Universidad Nacional Autónoma de México

Instituto de Energías Renovables

Modelación de estado de viento regionales a partir de datos anemométricos

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Ingeniero en Energías Renovables

PRESENTA:

Guillermo Olvera Guerrero

DIRECTOR DE TESIS

Dr. Miguel Robles Pérez

TEMIXCO, MOR. 2019

UNAM
CAMPUS MORELOS

IER
Instituto de Energías Renovables

OF/IER/LIER/079/2019
ASUNTO: Notificación de jurado y fecha para examen profesional.

LIC. IVONNE RAMÍREZ WENCE
DIRECTORA GENERAL DE ADMINISTRACIÓN ESCOLAR
P r e s e n t e .

Por medio de la presente le informo que el día viernes 31 de mayo de 2019, a las 16:00 hrs., en el Instituto de Energías Renovables, el **C. GUILLERMO OLVERA GUERRERO,** con número de cuenta 415012575 de la Licenciatura de Ingeniería en Energías Renovables, llevará a cabo la presentación del trabajo de tesis y examen profesional titulado:

"Modelación de estados de viento regionales a partir de datos anemométricos"

Debido a que el alumno ha cumplido con los requisitos que establece el Reglamento General de Exámenes, el Comité Académico de la Licenciatura de Ingeniería en Energías Renovables, le asigna el Jurado de Examen Profesional integrado por los siguientes académicos.

PRESIDENTE:	**DR. EDUARDO RAMOS MORA**
VOCAL:	**DR. ALBERTO REYES BALLESTEROS**
SECRETARIO:	**DR. MIGUEL ROBLES PÉREZ**
SUPLENTE:	**DR. OSVALDO RODRÍGUEZ HERNÁNDEZ**
SUPLENTE:	**DR. OSCAR ALFREDO JARAMILLO SALGADO**

Sin otro particular, le envío un cordial saludo.

A t e n t a m e n t e,
"Por mi raza hablará el espíritu"
Temixco, Mor., a 23 de mayo de 2019

Dr. Jorge Alejandro Wong Loya
Coordinador Académico de la LIER
IER-UNAM

UNAM
INSTITUTO DE ENERGÍAS RENOVABLES
COORDINACIÓN DE DOCENCIA

JAWL´mfp.

Priv. Xochicalco s/n,
Col. Centro,
62580 Temixco, Morelos, México
T. 01 (777) 362 00 90
(52) 55 56 22 97 00

ier.unam.mx

Man masters nature not by force
but by understanding.
J. Bronowski
Science and human values

A mis padres Adela y Guillermo, y a mis hermanos David y Ramón.

A Ramón Matías, un niño increíble, para que cuando crezca se de cuenta que todo lo que se proponga es posible con dedicación y esfuerzo.

Agradecimientos

Muchas son las personas que contribuyeron a la realización de esta tesis a través de su esfuerzo y convivencia. Por ello, quiero dedicar un fragmento de este trabajo para reconocerles y darles mi más profundo agradecimiento.

A mis padres Adela y Guillermo, que formaron una parte angular para hacer esto posible. Soy consciente de todas las desveladas y trabajos extra que tuvieron que hacer durante estos años y aunque no les puedo regresar ese tiempo y esfuerzo invertido, quiero agradecerles por acompañarme en este proyecto personal de forma tan desinteresada y, más que eso, por toda la formación que me dieron como ser humano y el amor que me hacen sentir.

Agradezco también a mis hermanos David y Ramón, ya que ellos como mis padres siempre estuvieron presentes para cualquier cosa que necesitara y me brindaron su apoyo en cada momento. Además, sus enseñanzas de vida son de las cosas que más valoro y estimo.

A mi asesor Miguel, por ser mi guía durante el trabajo de tesis y un gran amigo. Agradezco mucho su esfuerzo y todas las cosas que ha hecho por mí, entre ellas, descubrir los temas que me apasionan y motivarme a que las siga haciendo.

A todos mis amigos de la 4ta generación de la LIER (el orden no es importante, la amistad es más compleja que una jerarquía): Carlos, Ivette, Eros, Sergio, Juanico, Julio, Eira, Itzel, Caro, Sady, Luis, Sam, Juanca, Ade, Migue, Anali, Darinka, Clari, Adrián, Gesu, Fer y Sebas. Gracias por todas las experiencias increíbles que viví a su lado y ojalá sigan siendo muchas más.

Gracias a la Universidad Nacional Autónoma de México y en especial al Instituto de Energías Renovables por todas las herramientas y oportunidades que me han brindado para crecer académica, profesional y personalmente. A mis sinodales: Eduardo Ramos, Alberto Ballesteros, Osvaldo Rodriguez y Oscar Jaramillo, por el tiempo que se tomaron para hacer de esta tesis un mejor trabajo.

Resumen

En este trabajo se desarrolló una metodología para la definición de estados de viento regionales a partir de cuatro estaciones anemométricas como un primer paso para generar un modelo de predicción de viento para estimaciones de recurso eólico. Se compararon dos algoritmos: K-Means y Mezcla Gausiana, que presentan características útiles en la definición de los estados de viento. Se encontró que ambos algoritmos definen estados de viento similares, ya sea porque definen prácticamente al mismo estado o porque hacen una composición de estados de viento con otros. Se definió el algoritmo más fácil de automatizar, siendo entonces que Mezcla Gausiana permite ésto al no ser necesaria información *a priori* de los estados a definir.

Se usó la *sincronización de eventos* para comparar los estados de viento de diferentes estaciones anemométricas, considerando un tiempo de desfase que permita dar tiempo a que los estados se muevan de una estación a otra. Considerando los tiempos de desfase donde el coeficiente de correlación es máximo entre cada estado de viento de las cuatro estaciones se determinan las relaciones entre los estados de viento de las estaciones y se les representa con una estructura de red. Las relaciones que son consideradas en esta red de estados de viento se determinan limitando el coeficiente de correlación mínimo que puede ser aceptado y para ello se busca que la densidad de bordes de la red ρ tenga un valor aproximado a 0.01, de acuerdo a lo reportado en la literatura para redes de clima obteniendo un coeficiente de correlación mínimo de 0.73 para la red de estados de viento.

La red de estados de viento es agrupada por medio de métodos de detección de comunidades, los cuales buscan conjuntar los nodos que interactúan más fuertemente entre ellos. Se aplicó el método de detección de comunidades de Girvan-Newman optimizando la modularidad a la red de estados de viento definida. Se encontraron tres comunidades que son propuestas a representar comportamientos del viento regional o local dependiendo de su alcance geográfico y la interacción entre los estados de la comunidad. Los estados de viento no detectados por la red de estados de viento son candidatos a estados de viento locales por su falta de interacción con los estados de viento del resto de las estaciones.

Se analizaron las comunidades encontradas y se observó que son capaces de detectar corrientes de viento presentes en la región: la comunidad 1 detecta una corriente Sureste, la comunidad 2 detecta un cambio de dirección en la corriente Sureste y la comunidad 3 detecta otra corriente con dirección Sureste. Las comunidades tienen temporalidades promedio distintas entre ellas, lo que muestra que logran captar fenómenos distintos: la comunidad 1 tiene un tiempo de 2.7 horas, la comunidad 2 tiene un tiempo de 7.1 horas y la comunidad 3 tiene un tiempo de 27.5 horas. El método de detección de comunidades muestra gran potencial para detectar patrones ocultos en el comportamiento del viento.

Índice general

Índice de figuras

Índice de cuadros

Índice de algoritmos

Capítulo 1

Introducción

1.1. Antecedentes

El uso y manipulación de diferentes fuentes energéticas ha permitido al ser humano estar en el nivel de desarrollo en el que se encuentra actualmente, sin embargo, el uso de hidrocarburos como principal combustible ha causado un impacto ambiental de importantes consecuencias. En un contexto mundial, la producción de energía continúa con una gran dependencia a combustibles fósiles. El 59.1 % de la producción energética mundial fue debido a este tipo de combustibles para el 2016 [6], sin embargo, las energías renovables ya producían un 14 % de la producción energética mundial durante ese mismo año (figura **1-1**) con un incremento del 3.6 % con respecto al año anterior. Actualmente existen acuerdos internacionales para reducir las emisiones de gases de efecto invernadero (GEI) provocados por el uso de combustibles fósiles con la finalidad de reducir los impactos del cambio climático. Incrementar el uso de energías renovables promueve un desarrollo sustentable además de que proporciona fuentes de energía no contaminante a regiones de difícil acceso [7].

En México durante el 2017 la producción de energía primaria por fuentes renovables fue de 665.26 PJ, que corresponden a un 9.5 % de la producción nacional. En la figura **1-2** se muestra la estructura de la producción de energía primaria en México para dicho año, teniendo un incremento con respecto al año anterior del 1.53 %. La energía eólica durante ese mismo año produjo el 0.54 % de la energía primaria del país, que corresponde a 38.23 PJ, lo que se puede considerar una baja participación de esta fuente de energía considerando su potencial. El incremento porcentual para la energía eólica con respecto al 2016 fue del

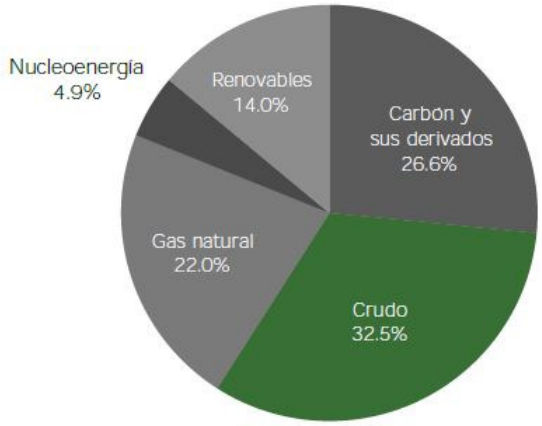


Figura **1-1**: Producción mundial de energía primaria para el 2016

2.33 %, lo que indica que junto con la energía solar (con un incremento porcentual de 36.68 % en relación al año anterior), es de las principales fuentes de energía renovable en crecimiento dentro del país [1].

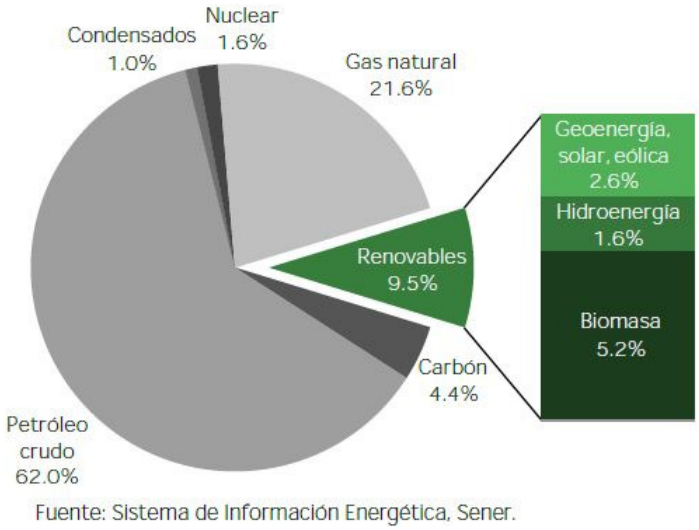


Figura **1-2**: Producción de energía primaria, 2016. Fuente [1]

Para el desarrollo de la energía eólica es importante generar modelos que permitan realizar predicciones cada vez más detalladas y que brinden mayor información de la dinámica del viento. De esta forma se pueden estudiar y reducir problemas de integración a la red eléctrica, las evaluaciones tecno-económicas resultan más precisas y fomentan una mayor confianza para la inversión y además permiten programar mejor los periodos de mantenimiento y el despacho de energía en los parques eólicos.

La metodología utilizada para la estimación del potencial eólico de una región consiste en un análisis estadístico de las velocidades del viento. Generalmente se hace el supuesto que las velocidades del viento tienen distribuciones estadísticas de tipo Rayleigh o Weibull. Sin embargo, existen regiones donde estas distribuciones estadísticas no representan correctamente las velocidades del viento como en el caso de La Ventosa, Oaxaca [8] y se deben realizar modelos más complejos. Esta metodología no hace uso de información de la dirección del viento, una pieza que se encuentra disponible en las estaciones anemométricas y que puede proporcionar un mejor conocimiento de la dinámica del viento.

Un estudio anterior [9, 10] define los *estados de viento* como una alternativa para realizar estimaciones del potencial por medio de una clasificación del viento con un mayor sentido físico. Se desarrolla además un modelo estocástico y se estima el potencial de generación de dos estaciones anemométricas en México . Este novedoso método brinda información acerca de la dinámica del viento y su potencial de generación, sin embargo, son análisis puntuales sin un alcance geográfico mayor.

En este trabajo se desarrolla una metodología para integrar el análisis de *estados de viento* de cuatro estaciones anemométricas ubicadas en Zacatecas y se genera una red de los estados de viento en la región para describir su dinámica. Tiene la finalidad de proporcionar un método alternativo a la tratamiento estadístico tradicional donde la dinámica del viento no es representada.

El objetivo de este trabajo no busca una descripción detallada de los fenómenos naturales y las dinámicas que se ven abordadas en la trama compleja del viento en la región. Más bien busca realizar una descripción que, aunque llega a tratar de describir la dinámica del viento, lo hace de una forma basada en datos reales y con algoritmos que no tomen muchos recursos computacionales de forma que puedan ser útiles para fines ingenieriles. El valor de este estudio recae en ser una nueva metodología de descripción de la dinámica del viento por medio de algoritmos computacionales que pueda ser funcional para evaluaciones de recurso eólico.

Conocer la naturaleza del viento se vuelve relevante al momento de estimar el potencial eólico de una región específica y es posible realizar estudios de relevancia para dicho fin que puedan mejorar el aprovechamiento de dicho recurso. Algunos ejemplos de este tipo de estudios son:

- Conocer los principales candidatos para instalar un parque eólico y evaluar en cuales es más conveniente profundizar en un proyecto.
- Una vez establecido el lugar permite conocer las variables necesarias para el diseño de las turbinas eólicas de forma que se seleccionen los perfiles y materiales convenientes a las cargas mecánicas y climáticas del lugar.
- Es posible seleccionar el sembrado óptimo de cada aerogenerador en una granja eólica

analizando la turbulencia provocada por la interacción del viento con los aerogeneradores.

- Ya instalados los aerogeneradores, conocer la naturaleza del viento con respecto a la producción de energía es un indicador del desempeño de las instalaciones y una nueva fuente de información para nuevas propuestas de mejora y decisiones de operación.
- Es posible analizar las variaciones climáticas al colocar aerogeneradores en un sitio y estimar si es realmente viable su instalación en términos ambientales [11].
- Tener modelos que puedan pronosticar el comportamiento del viento permite conocer la operación que se le debe proporcionar a la energía al incorporarla a la red eléctrica.

Estos y otro gran número de análisis son posibles conociendo el comportamiento del viento en la región de estudio. De ahí proviene la importancia de caracterizar la dinámica del viento en la región y la obtención de datos del viento.

1.2. Objetivos

1.2.1. Objetivo general

Generar un modelo en forma de red del viento usando la metodología de *estados de viento* para cuatro estaciones anemométricas ubicadas en el estado de Zacatecas.

1.2.2. Objetivos específicos

1. Comparar los algoritmos de agrupamiento de tipo K-Means y Mezcla Gaussiana para la determinación de estados de viento y seleccionar el algoritmo más adecuado a datos en múltiples estaciones metereológicas en una misma región geográfica.
2. Obtener indicadores para comparar los estados de viento de cuatro estaciones anemométricas diferentes.
3. Formar una red de estados de viento que modele el sistema de viento a escala regional.
4. Encontrar comunidades en los estados de viento de la red y clasificarlas en estados de viento regionales o locales.
5. Describir el comportamiento de las comunidades obtenidas.

1.3. Hipótesis

Es posible integrar un análisis de *estados de viento* de diferentes estaciones anemométricas para formar un modelo en red que permita obtener información de la dinámica del viento regional y local.

Capítulo 2

Marco Teórico

2.1. Recurso eólico

2.1.1. Propiedades del viento

El fenómeno básico que origina el viento es la diferencia de presión en masas de aire causadas por diferencias de temperatura. Aunque su fenómeno básico sea sencillo, el viento en el planeta es producto de una serie de variables que están estrechamente correlacionadas entre sí. Dependiendo de la escala geográfica en la que se está considerando hay diferentes fenómenos que influyen en su comportamiento, además de diferencias temporales. La estrecha correlación que existe entre estos fenómenos provoca que sea difícil generar un modelo del viento que funcione en cualquier escala, tanto geográfica como temporal, y es esta misma correlación la que hace de los vientos un fenómeno complejo.

De forma espacial, existen fenómenos que afectan los movimientos globales del viento y otros que se relacionan a una escala local. Existe una diferencia entre la energía del sol incidente debido a la variación de la latitud, este fenómeno afecta los vientos a escala global, como también lo hace el movimiento de rotación de la Tierra, que provoca la presencia de fuerzas de Coriolis. En una escala local la presencia de cuerpos de agua como océanos, la extención de terreno, la altitud del lugar, la vegetación y la interacción del terreno con la capa límite atmosférica son factores que influyen entre los vientos globales y los vientos locales [12].

Existen fenómenos más locales que también provocan movimientos de masas de aire. Algunos

ejemplos son el efecto valle-montaña y las brisas causadas por superficies tierra-mar. En ambos casos existe una diferencia de calentamiento en las superficies. En el efecto valle-montaña debido a la diferencia de altura y en el tierra-mar por las propiedades térmicas. Finalmente, los patrones de dirección del viento se ven modificados durante el día y la noche. Conocer los fenómenos locales es importante para la evaluación del potencial de un sitio en específico [13].

En cada lugar geográfico las condiciones de vientos son distintas por la combinación de los diferentes parámetros que influyen en su comportamiento. Además, como se mencionó antes, existen variaciones temporales del viento en un mismo sitio. Una clasificación de las diferencias temporales de viento se puede obtener de la práctica convencional, donde se dividen en cuatro categorías:

- Inter-anual: Escalas más allá del periodo de un año. Tienen un impacto a largo plazo en las variaciones del viento y puede presentar fenómenos como El niño o La niña que son perceptibles en escalas mayores.
- Anual: A lo largo del año existen periodicidades relacionadas a las estaciones o periodos mensuales.
- Diurno: Se refiere al periodo de tiempo de un día. Existen regiones donde la topografía u otros factores pueden causar variaciones considerables a lo largo del día y que dicho comportamiento se mantiene relativamente constante a lo largo del año.
- Corto plazo: Se refiere a ráfagas de viento y turbulencia. Estas variaciones ocurren en periodos de tiempo muy cortos, entre intervalos de pocos minutos.

2.2. Métodos de agrupamiento

El agrupamiento de datos es una técnica utilizada por diferentes disciplinas debido su gran utilidad práctica y de análisis. Su objetivo es clasificar grupos de una lista de datos donde los elementos pertenecientes tienen características similares.

Este tipo de técnicas es innata en los seres vivos y forman una primera herramienta para otros procesos cognitivos. De hecho, los seres vivos dependen de esta capacidad para sobrevivir y para entender su entorno. La capacidad de comunicarnos requiere un proceso de agrupamiento de objetos y se pueden desarrollar otros procesos de relaciones que resulten en un entendimiento más profundo del entorno. Por ejemplo, en el lenguaje asignamos significado a las palabras y las agrupamos en características específicas, que después se complementa al descubrir nuevas relaciones y de esa forma aprendemos. Relacionar fenómenos

separados que en realidad corresponden a uno mismo es una constante en importantes descubrimientos de la naturaleza del entorno, como la relación calor-trabajo, espacio-tiempo, presión-gas-temperatura, electricidad-magnetismo, por mencionar solo algunos ejemplos de variables físicas, pero conscientes de que no solo se delimitan a este tipo de datos.

