# REGIONALIZACIÓN DE LA VARIACIÓN TEMPORAL DEL FACTOR DE TURBIDEZ LINKE $T_L$ EN MÉXICO A PARTIR DE ALGORITMOS DE MACHINE LEARNING

**Salinas-González J.D.*, García-Hernández A.*, Riveros-Rosas D.**, Moreno-Chávez G.*, González-Cabrera A.E.**, Zarzalejo L.F.*****

* Maestría en procesamiento de la información, Universidad Autónoma de Zacatecas, Carretera Zacatecas-Guadalajara Km. 6. Ejido la Escondida, Zacatecas, 98160, (Zacatecas) México, alegarcia@uaz.edu.mx
**Instituto de Geofísica, Universidad Nacional Autónoma de México, Circuito de la investigación científica s/n, Cd. México, 04510, (Coyoacán) México.
*** Centro de Investigaciones Energéticas Medio Ambientales y Tecnológicas, Av. Complutense 40, 28040, (Madrid) España.

https://doi.org/10.34637/cies2020.1.6108

RESUMEN

El agrupamiento de áreas geográficas, por medio de análisis de *clusters*, es una tarea que permite identificar regiones de acuerdo a los comportamientos de las variables geoclimáticas. En este trabajo de investigación se agrupó al $T_L$ Linke en regiones similares de la república mexicana de acuerdo con su varianza temporal anual del año 2015. Mediante algoritmos y técnicas de *Machine Learning* (aprendizaje automático), como son las mezclas de gaussianos mixtos y análisis de componentes principales, se han obtenido dos mapas, de 10 y 9 regiones que permiten identificar de una manera visual dichas regiones y analizar el comportamiento del $T_L$ Linke en la república mexicana.

PALABRAS CLAVE: $T_L$ Linke, Análisis de *Clusters*, Reducción de Complejidad, *Machine Learning*, Turbidez.

ABSTRACT

The clustering of geographic zones, by cluster analysis, is a task which enables to identify geographic's regions according to the behavior of geoclimatic's variables. In this work, the $T_L$ Linke was grouped in similar regions in the Mexico country according to the temporal annual variance of the year 2015. Using machine learning algorithms and techniques like Gaussian mixture models and principal components analysis, we obtained two maps of 10 and 9 regions which enables identify and analyze the $T_L$ Linkes behavior on Mexico country.

KEYWORDS: $T_L$ Linke, Cluster Analysis, Complex Reduction, Machine Learning, Turbidity.

## INTRODUCCIÓN

La planeación e instalación de una red de medición solarimétrica requiere de un criterio climático para determinar puntos de medición representativos para la diversidad de ambientes naturales de la región a evaluar. Resulta conveniente que las estaciones de medición estén ubicadas de forma que representen la diversidad climática de la región a evaluar. Es por esto que a la fecha pueden encontrarse trabajos relevantes para la radiación solar. Por ejemplo, a partir de series de tiempo de imágenes de nubosidad (Zagouras et al., 2013) en Grecia, series de tiempo de radiación global horizontal (Journée et al., 2012) en Benlux, mediciones de estaciones terrestres de radiación solar de superficie (Watanabe, 2016) en Japón, estimaciones de radiación solar a partir de imágenes de satélite (Valenzuela et al., 2018; Vindel et al., 2018), así como el análisis variables meteorológicas que dependen de la ubicación geográfica de punto en el área de estudio (Riveros-Rosas et al., 2014) en México.

En este artículo, el objetivo general es regionalizar, a través de mapas, regiones de México similares en su variación de $T_L$ Linke respecto al año 2015 mediante algunas técnicas de análisis de *clusters*, de *Machine Learning* y a partir del conjunto óptimo de regiones. De esta forma, describir de forma estadística el comportamiento del $T_L$ Linke en dichas regiones. Este trabajo servirá posteriormente para el estudio en conjunto de otras variables geoclimáticas relacionadas a la radiación solar para la planeación de redes solarimétricas en México.

