Likas y Galatsanos (2005), y Fukunaga (1990), la distancia de Bhattacharyya puede ser utilizada para comparar fdp de modelos de mezclas Gaussianas y además tiene una expresión en forma cerrada. Esta distancia se utilizó para medir la distancia entre los grupos de los ciclones tropicales para los dos intervalos de estudio y se define como sigue:

$$B(f, g) = \frac{1}{8} (\mu_f - \mu_g)^T \left(\frac{\Sigma_f + \Sigma_g}{2} \right)^{-1} (\mu_f - \mu_g) + \frac{1}{2} \ln \left[\frac{|\Sigma_f + \Sigma_g|}{|\Sigma_f| |\Sigma_g|} \right] \tag{7}$$

donde μ_f , Σ_f y μ_g , Σ_g corresponden a los vectores de medias, y las matrices de varianzas y covarianzas de los núcleos de las densidades Gaussianas, respectivamente.

Estadístico de prueba para la comparación de las fdp del MMG

Se aplicó la técnica paramétrica de remuestreo para obtener el percentil 95 de la distribución

empírica del estadístico de prueba, que es la distancia de Bhattacharyya entre los componentes de los modelos de mezclas Gaussianas, para verificar si existen diferencias estadísticamente significativas entre ellos (Engel, 2010).

El método de remuestreo (Efron, 1979; Efron & Tibshirani, 1993) se usa para encontrar intervalos de confianza en situaciones donde es imposible obtener de manera analítica la distribución muestral del estimador (para mayor detalle, ver DiCiccio & Tibshirani, 1987; Hall, 1988). Esta técnica se sustenta teóricamente en dos consideraciones: 1) la función de distribución verdadera $F(x)$ se estima mediante la función de distribución empírica $\hat{F}(x)$ de acuerdo con el teorema Glivenko-Cantelli, que muestra que $\hat{F}(x) \xrightarrow{p} F(x)$ conforme $n \rightarrow \infty$ (Bickel & Freedman, 1981); y 2) de acuerdo con la propiedad de consistencia, la $F(\hat{\theta})$ de una muestra dada puede aproximarse mediante la distribución muestral del remuestreo $\hat{F}^*(\hat{\theta}^*)$ cuando el número de remuestreos es grande y puede también aproximarse $\hat{F}(x)$ a $F(x)$. Bajo estos supuestos, Babu y Singh (1983) demuestran que $\hat{F}^*(\hat{\theta}^*) \approx \hat{F}(\hat{\theta})$, cuando el número de remuestreos es suficientemente grande.

Procedimiento de remuestreo

1. Se genera una variable aleatoria $U \sim$ Uniforme (0,1).
2. Si $U \in \left[\sum_{j=1}^K \pi_j, \sum_{j=1}^{K+1} \pi_{j+1} \right]$, donde π_j corresponde a la probabilidad del j -ésimo componente del modelo de mezclas, entonces se generan variables aleatorias a partir de la distribución del j -ésimo componente. Dicho componente tiene una distribución normal bivariada, obtenida mediante el algoritmo EM.

Cuadro 1. Rango del coeficiente en silueta.

Rango del coeficiente de silueta	Interpretación
0.71-1.00	Estructura fuerte
0.51-0.70	Estructura razonable
0.26-0.50	Estructura débil y podría ser artificial
< 0.25	Ninguna estructura sustancial

Fuente: Kaufman y Rousseeuw (2005).