Existe una amplia posibilidad de relacionar los objetos y agruparlos de diferente forma de acuerdo al principio inductivo. Como ejemplo de esta posibilidad se hará uso de una imagen poco convencional pero útil para este fin. En la figura **2-1** se observan diferentes posibilidades para agrupar personas, criaturas u objetos (datos) en función del objetivo que se busca. Por ejemplo, es posible agrupar a los elementos debido a su cercanía entre ellos, pero hacerlo de esta forma no nos brinda mucha información debido a que la distancia no parece ser un parámetro clave en la descripción general. Parece ser que una mejor forma de agrupar los elementos es en base a la actividad que realizan, de esta forma encontramos quienes están enfrente del escenario con las manos levantadas, quienes tocan algún instrumento, quienes están quemando una bruja, quienes compran recuerdos, quienes esperan el baño, quienes se pelean, etc. Sin embargo, se vuelve evidente que la posibilidad de agrupar los elementos es muy variada tanto en objetivo como en la información que dichos grupos brindan al panorama general.

Figura **2-1**: Portada del disco *Lucha interior* de la banda Fraghor

2.2.1. Consideraciones para un buen agrupamiento de datos

A pesar de que el objetivo básico del agrupamiento de datos es sencillo, en la literatura no existe una definición concreta de "grupo" o de "método de agrupamiento" (*clustering*). El problema ampliamente discutido de una definición única es que depende del objetivo del agrupamiento, de la subjetividad del observador y que en ocasiones está fuera de lugar intentar una definición de este tipo [14–17]. Otro acercamiento es dado por Kleinberg [18] donde demuestra la imposibilidad de satisfacer simultáneamente tres características que se buscan en el agrupamiento: *invariante en escala*, *riqueza* requerida de todas las particiones y *consistencia* en el encogimiento y estiramiento de las distancias individuales.

Para hacer un buen agrupamiento de los datos es recomendable realizar diferentes actividades que ayudan a centrar el análisis que se busca y los métodos que pueden ser empleados. Algunos puntos guía son:

1. Identificar los objetivos del agrupamiento.
2. Selección de las variables que brindan información esencial de los datos.
3. Identificar las características deseables de los grupos que se formen.
4. Revisión de los diferentes algoritmos de agrupamiento y las características de los grupos que selecciona.
5. Seleccionar el algoritmo de agrupamiento que se ajusta mejor a las características deseadas de los grupos.
6. Evaluar si el agrupamiento cumple con los objetivos iniciales de forma cuantitativa y cualitativa.

Al utilizar los métodos de agrupamiento comúnmente hay dos tipos de objetivos que se buscan, uno enfocado a obtener información que no se conocía anteriormente en los datos y que permiten conocer mejor los fenómenos que suceden, y otra enfocada a una clasificación que permita manejar información de forma más sencilla. De acuerdo a Hennig [16] se puede clasificar en dos objetivos del agrupamiento:

- **Realista:** El objetivo desde una perspectiva realista es agrupar para obtener información subyacente de los datos y crear nuevas hipótesis acerca de su «naturaleza». Lo que se refiere con «natural» es una realidad del observador independiente, es decir, una característica que está presente fuera de la subjetividad y que representa características fundamentales del fenómeno.

- **Constructivista:** El objetivo en una perspectiva constructivista es la separación de datos para una actividad práctica, donde no se busca conocer un comportamiento más fundamental que la de cumplir con una tarea específica. Puede ser la división de datos para una organización más eficiente.

Al seleccionar correctamente las variables de los datos a agrupar se obtiene mayor información acerca de los grupos que resultarán [14]. Existen una gran variedad de clasificaciones de una misma serie de datos, por ejemplo: clasificar libros usando como variables el contenido o el color. Al agruparlos por contenido es más útil ya que obtienes mayor información de las características propias de los libros, en cambio, al agruparlos por el color está basado en una característica física de la cual difícilmente obtienes información fundamental del contenido de los libros.

Debido a que el objetivo del agrupamiento y los datos varían en cada aplicación, es necesario identificar las características del grupo que se buscan. En base a este criterio se pueden identificar los tipos de algoritmos de agrupamiento que se acoplan mejor a las características deseadas. El objetivo final, como se mencionó anteriormente, es obtener un método que agrupe los datos con características similares entre ellos (en donde es importante definir el tipo de características que se buscan) y que sean distintos entre grupos (de acuerdo al objetivo que se busca).

2.3. Algoritmos de agrupamiento

Existe una diferencia entre métodos de agrupamiento y algoritmos de agrupamiento [19]. Los métodos de agrupamiento son la filosofía por las cuales se realiza el agrupamiento de los elementos, mientras que los algoritmos de agrupamiento son la serie de pasos que con la estructura de agrupamiento seleccionada realizan la separación de los elementos en grupos. Existe una amplia colección de algoritmos de agrupamiento con formas distintas de agrupar los elementos y, por lo tanto, con características específicas de los grupos que forma. Los problemas de agrupamiento se pueden dividir en dos grupos:

- Agrupamiento duro (*Hard Clustering*): Se refiere a que los elementos solamente pueden ser parte de un grupo. Una clasificación general de los métodos que hacen agrupamiento duro es:

  - Métodos jerárquicos: Este método forman jerarquías entre los grupos. Es decir, dividen los elementos en una serie de grupos anidados, encontrando nuevos grupos. Una característica de este agrupamiento es que puede ser representado por un dendograma. Existen dos tipos de algoritmos de este tipo:

    - Algoritmos aglomerativos: Estos algoritmos comienzan considerando a cada dato como un grupo propio y comienzan a acumular diferentes elementos hasta que se quedan con uno o con un número de grupos especificado.
    - Algoritmos divisivos: Estos algoritmos funcionan de forma inversa a los algoritmos aglomerativos. Consideran inicialmente un solo grupo y hacen una separación de éstos hasta que cada elemento se convierte en un grupo o hasta quedarse con un número de grupos especificado.

  - Métodos particionales: Los algoritmos particionales dividen los elementos en una sola partición y su representación no es mostrada por un dendograma debido a que no hay grupos anidados. Es decir, considera todos los elementos para hacer la partición en lugar de fraccionar grupos.

- Agrupamiento borroso (*Fuzzy clustering*): En el agrupamiento borroso los elementos pueden pertenecer a un grupo y otro. Lo que define la pertenencia está expresado en probabilidades. Los métodos de agrupamiento duro generalmente puden modificarse para poder realizar un agrupamiento borroso.

Esta es una clasificación de los métodos de agrupamiento y cada uno de éstos puede tener diferentes algoritmos de agrupamiento. En este trabajo se analizan principalmente dos algoritmos de agrupamiento que son ampliamente conocidos: k-Means y Mezcla Gaussiana.

2.3.1. K-Means

El algoritmo k-Means es uno de los algoritmos de agrupamiento más utilizados por su sencillez computacional. Es clasificado como un algoritmo particional y se basa en definir un centroide en cada grupo, que se va modificando conforme se agregan elementos al grupo hasta tener un criterio de convergencia para una función error definida. Este algoritmo de agrupamiento necesita conocer *a priori* el número de grupos en los que se va a agrupar los datos y se pueden utilizar diferentes definiciones para la similitud o disimilitud.

Sea D la lista de datos a agrupar y sean $C_i = \{C_1, C_2, ..., C_k\}$ los grupos, donde k es el numero de grupos dados *a priori*. La función error está definida por la ecuación 2-1 donde $\mu(C_i)$ es el centroide del grupo C_i y $d(\mathbf{x}, \mu(C_i))$ es la distancia entre el vector o elemento $\mathbf{x}$ y el centroide $\mu(C_i)$. Esta distancia entre vector y centroide generalmente se toma como una distancia euclidiana definida por la ecuación 2-2 aunque no es la única distancia que puede tomarse como criterio de disimilitud. Considerar disminuir la función error a partir de una definición de distancia se considera el principio inductivo del algoritmo de agrupamiento.

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} d(\mathbf{x}, \mu(C_i)) \tag{2-1}$$

$$d_{euc}(\mathbf{x}, \mathbf{y}) = \left[\sum_{j=1}^{d} (x_j - y_j)^2 \right]^{\frac{1}{2}} \tag{2-2}$$

En este tipo algoritmo se pueden considerar dos fases principales: la fase de inicio y la fase de iteración. En la fase de inicio se determina de forma aleatoria los puntos iniciales que representan los centroides de cada grupo. Esta fase es de importancia debido a que este algoritmo es sensible a los valores iniciales de los centroides. En la fase de iteración se realiza una comparación de los centroides con los elementos hasta que se satisface un argumento mínimo de cambio entre los centroides y los elementos. Para este punto ya todos los elementos se han agrupado a alguno de los centroides que representa cada grupo. El algoritmo 1 muestra este proceso para agrupar los elementos.

Es importante notar nuevamente que el algoritmo permite definir tanto el método aleatorio para seleccionar los valores iniciales de los centroides así como la distancia de comparación entre centroides y elementos. La definición de estos métodos modifica la forma en la que el algoritmo agrupará los elementos.

Algoritmo 1: Algoritmo k-means convencional. Modificado de [4]

Datos: Lista de datos D, Número de grupos k, Dimensión d de los datos
1. $\{C_1, C_2, ..., C_j\}$=Partición inicial de D.
2. **repetir**
   d_{ij}= Distancia entre el elemento i y el grupo j:
   n_{ij} =arg $\min_{1 \leq j \leq k} d_{ij}$:
hasta que *No hay cambios significativos en los centroides de los grupos*;
devolver Lista de grupos C con elementos agrupados de la lista de datos D

Algunas características de este algoritmo es que fue diseñado para trabajar con valores numéricos y no requiere mucha exigencia computacional. Una de las ventajas es que puede funcionar con listas de datos muy grandes debido a que el algoritmo no se vuelve más complejo al incrementar los elementos a agrupar. Los grupos que determina tienden a tener formas convexas a partir de los espacios tridimensionales, mientras que en espacios bidimensionales los grupos tienden a tener una división más rígida. Algunas de las desventajas es que el algoritmo es sensible a los valores iniciales de los centroides, dando la posibilidad de dar resultados distintos en diferentes corridas y finalmente, una de las principales desventajas es que es necesario proporcionar *a priori* el número k de grupos.

2.3.2. Mezcla Gaussiana

El algoritmo de Mezcla Gaussiana es un algoritmo *basado en el modelo*. Los algoritmos basados en el modelo hacen la suposición de que los datos están representados por una mezcla de distribuciones de probabilidad y para el caso particular del algoritmo de mezclas Gaussianas se considera que la distribución de probabilidad es de tipo normal. Los algoritmos basados en el modelo toman el nombre de *modelo* de las restricciones que cada distribución puede tener para definir sus parámetros. En estos algoritmos se busca encontrar los parámetros que definen cada distribución de acuerdo a la lista de datos observada. Una de las ventajas de utilizar una mezcla de Gaussianas es su habilidad de manejar grupos con diferente forma, orientación y volumen [20], además de que no necesita información *a priori* de los datos a agrupar.

Se considerará la metodología de Mecla Gaussiana Multivariada para describir el funcionamiento del algoritmo pero cabe aclarar que para el algoritmo utilizado se se utiliza la Inferencia Variacional en Mezclas Gaussianas. Si se desea se puede consultar a Bishop [5] para una descripción más detallada y con un análisis más profundo de ambas metodologías.

Mezcla Gaussiana Multivariada

En un modelo de mezcla Gausiana multivariada, se asume que los datos $D = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ en un espacio de d dimensiones provienen de un vector aleatorio con densidad mostrada en la ecuación 2-3, en donde a los valores π_k se les llama *coeficientes de mezclado*; tienen las propiedades $0 < \pi_k < 1$ y $\sum_{k=1}^{K} \pi_k = 1$, y $\mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma})$ denota la densidad de una distribución normal mostrada en la ecuación 2-4 para d dimensiones con un vector promedio μ y una matriz de covarianza $\mathbf{\Sigma}$.

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\mu_k, \mathbf{\Sigma}_k) \tag{2-3}$$

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

En la ecuación 2-3 se trata de una superposición lineal de las diferentes mezclas de Gaussianas en las que se van a ajustar a los datos. Observando las propiedades del término π_k se observa que representa la ponderación que cada Gaussiana tiene sobre todos los datos.

Para encontrar el valor de los parámetros de la distribución normal (y de cualquier distribución estadística) que permitan una máxima probabilidad de que el modelo producido se ajuste a los datos observados se introduce un método llamado *Máxima verosimilitud* y que

representa el principio inductivo en este algoritmo [17]. En el caso de una distribución normal se utiliza el logaritmo natural de la función de probabilidad (ecuación 2-5) debido a que es más sencillo manejar la distribución como superposiciones que como multiplicaciones ya que de esta forma se obtienen los parámetros para cada Gaussiana de forma independiente. La máxima verosimilitud se obtiene maximizando esta función.

$$\ln p(\mathbf{X}|\pi), \mu, \mathbf{\Sigma}) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \mathbf{\Sigma}_k) \tag{2-5}$$

Las ecuaciones 2-6, 2-7 y 2-8 se utilizan para encontrar una maximización de la verosimilitud, sin embargo, no es tan sencillo resolver dichas ecuaciones debido no corresponden a un sistema de ecuaciones cerrado. Esto es porque las variables dependen de la responsividad $\gamma(z_{nk})$ de una forma compleja mostrada por la ecuación 2-9.

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \tag{2-6}$$

$$\mathbf{\Sigma}_k = \frac{1}{N_k} \gamma(z_{nk})(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T \tag{2-7}$$

$$\pi_k = \frac{N_k}{N} \tag{2-8}$$

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \mathbf{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\mu_j, \mathbf{\Sigma}_j)} \tag{2-9}$$

Para resolver el sistema de ecuaciones se usa el algoritmo Expectativa-Maximización. El algoritmo Expectativa-Maximización se divide en dos pasos y se considera una variable latente que servirá para re-evaluar los parámetros de las mezclas Gaussianas. La variable latente en este caso se trata de la responsividad definida en la ecuación 2-9. El algoritmo 2

describe la Expectativa-Maximización para la mezcla de Gaussianas multivariadas:

Algoritmo 2: Algoritmo de Expectatividad-Maximización para parámetros de Mezcla Gaussiana Multivariada. Modificado de [5]

Datos: Lista de datos D, Dimensión d de los datos, medias de las distribuciones μ_k, Covarianzas de las distribuciones $\mathbf{\Sigma}_k$ y coeficientes de mezclado π_k

1. Inicializar los parámetros iniciales y calcular el valor inicial del logaritmo natural de la máxima verosimilitud

{**E** Es el paso de Expectativa }

2. **E**: Evaluar las responsabilidades usando los valores de parámetros actuales:

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \mathbf{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\mu_j, \mathbf{\Sigma}_j)}$$

{ **M** Es el paso de Maximización }

3. **M**: Reestimar los parámetros usando las actulaes responsividades.

$$\mu_k^{\text{nueva}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n$$

$$\mathbf{\Sigma}_k^{\text{nueva}} = \frac{1}{N_k} \gamma(z_{nk})(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T$$

$$\pi_k^{\text{nueva}} = \frac{N_k}{N}$$

donde

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

4. Evaluar el logaritmo natural de la máxima verosimilitud

$$\ln p(\mathbf{x}|\mu, \mathbf{\Sigma}, \pi) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \mathbf{\Sigma}_k) \right\}$$

repetir

| Paso 2

hasta que *Exista convergencia de los parámetros o del logaritmo natural de la máxima verosimilitud*;

2.4. Estados de viento

El concepto de *estados de viento* es reciente dentro de la literatura. Surge la idea a partir de buscar hacer una discretización de los datos de viento con un mayor significado físico y se va dando como consecuencia de las observaciones realizadas en datos reales de viento [9].

Los *estados de viento* buscan definir regiones que se presenten con mayor probabilidad en un espacio de velocidades definido por la ecuación 2-10, donde u_i es la magnitud de la velocidad del viento y θ_i es la dirección del viento en el i-ésimo punto. Estas zonas representan los puntos donde la velocidad y la dirección del viento tienen estados "naturales", refiriéndose a que se presentan por condiciones geográficas y climáticas propias del viento en la zona. De esta forma, la discretización de los datos no es de forma arbitraria, sino que busca seguir las características propias del viento.

$$\mathbf{u}_i = u_i(\cos\theta_i, \sin\theta_i) \tag{2-10}$$

Los autores del concepto definen *estado de viento* como: "una región en el espacio fase de las velocidades que contiene las velocidades accesibles del viento teniendo una distribución de probabilidad que la caracteriza como un grupo" [9]. Se refiere como espacio fase al plano vectorial de velocidades de viento accesibles.

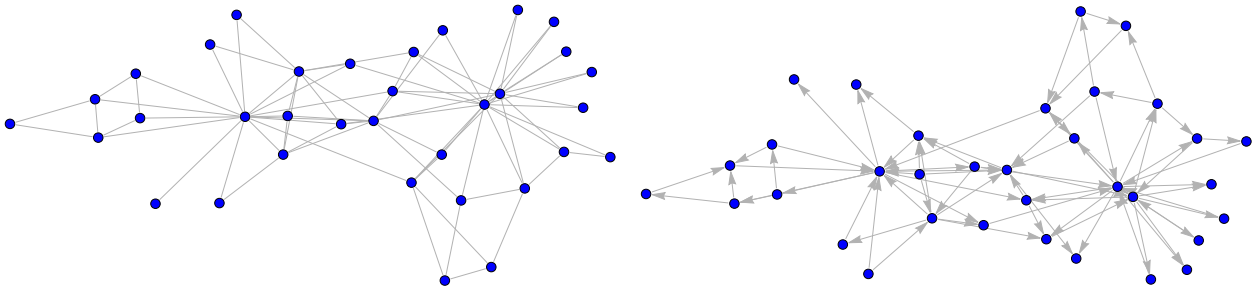
A partir de la representación del viento en el plano vectorial o plano de velocidades se utilizan algoritmos de agrupamiento para seleccionar los puntos que tengan un comportamiento similar entre ellos. Cada uno de estos grupos se puede caracterizar por una función de probabilidad. Seleccionar el algoritmo de agrupamiento toma importancia debido a que en función de la forma en la que realiza el agrupamiento, los *estados de viento* tendrán características distintas, así como el número de grupos puede ser diferente.

2.5. Análisis de redes

Las redes son sistemas en donde los elementos que la conforman (llamados nodos o vértices) están conectados por bordes que enlazan los diferentes elementos en una cierta topología o estructura [21]. Actualmente observamos e interactuamos en este tipo de sistemas de forma cotidiana y han tomado tal relevancia que ya se habla de un campo de la ciencia enfocada al estudio de estas estructuras denomidado *ciencia de redes* [22]. Su importancia recae en que permiten el estudio de sistemas que dejan de considerar un análisis reduccionista y en su lugar busca considerar la identidad de los elementos y los patrones de interacción entre ellos, lo que caracteriza a un sistema complejo [23].

Existen diferentes tipos de redes en función del tipo de relaciones de los bordes y características de los nodos.

- **Redes simples**: Redes donde los nodos son simples y los bordes no tienen dirección .
- **Redes direccionadas**: Los nodos son simples y los bordes tienen una dirección de un elemento a otro. Este tipo de redes lleva cierta jerarquía y también pueden ser representadas por dendogramas. Las redes que no tienen esta característica se les llama redes no direccionadas .
- **Redes ponderadas**: En este tipo de redes se hace diferencia entre importancia de los nodos o de los bordes, por lo que se les agrega una ponderación.



(a) Ejemplo de una red simple. (b) Ejemplo de una red direccionada.

Figura **2-2**: Ejemplos de las estructuras de red.