## DATOS

El análisis mensual del factor de turbidez atmosférica $T_L$ Linke fue propuesto debido a su relación con la radiación solar. $T_L$ Linke provee una aproximación de los efectos de absorción y esparcimiento de la radiación solar a su paso por la atmosfera. También se puede interpretarse como el número de atmósferas secas y limpias que produciría los efectos de atenuación y dispersión observados en la atmosfera real en ausencia de nubes, (J. Angles, L. M, O. Bauer et al., 1999), (Laguarda y Abal, 2016). Para los modelos como Heliosat 2 y ESRA (Rigollier C. Et al., 2000), el índice de turbidez de Linke es utilizado para la estimación de la radiación solar que alcanza la superficie en condiciones de cielo despejado. Algunos de los valores más frecuentes de $T_L$ Linke se describen en la Tabla 1. (Olcoz Larrayoz, 2014).

Tabla 1. Algunos valores frecuentes de $T_L$ Linke

| Tipo de atmósfera | TL |
|---|---|
| Describe una atmósfera de Rayleigh solo con efectos de dispersión molecular | =1 |
| Muy clara (pocas partículas en suspensión) | ~2 |
| Clara y cálida | ~3 |
| Húmeda y cálida | 4~6 |
| Con polución | >6 |

Para la investigación se usaron imágenes mensuales de la región mexicana con respecto a su valor de $T_L$ Linke en el año 2015 con una resolución original de $2127 \times 3066$ píxeles de la base de datos SODA (http://www.soda-pro.com/).

## METODOLOGÍA

El diagrama para la regionalización, mediante el análisis de clúster empleado, se describe en la Fig. 1. En la figura, $\mathbf{A}$ representa las imágenes de $T_L$ Linke como una matriz de $n \times m \times d$ de números reales. Donde $n$ es el número de filas, $m$ el número de columnas relacionadas a la geolocalización y $d$ la serie de tiempo analizado, en nuestro caso 12 meses. Las imágenes entran a la etapa de preprocesamiento con el objetivo de crear una base de datos $X_{BD}$ que contiene $píxeles\ x\ características$ para analizar de un rango de clúster que va desde $k_i$ á $k_r$. Esto da como resultado $\bar{k}$ vectores de clases (asignación del píxel a una región), estas son evaluadas para obtener un número óptimo $\bar{O}$ de regiones con respecto a la variación temporal de la turbidez atmosférica en México.

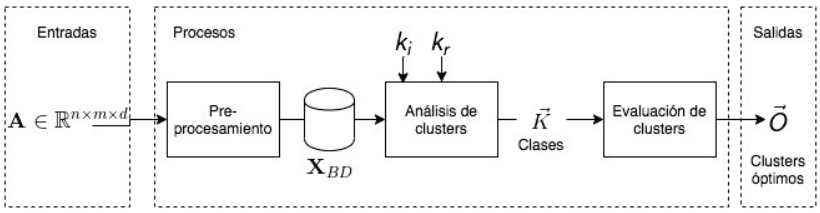


Fig. 1. Diagrama de actividades para la regionalización del $T_L$ Linke.

Preprocesamiento

Durante esta etapa se efectúan dos actividades, que consisten en un recorte de la imagen y el análisis de componentes principales (ACP). El objetivo general de este proceso es reducir la complejidad computacional de la evaluación y análisis de *cluster* por medio de algoritmos y técnicas de *Machine Learning*. Lo anterior se debe a que imágenes con una resolución espacial y análisis de series de tiempo muy grandes, dan como resultado análisis muy complejos. Con la resolución de las imágenes de $T_L$ Linke, para hacer un análisis de regiones de clúster con 12 meses, se tendrían que analizar $2127 \times 3066 \times 12$ píxeles, lo que sería un total de $7.8256584 \times 10^7$ píxeles.

El recorte es efectuado con el objetivo de reducir el número de píxeles a analizar. Esto permite crear una matriz reducida de la imagen $\boldsymbol{A}$ denotada como $\boldsymbol{a}$ el cual solo contiene píxeles de la superficie Mexicana. Por otro lado, ACP es efectuado para reducir el número de características, en este caso el número de meses a analizar.

