



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**INSTITUTO DE ENERGÍAS RENOVABLES**

**INSTITUTO DE INGENIERÍA**

**ESCUELA NACIONAL DE ESTUDIOS SUPERIORES-JURIQUILLA**

**Metodología para evaluación de modelos meteorológicos para el aprovechamiento del recurso eólico mediante métodos de aprendizaje de máquinas**

**TESIS**

Que para obtener el título de  
**Ingeniera en Energías Renovables**

**P R E S E N T A**

Karla Gutiérrez Ramírez

**TUTOR DE TESIS**

Dr. Miguel Robles Pérez

Temixco, Mor., a 26 de septiembre de 2023





Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**OF/IER/LIER/096/2023**  
**ASUNTO:** Notificación de jurado y  
fecha para examen profesional.

**LIC. IVONNE RAMÍREZ WENCE**  
**DIRECTORA GENERAL DE ADMINISTRACIÓN ESCOLAR**  
**Presente**

Por medio de la presente le informo que el día martes 26 de septiembre de 2023, a las 17:30 h, en el Instituto de Energías Renovables, la **C. KARLA GUTIÉRREZ RAMÍREZ**, con número de cuenta 418068234 de la Licenciatura de Ingeniería en Energías Renovables, llevará a cabo la presentación de su trabajo de titulación y examen profesional titulado:

***“Metodología para evaluación de modelos meteorológicos para el aprovechamiento del recurso eólico mediante métodos de aprendizaje de máquinas”***

Debido a que la alumna 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.

<b>PRESIDENTE:</b>	<b>DR. EDUARDO RAMOS MORA</b>
<b>VOCAL:</b>	<b>DRA. NADIA MARÍA SALGADO HERRERA</b>
<b>SECRETARIO:</b>	<b>DR. MIGUEL ROBLES PÉREZ</b>
<b>SUPLENTE:</b>	<b>DRA. MÓNICA BORUNDA PACHECO</b>
<b>SUPLENTE:</b>	<b>DR. ALBERTO REYES BALLESTEROS</b>

Sin otro particular, le envío un cordial saludo.

Atentamente,  
**“Por mi raza hablará el espíritu”**  
Temixco, Mor., a 18 de septiembre de 2023



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

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](http://ier.unam.mx)

## Resumen

El objetivo de este trabajo fue evaluar la precisión del modelo ERA5 en 7 sitios en México con disponibilidad de datos de viento a través de cadenas de Markov y el método de *clustering*. Se analizaron las series de datos con coordenadas idénticas a través de racimos o *clusters* de viento, donde se definieron cúmulos de datos por densidad. Para esta división se utilizó el método de K-medias para determinar los centroides de mayor densidad. Se comparó uno a uno los sitios tomando en cuenta la distribución de viento y sus *clusters* correspondientes entre series de datos y métricas de: Tiempo de vida total, tiempo de vida medio, probabilidad condicional de coincidencia por paso de tiempo y probabilidad de no coincidencia. De esta forma se obtuvieron resultados sobre las métricas en las que sobresale ERA5. En general, resulta bastante certero para generar un panorama anual del sitio. Calcula de forma relativamente certera la permanencia en cada *cluster*, específicamente para las divisiones en 3 estados. Mientras que en cuanto a coincidencia por paso de tiempo se desempeña mejor en la división en 2 *clusters* con porcentajes de error anuales cercanos al 10%. Se identificó además los mejores sitios modelados por ERA5 y el peor para el análisis de los posibles factores a considerar a la hora de realizar un análisis de ERA5 y determinar si el modelado será más o menos confiable. En particular, CERTE, en + Oaxaca y San Fernando, en Tamaulipas presentaron los resultados más precisos, mientras que Tepexi fue el menos certero. Pareciera que el potencial eólico del sitio juega un factor importante en el modelado de ERA5, posiblemente asociado a la cantidad de ruido o aleatoriedad del sitio.

**Palabras clave:** *Cluster*, ERA5, Atlas Eólico Mexicano (AEM), México.

## Abstract

The objective of this project was to evaluate the precision of the ERA5 model in 7 sites in Mexico with availability of wind data through Markov chains and the method of clustering. Data series with identical coordinates were analyzed through wind clusters, where these were defined by density. In this partitioning, the K-means method was used to determine the centroids of greatest density. The sites were compared one by one taking into account the wind distribution and their corresponding clusters between data series as well as metrics of: Total lifetime, mean lifetime, conditional probability of matching per time step, and probability of no match. In this way, results were obtained on the metrics in which ERA5 stands out. In general, it is quite precise in generating an annual site overview . It calculates in a relatively accurate way the permanence in each cluster, specifically for the divisions in 3 wind states where it performs better. While in terms of matching by time step, it performs better in the division of 2 clusters with annual error rates close to 10 %. The best sites modeled by ERA5 and the worst were also identified and the analysis of the possible factors to consider when carrying out an analysis of ERA5 and determine if the modeling would be more or less reliable was also carried out. In particular, CERTE in Oaxaca and San Fernando in Tamaulipas presented the most precise results, while Tepexi was the least accurate. It seems that the wind potential of the site plays an important factor in the modeling of ERA5, possibly associated with the amount of noise or randomness of the site.

**Keywords:** Cluster, ERA5, Wind Atlas for Mexico (WAM), Mexico.

## **Dedicatoria**

Quisiera dedicar este trabajo a mis padres, que me han acompañado toda mi vida y siempre han confiado en mí.

A mi mamá, que toda la vida me dio la seguridad de mi capacidad y que fue amable, atenta y comprensiva. Ella no me presionó, ni exigió, pero a su manera logró que yo siempre quisiera hacer más, mejorar y superarme. Hiciste todo lo que yo necesitaba, gracias mamá.

A mi papá, quien nunca me limitó y confió en mí. Por darme la posibilidad de tener una excelente educación, dejarme avanzar y tomar mis propias decisiones. Gracias por inspirarme a ser valiente; intento serlo todos los días.

## **Reconocimientos**

- Al Instituto de Energías Renovables de la UNAM, por la excelente formación y apoyo brindado.
- Al Dr. Miguel Robles, por su tiempo, guía y consejos en la realización de este trabajo.

# Índice general

<b>Lista de Figuras</b>	VI
<b>Lista de Tablas</b>	XI
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos	1
1.1.1. Objetivo general	1
1.1.2. Objetivos específicos	1
1.2. Estado del arte	2
1.2.1. Investigaciones Nacionales	2
1.3. Investigaciones Internacionales	6
1.4. Justificación	10
1.5. Hipótesis	11
1.6. Modelos meteorológicos	11
1.6.1. ERA5 global reanalysis	11
<b>2. Metodología y tratamiento de datos</b>	<b>17</b>
2.1. Datos de Viento Disponibles en México	17
2.1.1. Estaciones Meteorológicas Automáticas (EMAS)	17
2.1.2. Energía Eólica en el INEEL	18
2.1.3. Atlas Eólico Mexicano	18
2.2. Metodología y tratamiento de datos	20
2.2.1. Datos de Viento	20
2.2.2. Cadenas de Markov y método de <i>clustering</i> para definición de es-	
tados de viento	21
2.2.3. Tratamiento de datos	24

<b>3. Casos de estudio de las áreas geográficas de acuerdo al Atlas Eólico Mexicano</b>	<b>26</b>
3.1. Evaluación de ERA5 en México . . . . .	26
3.1.1. La Rumorosa, Baja California . . . . .	26
3.1.2. Mérida, Yucatán . . . . .	33
3.1.3. Cd. Cuahutémoc, Chihuahua . . . . .	37
3.1.4. CERTE, Oaxaca . . . . .	40
3.1.5. Ojuelos, Jalisco . . . . .	48
3.1.6. San Fernando, Tamaulipas . . . . .	52
3.1.7. Tepexi, Puebla . . . . .	59
<b>4. Discusión y Conclusiones</b>	<b>64</b>
<b>5. Referencias</b>	<b>71</b>

# Índice de figuras

3.1. Distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5, en rojo y el Atlas Eólico Mexicano, en azul. Fuente: Elaboración propia. . . . .	27
3.2. Histograma de función de densidad de probabilidad en 3D de la distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5, en rojo y el Atlas Eólico Mexicano, en azul. Fuente: Elaboración propia. . . . .	27
3.3. Distribución de datos de viento e histograma de función de densidad de probabilidad en 3D de la distribución de datos de viento en La Rumorosa, Baja California superpuestos para su comparación. Fuente: Elaboración propia. . . . .	28
3.4. Función de densidad de probabilidad basado en la densidad de la distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	29
3.5. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	30

3.6. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	30
3.7. Localización entre los 2 y 3 estados para cada paso o coordenada con el tiempo a lo largo de un año en La Rumorosa, Baja California; ERA5 es rojo y el AEM azul. Fuente: Elaboración propia. . . . .	33
3.8. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en La Rumorosa, Baja California. Fuente: Elaboración propia. . . . .	33
3.9. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en Mérida, Yucatán obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	34
3.10. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en Mérida, Yucatán obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	34
3.11. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en Mérida, Yucatán. Fuente: Elaboración propia. . . . .	36
3.12. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en Cd. Cuauhtémoc, Chihuahua obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	37
3.13. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en Cd. Cuauhtémoc, Chihuahua obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	38

3.14. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en Cd. Cuauhtémoc, Chihuahua. Fuente: Elaboración propia. . . . .	40
3.15. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en CERTE, Oaxaca 2018 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	42
3.16. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en CERTE, Oaxaca 2018 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	42
3.17. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en CERTE, Oaxaca 2018. Fuente: Elaboración propia. . . . .	44
3.18. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en CERTE, Oaxaca 2019 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	45
3.19. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en CERTE, Oaxaca 2019 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	45
3.20. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en CERTE, Oaxaca 2019. . . . .	48
3.21. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en Ojuelos, Jalisco obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	49

3.22. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en Ojuelos, Jalisco obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	49
3.23. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en Ojuelos, Jalisco. Fuente: Elaboración propia. . . . .	52
3.24. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en San Fernando, Tamaulipas 2018 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	54
3.25. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en San Fernando, Tamaulipas 2018 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	54
3.26. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en San Fernando, Tamaulipas 2018. Fuente: Elaboración propia. . . . .	56
3.27. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en San Fernando, Tamaulipas 2019 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	57
3.28. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en San Fernando, Tamaulipas 2019 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	57
3.29. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en San Fernando, Tamaulipas 2019. Fuente: Elaboración propia. . . . .	59

3.30. <i>Clusters</i> o estados obtenidos, divididos en dos subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en Tepexi, Puebla obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	61
3.31. <i>Clusters</i> o estados obtenidos, divididos en tres subconjuntos, a través del método de <i>k-means</i> de la distribución de datos de viento en Tepexi, Puebla obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia. . . . .	61
3.32. Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en Tepexi, Puebla. Fuente: Elaboración propia. . . . .	63

# Índice de tablas

1.1. Reanálisis atmosféricos de ECMWF. Fuente: Elaboración propia con información obtenida de Refs. [8]	13
3.1. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en La Rumorosa, Baja California. Fuente: Elaboración Propia	31
3.2. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en la Rumorosa, Baja California. Fuente: Elaboración Propia	31
3.3. Probabilidades condicionales de coincidencia y no coincidencia de clusters para el sitio en La Rumorosa, Baja California. Fuente: Elaboración propia.	32
3.4. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en Mérida, Yucatán. Fuente: Elaboración Propia	35
3.5. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en Mérida, Yucatan. Fuente: Elaboración Propia	35
3.6. Probabilidades de coincidencia y no coincidencia de clusters para el sitio en Mérida, Yucatán Fuente: Elaboración propia.	36

3.7. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en Cd. Cuauhtemoc, Chihuahua . Fuente: Elaboración Propia . . . . .	39
3.8. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en Cd. Cuauhtemoc, Chihuahua . Fuente: Elaboración Propia . . . . .	39
3.9. Probabilidades de coincidencia y no coincidencia de clusters para el sitio en Cd. Cuauhtemoc, Chihuahua. Fuente: Elaboración propia. . . . .	40
3.10. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada <i>cluster</i> (división en dos) correspondiente para ambas series de tiempo en CERTE, Oaxaca 2018. Fuente: Elaboración Propia . . . . .	43
3.11. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en CERTE, Oaxaca 2018. Fuente: Elaboración Propia . . . . .	43
3.12. Probabilidades de coincidencia y no coincidencia de <i>clusters</i> para el sitio en CERTE, Oaxaca 2018. Fuente: Elaboración propia. . . . .	44
3.13. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en CERTE, Oaxaca 2019. Fuente: Elaboración Propia . . . . .	46
3.14. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada <i>cluster</i> (división en tres) correspondiente para ambas series de tiempo en CERTE, Oaxaca 2019. Fuente: Elaboración Propia . . . . .	47
3.15. Probabilidades de coincidencia y no coincidencia de <i>clusters</i> para el sitio en CERTE, Oaxaca 2019. Fuente: Elaboración propia. . . . .	47

3.16. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en Ojuelos, Jalisco. Fuente: Elaboración Propia . . . . .	50
3.17. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en Ojuelos, Jalisco . Fuente: Elaboración Propia . . . . .	50
3.18. Suma de resultados del estado 2 y 3 entre el AEM y ERA5 para la evaluación de permanencia en cada <i>cluster</i> (división en tres) correspondiente para ambas series de tiempo en Ojuelos, Jalisco. Fuente: Elaboración Propia	51
3.19. Probabilidades de coincidencia y no coincidencia de <i>clusters</i> para el sitio en Ojuelos, Jalisco. Fuente: Elaboración propia. . . . .	51
3.20. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada <i>cluster</i> (división en dos) correspondiente para ambas series de tiempo en San Fernando, Tamaulipas 2018. Fuente: Elaboración Propia . . . . .	55
3.21. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en San Fernando, Tamaulipas 2018. Fuente: Elaboración Propia . . . . .	55
3.22. Probabilidades de coincidencia y no coincidencia de <i>clusters</i> para el sitio en San Fernando, Tamaulipas 2018. Fuente: Elaboración propia. . . . .	56
3.23. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en San Fernando, Tamaulipas 2019. Fuente: Elaboración Propia . . . . .	58

3.24. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en San Fernando, Tamaulipas 2019. Fuente: Elaboración Propia . . . . .	58
3.25. Probabilidades de coincidencia y no coincidencia de <i>clusters</i> para el sitio en San Fernando, Tamaulipas 2019. Fuente: Elaboración propia. . . . .	59
3.26. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada <i>cluster</i> (división en dos) correspondiente para ambas series de tiempo en Tepexi, Puebla . Fuente: Elaboración Propia . . . . .	61
3.27. Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en Tepexi, Puebla . Fuente: Elaboración Propia . . . . .	62
3.28. Probabilidades de coincidencia y no coincidencia de <i>clusters</i> para el sitio en Tepexi, Puebla. Fuente: Elaboración propia. . . . .	63
4.1. Porcentaje de error de ERA5 respecto al AEM de la permanencia anual. Fuente: Elaboración propia. . . . .	65
4.2. Porcentaje de error del Tiempo de vida total anual. Fuente: Elaboración propia. . . . .	65
4.3. Porcentaje de error con respecto al AEM de ERA5 del Tiempo de vida medio. Fuente: Elaboración propia. . . . .	66
4.4. Probabilidad condicional de coincidencia por paso de tiempo en todos los sitios analizados. Fuente: Elaboración propia. . . . .	68
4.5. Porcentaje de no coincidencia por paso de tiempo a lo largo del año entre ERA5 y el AEM. Fuente: Elaboración propia. . . . .	68

# Capítulo 1

## Introducción

### 1.1. Objetivos

#### 1.1.1. Objetivo general

Evaluar la confiabilidad del modelo ERA5 en sitios con disponibilidad de datos reales en México utilizando métodos novedosos de aprendizaje de máquina (*machine learning*).

#### 1.1.2. Objetivos específicos

- Realizar una metodología definida para comparar el modelo ERA5 con datos reales de viento.
- Determinar la precisión de ERA5 para modelar el viento en las coordenadas específicas de 7 torres anemométricas, principalmente en el año 2018.
- Determinar los posibles factores que afectan la precisión de ERA5.
- Determinar si ERA5 podría ser utilizado para un análisis preliminar en un posible sitio productor.

## 1.2. Estado del arte

El estado del arte que a continuación se presenta contiene información relevante a este trabajo relacionado especialmente hacia ERA5 y otros modelos de reanálisis. En los siguientes artículos se trabajó con series de tiempo de viento, en la mayoría de los casos, para determinar el desempeño de ERA5 o para su uso en la estimación del recurso eólico en zonas donde hay poca accesibilidad a mediciones reales. La estructura presenta un resumen general de los trabajos de investigación, así como sus hallazgos más relevantes a este trabajo. Además, se incluyó una breve síntesis de la metodología utilizada para la comparación de datos con ERA5.

### 1.2.1. Investigaciones Nacionales

#### *Advancing global storm surge modelling using the new ERA5 climate reanalysis*

Dullaart, J. C., Muis, S., Bloemendaal, N., & Aerts, J. C. (2020). Advancing global storm surge modelling using the new ERA5 climate reanalysis. *Climate Dynamics*, 54, 1007-1021. [\[2\]](#)

El artículo se centra en analizar ocho eventos históricos de oleadas debido a tormentas derivadas de ciclones tropicales. Para realizarlo, se extrajeron series de datos de viento tanto de ERA5 como del *Advanced SCATterometer*, una herramienta que utiliza información satelital para modelar variables meteorológicas sobre los océanos. En cuatro de los 8 eventos, se validó la altura del oleaje a través de mediciones observadas. En el artículo también se comparó contra la versión anterior de ERA5: ERA-Interim. ERA5 y ERA-Interim tienen una resolución de  $(0.25^\circ \times 0.25^\circ)$  y  $(0.75^\circ \times 0.75^\circ)$  respectivamente. Estos igualmente se compararon con el *Integrated Forecasting System* operacional  $(0.125^\circ \times 0.125^\circ)$  y se encontró que a mayor resolución en el modelo, la representación de los campos de viento y el oleaje asociado es mejor, especialmente para ciclones tropicales de menor tamaño. Como conclusión, el artículo afirma que los recientes avances en el modelado climático global tienen el potencial de incrementar la precisión de los sistemas

de alerta temprana y la evaluación de riesgo en las costas.

Metodología: El tipo de comparación que se llevó a cabo fue con información satelital de ASCAT. Este funciona a través de microondas. ASCAT se centra en océanos y funciona, ya que estas microondas no son susceptibles a las dispersión atmosférica, lo cual permite que atraviesen nubes, neblina y polvo. Se mide la retro dispersión a través de tres ángulos distintos y de este modo es posible obtener rugosidad de la superficie marina y las velocidades de viento. Para determinar el parecido con los modelos meteorológicos se calculó el error de raíz cuadrada media (RMSE), el error medio absoluto (m/s), el sesgo relativo %, y el coeficiente de correlación de Pearson.

### ***Drivers of extreme wind events in Mexico for windpower applications***

Thomas, S. R., Martínez-Alvarado, O., Drew, D., & Bloomfield, H. (2021). Drivers of extreme wind events in Mexico for windpower applications. *International Journal of Climatology*, 41, E2321-E2340. [\[16\]](#)

El artículo utiliza los 40 años disponibles de ERA5 y los compara a través de *clusters* usando el método de *k-means* con el objetivo de analizar principalmente vientos extremos. Éste es exitoso caracterizando lo patrones climáticos dominantes. Los periodos de mayores vientos ocurren en diferentes periodos en el país debido a marcadas divisiones geográficas. En el sur, el incremento de viento se asocia a mareas por anticiclones en el Golfo de México, que posteriormente ocasiona vientos hacia el norte a través del Istmo de Tehuantepec. En el noreste, son los vientos alisios del este los que ocasionan los eventos de mayores vientos. Se busca caracterizar los patrones climáticos dominantes que ocasionan los vientos cercanos a la superficie a través del país. Sin embargo, la falta de observaciones continuas en el país, especialmente de viento y también la falta de modelos de reanálisis atmosférico, dificultan un consenso de los patrones climáticos en las distintas regiones. Para el estudio se utilizaron ocho estaciones anemométricas a dos alturas distintas de entre 20 y 40 m sobre el nivel del suelo para el análisis en: Chiapas, Oaxaca, Baja California Sur, Sinaloa, Tamaulipas, Veracruz, Yucatán y Puebla.

Metodología: ERA5 tiene pasos de tiempo de una hora, mientras que las estaciones meteorológicas miden cada 10 minutos, asimismo las alturas son distintas. Se utilizó

una interpolación bilinear para estimar la velocidad de viento en el sitio y un perfil logarítmico para ajustar las alturas de observación utilizadas en ERA5 a las alturas de medición de las torres. Con esto se produjo una serie de tiempo que puede ser comparada con los datos reales. Una vez obtenido, se realizó un *scatter plot* con los datos de ERA5 ajustados y los datos reales de las torres, donde posteriormente se ajustó la línea de mejor ajuste. Asimismo, la correlación de Pearson resultó altísima, alrededor de 0.9 en las comparaciones. Sin embargo, ERA5 tiende a subestimar las velocidades de viento por lo cual se le aplicó una corrección de sesgo. Para definir los patrones de vientos dominantes se usó *k-means* y *Empirical Orthogonal Functions* (EOF).

### **Estimación del recurso eólico marino en la península de Baja California, México, utilizando la base de datos de ERA5**

Arredondo Gámez, M., García Nava, H., Hernández, R., & Lámbert-Arista, A. (2022). Estimación del recurso eólico marino en la península de Baja California, México, utilizando la base de datos ERA5. [\[1\]](#)

El artículo explora la posibilidad de abastecer la costa de Baja California y Baja California Sur con energía eólica *offshore*. Esto debido a que esta es una zona pobremente interconectada con el resto de país. Los pobladores de la zona usualmente recurren a la quema de combustibles fósiles para abastecerse de energía lo cual contribuye a la generación de emisiones fósiles y con esto al cambio climático. Se utilizarán datos de ERA5 en los últimos 20 años para estimar las velocidades de viento marítimas y posteriormente la energía extraíble con diferentes aerogeneradores. Los resultados indican que existen velocidades promedio de alrededor de 6 m/s en la zona, lo cual resulta favorable para la producción de energía eólica.

Metodología: Se utilizaron 20 años del reanálisis ERA5, del año 2000 al 2019. Para ajustar la altura de la observación a la requerida, se utilizó una ley de potencia tomando en cuenta el coeficiente de rugosidad del océano. Al tener la dirección y magnitud ajustada a la altura necesaria, se realizó el cálculo de densidad de potencia a distintas alturas y para distintos aerogeneradores. Asimismo, se comparó ERA5 usando el coeficiente de correlación de Pearson entre los datos de ERS2, METOP A, METOP B, OCEANSAT-2, QUICKCAT y RAPIDSCAT. En general los coeficientes oscilaron entre 0.8 y 1, por lo

que el artículo consideró aceptable entre el conjunto de datos. Además se confirmó que al realizar diagramas de dispersión entre ERA5 y la información satelital anteriormente mencionada, los datos siguen el mismo patrón en cuanto a la densidad de datos, aunque algunos satélites tienen una mayor concentración alrededor de la línea de ajuste perfecto.

***How well do atmospheric reanalyses reproduce observed winds in coastal regions of Mexico?***

Thomas, S. R., Nicolau, S., Martínez-Alvarado, O., Drew, D. J., & Bloomfield, H. C. (2021). How well do atmospheric reanalyses reproduce observed winds in coastal regions of Mexico?. *Meteorological Applications*, 28(5), e2023. [\[17\]](#)

El artículo analiza tres modelos de reanálisis global: ERA-Interim, ERA5 y MERRA-2, para determinar si son capaces de reproducir datos reales de viento en: Baja California Sur (San Hilario), Baja California Sur (El Paso), Sinaloa, Tamaulipas, Veracruz, Chiapas, Oaxaca y Yucatán; con el fin de abarcar la mayoría de las regiones del país. Se encontró que la capacidad de los modelos para reproducir los datos reales varía significativamente en distintas zonas de México. En sitios donde se encuentran los vientos más fuertes, como en Oaxaca, al sur del país, los coeficientes de correlación son de alrededor de 0.9, mientras que en zonas con menores vientos, como en Baja California Sur, los coeficientes decrecen a 0.5, atribuido a la orografía compleja de la región. Se encontró que ERA5 se desempeña mejor que ERA-Interim y MERRA-2 en la mayoría de los sitios. El artículo también llega a la conclusión de que esto se debe a la mayor resolución con la que cuenta ERA5. Sin embargo, en el sureste de México, ERA5 también contó con el mayor sesgo. Además subestimó las velocidades de viento en la mayor parte del país. En general, se recomienda el uso de ERA5 por sobre los demás modelos, sin embargo es importante tener cuidado en zonas con una orografía más complicada, ya que es en estos casos donde ERA5 tiene dificultad para estimar vientos locales.

