

# Agrupamiento Espacial de Emisiones de Metano por Ganado Vacuno en Pastoreo en México (Junio 2024)

Erick Walter Trejo Duran, 166770 y Heriberto Espino Montelongo, 175199

**Abstract**—En este trabajo se realiza un análisis de agrupamiento espacial de las emisiones de metano generadas por ganado vacuno en pastoreo, utilizando datos por estación y por estado de la República Mexicana durante el mes de junio de 2024. Se exploran diferentes esquemas de pesos espaciales, contigüidad tipo Queen y distancia euclíadiana entre centroides con  $k$ -vecinos ( $k = 4$ ), y se comparan dos algoritmos de agrupamiento: *Max-p* y *Spatially Constrained K-Means* (SC K-Means). Se analiza la estructura de los clústeres, la proporción de variabilidad entre versus total (BSS/TSS).

## I. INTRODUCCIÓN

La emisión de metano ( $\text{CH}_4$ ) por ganado vacuno en pastoreo constituye una de las principales fuentes de gases de efecto invernadero en el sector agropecuario. En México, con sus 32 entidades federativas, existen diferencias marcadas en prácticas de manejo, tipo de vegetación y densidad de ganado, lo cual justifica un análisis de agrupamiento espacial para identificar regiones con patrones similares de emisión. El uso de técnicas de agrupamiento espacial permite considerar la dependencia espacial de los datos y facilita la configuración de políticas de mitigación regionalizadas.

El problema central consiste en agrupar espacialmente los estados de la República Mexicana según las emisiones promedio de metano registradas por estación en junio de 2024, a fin de identificar zonas críticas y sus posibles determinantes ambientales y ganaderos. El objetivo general es aplicar métodos de agrupamiento espacial bajo distintos esquemas de pesos y algoritmos para cuantificar la calidad de los clústeres mediante la proporción BSS/TSS.

## II. SISTEMA DE REFERENCIA

Se adoptó la proyección EPSG:6372 debido a que preserve áreas a escala nacional, minimiza distorsiones en la medición de superficies y principalmente mantiene las distancias, a diferencia de EPSG:4326, que emplea coordenadas geográficas con grados de latitud y longitud y no define un sistema métrico uniforme, EPSG:6372 posibilita el cálculo directo de distancias euclidianas y la aplicación de algoritmos de agrupamiento espacial basados en métricas.

## III. PREPROCESAMIENTO DE LOS DATOS

Los datos de emisiones provenían de estaciones de monitoreo distribuidas en 2,305 estaciones. Para solucionar la falta de representatividad en municipios sin estaciones, ya que hay 2,478 municipios (INEGI, 2024), se agruparon los registros por estado y se calculó la media de emisiones de metano, además, hace más sencilla la interpretación y el menos costoso el tiempo computacional. Únicamente se ocupará la variable de cantidad de emisiones, porque, una vez agrupados por la media, todas las variables estaban correlacionadas linealmente ( $r = 1.0$ ).

En la Figura 1 se observa que Baja California Sur presenta el promedio más alto, siendo que sus estaciones promedian 1,060 toneladas por mes, duplicando el valor de Baja California Norte de 524, teniendo valores que superan por cinco veces el promedio nacional de 113. Visualmente, y dado que únicamente trataremos una variable, se distinguen entre dos y tres clústeres, uno en la península de Baja California con el noroeste de México, que contrasta con el resto del país, sugiriendo que factores como densidad ganadera y clima árido influyen en las emisiones elevadas de metano en el noroeste mexicano, posiblemente debido a la calidad y tipo de forraje, es decir, vegetación más seca, fibrosa y menos digestible para el ganado vacuno en pastoreo.

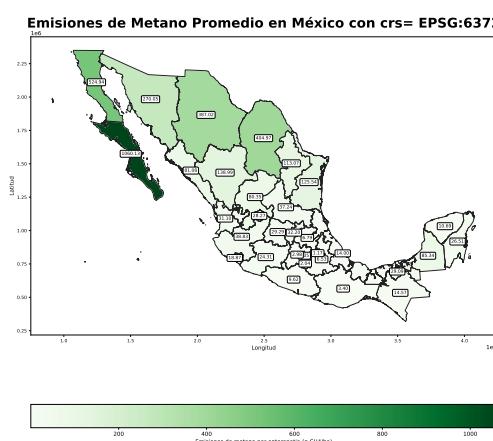


Fig. 1. *Elaboración Propia*. Descripción de la imagen.