ACP es una transformación lineal, que escoge un nuevo sistema de coordenadas, para el conjunto de datos. En dichos datos, la varianza de mayor tamaño del conjunto es almacenada como el primer componente y se ordenan de forma decreciente. Para aplicar ACP, es necesario normalizar la matriz $\boldsymbol{a}$ aplicando la Ec. (1). Donde $x$ es un píxel de la matriz $\boldsymbol{a}$ , $x_{min}$ es el valor mínimo, $x_{max}$ el valor máximo de la matriz y $x'$ es el nuevo valor del píxel, cuyo valor se encuentra entre el rango [0,1]. El resultado de normalizar es $\hat{\boldsymbol{a}}$ normalizada.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

ACP se obtiene a partir de la Ec. (2) y (3), donde $|\boldsymbol{a}|$ es la matriz normalizada, $\boldsymbol{w}$ es un eigenvector de la matriz de covarianza $\boldsymbol{\Sigma}$, $\lambda$ los eigenvalores, $\mathbf{I}$ la matriz identidad y $det$ la determinante de la matriz.

$$\boldsymbol{\Sigma}\mathbf{w} = \lambda\mathbf{w} \qquad (2)$$

$$\det(\lambda\mathbf{I} - \hat{\boldsymbol{a}}) = 0 \qquad (3)$$

El resultado de aplicar ACP es la sumatoria de los eigenvalores, donde en el primer eje es aquel que contiene la mayor varianza.

Evaluación de *clusters*

El análisis de clúster es una tarea no supervisada de *machine learning* que separa los datos en grupos (o clases de objetos similares) basados en un criterio de similaridad (Tobergete y Curtis, 2013). El algoritmo más usado empleado para determinar regiones climatológicas es *Kmeans* o modificaciones del mismo (Zagouras et.al., 2013). El problema con dicho algoritmo es que la asignación de grupos depende de la distancia entre los datos y un valor central llamado centroide. Como la distancia euclidiana es esférica, el algoritmo tiene deficiencias para agrupar datos alejados de dicho centroide (VanderPlas,2016).

Con el objetivo de regionalizar zonas geográficas con respecto a la varianza temporal de $T_L$ Linke, y evitar las deficiencias, el algoritmo de Mezclas Mixtas Gaussianas (GMM) es un algoritmo de agrupamiento suave, lo cual significa que un píxel tiene la probabilidad de pertenecer a uno o más clúster a partir de un conjunto finito de distribuciones gaussianas de parámetros desconocidos.

Una distribución de probabilidad gaussiana multivariada está dada por la Ec. (4). Donde $\mu$ es la media, $\Sigma$ la matriz de covarianza, $\pi$ es un parámetro que define la ponderación de la distribución gaussiana (Contreras, 2019).

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu)\right) \qquad (4)$$

Para determinar la probabilidad de $x$ pertenezca a una distribución gaussiana, se usa el algoritmo de EM (máxima de expectación).

Durante el paso de estimación se inicializa valores aleatorios $\mu_k$, $\Sigma_k$y $\pi_k$ y estima que la probabilidad de que el elemento pertenezca a una distribución aplicando la Ec. (5). Donde $z_{nk} \leq 1$ es la variable latente que toma $k$ posibles valores cuando $x$ viene de una distribución gaussiana $k$ y cero en otro caso.

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{k} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \quad (5)$$

Durante el paso de maximización se actualizan los parámetros $\mu_k$, $\Sigma_k$y $\pi_k$ aplicando las Ec (6), (7) y (8).

$$\pi_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})}{N} \quad (6)$$

$$\mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})} \quad (7)$$

$$\Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T)}{\sum_{n=1}^{N} \gamma(z_{nk})} \quad (8)$$

El algoritmo se itera sucesivamente, hasta un número de iteraciones definido, o bien cuando el algoritmo converge, es decir, cuando los valores de $\mu_k$, $\Sigma_k$y $\pi_k$ no sufren cambios o los cambios son menores a un valor de tolerancia.

Evaluación de *clusters*

Para agrupar regiones similares, debemos establecer/definir a priori un número de clases a evaluar. El número óptimo de clases puede ser una tarea ambigua y compleja. Por ello, existen técnicas de evaluación de clúster cuya meta es determinar el número de clases o regiones óptimas en las que se ajustan nuestros datos. Estas pueden ser validaciones externas (previo conocimiento del agrupamiento de los datos) y/o validaciones internas que consideran la información intrínseca de la estructura geométrica de los datos. El segundo tipo de validación es muy apropiada para analizar los datos sin necesidad de hacer suposiciones del comportamiento de los datos. Los índices Davies-Bouldin (DB) y Calinski-Harabaz (CH), así como el método L, son algoritmos bien conocidos por su aplicación en la regionalización (Zagouras et al., 2013; Zagouras et al., 2014a).