3. Se repiten los pasos 1 y 2 hasta que se tenga la cantidad deseada de muestras de la mezcla de la distribución.
 4. Se definen las variables aleatorias $X_i \sim N_2(\mu, \Sigma)$, $i = 1, \dots, n$ obtenidas en el paso anterior como muestra aleatoria.
 5. Se obtiene una muestra bootstrap $X^* = \{x_1^*, \dots, x_n^*\}$ muestreando aleatoriamente con reemplazo n veces los datos originales x_1, \dots, x_n ; el tamaño de la muestra aleatoria es el mismo que el de la muestra de remuestreo, y las X_i^* tienen probabilidad n^{-1} , siendo igual en cada una de las X_i .
 6. Se calcula el estadístico $\hat{d}_B(f, g)$ (distancia de Bhattacharyya) de este remuestreo, produciendo $\hat{d}_B^*(f, g)$.
 7. Se repite el paso 4 B veces. La ley de los números grandes indica que si B es lo suficientemente grande, se obtendrá una buena aproximación a la densidad verdadera $d_B(f, g)$. En este trabajo se fijó $B = 1\ 000$.
 8. Se construye la distribución de probabilidad de las B $\hat{d}_B^*(f, g)$, asignando probabilidad B^{-1} a cada $\hat{d}_B^*(f, g)$. Esta es la estimación de la distribución $\hat{d}_B(f, g)$, $\hat{F}^*(\hat{d}_B^*(f, g))$.
4. Se calcula el estadístico $\hat{d}_B(f, g)$ para cada muestra de remuestreo, y con ellos se construye la función de distribución empírica de $\hat{d}_B(f, g)$.
 5. La prueba de hipótesis de la distancia de las fdp de los grupos es $H_0: \hat{d}_B(f, g)$ con $\alpha = 0.05$; esto equivale a obtener el percentil 95, y rechazar H_0 si $\hat{d}_B(f, g)$ es mayor que éste.

Resultados

Estimación de la función de densidad de probabilidad

De acuerdo con la figura 3, el número de grupos por intervalo (1951-1975 *versus* 1976-2013 y 1951-1989 *versus* 1990-2013) es dos (es decir, $K = 2$ componentes). Su ancho promedio de silueta fue 0.585 para ambos intervalos (ver figura 4), lo cual significa que tienen una estructura razonable (ver cuadro 1).

Los valores de los parámetros iniciales para cada uno de los componentes de las mezclas ($K = 2$) se determinaron mediante el algoritmo implementado por Fraley *et al.* (2012). El vector de parámetros ϕ se estimó iterativamente mediante el algoritmo EM. La función de densidad estimada para cada uno de los modelos de mezclas Gaussianas se ajusta a la distribución espacial de los puntos de ubicación de ocurrencia de los ciclones tropicales en ambos intervalos (figura 5). Los puntos que están muy alejados de los centroides exhiben un comportamiento atípico desde el punto de vista estadístico, pero dado que se sabe que son datos reales, se decidió mantenerlos como muestras legítimas al momento de realizar los análisis. Otra opción para manejo de este tipo de datos es el uso de mezclas con colas pesadas, como por ejemplo la distribución t -bivariada.

Además, en la figura 6 se muestra la ubicación de los centroides de cada uno de los grupos para ambos periodos de estudio, así como su función de densidad estimada por periodo. Los puntos negros corresponden a los datos. Los dos grupos en los dos periodos de estudio aparentemente indican que tienen la misma orientación.

Prueba de hipótesis a partir del remuestreo

La técnica de remuestreo permite realizar la prueba de hipótesis de similitud de las fdp de los grupos de los modelos de mezclas Gaussianas entre los diferentes intervalos, es decir, $H_0: d_B(f, g) = 0$. La regla de decisión es rechazar H_0 si $d_B(f, g)$ es grande. El procedimiento consiste en:

1. De los datos originales de la muestra se obtienen los estimadores (π, θ) de los componentes del modelo de mezclas normal bivariado mediante el algoritmo EM.
2. Mediante el procedimiento de remuestreo descrito anteriormente se obtiene $d_B(f, g)$ bajo H_0 (es decir, $\hat{d}_B(f, g)_{H_0}$), realizando 1 000 muestras de remuestreo de la distribución normal bivariada bajo $H_0: d_B(f, g) = 0$.
3. Cada remuestreo deberá ser del mismo tamaño que el de la muestra inicial.