#### IV. INTERPOLACIÓN ESPACIAL COMO APOYO VISUAL

Se implementó interpolación de IDW ( $P = 1$ ) únicamente como herramienta de visualización para confirmar que la utilización de la media concordara con los datos interpolados. El mapa de interpolación 2 revela cuatro zonas principales, que también se mantienen en la Figura 1, pero con mayor dificultad de observar en la primera figura debido a la paleta de colores: la región noroeste tiene valores máximos de hasta 2,632 toneladas por mes, que descienden gradualmente conforme se acerca a la zona metropolitana del país.



Fig. 2. *Elaboración Propia. Descripción de la imagen.*

#### V. ESQUEMAS DE PESOS ESPACIALES

Se definieron dos tipos de pesos con la finalidad de compararlos y rescatar el de mejor desempeño.

##### A. Contigüidad Queen ( $peso=1$ )

Fueron utilizados los pesos de la contigüidad de la reina con los vecinos con los que comparte fronteras. Se reportó un mínimo de 1 vecino, máximo de 8, media de 4.19 y mediana de 4 vecinos, las celdas no vacías fueron 13.09%. Fig 3.

Property	Value
type	queen
symmetry	symmetric
file	contiguity-queen-1.gal
id variable	idx
order	1
# observations	32
min neighbors	1
max neighbors	8
mean neighbors	4.19
median neighbors	4.00
% non-zero	13.09%

Fig. 3. *Elaboración Propia. Descripción de la imagen.*

El histograma de distribución de vecinos muestra una cola derecha, con frecuencia mayor en 2 a 4 vecinos. Fig 4.

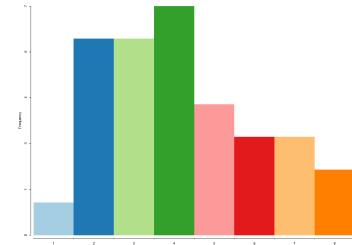


Fig. 4. *Elaboración Propia. Descripción de la imagen.*

##### B. Distancia euclíadiana entre centroides ( $KNN, k = 4$ )

Cada centroide de un polígono de estado se conecta con sus cuatro centroides más cercanos. En este esquema, todos los estados tienen exactamente 4 vecinos y las celdas no vacías fueron 12.50%. Estos datos se presentan en la Fig 5, y una visualización de las conexiones en Fig 6.

Property	Value
type	k-NN
symmetry	asymmetric
file	distance-euclidean-xycentroids-knn-4.gwt
id variable	idx
distance metric	Euclidean
distance vars	centroids
neighbors	4
# observations	32
min neighbors	4
max neighbors	4
mean neighbors	4.00
median neighbors	4.00
% non-zero	12.50%

Fig. 5. *Elaboración Propia. Descripción de la imagen.*



Fig. 6. *Elaboración Propia. Descripción de la imagen.*

## VI. MAX-P Y PESOS QUEEN=1

Se fijó `random_seed=42` para reproducibilidad. El método *Max-p* empleó búsqueda local greedy, con un tamaño mínimo de 4 regiones por clúster y distancia euclídea sin transformación. Generó ocho clústeres, con centros de emisión que oscilan entre 2.97 y 484.05 toneladas por mes. La proporción  $BSS/TSS = 0.567$  indica baja separación entre clústeres y alta varianza dentro de los clústeres, sugiriendo que no se obtuvo una buena discriminación de patrones. Fig 7.



Fig. 7. *Elaboración Propia*. Descripción de la imagen.

## VII. SC K-MEANS CON $k = 4$ Y PESOS KNN $k = 4$

SC K-Means se ejecutó con inicialización *KMeans++* (150 re-ejecuciones) y máximo 1000 iteraciones, sobre pesos euclidianos ( $k = 4$ ). Los cuatro centros resultantes (28.81, 100.14, 396.74, 1060.13) y un  $BSS/TSS = 0.958$  reflejan una excelente separación espacial y baja varianza interna. En los clústeres se agrupa a Baja California Sur como clúster único, debido a sus altas emisiones; BC Norte, Sonora, Chihuahua y Coahuila por sus emisiones medias-altas; y a Sinaloa, Durango y Zacatecas por sus zonas medias-bajas, separando al resto del país. Fig 8.