Metodología: Se obtuvieron los datos reales a través de estaciones anemométricas a tres alturas diferentes de entre 10 y 50 metros sobre el nivel del suelo. Como en la mayoría de las mediciones, los datos fueron tomados cada 10 minutos. Fue necesario interpolar los datos de los modelos para que la altura correspondiera con las estaciones. Para esto, se utilizó un perfil logarítmico. Una vez que se ajustaron los datos a la altura necesaria, se

obtuvieron los promedios de los dato reales para ajustarlos a 6 horas, 12 horas y promedios diarios para comparación directa contra los tres reanálisis. Los mismos promedios se aplicaron a las series de tiempo de los modelos para poder comparar directamente. Una vez realizado esto, se compararon las series de tiempo y se obtuvieron las correlaciones de Pearson para cada una. Además se obtuvo el error cuadrático (MSE) entre los datos reales y teóricos.

### 1.3. Investigaciones Internacionales

#### *ERA5: The new champion of wind power modelling?*

Olauson, J. (2018). ERA5: The new champion of wind power modelling?. *Renewable energy*, 126, 322-331. [\[12\]](#)

El artículo compara el desempeño de ERA5 y MERRA-2 en términos de modelado de viento en 5 países distintos. Los países analizados fueron: Alemania, Dinamarca, Francia, Suecia, *Bonneville Power Administration* en Estados Unidos. Los datos teóricos fueron comparados contra mediciones reales. El artículo concluye que ERA5 se desempeña mejor que MERRA-2. Las métricas utilizadas, en las cuales ERA5 se desempeñó mejor en todas son: Coeficientes de correlación, media absoluta y raíz del error cuadrático medio. Además, la incertidumbre relacionada a correcciones a largo plazo, es decir, a través de un año de datos, predecir los siguientes 1-5 años, fue un 20 % menor para ERA5.

Metodología: El modelo para cada unidad modelada se mantuvo relativamente sencillo. No se utilizaron datos de validación ni entrenamiento, tampoco fue utilizado un procedimiento de optimización, sino que prueba y error. En horizontal, se utilizó una interpolación bi-linear para puntos de red adyacentes. Para este trabajo no se interpolaron las alturas.

#### *Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and AROME to simulate wind power production over France*

Jourdier, B. (2020). Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and

AROME to simulate wind power production over France. *Advances in Science and Research*, 17, 63-77. [10]

El artículo evalúa cómo se desempeñan distintas series de datos modeladas, comparadas contra datos a escala local, cuando no hay observaciones para realizar métodos de corrección de sesgo. El estudio evalúa ERA5, MERRA-2 de la NASA, COSMO-REA6, AROME NWP y el *New European Wind Atlas*. Se compararon datos de Francia continental a alturas de entre 55 y 100 metros de altura sobre el nivel del suelo. Además, de realizar cálculos correspondientes a potencia en granjas eólicas de Francia. Se concluyó que si bien ERA5 es hábil en el cálculo de las series de viento, tiende a subestimar la magnitud de la velocidad del viento, especialmente en áreas montañosas. AROME y COSMO-REA6 tuvieron mejores resultados en áreas complejas. Por otro lado, MERRA-2 y NEWA tienen sesgos grandes y sobreestiman las velocidades de viento. El artículo también encontró errores en ERA5 en ciclos diurnos. Se reportó el error y el ECMFW aceptó el problema que está asociado al proceso de asimilación, el cual afecta principalmente regiones oceánicas, Europa y Norte América.

Metodología: Para cada sitio, las velocidades de viento se interpolaron. A través de interpolación bi-linear se obtuvieron las coordenadas geográficas deseadas, excepto para NEWA y AROME, en los cuales se utilizó el punto más cercano disponible. También se ajustaron las series en tiempo, a través de una interpolación linear para tener intervalos de tiempo de 30 minutos. Y la última interpolación fue verticalmente para obtener la altura del anemómetro o turbina usando la ley de perfil de potencia para el ajuste. Para la comparación de los datos de viento se utilizaron correlaciones y sesgos locales.

***Wind power assessment in the Caribbean region of Colombia, using ten-minute wind observations and ERA5 data***

Ruiz, S. A. G., Barriga, J. E. C., & Martínez, J. A. (2021). Wind power assessment in the Caribbean region of Colombia, using ten-minute wind observations and ERA5 data. *Renewable Energy*, 172, 158-176. [13]

El artículo resalta la importancia estratégica del uso del recurso eólico en la región del Caribe, así como las dificultades para la obtención de datos debido a estaciones de baja

resolución y muy lejanas entres sí, que además tienen menor cantidad de información que la requerida para un análisis de recurso eólico. En el trabajo se comparan los resultados de 13 estaciones meteorológicas contra los resultados de ERA5. Posteriormente se analiza con distribuciones de Weibull y densidad de potencia para encontrar sitios con potencial eólico comercial. Además, se encontró que hay potencial para energía eólica *offshore*. Se encontraron además factores de capacidad de alrededor del 50 %. En cuanto al desempeño de ERA5, se concluye que si bien ERA5 es capaz de entregar información útil sobre la disponibilidad de viento en tiempo y espacio, se le dificulta estimar la influencia de efectos locales en los patrones de viento. ERA5 tiende a subestimar la magnitud del viento en la mayoría de los sitios analizados, con la excepción de las estaciones en San Andrés donde al contrario sobreestima las velocidades de viento.

Metodología: Se analizaron 13 estaciones meteorológicas con pasos de tiempo de 10 minutos, a una altura de 10m sobre el nivel del suelo. El periodo que cubren estas estaciones es de entre uno y seis años. Se estudiaron los patrones en el área de estudio usando ERA5. Para estimar la densidad de potencia se utilizaron variables de temperatura, presión, humedad específica y temperatura de bulbo húmedo de ERA5. Se utilizó el RMSE o el error de raíz cuadrática media para estimar la precisión de ERA5, además se utilizó una medición de sesgo para estimar si ERA5 subestima o sobreestima los valores reales. Por último, se utilizó el coeficiente de correlación de Pearson. Asimismo, se caracterizaron los datos de viento usando una exploración con medias, desviación estándar y cuartiles.

***Downscaling ERA5 wind speed data: a machine learning approach considering topographic influences*** [\[9\]](#)

Hu, W., Scholz, Y., Yeligeti, M., von Bremen, L., & Deng, Y. (2023). Downscaling ERA5 wind speed data: a machine learning approach considering topographic influences. *Environmental Research Letters*, 18(9), 094007.

El artículo presenta ERA5 como un modelo global meteorológico con alta cobertura pero con deficiencias al examinar efectos en sistemas energéticos a escala local. En el artículo se propone un enfoque de reducción de escala estadística para transformar la resolución de ERA5 en velocidad de viento de 31km x 31 km a 1km x 1km. Como pre-procesamiento se clasificaron las regiones en tres clases basado en la calidad del estimado

de ERA5. Posteriormente se utilizó un método de regresión en cada clase para reducir la escala de la serie de tiempo de viento considerando topografía y datos reales de estaciones meteorológicas. Como resultado se comprobó que la metodología mejora el rendimiento de ERA5 especialmente para terrenos complejos, sin embargo es más precisa para la clase 1 y 2, que son las que tienen un orografía menos compleja y datos ya bastante precisos desde un inicio. El ajuste sí mejora los resultados de las zonas de clase 3, pero no es tan pronunciado como en las otras dos clases. Esto comprueba que la orografía complicada juega un papel fundamental en la precisión de ERA5.

Metodología: Como variables de entrada para la regresión de las series de tiempo de viento, se utilizó: Velocidad de viento de ERA5, dirección de viento de ERA5 y disipación de ondas de gravedad (GWD) de ERA5. Las variables topográficas de los sitios, al no ser dependientes del tiempo se introdujeron como constantes, además de 5 km y 75 km como índice de posición topográfica (TPI). La variable objetivo es la velocidad del viento. Se incluyó la disipación de ondas de gravedad ya que este puede agregar información sobre el impacto de valles, lomas y montañas a una escala de entre 5 km y los 31 km de la red de ERA5. Con esto se puede agregar más información sobre la influencia de la topografía en las velocidades de viento a una escala local. Asimismo, las variables predictoras son normalizadas antes de introducirse al modelo de *machine learning*. Las variables utilizadas para estimar el rendimiento del modelo son: Error de raíz cuadrática media, el coeficiente de correlación de Pearson y el estadístico de Kolmogorov-Smirnov D (KSD). Estas se compararon con mediciones reales.

### Conclusiones

En la literatura analizada se realizan estudios sobre el rendimiento de ERA5, así como su uso potencial para la estimación de potencia eólica o incluso la asignación de aerogeneradores. En este trabajo no se realizarán cálculos de densidad de potencia ni de potencial eólico en la zonas, sino que es enfocado completamente al desempeño de ERA5 y la asignación de estados de vientos. Otros artículos tocan temas de *machine learning* o *clustering*, sin embargo se utilizan en uno de los casos para proponer un modelo para aumentar la resolución de ERA5 especialmente para zonas con topografía complicada o para definir patrones climáticos en regiones. En este trabajo se utiliza el MAPE o error de

porcentaje medio absoluto, tiempos de vida medio, tiempos de vida totales, probabilidad condicional de coincidencia y porcentajes de permanencia para la evaluación del modelo y los estados de viento. En la literatura, comúnmente se utilizan el error cuadrático medio (MSE) y los coeficientes de correlación de Pearson. Como metodología, este trabajo explora el uso de cadenas de Markov para comparar la dinámica de viento entre *clusters*, mientras que la literatura compara las series de tiempo lado a lado, o realiza regresiones para obtener las líneas de mejor ajuste. En general, los autores coinciden que ERA5 supera a su antecesor ERA-Interim, y que también es mejor que MERRA-2, otro modelo de reanálisis comúnmente utilizado. Además se concluye que una orografía complicada disminuye la precisión de ERA5, así como que se le dificulta estimar vientos locales. Por último, la mayoría está de acuerdo en que ERA5 tiende a subestimar las magnitud de la velocidad de viento.

## 1.4. Justificación

El viento juega un papel muy importante en los cambios climáticos y en la toma de decisiones con base en ellos. Hablando de energía, en específico la energía eólica, es imprescindible conocer la distribución del viento en un sitio con aparente potencial eólico, las direcciones preferentes y la magnitud para primero determinar si el sitio es favorable y después tomar decisiones sobre el tipo de aerogenerador necesario, su potencia y distribución. Actualmente no se cuenta con suficientes estaciones meteorológicas para realizar análisis de recurso eólico en gran parte del país. Para realizar dichos análisis, es necesario que el proyecto instale una torre en el sitio con potencial recurso eólico, y que posteriormente espere un tiempo considerable de al menos 1-3 años para tener datos confiables. Después de este tiempo y al realizar el análisis de los datos, se podría llegar incluso a la conclusión de que el sitio no es adecuado y no cuenta con un buen recurso eólico para producir energía eólica. ERA5 es una herramienta de modelado de viento que permite obtener datos en cualquier parte del mundo utilizando coordenadas precisas. Conocer la confiabilidad de ERA5 y en qué condiciones, permitirá hacer un uso de él más acertado tomando en consideración los posibles errores y también sus áreas más fuertes. De esta forma, será posible conocer la distribución de viento aproximada en cualquier

lugar, pero sin perder de vista las consideraciones a tomar en cuanto a precisión.

## 1.5. Hipótesis

ERA5 presenta estados de viento similares a datos reales del Atlas Eólico Mexicano, con los que se puede estimar la precisión del modelo.

## 1.6. Modelos meteorológicos

Los modelos meteorológicos se basan en programas informáticos que calculan información correspondiente meteorológica en el futuro para puntos específicos en el planeta. Utilizan ecuaciones matemáticas para definir el modelo que pretende predecir el comportamiento atmosférico en las siguientes horas, semanas, meses e incluso años. Los modelos se definen por su extensión geográfica. Dichos modelos pueden ser regionales, es decir que solo abarcan una parte del planeta o globales, los cuales son capaces de hacer estimaciones en todo el planeta.

El pronóstico meteorológico utiliza datos obtenidos satelitalmente, a través de barcos, aviones, estaciones meteorológicas, y otras estrategias. Las predicciones utilizadas son determinísticas y probabilísticas. Una predicción determinista predice un evento específico por ocurrir en una localización precisa como un huracán. Las predicciones probabilísticas se basan en eventos que podrían ocurrir en una región durante un periodo de tiempo.

### 1.6.1. ERA5 global reanalysis

#### Introducción

El servicio de *Copernicus Climate Change* está produciendo el analizador ERA5 que pretende comprender un registro detallado de la atmósfera global, la superficie terrestre y el oleaje marino a partir de 1950. ERA5 está basado en el *Integrated Forecasting Sys-*

tem(IFS) [4] utilizando variables de datos 4D-Var (*Four-Dimensional Variational data assimilation*) para su asimilación y modelos de predicción en CY41R2 (Cycle 41R2). El modelo opera utilizando modelos físicos, dinámica central y asimilación de datos. La resolución espacial es de 31 km y tiene un paso de tiempo por hora. Estas medidas son de alta resolución y se conocen como HRES (*High Resolution Components*). Sin embargo también se tiene una resolución reducida compuesta por diez miembros en conjunto referida como EDA (*Ensemble Data Assimilation*). Hasta ahora ERA5 tiene bases de datos a partir de 1979, pero sigue avanzando para contar con el resto de años faltantes en la versión anterior de ERA5 (antes en *preliminary back extension*) [8].

Al combinar óptimamente modelos y observaciones, el reanálisis provee “mapas sin interrupciones” consistentes para variables climatológicas esenciales. De este modo se pretende asegurar la integridad y coherencia en la representación de los principales sistemas de ciclos terrestres (tales como agua y energía). El reanálisis tiene una gran aplicación en las ciencias atmosféricas, donde se utiliza para evaluar el impacto sobre la observación de cambios en el sistema para estimar el progreso en el modelaje y la asimilación de capacidades para obtener climatologías de vanguardia para evaluar anomalías en las predicciones de errores.

ECMWF (*European Centre for Medium-Range Weather Forecasts*) o el Centro Europeo para Predicciones Climáticas de Mediano Rango pretende crear pronósticos para conocer cómo es que el clima evolucionará de manera más probable. Se basa en producir un conjunto de predicciones. Individualmente son descripciones completas sobre la evolución del tiempo. Colectivamente indican la probabilidad de un rango de posibles escenarios climáticos [5]. ERA5 es la quinta generación producida de reanálisis atmosférico. La producción de reanálisis atmosférico comenzó en 1979 con el proyecto FGGE, seguido de ERA-15 a la mitad de los 90s, ERA-40 de 2001 a 2003 y ERA-Interim del 2006 a 2019. En la Tabla [1.1] se muestran los reanálisis que ha tenido ECMWF. Cada reanálisis ha presentado cada vez una mayor resolución horizontal, estrategias de asimilación de datos más sofisticadas y se han visto beneficiados por los avances y desarrollo de los modelos de predicción. Todos los reanálisis ya mencionados cuentan con una componente en tierra y a partir de ERA-40, también tienen oleaje de la superficie del océano y productos de ozonos atmosféricos. Además de ECMWF otros grupos también producen reanálisis

globales atmosféricos como: *MERRA-2 reanalysis* de la NASA GMAO (*Global Modeling and Assimilation Office*), JRA-55 producido por la JMA (*Japan Meteorological Agency*) y CFSR (versión 2) producida por NCEP (*National Centers for Environmental Prediction*) [8].

Tabla 1.1: Reanálisis atmosféricos de ECMWF. Fuente: Elaboración propia con información obtenida de Refs. [8]

Reanálisis	Periodo	Resolución de Malla	Estrategia de Asimilación
FGGE	1979	208 km	OI ( <i>Optimal Interpolation</i> )
ERA-15	1979-1994	125 km	OI
ERA-40	1957-2002	125 km	3D-Var
ERA-Interim	1979-2019	80 km	4D-Var
ERA5	1950-presente	31 km	4D-Var

ERA5 es el sucesor de ERA-Interim, el cual estaba desactualizado y fue detenido a finales de agosto 2019. Dentro de *Copernicus Climate Change Service* ERA5 provee un registro consistente y mejorado para una gran número de variables climáticas esenciales. ERA5 también provee una mejor resolución tanto horizontal como vertical, tiene 10 años de asimilación de datos y modelos además de mejores parámetros de salida como datos de viento a 100 metros de altura. ERA5 además provee información preliminar disponible aproximadamente en 5 días del tiempo real. Esta información es reemplazada a los 2 meses tras un chequeo más exhaustivo. Se espera que ambos resultados raramente difieran pero en caso de hacerlo se notifica a los usuarios.

### ¿Qué es el reanálisis?

El reanálisis combina datos de modelos con observaciones en un sitio específico o en todo el mundo; de este modo se crea un conjunto de datos globales completo y consistente utilizando las leyes de la física. El principio se llama asimilación de datos y se basa en metodologías numéricas utilizadas en centros de predicción de clima. En ese proceso, cada cierto tiempo (12 horas en ECMWF) los pronósticos previos se combinan con observaciones nuevas disponibles. Con esto se pretende crear un nuevo estimado más acertado del estado de la atmósfera. De este análisis se crea un pronóstico actualizado y mejorado. La asimilación de datos es una herramienta matemática la cual busca combinar datos

teóricos, normalmente obtenidos a través de modelos numéricos, con observaciones. La asimilación de datos, a diferencia de otros métodos de *machine learning*, se distingue por utilizar modelos dinámicos del sistema analizado [7].

El reanálisis funciona de la misma manera, pero a una menor resolución para garantizar suministro de datos con un alcance de años en el pasado e incluso décadas. El reanálisis no está condicionado por el tiempo, por lo que hay más posibilidad de recolectar observaciones por mayores periodos para un análisis más certero. De esta forma también es posible mejorar las predicciones realizadas en el pasado a través de la mejora en la calidad del reanálisis en el presente. El sistema de asimilación es capaz de estimar sesgos entre observaciones y de este modo filtrar y encontrar qué datos son de buena calidad y cuáles no. Las leyes físicas permiten realizar estimados en zonas donde la disponibilidad de datos es pobre. El hecho de utilizar una metodología similar alrededor del mundo para ventanas de tiempo regulares utilizando el mismo formato hace que el reanálisis sea muy conveniente de utilizar y pueda estar en mejora continua [7].

### **Frecuencia de Actualización de los Datos**

Los datos lanzados inicialmente se llaman ERA5T y son aquellos datos de 3 meses o menos atrás del tiempo real. Si se detectan fallas serias en ERA5T, estos datos podrían ser distintos a la versión final en los datos de ERA5. Esto es muy raro que ocurra ya en la práctica y se estima que solo sucede una vez cada cierta cantidad de años si es que sucede alguna vez; sin embargo los usuarios son notificados si es que existe algún error. En la CDS (*Climate Data Store*), las actualizaciones son diarias y de alrededor de 5 días atrás del tiempo real y las actualizaciones de las medias mensuales se realizan alrededor de 5 días después del término de cada mes [4].

### **Asimilación de datos y el Sistema Integrado de Predicción, IFS (*Integrated Forecast System*)**

La asimilación de datos variacionales de cuatro dimensiones (4D-Var) utiliza una ventana de 12 horas desde las 09 UTC (*Coordinated Universal Time*) hasta las 21 UTC y

de las 21 UTC hasta las 9 UTC del siguiente día. El paso de tiempo del modelo es de 12 minutos a alta resolución (HRES) y 20 minutos para el modo de menor resolución (EDA) [4].

### **Precisión e Incertidumbre**

Los pronósticos realizados en el pasado y las observaciones que caen dentro de las ventanas de tiempo de 12 horas antes mencionadas son utilizados para especificar todos los análisis durante la ventana. Sin embargo la precisión del análisis no es uniforme a lo largo de cada ventana. Si el modelo y las observaciones no están sesgadas, los errores siguen una distribución Gaussiana y las observaciones son homogéneas en espacio y tiempo, entonces el error será menor en la mitad (cerca del centro) de la ventana de asimilación. Sin embargo en el Sistema Integrado de Predicción, IFS, ninguna de las asunciones anteriores son ciertas por lo que el parámetro en particular y la ubicación son importantes también. Sabiendo esto, es necesario realizar un estudio exhaustivo para determinar en qué zona de la ventana de tiempo el análisis es más preciso [4].

La incertidumbre de ERA5 esta definida por el Conjunto de Asimilación de Datos (EDA). El EDA toma como consideración mayormente incertidumbres aleatorias en las observaciones, en la temperatura de la superficie marina y las parametrizaciones físicas del modelo. Mientras las incertidumbres estén descritas adecuadamente y no existan más fuentes de variación, entonces el EDA será capaz de describir adecuadamente la incertidumbre del reanálisis. Los errores sistemáticos del modelo no son considerados por el EDA y los errores o incertidumbres tampoco tienen correlación. Asimismo el EDA tiene un menor resolución que ERA5, por lo que el sistema EDA no es capaz de detectar todos los errores o incertidumbres de ERA5, esto quiere decir que existen limitaciones en la estimación de errores. Esto se debe a cuestiones económicas y a que el sistema de EDA no fue diseñado para estimación de incertidumbre. Aun así, estas estimaciones otorgan información de dónde y cuándo los resultados del reanálisis son más o menos certeros y dónde existen errores más considerables. Usualmente los sitios con menores incertidumbres son aquellos que cuentan con observaciones y datos más recientes, mientras que los lugares de mayores errores son aquellos más cercanos a ciclones tropicales o en rutas de tormentas [3].

## Componentes de Viento

ERA5 cuenta con la posibilidad de descargar una serie de variables climatológicas para un periodo de tiempo y sitio específico. En cuanto al viento ERA5 entrega las componentes del viento en coordenadas rectangulares y las llama componentes  $U$  y  $V$ .

- **Componente  $U$  del viento:** Este parámetro es la componente en dirección al este del viento y está en unidades de  $\text{m s}^{-1}$ . Es la rapidez del viento horizontal moviéndose hacia el este. Esto quiere decir que valores positivos se dirigen al este y valores negativos al oeste.
- **Componente  $V$  del viento:** Este parámetro es la componente en dirección al norte del viento y está en unidades de  $\text{m s}^{-1}$ . Es la rapidez del viento horizontal moviéndose hacia el norte. Esto quiere decir que valores positivos se dirigen al norte y valores negativos al sur.

La combinación de ambas componente permite conocer la rapidez y dirección del viento horizontal.

# Capítulo 2

## Metodología y tratamiento de datos

### 2.1. Datos de Viento Disponibles en México

En México hay cierta disponibilidad de datos de viento para su uso y análisis para el público en general.

- **Estaciones Meteorológicas Automáticas (EMAS)**
- **Atlas Eólico Mexicano**

#### 2.1.1. Estaciones Meteorológicas Automáticas (EMAS)

Estas estaciones son un conjunto de dispositivos que realizan mediciones de diversas variables de forma automática. Están conformadas por una serie de sensores que pueden transmitir la información meteorológica automáticamente o en algunos casos se almacena la información y luego se recolecta. Su función es recopilar y monitorear algunas variables meteorológicas. Se generan bases de datos con intervalos de 10 minutos y se envía satelitalmente para su actualización cada 1-3 horas por estación. La hora que se utiliza en dichas estaciones es UTC o Tiempo Universal Coordinado; esto debe ser considerado para la correcta interpretación de los datos. El área representativa de las estaciones es de

5 km de radio aproximadamente en terreno plano. Las EMAS miden: velocidad de viento, dirección de viento, presión atmosférica, temperatura y humedad relativa, radiación solar y precipitación. A través de la página web del sistema<sup>[1]</sup> es posible acceder al mapa de México con todas las estaciones disponibles. Se pueden descargar datos de una hora anterior, las últimas 24 horas y los últimos tres meses <sup>[6]</sup>.

Para acceder a periodos más extensos de datos, que es lo que se necesita para un análisis de recurso eólico, se pide ponerse en contacto por teléfono<sup>[2]</sup> o correo electrónico<sup>[3]</sup>.

### 2.1.2. Energía Eólica en el INEEL

En 2015 se diseñó el proyecto Atlas Eólico Mexicano (AEM) con la participación del INEEL en la cooperación entre la Secretaría de Energía y el Gobierno de Dinamarca para la creación y desarrollo de *A roadmap for cooperation on wind resource mapping and forecasting*, es decir una guía para la cooperación en el mapeo del recurso eólico y estimación. El proyecto fue aprobado en 2015 tras presentar la propuesta al Fondo para la Transición Energética y el Aprovechamiento Sustentable de la Energía. De este modo el proyecto comenzó el primero de octubre del mismo año <sup>[11]</sup>.

### 2.1.3. Atlas Eólico Mexicano

“El proyecto Atlas Eólico Mexicano (AEM) es coordinado por el INEEL y en él participan la Gerencia de Estudios de Ingeniería Civil y Ciencias de la Tierra de la CFE, el Instituto de Geografía de la UNAM y la Universidad Técnica de Dinamarca, la cual funge como apoyo científico del proyecto” <sup>[11]</sup>. Los principales objetivos del proyecto son:

- El desarrollo de un atlas eólico nacional para desarrollar la capacidad de planeación y explotación del recurso eólico en el país para la generación de energía eléctrica.

---

<sup>1</sup>[https://smn.conagua.gob.mx/tools/PHP/sivea\\_v2/sivea.php](https://smn.conagua.gob.mx/tools/PHP/sivea_v2/sivea.php)

<sup>2</sup>Tel (55) 2636-4600

<sup>3</sup>[ventanillaunica.smn@conagua.gob.mx](mailto:ventanillaunica.smn@conagua.gob.mx)

- Adecuada valoración del recurso eólico y aplicación de herramientas de localización para la planeación, es decir un atlas de viento numérico y una base de datos mexicana.