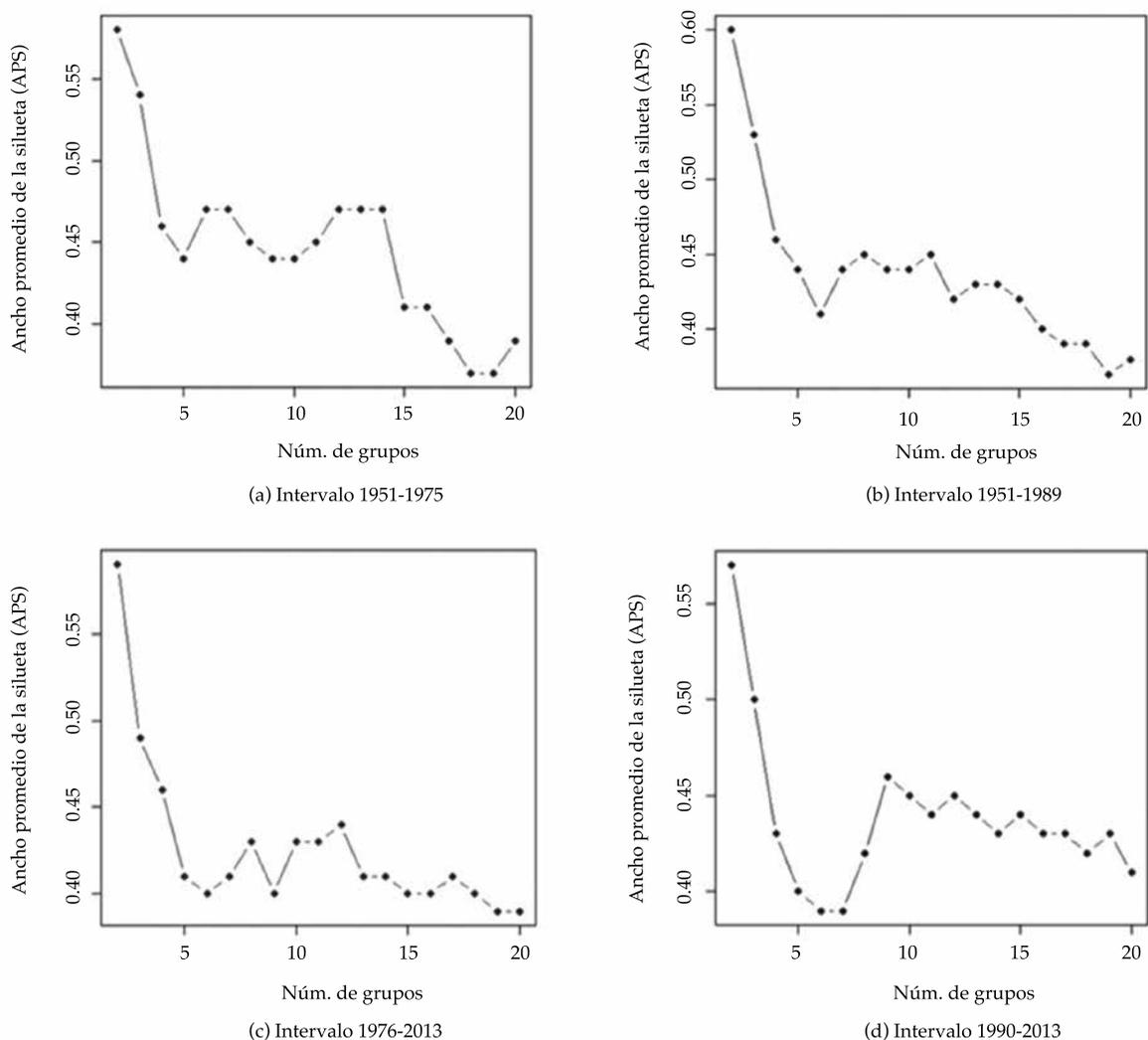


Figura 3. La gráfica número de grupos *versus* amplitud de silueta. Esta gráfica muestra que a medida que aumenta el número de grupos por intervalo disminuye el APS en las cuatro gráficas, pero disminuye la consistencia en su estructura de acuerdo con el cuadro 1.