Existen otra clasificación de las redes de acuerdo a la teoría por las cuales son modeladas:

- **Teoría de grafos aleatorios**: La teoría de grafos aelatorios es introducida por Paul Erdös, el matemático más relacionado y uno de sus colaboradores Alfréd Rényi y tiene como fundamento el uso de métodos probabilísticos para la teoría de grafos.
- **Redes de mundo pequeño**: Las redes de mundo pequeño provienen de estudiar redes en donde se describe el fenómeno en que la distancia entre dos nodos es relativamente corto con respecto a toda la red. Uno de los ejemplos más característicos de este tipo de redes es el concepto de "seis grados de separación" de Milgram [24] donde los grados de separación no necesariamente se refieren a distancias euclidianas. Es posible utilizar la teoría de grafos aleatorios para generar redes de mundo pequeño agregando la restricción de una distancia entre nodos que sea relativamente pequeña. Un modelo típico de estas redes es el de Watts-Strogatz [25].

- **Redes libres de escala**: Las redes de escala libre se desarrollan para realizar redes de datos reales y descubrir que las interacciones entre los nodos no corresponden del todo a las teorías ya realizadas, en específico, que la distribución de los grados de los nodos no siguen una ley de potencias. El modelo más utilizado de este tipo es el de Barabási-Albert [26].

2.5.1. Conceptos básicos

A partir de la teoría de grafos se utilizan diferentes conceptos para describir y modelar las redes.

- **Grado**: El concepto de grado es de carácter fundamental para la descripción de redes. El grado de un nodo i en una red no direccionada se define como el número k_i de sus bordes existentes, es decir, el número de bordes que salen del nodo. Se define también el *grado promedio* de una red como el promedio del grado de todos los nodos de la red y se denota como $\langle k \rangle$.

- **Asortatividad-Disasortatividad**: La asortatividad se refiere a la forma en la que los nodos se relacionan entre ellos. De la definición de grado en una red, se obtiene que hay nodos con alto grado (aquellos que tienen una gran número de bordes) y nodos de bajo grado (aquellos con un menor número de bordes), la asortatividad se refiere a la tendencia en la red en que nodos de alto grado se unen con nodos de alto grado, mientras que la disasortatividad es el caso contrario, nodos de alto grado que se unen a nodos de bajo grado.

  Formalmente se define de acuerdo a la ecuación 2-11 donde j_i, k_i son los grados de los vértices al final del i-ésimo borde, con $i = 1, ..., M$ [27].

$$r = \frac{M^{-1}\sum_i j_i k_i - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1}\sum \frac{1}{2}(j_i + k_i)]^2} \tag{2-11}$$

  Donde $r \in [-1, 1]$. Si $r > 0$ la red es asortativa mientras que si $r < 0$ la red es disasortativa. Si $r = 0$ los grados de los nodos en la red están en promedio sin correlación.

- **Coeficiente de agrupamiento**: Existe la posibilidad que dos nodos conectados tengan en común un nodo conectado entre ellos. Dicho de otra forma, es posible que en una red entre amigos, dos amigos de alguien puedan o no compartir una amistad entre ellos. Este concepto se conoce en redes como agrupamiento (notase que es algo distinto al concepto de agrupamiento de datos).

En una red, sea i el nodo con k_i bordes conectando a otros k_i nodos, que se les denomina *vecinos* del nodo i y sea E_i el número de bordes existentes entre estos k_i nodos. Se define un radio de actuales y posibles números de bordes en el grupo de los k_i nodos y se define como *coeficiente de agrupamiento* del nodo i y se denota como C_i (ecuación 2-12).

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{2-12}$$

Donde $0 \leq C_i \leq 1$, si $C_i = 0$ significa que todos los vecinos del nodo i están desconectados de él mientras que si $C_i = 1$ todos los nodos vecinos del nodo i están conectados unos a otros (formando un grafo completo). Se puede definir también el *coeficiente de agrupamiento promedio* como el promedio del coeficiente de agrupamiento de todos los nodos de la red y se denota como $\langle C \rangle$.

- **Densidad de bordes**: La densidad de bordes de una red es un concepto que ayuda a la determinación de la significancia en los elementos de una red. Es el promedio del grado de la red $\langle k \rangle$ dividido entre el número total de bordes en la misma red $|E|$. Formalmente se describe en la ecuación 2-13.

$$\rho = \frac{|E|}{\binom{N}{2}} = \frac{\langle k \rangle}{N} \tag{2-13}$$

2.5.2. Estructura de comunidades

Mi generación de la universidad estaba integrada por 25 personas, una generación pequeña si se compara con otras licenciaturas y otras facultades. Sin embargo, a pesar del tamaño relativamente corto de la generación existían pequeños grupos de personas que interactuaban más frecuentemente, hacían equipos de forma más regular y tenían intereses similares. Por ejemplo, existían "las niñas" que era un grupo caracterizado porque todas eran mujeres aunque no incluía a todas las mujeres de la generación, existían "los populares" caracterizados por hacer planes secretos dando una sensación de exclusividad, estaban también "los geniales" que era un grupo que hacía salidas en bicicleta, entre otros. Esta característica de grupos dentro de una red o *comunidades* (figura **2-3**) no solamente se limita a fenómenos sociales, sino que tiene ejemplos también otras áreas como la biología [28], ciencias computacionales, ingeniería, economía [29], política y para buscar patrones emergentes en datos de clima [30–35].

Las comunidades se pueden definir como "un grupo de nodos que son densamente conectados entre ellos pero conectados de forma dispersa con otros grupos densos de la red" [36], una definición similar a la de agrupamiento de datos. Las comunidades en ocasiones también son

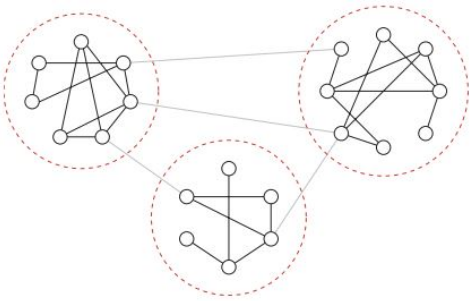


Figura **2-3**: Estructura básica de las comunidades en una red. Fuente [2]

llamadas *módulos* o *grupos* (término que se evitará a fin de no confundirlo con el agrupamiento de datos) y "son grupos de vértices donde la probabilidad de compartir propiedades y/o asumir roles similares en el grafo" [37]. La segunda definición amplía el concepto de comunidades al determinar que no se basan únicamente en la densidad de las conexiones sino que deben tener características similares en el comportamiento de la red, de esta forma se esperaría que las comunidades puedan dar información acerca de la forma en la que los elementos interactúan en la red y la función que realizan en ella.

2.5.3. Métodos de detección de comunidades

Se puede entender la detección de comunidades como métodos de agrupamiento aplicados a grafos o redes. De esta forma, existe también una colección de métodos y algoritmos que encuentran la comunidades en los grafos y que de hecho muchos de ellos se comparten entre el agrupamiento de datos y la detección de comunidades. Una clasificación de los métodos tradicionales [37] para encontrar comunidades es:

- Partición del grafo: Buscan dividir el grafo en vértices de g comunidades de tamaño predefinido, de tal forma que el número de bordes entre los grupos sea mínimo. Se le denomina *tamaño de bordes* al número de bordes entre las comunidades. En la figura **2-4** se muestra una partición del grafo con $g = 2$ y comunidades de igual tamaño. Este método requiere información *a priori* del número de comunidades que se buscan.

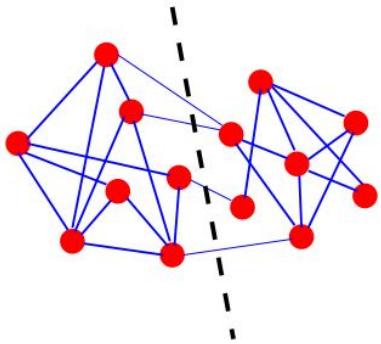


Figura **2-4**: Ejemplo de una partición de grafo. Fuente [3]

- Jerárquicos: Como en el agrupamiento de datos, los métodos jerárquicos para encontrar comunidades se basa en un ordenamiento con la finalidad de encontrar varios niveles de agrupamiento de los vértices. Para este método se debe definir una medida de similaridad entre los vértices. Este método puede ser dividido en dos algoritmos: Aglomerativos y Divisivos, con definiciones similares al discutido para el agrupamiento de datos.

- Particional: En este método el número de comunidades es pre-asignada por un número k y los elementos se colocan en una medida espacial, de forma que cada vértice es un punto con una distancia de medida (una medida de la disimilaridad entre vértices) a definir entre los pares de puntos en el espacio. El objetivo del método es separar los puntos en k grupos de forma que se minimice/maximice una función basada en las distancias entre puntos y/o de los puntos a centroides. Algunos ejemplos de funciones utilizadas son: Minimización de k grupos, suma de k grupos, k centros, k medias y k-Means para redes.

Algoritmo de Girvan y Newman basado en la modularidad

Uno de los algoritmos más utilizados para encontrar comunidades es el propuesto por Girvan y Newman [38] que es un algoritmo divisivo. A partir de un ordenamiento de la red en un orden jerárquico resultando en un dendograma, es posible dividir la red en los grupos anidados que se forman, lo que formaría las comunidades. Sin embargo, una de las preguntas es: ¿qué criterio usar para cortar la jerarquía?. Este algoritmo se enfoca en detectar los bordes que son menos centrales a las comunidades, es decir, los bordes que sirven de conexión entre comunidades o que están entre ellas.

Se utiliza un concepto definido como centralidad de la intermedialidad (*betweenness centrality*) que se define sobre un vértice k y se refiere al número de caminos más cortos entre los vértices i y j que pasan por el punto k [39]. Esta medida describe la influencia que tiene

un nodo en el flujo de información entre otros nodos o, en un ejemplo de una red social, es la influencia que tiene Irving (nodo k) en la comunicación entre Luis (nodo i) y Carlos (nodo j). La generalización de este concepto se puede denominar la intermedialidad de un borde (*edge betweenness*) como el número de caminos más cortos entre pares de vértices que pasan por él, es decir, ahora en lugar de tomar los vértices se toman un canal de conexión específico para medir la influencia que tiene en el flujo de información. En caso de que se tenga más de un camino corto entre pares de vértices, cada camino toma una ponderación igual de manera que la ponderación total sea la unidad.

En la definición de intermedialidad se considera "el camino más corto" como una variable del parámetro, sin embargo esta frase queda ambigua al no definirla con precisión. De hecho, puede existir más de un concepto de intermedialidad de acuerdo a la definición que se tome como medida de los caminos entre vértices, en este trabajo se mencionan tres definiciones pero pueden existir otras:

1. **Intermedialidad de camino más corto**: Se refiere a una medida geodésica, que es la generalización de una línea recta en espacios curvos.
2. **Intermedialidad de camino aleatorio**: Es una medida considerando ya no una geodesia sino el número neto de veces que un caminante aleatorio pase entre un par de vértices por un borde en particular y la suma de todos los pares de vértices.
3. **Intermedialidad de flujo de corriente**: Este concepto se basa de un análisis del flujo de corriente en un circuito eléctrico. Suponiendo que se coloca una resistencia en lugar de los bordes, la corriente eléctrica tenderá a seguir el camino con la menor resistencia. Esta medida se puede obtener utilizando las leyes de Kirchhoff.

El algoritmo Girvan-Newman para comunidades se puede observar en el algoritmo 3. Los bordes con mayor intermedialidad son aquellos que están entre las comunidades o dicho de otra forma, los que sirven de conexión entre comunidades. Es necesario el paso 3. debido a que al remover el borde la estructura de la red se ve modificada por lo que la intermedialidad de los bordes también cambia, asegurando siempre que el borde con mayor influencia en las redes sea el que se remueva.

Algoritmo 3: Algoritmo de Girvan y Newman para encontrar comunidades en una red.

1. Calcular la intermedialidad de todos los bordes de la red.
2. Remover el borde con la mayor intermedialidad.
3. Recalcular la intermedialidad de todos los bordes efectados por la remoción.
4. **repetir**
   | Paso 2. Paso 3.

hasta que *No queden bordes*;

Newman y Girvan para complementar su trabajo definen un nuevo concepto llamado modularidad (*modularity*) que funciona como una medida de la calidad de la división en comunidades [40] y está basado en la medida de asortatividad [27]. La modularidad compara la división de comunidades realizada por cualquier algoritmo con un modelo nulo (*null model*) que es una red sin comunidades. Se define de acuerdo a la ecuación 2-14 donde M es el número total de bordes de la red, A es la matriz de adyacencia, P_{ij} representa el número esperado de bordes entre los vértices i y j en el modelo nulo y la función δ es 1 si los vértices i y j están en la misma comunidad y 0 en caso contrario.

$$Q = \frac{1}{2M} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j) \tag{2-14}$$

En caso de que las comunidades que forma el algoritmo no sean muy diferentes a las que se pueden formar de forma aleatoria la modularidad será baja, en cambio cuando las comunidades que tengan una marcada estructura en conjunto tendrán alta modularidad. En el algortimo 3 se puede recalcular la modularidad conforme se retiran los bordes con la mayor intermedialidad y se espera que los picos que presente este valor indicarán la mejor división de las comunidades.

Capítulo 3

Metodología

En este capítulo se describirá la metodología usada para encontrar una correlación entre los estados de viento de las diferentes estaciones anemométricas. En forma de listado se muestran los pasos realizados y después se describirán a detalle:

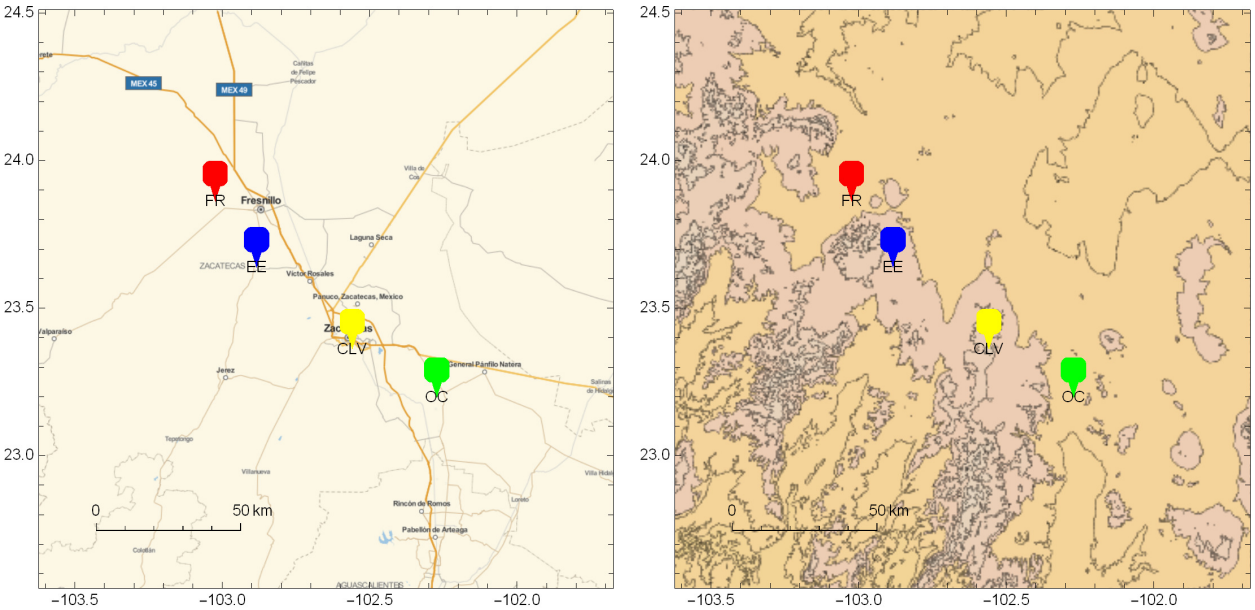
1. Definición de la región de estudio y obtención de los datos históricos.
2. Definición de los estados de viento de cada estación anemométrica usando los algoritmos de agrupamiento Mezcla Gaussiana y K-Means y su comparación entre ellos.
3. Selección del mejor algoritmo de agrupamiento para los objetivos buscados.
4. Descripción de la sincronización de eventos.
5. Formación de una red de estados de viento y determinación de comunidades.

El análisis de los datos realizados durante este trabajo se hizo utilizando el software *Wolfram Mathematica* del cual se cuenta con licencia de estudiante.

3.1. Región de estudio

Se definió el área de estudio del presente trabajo después de localizar una región en el estado de Zacatecas donde se encuentran cuatro estaciones anemométricas ubicadas en los

municipios de Fresnillo, Gral. Enrique Estrada, Zacatecas y Ojo Caliente. En la figura **3-1**a se muestra la ubicación geográfica de las cuatro estaciones anemométricas dentro del estado de Zacatecas y en la figura **3-1**b se muestra el relieve topográfico de la misma zona. Para determinar este conjunto de estaciones se consideró una ubicación geográfica en línea recta, una relativa cercanía entre ellas, para facilitar la interpretación de los resultados, y una orografía del terreno compleja que permita encontrar patrones no obvios.



(a) Ubicación geográfica de las estaciones.

(b) Relieve topográfico de la región.

Figura **3-1**: Ubicación geográfica y relieve topográfico de las cuatro estaciones anemométricas.

Existe un interés comercial en esta zona por sus velocidades de viento para la generación de energía eólica. Aquí se encuentra el parque eólico "La Bufa" ubicada en el Cerro La Virgen, a poca distancia de donde se encuentra la estación anemométrica que se considera para este estudio. Este interés se basa en una comprobada factibilidad económica, siendo un indicador que los vientos de la región son aptos para la generación de energía eólica y que presentan características interesantes para el presente estudio.

Se obtuvieron datos históricos de velocidad y dirección de viento de las estaciones anemométricas. Es importante considerar solo estaciones anemométricas que cuenten con ambas piezas de información ya que, es fundamental para hacer un análisis de estados de viento. Los datos históricos obtenidos corresponden a mediciones con resolución temporal de promedios de 10 minutos durante el año 2011. Es importante aclarar que para la estación Ojo Caliente no se pudo encontrar la información de los primeros dos meses (enero y febrero), por lo que durante el trabajo se tuvieron que ajustar los datos temporales de esta estación que se explicarán a

detalle en la sección 3.4.

En el cuadro **3-1** se muestran algunas características de cada estación anemométrica. Para el estudio se usa una altura de 50 metros sobre el nivel del suelo. Debido a que no todas las estaciones cuentan con mediciones a esta altura se hizo una extrapolación logarítmica para las velocidades de viento tal como se muestra en la ecuación 3-1, donde z_r es la altura de referencia, $U(z_r)$ es la velocidad de viento a la altura de referencia, z_0 es la longitud de la rugosidad de la superficie que depende del tipo de terreno. Se considera la dirección del viento de la altura de medición más cercana a 50 metros debido a que no se encontró forma de extrapolar dicha información.

Cuadro **3-1**: Características principales de las estaciones anemométricas

Estación	Ubicación	Alturas de medición* [m]	Tipo de terreno	z_0 [mm] [13]
Fresnillo (FR)	N: 23°11′57″, O: 103°1′26″	**20**	Cultivo agrícola	50
Cerro La Virgen (CLV)	N: 22°44′14″, O: 102°33′33″	20-**40**	Pedregoso	10
Enrique Estrada (EE)	N: 22°59′36″, O: 102°53′2″	30-**50**	Cultivo agrícola	50**
Ojo Caliente (OC)	N: 22°35′12″, O: 102°16′21″	**50**-80	Cultivo agrícola	50**

*Se denotan en negro las alturas de donde se extraen los datos.