El índice DB se basa en la relación de distancias de los inter e intra-grupos donde un valor pequeño de DB indica *clusters* compactos y bien separados (minimiza el potencial de similaridad entre cada uno) mientras que el índice CH se basa en la posición de los centroides del clúster. Un valor grande de CH está relacionado a la partición de *clusters* bien separados en donde los centroides o medias de los *clusters* se encuentran distantes en el espacio y mantienen distancias compactas inter-*cluster*. Las evaluaciones DB y CH pueden ser visualizadas como un gráfico de evaluación, donde el eje $x$ representa la cantidad de clases que se están evaluando y el eje $y$ la magnitud de la evaluación.

Para los dos tipos de índices es necesario evaluar sobre un rango grande de clases, los cuales generalmente producen comportamientos asintóticos. Por esta razón para determinar un número razonable de clases, el Método L puede ser aplicado a los índices DB y CH. El método L se basa en encontrar un punto crítico o "rodilla" en el gráfico de evaluación que determina el apropiado número de clases. Este punto crítico es donde la intersección de los dos mejores ajustes lineales al lado izquierdo y derecho del punto crítico dan el error cuadrático medio (RMSE) más pequeño y cubren la mayoría de los puntos. El método L se encuentra definido por la Ec. (9), donde $K$ es el número total de *clusters*, $Lc$ y $Rc$ son secuencias del lado izquierdo y derecho de los datos particionados por el punto $c$, $RMSE_{Lc}$ y $RMSE_{Rc}$ son los errores cuadráticos medios de $Lc$ y $Rc$.

$$RMSE_C = \frac{c-1}{K-1} \cdot RMSE_{Lc} + \frac{c-1}{K-1} \cdot RMSE_{Rc} \quad (9)$$

Dicho esto, el número de clases más adecuado es aquel que corresponde con el mejor ajuste determinado por el mínimo valor del $RMSE_C$.

## IMPLEMENTACIÓN Y RESULTADOS

En la etapa de preprocesamiento, las matrices de $T_L$ Linke fueron recortadas y el algoritmo ACP fue aplicado. Gracias a las dos actividades fue posible crear una base de datos $X_{BD}$ de 1130253 píxeles × 3 componentes principales, siendo ésta una reducción significativa de los $7.8256584 \times 10^7$ píxeles originales a analizar. La base de datos almacena solamente píxeles de la superficie mexicana y 3 eigenvalores que describen el 95.53% de la varianza total del $T_L$ Linke durante los 12 meses de estudio como se puede observar en Fig. (2).

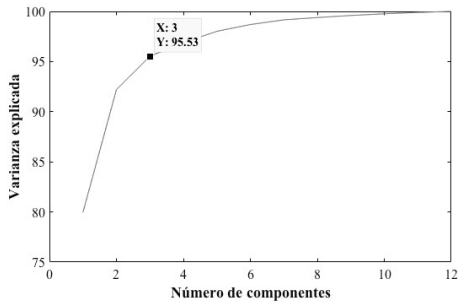


Fig. 2. Relación de componentes principales / varianza explicada.

Para regresar a la imagen original, las coordenadas $x$ e $y$ fueron almacenadas en una base de datos de posiciones, con los valores nuevos de nuestras matrices. En la Fig. 3, se construye una imagen de $T_L$ Linke como la suma de los primeros 3 componentes principales.

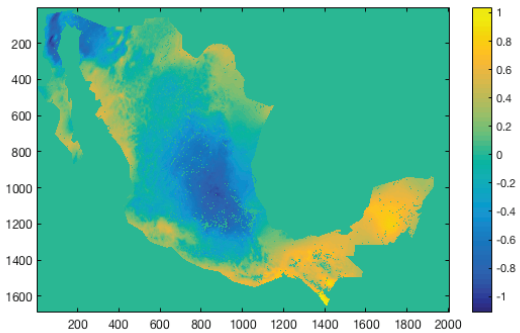


Fig. 3. Suma de los primeros 3 componentes de $T_L$ Linke.

Como se puede observar en la Fig. 3. Mediante la aplicación ACP es posible observar las magnitudes y direcciones de los primeros tres pca's de $T_L$ Linke con respecto a su varianza temporal en diversas zonas geográficas.

En la etapa de análisis de *clusters*, el algoritmo GMM fue aplicado a la base de datos $X_{BD}$ en un rango de $2 \leq k \leq 50$ clases. Mientras que en la etapa de evaluación los resultados del método L, para los índices DB y CH, se pueden observar en la Fig. 4. Y la Fig. 5. El número óptimo de clases de acuerdo con el método L es 9 y 10 respectivamente.