Asimismo se tiene como objetivos particulares:

- “Desarrollar capacidades de las instituciones mexicanas para la utilización de modelos de validación, así como la actualización de los datos en años subsecuentes.”
- “Poner al alcance de la sociedad en general, desarrolladores de proyectos eólicos e inversionistas, en una plataforma accesible y sin costo para el usuario, información confiable y validada bajo los estándares aceptados por la industria eólica internacional sobre el recurso eólico disponible en México.”
- “Reducir las barreras de acceso a la información, lo que permitirá acelerar inversiones en el desarrollo de parques eólicos en México.” [11]

El proyecto cuenta con siete torres de medición en siete sitios en la República Mexicana. Las torres tienen una altura de 80 m y realizan mediciones a 20, 40, 60, y 80 m. Dichas torres se encuentran operando ubicadas estratégicamente en los estados de: Baja California (La Rumorosa), Chihuahua (Cd. Cuauhtémoc), Tamaulipas (San Fernando), Jalisco (Ojuelos), Puebla (Tepexi de Rodríguez), Oaxaca (Juchitán, CERTE-INEEL) y Yucatán (Mérida). Con esto se ha logrado tener una base de datos desde 2017 en general de viento y otros parámetros tales como: temperatura, humedad, presión y radiación solar. Dicha base de datos puede ser descargada por cualquier usuario sin ningún costo. Además el proyecto cuenta con la participación de personal de instituciones tales como UNAM, UV, CFE, etc. capacitado para utilizar el modelo *Weather Research Forecasting* en aplicaciones de energía eólica.

Información relevante sobre el proyecto, además de la página de descarga de datos se encuentra disponible a través de la página del INEEL <sup>[4]</sup>. Los datos de monitoreo muestran los valores promedio, mínimos y máximos de la velocidad y dirección de viento.

---

<sup>4</sup><http://aems.ineel.mx/>

Todos los datos se monitorean cada segundo y se promedian en intervalos de 10 minutos descargables. Es posible consultar los datos por día, semana o mes. El AEM sigue en desarrollo y se pretende la instalación de tres torres más para conformar 10 torres en 10 estaciones en total. El proyecto busca realizar un análisis de los datos medidos para realizar simulaciones de viento en todo el país en valores horarios cada 5 km a varias alturas. Se espera tener una primera versión de los datos simulados montados en mapas georreferenciados, para su libre consulta vía internet [11]. Dichas bases de datos tienen una resolución de malla de 4 a 10 km en lapsos de 10 minutos, los que se pretende sean procesados en un software desarrollado por la Universidad Técnica de Dinamarca con el fin de facilitar la planeación y explotación del recurso en México [15]. El Atlas Eólico ha sido promovido por el SENER con el fin de ofrecer a la sociedad, inversionistas y desarrolladores de proyectos eólicos, una plataforma sin costo, validada bajo los estándares de la industria eólica internacional y accesible sobre el recurso eólico disponible en México. Esto permitirá acelerar las inversiones el desarrollo de parques eólicos en México.

## 2.2. Metodología y tratamiento de datos

### 2.2.1. Datos de Viento

Se trabajó con series de viento de un año tanto para ERA5 como el AEM. Como se mencionó anteriormente, en ERA5 es posible descargar las coordenadas  $(x,y)$  de viento que el modelo nombró como  $(u,v)$ . La componente  $u$  de viento representa la dirección hacia el este, mientras que la componente  $v$  la dirección hacia el norte. Angulos negativos de  $u$  y  $v$  representan la dirección hacia el oeste y el sur respectivamente. La combinación de ambas componentes dan como resultado la dirección y rapidez del viento. Se grafican directamente sin necesidad de manipularlas, ya que son coordenadas cartesianas.

En cuanto a los datos disponibles en el AEM, las coordenadas están en forma polar. Se entrega la magnitud de la dirección del viento y un ángulo con respecto al norte, el cual parece ser medido a favor de las manecillas del reloj. Además el intervalo de tiempo es de 10 minutos, en lugar de una hora como en ERA5. Para la comparación de ambas series

de datos es necesario que su paso de tiempo sea el mismo, por lo que se promediaron los datos del AEM cada 60 minutos para obtener intervalos cada hora. Ya que los datos del AEM son experimentales, es necesario limpiarlos antes de trabajar con ellos. Por ello se realizó una depuración condicional. Una vez que se limpiaron los datos se cuantificó la cantidad de datos restantes (en total deberían ser 8760) y se realizó una interpolación en los intervalos donde se encontraban los espacios faltantes. Para esto primero se convirtió tanto la magnitud como la dirección en series de tiempo con su columna de fecha y hora. Una vez obtenidos ambos valores, fue posible realizar un *resample* o interpolación lineal necesaria para cada punto vacío. Asimismo, para el tratamiento de los datos del AEM también es necesario realizar la conversión a coordenadas cartesianas como se muestra a continuación, donde  $v$  es la rapidez.

$$\bar{v} = -v(\text{sen}\theta, \text{cos}\theta) \quad (2.1)$$

Una vez convertidas es necesario graficar los datos para observar cualquier anomalía o incongruencia. De ser el caso se localiza el intervalo anómalo y se elimina, el mismo intervalo debe también ser eliminado de la otra serie temporal correspondiente (ERA5) aunque no sea erróneo, con el fin de no causar sesgos en el análisis o resultados incorrectos. Un ejemplo sería observar una línea sólida de puntos sin aleatoriedad.

### 2.2.2. Cadenas de Markov y método de *clustering* para definición de estados de viento

Una cadena de Markov puede ser descrita como un proceso estocástico sin memoria. Este modelo asume que el sistema solo es capaz de visitar un número finito de estados y solo depende del estado en el que se encontraba anteriormente y no de su historia. En este modelo el sistema es capaz de moverse entre estados con cierta probabilidad. Frecuentemente las predicciones de rapidez de viento se realizan utilizando cadenas de Markov como base. Estas también se pueden utilizar para generar series de tiempo de

viento sintéticas para simulaciones.

Para aplicar cadenas de Markov a los datos es necesario separar la información en conjuntos o *clusters* para poder convertir la información en estados discretos. Es sabido que el viento sí tiene cierto patrón estacional, por lo que el tiempo sí tiene influencia sobre su comportamiento, sin embargo también cuenta con una gran aleatoriedad que permite utilizar las cadenas de Markov para representarlo. Con esto es posible predecir el comportamiento o la tendencia del sistema a través del tiempo.

Una vez que se tiene la información separada en estados de viento discretos, los puntos de viento se ordenan en solo una partición. Es decir, se asocia cada uno a un estado de viento. En este trabajo se les numeró y asoció a un color: Azul es 1, naranja es 2 y verde es el estado de viento 3. La cadena de Markov se ensambla considerando la asociación de cada vector con el estado de viento al que pertenece. Cada estado de viento representa un conjunto de datos que cuentan con una magnitud y dirección preferencial similar. Utilizando cadenas de Markov es posible calcular el tiempo de vida total y el tiempo de vida medio en cada estado de viento [14].

La división en subconjuntos, *clusters* o estados de viento puede realizarse a través de distintos métodos de partición, sin embargo uno de los algoritmos más utilizados es el de *k-means*. Este algoritmo se basa en particionar información no etiquetada en  $k$  grupos. Es un algoritmo iterativo que asigna cada dato de información (cada coordenada) al *cluster* o grupo más cercano con respecto a su centroide. Los  $k$  centroides iniciales se escogen de manera aleatoria y van ajustando su posición hacia el sitio de mayor concentración dentro del grupo en cada iteración. El algoritmo continúa con los cálculos hasta que converge y se tiene la asignación final de todos los puntos en su respectivo grupo o *cluster*. Este método indirectamente crea o define los estados de viento. En este caso se analizarán todos los sitios utilizando dos y tres *clusters* como el número  $k$ ; sin embargo también se inspeccionará visualmente a través de histogramas de densidad cuántos conjuntos o estados de vientos pareciera presentar cada sitio a analizar.

Los histogramas de densidad generan un escala de color de un función de densidad de probabilidad de la serie de datos. De este modo es relativamente sencillo a través de una

inspección visual definir *clusters* de acuerdo a los sitios donde hay una mayor densidad de datos y donde se distinguen concentraciones que pueden representar un área aproximada donde se encuentra el centroide. Asimismo observar las zonas de baja densidad ayuda a determinar las posibles fronteras entre *clusters* o estado de viento para este tipo de datos y análisis en específico. Los histogramas en 3D son otra forma de comparar y determinar las zonas de mayor concentración. En el plano  $(x,y)$  la distribución se ve igual que en un gráfico normal pero las alturas normalizadas son una gran herramienta no solo para la determinación de posibles centroides sino para la comparación entre series de datos; en este caso, entre ERA5 y el AEM.

El **tiempo de vida total** representa la cantidad de tiempo, en fracción de días, que el viento se mantuvo en un estado de viento. Para su cálculo es necesario sumar todos los intervalos de tiempo que el viento estuvo en un *cluster* durante un año.

$$\sum_{i=1}^n t_{n,i} \quad (2.2)$$

donde  $n$  es el número de intervalos.

El **tiempo de vida medio** es la media de tiempo que el viento se mantuvo en un estado durante todo el año. Para su cálculo se debe promediar el tiempo de permanencia en cada estado durante toda la serie de datos.

$$\left( \sum_{i=1}^n t_{n,i} \right)^{-1} * n \quad (2.3)$$

El **porcentaje de permanencia** representa el porcentaje de tiempo que el viento se mantuvo en cada estado durante todo el año. Para esto es necesario realizar el cociente entre el tiempo de vida total y el tiempo total que abarca la serie de datos.

$$\left( \sum_{i=1}^n t_{n,i} \right) * \left[ \sum_{j=1}^{ws} \left( \sum_{i=1}^n t_{n,i} \right)_j \right]^{-1} \quad (2.4)$$

donde  $ws$  es el número de estados.

### 2.2.3. Tratamiento de datos

Se separó en estados de viento la serie de datos a través del algoritmo de *k-means*. Una vez obtenidos los subconjuntos se realizó una tabla para encontrar en qué estado se encuentra cada coordenada de viento. Para esto se comparó la serie de datos con el arreglo que contiene los datos separados en estados de viento. Se busca cada dato en la serie de tiempo dentro de uno de los estados de viento. En el caso de la separación en dos estados, para cada paso se crea un sub arreglo con dos posiciones, donde una contiene cero y la otra un uno dependiendo de en qué estado se localice la coordenada. De este modo es posible saber en qué estado de viento se localiza la serie temporal en cada paso o intervalo. Posteriormente se realizó una tabla que para cada paso de tiempo indicara con un solo número (uno y dos en el caso de dos estados de viento), en qué estado de viento se encuentra la coordenada en dicha posición. Posteriormente se creó una arreglo con subarreglos donde la primera coordenada representa el paso de tiempo en fracción de 365 días (para esta serie de tiempo anual) y la segunda coordenada del subarreglo es el estado en el que se encuentra. De este modo se puede obtener el tiempo de vida total, el tiempo de vida medio y el porcentaje de permanencia de cada estado. Además es posible graficar la dispersión de las velocidades en cada nube según aparecen en el año. El mismo proceso se realiza en el caso de tener 3 estados de viento.

Una vez obtenida la información de a qué estado de tiempo pertenece cada paso de tiempo se realizó una comparación punto a punto. Para cada paso de tiempo de ERA5 y en el AEM se realizó una comparación con respecto al AEM. Si el paso de tiempo  $s$  se encuentra en el estado 1 en el AEM, se evaluó si ese mismo paso  $s$  en ERA5 también se encuentra en el estado equivalente al estado 1 del AEM. De este modo no solo es posible analizar la predicción de ERA5 en cuanto a la permanencia del viento en cada estado, sino su precisión temporal con respecto al original. Si hay coincidencia con respecto al estado 1 del AEM se asigna un 1 en un nuevo arreglo, si hay coincidencia con el estado 2, se asigna un 2 al nuevo arreglo y así sucesivamente. En caso de no haber coincidencia se asigna un cero. De este modo se puede graficar la serie temporal para observar cualitativamente las coincidencias temporales de cada estado, así como los periodos de tiempo de no coincidencia. También es posible cuantificar las incidencias de coincidencia para cada estado, así como las no coincidencias. Una vez obtenidos dichas cifras se calcula

la probabilidad condicional con respecto al AEM de coincidencia. Es decir, de todas las veces que una coordenada en la serie de datos del AEM estaba en un estado determinado, cuántas veces o qué porcentaje ERA5 fue certero.

$$P(\text{ERA5}|\text{AEM}) = \frac{P(\text{ERA5} \cap \text{AEM})}{P(\text{AEM})} \quad (2.5)$$

En cuanto al porcentaje de no coincidencia se realizó el cociente de los eventos de no coincidencia, entre el número total de eventos. Se usó el software *Wolfram Mathematica* para hacer estos cálculos. Cabe destacar que *Mathematica* tienen un código de color semi aleatorio cuando realiza la división en racimos o *clusters*. Para la división en 2 utiliza el color azul y naranja, mientras que al agregar un tercer *cluster* toma el color verde. Estos colores no tienen relevancia en cuanto a la densidad, sin embargo *Mathematica* asocia cada color a un número de *cluster* donde: azul es 1, naranja es 2 y verde 3. Por conveniencia se respetará dicha numeración al referirse a alguno de los *clusters* de cada serie de tiempo. Es importante recordar que la correspondencia por paso de tiempo entre racimos se hizo con respecto al Atlas Eólico Mexicano.

## Capítulo 3

# Casos de estudio de las áreas geográficas de acuerdo al Atlas Eólico Mexicano

### 3.1. Evaluación de ERA5 en México

La metodología y análisis realizado para cada sitio se ejemplifica en este capítulo con el sitio de La Rumorosa en Baja California como caso representativo. Un análisis similar se realizó en todos los sitios descritos a continuación. Los sitios con ciertas peculiaridades se describen más a detalle en su respectiva sección.

#### 3.1.1. La Rumorosa, Baja California

La Rumorosa se encuentra en el municipio de Tecate, Baja California México. Las coordenadas de la estación meteorológica (y utilizadas para obtener la información de ERA5) son  $32^{\circ} 28' 50.47''\text{N}$  y  $116^{\circ} 6' 46.66'' \text{O}$ ; y se encuentra a una altura de 1343 metros sobre el nivel de mar. Se obtuvieron datos del Atlas Eólico Mexicano y ERA5 del año 2018.

La distribución de los datos de viento para ERA5 (en rojo) y el Atlas Eólico Mexicano

(en azul) se muestran en las figuras 3.1. Los datos de ERA5 se graficaron directamente al encontrarse en coordenadas cartesianas, por otro lado se convirtieron las coordenadas de viento polares del AEM en coordenadas cartesianas. Por lo tanto la conversión se realizó como:

$$\vec{v} = -v(\text{sen}\theta, \text{cos}\theta) \quad (3.1)$$

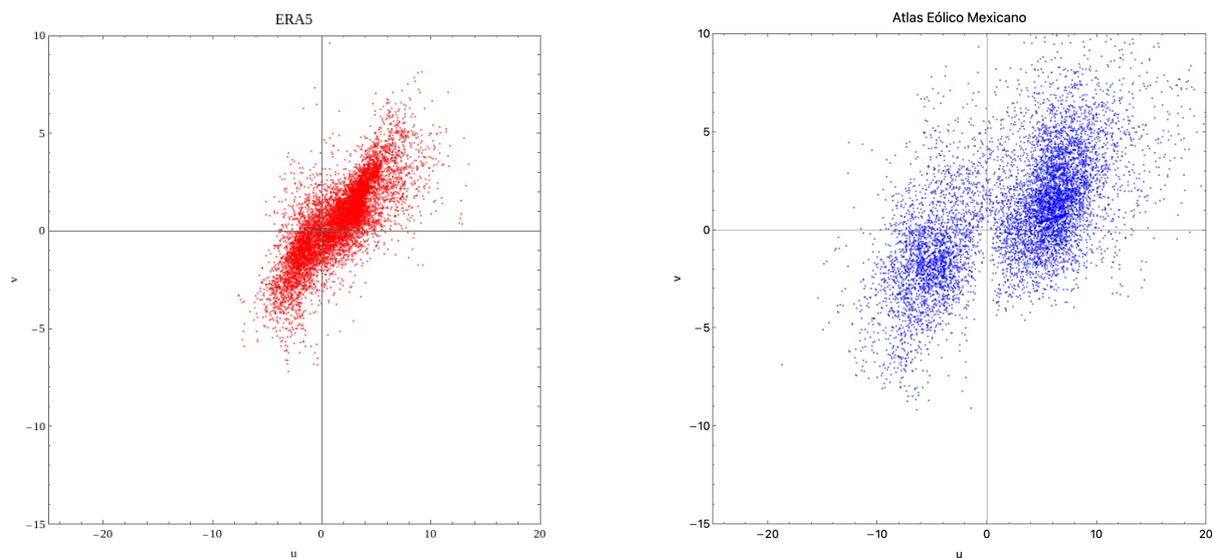


Figura 3.1: Distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5, en rojo y el Atlas Eólico Mexicano, en azul. Fuente: Elaboración propia.

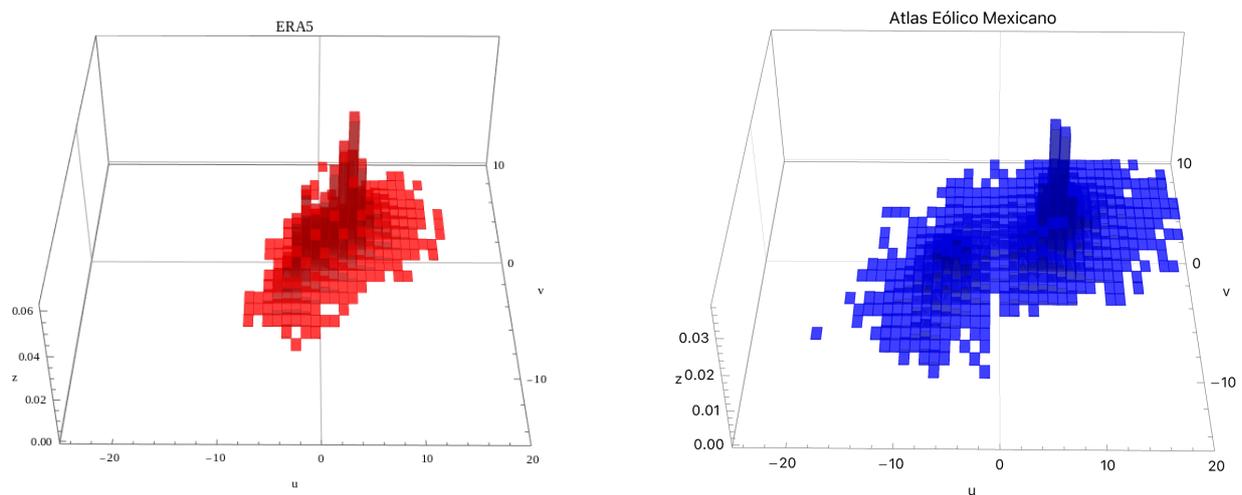


Figura 3.2: Histograma de función de densidad de probabilidad en 3D de la distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5, en rojo y el Atlas Eólico Mexicano, en azul. Fuente: Elaboración propia.

Se puede observar cualitativamente que en el caso de ERA5 los datos están más

concentrados en la zona central, mientras que en el AEM tienen una dispersión más amplia radialmente y cerca del origen no hay una gran concentración de datos. Se distinguen a simple vista dos cúmulos principales mientras que en ERA5 solo hay un cúmulo bien definido. En la Figura 3.3 se muestran ambas series de datos superpuestas para facilitar su comparación. Se observa que la dirección de la distribución de los datos parece ser relativamente correcta, ya que la mayoría de los puntos tienen una dirección dirigida hacia el noreste y otra hacia el suroeste. ERA5 parece no detectar la existencia de algunas corrientes en dirección noroeste y sobreestima las máximas velocidades de viento en el sitio. En la Figura 3.2 es más sencillo observar la concentración de datos con base en la altura de los histogramas de densidad de probabilidad. Tanto para ERA5, como para el AEM la mayor concentración de datos se encuentra en dirección noreste. En cuanto a la velocidad del viento, en el eje  $z$ , ERA5 tiene una concentración de datos de viento cercana a 0.06, mientras que el AEM se queda por la mitad con concentraciones de 0.03, por lo que ERA5 sobreestima las máximas velocidades como se comentó anteriormente.

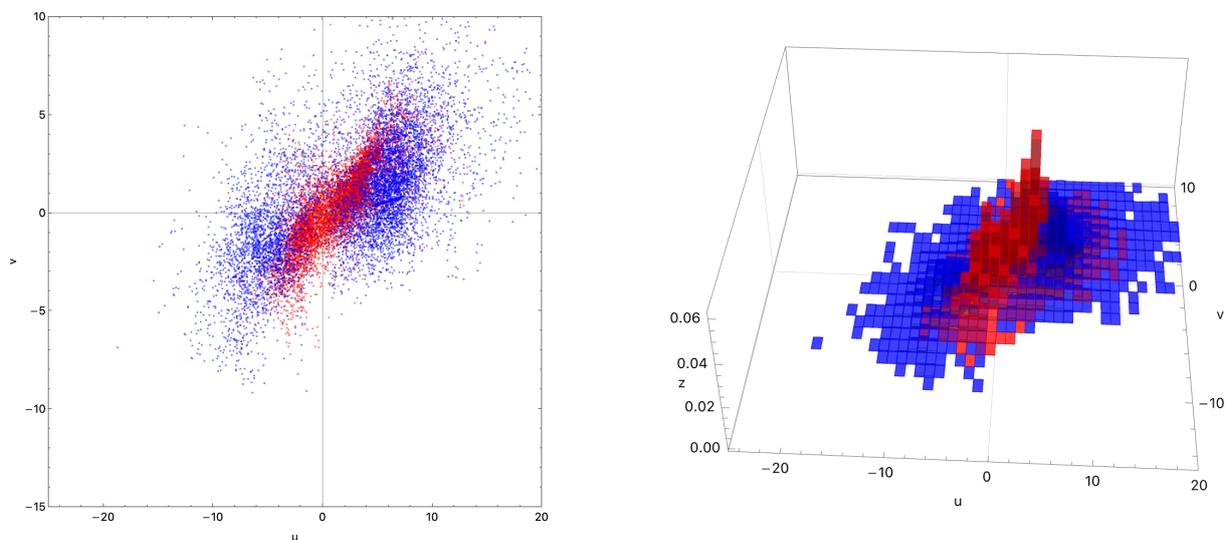


Figura 3.3: Distribución de datos de viento e histograma de función de densidad de probabilidad en 3D de la distribución de datos de viento en La Rumorosa, Baja California superpuestos para su comparación. Fuente: Elaboración propia.

Los datos estimados por el modelo cuentan con mayor orden, mientras que los datos obtenidos a través de mediciones tienen mucha más dispersión y aleatoriedad, así como menor incidencia en el mismo punto. En ambas series de tiempo (ERA5 y el AEM) se observa que alrededor de los puntos de mayor concentración existe un área radial de

desorden o aleatoriedad. En ambos casos coincide principalmente en el noreste con su mayor pico de concentración.

En la Figura 3.4 se muestran histogramas de función de densidad de probabilidad que muestran concentraciones con base a una escala de color. Son muy útiles para localizar visualmente dónde pueden haber subconjuntos, estados de viento o *clusters*. En el AEM es muy evidente la presencia de dos *clusters*, mientras que en ERA5 no se hace tan evidente, sin embargo sí se podrían definir los mismos dos cúmulos en el primer y tercer cuadrante, pero ERA5 sí tiene concentración cerca del origen mientras que en el AEM es casi nula.

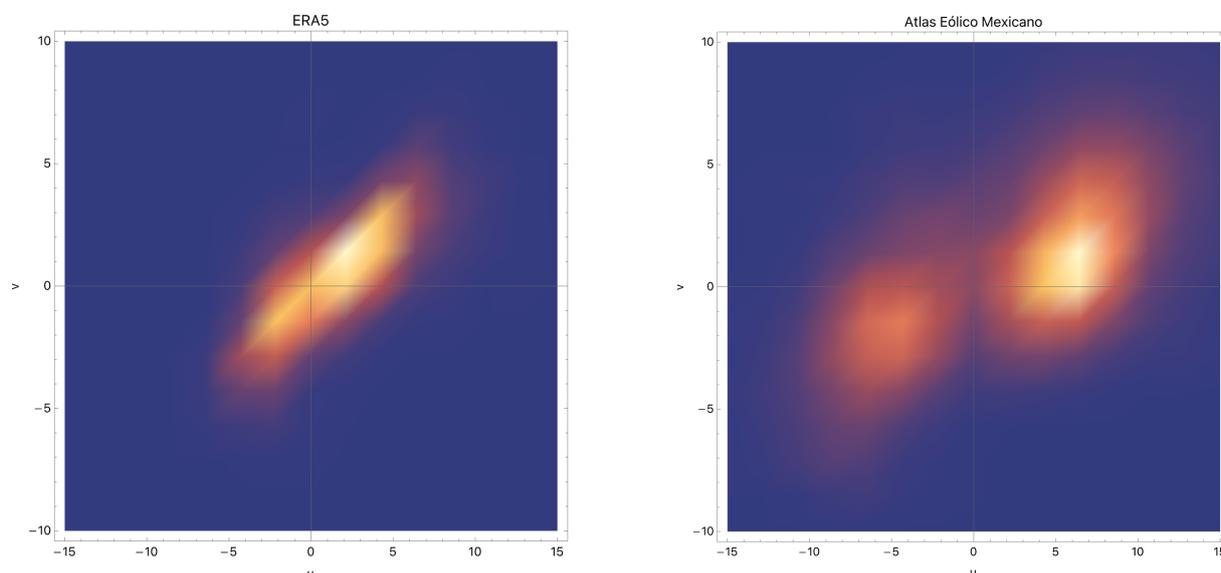


Figura 3.4: Función de densidad de probabilidad basado en la densidad de la distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

La Figura 3.5 muestra la división en dos *clusters* de la serie de datos a través del método de *k-means*. En el AEM la partición se encuentra casi exactamente cruzando el origen, mientras que en ERA5 está un poco desfasada hacia el primer cuadrante. Ambas particiones son en la misma dirección. En la Figura 3.6 se muestra la división en tres *clusters* por el método de *k-means*. En este caso se puede observar más fácilmente la disparidad o similitud entre los *clusters* de datos. Ambos coinciden con una densidad media cerca del origen.