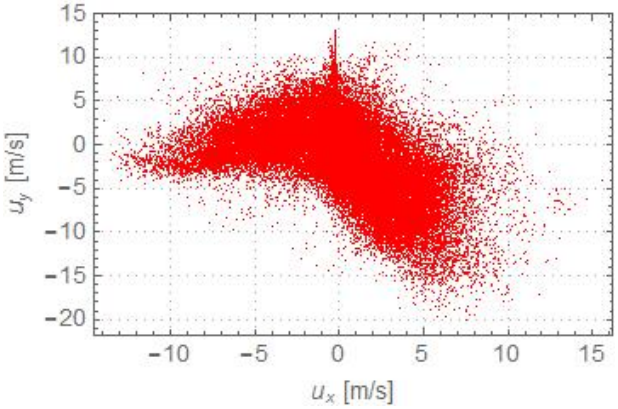
**No aplican en la extrapolación debido a que se toman los datos directo de la altura de medición.

$$U(z) = U(z_r) \ln\left(\frac{z}{z_0}\right) / \ln\left(\frac{z_r}{z_0}\right) \tag{3-1}$$

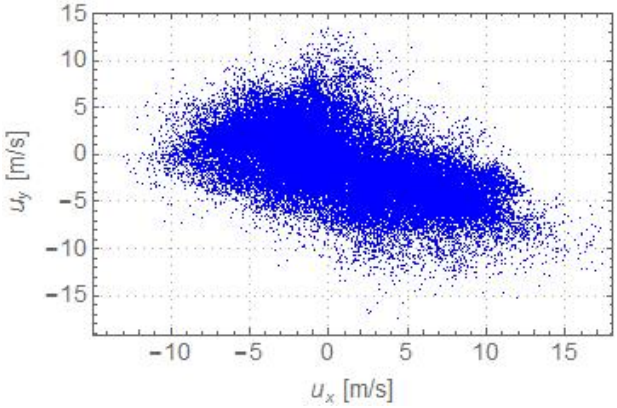
3.2. Plano de velocidades de viento de las cuatro estaciones

Una vez obtenidos y extrapolados los datos a una altura de 50 metros se generó el vector de velocidades para cada estación anemométrica de acuerdo a la ecuación 2-10. A partir del vector de velocidades se realizó el plano de velocidades de viento y un diagrama de densidad de dicho espacio para cada estación. En la figura **3-2** se muestran los planos de velocidades de cada una de las estaciones y en la figura **3-3** los histogramas suavizados de densidades.

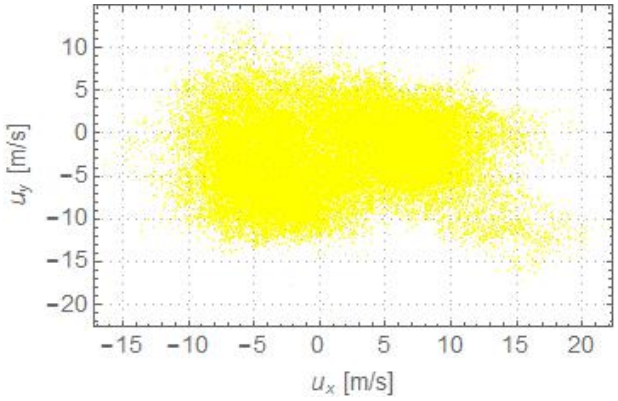
El comportamiento ideal en el plano de velocidades es que alcance regiones de altas velocidades en una dirección preferencial, ya que esto implica que hay corrientes de viento donde no es necesario cambiar la posición del aerogenerador. En las cuatro estaciones se observa la presencia de una corriente en dirección Noroeste y Sureste al observar los alcances que tienen los estados en dichas direcciones.



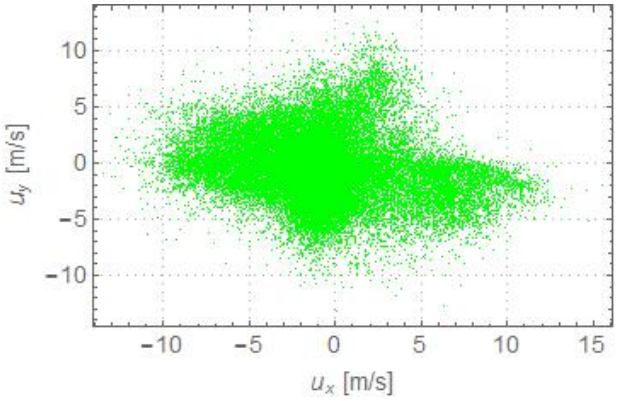
(a) Plano de velocidades de la estación Fresnillo (FR)



(b) Plano de velocidades de la estación Enrique Estrada (EE)

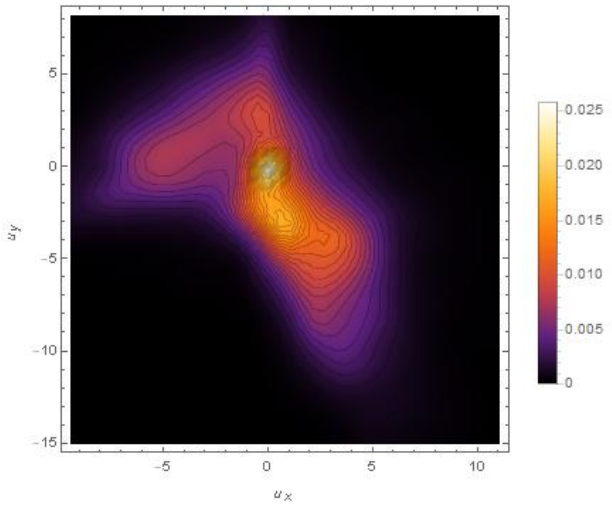


(c) Plano de velocidades de la estación Cerro La Virgen (CLV)

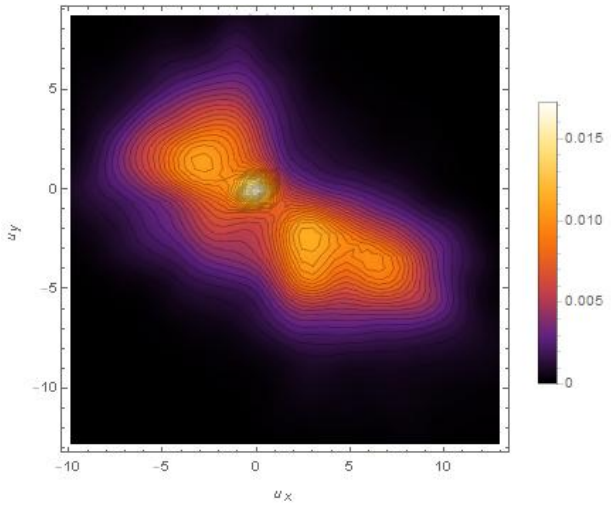


(d) Plano de velocidades de la estación Ojo Caliente (OC)

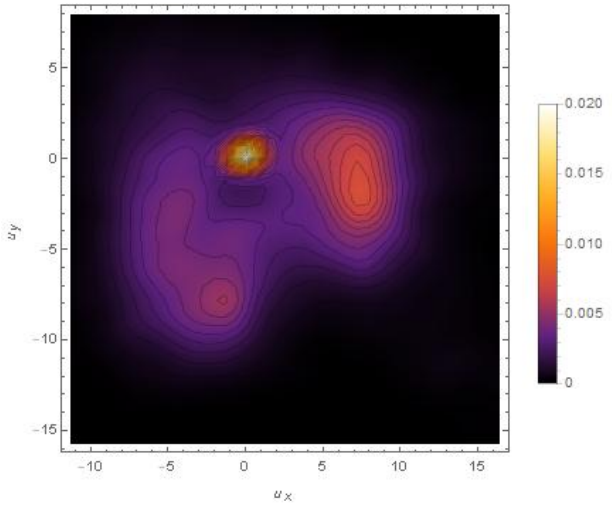
Figura **3-2**: Planos de velocidades de viento de las cuatro estaciones anemométricas.



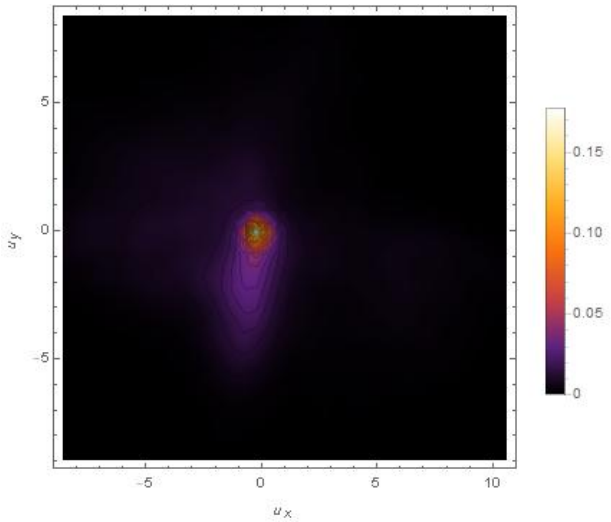
(a) Histograma suavizado de densidades de velocidad en la estación Fresnillo (FR)



(b) Histograma suavizado de densidades de velocidad en la estación Enrique Estrada (EE)



(c) Histograma suavizado de densidades de velocidad en la estación Cerro La Virgen (CLV)

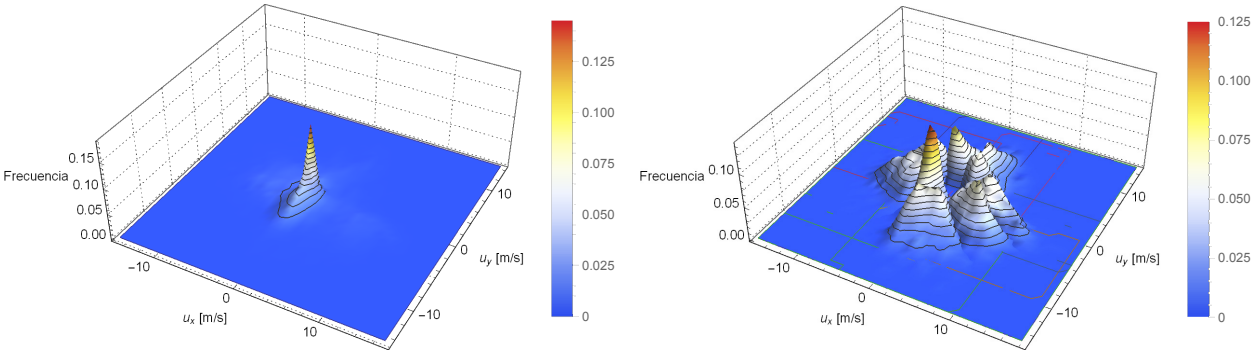


(d) Histograma suavizado de densidades de velocidad en la estación Ojo Caliente (OC)

Figura **3-3**: Histograma suavizado de densidades de las velocidades de veinto para las cuatro estaciones.

Sánchez-Pérez *et. al* [9] utilizan la información de la figura **3-3** para determinar el número de estados de viento al identificar y contar las regiones con mayor densidad. Sin embargo, para el propósito de este trabajo este procedimiento no satisface las características deseadas para dicho agrupamiento. Debido a que se busca definir los vientos regionales y locales de cada estación, las regiones en el plano de velocidades con mayor densidad representan únicamente "macro-estados" y desprecia posibles estados locales, despreciando los estados de viento debidos a fenómenos locales. En este trabajo se busca determinar ambos tipos de estados, los locales y los regionales. Por lo que se busca que los estados de viento en cada estación sean de un número mayor a los que se pueden identificar en los histogramas suavizados de densidad.

para ejemplificar este caso se usará la estación Ojo Caliente. En la figura **3-3**d se muestra su histograma suavizado del plano de velocidades, en donde se observa una región de frecuencia dominante. La presencia de esta región se observa también en el diagrama de densidades lo cual impide observar otros grupos que pueden estar presentes (figura **3-4**a). En la figura **3-4**b se muestra un diagrama donde se ha suprimido este macro-estado y donde ahora es posible observar nuevos grupos, que aunque no son los grupos dominantes se observa que tienen propiedades distintas que pudieran representar fenómenos locales.



(a) Histograma de densidades para la estación Ojo Caliente.

(b) Histograma de densidades sin el grupo dominante para la estación Ojo Caliente

Figura **3-4**: Histogramas de densidades para la estación Ojo Caliente con y sin el grupo dominante.

3.3. Comparación de k-Means y Mezcla Gaussiana para determinar estados de viento

Es importante enumerar las características deseadas de los estados de viento ya que con base a éstas se comparan dos algoritmos que se considera pueden definir los estados de viento de las cuatro estaciones anemométricas, Mezcla Gaussiana y k-Means:

1. Los estados de viento formados deben de ser representados por una función de probabilidad Gaussiana: Es deseable esta característica ya que permite tener una definición formal del estado de viento a partir de una expresión matemática bien definida e incluirla en modelos de predicción de viento.

2. El numero de estados de viento no debe ser muy grande pero tampoco muy pequeño: Como se mencionó anteriormente, es recomendable considerar un número medio de estados de viento para que cada comportamiento pueda ser representado.

3. La disimilaridad dentro de los estados de viento debe ser baja: Esto quiere decir que los elementos del estado se viento tengan un comportamiento similar entre ellos.

4. La disimilaridad entre estados de viento debe ser alta: Los estados de viento de diferentes grupos deben tener características diferentes entre ellos.

5. Los estados de viento deben corresponder a áreas en el espacio de los datos con alta densidad: Los estados de viento deben captar los fenómenos que son más influyentes en el comportamiento del viento, lo que quiere decir tomar las regiones con mayor densidad (sin contradecir el punto 2).

6. Los estados de viento deben ser caracterizados por un número pequeño de variables: Es decir, que sean representados por una función de densidad de probabilidad que no requiera de muchos parámetros.

Debido a que el algoritmo de Mezcla Gaussiana no requiere un conocimiento *a priori* del número de grupos y para hacer una comparación más directa entre los algoritmos se dejó que éste definiera el número K de grupos y ese mismo número K fue utilizado en el algoritmo k-Means. Para el algoritmo de Mezcla Gaussiana se seleccionó como medida de disimilaridad la distancia euclidiana definida por la ecuación 2-2 y como parámetro extra disponible en el software Mathematica, se buscó priorizar la calidad de los grupos sobre el tiempo de cómputo. En el cuadro **3-2** se muestran el número de estados de viento encontrados. Considerando las características deseadas se encuentra que el número de estados es bueno para los fines de este trabajo.

Cuadro **3-2**: Número de estados de viento

Estación	No. de estados de viento
FR	7
EE	6
CLV	8
OC	9

3.4. Sincronización de eventos

Antes de comparar los algoritmos de agrupamiento y de generar la red de estados de viento es necesario determinar un parámetro que indique que tan parecidos son dos estados entre sí. Para este fin se utiliza el coeficiente de correlación que por su definición determina el grado de correlación que existe entre dos variables. En este caso se denotará el coeficiente de correlación como c para evitar confusión con la asortatividad de una red denotada como r.

Los estados de viento pueden ser representados como una cadena de eventos. Esta cadena de eventos tiene la característica de indicar los momentos en los que un estado de viento está presente a lo largo del tiempo. Con base a esto se puede definir una correlación que esté basada en la comparación de eventos de los estados de viento.

Quian Quiroga *et. al.* definen una función de correlación que tiene como objetivo capturar la *sincronización de eventos* [41]. Una ventaja de este método es que también permite obtener información a diferentes tiempos de desfase. Este parámetro puede funcionar entonces para la comparación entre dos estados de viento de una misma estación anemométrica obtenidos con dos algoritmos distintos y también para la comparación de dos estados de viento de estaciones distintas que puedan tener un tiempo de desfase entre ellas.

Sean dos series de tiempo discretas y medidas en forma simultánea x y y, se definen las cadenas de eventos t_i^x y $t_{i+\tau}^y$ $(i = 1, ..., N)$ donde $\pm\tau$ es el tiempo de desfase entre ambas cadenas. Se denota como $c^\tau(x|y)$ como el número de veces que un evento en la cadena x sucede en la cadena y con el desfase τ y se obtiene a partir de la ecuación 3-2.

$$c^\tau(x|y) = \sum_{i=1}^{N} J_{i,i+\tau}^\tau \tag{3-2}$$

Donde a cada tiempo de desfase τ se cumple:

$$J_{i,i+\tau}^\tau = \begin{cases} 1 & \text{si} \quad t_i^x = t_{i+\tau}^y \\ -1 & \text{si} \quad t_i^x \neq t_{i+\tau}^y \end{cases} \tag{3-3}$$

Con base a esta definición, se define el coeficiente de correlación para cada tiempo de desfase τ normalizando con el número de elementos en la serie de tiempo N (ecuación 3-4) con $C(x|y) \in [-1, 1]$, donde si $C(x|y) = 1$ quiere decir que están perfectamente correlacionados, $C(x|y) = -1$ están inversamente correlacionados (cuando un elemento se presenta el otro no) y $C(x|y) = 0$ quiere decir que son perfectamente no correlacionados.

$$C^\tau(x|y) = \frac{c^\tau(x|y)}{N} \tag{3-4}$$

En el caso de los estados de viento esta metodología puede funcionar para comparar tanto los estados resultantes de diferentes algoritmos de agrupamiento como los estados de viento de diferentes estaciones anemométricas. Se toma la resolución mínima de los datos recolectados, que es de promedios cada 10 minutos, para definir los pasos con que se incrementa τ. De esta forma para cada unidad de incremento de τ se incrementan 10 minutos en la cadena de eventos de viento.

Debido a que no se lograron obtener los datos de los dos primeros meses de la estación Ojo Caliente se consideró que para dicha estación solament se usan los datos de los últimos 10 meses ($N = 44064$ datos en total), mientras que para el resto de las estaciones se usan los datos de todo el año ($N = 52560$ datos en total).

Para la comparación entre algoritmos de agrupamiento se utiliza este coeficiente de correlación con $\tau = 0$, es decir la cadena de eventos original, ya que se busca tener una comparación de lo que realizan los algoritmos de agrupamiento. Por otro lado, para la comparación de estados de viento entre diferentes estaciones anemométricas se utiliza el coeficiente de correlación con $\tau \neq 0$. Para definir a que tiempo de desfase se realizarán estas comparaciones se determina que sea para el tiempo de desfase donde la correlación sea máxima. En un sentido físico es como si se tomaran fotografías del sistema de vientos a diferentes tiempos y se superpusieran para formar un nuevo sistema. Se hace de esta forma porque se observa que hay estados de viento en dos estaciones que pueden corresponder a uno solo pero que tardan un cierto tiempo en moverse de una estación anemométrica a otra. Si se realiza la red en tiempos de desfase cero no se podrían percibir este movimiento de los estados de viento. En la figura **3-5** se muestra un ejemplo de dos estados de viento, para los que sus cadenas de eventos son muy similares, pero que están desfasadas un cierto tiempo. Los tiempos de desfase máximos encontrados en para las estaciones anemométricas fueron en orden de minutos a algunos días, pudiendo representarse en una escala de horas.

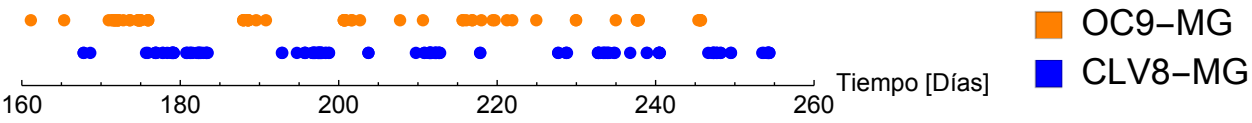


Figura **3-5**: Cadena de eventos de los estado de viento CLV8-MG y OC9-MG donde se observa un desfase en el coportamiento.

3.5. Red de estados de viento

A partir de la correlación de los estados de viento de las diferentes estaciones anemométricas se forman relaciones de los estados de viento que pueden corresponder a un mismo estado general o estados locales. Es posible formar una red que represente el sistema de los estados

de viento en las cuatro estaciones anemométricas considerando estas relaciones.