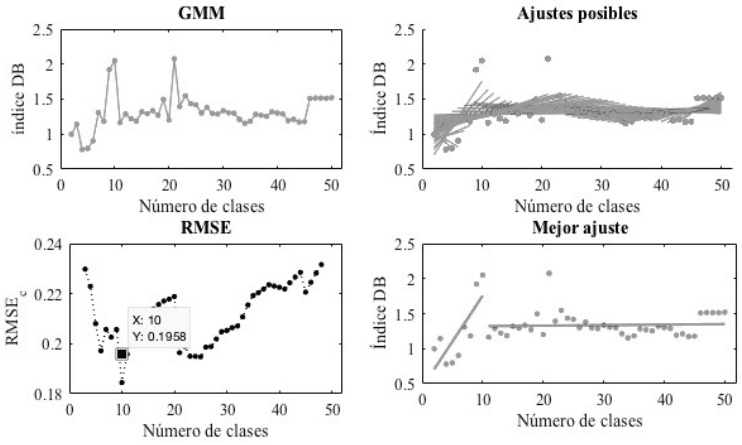


Fig. 4. Método L. Para índice DB

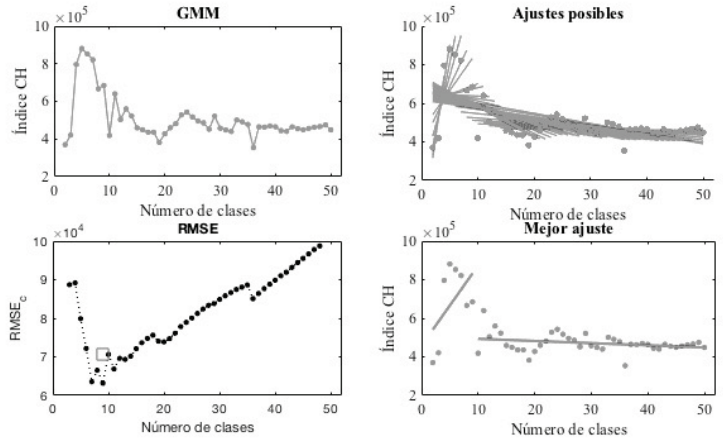


Fig. 5. Método L.  Para índice CH.

En la Fig. 6 muestra la regionalización a 10 clases determinadas (usando al método L) a partir de los indices DB. La Fig. 7 muestra la regionalización a 9 clases determinadas mediante el índice CH.

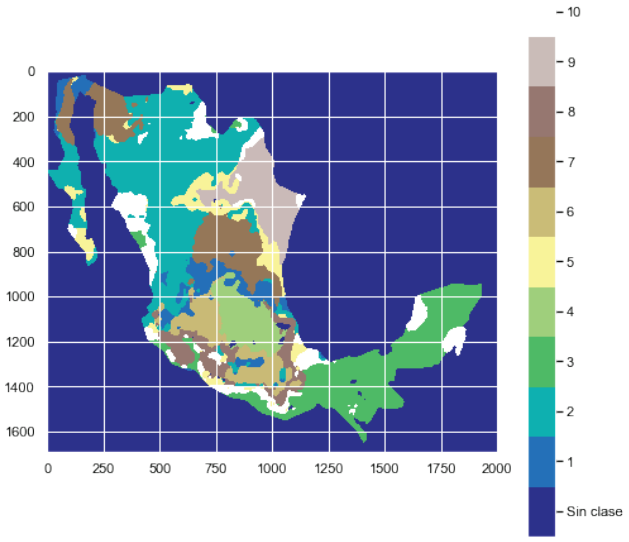


Fig. 6. Regionalización a 10 clases de $T_L$ Linke.

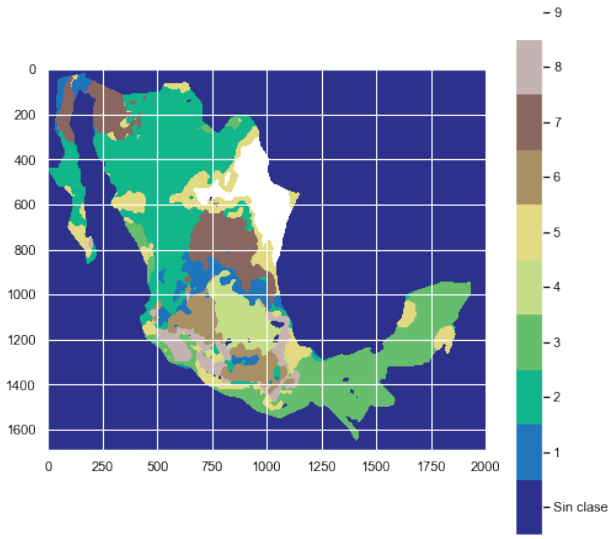


Fig. 7. Regionalización a 9 clases de $T_L$ Linke.