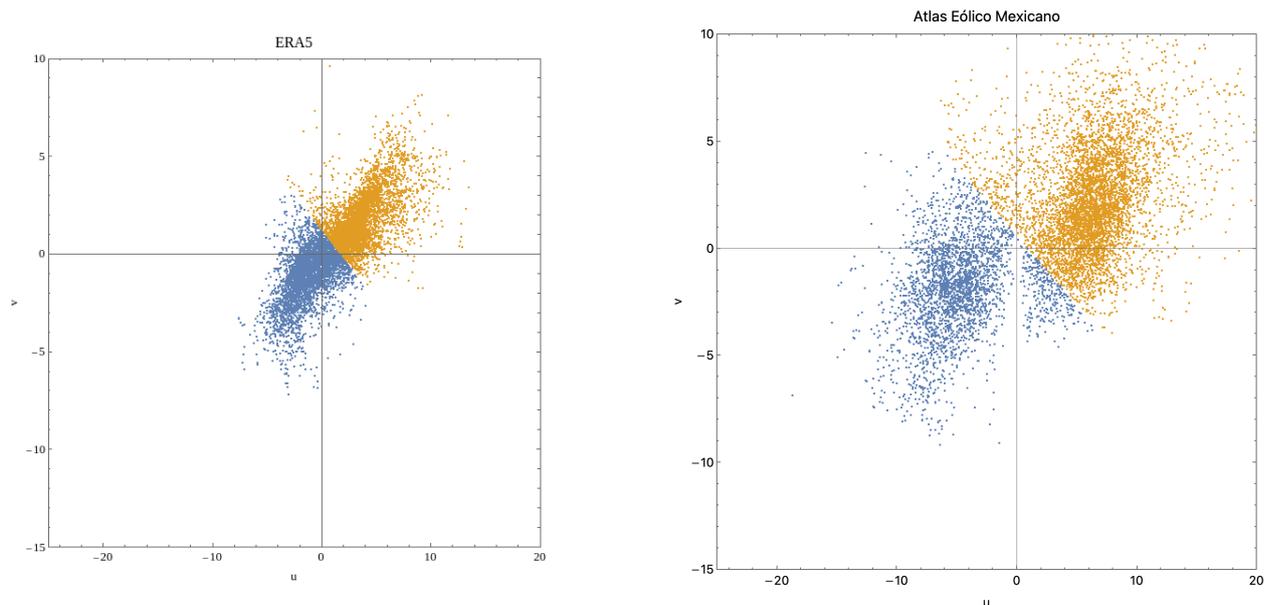


Figura 3.5: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de *k-means* de la distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

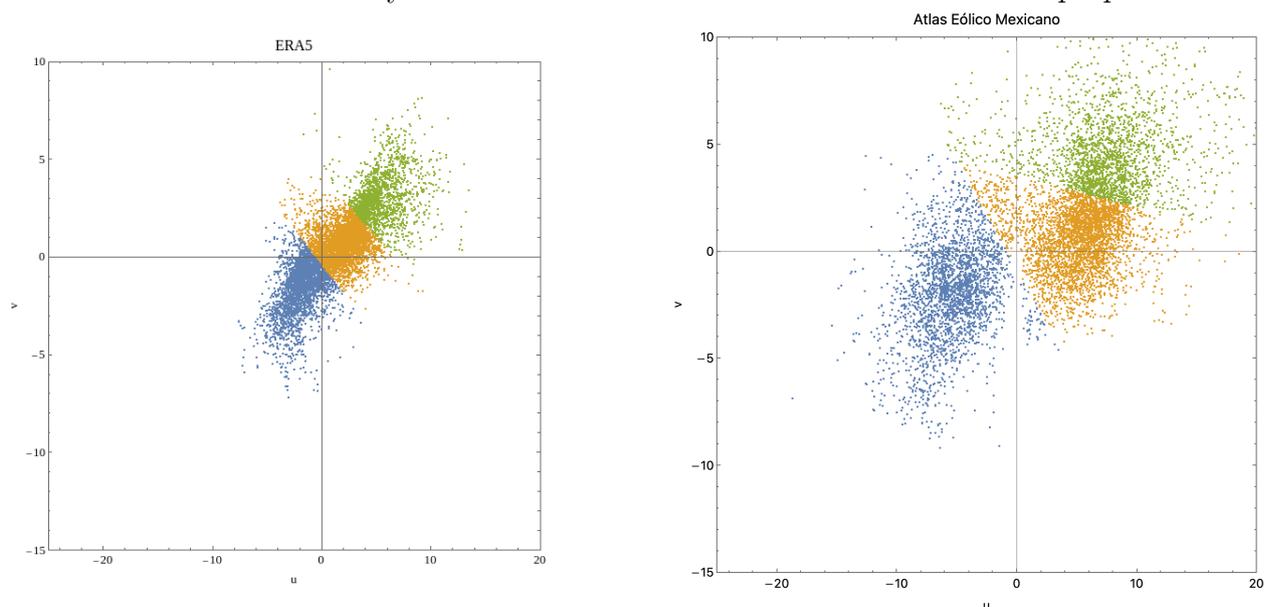


Figura 3.6: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de *k-means* de la distribución de datos de viento en La Rumorosa, Baja California obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

La correspondencia de los *clusters* se realiza con respecto al Atlas Eólico Mexicano. Para el caso de dos estados, el *cluster* 1 (en azul) corresponde en posición al *cluster* 1 de ERA5 (en naranja). El estado 2 del AEM (naranja) corresponde al estado 2 de ERA5 (azul). Para el caso de tres estados la correspondencia con respecto al AEM es: 1:1, 2:2, 3:3, por lo que para ambas particiones en este sitio la correspondencia es directa. Los

tiempos de vida totales, medio y el porcentaje de permanencia con respecto al año son una buena medida de comparación entre ambas señales y se muestran en la Tabla 3.1 y 3.2 para la división en dos y tres estados respectivamente. La asignación de colores en ambas tablas es para facilitar su lectura, recordando que el estado 1 es azul, 2 naranja y 3 verde. En la tabla se muestra el estado que corresponde de ERA5 para cada estado del AEM. Como se había especulado anteriormente el estado 2 del AEM es el que tiene mayor concentración de datos y su tiempo de vida medio también es mayor; por otro lado, los estados correspondientes de ERA5 también siguen este comportamiento. El estado correspondiente al estado 2 del AEM y con mayor densidad, es el estado 2 de ERA5 el cual es el de mayor concentración también, esto para el caso de división en dos *clusters*. Para el caso de división en tres *clusters* el orden de densidad del AEM de mayor a menor para el tiempo de vida total en los estados de viento es: 2, 1 y 3. Para ERA5 el orden también es 2, 1 y 3. Recordemos que la correspondencia con respecto al AEM es directa: 1:1, 2:2, 3:3, por lo que los *clusters* número 2 coinciden en ser lo más densos en ambas series de datos con valores bastante más similares entre sí.

Tabla 3.1: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en La Rumorosa, Baja California. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	32.97 %	44.26 %	67.03 %	55.74 %
Tiempo de Vida Total (días)	120.08	161.17	244.67	203.46
Tiempo de Vida Medio (días)	0.43	0.74	0.88	0.94

Tabla 3.2: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en la Rumorosa, Baja California. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2	3	3
Porcentaje de permanencia anual	29.5 %	31.68 %	46.3 %	45.69 %	24.2 %	22.63 %
Tiempo de Vida Total (días)	107.46	115.25	169	166.79	88.33	82.58
Tiempo de Vida Medio (días)	0.56	0.61	0.37	0.5	0.27	0.58

Tabla 3.3: Probabilidades condicionales de coincidencia y no coincidencia de clusters para el sitio en La Rumorosa, Baja California. Fuente: Elaboración propia.

AEM	Probabilidades			
	1	2	3	No
Coincidencias o no 2 <i>clusters</i>	94.94 %	80.76 %	-	14.57 %
Coincidencias o no 3 <i>clusters</i>	88.74 %	76.27 %	67.56 %	22.16 %

Una vez obtenida la correspondencia entre los estados, se calculó la probabilidad condicional con respecto al Atlas Eólico Mexicano por cada paso de tiempo de coincidencia con ERA5; es decir, se calculó el porcentaje de las veces que ERA5 acertó el estado en el que se encuentran las coordenadas del viento en el AEM para un paso de tiempo específico. La comparación de la incidencia en cada estado para cada serie de tiempo por cada paso de tiempo se muestra en la Figura 3.7, es decir, en qué estado de viento se encontró cada serie de datos a lo largo del año; y en la Figura 3.8 se observa una comparación de incidencia mutua en los estados de viento, donde el cero representa no coincidencia para dicho estado de tiempo. Es decir, con respecto al *cluster* en el que se encuentra para cada paso de tiempo el Atlas Eólico Mexicano, se contabilizó las veces que ERA5 acertó para dicho estado de tiempo correspondiente. En la Tabla 3.3 se muestran los resultados de las probabilidades condicionales y la probabilidad de no acertar tomando en cuenta el paso de tiempo. Para la división en dos *clusters* se observa que la probabilidad de que el modelo meteorológico acierte el *cluster* en el que se encontrará la coordenada de viento es bastante alta para ambos casos; asimismo la probabilidad de que ERA5 falle y no acierte es baja, del 14.57%. Para el caso de tres *clusters* la probabilidad de no acertar sigue siendo bastante baja, sin embargo, a pesar de que la probabilidad condicional de que ERA5 acierte para el *cluster* 1 aumenta con respecto a la división en dos, la probabilidad de acierto en el *cluster* 2 y 3 es ligeramente menor en comparación con los resultados obtenidos en la división en dos. En este caso pareciera que la estimación del modelo es más precisa antes de aumentar el número de divisiones, es decir, manteniendo solo dos. Esto podría suceder, ya que la Rumorosa muestra más claramente una división en 2 en los mapas de densidad (Figura 3.4) y el tercer *cluster* no se distingue de forma tan evidente como en el caso de solamente dos divisiones. De ambas divisiones, el *cluster* con mayor acierto es para el número 1 correspondiente a la división en dos con un acierto del 94.94%.

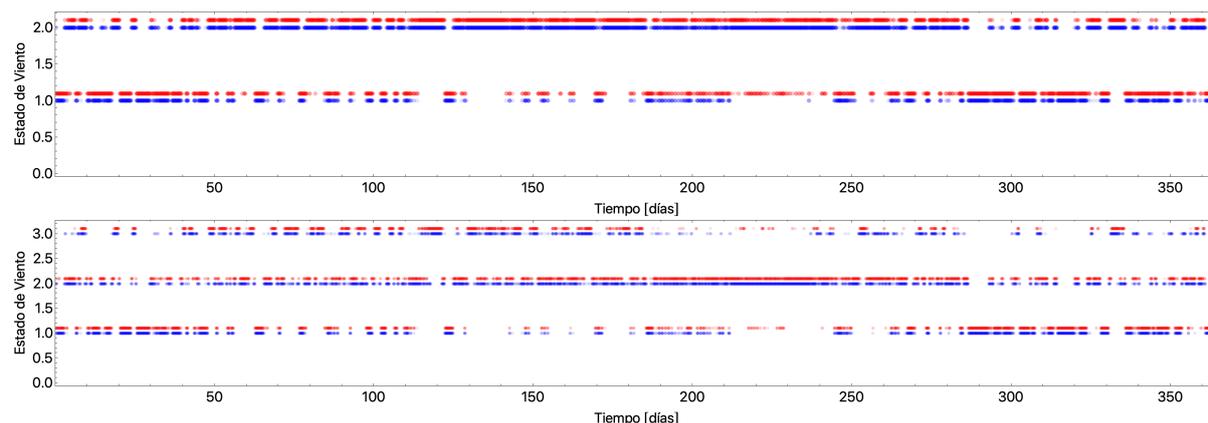


Figura 3.7: Localización entre los 2 y 3 estados para cada paso o coordenada con el tiempo a lo largo de un año en La Rumorosa, Baja California; ERA5 es rojo y el AEM azul. Fuente: Elaboración propia.

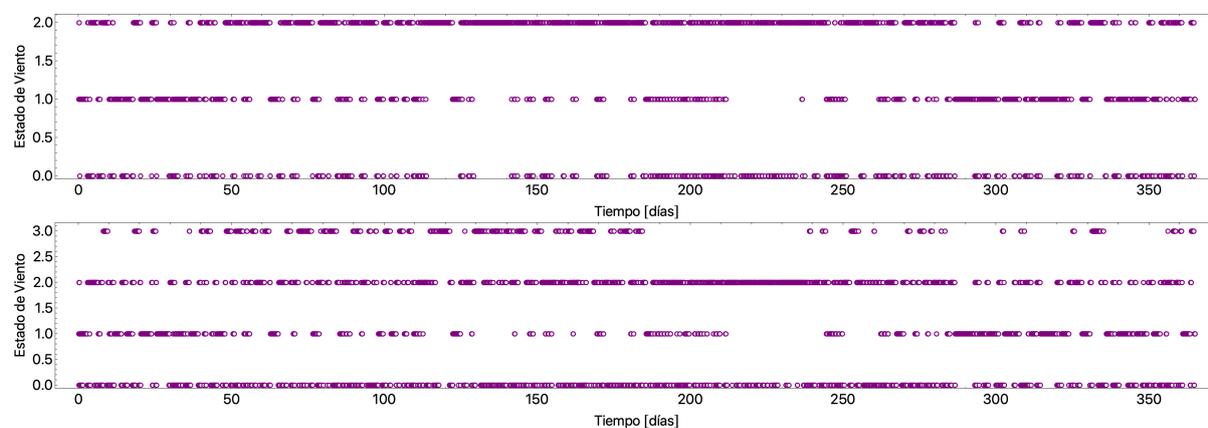


Figura 3.8: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en La Rumorosa, Baja California. Fuente: Elaboración propia.

### 3.1.2. Mérida, Yucatán

Mérida se encuentra en el estado de Yucatan México. Las coordenadas de la estación meteorológica (y utilizadas para obtener la información de ERA5) son  $21^{\circ} 8' 14.92''$  N y  $89^{\circ} 47' 7.36''$  O; y se encuentra a una altura de 3 metros sobre el nivel de mar. Se obtuvieron datos del Atlas Eólico Mexicano y ERA del año 2018.

Los datos del AEM se tuvieron que convertir de coordenadas polares a cartesianas. Ambas distribuciones se aprecian en las figuras [3.9](#) y [3.10](#). Por lo tanto la conversión se realizó como:

$$\vec{v} = -v(\text{sen}\theta, \text{cos}\theta) \quad (3.2)$$

Una vez obtenida la conversión de los datos teóricos se realizó la división en dos y tres *clusters* a través del método de  $k$  medias como se muestra en las figuras 3.9 y 3.10 respectivamente. Para el caso de la división en dos estados, la correspondencia de los *clusters* es directa, donde el estado 1 del AEM corresponde al 1 de ERA5 y 2:2, para el caso de la división en tres *clusters* la correspondencia con respecto al AEM es: 1:1, 2:3 y 3:2. Tanto para los datos de AEM y ERA5 la dirección preferencial es con dirección al noroeste y suroeste con muy poco ruido y aleatoriedad hacia las direcciones contrarias.

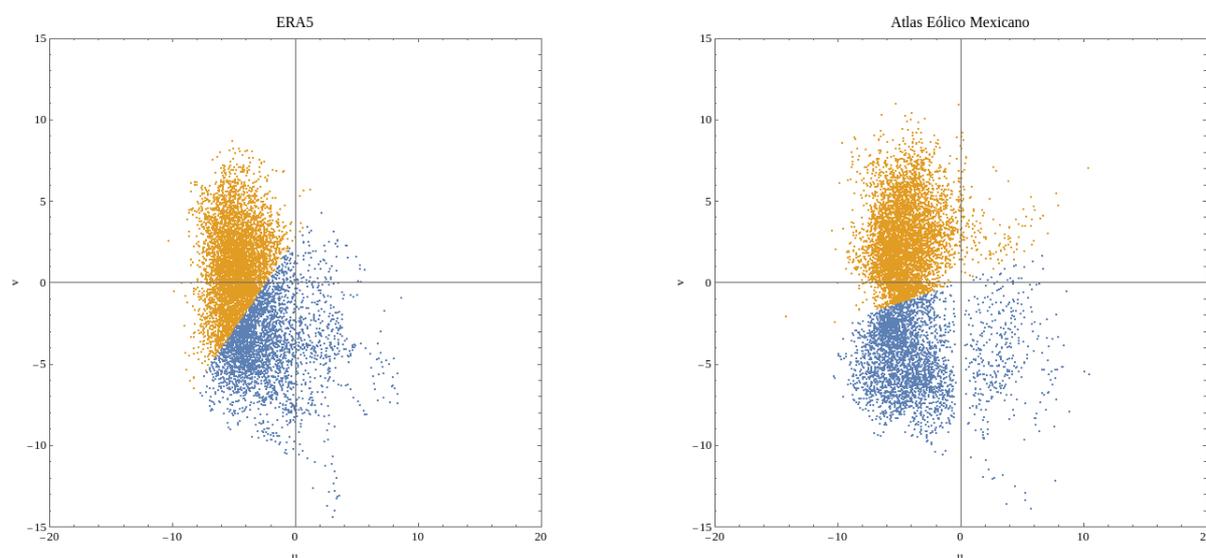


Figura 3.9: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de  $k$ -means de la distribución de datos de viento en Mérida, Yucatán obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

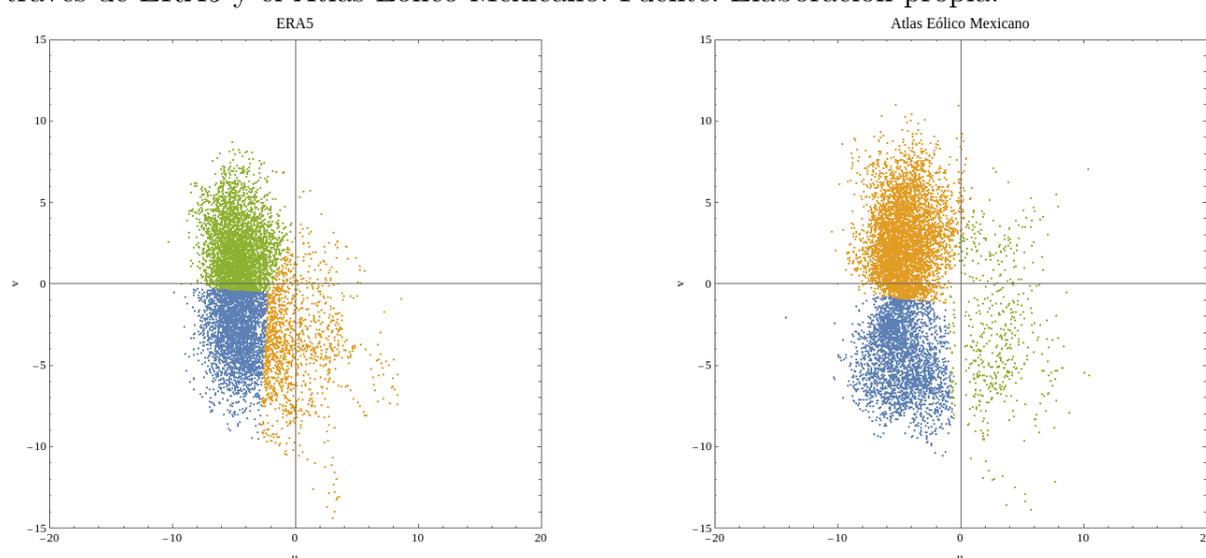


Figura 3.10: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de  $k$ -means de la distribución de datos de viento en Mérida, Yucatán obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

La densidad de datos se puede contabilizar a través de la medida del tiempo de vida en cada estado. En las tablas 3.4 y 3.5 se muestran los resultados de los tiempos de vida de los datos de viento para la división en dos y tres estados. Para el caso de división en dos el *clusters*, el estado con mayor densidad del AEM es el 2, esto coincide con el estado correspondiente (también 2 en este caso) en ERA5. Ambos tienen un porcentaje de permanencia en el año muy similar, el tiempo de vida total también es muy similar y el tiempo de vida medio podría considerarse menos certero, pero sigue siendo razonablemente bueno. El *cluster* 1 del AEM y su correspondiente en ERA5 se acercan aún más en similitud que el anterior. Se puede afirmar entonces que para este caso el modelo meteorológico muestra resultados bastante favorables.

Para el caso de división en tres estados el orden de densidad de mayor a menor del AEM es 2, 1 y 3; y el de ERA5 es 3, 1 y 2. En este caso el orden también coincide (con respecto a la correspondencia de ambas series) y el estado que muestra más similitudes es el correspondiente al 1 en el AEM. El más alejado es el 3, ya que ERA5 sobreestima la incidencia en dicho estado.

Tabla 3.4: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en dos) correspondiente para ambas series de tiempo en Mérida, Yucatán. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	38.01 %	36.64 %	61.99 %	63.36 %
Tiempo de Vida Total (días)	138.75	133.75	225	231.13
Tiempo de Vida Medio (días)	0.3	0.46	0.49	0.79

Tabla 3.5: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en tres) correspondiente para ambas series de tiempo en Mérida, Yucatán. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	3	3	2
Porcentaje de permanencia anual	34.63 %	37.57 %	59.16 %	46.32 %	6.21 %	16.12 %
Tiempo de Vida Total (días)	126.46	137.13	214.67	169.04	22.67	58.79
Tiempo de Vida Medio (días)	0.29	0.37	0.45	0.53	0.21	0.37

Los resultados de las probabilidades condicionales se muestran en la Tabla 3.6. En este sitio, en ambas divisiones, dos y tres estados, los resultados son muy certeros. El error

para ambos casos es bajo, sin embargo la no coincidencia aumenta para la división en tres estados. Para este sitio podría resultar más favorable limitarse a la división en dos. Para el *cluster* 3, en la división en tres estados, la probabilidad condicional asciende al 90.68 %, siendo que anteriormente en la Tabla 3.5 el estado 3 era el que tenía la disparidad más considerable. Esto se debe a que ERA5 sobreestimó la incidencia en dicho *cluster* por lo que la probabilidad condicional asciende al tener más oportunidades de acertar cada vez que una coordenada en un paso de tiempo se encuentra en el estado 3 de ERA5; sin embargo el resultado sigue siendo bastante favorable, ya que de 22.67 días que el AEM incide en ese estado, ERA5 fue capaz de acertarlos en más del 90 %.

Tabla 3.6: Probabilidades de coincidencia y no coincidencia de clusters para el sitio en Mérida, Yucatán Fuente: Elaboración propia.

AEM	Probabilidades			
	1	2	3	No
Coincidencias o no 2 <i>clusters</i>	78.71 %	89.15 %	-	14.82 %
Coincidencias o no 3 <i>clusters</i>	73.24 %	74.89 %	90.8 %	24.69 %

En la Figura 3.11 se muestra en resumen de los resultados discutidos anteriormente. La Figura muestra a lo largo del año por paso de tiempo, en donde el AEM y ERA5 coincidieron o no en un *cluster*. El valor cero representa no coincidencia en ningún estado para ese paso de tiempo.

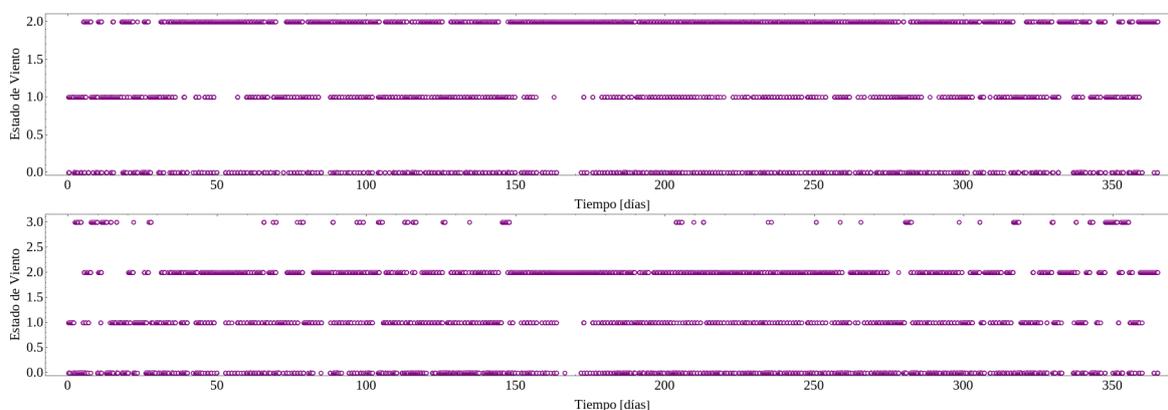


Figura 3.11: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en Mérida, Yucatán. Fuente: Elaboración propia.