En la formación de esta red se hacen varias suposiciones *a priori* y se determina el parámetro que afectan la formación de esta red. Ambas suposiciones hacen que la red formada sea una red simple.

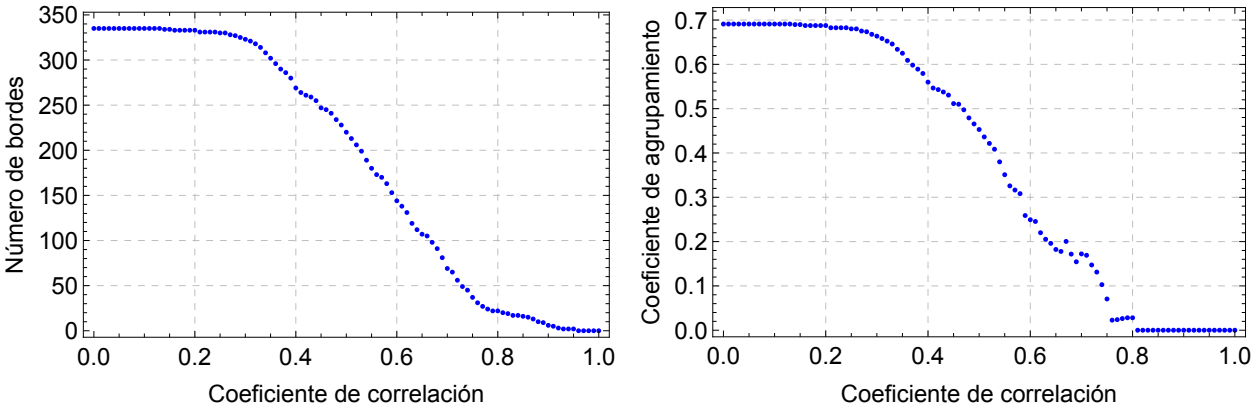
1. Los bordes no son direccionados: En esta suposición los estados de viento no tienen algún tipo de relación diferente entre ellos como sucede en otras redes de tipo social, donde el tipo de vínculo entre agentes es esencial. Esta suposición conduce entonces a que la red no tiene ningún tipo de orden jerárquico establecido, aunque se puede realizar un dendograma para encontrar las comunidades. Es posible que pueda determinarse una dirección en los bordes si se conoce la dirección del viento de una estación a otra, teniendo mayor jerarquía la que se presenta primero, sin embargo para este estudio el viento puede tomar diferentes direcciones por lo que sería necesario rastrear la dirección del viento en cada momento y cambiar la jerarquía de forma dinámica, aunque en este caso no se podrían considerar fenómenos locales que no presenten la dirección preferencial de los vientos generales.

2. Los nodos y los bordes no son ponderados: Para la formación de la red de estados de viento se considera que cada estado de viento es un nodo y que no tiene ponderación alguna. Es posible determinar una ponderación a partir del coeficiente de correlación, pero para simplificar el análisis en este trabajo no se hace tal consideración.

3.5.1. Determinación del coeficiente de correlación mínimo

La determinación del rango del coeficiente de correlación para la red se convierte en un parámetro clave ya que se busca que la red tenga gran riqueza estructural y un nivel de significancia estadística considerable. El coeficiente de correlación debe ser tomado de forma que balancee ambos aspectos y por lo tanto no es trivial.

En la figura **3-6**a y **3-6**b se muestran el número de bordes y el coeficiente de agrupamiento respectivamente en función del coeficiente de correlación. Se observa que para ambos casos la tendencia es que el valor de estos parámetros disminuya conforme crece el coeficiente de correlación lo que indica una pérdida de riqueza en la red al mismo tiempo que disminuye la significancia estadística.



(a) Número de bordes y correlación. (b) Coeficiente de agrupamiento y de correlación.

Figura **3-6**: Variación de la riqueza de estructura de la red de estados de viento en función del coeficiente de correlación.

Steinhaeuser determina un coeficiente de correlación cruzada, hace una comparación a diferentes tiempos de desfase y determina que el 1 % de las relaciones que obtiene son suficientes para representar su red de clima [31]. Por otro lado, Tsonis utiliza el coeficiente de correlación de Pearson y encuentra que para un $r > 0.5$ se tiene un nivel de significancia del 99 % usando una prueba t-student [34]. La determinación del coeficiente de correlación más adecuado y el nivel de significancia mínimo permitido sigue siendo un tema no del todo resuelto para el armado de redes de clima.

Donges hace un análisis de diferentes parámetros para definir la correlación lineal o no lineal para el armado de una red de clima [42]. De forma general determina que la densidad de bordes ρ es un parámetro útil para la determinación del factor limitante para el armado de la red. Menciona que para redes de clima al tener $\rho \geq 0.1$ se espera que las información contenida no sea muy significativa. Menciona que en los trabajos de Tsonis la densidad de bordes es $\rho \approx 0.01$. Para la red de estados de viento se buscará entonces que la densidad de bordes corresponda aproximadamente a lo reportado por Tsonis.

Capítulo 4

Resultados y discusión

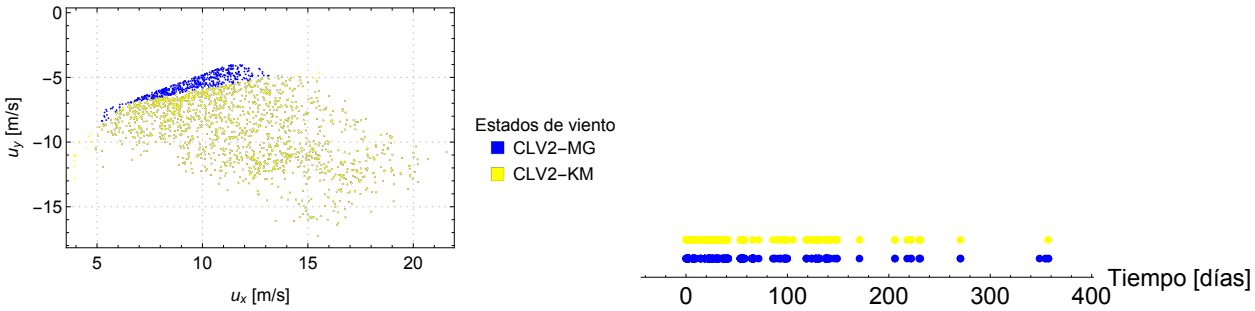
4.1. Comparación de algoritmos de agrupamiento

En las figuras **4-1** y **4-2** se muestran en el plano de velocidades los estados de viento obtenidos usando los algoritmos de agrupamiento. Al realizar la comparación de las cadenas de eventos se obtienen cuatro matrices cuadradas, una por cada estación, donde se representa el coeficiente de correlación de los estados usando el algoritmo k-Means y Mezcla Gaussiana. En las matrices de correlación mostradas en el anexo A.1 se observa el muestreo completo en la búsqueda de estados de viento equivalentes entre ambos algoritmos en base a la sincronización de eventos con $\tau = 0$ (cadenas de evento sin desfase). Observando estas matrices y comparando los estados de viento en el plano de velocidades así como sus cadenas de tiempo se pudieron dentificar dos comportamientos:

1. Relación directa entre un estado de viento y otro: Existen relaciones directas entre un estado de viento de un algoritmo y otro. Este comportamiento se puede identificar cuando los coeficientes de correlación tienen un número relativamente mayor al resto, al comparar estos estados de viento entre sí se observa que prácticamente se trata del mismo estado de viento.

2. Composición de estados de viento: Existen otras relaciones donde los coeficientes de correlación entre varios estados son muy similares entre ellos. Estas relaciones muestran una especie de composición ya que un estado de viento comparte características de otros estados de viento del algoritmo contrario.

En la figura **4-3** se muestra un ejemplo de una relación directa entre los estados de viento de la estación Cerro La Virgen. Se puede observar como tanto en el plano de velocidades **4-3**a como en la cadena de eventos **4-3**b estos estados tienen un comportamiento muy similar. Esto quiere decir que los algoritmos de agrupamiento definen prácticamente el mismo estado de viento.

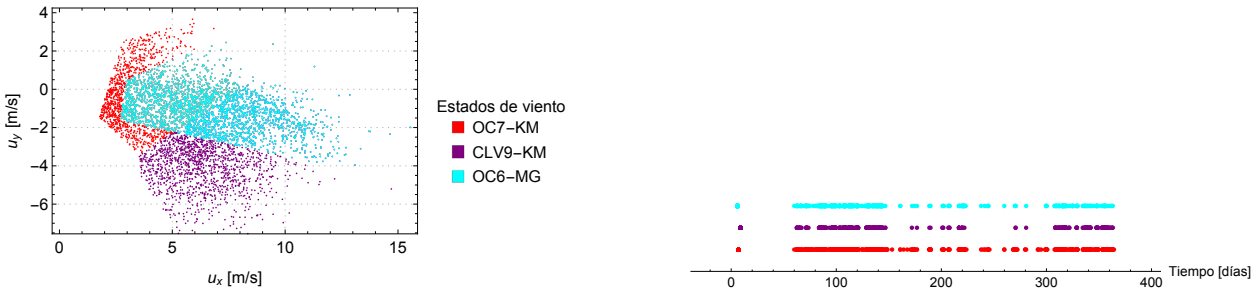


(a) Relación directa entre los estados de viento en la estación Cerro La Virgen.

(b) Cadena de eventos de los estados comparados.

Figura **4-3**: Ejemplo de una comparación directa entre estados de viento de los algoritmos k-Means (KM) y Mezcla Gaussiana (MG)

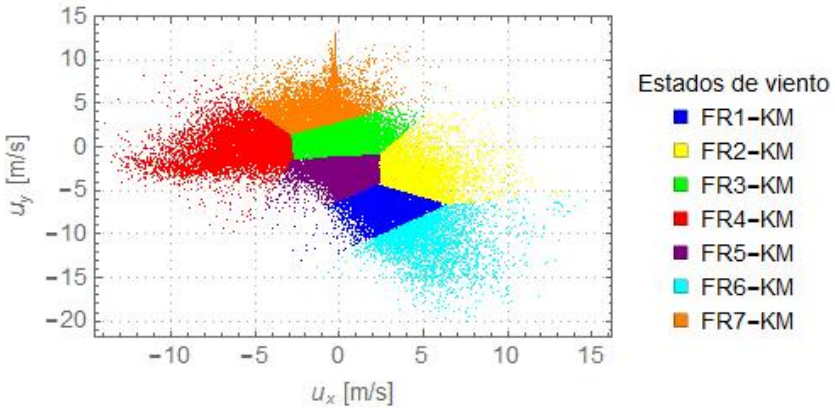
Por otro lado, en la figura **4-4** se muestra una composición de estados de viento de la estación Ojo Caliente. En el plano de velocidades **4-4**a se puede observar que el estado obtenido por la Mezcla Gaussiana (OC6-MG) comparte regiones con otros estados de viento obtenidos por k-Means (OC7-KM y OC9-KM), en cuanto a las cadenas de eventos **4-4**b se observa que tienen un comportamiento similar entre ellas. Esto indica que en estos casos los algoritmos de agrupamiento obtienen los estados de viento de una forma distinta pero que pueden formar conjuntos donde su comportamiento es estadísticamente equivalente.



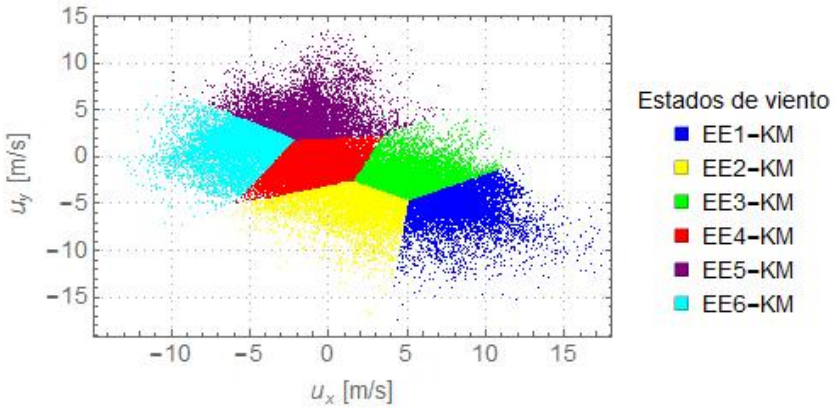
(a) Composición de estados de viento en la estación Ojo Caliente.

(b) Cadena de eventos de los estados compuestos.

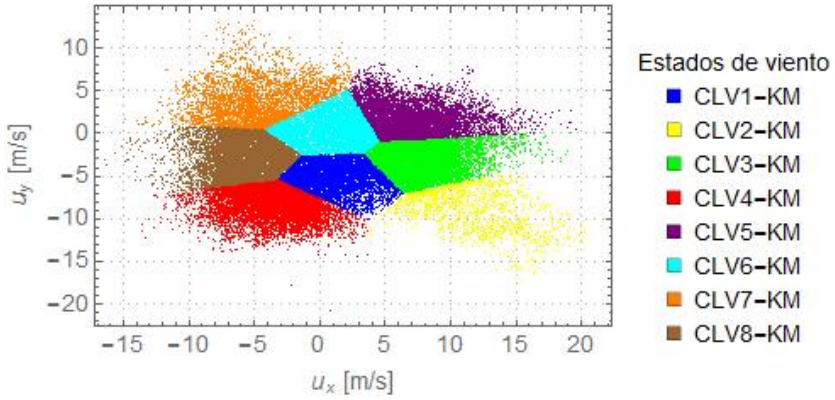
Figura **4-4**: Ejemplo de una comparación por composición entre los estados de viento de los algoritmos k-Means y Mezcla Gaussiana.



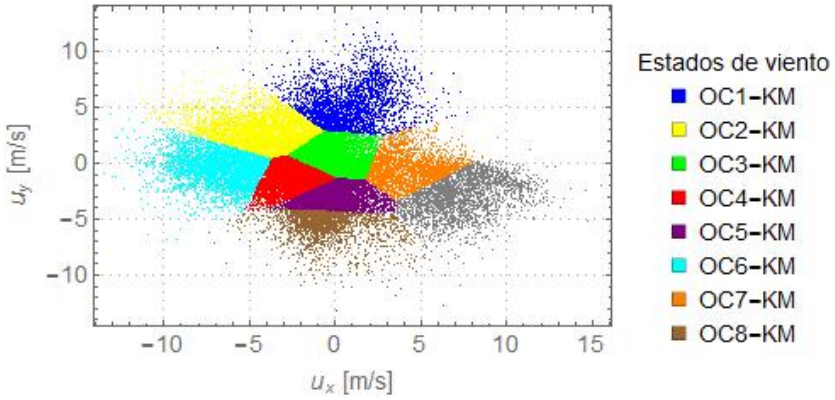
(a) Estados de viento de la estación FR



(b) Estados de viento de la estación EE

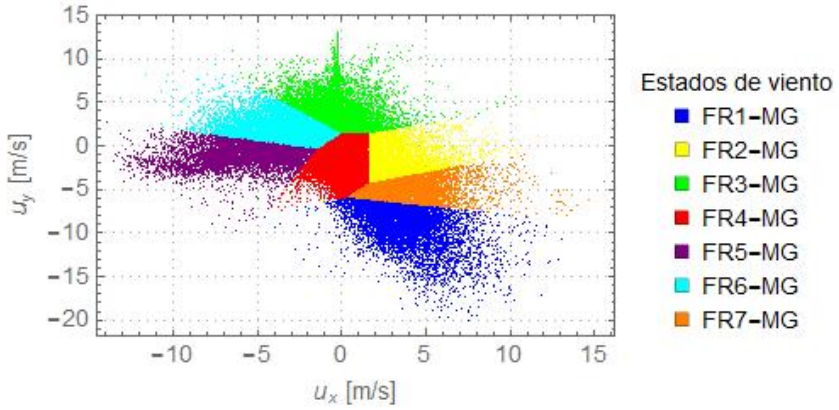


(c) Estados de viento de la estación CLV

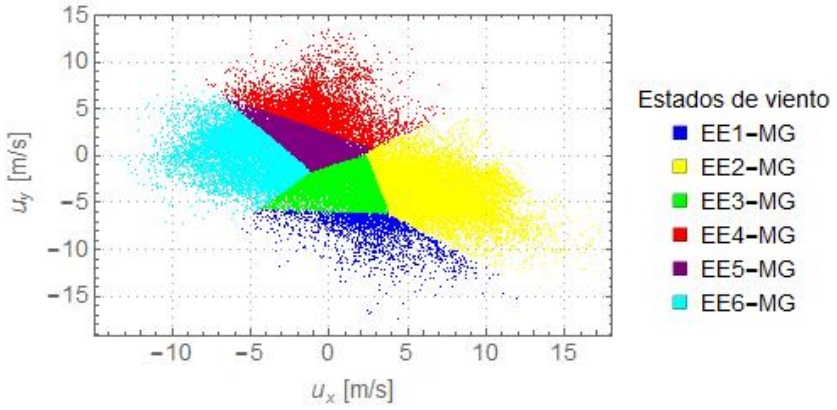


(d) Estados de viento de la estación OC

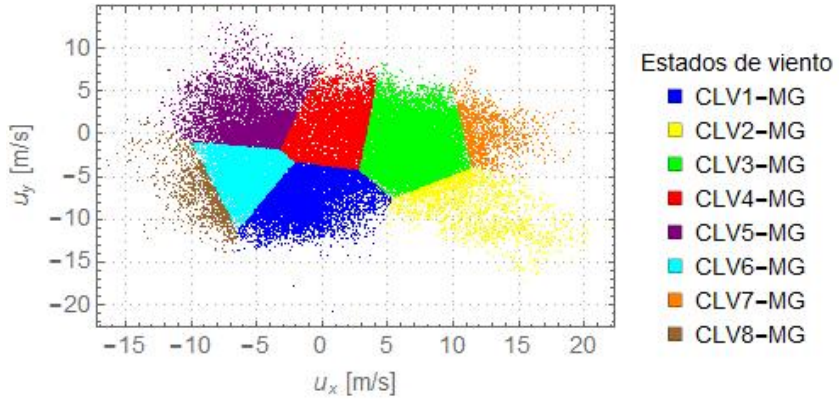
Figura **4-1**: Estados de viento de las cuatro estaciones usando el algoritmo k-Means



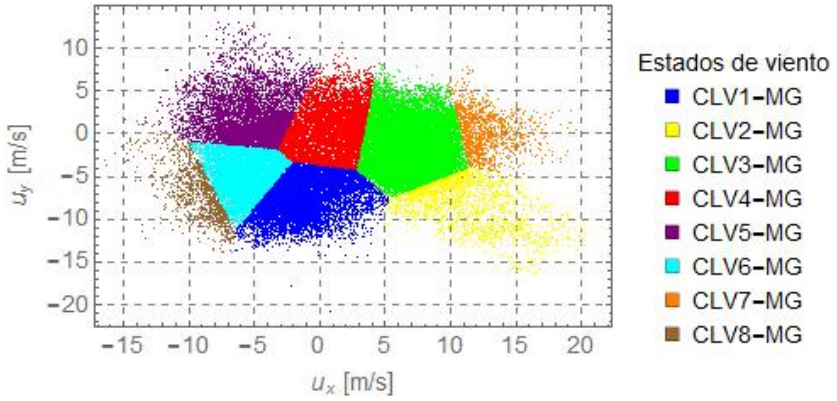
(a) Estados de viento de la estación FR



(b) Estados de viento de la estación EE



(c) Estados de viento de la estación CLV



(d) Estados de viento de la estación OC

Figura **4-2**: Estados de viento de las cuatro estaciones usando el algoritmo Mezcla Gaussiana

La existencia de estados de viento con relación directa entre los algoritmos de agrupamiento permite ver que hasta un cierto nivel ambos algoritmos están realizando un agrupamiento estadísticamente equivalente. Es natural pensar que no todos los estados de viento tendrán este tipo de relaciones directas porque sería pensar que no hay diferencia en lo que realizan los algoritmos de agrupamiento. De acuerdo con Bishop [5], haciendo unas suposiciones (como una matriz de covarianza uno y definiendo como parámetro de similitud una distancia Euclidiana) los algoritmos de Mezcla Gaussiana Multivariada y K-Means son equivalentes. A pesar de que en este trabajo se utiliza el algoritmo de la Inferencia Variacional de Mezclas Gaussianas Multivariadas las suposiciones aún pueden definir grupos similares.