Prueba de remuestreo no paramétrica para comparar las fdp de los MMG

La obtención del percentil 95 de la distribución empírica del estadístico $\hat{d}_B(f, g)$ equivale a rechazar H_0 si $\hat{d}_B(f, g) > q_{0.95}$. Por lo que con $\alpha = 0.05$ y $\hat{d}_B(f, g) = 0.081$ se rechazó la hipótesis nula, indicando que existen diferencias estadísticamente significativas entre las fdp de las mezclas.

Discusión y conclusiones

Con base en los resultados obtenidos, en cada uno de los periodos de estudio se encontraron dos regiones de génesis de ciclones tropicales en la cuenca oceánica del Atlántico Norte; esto significa que no existen indicios de que se estén formando nuevas regiones ciclogénicas debido al cambio climático. Por lo tanto, se concluye que solamente hay dos regiones que

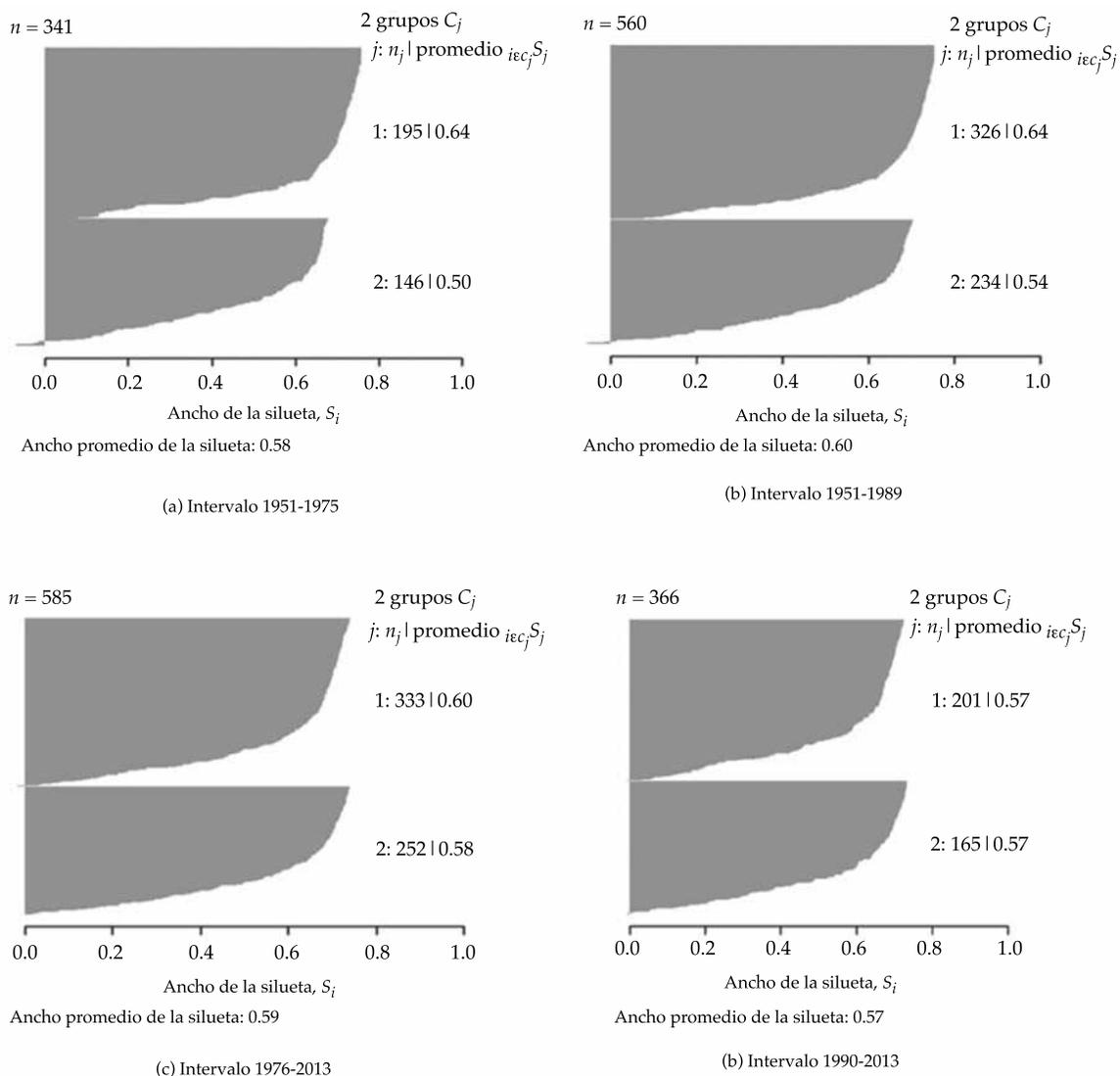


Figura 4. El Ancho promedio de la silueta, de manera general, indica que hay una buena estructura en los grupos elegidos, con la mayoría de las observaciones que parecen pertenecer a la agrupación en que están de acuerdo con el cuadro 1.

generan los ciclones tropicales desde 1951 hasta el 2013.

