

Universidad Autónoma de Querétaro

Facultad de Informática

Maestría en Ciencias de la Computación

Desarrollo de una metodología para predicción del recurso eólico en
Querétaro mediante redes bayesianas

Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Ciencias de la Computación

Presenta

I.C. Daniel Cantón Enríquez

Dirigido por:

Dr. Hugo Jiménez Hernández

Dr. Hugo Jiménez Hernández

Presidente

Dra. Ana Marcela Herrera Navarro

Secretario

Dra. Diana Margarita Córdova Esparza

Vocal

Dra. Ma. Teresa García Ramírez

Suplente

Dra. Sandra Luz Canchola Magdaleno

Suplente

Centro Universitario, Querétaro, Qro.

Agosto 2019

México

DEDICATORIA

Dedico esta tesis a las siguientes personas:

A mis padres Edgar Cantón Girón y Claudia Enríquez Méndez, por el apoyo que siempre me han brindado, por la motivaciones recibidas y las enseñanzas de vida las cuales han formado mi persona.

A mi pareja y compañera Lirio del Carmen Luis Guzmán por su comprensión, cariño y amor para alentarme durante este proyecto de tesis. *Nadxiee lii stale xheela' stine'*.

A mi hermana Claudia Cantón Enríquez con quien siempre he contado, especialmente durante estos dos años intensos.

AGRADECIMIENTOS

A mi director de tesis el Dr. Hugo Jiménez Hernández, por su disposición y enseñanzas recibidas durante estos dos años.

A la doctora Ana Marcela Herrera Navarro, por su apoyo en el ingreso y seguimiento académico durante el posgrado.

A los profesores del programa en Ciencias de la Computación por compartir sus conocimientos y experiencias.

A mis compañeros de generación Aldo Valencia, Vania Hinojosa y Sergio Carrillo; especialmente a mi amigo Adolfo Solís por su amistad y experiencias compartidas.

A los compañeros del Laboratorio de Modelado y Simulación en CIDESI por los conocimientos compartidos y los buenos ratos de convivencia.

A la Universidad Autónoma de Querétaro (UAQ) por el conocimiento y la oportunidad de superarme académicamente.

Por ultimo, quiero agradecer al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo financiero recibido a través de la beca escolar 494040.

RESUMEN

La estimación del recurso eólico ha tenido un gran incremento en los últimos años debido a la fuerte demanda de energías renovables. En lo particular la estimación del viento es volátil debido a su naturaleza estocástica. Por lo tanto, es importante estudiar la velocidad del viento en diferentes sitios para conocer la asociación que existe entre ellos. En este proyecto de tesis, se desarrolló un proceso Bayesiano que analiza las dependencias geográficas en ocho sitios de la ciudad de Querétaro utilizando un método estadístico, el algoritmo de Expectation-Maximization útil para trabajar con datos incompletos. Posteriormente se utiliza el coeficiente de correlación de Spearman para determinar la intensidad en la relación entre cada uno de los sitios. Los resultados de la experimentación se realizaron en tres escenarios, en periodos distintos, para predecir el recurso eólico en los sitios durante cada periodo. De esta manera, se conocen aquellos sitios en donde se puede aprovechar la fuerza del viento para la obtención de energía eólica.

Palabras Clave: análisis de dependencias geográficas, coeficiente de correlación de Spearman, Expectation-Maximization, velocidad del viento.

SUMMARY

The estimation of the wind resource had a large increase in recent years due to the strong demand for renewable energy. Wind estimation is a dynamic problem due to its stochastic nature. Consequently, the study of wind speed in different places helps to know the association that exist between them. In this thesis project, we present a Bayesian process that analyzes the geographic dependencies in eight sites in the city of Queretaro using a statistical method, the Expectation-Maximization algorithm useful to working with incomplete data. The Spearman correlation coefficient is then used to determine the intensity in the relationship between each of the sites. The results of the experimentation were carried out in three scenarios, in different periods, to predict the wind resource in the sites during each period. In this way, the sites where wind power can be used to obtain wind energy are known.