Se considera entonces que tanto la información *a priori* como la búsqueda de distribuciones normales variadas son el criterio para la selección del mejor algoritmo de agrupamiento de estados de viento. La principal desventaja que se tiene al usar el método k-Means es un conocimiento *a priori* de los estados de viento que se desean, que si se busca sistematizar el análisis de estados de viento, requiere de intervención humana en el proceso. Esto, además de imposibilitar la completa automatización, no permite ver estados de viento que puedan estar escondidos en el procedimiento descrito por Sanchéz-Perez *et. al.* [9] y que podrían corresponder a fenómenos locales como se argumentó en la metodología 3.2. La definición de los estados de viento por medio de una mezcla de distribuciones normales simplifica el proceso de detección y caracterización de éstas para tratamientos estadísticos posteriores, además de que al ser una distribución estadística también se cuentan con las incertidumbres asociadas.

4.2. Red de estados de viento

Después de comparar los estados de viento de cada estación anemométrica se obtienen seis matrices de comparación, en las que se indican el máximo coeficiente de correlación entre los estados y a que tiempo de desfase corresponden. En el anexo A.2 se pueden encontrar las matrices completas. El algoritmo seleccionado para el armado de esta red es Mezcla Gaussiana de acuerdo al cirterio discutido con anterioridad.

Para determinar el coeficiente de correlación mínimo aceptado para la formación de la red se utiliza la densidad de bordes como criterio de selección. Se busca que el coeficiente de correlación mínimo tenga una densidad de bordes $\rho \approx 0.01$. Se define este número de acuerdo a lo encontrado en la literatura como un indicador de redes climáticas reportadas. La densidad de bordes se refiere a la cantidad de bordes que salen de un nodo con respecto a todas las combinaciones posibles de pares en la red, de esta forma se expresa como la probabilidad mínima que le estás permitiendo a la red hacer conexiones entre los nodos. Sin embargo, sigue existiendo una colección de coeficientes posibles y que entran en el criterio. En la figura **4-5** se observa la densidad de bordes en función del coeficiente de correlación.

La línea amarilla representa una densidad de bordes $\rho = 0.01$. Se puede observar que existen aún varias posibilidades de definición del coeficiente de correlación mínimo para la red.

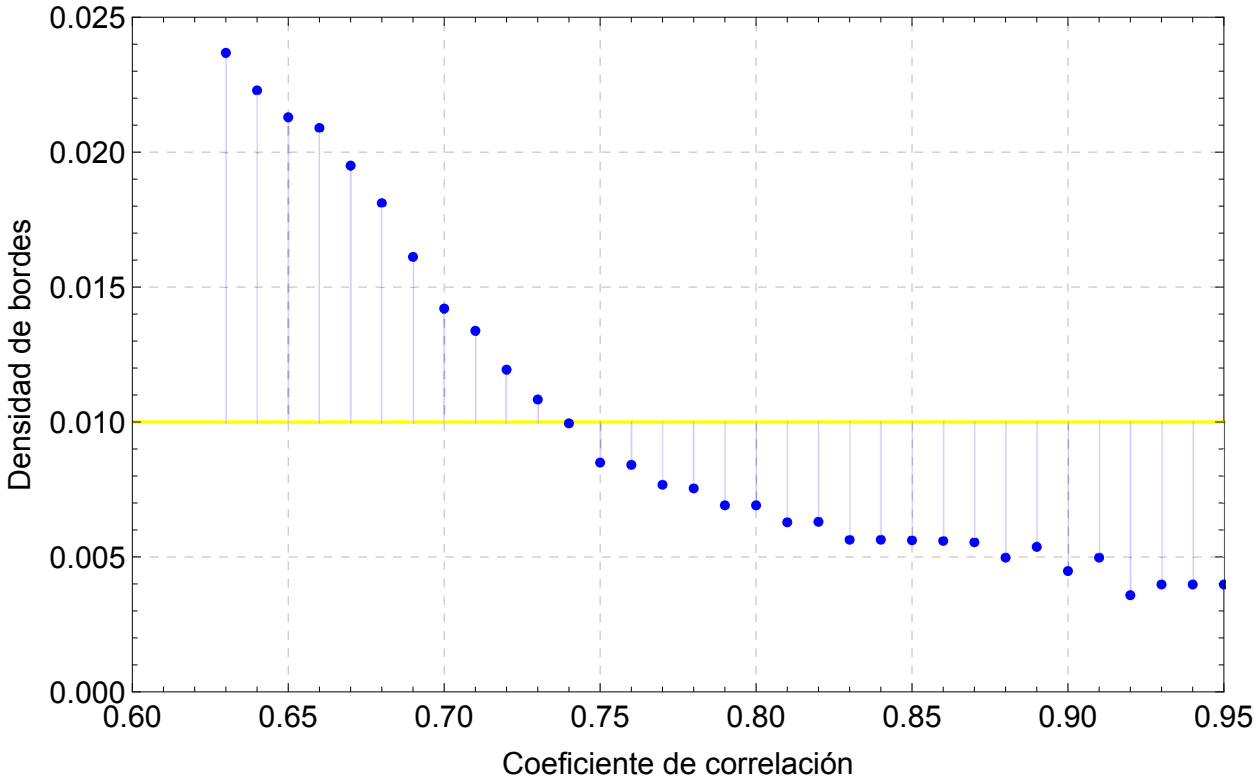


Figura **4-5**: Densidad de bordes en función del coeficiente de correlación.

Para definir el mejor coeficiente de correlación se toman los dos más cercanos a la línea de $\rho = 0.01$, que son $C = 0.73$ con $\rho = 0.0108$ y $C = 0.74$ con $\rho = 0.0099$. Para la determinación entre ambos se utilizó un criterio empírico pero que tiene sentido para la estructura de la red de estados de viento. Si se toma $C = 0.74$ como el coeficiente de correlación mínimo se pierde la conexión directa entre los estados de dos estaciones anemométricas contiguas (FR y EE) mientras que para $C = 0.73$ aún existe una correlación entre estas estaciones. Se considera que perder esta conexión directa entre ambas estaciones evita la formación de estados de viento regionales que pueden ser significativos. Aceptar este valor puede significar tomar un conexión que no sea estadísticamente significativa, sin embargo la diferencia de la densidad de bordes entre ambos coeficientes de correlación no es muy grande y el cambio en la estructura de la red sí lo es (el algoritmo para la detección de comunidades es muy sensible a los cambios de elementos), por lo que se le da prioridad a encontrar fenómenos regionales usando el coeficiente de correlación mínimo $C = 0.73$. La red obtenida se muestra en la figura **4-6**.

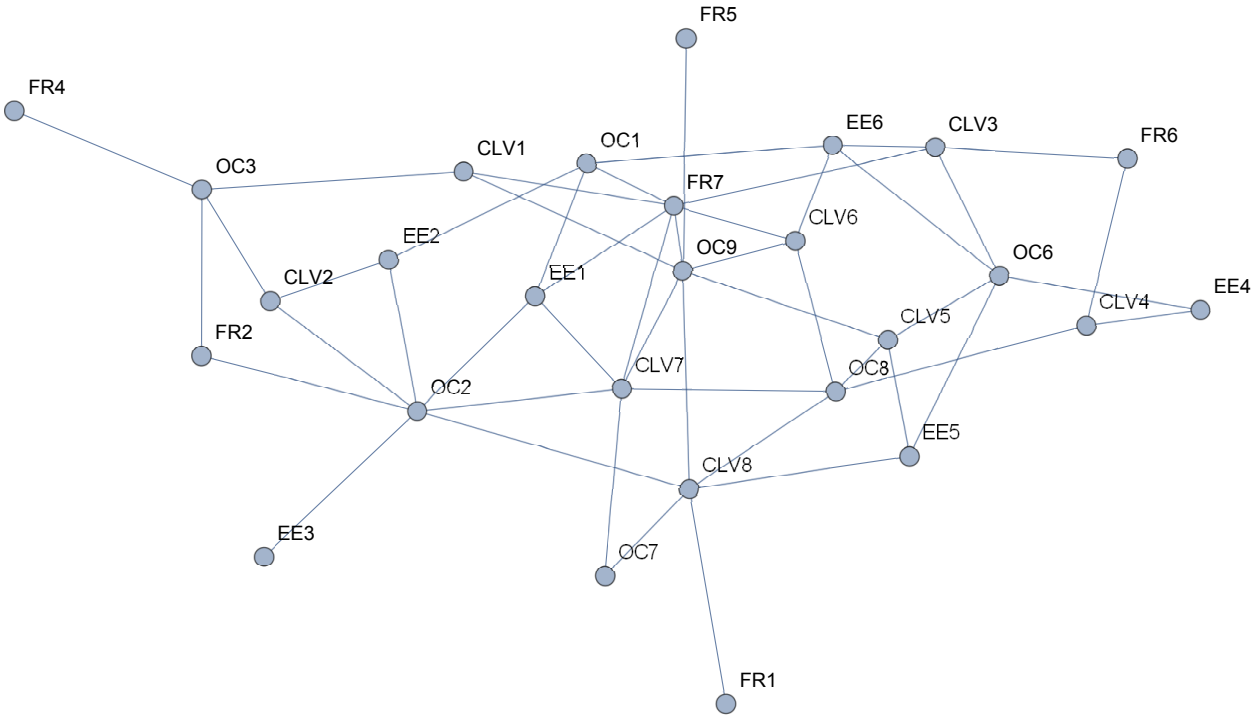


Figura **4-6**: Red de estados de viento.

Esta red es un modelo del sistema de estados de viento, donde se muestra el tipo de interacción que tienen entre estados de acuerdo a la sincronización de eventos. Las interacciones de este modelo describe la forma en la que los estados de viento, con regiones delimitadas de alcance de magnitud y dirección de viento, interactúan en las cuatro estaciones de la región. La red consta de 27 estados de viento con 49 bordes entre ellos, tiene un coeficiente de agrupamiento de 0.13, la asortatividad de la red es $r = -0.152$ que es una red ligeramente disasortativa indicando que los nodos de alto grado tienden a unirse con nodos de bajo grado pero con un comportamiento no muy marcado.

4.3. Comunidades de red de estados de viento

Se observa que existen regiones donde las conexiones entre estados son más densas y otros estados que son más alejados. Este comportamiento puede representar que en el sistema existen comunidades que pueden describir comportamientos regionales o locales en los vientos de la región debido a que la interacción entre ellos es mayor. En la figura **4-7** se muestran las comunidades para la red **4-6** utilizando el algoritmo de Girvan-Newman y optimizando la modularidad.

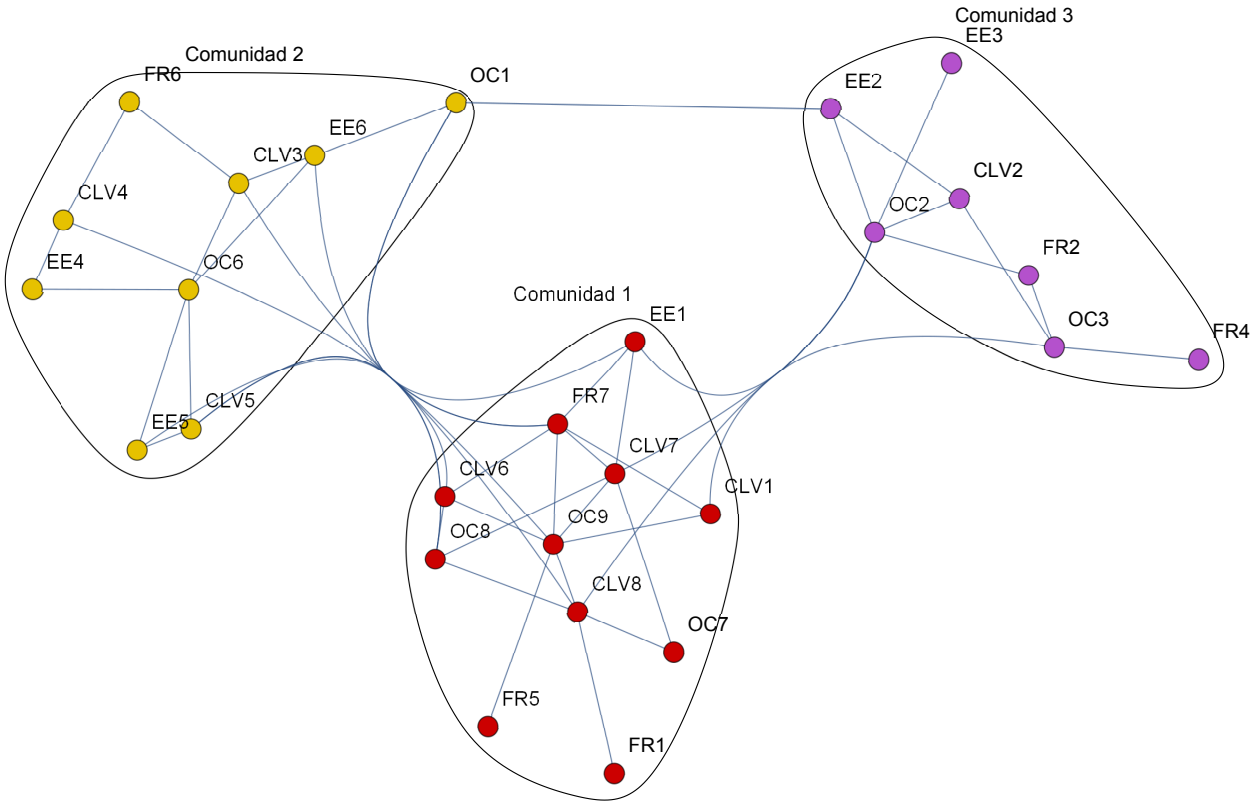


Figura **4-7**: Comunidades de la red de estados de viento.

Se encontraron tres comunidades en la red, lo que podría representar tres estados de viento regionales o locales, dependiendo del alcance geográfico de cada una y de la forma en la que se correlacionan los estados. Existe interacción entre las tres comunidades, es decir, cada una de ellas contiene conexiones que permiten la interacción con las otras comunidades, siendo la comunicación entre la comunidad 2 y la comunidad 3 la menor ya que solamente están relacionadas por una conexión. De acuerdo a esta estructura puede ser que la comunidad 1 actúe como una comunidad de transición de información entre las otras dos. Esta red también tiene la característica de que puede ser recursiva.

La red de estados de viento describe la interacción permitida entre los diferentes estados de viento de acuerdo al criterio de correlación seleccionado y esto permite conocer la dinámica del viento a lo largo de las estaciones y de los estados de viento. Sin embargo, se espera que las comunidades no muestren solamente el comportamiento permitido de los estados de viento sino que también representen como conjunto en general el comportamiento del viento a lo largo de las estaciones que considera y de los principales tiempos de desfase.

Se analizarán cada una de las comunidades y se buscará hacer una interpretación física al considerar las conexiones encerradas por la comunidad, sus tiempos de desfase y el desplazamiento geográfico en la región. Para la representación de las comunidades se obtiene el

promedio de cada uno de los puntos que componen al estado de viento y se le representa en un mapa cartográfico el vector resultante de este promedio. El vector obtenido tiene la característica de representar el comportamiento típico del estado de viento y al representar todos los estados de viento de la comunidad se puede buscar una interpretación física de los fenómenos que captura.

4.3.1. Comunidad 1

En la figura **4-8** se muestra el mapa cartográfico de la comunidad 1 con los promedios de los estados de viento. Esta comunidad es candidato a representar un estado de viento regional debido a que cuenta con un alcance de las cuatro estaciones anemométricas, aunque para la estación EE solamente se cuenta con un estado de viento que lo representa. De forma general se puede observar que esta comunidad tiene un comportamiento de una corriente Sureste que parece comenzar en la estación FR y pasa por todas las estaciones. Al llegar a las estaciones CLV y OC la corriente parece ampliar su dirección al tener disponibles estados de viento con mayor apertura y al llegar a la estación OC muestra una mayor dirección al Este.

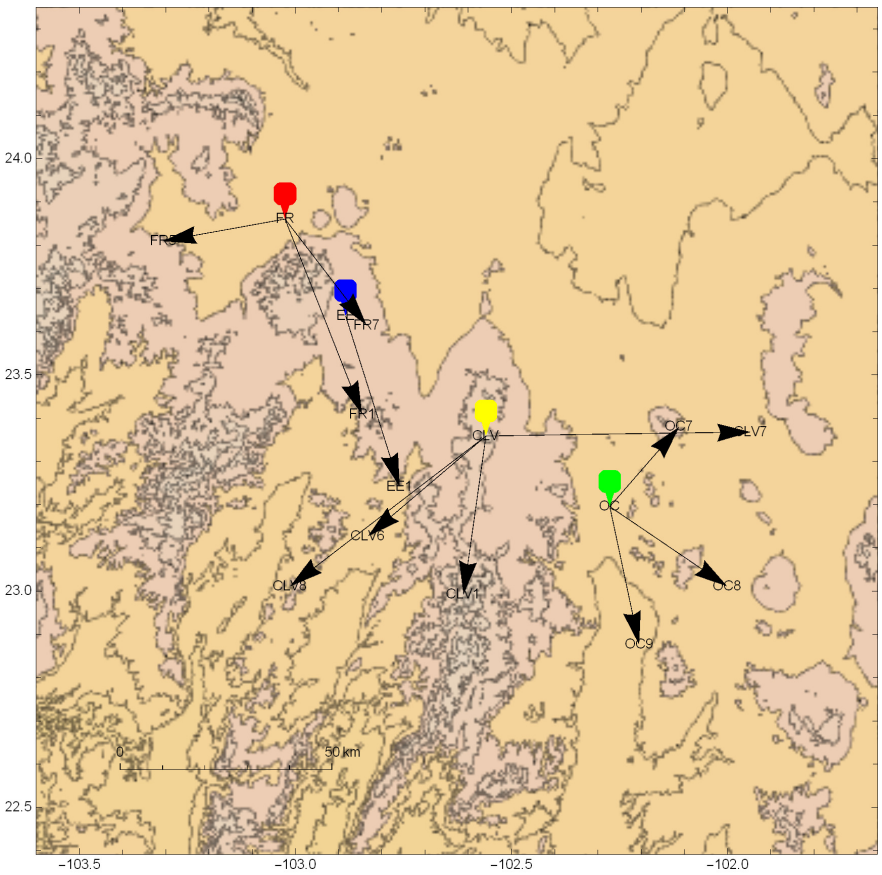


Figura **4-8**: Mapa cartográfico de la comunidad 1.

En la tabla **4-1** se tienen los tiempos de desfase en minutos y el coeficiente de correlación de la interacción de estados de viento dentro de la comunidad 1. Esta tabla nos da información acerca de la dinámica dentro de la comunidad, donde se puede observar que no todos los movimientos entre los estados de viento de la comunidad están permitidos y que, de hecho, de toda la posibilidad de movimientos la comunidad solamente se permite una fracción.

Se observa que el estado FR7 interactúa fuertemente con los estados de viento de la estación CLV y que los estados de las estaciones CLV y OC también tienen una interacción importante. Esto parece indicar que la estación CLV tiene una importante comunicación dentro de la comunidad y que sirve como puente de interacción entre las estaciones más alejadas, efecto que también se observa de acuerdo a la ubicación geográfica. En promedio el tiempo de desfase de la comunidad 1 es de 2.75 horas, lo que quiere decir esa es aproximadamente su resolución temporal.