### 3.1.3. Cd. Cuahutémoc, Chihuahua

Cd. Cuahutémoc se encuentra en el estado de Chihuahua, México. Las coordenadas de la estación meteorológica (y utilizadas para obtener la información de ERA5) son  $29^{\circ} 1' 14.07''$  N y  $106^{\circ} 57' 7.92''$  O; y se encuentra a una altura de 2128 metros sobre el nivel de mar. Se obtuvieron datos del Atlas Eólico Mexicano y ERA5 del año 2018.

Los datos del AEM se convirtieron de coordenadas polares a cartesianas. Por lo tanto la conversión se realizó como:

$$\vec{v} = -v(\text{sen}\theta, \text{cos}\theta) \quad (3.3)$$

La división en dos y tres *clusters* se muestra en las figuras [3.12](#) y [3.13](#) respectivamente. Para el caso de la división en dos estados, la correspondencia es directa: 1:1 y 2:2 con respecto al AEM; para la división en tres estados la correspondencia se mantiene directa como: 1:1, 2:2 y 3:3. Se observa además, que el viento no tiene una dirección preferencial marcada, sino que la distribución es más bien radial con respecto al origen.

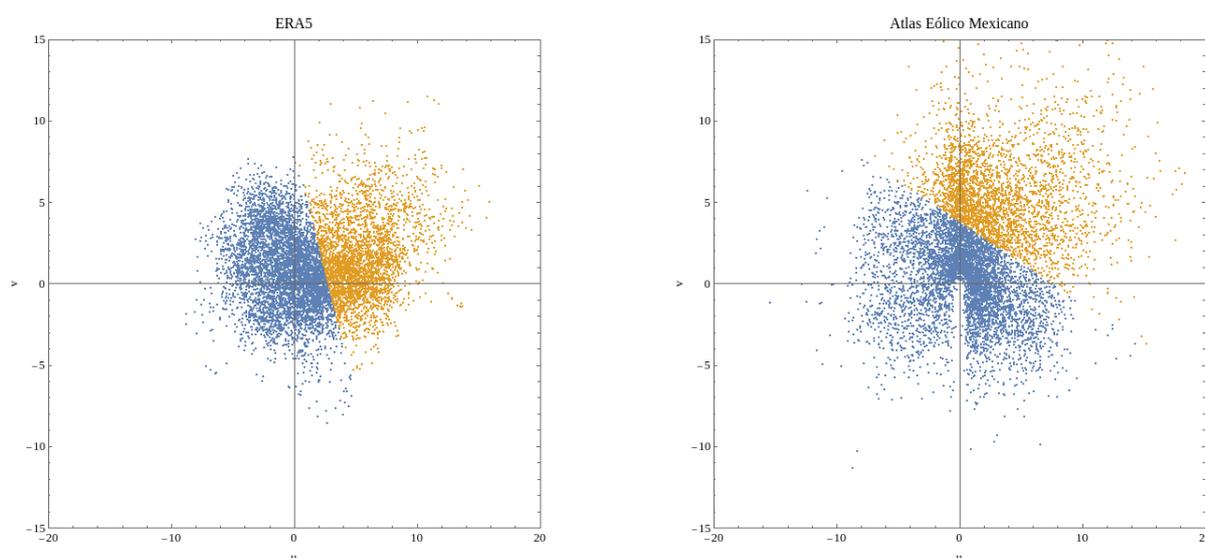


Figura 3.12: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de *k-means* de la distribución de datos de viento en Cd. Cuahutémoc, Chihuahua obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

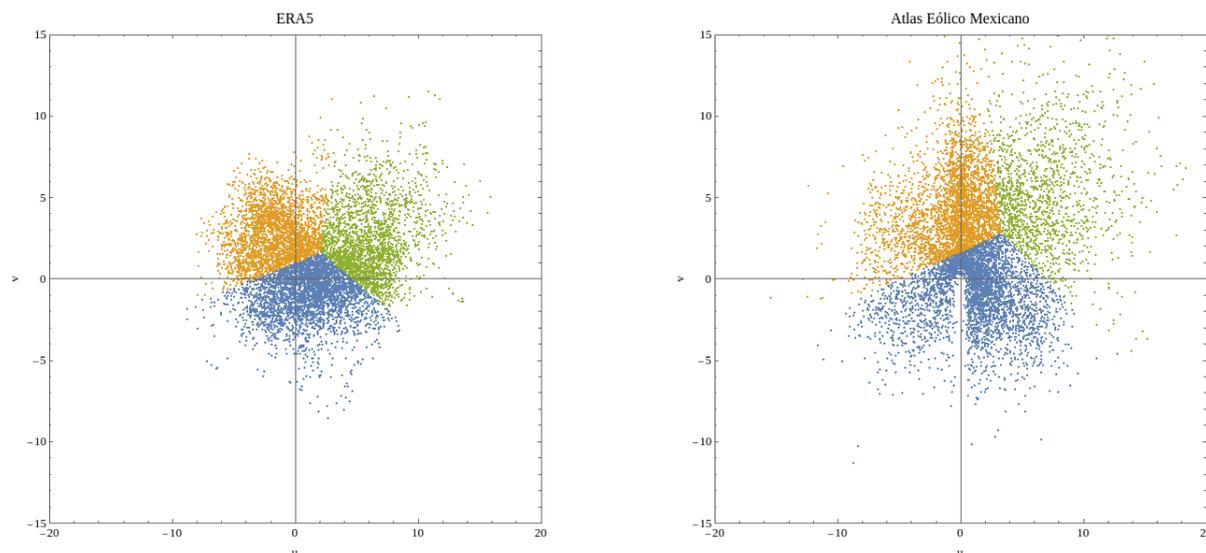


Figura 3.13: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de *k-means* de la distribución de datos de viento en Cd. Cuauhtémoc, Chihuahua obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

Los resultados de los tiempos de vida para ambas distribuciones de datos se muestran en las tablas 3.7 y 3.8. Éstas también son una medida de la densidad de datos de cada estado. Para la división en dos estados la correspondencia del tiempo de vida total es muy similar entre las series de tiempo, en cuanto a la media los valores se alejan. El *cluster* con mayor densidad es el 1 y coincide con el estado correspondiente de ERA5 (el estado 1 también para este caso).

Para el caso de división en tres estados la estimación del tiempo de vida medio sigue sin ser demasiado certera, sin embargo para los estados 1 y 2, el tiempo de vida total tiene una buena correspondencia. El estado que presenta la menor similitud entre los resultados de ambas series es el 3, sin embargo también es el *cluster* de menor densidad o incidencia. En este caso la densidad entre el estado 1 y 2 (recordemos que todo es con respecto al AEM) es similar en magnitud, sin embargo el más denso es el 1, seguido del 2. Esta podría ser una forma de acotar las zonas de mayor producción del sitio y descartar zonas donde las pérdidas serían mayores al tener poca probabilidad de vientos favorables.

Tabla 3.7: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en Cd. Cuauhtemoc, Chihuahua . Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	64.06 %	61.88 %	35.94 %	38.11 %
Tiempo de Vida Total (días)	233.83	225.88	130.75	138.58
Tiempo de Vida Medio (días)	0.42	1.14	0.24	0.7

Tabla 3.8: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en Cd. Cuauhtemoc, Chihuahua . Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2	3	3
Porcentaje de permanencia anual	46.12 %	41.78 %	36.94 %	31.11 %	16.94 %	27.1 %
Tiempo de Vida Total (días)	168.38	152.5	134.75	113.58	61.83	98.38
Tiempo de Vida Medio (días)	0.22	0.49	0.17	0.48	0.18	0.52

Los resultados de las probabilidades condicionales se muestran en la Tabla [3.9](#). Para el caso de división en dos estados, el *cluster* 1 tiene la mayor probabilidad condicional, mientras que la probabilidad de error es del 33%. En el caso de división en tres, el *cluster* 1 y 2 presenta probabilidades razonables y similares, de nuevo se repite el caso, como en Mérida, donde el estado de menor densidad, el 3 en este caso, presenta la mejor probabilidad condicional; esto posiblemente se debe a que ERA5 ha sobre estimado la incidencia en dicho estado. A pesar de la sobre estimación, ERA5 ha sido capaz de detectar los momentos del año de incidencia en un 87% aproximadamente, lo cual es muy favorable. En este caso la probabilidad de error disminuyó, por lo que se puede afirmar que en este caso es más certera la división a tres estados. En la Figura [3.14](#) se observa un resumen de la información presentada anteriormente. Cada punto representa el paso de tiempo donde ambas series de tiempo coincidieron o no, para el caso de cero, con respecto al AEM.

Tabla 3.9: Probabilidades de coincidencia y no coincidencia de clusters para el sitio en Cd. Cuauhtemoc, Chihuahua. Fuente: Elaboración propia.

AEM	Probabilidades			
	1	2	3	No
Coincidencias o no 2 clusters	72.5 %	57.1 %	-	33 %
Coincidencias o no 3 clusters	67.33 %	62.4 %	87.53 %	31.1 %

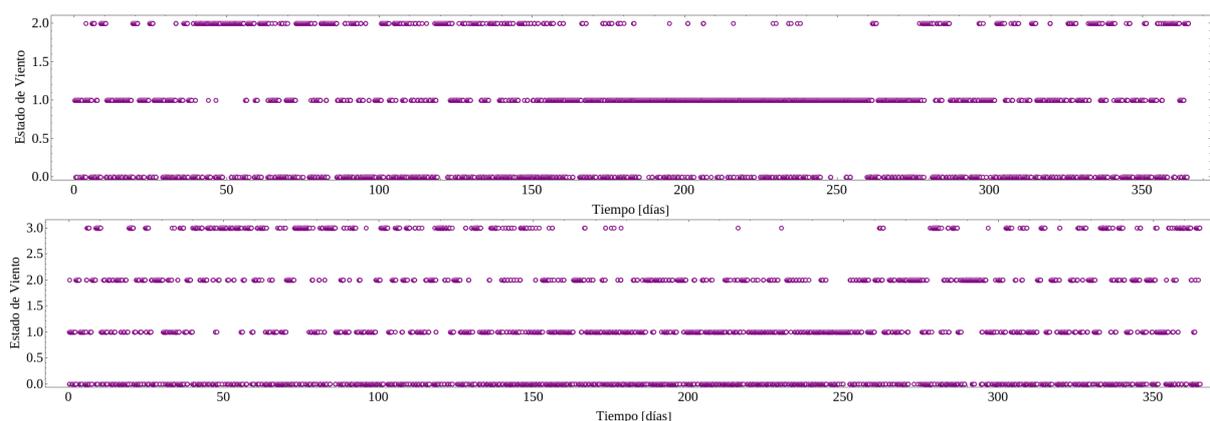


Figura 3.14: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en Cd. Cuauhtémoc, Chihuahua. Fuente: Elaboración propia.

### 3.1.4. CERTE, Oaxaca

CERTE se refiere a las siglas para el Centro Regional de Tecnología Eólica del INEEL y se encuentra en el estado de Oaxaca, México en la Ventosa. Las coordenadas de la estación meteorológica (y utilizadas para obtener la información de ERA5) son  $16^{\circ} 32' 49.27''$  N y  $94^{\circ} 57' 20.83''$  O; y se encuentra a una altura de 31 metros sobre el nivel de mar. Se obtuvieron datos del Atlas Eólico Mexicano y ERA5 del año 2018 y 2019. Esto debido a que al analizar la serie de datos de 2018, se observó un par de líneas definidas en el primer y cuarto cuadrante como se muestra en las figuras [3.15](#) y [3.16](#). Ya que los datos de viento son relativamente aleatorios, no es posible tener líneas definidas en la naturaleza. Este comportamiento en los datos posiblemente se debe a un error en la medición. Podría ser una mala calibración del anemómetro o que algún factor externo interrumpió su funcionamiento normal durante ese periodo de tiempo. Se analizó el año 2019 para ver si este comportamiento se mantenía. Se puede observar en las figuras [3.18](#)

y [3.19](#) que este año también tuvo problemas de medición. Las líneas definidas del primero y cuarto cuadrante se mantiene y encima aparece una línea más concentrada en el tercer cuadrante.

Los datos del AEM se convirtieron de coordenadas polares a cartesianas. Además se utilizó una matriz de rotación a  $-0.2$  radianes para lograr una mejor coincidencia. Ambas distribuciones se aprecian en las figuras [3.12](#) y [3.13](#). Por lo tanto la conversión se realizó como:

$$\vec{v} = R(-11.46^\circ) * -v(\text{sen}\theta, \text{cos}\theta) \quad (3.4)$$

donde  $R(\theta)$  es la matriz de rotación. La rotación fue necesaria tanto para el año 2018 como para el año 2019, lo que sugiere un error ya sea en la dirección de anemómetro en la estación o de los datos teóricos de ERA5, lo cual resulta menos probable. Si los datos fueran confiables esta rotación representaría una metodología para corregir los datos de ERA5. Otro aspecto notorio es la concentración de datos en líneas. En el 2018 hay dos apuntando al noreste y sureste; este patrón se repite en el año 2019 y además tiene otra línea apuntando en dirección suroeste. No se sabe a qué se deben y se revisó que no fuese error de la limpieza de datos o de interpolaciones, incluso se graficaron los datos crudos y no desaparecieron, por lo que podría ser un fenómeno del sitio o error en las mediciones.

### Año 2018

Para las mediciones tomadas en el año 2018 los resultados se muestran a continuación. En la Figura [3.15](#) y [3.16](#) se observa la división en dos y tres *clusters* respectivamente para CERTE, Oaxaca. En el caso de la división en dos subconjuntos, las correspondencias entre *clusters* son directas, es decir: 1:1 y 2:2. Con respecto a la división en tres *clusters*, la correspondencia es: 1:2, 2:1 y 3:3, tomando como referencia el AEM.

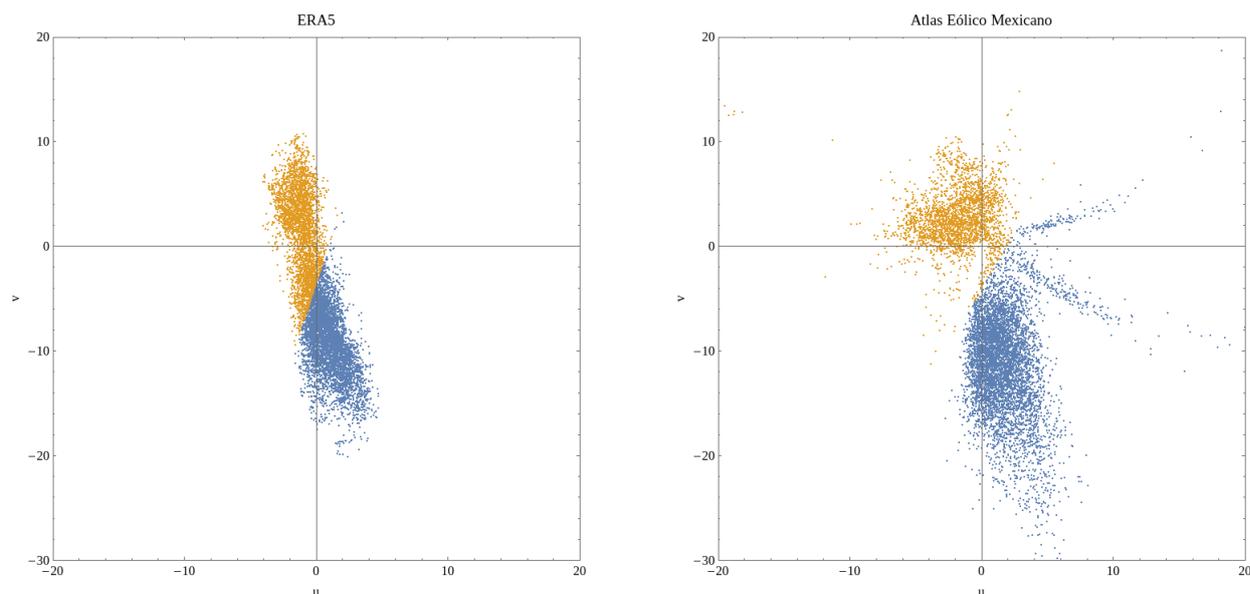


Figura 3.15: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de *k-means* de la distribución de datos de viento en CERTE, Oaxaca 2018 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

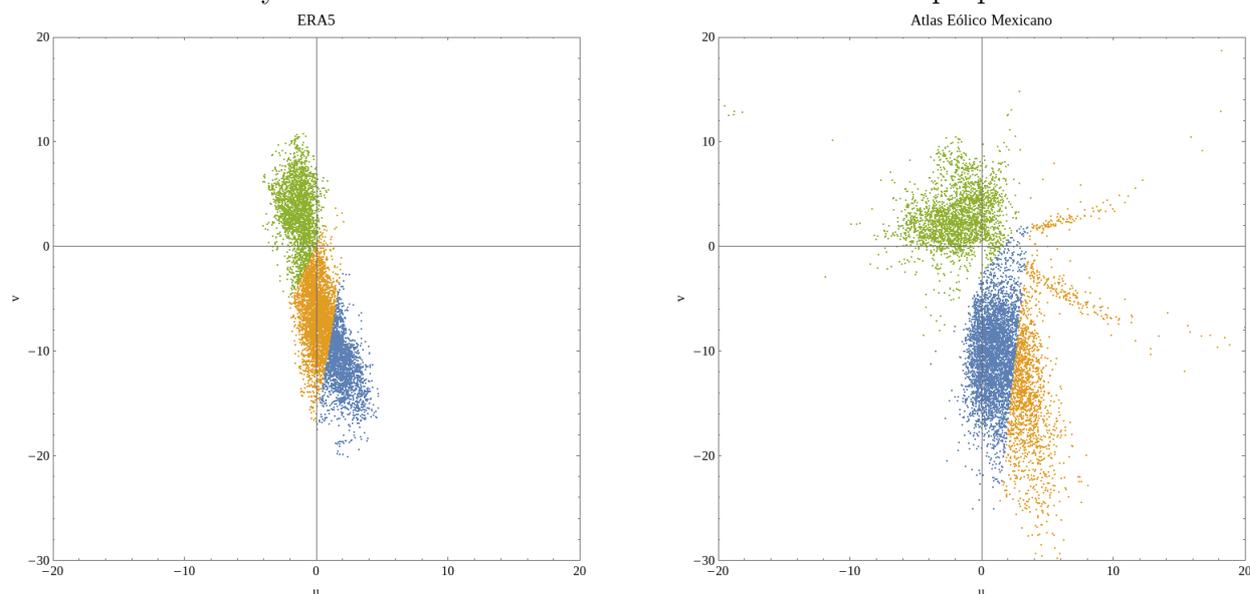


Figura 3.16: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de *k-means* de la distribución de datos de viento en CERTE, Oaxaca 2018 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

En la Tabla [3.10](#) y [3.11](#) se pueden observar los cálculos para la división en 2 y 3 estados respectivamente, del porcentaje de permanencia anual, tiempo de vida total y tiempo de vida medio. Para el caso de la división en 2, los estados más densos son el número 1, los cuales corresponden entre sí. La similitud en permanencia anual es razonable; lo mismo sucede con el tiempo de vida total. Con respecto a estas dos métricas, ERA5 subestima

los valores, pero sobreestima el tiempo de vida medio. Para el caso del *cluster* número 2, ERA5 sobre estima todos los valores.

Para el caso de división en 3, el *cluster* más denso del AEM es el número 1, seguido del 3 y luego el 2. En ERA5, el orden de densidad de mayor a menor es: 2, 3, 1. Estos resultados son favorables, ya que la densidad coincide en los estado correspondientes. Entre los *clusters* 1 y 2 del AEM y ERA5 respectivamente, ERA5 subestima el porcentaje de permanencia anual y calcula aproximadamente 22 días menos de permanencia. El tiempo de vida es relativamente cercano, se compara con los otros estados. Para la correspondencia del estado 2 del AEM, ERA5 esta vez sobre estima el porcentaje de permanencia anual, el tiempo de vida total y e tiempo de vida medio. Es el *cluster* número 3 el que tiene los resultados mas precisos. El porcentaje de permanencia anual, así como el tiempo de vida medio, son muy similares entre sí. Solamente el tiempo de vida medio se ve más alejado.

Tabla 3.10: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en dos) correspondiente para ambas series de tiempo en CERTE, Oaxaca 2018. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	70.35 %	61.26 %	29.65 %	38.74 %
Tiempo de Vida Total (días)	253.63	223.33	108.21	141.42
Tiempo de Vida Medio (días)	1.19	1.3	0.51	0.82

Tabla 3.11: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en tres) correspondiente para ambas series de tiempo en CERTE, Oaxaca 2018. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	2	2	1	3	3
Porcentaje de permanencia anual	50.26 %	44.22 %	20.84 %	27.5 %	28.89 %	28.27 %
Tiempo de Vida Total (días)	183.5	161.42	76.04	100.25	105.46	103.21
Tiempo de Vida Medio (días)	0.31	0.49	0.15	0.50	0.49	0.82

En la Tabla [3.12](#) se pueden observar los resultados de las probabilidades condicionales para cada estado con respecto al AEM. Para este caso pareciera que la división en dos

estados es más favorable, ya que los porcentajes de acierto son altos, y el porcentaje de error es más bajo que en el caso de división en 3. El estado 2 de la división en 2 *clusters* es el que tiene mayor porcentaje de acierto, esto también se debe en parte a que ERA5 para este caso sobre estima la incidencia en dicho estado, por lo que tiene más probabilidades de acertar cuando el AEM se encuentra ahí. A pesar de esto los resultados son bastante positivos.

Tabla 3.12: Probabilidades de coincidencia y no coincidencia de *clusters* para el sitio en CERTE, Oaxaca 2018. Fuente: Elaboración propia.

AEM	Probabilidades			
	1	2	3	No
Coincidencias o no 2 <i>clusters</i>	85.17 %	95.49 %	-	11.77 %
Coincidencias o no 3 <i>clusters</i>	71.22 %	72.07 %	88.23 %	23.69 %

En la Figura 3.17 se observa la información extendida de la tabla anterior para la división en 2 y 3 estados de viento para cada paso de tiempo anual. Los puntos representan coincidencia en algún estado, y en cero se muestra cuando hubo una no coincidencia.

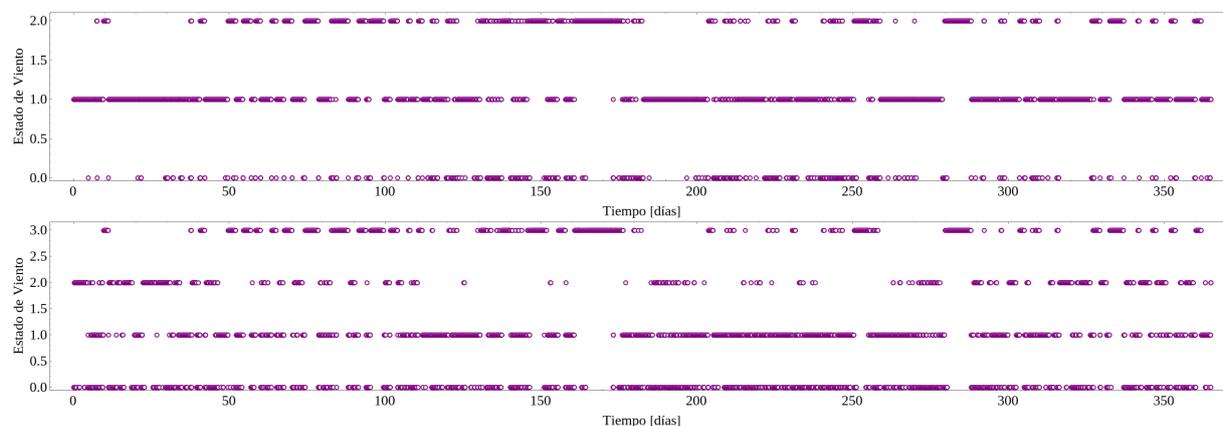


Figura 3.17: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en CERTE, Oaxaca 2018. Fuente: Elaboración propia.

### Año 2019

En las figuras 3.18 y 3.19 se observa la división en dos y tres *cluters* respectivamente para CERTE, Oaxaca en el año 2019. En el caso de las división en dos subconjuntos las

correspondencias entre *clusters* de nuevo son directas, es decir: 1:1 y 2:2. Con respecto a la división en tres *clusters* esta vez la correspondencia sí es directa a diferencia del año 2018 y es: 1:1, 2:2 y 3:3, tomando como referencia el AEM. Cabe destacar que este sitio tuvo un error en los datos experimentales, donde dos intervalos se encontraban vacíos.