Keywords: Geographical dependency analysis, Spearman's correlation coefficient, Expectation-Maximization, wind speed.

TABLA DE CONTENIDOS

ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABLAS.....	IX
1 INTRODUCCIÓN.....	1
1.1 PLANTEAMIENTO DEL PROBLEMA.....	1
1.2 JUSTIFICACIÓN.....	3
1.3 OBJETIVOS	4
1.3.1 <i>Objetivo general</i>	4
1.3.2 <i>Objetivo particulares</i>	4
1.4 HIPÓTESIS	5
1.5 DELIMITACIONES	5
1.6 CONTRIBUCIONES	5
1.7 ESTRUCTURA DE LA TESIS	6
2 FUNDAMENTOS TEÓRICOS.....	7
2.1 CLASIFICACIÓN DE ESCALAS DE TIEMPO EN LA PREDICCIÓN	7
2.2 MÉTODOS DE PREDICCIÓN DEL POTENCIAL EÓLICO.....	7
2.2.1 <i>Métodos físicos</i>	8
2.2.2 <i>Métodos estadísticos</i>	8
2.2.3 <i>Métodos híbridos</i>	9
2.3 ALGORITMO DE EXPECTATION-MAXIMIZATION	9
2.4 MÉTODOS DE PREDICCIÓN DE ACUERDO A SU FUNCIÓN DE DENSIDAD..	10
2.4.1 <i>Métodos no paramétricos</i>	11
2.4.2 <i>Correlación</i>	11
2.4.3 <i>Coefficiente de correlación de Spearman</i>	12
2.5 COMENTARIOS	12
3 METODOLOGÍA.....	14

3.1	ÁREA DE ESTUDIO Y RECOLECCIÓN DE LOS DATOS.....	14
3.1.1	<i>Descripción de la región de estudio.....</i>	14
3.1.2	<i>Descripción de los sitios.....</i>	15
3.2	PROCESAMIENTO DE DATOS DEL VIENTO.....	16
3.2.1	<i>Filtrado de las mediciones.....</i>	16
3.2.2	<i>Ajuste de los datos.....</i>	17
3.3	ESTIMACIÓN DE LOS PARÁMETROS.....	17
3.4	ANÁLISIS DE CORRELACIÓN.....	18
3.5	COMENTARIOS.....	19
4	RESULTADOS.....	20
4.1	ANÁLISIS DE LOS SITIOS.....	20
4.2	RESULTADOS DE LOS ESCENARIOS.....	21
4.2.1	<i>Escenario Primavera-Verano.....</i>	21
4.2.2	<i>Escenario Otoño-Invierno.....</i>	23
4.2.3	<i>Escenario Anual.....</i>	26
4.3	COMENTARIOS DE LOS ESCENARIOS.....	29
4.4	COMPARACIÓN CON LOS TRABAJOS CONSULTADOS.....	30
5	CONCLUSIONES Y TRABAJO FUTURO.....	33
5.1	CONCLUSIONES.....	33
5.2	TRABAJO FUTURO.....	34
6	REFERENCIAS.....	35