Algo interesante de observar es que el estado de viento EE1, único de la estación EE dentro de la comunidad 1, no tiene una conexión con ningún estado de la estación FR y la única relación presente (con una correlación alta) es con el estado CLV7. Esto quiere decir que no necesariamente tiene que existir un movimiento contiguo de estaciones de acuerdo a las interacciones permitidas en la red y parece indicar que en algunos casos existe un fenómeno de teleconexión que siguen un mismo patrón a nivel comunidad. La presencia de teleconexiones en redes de clima (altas correlaciones con una considerable distancia geográfica) son las responsables de características no triviales e interesantes. La inclusión de este tipo de teleconexiones resultan necesarias si lo que se busca es encontrar patrones ocultos en el comportamiento del clima [42].

Cuadro **4-1**: Tiempos de desfase y coeficiente de correlación entre los estados de viento de la comunidad 1.

(τ, C_{ij}^{τ})	CLV1	CLV6	CLV7	CLV8	FR1	FR5	FR7	EE1	OC7	OC8	OC9
CLV1	-	-	-	-	-	-	(690, 0.73)	-	-	-	(980, 0.76)
CLV6	-	-	-	-	-	-	(90, 0.76)	-	-	(0, 0.91)	(40, 0.91)
CLV7	-	-	-	-	-	-	(0, 0.73)	(0, 0.89)	(120, 0.85)	(20, 0.85)	-
CLV8	-	-	-	-	(40, 0.75)	-	-	-	(0, 0.78)	(250, 0.82)	(70, 0.92)
FR1	-	-	-	-	-	-	-	-	-	-	-
FR5	-	-	-	-	-	-	-	-	-	-	(0. 0.74)
FR7	-	-	-	-	-	-	-	-	-	-	(180, 0.74)
EE1	-	-	-	-	-	-	-	-	-	-	-
OC7	-	-	-	-	-	-	-	-	-	-	-
OC8	-	-	-	-	-	-	-	-	-	-	-
OC9	-	-	-	-	-	-	-	-	-	-	-

Desde la observación de los estados de viento se podía notar que en todos existía la presencia de una corriente de viento en dirección Sureste. La determinación de las comunidades fue capaz de detectar parte de este comportamiento y tomar los estados de viento que interactúan en ella. Una interpretación física es justamente la presencia de una corriente de viento direc-

ción Sureste que se mueve rápidamente (caracterizada en promedio por el tiempo de desfase $\tau = 2.7$ horas) y que tiene presencia a lo largo del año.

4.3.2. Comunidad 2

La comunidad 2 es candidata a representar un estado de viento regional debido a que tiene una extensión geográfica que abarca todas las estaciones. En la figura **4-9** se puede observar que la comunidad 2 tiene características interesantes. De forma general muestra un comportamiento de cambio de dirección en donde la estación CLV actúa como intermediaria, hacia el Norte de la estación se observa una posible corriente Noroeste, la estación CLV tiene dos estados en dirección Este-Oeste y un estado central y en la estación OC al Sur de la estación CLV se presentan un estado dirección Este y otro norte. El movimiento general que se observa es de un cambio de dirección del viento con la estación CLV como pivote. El tiempo de desfase promedio de la comunidad 2 es de 7.1 horas, mayor al de la comunidad 1.

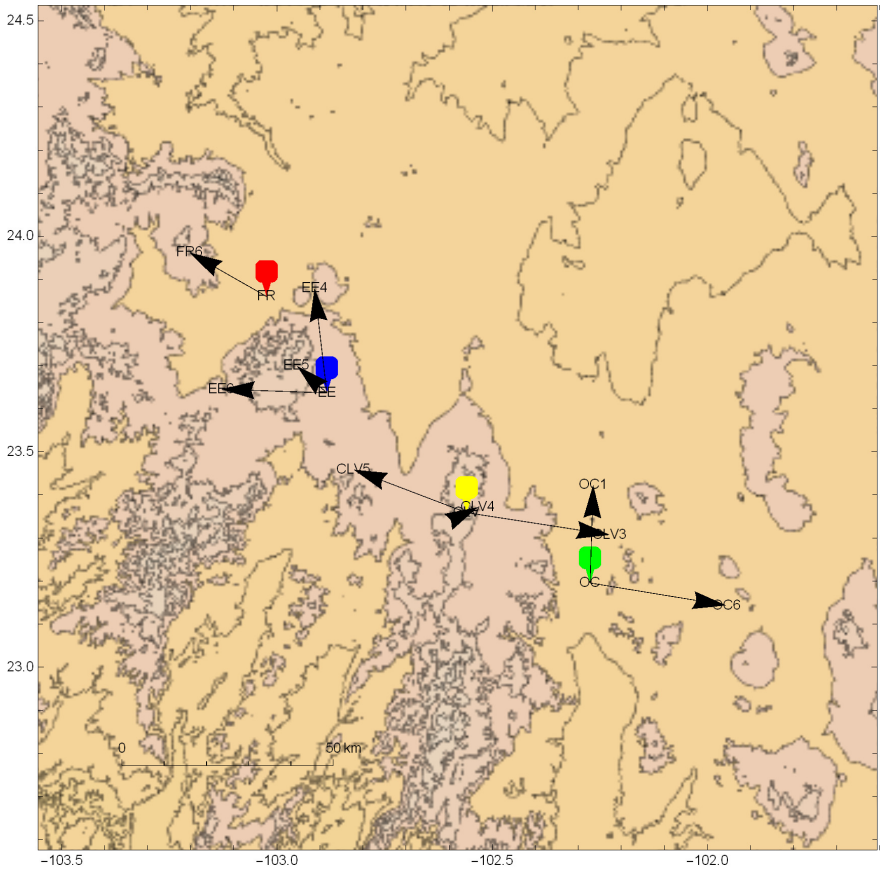


Figura **4-9**: Mapa cartográfico de la comunidad 2.

En la tabla **4-2** se muestran las conexiones permitidas dentro de la comunidad 2. Se observa que la estación CLV tiene una importante interacción con todas las estaciones y en forma de

teleconexión se presenta también una interacción considerable entre el estado OC6 con los estados de la estación EE en el que sucede un cambio de dirección en un promedio de 7.2 horas. Es decir, de la estación EE a el estado OC6 generalmente sucede en un promedio de 7.2 horas, tiempo que caracteriza este movimiento de cambio de dirección de la comunidad 2.

Cuadro **4-2**: Tiempos de desfase y coeficiente de correlación entre los estados de viento de la comunidad 2.

(τ, C_{ij}^{τ})	CLV3	CLV4	CLV5	FR6	EE4	EE5	EE6	OC1	OC6
CLV3	-	-	-	(1080, 0.75)	-	-	(250, 0.88)	-	(370, 0.75)
CLV4	-	-	-	(0, 0.74)	(0, 0.76)	-	-	-	
CLV5	-	-	-	-	-	(0, 0.89)	-	-	(1290, 0.75)
FR6	-	-	-	-	-	-	-	-	
EE4	-	-	-	-	-	-	-	-	(90, 0.77)
EE5	-	-	-	-	-	-	-	-	(1210, 0.74)
EE6	-	-	-	-	-	-	-	(390, 0.81)	(10, 0.77)
OC1	-	-	-	-	-	-	-	-	-
OC6	-	-	-	-	-	-	-	-	-

Una posible interpretación física de la comunidad 2 es que caracteriza cambios de dirección provocados por los efectos valle-montaña. La comunidad tiene una temporalidad de aproximadamente un cuarto del día, momentos en los que se hace presente cambios en la dirección de viento por el efecto valle-montaña y la orografía de la región está adecuada para que este fenómeno pueda llevarse a cabo.

4.3.3. Comunidad 3

La comunidad 3 también es candidato a ser un estado de viento regional ya que cuenta con estados de viento de todas las estaciones anemométricas. En la figura **4-10** se observa que de forma general la comunidad tiene una corriente Sureste y a diferencia de la comunidad 1 esta corriente no tiende a ampliar su dirección en las otras estaciones pero sí tiende a tener una orientación más Este. Sin embargo, al llegar a la estación OC la corriente de viento cambia completamente de dirección hacia una dirección Noroeste. También se observa que las estaciones FR y EE tienen estados de viento muy similares en dirección pero con una mayor magnitud para el momento en que se presentan en la estación EE y, al llegar a la estación CLV la magnitud es considerablemente mayor.

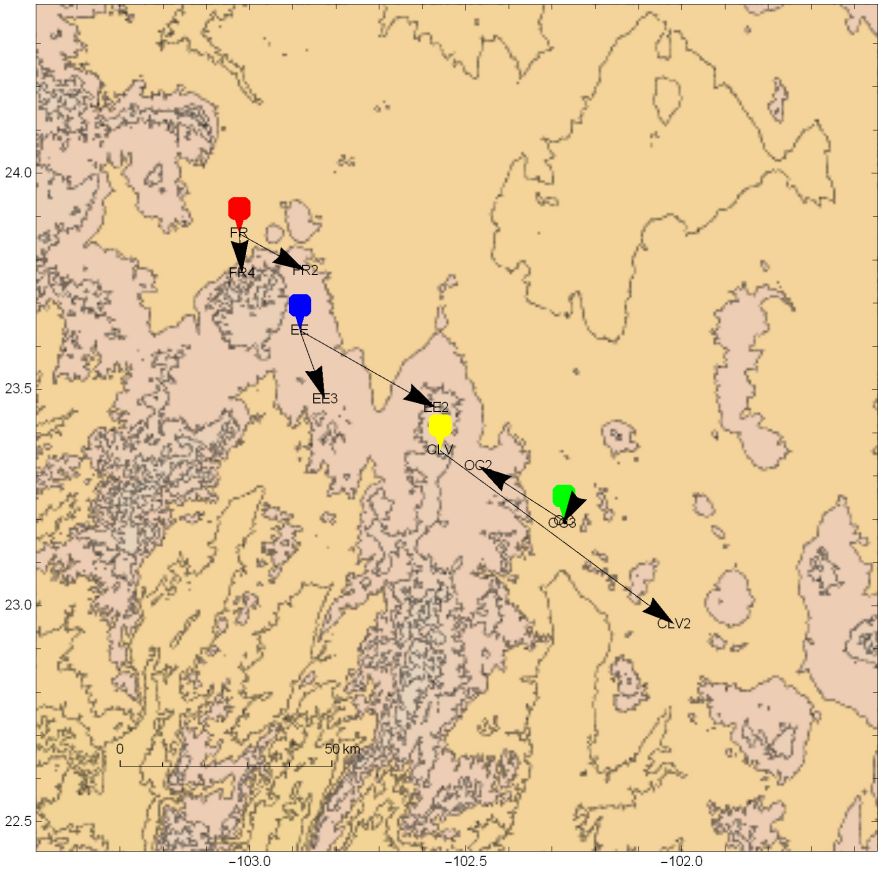


Figura **4-10**: Mapa cartográfico de la comunidad 3.

En la tabla **4-3** se observan que la cantidad de conexiones disponibles es menor considerando la de las otras comunidades y los estados de la estación OC muestran una amplia interacción con el resto de los estados. Nuevamente eso puede representar una teleconexión que es detectada. El promedio de tiempos de desfase de la comunidad 3 es 27.54 horas, siendo la mayor de todas las estaciones. Esta información sobre los tiempos de desfase da entonces una indicación de lo que realiza el método de comunidades de forma indirecta, logra agrupar elementos de la red que tienden a tener temporalidades mayores.

La comunidad 3 nuevamente muestra esta corriente Sureste que es perceptible en los estados de viento y el método de comunidades nuevamente logra captar los estados que tienen este comportamiento. La diferencia con la comunidad 1 es que los tiempos de desfase en este caso son mayores lo que representa movimientos en los estados de viento más lentos, con una resolución de poco más de un día.

Cuadro **4-3**: Tiempos de desfase y coeficiente de correlación entre los estados de viento de la comunidad 3.

(τ, C_{ij}^{τ})	CLV2	FR2	FR4	EE2	EE3	OC2	OC3
CLV2	-	-	-	(9360, 0.74)	-	(1110, 0.78)	(450, 0.87)
FR2	-	-	-	-	-	(10, 0.74)	(880, 0.75)
FR4	-	-	-	-	-	-	(740, 0.75)
EE2	-	-	-	-	-	(220, 0.86)	-
EE3	-	-	-	-	-	(450, 0.9)	-
OC2	-	-	-	-	-	-	-
OC3	-	-	-	-	-	-	-

Un caso interesante es lo que sucede con el estado de viento OC2 que tiene un cambio de dirección muy marcado. Como se mencionó antes este estado interactúa fuertemente con el resto de las los estados de la comunidad. Tiene una fuerte correlación con los estados EE2 y EE3, con tiempos de desfase de 3.6 y 7.5 horas respectivamente. Esta fuerte correlación muestra un patrón de movimiento del viento que no es muy clara en su interpretación física y no fácil de detectar. Los fenómenos físicos que originan este cambio de dirección no son claros pero la correlación por sincronización de eventos y la detección de comunidades fueron capaces de captar este movimiento.

Una interpretación física de la comunidad 3 en general es nuevamente la presencia de una corriente Sureste con una temporalidad mayor que la de la comunidad 1 y que tiende a tomar una dirección más al Este.

4.4. Posibles estados locales

La red de estados de viento integra a casi todos los elementos con lo que puede contar, pero no integra absolutamente a todos los estados. Los estados de viento FR3, OC4 y OC5 no aparecen en la red de estados de viento por lo que puede representar que no están conectados con el resto de los estados de viento circundantes, esto puede ser una característica de un estado de viento local, que tiene fenómenos que solamente representan fenómenos cercanos a la estación de donde se obtienen. En la figura **4-11** se muestra un mapa cartográfico de la región donde se muestra el vector promedio de velocidad de estos estados de viento. Aunque se representen en la misma figura cabe aclarar que en efecto lo que se está proponiendo es que estos estados no se pueden correlacionar con ningún otro porque corresponden a fenómenos aislados de viento en la localidad.

En el caso del estado FR3 se trata de un estado con dirección Norte, mientras que los estados OC4 y OC5 presentan direcciones Suroeste. Estos estados se presentan en las estaciones más externas de la región de estudio y puede que esto provoque que no se correlacionen lo suficiente con el resto de los estados, pero que si se pudiera considerar una región más amplia puedan ser correlacionados con otra extensión geográfica.

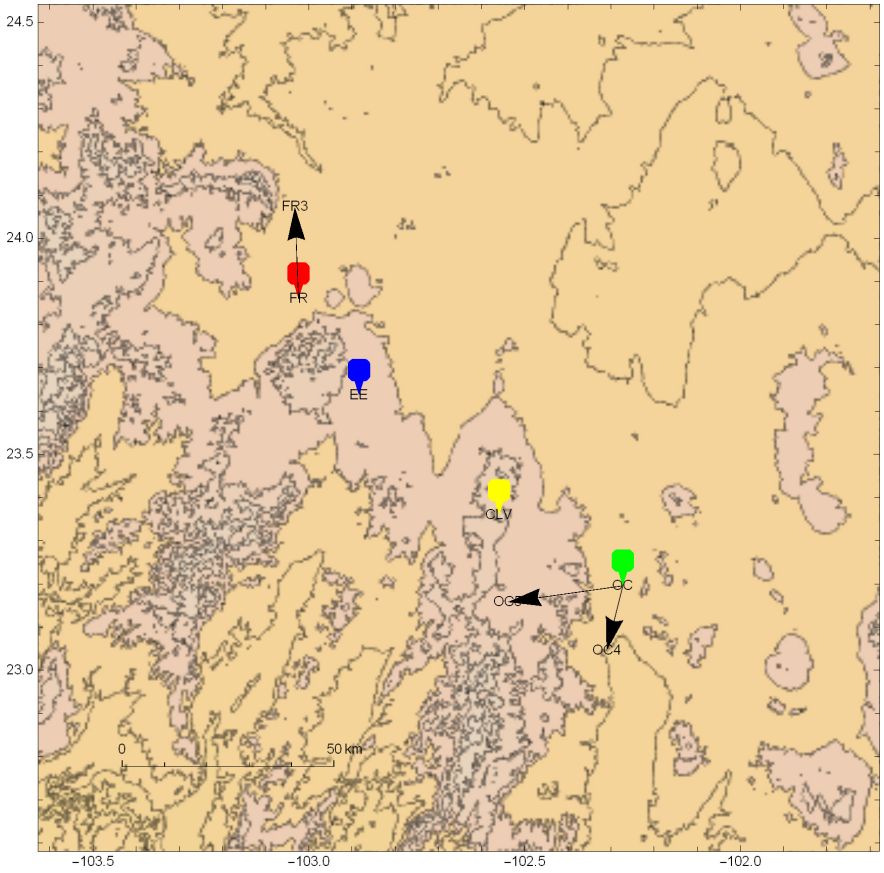


Figura **4-11**: Mapa cartográfico de los posibles estados locales de viento.

Es muy arriesgado determinar que estos estados de viento son debidos únicamente a fenómenos locales ya que su comportamiento no es extraordinario o único en cuanto a movimientos de viento. Para el caso del estado FR3 pareciera como si el estado pudiera pertenecer a un estado de la comunidad 2, donde se tiene un movimiento de cambio de dirección. Los estados OC4 y OC5 tienen claramente un comportamiento distinto a las tendencias del resto de las comunidades en esta estación. Sin embargo, en este análisis se consideran también las temporalidades en las que se presentan los estados de viento con respecto a los otros y, aunque se determinó el tiempo de desfase a la mayor determinación no tienen la suficiente correlación para estar incluidas en la red. De esta forma, aunque tengan un comportamiento similar a los estados de viento en otras comunidades su presencia no es igual de representativa que los otros.

Capítulo 5

Conclusiones y trabajos futuros

5.1. Conclusiones

Los algoritmos de agrupamiento K-Means y Mezcla Gaussiana producen algunos estados de viento con una equivalencia estadística comprobable. La presencia de estos estados equivalentes entre ambos algoritmos indica que ambos son útiles para la definición de estados de viento. En este trabajo se decide utilizar el algoritmo Mezcla Gaussiana ya que permite definir los estados de viento como una mezcla de distribuciones normales que son estadísticamente representativas y simples de operar. Este algoritmo también permite la automatización de la metodología propuesta debido a que no es necesario el conocimiento *a priori* de los grupos, describen las propiedades estadísticas de los estados de viento y reconocen mejor los posibles estados de viento locales.

La red de estados de viento es un buen mecanismo para modelar el sistema de correlaciones entre los estados de viento de las diferentes estaciones debido a que se hace natural la organización de esta forma por la compleja trama de relaciones presentes. Para la determinación de un criterio del coeficiente de correlación mínimo en la formación de la red de estados de viento se utilizó la densidad de bordes como el parámetro que permite definir esta limitante, reportando que para redes de clima $\rho = 0.01$ tiene un nivel de significancia aceptable.

Los métodos de detección de comunidades fueron capaces de identificar fenómenos con temporalidades promedio distintas y permite agrupar los estados de viento que participan en corrientes de viento presentes en la región. Este método tiene un buen potencial para la

detección de estados de viento regionales y es la primera vez que se realiza un trabajo de este tipo que realiza un análisis en conjunto de estados de viento, redes y comunidades de redes. Sin embargo, es necesaria una profundización mayor de su aplicación para este fin, con la finalidad de optimizar su funcionamiento y validar sus resultados.

5.2. Trabajos futuros

En el armado de la red de estados de viento se trató de simplificar en análisis como una primera aproximación del estudio, pero es posible formar redes más elaboradas que modelen mejor el sistema. Por ejemplo, es posible integrar bordes direccionados si se conoce la interacción que hay entre un estado de viento y otro. Además, con base al coeficiente de sincronización es posible determinar bordes ponderados.