Para analizar los valores anuales de $T_L$ Linke en las clases resultantes (de cada regionalización) fueron aplicados los siguientes estadísticos sobre el $T_L$ Linke: media, desviación estándar (Std.), varianza (Var.), valor mínimo (Min.), máximo (Máx.) (ver Tabla 2). Para diferenciar los estadísticos, de las dos regionalizaciones, DB índica la regionalización a 10 clases y CH la regionalización a 9 clases. Las celdas vacías son debido a que la regionalización descubierta mediante el método L, al relacionar los números de clases e índices CH, solo tienen 9 clases.

Tabla 2. Comportamientos anuales de $T_L$ Linke por región.

| Clase | media | | Std. | | Var. | | Min. | | Máx. | |
|---|---|---|---|---|---|---|---|---|---|---|
| | DB | CH | DB | CH | DB | CH | DB | CH | DB | CH |
| 1 | 3.243 | 3.158 | 0.0291 | 0.03 | 0.0002 | 0.0002 | 1.89 | 1.8566 | 4.62 | 4.5486 |
| 2 | 3.925 | 3.925 | 0.0465 | 0.046 | 0.0004 | 0.0004 | 2.42 | 2.4214 | 5.34 | 5.4397 |
| 3 | 3.6970 | 3.705 | 0.0341 | 0.332 | 0.0003 | 0.0332 | 2.30 | 2.3056 | 5.54 | 5.5486 |
| 4 | 3.261 | 3.259 | 0.0271 | 0.026 | 5.0E-5 | 4.8E-5 | 2.31 | 2.3251 | 4.29 | 4.2993 |
| 5 | 3.909 | 3.833 | 0.0319 | 0.035 | 0.0001 | 0.0002 | 2.37 | 2.3575 | 5.95 | 5.9529 |
| 6 | 3.233 | 3.269 | 0.1084 | 0.116 | 0.0091 | 0.0177 | 1.8 | 1.8 | 5.3 | 5.3 |
| 7 | 3.640 | 3.64 | 0.0499 | 0.041 | 0.0018 | 0.001 | 1.85 | 1.9512 | 5.34 | 5.5069 |
| 8 | 3.563 | 3.488 | 0.0282 | 0.026 | 8.1E-5 | 5.2E-5 | 2.25 | 2.2567 | 4.85 | 4.8018 |
| 9 | 4.189 | 4.148 | 0.0335 | 0.044 | 0.0001 | 0.0003 | 2.90 | 2.7237 | 5.44 | 5.7917 |
| 10 | 3.243 | | 0.0669 | | 0.0015 | | 2.45 | | 5.79 | |

Para evaluar que los valores promedios de cada clase tuvieran relación con los centroides de cada clúster, el coeficiente de correlación Pearson fue empleado. Esto implicó la transformación inversa de los centroides, definidos como componentes principales, a valores de $T_L$ Linke originales. El resultado de la correlación de 10 clases fue de 0.99678328 y a 9 clases de 0.99863388, lo que implica que los valores anuales de $T_L$ Linke de cada región son prácticamente los mismos a los valores de los centroides correspondientes. Los 12 valores anuales de $T_L$ Linke (para cada pixel) se pueden mapear a tres componentes principales. Esto sin tener diferencias significativas en los datos. Lo anterior resulta conveniente para análisis más complejos, e indican a su vez poca varianza con respecto al $T_L$ Linke en México.