Los centroides de las dos regiones ciclogénicas ubicadas en la cuenca oceánica del Atlántico Norte han experimentado cambios en su localización en el intervalo de estudio; es decir, se han movido hacia el centro y hacia los polos de la cuenca, lo cual coincide con Knutson *et al.* (2010) y Kossin *et al.* (2014), quienes mencionan

que los cambios ambientales sistemáticos en los últimos 50 años han provocado que los ciclones tropicales se desplacen hacia los polos en ambos hemisferios, esto es, fuera de los trópicos. Dichos desplazamientos encontrados en este trabajo también coinciden con las proyecciones estadísticas hechas por Mori *et al.* (2013) mediante datos generados a través de modelos de circulación general para finales del siglo XXI

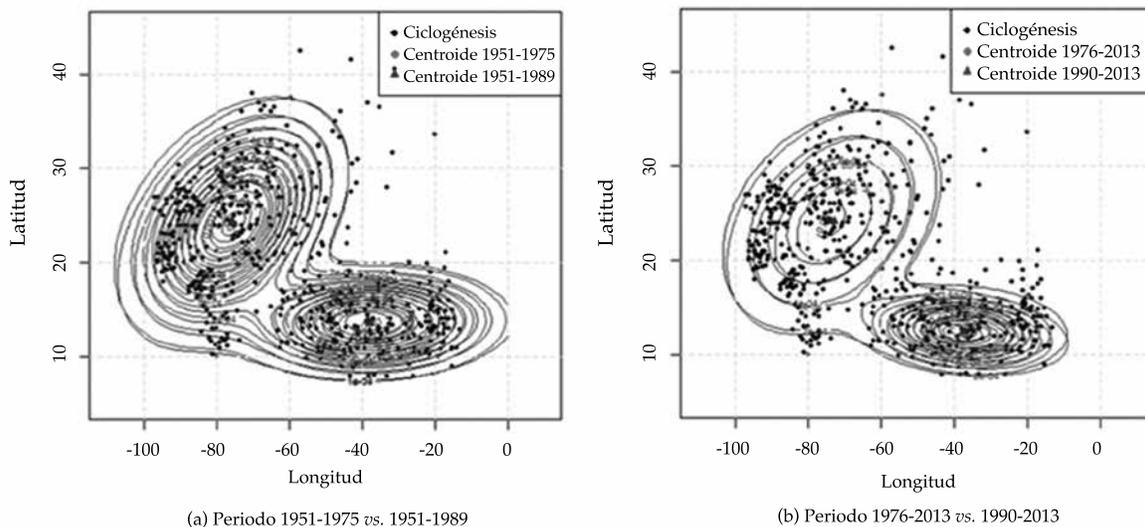


Figura 5. Función de densidad estimada por intervalo; a) muestra la marginal de la longitud y la latitud para los intervalos 1951-1975 y 1951-1989; b) indica la marginal de la longitud y la latitud, pero para los intervalos 1976-2013 y 1990-2013. Los puntos negros son los puntos de los datos. Los dos grupos en los dos periodos de estudio aparentemente indican que tienen la misma orientación.