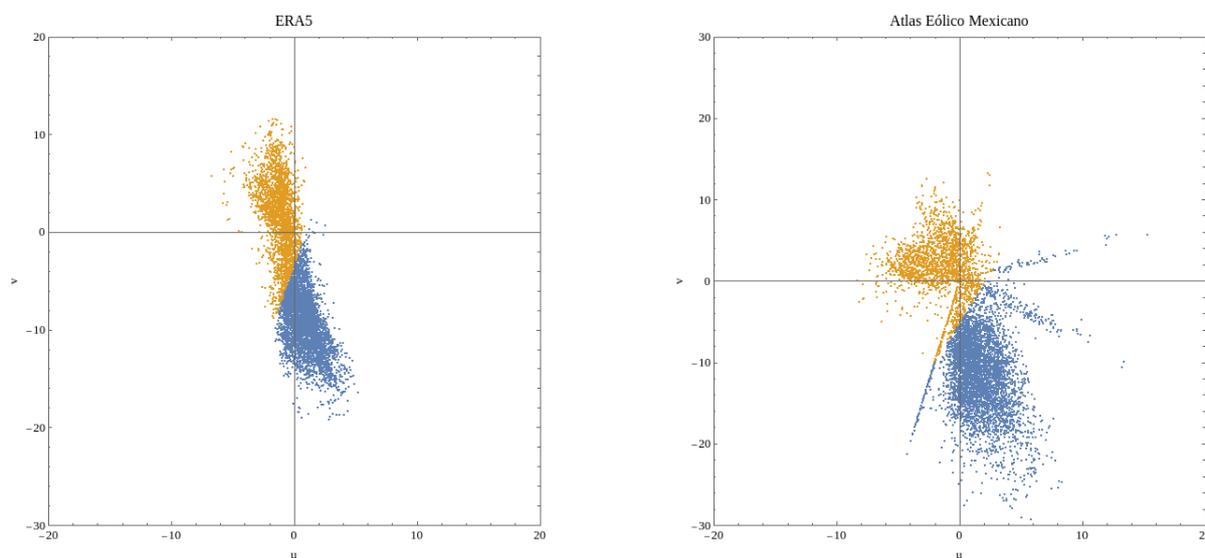


Figura 3.18: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de *k-means* de la distribución de datos de viento en CERTE, Oaxaca 2019 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

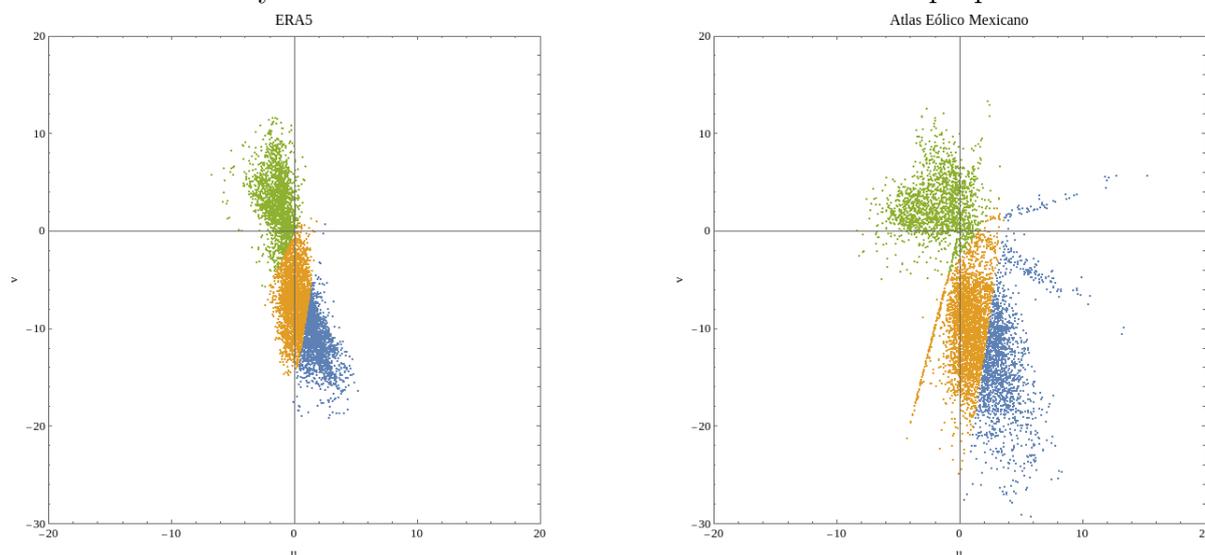


Figura 3.19: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de *k-means* de la distribución de datos de viento en CERTE, Oaxaca 2019 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

La serie de datos del AEM en lugar de contener 8760 datos, solo cuenta con 6925. Por lo tanto se tiene el 79.05 % de un año. Para realizar una comparación adecuada con

ERA5, sólo se utilizaron los datos de ERA5 en los pasos de tiempo donde existían también datos en el AEM. Es decir, se creó una segunda serie de datos de ERA5 donde no se guardaron los pasos de tiempo donde el AEM tenía ausencia de información. De esta forma, ambas series de datos tienen un total de 6925 datos y los cálculos se hicieron con respecto a este número, por ser el total de datos disponibles.

En las figuras 3.18 y 3.19 se muestran ambas series de datos (ERA5 y el AEM) con un *clustering* en 2 y 3 estados respectivamente. Se observa que las líneas anómalas en el primer y cuarto cuadrante se mantienen en este año y además se creó una tercera línea, esta aun más definida, en el tercer cuadrante. Estos probablemente son errores de medición en el sitio. Con respecto a los estados, este año la correspondencia es directa. Para el caso de dos *clusters*, la correspondencia con respecto al AEM, es: 1:1 y 2:2. Lo mismo sucede con la división en 3 *clusters*: 1:1, 2:2, 3:3.

En las tablas 3.13 y 3.14 se muestran los resultados de porcentaje de permanencia anual, tiempo de vida total y medio para la división en 2 y 3 *clusters* respectivamente. En el caso de división en 2, el estado 1 tiene mayor cercanía en sus métricas entre el AEM y ERA5. El estado 2 es similar, pero con diferencias un poco más grandes que el estado 1.

Para la división en 3, el estado más parecido es el 3. El porcentaje de permanencia anual y el tiempo de vida total son muy similares. La precisión se pierde un poco en el tiempo de vida medio. En este último caso, ERA5 sobre estima todas las métricas. Es el estado 1 en este caso, el que tiene mayor error. Sin embargo los resultados siguen siendo similares.

Tabla 3.13: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en CERTE, Oaxaca 2019. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	69.2 %	64.17 %	30.8 %	35.83 %
Tiempo de Vida Total (días)	197.79	183.33	88.88	103.38
Tiempo de Vida Medio (días)	1.05	1.28	0.47	0.72

Tabla 3.14: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en tres) correspondiente para ambas series de tiempo en CERTE, Oaxaca 2019. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2	3	3
Porcentaje de permanencia anual	24.92 %	28.1 %	47.72 %	44.17 %	27.35 %	27.72 %
Tiempo de Vida Total (días)	71.79	81.04	137.708	127.46	78.92	80.0
Tiempo de Vida Medio (días)	0.18	0.49	0.29	0.49	0.46	0.83

En la Tabla 3.15 se muestran los resultados del cálculo de la probabilidad condicional de que ERA5 acierte en un paso de tiempo determinado el *cluster* de incidencia con respecto al AEM. Tanto la división en 2 como en 3 estados, la probabilidad de acierto de ERA5 es bastante certera. Para este sitio, este año también tienen una probabilidad de acierto más alta para el caso de la división en 2 *clusters*. El porcentaje de no acierto también es bastante bajo para este caso. Si bien, los resultados de la división en 2 son muy favorables, la probabilidades de acierto en el caso de división en 3 también son altos.

Tabla 3.15: Probabilidades de coincidencia y no coincidencia de *clusters* para el sitio en CERTE, Oaxaca 2019. Fuente: Elaboración propia.

	Probabilidades			
AEM	1	2	3	No
Coincidencias o no 2 <i>clusters</i>	89.04 %	91.7 %	-	10.13 %
Coincidencias o no 3 <i>clusters</i>	72.88 %	72.38 %	87.27 %	23.42 %

La Figura 3.20 se muestra para ambas divisiones en qué paso de tiempo equivalente coincidieron en cierto estado de tiempo. Los resultados en cero representan no coincidencia en ningún estado de tiempo.

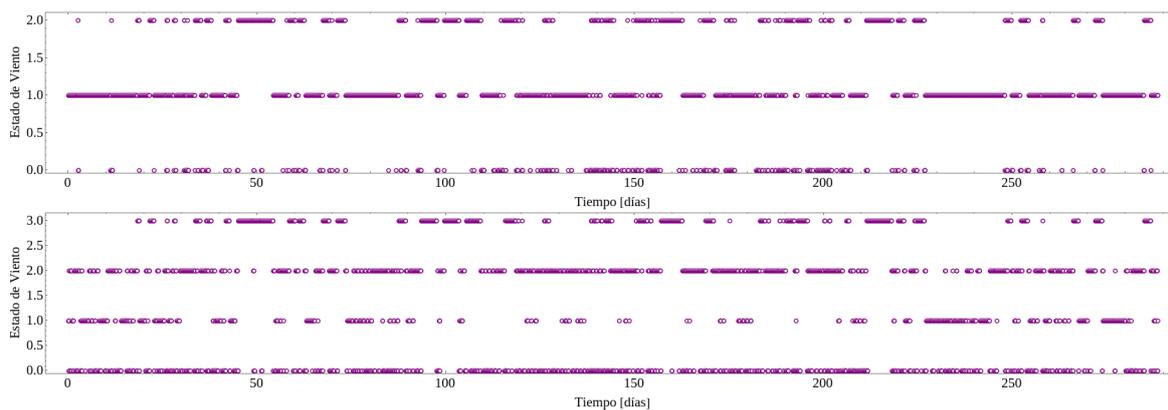


Figura 3.20: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en CERTE, Oaxaca 2019.

### 3.1.5. Ojuelos, Jalisco

Ojuelos se encuentra en el estado de Jalisco, México. Las coordenadas de la estación meteorológica (y utilizadas para obtener la información de ERA5) son  $21^{\circ} 39' 24.12''$  N y  $101^{\circ} 42' 54''$  O; y se encuentra a una altura media de 2221 metros sobre el nivel del mar. Se obtuvieron datos del Atlas Eólico Mexicano y ERA5 del año 2018.

Los datos del AEM se convirtieron de coordenadas polares a cartesianas. Por lo tanto la conversión se realizó como:

$$\vec{v} = -v(\text{sen}\theta, \text{cos}\theta) \quad (3.5)$$

La división en dos y tres *clusters* se muestra en las figuras [3.21](#) y [3.22](#) respectivamente. Para el caso de la división en dos, la correspondencia es directa: 1:1 y 2:2 con respecto al AEM; para la división en tres la correspondencia ya no es tan clara, ya que los datos teóricos están bien definidos de una forma lineal, mientras que en el AEM la concentración de datos para el *cluster* 2 y 3 es hacia el noreste y sureste en lugar de que uno se posicionara alrededor del origen. En este sitio parece que ERA5 tiene problemas identificando el ruido hacia el cuadrante número 4. Esta falla es importante, ya que tiene una concentración considerable al haber sido aislado por un *cluster* debido a que un centroide se encuentra en esta área. Debido a que para la división en 3 no existe una correspondencia clara de ERA5 con respecto al AEM, se tomará la siguiente correspondencia: Tal y como en el

caso de división en 2, para el *cluster* 1, la correspondencia es directa 1:1. Para el caso del *cluster* 2 y 3 del AEM, se considerarán como uno solo, por lo que el estado 1 y 2 del AEM corresponderá a los estados 1 y 2 de ERA5. El cambio en división con esta segunda estrategia se encuentra en que el estado 1 es más pequeño que en la división en 2 *clusters*.

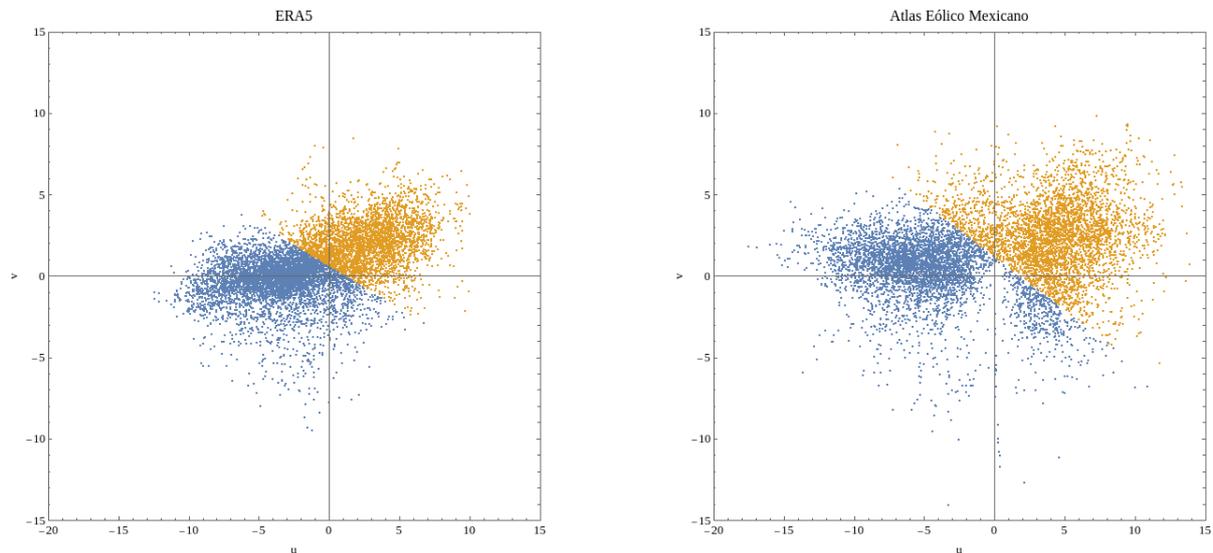


Figura 3.21: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de *k-means* de la distribución de datos de viento en Ojuelos, Jalisco obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

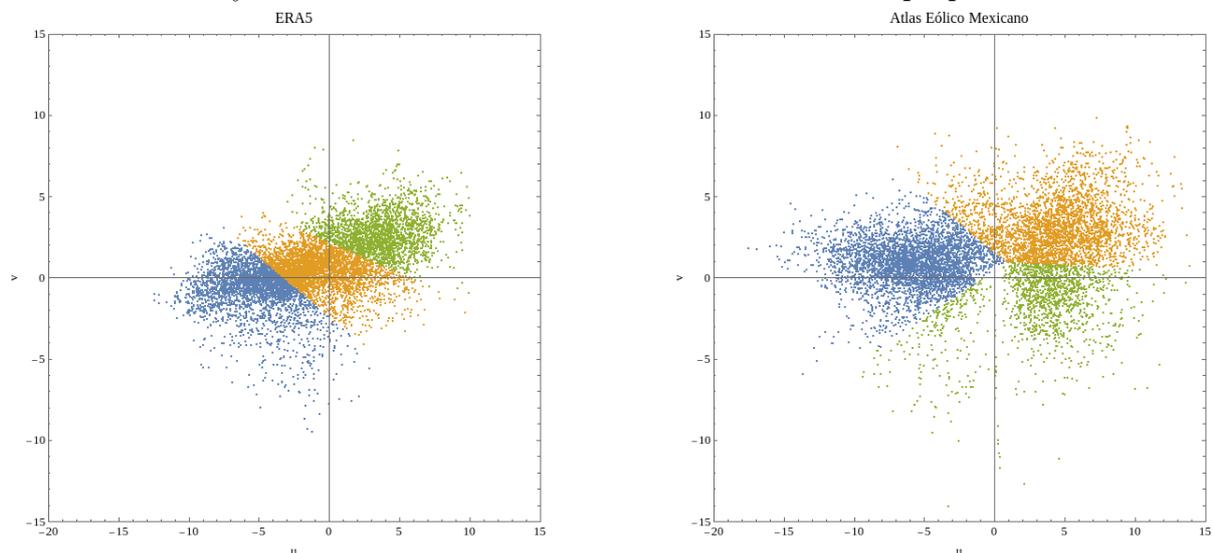


Figura 3.22: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de *k-means* de la distribución de datos de viento en Ojuelos, Jalisco obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

Los resultados de los tiempos de vida para ambas distribuciones de datos se muestran en las tablas [3.16](#) y [3.17](#). Éstas también son una medida de la densidad de datos de cada

estado. Para la división en dos *clusters* los porcentajes de permanencia anual y el tiempo de vida total son bastante parecidos entre sí. Para el estado 1, la permanencia es mayor en el caso de ERA5 y se invierten los resultados para el caso del estado 2. El tiempo de vida medio es el que difiere más, pero tiene un mejor resultado para el estado 2.

Para el caso de división en tres estados se observa que el tiempo de vida medio del *cluster* 1 es muy certero. Son los tiempos de vida total los que se alejan más, sin embargo los resultados no son tan distintos entre sí. Los estados 2 y 3 se analizarán en conjunto. Los resultados de la suma de ambos estados se muestra en la Tabla 3.18. ERA5 tiene mayor densidad en estos estados, ya que sobre estima la permanencia anual, el tiempo de vida total e incluso el tiempo de vida medio. La sobre estimación en todas las métricas es considerable, y los resultados son menos certeros que los obtenido con la división en 2 estados. Para este caso en específico, probablemente sea más conveniente mantener la división en 2 estados en lugar de la de 3.

Tabla 3.16: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en Ojuelos, Jalisco. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	52.24 %	59.83 %	45.94 %	40.17 %
Tiempo de Vida Total (días)	190.5	218.29	167.67	146.63
Tiempo de Vida Medio (días)	0.39	0.83	0.34	0.56

Tabla 3.17: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en Ojuelos, Jalisco . Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2	3	3
Porcentaje de permanencia anual	45.34 %	33.69 %	35.58 %	40.16 %	17.25 %	26.14 %
Tiempo de Vida Total (días)	165.54	123.0	129.88	146.33	62.75	95.42
Tiempo de Vida Medio (días)	0.43	0.47	0.23	0.32	0.13	0.5

Tabla 3.18: Suma de resultados del estado 2 y 3 entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en tres) correspondiente para ambas series de tiempo en Ojuelos, Jalisco. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2 y 3	2 y 3
Porcentaje de permanencia anual	45.34 %	33.69 %	52.83 %	66.3 %
Tiempo de Vida Total (días)	165.54	123.0	192.63	241.75
Tiempo de Vida Medio (días)	0.43	0.47	0.36	0.82

Los resultados de las probabilidades condicionales se muestran en la Figura 3.19. En cuanto a la división en dos *clusters* se observa que la probabilidad condicional de coincidencia en cada paso de tiempo con respecto a el AEM es mayor para el estado 1; se encuentra por arriba del 65 % lo cual es un resultado favorable. En cuanto al estado 2, la probabilidad condicional es casi del 44 %, lo cual es bastante bajo con respecto al *cluster* 1 y otros sitios analizados. La probabilidad de no coincidencia de toda la serie de datos es casi del 45 %.

Tabla 3.19: Probabilidades de coincidencia y no coincidencia de *clusters* para el sitio en Ojuelos, Jalisco. Fuente: Elaboración propia.

	Probabilidades			
AEM	1	2	2 y 3	No
Coincidencias o no 2 <i>clusters</i>	65.03 %	43.76 %	-	44.92 %
Coincidencias o no 3 <i>clusters</i>	36.78 %	-	67.78 %	46.53 %

Para el caso de la división en 3, la probabilidad condicional de coincidencia para cada paso de tiempo del estado 1 es casi del 37 %, lo cual es mucho menor a los dos casos analizados para la división en 2 estados. A pesar de ser una zona similar la que se analiza en los *clusters* 1 de ambas divisiones, el caso de partición en dos resulta mejor para este sitio. Para el estado en conjunto 2 y 3 se analizó la probabilidad condicional de que ERA5 se encontrara en el estado 2 o 3, cuando el AEM también se encontrara en cualquiera de dichos estados. La probabilidad para este caso es más alta y tiene coherencia debido a que ERA5 sobreestimó la incidencia en este caso, por lo que las probabilidades de acertar en este caso aumentó. Sin embargo, la probabilidad de error sigue siendo alta para ambas

particiones.

En la Figura 3.23 se observa un resumen de la información presentada anteriormente. Cada punto representa el paso de tiempo donde ambas series de tiempo coincidieron o no, para el caso de cero, con respecto al AEM. Para el caso de división en 3, se representó la incidencia en el estado 2 y 3 colocando los valores en 2.5 en el gráfico.

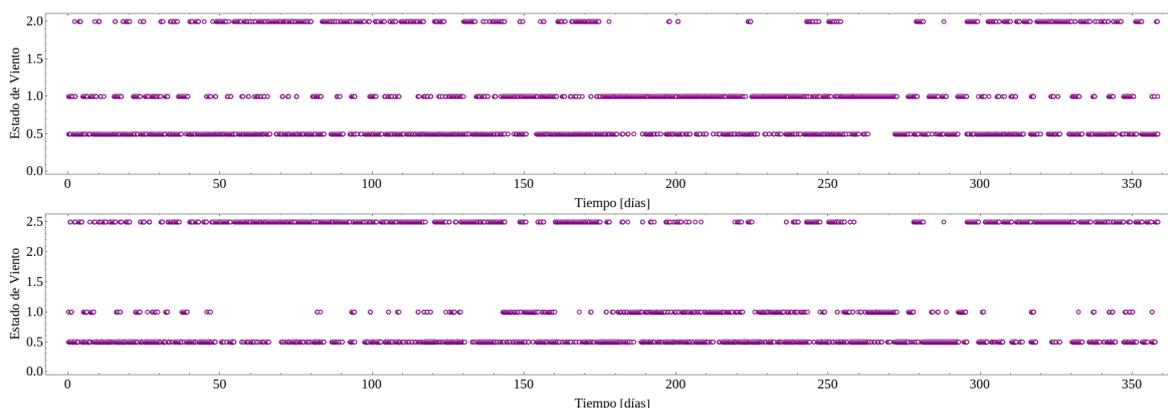


Figura 3.23: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en Ojuelos, Jalisco. Fuente: Elaboración propia.

### 3.1.6. San Fernando, Tamaulipas

San Fernando se encuentra en el estado de Tamaulipas, México. Las coordenadas de la estación meteorológica (y utilizadas para obtener la información de ERA5) son  $25^{\circ} 1' 12''$  N y  $98^{\circ} 5' 24''$  O; y la torre se encuentra a una altura media de 32 metros sobre el nivel del mar. Se obtuvieron datos del Atlas Eólico Mexicano y ERA5 del año 2018 y 2019.

Los datos del AEM se convirtieron de coordenadas polares a cartesianas, sin embargo fue necesario rotar la conversión para hacer coincidir la distribución de los datos de viento del AEM, con los datos teóricos de ERA5. Esto puede suceder debida a un error de configuración en los equipos de medición. Por lo tanto la conversión se realizó como:

$$\vec{v} = v(\text{sen}\theta, \text{cos}\theta) \quad (3.6)$$

La razón para el análisis de dos años para este sitio es por un error encontrado en los datos experimentales. Se analizaron dos años para ver si el error desaparecía, sin embargo

tanto 2018 como 2019 presentan un error similar. En 2018 se empezó a crear una línea pronunciada en el dato 4044 hasta el dato 5045. Al revisarlo se encontró que sí hay datos y las magnitudes sí cambian, sin embargo es la dirección la que tiene cambios mínimos de entre  $295.33^\circ$  y  $308.55^\circ$ . Dicho rango de datos se eliminó tanto en el AEM como en ERA5 para que las series de datos coincidieran adecuadamente. En el año 2019 sucede algo similar; la línea se empieza a crear en el rango de datos 4201-7115. La magnitud sí cambia, pero la dirección oscila entre  $284.62^\circ$  y  $309.23^\circ$ . En esta caso también se descartó este intervalo en ambas series de datos.

### Año 2018

En las figuras [3.24](#) y [3.25](#) se puede apreciar la forma de las distribuciones de datos en el plano, su comparación entre las series de datos y la división en 2 y 3 *clusters*. A primera vista, ERA5 hace un excelente trabajo simulando la distribución de datos. Ambas series son sumamente parecidas, con la excepción de que el AEM tiene un poco más de ruido o aleatoriedad. La correspondencia de estados tanto para la división en 2 como la división en 3 es directa, donde el *cluster* 1 del AEM coincide con el *cluster* 1 de ERA5 y así sucesivamente.