Durante este trabajo se utilizó el algoritmo de Girvan y Newman para encontrar comunidades. Este algoritmo ha demostrado su capacidad para determinar correctamente comunidades en redes de diferentes campos científicos, sin embargo, no es el único algoritmo posible. Así como en el agrupamiento de datos, se propone realizar un análisis de las características deseadas de las comunidades para la detección de estados de viento regionales y en base a esto determinar el algoritmo que mejor funcione.

Profundizar en la detección de estados de viento locales permitirá tener un mejor panorama para la caracterización del viento en la región permitiendo realizar un modelo estocástico de la red de estados de viento considerando las comunidades como un nuevo nivel de interacción. De esta forma, se modela lo que sucede a nivel regional y a nivel local al considerar tanto la interacción entre comunidades, entre estados de una misma comunidad y en fenómenos locales. Un ejemplo de esta red puede ser una cadena de Markov anidada [43] o una cobija de Markov que contenga todas las variables que encierran cada nodo del resto de la red.

Finalmente, esta metodología solamente es una primera exploración de la capacidad de utilizar algoritmos de aprendizaje no supervisado para la detección de patrones de viento de datos reales y que no son fáciles de detectar de forma tradicional. Profundizar más en las diferentes suposiciones realizadas en este trabajo permitirá tener un algoritmo más fino. Contar con un área de estudio que ya esté caracterizada o utilizar herramientas de modelación a nivel de mesoescala como MERRA-2 permitiría además poner a prueba los algoritmos resultantes de este trabajo y otros con mayor certeza en los resultados obtenidos.

Apéndice A

Apéndice de resultados

A.1. Matrices de correlación

Se muestran las matrices de comparación entre estados de viento de los algoritmos k-Means y Mezcla Gaussiana por estación anemométrica.

Cuadro **A-1**: Matriz de correlación de la estación Fresnillo.

	FR1-MG	FR1-MG	FR3-MG	FR4-MG	FR5-MG	FR6-MG	FR7-MG
FR1-KM	0.756	0.515	0.466	0.238	0.494	0.464	0.729
FR2-KM	0.485	0.804	0.496	0.265	0.522	0.491	0.737
FR3-KM	0.354	0.477	0.515	0.498	0.424	0.486	0.389
FR4-KM	0.384	0.441	0.392	0.164	0.845	0.648	0.419
FR5-KM	0.341	0.505	0.348	0.745	0.395	0.346	0.393
FR6-KM	0.845	0.653	0.605	0.377	0.633	0.603	0.641
FR7-KM	0.460	0.517	0.846	0.240	0.496	0.619	0.496

Cuadro **A-2**: Matriz de correlación de la estación Enrique Estrada

	EE1-MG	EE2-MG	EE3-MG	EE4-MG	EE5-MG	EE6-MG
EE1-KM	0.655	0.628	0.374	0.507	0.335	0.325
EE2-KM	0.755	0.157	0.738	0.506	0.333	0.329
EE3-KM	0.559	0.619	0.404	0.444	0.266	0.256
EE4-KM	0.493	-0.065	0.439	0.387	0.650	0.380
EE5-KM	0.627	0.066	0.373	0.883	0.540	0.332
EE6-KM	0.599	0.038	0.345	0.478	0.392	0.855

Cuadro **A-3**: Matriz de correlación de la estación Cerro La Virgen.

	CLV1-MG	CLV2-MG	CLV3-MG	CLV4-MG	CLV5-MG	CLV6-MG	CLV7-MG	CLV8-MG
CLV1-KM	0.724	0.697	0.202	0.483	0.554	0.523	0.726	0.734
CLV2-KM	0.584	0.979	0.303	0.574	0.722	0.691	0.894	0.901
CLV3-KM	0.293	0.611	0.630	0.284	0.431	0.400	0.645	0.610
CLV4-KM	0.745	0.621	0.061	0.332	0.480	0.598	0.652	0.707
CLV5-KM	0.326	0.606	0.593	0.349	0.464	0.433	0.689	0.643
CLV6-KM	0.319	0.599	0.084	0.873	0.495	0.426	0.629	0.637
CLV7-KM	0.492	0.773	0.213	0.520	0.893	0.600	0.803	0.811
CLV8-KM	0.368	0.645	0.085	0.382	0.643	0.825	0.675	0.715

Cuadro **A-4**: Matriz de correlación de la estación Ojo Caliente.

	OC1-MG	OC2-MG	OC3-MG	OC4-MG	OC5-MG	OC6-MG	OC7-MG	OC8-MG	OC9-MG
OC1-KM	0.942	0.656	0.262	0.402	0.583	0.733	0.873	0.786	0.827
OC2-KM	0.673	0.964	0.190	0.330	0.520	0.661	0.752	0.713	0.755
OC3-KM	0.360	0.267	0.859	-0.012	0.170	0.320	0.420	0.372	0.414
OC4-KM	0.592	0.523	0.266	0.478	0.629	0.598	0.689	0.651	0.692
OC5-KM	0.562	0.487	0.120	0.791	0.418	0.568	0.659	0.637	0.667
OC6-KM	0.673	0.602	0.208	0.347	0.900	0.679	0.770	0.731	0.772
OC7-KM	0.746	0.670	0.317	0.421	0.602	0.898	0.869	0.826	0.846
OC8-KM	0.742	0.666	0.277	0.580	0.601	0.748	0.839	0.801	0.912
OC9-KM	0.738	0.662	0.273	0.413	0.594	0.864	0.835	0.919	0.841

A.2. Matrices de correlación entre estaciones

Se muestran las matrices de comparación de estados de viento entre estaciones anemométricas. Dentro de la matriz de muestra el tiempo de desfase τ a la cual ocurre la máxima sincronización.

Cuadro **A-5**: Matriz de correlación entre estaciones Cerro La Virgen y Fresnillo.

	FR1		FR2		FR3		FR4		FR5		FR6		FR7	
	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación
CLV1	95.67	0.42	22.50	0.72	1068.83	0.32	0.00	0.43	127.17	0.61	0.00	0.67	11.50	0.73
CLV2	53.33	0.50	36.50	0.67	699.00	0.20	132.17	0.49	0.67	0.63	80.17	0.55	46.83	0.69
CLV3	84.00	0.53	305.67	0.46	1270.00	0.30	77.33	0.48	67.17	0.59	0.00	0.74	12.33	0.74
CLV4	1.67	0.45	2.50	0.68	1260.17	0.26	83.67	0.50	1248.83	0.49	18.00	0.75	2.67	0.70
CLV5	0.00	0.60	83.00	0.66	0.00	0.48	214.83	0.55	75.83	0.57	24.00	0.69	82.83	0.49
CLV6	11.50	0.57	20.00	0.72	897.00	0.47	41.00	0.61	95.00	0.58	207.67	0.49	1.50	0.76
CLV7	49.17	0.47	0.17	0.44	0.00	0.54	0.00	0.70	179.83	0.41	57.33	0.72	0.00	0.73
CLV8	0.67	0.75	5.50	0.37	4.33	0.52	158.83	0.44	2.50	0.68	53.33	0.69	73.33	0.72

Cuadro **A-6**: Matriz de correlación entre estaciones Cerro La Virgen y Enrique Estrada.

	EE1		EE2		EE3		EE4		EE5		EE6	
	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación
CLV1	1.33	0.59	68.17	0.62	1345.17	0.20	103.67	0.72	132.67	0.51	36.17	0.59
CLV2	3522.67	0.34	156.00	0.74	1260.17	0.24	3557.33	0.33	98.17	0.62	49.00	0.58
CLV3	1.17	0.53	74.00	0.57	120.83	0.58	252.00	0.54	16.50	0.53	4.17	0.88
CLV4	86.17	0.54	79.17	0.55	3468.33	0.29	0.00	0.76	0.33	0.67	3389.00	0.38
CLV5	5.33	0.46	712.67	0.37	0.00	0.47	0.50	0.56	0.00	0.89	61.17	0.64
CLV6	14.33	0.59	0.00	0.63	5.00	0.56	145.33	0.53	2.83	0.40	0.00	0.77
CLV7	0.00	0.89	701.17	0.28	1.17	0.49	55.00	0.68	89.50	0.64	57.17	0.60
CLV8	4.83	0.39	572.33	0.31	124.50	0.42	3434.33	0.36	307.17	0.76	0.50	0.65

Cuadro **A-7**: Matriz de correlación entre estaciones Cerro La Virgen y Ojo Caliente.

	OC1		OC2		OC3		OC4		OC5		OC6		OC7		OC8		OC9	
	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación
CLV1	367.00	0.53	12.67	0.60	0.33	0.84	14.50	0.40	225.33	0.57	3.00	0.58	163.83	0.39	13.83	0.39	16.33	0.76
CLV2	14.67	0.51	18.50	0.78	0.75	0.87	13.67	0.38	196.67	0.62	3.17	0.68	1.83	0.54	13.33	0.54	83.50	0.39
CLV3	602.17	0.37	14.33	0.71	19.83	0.36	2.33	0.40	54.83	0.61	6.17	0.75	23.50	0.62	3.17	0.71	0.17	0.51
CLV4	334.83	0.32	87.00	0.36	11.67	0.34	76.67	0.57	200.50	0.63	0.67	0.72	0.33	0.69	0.33	0.86	23.50	0.70
CLV5	13.67	0.47	44.00	0.50	0.50	0.13	391.33	0.52	5.33	0.68	21.50	0.75	0.00	0.66	0.00	0.95	5.17	0.82
CLV6	18.33	0.54	5.00	0.64	4.50	0.27	163.00	0.33	2.50	0.62	214.00	0.63	0.00	0.69	0.00	0.91	0.67	0.91
CLV7	23.83	0.59	3.67	0.80	113.33	0.34	192.33	0.38	63.83	0.35	71.17	0.60	2.00	0.85	0.33	0.95	2.67	0.87
CLV8	16.17	0.58	0.00	0.87	0.00	0.38	226.50	0.49	58.33	0.43	37.17	0.38	0.00	0.78	4.17	0.82	1.17	0.92

Cuadro **A-8**: Matriz de correlación entre estaciones Fresnillo y Enrique Estrada.

	EE1		EE2		EE3		EE4		EE5		EE6	
	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación
FR1	0.00	0.71	0.00	0.35	17.83	0.50	15.33	0.44	9.00	0.51	2.67	0.62
FR2	3.00	0.55	0.00	0.58	0.83	0.69	0.67	0.41	0.00	0.68	0.00	0.71
FR3	159.00	0.51	16.00	0.63	0.00	0.52	204.00	0.36	36.00	0.66	0.00	0.47
FR4	277.00	0.58	48.17	0.48	72.50	0.50	70.17	0.69	3436.83	0.35	5.33	0.52
FR5	107.33	0.44	105.17	0.48	77.83	0.48	3406.00	0.35	86.50	0.50	23.83	0.61
FR6	1341.67	0.39	95.83	0.67	3453.50	0.26	54.17	0.53	0.00	0.62	79.67	0.46
FR7	4.17	0.73	655.33	0.28	0.50	0.47	88.00	0.61	17.17	0.53	86.17	0.44

Cuadro **A-9**: Matriz de correlación entre las estaciones Fresnillo y Ojo Caliente.

	OC1		OC2		OC3		OC4		OC5		OC6		OC7		OC8		OC9	
	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación
FR1	23.17	0.67	0.00	0.71	9.67	0.69	160.17	0.41	537.33	0.44	8.17	0.52	4.33	0.71	3.00	0.53	2.67	0.31
FR2	20.50	0.60	0.17	0.74	14.67	0.75	366.83	0.55	23.67	0.35	87.83	0.64	19.83	0.68	77.00	0.62	389.17	0.43
FR3	57.67	0.31	5.50	0.68	9.33	0.72	52.17	0.62	458.83	0.33	113.33	0.62	18.33	0.71	8.17	0.68	142.17	0.57
FR4	16.50	0.45	4.67	0.62	12.33	0.75	6.33	0.68	470.50	0.39	9.50	0.36	0.17	0.62	3.00	0.66	19.00	0.69
FR5	169.00	0.57	139.83	0.32	1.83	0.63	1.17	0.66	14.00	0.52	37.17	0.39	98.50	0.59	1.67	0.69	0.00	0.74
FR6	2.50	0.69	95.50	0.44	2.33	0.60	23.50	0.69	42.33	0.51	0.00	0.56	63.33	0.38	40.83	0.67	18.17	0.71
FR7	11.67	0.73	83.67	0.58	62.50	0.34	99.50	0.47	10.67	0.53	23.50	0.65	261.00	0.39	12.67	0.61	3.00	0.74

Cuadro **A-10**: Matriz de correlación entre las estaciones Enrique Estrada y Ojo Caliente.

	OC1		OC2		OC3		OC4		OC5		OC6		OC7		OC8		OC9	
	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación	τ [hr]	correlación
EE1	21.17	0.80	0.17	0.89	79.83	0.29	78.00	0.59	22.50	0.66	157.83	0.44	227.86	0.54	18.50	0.59	2697.00	0.30
EE2	12.17	0.74	3.67	0.86	139.33	0.34	58.83	0.55	8.83	0.64	4.67	0.58	106.50	0.49	297.83	0.58	24.00	0.46
EE3	102.67	0.37	7.50	0.90	0.00	0.39	22.83	0.32	12.50	0.66	45.67	0.70	93.00	0.33	7.00	0.59	2017.00	0.48
EE4	9.83	0.51	22.50	0.36	20.33	0.40	0.00	0.39	0.00	0.71	1.50	0.77	816.00	0.34	122.17	0.49	6.33	0.54
EE5	18.50	0.67	42.50	0.33	17.50	0.39	107.33	0.52	23.67	0.63	20.17	0.74	11.17	0.43	95.50	0.50	0.17	0.53
EE6	6.50	0.81	6.50	0.15	0.67	0.40	36.50	0.61	79.17	0.33	0.17	0.77	12.33	0.55	324.00	0.35	19.33	0.54

Bibliografía

[1] Sener. Balance nacional de energía 2017. page 184, 2017.

[2] M. E.J. Newman. Detecting community structure in networks. *European Physical Journal B*, 38(2):321–330, 2004.

[3] Santo Fortunato and Claudio Castellano. Community structure in graphs. *Computational Complexity: Theory, Techniques, and Applications*, 9781461418:490–512, 2012.

[4] Guojun Gan, Chaoqun Ma, and Jianhong Wu. *Data Clustering: Theory, Algorithms and Applications*, volume 106. American Stadistical Association and Society for Industrian and Applied Mathematics, Philadelphia, 1979.

[5] Bishop Christopher M. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.

[6] International Energy Agency. World Energy Balances: Overview. Technical report, International Energy Agency, Paris, 2018.

[7] Ibrahim Dincer. Renewable energy and sustainable development: a crucial review. *Renewable and Sustainable Energy Reviews*, 4(2):157–175, 2000.

[8] O. A. Jaramillo and M. A. Borja. Wind speed analysis in La Ventosa, Mexico: A bimodal probability distribution case. *Renewable Energy*, 29(10):1613–1630, 2004.

[9] P. A. Sánchez-Pérez, M. Robles, and O. A. Jaramillo. Real time Markov chains: Wind states in anemometric data. *Journal of Renewable and Sustainable Energy*, 8(2), 2016.

[10] Pedro Andrés Sánches Pérez. *Estados de viento para estimación del potencial eólico*. Temixco, 2016.

[11] Lee M Miller and David W Keith. Climatic Impacts of Wind Power Climatic Impacts of Wind Power. *Joule*, (2):1–15, 2018.

[12] Tony Burton, Sharpe David, Jenkins Nick, and Bossanyi Ervin. *Wind Energy Handbook*. John Wiley & Sons, West Sussex, 2001.

[13] J. F. Manwell; J. G. McGowan; A. L. Rogers. *Wind Energy Explained*. Jhon Wiley & Sons, 2nd edition, 2009.

[14] Brian S Everitt, Sabine Landau, Morven Leese, and Daniel Stahl. *Cluster Analysis*. Wiley, West Sussex, 5th edition, 2011.

[15] Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in data*. WILEY-INTERSCIENCE, Hoboken, NJ, 1990.

[16] Christian Hennig. What are the true clusters ? *Pattern Recognition Letters*, 64:53–62, 2015.

[17] Vladimir Estivill-Castro. Why so many clustering algorithms. *ACM SIGKDD Explorations Newsletter*, 4(1):65–75, 2002.

[18] Jon Kleinberg. An impossibility theorem for clustering. *Advances in NIP 15*, pages 446–453, 2017.

[19] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. PRENTICE-HALL, Englenwood Clifs, 1988.

[20] J Franzén. *Bayesian Cluster Analysis: Some Extensions to Non-standard Situations*. Phd thesis, Stockholm University, Stockholm, 2008.

[21] Cheng Guanrong, Wang Xiofan, and Li Xiang. *Fundamentals of complex networks: models, structures, and dynamics*. WILEY, Singapore, 1st edition, 2015.

[22] Ulrtk Brandes, Garry Robins, Ann Mrcr Anif, and Stanley Wasserman. What is network science? *Network Science*, 1(1):1–15, 2013.

[23] Warren Weaver. Science and complexity. *American Scientist*, 36(536):13, 1948.

[24] Stanley Milgram. The small-world problem. *Psychology today*, 1(1):61–67, 1967.

[25] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(4 June):440–442, 1998.

[26] Barabasi Albert-Laszlo and Albert Reka. Emergence of Scaling in Random Networks. *Science*, 286(October):509–512, 1999.

[27] M. E.J. Newman. Assortative Mixing in Networks. *Physical Review Letters*, 89(20):1–4, 2002.

[28] Aidong Zhang. *Protein interaction Networks: computational analysis*, volume 91. Cambridge University Press, New York, 1st edition, 2009.

[29] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann. The product space conditions the development of nations. *Science*, 317(July 2007):482–488, 2007.

[30] Anastasios A. Tsonis, Kyle L. Swanson, and Paul J. Roebber. What do networks have to do with climate? *Bulletin of the American Meteorological Society*, 87(5):585–595, 2006.

[31] Karsten Steinhaeuser, Nitesh V. Chawla, and Auroop R. Ganguly. An exploration of climate data using complex networks. *Proceedings of the Third International Workshop on Knowledge Discovery from Sensor Data - SensorKDD '09*, page 23, 2009.

[32] Mohamed Laib, Fabian Guignard, Mikhail Kanevski, and Luciano Telesca. Community detection analysis in wind speed-monitoring systems using mutual information-based complex network. *arXiv Data Analysis, Statistics and Probability*, pages 1–25, 2018.

[33] A. Agarwal, N. Marwan, R. Maheswaran, B. Merz, and J. Kurths. Quantifying the roles of single stations within homogeneous regions using complex network analysis. *Journal of Hydrology*, 563(May):802–810, 2018.

[34] A. A. Tsonis and P. J. Roebber. The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications*, 333(1-4):497–504, 2004.

[35] V. Stolbova, P. Martin, B. Bookhagen, N. Marwan, and J. Kurths. Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. *Nonlinear Processes in Geophysics*, 21(4):901–917, 2014.

[36] Mason A Porter, Jukka-Pekka Onnela, and Peter J Mucha. Communities in Networks. *Notices of the American Mathematical Society*, 56(9):19, 2009.

[37] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[38] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the National Acadey of Science of the United States of America*, 12(99):7821–7826, 2002.

[39] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[40] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review Letters*, 69(026113):15, 2004.

[41] R. Quian Quiroga, T. Kreuz, and P. Grassberger. Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 66(4):9, 2002.

[42] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods. *European Physical Journal: Special Topics*, 174(1):157–179, 2009.

[43] F. Tagliaferri, B. P. Hayes, I. M. Viola, and S. Z. Djokić. Wind modelling with nested Markov chains. *Journal of Wind Engineering and Industrial Aerodynamics*, 157:118–124, 2016.