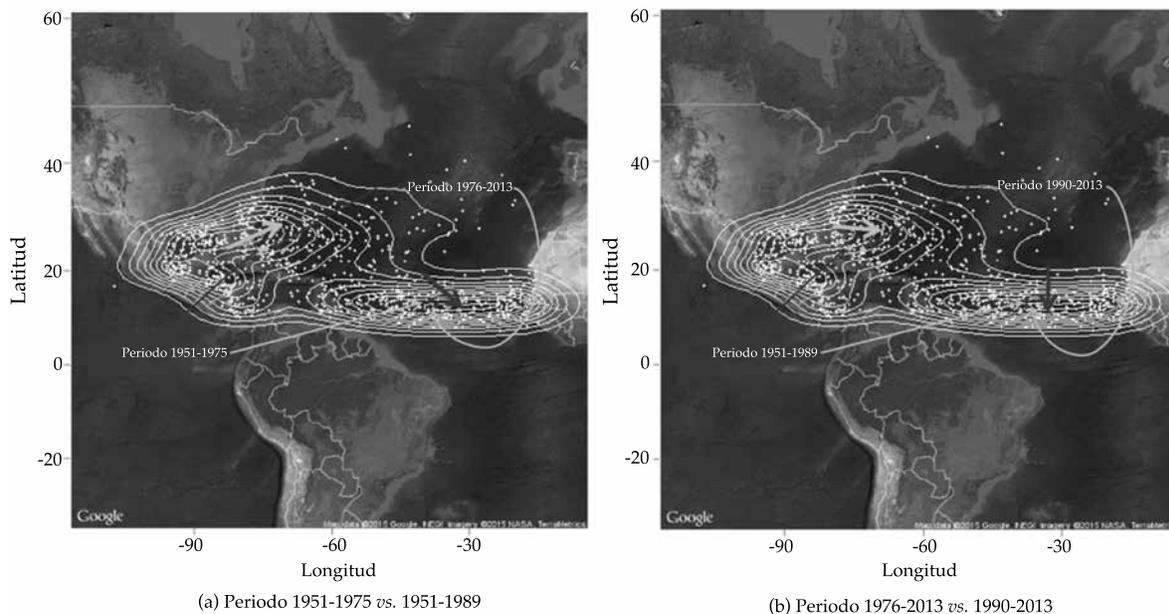


Figura 6. Regiones de génesis de los ciclones tropicales y sus centroides aparentemente se han movido hacia el centro y hacia los polos de la cuenca oceánica del Atlántico Norte en los dos intervalos de estudio.

respecto al impacto del cambio climático sobre el desplazamiento de las regiones de génesis de los ciclones tropicales.

En específico, el desplazamiento del centroide de la región de génesis del Golfo de México del periodo 1951-1975 *versus* 1951-1989 fue de 1.1942 grados al este y 0.1279 al norte (ver panel a) de la figura 6); mientras que el centroide de la misma región del periodo 1976-2013 *versus* 1990-2013 se desplazó 3.0021 grados al este y 0.3067 al sur (ver panel b) de la figura 6). Respecto al desplazamiento del centroide de la región del Atlántico-tropical del periodo 1951-1975 *versus* 1951-1989, éste fue de 1.25085 grados al este y 1.18242 al sur, en tanto que el desplazamiento del centroide de la misma región para el periodo 1975-2013 *versus* 1990-2013 fue de 0.06608 grados al este y 1.15809 al sur.

Además, el hecho de que los centroides de las regiones ciclogénicas se desplacen hacia el noreste y sur de la cuenca oceánica provocaría que la precipitación inducida por los ciclones tropicales sufriría cambios tanto en distribución espacial como en eventos y cantidad, tal y como lo mencionan Kim *et al.* (2006) y Lau *et al.* (2008). Esto significa que en la región central del continente americano, zona donde se ubica México, su precipitación disminuiría ligeramente, como indican Houghton *et al.* (2001) en sus proyecciones de precipitación para el siglo XXI debidas al cambio climático.