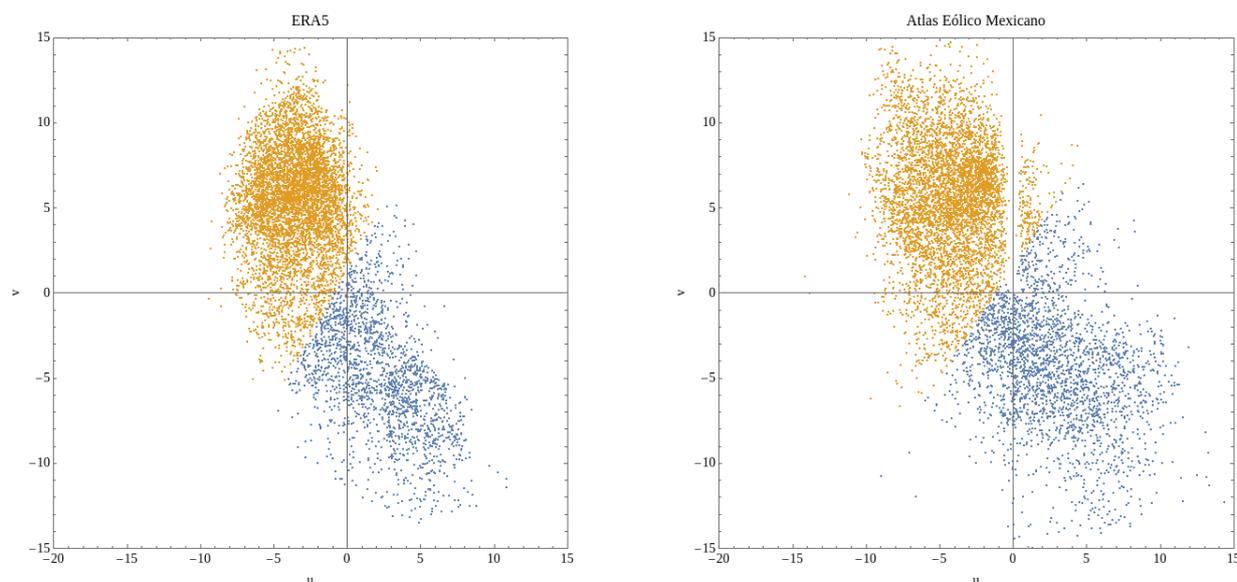


Figura 3.24: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de *k-means* de la distribución de datos de viento en San Fernando, Tamaulipas 2018 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

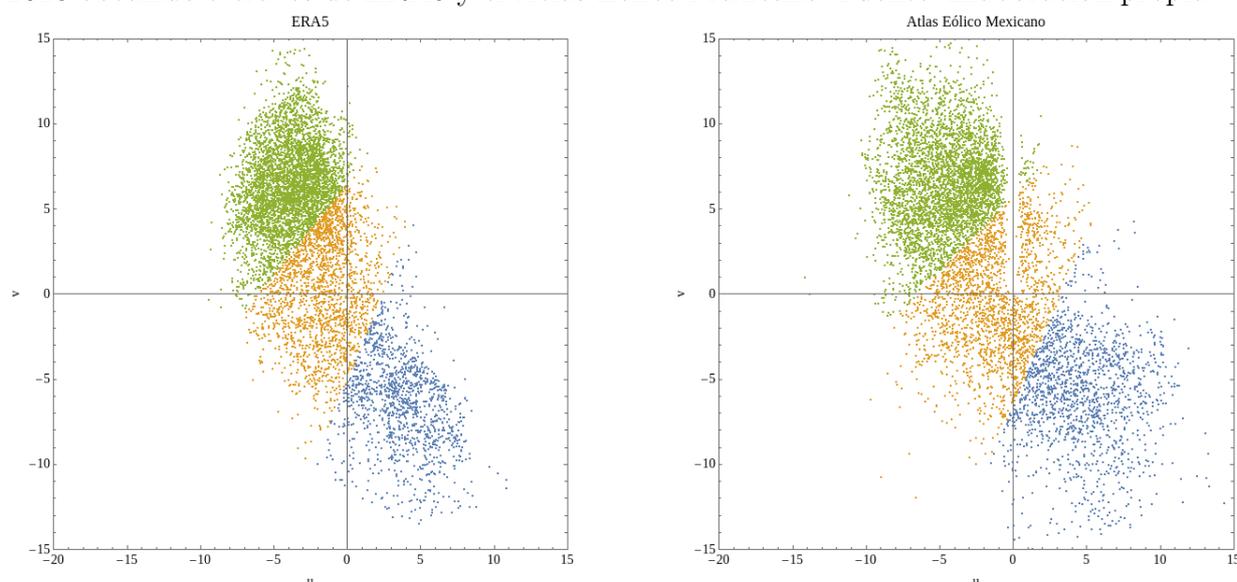


Figura 3.25: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de *k-means* de la distribución de datos de viento en San Fernando, Tamaulipas 2018 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

En la Tabla [3.20](#) se muestran los resultados del cálculo de los porcentajes de permanencia anual, tiempo de vida total y medio para la división en 2 estados. Cabe mencionar que los porcentajes de permanencia anual se calcularon con respecto a la cantidad de datos disponibles en la serie de tiempo en lugar de los datos totales que debería tener el año. ERA5 identifica adecuadamente qué estado es más denso que el otro. El error recae en que ERA5 subestima el tiempo de permanencia y el tiempo de vida total en el estado

1 y en el estado 2 lo sobre estima. Además sobreestima los tiempos de vida medio por más del doble en ambos casos.

Tabla 3.20: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en dos) correspondiente para ambas series de tiempo en San Fernando, Tamaulipas 2018. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	33.4 %	25.5 %	66.6 %	74.5 %
Tiempo de Vida Total (días)	107.83	82.13	215.21	241.08
Tiempo de Vida Medio (días)	0.47	1.02	0.91	2.95

En la Tabla [3.21](#) se muestran los resultados del análisis para la división en 3 *clusters*. En este caso los resultados son más precisos. ERA5 sigue identificando adecuadamente en qué zonas, o estados, se encuentran mayores y menores densidades. El *cluster* más denso es el número 3, en verde. ERA5 sobreestima la incidencia un poco, pero sigue siendo cercana. El estado más similar es el número 2. ERA5 tiene casi los mismo resultados para el porcentaje de permanencia anual y en el tiempo de vida total. Además el tiempo de vida medio es relativamente similar también; en realidad, éste es el más certero de todos los estados tanto para la división en 2 como en 3 *clusters*. ERA5 subestima los valores del estado 1 pero el error es aceptable.

Tabla 3.21: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en tres) correspondiente para ambas series de tiempo en San Fernando, Tamaulipas 2018. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2	3	3
Porcentaje de permanencia anual	20.07 %	16.52 %	26.43 %	26.55 %	53.5 %	56.92 %
Tiempo de Vida Total (días)	62.29	53.63	85.92	85.04	174.46	184.21
Tiempo de Vida Medio (días)	0.39	0.89	0.3	0.43	0.78	1.22

A continuación se muestran los resultados del cálculo de probabilidades condicionales de coincidencia y no coincidencia en la Tabla [3.22](#). Si bien los resultados de permanencia fueron más certeros para la división en 3 estados, la probabilidad condicional de coincidencia en el tiempo es más certera para la división en 2 estados. También el porcentaje de error es menor en este caso. Para la división en 2 estados, el *cluster* número 2, tiene

una coincidencia altísima del 96.25%. Para el caso de división en 3, el *cluster* que más certero fue en cuanto permanencia y vida total y media fue el 2, sin embargo este no es el de mejor coincidencia con respecto al tiempo. En este caso el más certero fue el estado 3.

Tabla 3.22: Probabilidades de coincidencia y no coincidencia de *clusters* para el sitio en San Fernando, Tamaulipas 2018. Fuente: Elaboración propia.

AEM	Probabilidades			
	1	2	3	No
Coincidencias o no 2 <i>clusters</i>	69.01 %	96.36 %	-	12.77 %
Coincidencias o no 3 <i>clusters</i>	70.5 %	61.49 %	88.89 %	22.04 %

En la Figura 3.26 se muestra de forma visual para cada paso de tiempo disponible del sitio, si las dos series de datos coincidieron en un estado de tiempo y en cuál. En el eje  $y$ , 0.5 representa que no hubo coincidencia en algún estado de viento para ese paso de tiempo.

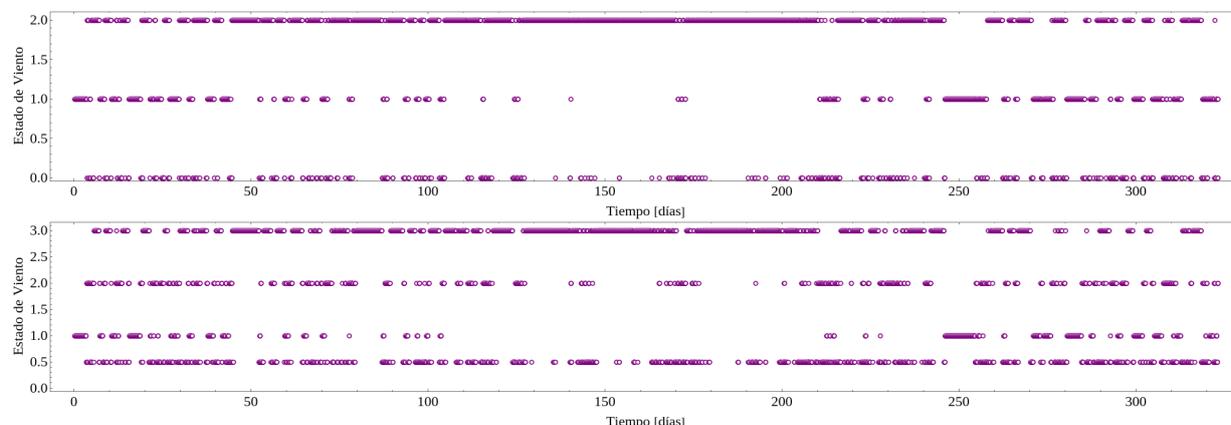


Figura 3.26: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en San Fernando, Tamaulipas 2018. Fuente: Elaboración propia.

### Año 2019

Para el año 2019 ocurre un comportamiento muy similar al del año 2018. Visualmente ambas series de datos son muy parecidas. Éstas se pueden observar en las figuras 3.27 y 3.28 para la división en 2 y 3 *clusters* respectivamente. Para ambas divisiones la corres-

pondencia de datos es directa donde el *cluster* 1 coincide con el 1 de la segunda serie de datos, el 2 con el 2 y así sucesivamente.

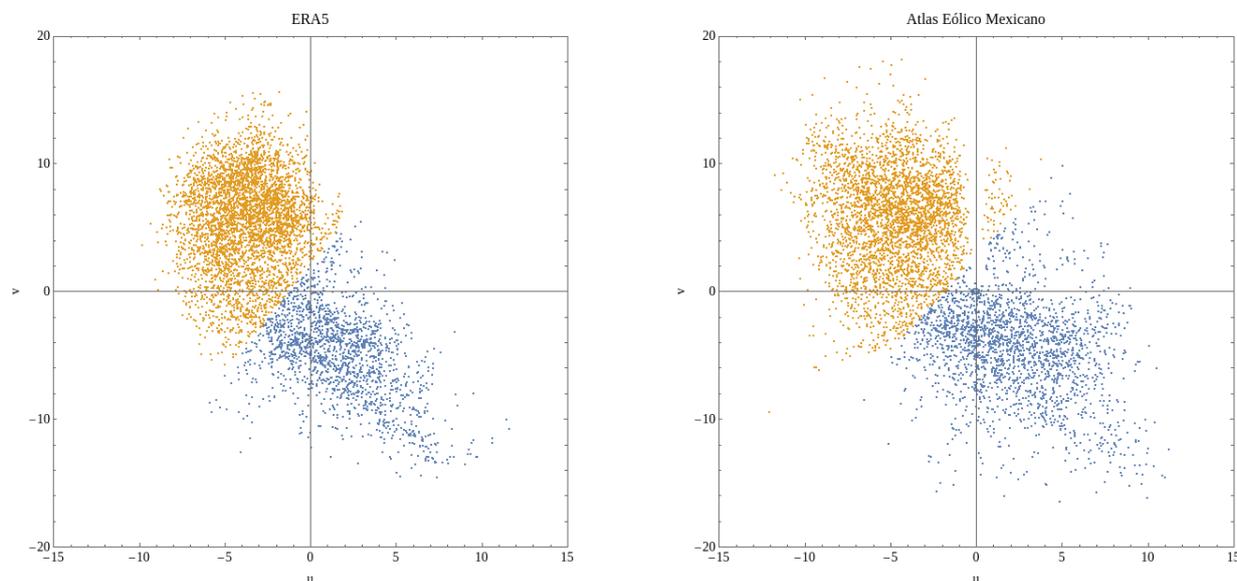


Figura 3.27: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de *k-means* de la distribución de datos de viento en San Fernando, Tamaulipas 2019 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

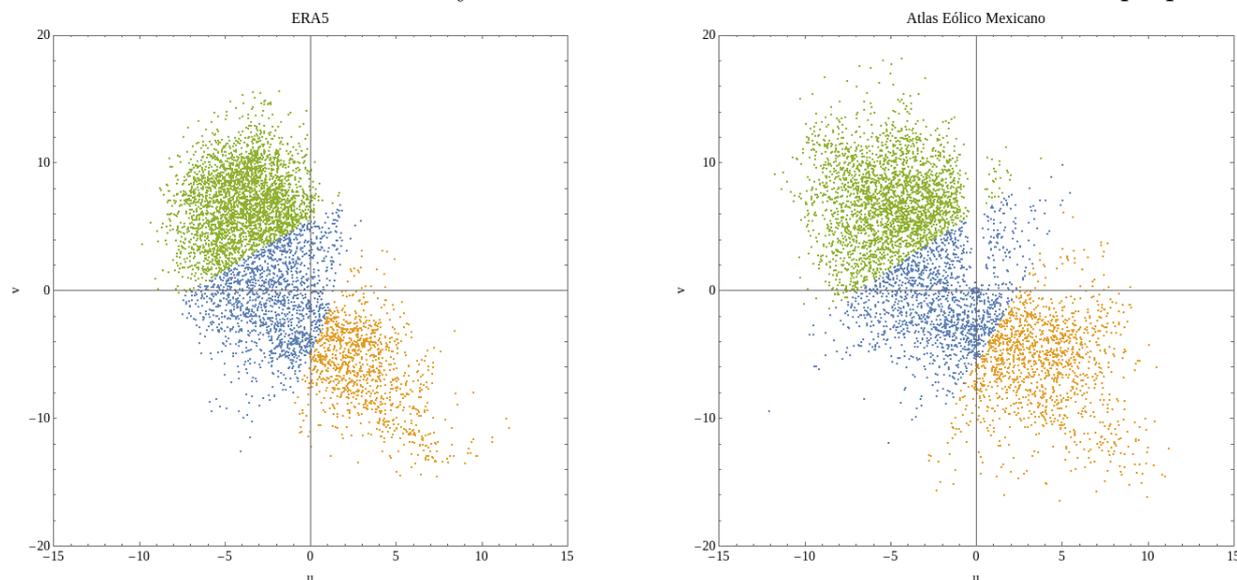


Figura 3.28: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de *k-means* de la distribución de datos de viento en San Fernando, Tamaulipas 2019 obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

En la Tabla [3.23](#) se muestran los resultados del cálculo de los porcentajes de permanencia anual, tiempo de vida total y medio para la división en 2 estados. Cabe mencionar que los porcentajes de permanencia anual de este año también se calcularon con respecto

a la cantidad de datos disponibles en la serie de datos en lugar de los datos totales que debería tener el año. En este año ERA5 también identifica correctamente qué estados tienen mayor y menor densidad, sin embargo 2019 tiene mayor error que 2018 en cuanto al cálculo del porcentaje de permanencia anual. Sin embargo, ERA5 se mantiene consistente en el análisis, subestimando los valores para el estado 1 y sobre estimándolos para el estado 2. Además, también sobre estima los valores del tiempo de vida medio para este año.

Tabla 3.23: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en dos) correspondiente para ambas series de tiempo en San Fernando, Tamaulipas 2019. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	40.14 %	30.09 %	59.86 %	69.9 %
Tiempo de Vida Total (días)	97.54	73.29	145.79	169.75
Tiempo de Vida Medio (días)	0.60	0.92	0.89	2.18

En la Tabla [3.24](#) se muestran los resultados para la división en 3 estados. Así como en 2018, los valores de ambas series de datos son más cercanos para la división en 3 estados. El estado con los resultados más precisos es el número 1 en todas las métricas. El más denso de los estados en este año sigue siendo el verde, lo cual tiene sentido al ser el mismo sitio.

Tabla 3.24: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en San Fernando, Tamaulipas 2019. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2	3	3
Porcentaje de permanencia anual	28.38 %	26.43 %	24.0 %	19.25 %	47.61 %	54.32 %
Tiempo de Vida Total (días)	68.58	62.79	58.46	46.88	115.96	132.29
Tiempo de Vida Medio (días)	0.23	0.38	0.36	0.68	0.66	1.31

En la Tabla [3.25](#) se muestran los resultados del cálculo de la probabilidad condicional de coincidencia para cada *cluster*. De nuevo el estado 2 de la división en 2 tienen una incidencia altísima. Y también la probabilidad del estado 3 de la división en 3 es muy alta, más que en el caso de 2018. Todos los resultados son bastante prometedores, y para este sitio ERA5 hace un buen trabajo en aproximar el comportamiento del sitio. El hecho

de que estos resultados de probabilidad sean buenos significa que también ERA5 logró predecir cuándo en la línea de tiempo la coordenada de viento va a cambiar entre estados de viento.

Tabla 3.25: Probabilidades de coincidencia y no coincidencia de *clusters* para el sitio en San Fernando, Tamaulipas 2019. Fuente: Elaboración propia.

AEM	Probabilidades			
	1	2	3	No
Coincidencias o no 2 <i>clusters</i>	73.57 %	99.05 %	-	11.17 %
Coincidencias o no 3 <i>clusters</i>	65.52 %	71.34 %	96.08 %	18.53 %

En la Figura [3.29](#) se muestra de forma visual para cada paso de tiempo disponible del sitio, si las dos series de datos coincidieron en un estado de tiempo y en cuál. En el eje *y*, 0.5 representa que no hubo coincidencia en algún estado de viento para ese paso de tiempo.

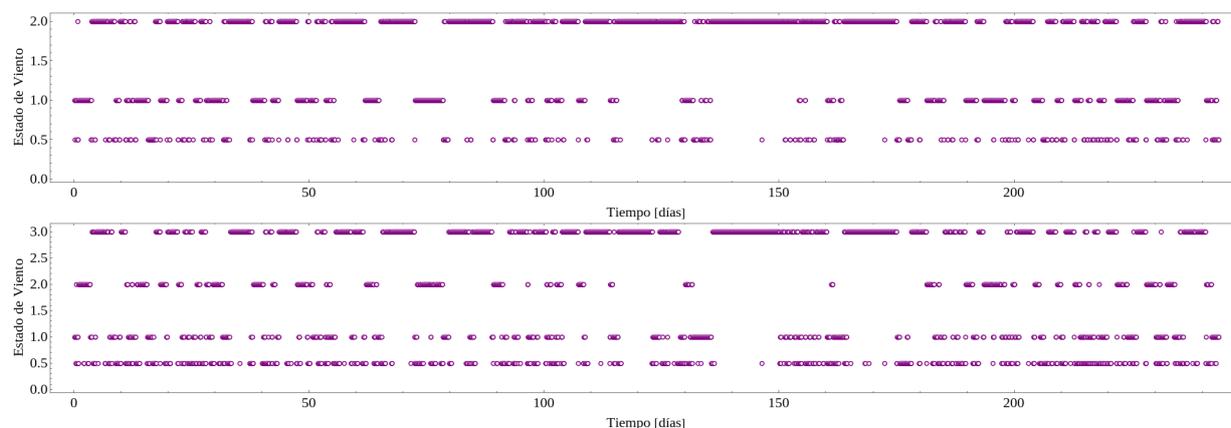


Figura 3.29: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en San Fernando, Tamaulipas 2019. Fuente: Elaboración propia.

### 3.1.7. Tepexi, Puebla

Tepexi se encuentra en el estado de Puebla, México. Las coordenadas de la estación meteorológica (y utilizadas para obtener la información de ERA5) son  $18^{\circ} 35' 24''$  N y  $97^{\circ} 56' 24''$  O; y la torre se encuentra a una altura de 1936 metros sobre el nivel del mar. Se obtuvieron datos del Atlas Eólico Mexicano y ERA5 del año 2018.

Los datos del AEM se convirtieron de coordenadas polares a cartesianas. Por lo tanto la conversión se realizó como:

$$\vec{v} = -v(\text{sen}\theta, \text{cos}\theta) \quad (3.7)$$

Este sitio tienen una representación muy deficiente. La serie de datos de ERA5 tiene poca a nada de similitud con respecto a la serie de datos real. Los *clusters* no coinciden y compararlos no será muy representativo, ya que ERA5 no está describiendo el fenómeno real. Como se observa en la Figura [3.30](#) los *clusters* de datos no coinciden entre sí, por lo que no vale la pena compararlos. Aún, así, la Tabla [3.26](#) muestra los resultados para cada *cluster* en el caso de la división en dos estados.

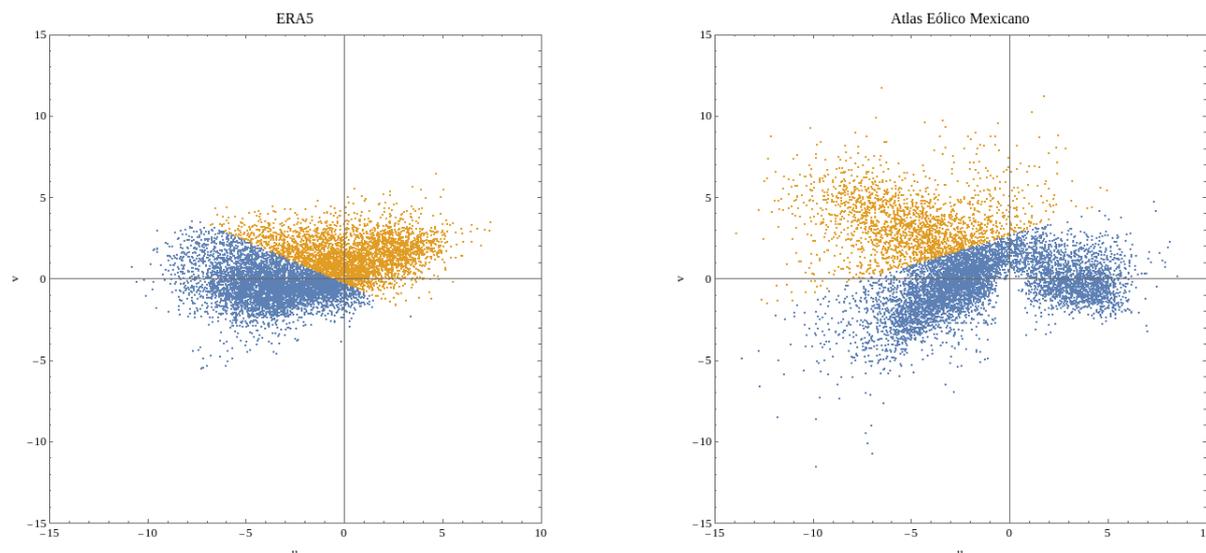


Figura 3.30: *Clusters* o estados obtenidos, divididos en dos subconjuntos, a través del método de *k-means* de la distribución de datos de viento en Tepexi, Puebla obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

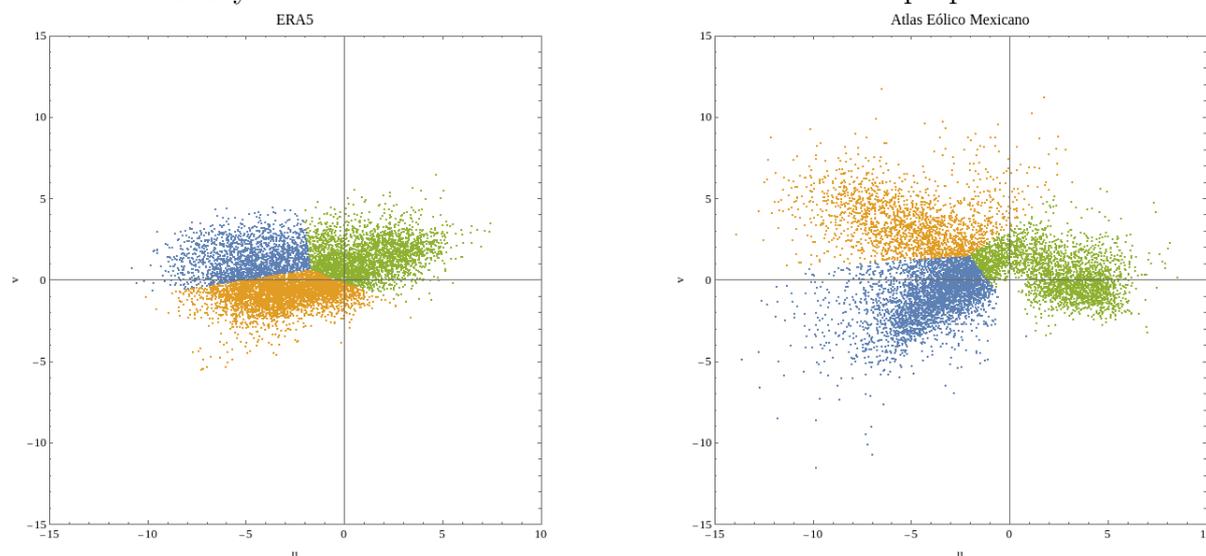


Figura 3.31: *Clusters* o estados obtenidos, divididos en tres subconjuntos, a través del método de *k-means* de la distribución de datos de viento en Tepexi, Puebla obtenido a través de ERA5 y el Atlas Eólico Mexicano. Fuente: Elaboración propia.