Fig. 8. *Elaboración Propia*. Descripción de la imagen.

El alto ratio entre-varianza a total implica que los clústeres explican el 95.8% de la variabilidad global. Esto sugiere que el número de vecinos y la homogeneidad de los clústeres son óptimas para distinguir regiones con distinto nivel de emisiones.

## VIII. SC K-MEANS CON $k = 4$ Y PESOS QUEEN=1

Bajo contigüidad Queen=1, se obtuvieron también cuatro clústeres con centros similares y un  $BSS/TSS = 0.970$ . La incorporación de la conexión por vértices mejoró ligeramente la separación (1.2%), al restringir agrupaciones a adyacencias reales, concatenando a Nuevo León y Tamaulipas a la región el  $BSS/TSS$  aumenta, indicando que estos estados comparten características de emisión y vecindad con la región noroeste. Fig 9.



Fig. 9. *Elaboración Propia*. Descripción de la imagen.

Como el ratio  $BSS/TSS$  es mayor con los pesos Queen=1 que con los de los pesos de KNN=4, se realizarán nuevos k-means con diferentes números de medias para los pesos de Queen=1.

### IX. SC K-MEANS CON $k = 2$ Y PESOS QUEEN=1

Con dos clústeres y pesos Queen-1, los centros fueron 68.47 y 792.53, presentando un  $BSS/TSS = 0.678$ . La reducción del número de clústeres sacrifica información inherente a los datos, pues solo distingue a Baja California Sur y Norte del resto del país. Fig 10.



Fig. 10. *Elaboración Propia.* Descripción de la imagen.

### X. SC K-MEANS CON $k = 3$ Y PESOS QUEEN=1

Con tres clústeres y los pesos de Queen-1, los centros (36.74, 396.74, 1060.13) y un  $BSS/TSS = 0.949$  equilibran la calidad de separación con el número de particiones, distinguiendo tres niveles de emisión: bajo, medio y alto, por lo que podría ser el mejor número de clústeres que pueden ser escogidos. Fig 11.



Fig. 11. *Elaboración Propia.* Descripción de la imagen.

### XI. SC K-MEANS CON $k = 5$ Y PESOS QUEEN=1

Con cinco clústeres y Queen-1, los centros (20.59, 107.81, 354.01, 524.94, 1060.13) y un  $BSS/TSS = 0.985$  indican una separación mejor a todos los demás, pero esta mejora en la varianza explicada se debe principalmente a la subdivisión de clústeres ya pequeños. En particular, el clúster correspondiente a la península de Baja California fue separado en dos, lo cual puede sobreajustar el modelo sin aportar una mejora interpretativa sustancial. Fig 12



Fig. 12. *Elaboración Propia.* Descripción de la imagen.

### XII. CONCLUSIONES

El análisis de agrupamiento espacial sobre las emisiones de metano por ganado vacuno en pastoreo durante junio de 2024 en los estados de México permitió identificar patrones geográficos. Se utilizaron diferentes configuraciones de SC K-Means con diferentes pesos espaciales, revelando que los clústeres más consistentes corresponden a estados del norte con altas emisiones, especialmente Baja California Sur y Norte y estados del centro-sur y sureste con emisiones bajas.

La heterogeneidad en las regiones responde a diferencias en densidad ganadera y condiciones climáticas. Regiones áridas del noroeste concentran las mayores emisiones, mientras que el resto de la república presenta valores menores, posiblemente por menor carga animal o diferente tipo de forraje.

Se observó que mientras mayor es el número de clústeres, mayor es el  $BSS/TSS$ , indicando una mejor separación estadística. Sin embargo, a partir de cuatro o cinco clústeres, se comienza a subdividir grupos que ya eran pequeños (como la península de Baja California), lo cual reduce la interpretabilidad del modelo.

El modelo con tres o cuatro clústeres ofrece un equilibrio adecuado entre varianza explicada y capacidad interpretativa, señalando de forma robusta que las emisiones más intensas se concentran en el noroeste del país, particularmente en Baja California Sur.