Chan (2007), y Yokoi y Takayabu (2009) indican que una consecuencia importante que se generaría por el desplazamiento de los centroides de la cuenca es que los ciclones tropicales durarán más tiempo debido a que se generarán más hacia el centro de la cuenca oceánica; las temperaturas de la superficie del mar en el centro son por lo general más cálidas que en las regiones externas. El alcance de nuestro estudio es limitado, ya que no es posible saber si el cambio en los puntos de ciclogénesis impactará sobre la duración de los mismos o los cambios en los puntos de llegada a tierra.

Referencias

- Babu, G. J., & Singh, K. (1983). Inference on means using the bootstrap. *Annals of Statistics*, 11(3), 999-1003.
- Bickel, O., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9(6), 1196-1217.
- Chan, J. C. L. (2006). Comment on changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, 311, 1713b.
- Chan, J. C. L. (2007). Inter-annual variations of intense typhoon activity. *Tellus*, 59(A)(4), 455-460.
- Dempster, A. P., Laird, N., & Rubin, D. (1977). Maximum likelihood for incomplete data via The EM Algorithm. *Journal of the Royal Statistical Society*, 39(1), 1-38.
- DiCiccio, T., & Tibshirani, R. (1987). Bootstrap confidence intervals and bootstrap approximations. *Journal of American Statistical Association*, 82(397), 163-170.
- Efron, B. (1979). Bootstrap method: Another look at the jackknife. *Annals of Statistics*, 7(1), 1-26.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, London: Chapman and Hall.
- Elsner, J. B., Kossin, J. P., & Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, 455, 92-95.
- Emanuel, K. (2005). Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, 436, 686-688.
- Engel, J. (2010). *On teaching bootstrap confidence intervals*. ICOTS8. Consulted at http://new.censusatschool.org.nz/wp-content/uploads/2012/09/ICOTS8_4B2_ENGEL.pdf.
- Figueiredo, M. A. T., & Jain, A. K. (2000). Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3), 381-396.
- Fraley, C., Raftery, A. E., Murphy, T. B., & Scrucca, L. (2012). *M-CLUST Version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation*. Technical Report No. 597. Washington, DC: Department of Statistics, University of Washington.
- Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (2nd ed.). New York: Academic Press Professional, Inc.
- Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, 16, 3, 927-953.
- Holland, G., & Bruyere, C. L. (2013). Recent intense hurricane response to global climate change. *Climate Dynamics*, 42, 617-627.
- Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., Der Linden, P. J. V., Dai, X., Maskell, K., & Johnson, C. A. (2001). *Climate Change 2001: The scientific basis is the most comprehensive and up-to-date scientific assessment of past, present and future climate change* (pp. 1-83). Cambridge/ New York: Cambridge University Press.

- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.
- Kim, J. H., Ho, C. H., Lee, M. H., Jeong, J. H., & Chen, D. (2006). Large increase in heavy rainfall associated with tropical cyclone landfalls in Korea after the late 1970s. *Geophysical Research Letters*, 33:L18706, 1-5.
- Klotzbach, P. J. (2006). Trends in global tropical cyclone activity over the past twenty years (1986-2005). *Geophysical Research Letters*, 33:L10805, 1-4.
- Knaff, J. A., & Zehr, R. M. (2007). Reexamination of tropical cyclone wind-pressure relationships. *Weather and Forecasting. American Meteorological Society*, 22, 71-88.
- Knudsen, M. F., Seidenkrantz, M.-S., Jacobsen, B. H., & Kuijpers, A. (2011). Tracking the Atlantic multidecadal oscillation through the last 8,000 years. *Nature Communications*, 2(178), 1-8.
- Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J. P., Srivastava, A. K., & Sugi, M. (2010). Tropical cyclones and climate change. *Nature Geoscience*, 3, 157-163.
- Kossin, J. P., Knapp, K. R., Vimont D. J., Murnane R. J., & Harper, B. A. (2007). A globally consistent reanalysis of hurricane variability and trends. *Geophysical Research Letters*, 34:L04805, 1-6
- Kossin, J. P., Emanuel, K. A., & Vecchi, G. A. (2014). The poleward migration of the location of tropical cyclone maximum intensity. *Nature*, 509, 349-352.
- Lau, K. M., Zhou, Y. P., & Wu, H. T. (2008). Have tropical cyclones been feeding more extreme rainfall? *Journal of Geophysical Research*, 113, 1-12.
- Maitra, R. (2009). Initializing partition-optimization algorithms. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 6(1), 144-157.
- Maitra, R., & Melnykov V. (2010). Simulating data to study performance of finite mixture modeling and clustering algorithms. *Journal of Computational and Graphical Statistics*, 19(2), 354-376, DOI: 10.1198/jcgs.2009.08054.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering*. New York: Marcel Dekker.
- Mori, N., Kuniyoshi, S., Nakajo, S., Yasuda, T., & Mase, H. (2013). Projection of future tropical cyclone activity and extreme waves (pp. 1229-1240). *Proceedings of Coastal Dynamics 2013*, Arcachon, France # 111.
- Ncdc.noaa.gov (2017). *IBTrACS-International Best Track Archive for Climate Stewardship*. Accessed 19 Jan. 2017. Recovered from: <https://www.ncdc.noaa.gov/ibtracs/index.php?name=ibtracs-data-access>.
- Redner, R. A., & Walker, H. F. (1984). Mixtures densities, maximum likelihood and the EM Algorithm. *SIAM Review*, 26(2), 195-239.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Recovered from: <https://www.R-project.org/>.
- Seidel, W., Mosler, K., & Alker, M. (2000). A cautionary note on likelihood ratio tests in mixture models. *Annals of the Institute of Statistical Mathematics*, 58(3), 481-487.
- Sfikas, G., Constantinopoulos, C., Likas, A., & Galatsanos, N. (2005). *An analytic distance metric for Gaussian mixture models with application in image retrieval* (pp. 835-840). Berlin/ Heidelberg: Springer-Verlag.
- Villarini, G., Vecchi, G. A., & Smith, J. A. (2011). *U.S. Landfalling and North Atlantic Hurricanes: Statistical modeling of their frequencies and ratios* (pp. 1-72). Unprinted.
- Walsh, K. (2004). Tropical cyclones and climate change: Unresolved issues. *Climate Research*, 27, 77-83.
- Webster, P. J., Holland, G. J., Curry, J. A., & Chang, H. R. (2005). Changes in tropical cyclone number, duration, and intensity in warming environment. *Science*, 309, 1844-1846.
- Yokoi, S., & Takayabu, Y. N. (2009). Multi-model projection of global warming impact on tropical cyclone genesis frequency over the Western North Pacific. *Journal of the Meteorological Society of Japan*, 87(3), 525-538.

Dirección institucional de los autores

Dr. Julio César Buendía-Espinoza

Universidad Autónoma Chapingo
Departamento de Suelos
Km 38.5 de la carretera federal México-Texcoco
56230 Chapingo, Texcoco, Estado de México, MÉXICO
Teléfono: +52 (595) 9521 500, ext. 6207 y 6281
jcesar732000@yahoo.com.mx

Dr. Paulino Pérez-Rodríguez
Dr. Juan Manuel González-Camacho
Dr. Sergio Pérez-Elizalde

Posgrado en Socioeconomía, Estadística e Informática
Colegio de Postgraduados, Campus Montecillo
Km 36.5 de la carretera federal México-Texcoco,
56230 Montecillo, Texcoco, Estado de México, MÉXICO
Teléfono: + 52 (55) 5804 5900, ext. 1431
perpdgo@colpos.mx
jmgc@colpos.mx
sergiop@colpos.mx

Dr. Adolfo Exebio García

Colegio de Postgraduados, Campus Montecillo
Posgrado en Hidrociencias
Km 36.5 de la carretera federal México-Texcoco
56230 Montecillo, Texcoco Estado de México, MÉXICO
exebio@colpos.mx

Dr. Michel Rosengaus Moshinsky

Consultor independiente

Dr. Malaquías Peña

National Oceanic and Atmospheric Administration
Environmental Modeling Center (EMC)
5830 University Research Court
College Park, MD 20740, Washington, D.C., UNITED STATES
Teléfono: +001 (301) 6833 711
malaquias.pena.mendez@noaa.gov