Tabla 3.26: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada *cluster* (división en dos) correspondiente para ambas series de tiempo en Tepexi, Puebla. Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	1	2	2
Porcentaje de permanencia anual	75.83 %	59.86 %	24.17 %	40.13 %
Tiempo de Vida Total (días)	274.71	218.375	88.21	146.5
Tiempo de Vida Medio (días)	0.47	0.56	0.15	0.38

La división en 3 estados se puede observar en la Figura 3.31 y cobra algo más de sentido. Si bien la distribución del viento no es tan concordante, los estados logran estimar en dos casos los cuadrantes, es decir, la dirección del viento. Tomando como referencia el AEM, los estados 1 y 2 corresponden a los estados 2 y 1 de ERA5. En este caso ERA5 estima correctamente la dirección. El estado 3 de ERA5 tiene más margen de error en cuanto a la distribución.

En la Tabla 3.27 se pueden observar los resultados de la comparación entre los *clusters* de ERA5 y el AEM. Las métricas son muy certeras, al punto que para el caso del estado 2 con respecto al AEM a dos cifras decimales los resultado del porcentaje de vida anual y el tiempo de vida total anual son iguales. Además el estado 1 y 3 también tienen muy buenos resultados

Tabla 3.27: Comparación de resultados entre el AEM y ERA5 para la evaluación de permanencia en cada cluster (división en tres) correspondiente para ambas series de tiempo en Tepexi, Puebla . Fuente: Elaboración Propia

	AEM	ERA5	AEM	ERA5	AEM	ERA5
<i>Cluster</i>	1	2	2	1	3	3
Porcentaje de permanencia anual	43.55 %	44.28 %	22.79 %	22.79 %	33.66 %	32.93 %
Tiempo de Vida Total (días)	158.88	161.5	83.17	83.17	122.88	120.21
Tiempo de Vida Medio (días)	0.21	0.32	0.14	0.23	0.24	0.36

La Tabla 3.28 muestra la probabilidad condicional de que a lo largo del año ERA5 concuerde en el mismo estado en el que se encuentra el AEM para dicho paso de tiempo. Como se discutió previamente para el caso de la división en 2 estados no hay una división clara, pero para la división en 3 sí y podemos observar que si bien en la Tabla 3.27 el *cluster 2* fue el que tuvo mejores resultados comparativos, en este caso es el *cluster 1* el que coincide mejor si se analiza por paso de tiempo a lo largo del año. El *cluster 3* tiene un valor ligeramente menor pero sigue siendo alto como el ya mencionado y en la comparativa de tiempos de vida y permanencia anual también tiene muy buenos resultados.

En la Figura 3.32 se puede observar la Tabla 3.28 de forma visual por paso de tiempo.

Tabla 3.28: Probabilidades de coincidencia y no coincidencia de *clusters* para el sitio en Tepexi, Puebla. Fuente: Elaboración propia.

AEM	Probabilidades			
	1	2	3	No
Coincidencias o no 2 <i>clusters</i>	57.16 %	31.64 %	-	49.0 %
Coincidencias o no 3 <i>clusters</i>	71.98 %	58.57 %	71.52 %	31.23 %

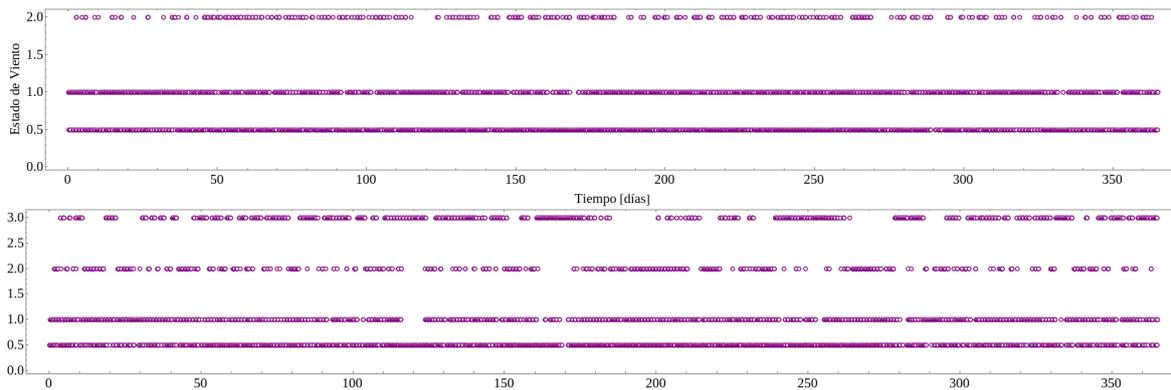


Figura 3.32: Coincidencia y no coincidencia entre dos y tres estados a lo largo de un año en Tepexi, Puebla. Fuente: Elaboración propia.

# Capítulo 4

## Discusión y Conclusiones

Con el fin de analizar todos los resultados del capítulo anterior en conjunto, se calculó el porcentaje de error con respecto al AEM utilizando la fórmula:

$$\% \text{ de error} = \left| \frac{\text{Resultado ERA5} - \text{Resultado AEM}}{\text{Resultado AEM}} \right| * 100 \quad (4.1)$$

Los resultados se muestran en las tablas [4.1](#), [4.2](#), [4.3](#). Las tablas [4.4](#) y [4.5](#) muestran los resultados de la probabilidad condicional de coincidencia y no coincidencia por paso de tiempo en de todos los sitios.

En las tablas [4.1](#) y [4.2](#) de permanencia anual y tiempo de vida total anual respectivamente se observan buenos resultados para la mayoría de los sitios. En verde se resaltan aquellos resultados con un error menor al 10%. Teniendo en cuenta que se están analizando 8760 datos en cada año, ERA5 es bastante bueno dando una aproximación o predicción del comportamiento anual del sitio. Solamente un sitio no cumple con tener un error menor al 10% en al menos uno de sus *clusters*: Ojuelos, Jalisco; sin embargo sigue en el rango de 10% - 30%. En cuanto a este par de métricas, la mayoría de los sitios presentan mejores resultados en su división en 3 *clusters*. Incluso en sitios donde pareciera haber solamente dos cúmulos de datos distinguibles o una corriente principal como la Rumorosa, CERTE y San Fernando, la división en 3 resulta más favorable. Esto posiblemente se deba a que se acota la información en el *cluster* de mayor tamaño al sub-

dividirlo y analizarlo a más detalle. Quizás de esta forma ERA5 logra hacer una mejor distinción de dónde y cuánto tiempo en el año el viento pasa en una zona o racimo.

En el caso de Mérida, Cd. Cuahutémoc y Ojuelos la división en 2 *clusters* es más favorable. Cd. Cuahutémoc tiene una forma relativamente circular con poco ruido alrededor. Sin embargo Mérida y Ojuelos no tienen una distribución concentrada al centro similar. Aún así la división en 3 *clusters* para estos sitios sigue siendo relativamente precisa si se excluye el racimo menos denso, el racimo 3.

Tabla 4.1: Porcentaje de error de ERA5 respecto al AEM de la permanencia anual. Fuente: Elaboración propia.

Sitio	División en 2		División en 3		
	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
La Rumorosa, Baja California	34.24 %	16.84 %	7.39 %	1.32 %	6.49 %
Mérida, Yucatán	3.60 %	2.21 %	8.49 %	21.7 %	159.58 %
Cd. Cuahutémoc, Chihuahua	3.40 %	6.04 %	9.41 %	15.78 %	59.98 %
CERTE, Oaxaca 2018	12.92 %	30.66 %	12.02 %	31.96 %	2.15 %
CERTE, Oaxaca 2019	7.27 %	16.33 %	12.76 %	7.44 %	1.35 %
Ojuelos, Jalisco	14.53 %	12.56 %	25.69 %	25.50 %	
San Fernando, Tamaulipas 2018	23.65 %	11.86 %	17.69 %	0.45 %	6.39 %
San Fernando, Tamaulipas 2019	25.04 %	16.77 %	6.87 %	19.79 %	14.09 %
Tepexi, Puebla	21.06 %	66.03 %	1.68 %	0.00 %	2.17 %

Tabla 4.2: Porcentaje de error del Tiempo de vida total anual. Fuente: Elaboración propia.

Sitio	División en 2		División en 3		
	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
La Rumorosa, Baja California	34.22 %	16.84 %	7.25 %	1.31 %	6.51 %
Mérida, Yucatán	3.60 %	2.72 %	8.44 %	21.26 %	159.33 %
Cd. Cuahutémoc, Chihuahua	3.40 %	5.99 %	9.43 %	15.71 %	59.11 %
CERTE, Oaxaca 2018	11.95 %	30.69 %	12.03 %	31.84 %	2.13 %
CERTE, Oaxaca 2019	7.31 %	16.31 %	12.88 %	7.44 %	1.37 %
Ojuelos, Jalisco	14.59 %	12.55 %	25.70 %	25.50 %	
San Fernando, Tamaulipas 2018	23.83 %	12.02 %	13.90 %	1.02 %	5.59 %
San Fernando, Tamaulipas 2019	24.86 %	16.43 %	8.44 %	19.81 %	14.08 %
Tepexi, Puebla	20.51 %	66.08 %	1.65 %	0.00 %	2.17 %

Tomando en cuenta los resultados y la discusión anterior, se puede concluir que al realizar un análisis de un sitio con ERA5 es preferible confiar más en la división en 3 *clusters* y tratar con más cautela o desconfianza el racimo que represente la menor densidad de

datos. Asimismo, se puede esperar en la división en 3 racimos tener porcentajes de error menores al 30 % en la mayoría de los casos, es decir, una precisión del 70 %. Sin embargo, si el racimo menos denso, es decir, el estado menos frecuente, es considerablemente menos denso que los demás, como es el caso de Mérida, donde prácticamente representa ruido y una dispersión no muy clara, se puede esperar que el error ascienda dramáticamente por lo que no será recomendable confiar o tomar como referencia dicho *cluster*. Igualmente hay una posibilidad de que el *cluster* de menor densidad sea poco preciso, aunque se requeriría el análisis de más sitios con un comportamiento similar, con distribuciones de viento concentradas al centro, es decir, sin una dirección preferente clara y con una forma más circular como es el caso de Cd. Cuahutémoc para poder confirmar que este tipo de distribuciones tienden a ser menos certeras en ERA5.

En la Tabla 4.3 se presentan los errores en el cálculo del tiempo de vida medio para cada *cluster* de ambas series de tiempo. En este caso ERA5 es muy deficiente en el cálculo. Esto implica que la predicción de la dinámica del viento de ERA5 es poco precisa. En verde se muestran los sitios y los *clusters* donde la aproximación fue más certera, sin embargo en todos los demás casos la aproximación no es buena.

Tabla 4.3: Porcentaje de error con respecto al AEM de ERA5 del Tiempo de vida medio. Fuente: Elaboración propia.

Sitio	División en 2		División en 3		
	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
La Rumorosa, Baja California	72.09 %	6.82 %	8.93 %	35.14 %	114.81 %
Mérida, Yucatán	53.33 %	61.22 %	27.59 %	17.78 %	76.19 %
Cd. Cuahutémoc, Chihuahua	171.43 %	191.67 %	122.73 %	182.35 %	188.89 %
CERTE, Oaxaca 2018	9.24 %	60.78 %	58.06 %	233.33 %	67.35 %
CERTE, Oaxaca 2019	21.90 %	53.19 %	172.22 %	68.97 %	80.43 %
Ojuelos, Jalisco	112.82 %	64.71 %	9.30 %	127.78 %	
San Fernando, Tamaulipas 2018	117.02 %	224.18 %	128.21 %	43.33 %	56.41 %
San Fernando, Tamaulipas 2019	53.33 %	144.94 %	65.22 %	88.89 %	98.48 %
Tepexi, Puebla	19.15 %	153.33 %	52.38 %	64.29 %	50.00 %

En la Tabla 4.4 se muestra la probabilidad de que, dado que una coordenada de viento en el AEM se encuentra en un *cluster* en un paso de tiempo específico, ERA5 también se encuentre en dicho *cluster* para dicho paso de tiempo. Cabe recalcar que esta es una probabilidad condicional. En la Tabla 4.5 se muestra el porcentaje total anual en el cual

el AEM y ERA5 no coincidieron por paso de tiempo en el mismo *cluster*.

Los valores mostrados en la Tabla 4.4 son muy favorables en la mayoría de los casos. Verde oscuro representa valores arriba del 90 % de coincidencia, verde claro son valores arriba del 80 %, amarillo fuerte arriba del 70 % y amarillo claro arriba del 60 %. En la mayoría de los sitios y divisiones, la probabilidad de coincidencia es alta; y la división en 2 *clusters* pareciera ser más favorable para esta métrica. Los sitios con probabilidades más favorables son CERTE, Tamaulipas y la Rumorosa. En el caso de CERTE, este tiene bien definida una dirección preferente, mientras que San Fernando y la Rumorosa tienen dos direcciones bien definidas y prácticamente contrarias. Se podría decir que una condición o aspecto favorable para que esta métrica sea más certera es tener una o dos direcciones de viento bien definidas.

Mérida, a pesar de haber tenido valores poco favorables en el caso de permanencia total anual y tiempos de vida total y medio, en este caso tiene muy buenos resultados, por lo que es posible que una dirección bien definida no sea el único de los posibles indicios de que ERA5 será certero en esta métrica. Los sitios con peores resultados son Ojuelos y Tepexi, aunque para Tepexi la división en 2 no es representativa ya que los *clusters* no tuvieron una correspondencia clara, por lo que se puede omitir; para su división en 3 *clusters* los resultados mejoran, sin embargo, dado que Tepexi no es muy bueno modelando la distribución y su división en 2 no se pudo utilizar por la concordancia casi nula, los resultados se pueden considerar muy pobres. A pesar de que Ojuelos no obtuvo resultados tan favorables, al menos uno de sus *clusters* en ambas divisiones tiene valores de probabilidad mayores al 65 %. Para la probabilidad condicional se puede concluir que ERA5 hace un buen trabajo determinando por paso de tiempo en qué estado se encontrará el viento con respecto al AEM, es decir, que en el mismo paso de tiempo las coordenadas de ERA5 y el AEM coincidan en el mismo estado o *cluster*. La división en 2 *clusters* es la que muestra los resultados más favorables para tres sitios: CERTE, San Fernando y la Rumorosa. Es preferible escoger la división en 2 para análisis por paso de tiempo, ya que también se complementa con el bajo error de no coincidencia anual mostrado en la Tabla 4.5.

Tabla 4.4: Probabilidad condicional de coincidencia por paso de tiempo en todos los sitios analizados. Fuente: Elaboración propia.

Sitio	División en 2		División en 3		
	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
La Rumorosa, Baja California	94.94 %	80.76 %	88.74 %	76.27 %	67.56 %
Mérida, Yucatán	78.71 %	89.15 %	73.24 %	74.89 %	90.8 %
Cd. Cuauhtémoc, Chihuahua	72.50 %	57.1 %	67.33 %	62.4 %	87.53 %
CERTE, Oaxaca 2018	85.17 %	95.49 %	71.22 %	72.07 %	88.23 %
CERTE, Oaxaca 2019	89.04 %	91.7 %	72.88 %	72.38 %	87.27 %
Ojuelos, Jalisco	65.03 %	43.76 %	36.78 %	67.78 %	
San Fernando, Tamaulipas 2018	69.01 %	96.36 %	70.5 %	61.49 %	88.89 %
San Fernando, Tamaulipas 2019	73.57 %	99.05 %	65.52 %	71.34 %	96.08 %
Tepexi, Puebla	57.16 %	31.64 %	71.98 %	58.57 %	71.52 %

Si se toma en consideración la Tabla [4.5](#) entonces la división en 2 *clusters* es la que presenta menores porcentajes de no coincidencia por paso de tiempo, sin embargo la división en 3 tampoco tiene errores demasiado grandes. Se puede decir que tanto la división en 2 *clusters* como en 3 *clusters* resulta favorable y arroja buenas predicciones de parte de ERA5, pero en general la división en 2 es más precisa.

Tabla 4.5: Porcentaje de no coincidencia por paso de tiempo a lo largo del año entre ERA5 y el AEM. Fuente: Elaboración propia.

Sitio	División en 2	División en 3
La Rumorosa, Baja California	14.57 %	22.16 %
Mérida, Yucatán	14.82 %	24.69 %
Cd. Cuauhtémoc, Chihuahua	33 %	31.10 %
CERTE, Oaxaca 2018	11.77 %	23.69 %
CERTE, Oaxaca 2019	10.13 %	23.42 %
Ojuelos, Jalisco	44.92 %	46.53 %
San Fernando, Tamaulipas 2018	12.77 %	22.04 %
San Fernando, Tamaulipas 2019	11.17 %	18.53 %
Tepexi, Puebla	49 %	31.23 %

CERTE y Tamaulipas son los sitios que mejores resultados tuvieron considerando todas las métricas. El porcentaje de permanencia anual y el tiempo de vida total presentan buenos resultados para casi todos los *clusters*, además ningún error es demasiado alto como sí sucede en algunos racimos de otros sitios. En cuanto al tiempo de vida medio, los resultados son muy pobres, pero esto no es algo característico de estos sitios, sino que ERA5 no tienen mucha precisión en el cálculo de esta métrica. Por otro lado, la

probabilidad condicional de coincidencia sí es muy favorable. Ninguno de los *clusters* presenta, para ambos sitios, una precisión menor al 60 %, además que para ambos casos la división en 2 *clusters* es más precisa; solo San Fernando en 2018 tienen un valor menor al 70 % en su racimo 1, pero para el resto de los *clusters* en la división en 2, los resultados oscilan entre el 73.57 % y 99.05 %. Además sus porcentajes de no coincidencia por paso de tiempo anual total son también muy bajos para este par de sitios, sobre todo para la división en 2. Como ya se ha discutido, CERTE tiene un corriente principal de viento hacia el tercer cuadrante y una corriente menor hacia el segundo cuadrante. Por otro lado San Fernando tiene dos corrientes de viento similares entre sí en direcciones contrarias, de nuevo hacia el segundo y tercer cuadrante. Si bien la Rumorosa no tiene resultados en todas sus métricas tan favorables como CERTE y San Fernando aún así presenta buenos resultados en general. Este además también tiene dos corrientes definidas en direcciones contrarias como San Fernando. Por lo tanto, un indicador de que ERA5 será más preciso podría ser que se aprecien corrientes de viento relativamente bien definidas, con poco ruido hacia otros cuadrantes que el de la misma corriente y además que su orografía no sea tan accidentada, ya que esto pudiese afectar al modelo debido a vientos locales.

Tepexi, Puebla es uno de los sitios que tuvo resultados poco favorables. Su división en dos se tiene que descartar completamente por su concordancia nula y aunque su división en tres se desempeña bien tanto para los tiempos de permanencia anual y probabilidad condicional, la distribución de viento y las direcciones no concuerdan del todo. Es decir, ERA5 falla en definir la distribución de viento, por lo que al intentar definir los estados de viento, estos no coinciden. Tepexi tiene magnitudes de viento pequeñas, o al menos no es claramente un sitio productor con direcciones definidas y rapidez de viento considerables como CERTE, San Fernando y la Rumorosa. Los sitios que no se desempeñan tan bien coinciden con no ser sitios con alto potencial de producción, por lo que al tener más aleatoriedad ERA5 podría tener más error.

En general ERA5 demostró ser bastante bueno en el modelado de varios sitios, pero no se tiene unas características exactas o precisas de condiciones a cumplirse para saber que ERA5 funcionará con propiedades estadísticas parecidas. Si se desea observar el comportamiento anual del viento, es más conveniente inclinarse por la división en 3 *clusters*, ya que la probabilidad de tener racimos con buena precisión es mayor. Por otro lado, si

se desea analizar el comportamiento por paso de tiempo, la división en 2 *clusters* es la que tiene mejores probabilidades de arrojar resultados más certeros. Asimismo existe una posibilidad de que una alta aleatoriedad en sitios no productores juegue un papel importante en la precisión de ERA5 en su contra, pero esta teoría tendría que ser analizada con más sitios con estas condiciones para poder comprobarla.

Si se desea realizar un análisis de un sitio donde no hay datos reales de viento y utilizar un método de *clustering*, no es recomendable utilizar los datos como base para análisis de tiempo de vida medio, ya que lo más probable es que sean erróneos y no se acerquen a la realidad. El tiempo de vida total sí es una métrica más confiable y se pueden esperar valores con errores por debajo del 25-30 % en general al utilizar la división en 3 *clusters*. Sin embargo, graficar los datos puede también facilitar ver si ERA5 será más o menos preciso en el cálculo. Si los datos de viento en general van bien direccionados y el sitio tienen indicios de ser un buen productor de energía eólica, es decir, cuenta además con magnitudes de viento considerables, además de estar bien direccionadas, es más probable que los datos de ERA5 sean mejores. En cuanto a la probabilidad condicional de coincidencia por paso de tiempo, las condiciones anteriores de magnitud y datos direccionados también son buen indicio de resultados favorables por parte de ERA5, pero en este caso convendría más utilizar la división de 2 *clusters* para obtener resultados en promedio del 78.75 %.

Asimismo, ERA5 también puede ser una buena herramienta para encontrar problemas en los datos, ya que en San Fernando fue necesario rotar los datos del AEM para que coincidieran con los datos teóricos.

Para un trabajo a futuro, se podrían analizar sitios con bajo potencial de producción o alta aleatoriedad, para ver su efecto en el modelado de ERA5. Además, se podría experimentar con más divisiones para ver si la precisión en las métricas aumenta o disminuye o incluso si hay una división óptima donde los resultados son más favorables. Por último, analizar más sitios con alto potencial eólico como San Fernando, CERTE y la Rumorosa, para confirmar la teoría de que un sitio productor, con magnitudes de viento altas y datos bien direccionados, efectivamente tendrá mejores resultados por parte de ERA5.

# Capítulo 5

## Referencias

## Bibliografía

- [1] Arredondo Gámez, M., García Nava, H., Hernández, R., Lámbert-Arista, A., et al. (2022). Estimación del recurso eólico marino en la península de baja california, méxico, utilizando la base de datos era5.
- [2] Dullaart, J. C., Muis, S., Bloemendaal, N., and Aerts, J. C. (2020). Advancing global storm surge modelling using the new era5 climate reanalysis. *Climate Dynamics*, 54:1007–1021.
- [3] ECMWF (2021a). Era5: uncertainty estimation. <https://confluence.ecmwf.int/display/CKB/ERA5%3A+uncertainty+estimation>.  
Accedido por última vez: 21-06-2021.
- [4] ECMWF (2021b). Era5:data documentation. <https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#app-switcher>.  
Accedido por última vez: 20-06-2021.

- [5] ECMWF (s.f.). Forecast charts and data. <https://www.ecmwf.int/en/forecasts>. Accedido por última vez: 20-06-2021.
- [6] Gobierno de México (s.f.). ¿qué es una estación meteorológica automática? [https://smn.conagua.gob.mx/es/?option=com\\_content&view=article&id=38&Itemid=46](https://smn.conagua.gob.mx/es/?option=com_content&view=article&id=38&Itemid=46). Accedido por última vez: 19-06-2021.
- [7] Hersbach, H. (2017). Era5: the new reanalysis of weather and climate data. <https://www.ecmwf.int/en/about/media-centre/science-blog/2017/era5-new-reanalysis-weather-and-climate-data>. Accedido por última vez: 21-06-2021.
- [8] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.
- [9] Hu, W., Scholz, Y., Yeligeti, M., von Bremen, L., and Deng, Y. (2023). Downscaling era5 wind speed data: a machine learning approach considering topographic influences. *Environmental Research Letters*, 18(9):094007.
- [10] Jourdier, B. (2020). Evaluation of era5, merra-2, cosmo-rea6, newa and arome to simulate wind power production over france. *Advances in Science and Research*, 17:63–77.
- [11] Miranda, U. and Saldaña, R. (2019). El atlas eólico mexicano, avances. <https://transicionenergetica.ineel.mx/Revista.mvc/CD2n2v1>. Accedido por última vez:15-06-2021.
- [12] Olauson, J. (2018). Era5: The new champion of wind power modelling? *Renewable energy*, 126:322–331.
- [13] Ruiz, S. A. G., Barriga, J. E. C., and Martínez, J. A. (2021). Wind power assessment in the caribbean region of colombia, using ten-minute wind observations and era5 data. *Renewable Energy*, 172:158–176.
- [14] Sánchez-Pérez, P. A., Robles, M., and Jaramillo, O. A. (2016). Real time Markov chains: Wind states in anemometric data. *Journal of Renewable and Sustainable Energy*, 8(2).

- [15] Secretaría de Energía (2018). Se pone en marcha el atlas eólico mexicano. <https://www.gob.mx/sener/prensa/se-pone-en-marcha-el-atlas-eolico-mexicano>. Consultado por última vez:19-06-2021.
- [16] Thomas, S. R., Martínez-Alvarado, O., Drew, D., and Bloomfield, H. (2021a). Drivers of extreme wind events in Mexico for windpower applications. *International Journal of Climatology*, 41:E2321–E2340.
- [17] Thomas, S. R., Nicolau, S., Martínez-Alvarado, O., Drew, D. J., and Bloomfield, H. C. (2021b). How well do atmospheric reanalyses reproduce observed winds in coastal regions of Mexico? *Meteorological Applications*, 28(5):e2023.