## CONCLUSIONES

Gracias a las actividades de *Machine Learning*, como el procesamiento (recortes y el ACP) podemos determinar áreas de interés y reducir el número de pixeles. En nuestro caso describió el 95.53% de la varianza temporal de $T_L$ Linke. Esto significa que el preprocesamiento puede reducir en varios ordenes de magnitud la cantidad de información a analizar sin que se pierda información significativa al realizar la regionalización. Se condujo un análisis de clúster y se evaluó mediante el algoritmo GMM y el método L. Los cuales ayudaron a visualizar regiones significativas de $T_L$

Linke (en la superficie de México) con poca varianza, y desviación estándar sobre la media de cada región. Debido a que los valores promedios de cada clase tienen poca variación con respecto a los centroides, la agrupación de las regiones tiene un margen de error muy pequeño sobre sus valores anuales, y esto implica que dichas regiones representan un buen estimado del $T_L$ Linke en México. Además de lo anterior, cabe destacar la buena correlación entre los valores de Linke reconstruidos a partir de los centroides (definidos como componentes principales) y los valores de $T_L$ Linke originales. Los coeficientes de correlación de Pearson, entre los valores reconstruidos y originales, fueron superiores a 0.998 y 0.996, para un número de clases óptimo de 9 y 10 respectivamente.

Como trabajo futuro, se desea aplicar esta metodología a otras variables geoclimáticas en conjunto con el $T_L$ Linke. Esto a fin de regionalizar la república mexicana con información de imágenes de altitud y series de tiempo de albedo e índice de nubosidad. El objetivo es crear un mapa donde las clases agrupen de mejor forma la diversidad climática de la república mexicana y relacionarlos con la radiación solar.

## REFERENCIAS

Contreras Carrasco, O. (2019). Gaussian Mixture Models Explained. Retrieved February 15, 2020, from https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95

Joel Angles, Lionel Ménard, Olivier Bauer, Christelle Rigollier, Lucien Wald. A climatological database of the Linke turbidity factor. ISES Solar World Congress 1999, Jul 1999, Jerusalem, Israel. pp. 432-434. hal-00465886

Journée, M. & Demain, C. & Müller, Richard & Bertrand, C. (2012). Towards a climatology of surface incoming solar radiation over the Benelux by merging long time series of Meteosat-derived estimations and ground-based measurements. 2664-.

Laguarda, A., & Abal, G. (2016). Índice de turbidez de Linke a partir de irradiación solar global en el Uruguay.

Olcoz Larrayoz, A. (2014). Implementación del método heliosat para la estimación de la radiación solar a partir de imágenes de satellite. Reporte técnico, Universidad Pública de Navarra, Pamplona

Rigollier C. Bauer O. and Wald L. (2000). On the clear sky model of the ESRA — European Solar Radiation Atlas — with respect to the heliosat method, Solar Energy, 68 (1), pp33-48.

Riveros-Rosas D, Bonifaz R., Valdes M., Rivas R. "Análisis por Región de Información Solarimétrica en la República Mexicana" XI congreso Iberoamericano de energía solar y XXXVIII Semana Nacional de Energía Solar, Querétaro, Querétaro, México, octubre, 2014.

Tobergte, D. R. and Curtis, S. (2013). Machine learning with R, volume 53, Packt Publishing Ltd., Birmingham, second edition.

Valenzuela, R. X., Vindel, J. M., Navarro, A. A., Zarzalejo, L. F., Paz-Gallardo, A. y Ferrera-Cobos, F., 2018. GeoPAR - Red de estaciones de medida de Radiación Fotosintéticamente Activa. Ponencias de: XVI Congreso Ibérico y XII Iberoamericano de Energía Solar, 21-6-2018, Madrid (España)

VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, Inc., 1st edition.

Vindel, J. M., Valenzuela, R. X., Navarro, A. A. y Zarzalejo, L. F., 2018. Methodology for optimizing a photosynthetically active radiation monitoring network from satellite-derived estimations: A case study over mainland Spain. Atmospheric Research 212, 227-239

Watanabe, Takeshi & Takamatsu, Takahiro & Nakajima, Takashi. (2016). Evaluation of Variation in Surface Solar Irradiance and Clustering of Observation Stations in Japan. Journal of Applied Meteorology and Climatology. 55. 10.1175/JAMC-D-15-0227.1.

Zagouras, A., Kazantzidis, A., Nikitidou, E., & Argiriou, A. A. (2013). Science Direct Determination of measuring sites for solar irradiance, based don cluster analysis of satellite-derived cloud estimations. Solar Energy, 97, 1-11. https://doi.org/10.1016/j.solener.2013.08.005.

Zagouras, A., Inman, R. H., and Coimbra, C. F. (2014a). On the determination of coherent solar microclimates for utility planning and operations. Solar Energy, 102: 173-188.