ÍNDICE DE FIGURAS

1.1	Capacidad eólica instalada: (a) a nivel mundial en diciembre de 2017 y (b) en México a finales de 2017.	2
3.1	Esquema de la metodología utilizada.....	14
3.2	Localización geográfica de los municipios de Querétaro (morado) y El Marqués (azul).	15
3.3	Estaciones meteorológicas en Querétaro y El Marqués.....	16
4.1	Series de tiempo de la velocidad del viento en m/s en diciembre 2017..	20
4.2	Variaciones en la velocidad del viento mensualmente.	21
4.3	Ubicación de las dependencias geográficas durante el periodo mayo 2017 - octubre 2017.	22
4.4	Curvas de nivel de las dependencias geográficas durante el periodo mayo 2017 - octubre 2017.	23
4.5	Grafos de dependencias de los sitios durante el periodo mayo 2017 - octubre 2017.....	24
4.6	Ubicación de las dependencias geográficas durante el periodo noviembre 2017 - abril 2018.	25
4.7	Curvas de nivel de las dependencias geográficas durante el periodo noviembre 2017 - abril 2018.....	25
4.8	Grafos de dependencias de los sitios durante el periodo noviembre 2017 - abril 2018.	26
4.9	Ubicación de las dependencias geográficas durante el periodo mayo 2017 - abril 2018.	27
4.10	Curvas de nivel de las dependencias geográficas durante el periodo mayo 2017 - abril 2018.	28
4.11	Grafos de dependencias de los sitios durante el periodo mayo 2017 - abril 2018.....	28

4.12 Curvas de nivel rellenas que representan los valles de Querétaro, Amazcala y Buenavista.....	29
4.13 Histogramas de la velocidad del viento en m/s para el periodo mayo 2017 a abril 2018.	30

ÍNDICE DE TABLAS

2-1	Escala de tiempo en la predicción del recurso eólico.	7
2-2	Valores críticos del coeficiente de correlación de rangos de Spearman.	13
3-1	Información geográfica de los ocho sitios y sus velocidades del viento..	16
4-1	Coeficientes de correlación entre las velocidades del viento de los diferentes sitios para el periodo mayo 2017 a octubre 2017.....	23
4-2	Coeficientes de correlación entre las velocidades del viento de los diferentes sitios para el periodo noviembre 2017 a abril 2018.....	26
4-3	Coeficientes de correlación entre las velocidades del viento de los diferentes sitios para el periodo mayo 2017 a abril 2018.	27
4-4	Claves de referencia de los trabajos consultados.....	31
4-5	Comparación de los trabajos consultados.	31

1. INTRODUCCIÓN

El constante aumento en la demanda de energía eléctrica a nivel mundial ha tenido como efecto el incremento en un 60 % en la generación de energía eléctrica (IEA, 2018). Actualmente se sigue dependiendo en gran medida de los combustibles fósiles como fuente de energía principal (Mikati et al., 2012), teniendo en cuenta el impacto ambiental que estos ocasionan por medio del calentamiento global.

Estas circunstancias han propiciado el rápido crecimiento de las energías renovables [Hernández-Escobedo et al. (2014); IEA (2018)]; entre ellas la energía eólica. De acuerdo con estadísticas del Consejo Mundial de Energía Eólica (Sawin et al., 2018), el total de energía eólica instalada a nivel mundial a finales del 2017 es de 539 GW, lo cual es 5.7 veces más grande que lo instalado en 2007.

Sin embargo, a pesar de que existen diferentes enfoques y métodos para predecir el recurso eólico, resulta difícil encontrar una metodología que analice las dependencias geográficas de la velocidad del viento mediante un procesos Bayesiano que trabaje con datos incompletos. En este trabajo se presenta una metodología para predecir el recurso eólico mediante redes Bayesianas en el estado de Querétaro. Esta metodología representa un estudio preliminar de las características de la velocidad del viento a largo plazo en ocho sitios del área metropolitana de Querétaro.

1.1 Planteamiento del problema

La energía eólica ha recibido un gran crecimiento a nivel mundial, los países líderes en cuanto a capacidad instalada son China, Estados Unidos, Alemania, India y España (Sawin et al., 2018), generando la instalación de parques eólicos para satisfacer la demanda de energías renovables, véase Figura 1.1a.

En México a finales de 2017 se ubican 45 parques eólicos cuya capacidad

instalada alcanza los 4,199 MW, que representa el 6 % de la capacidad total instalada y el 18 % de la capacidad generada mediante energías limpias (SENER, 2018). De acuerdo con AMDEE (2019), la mayor concentración de parques eólicos se localiza en el estado de Oaxaca, con el 58.9 % de la capacidad instalada, véase Figura 1.1b.

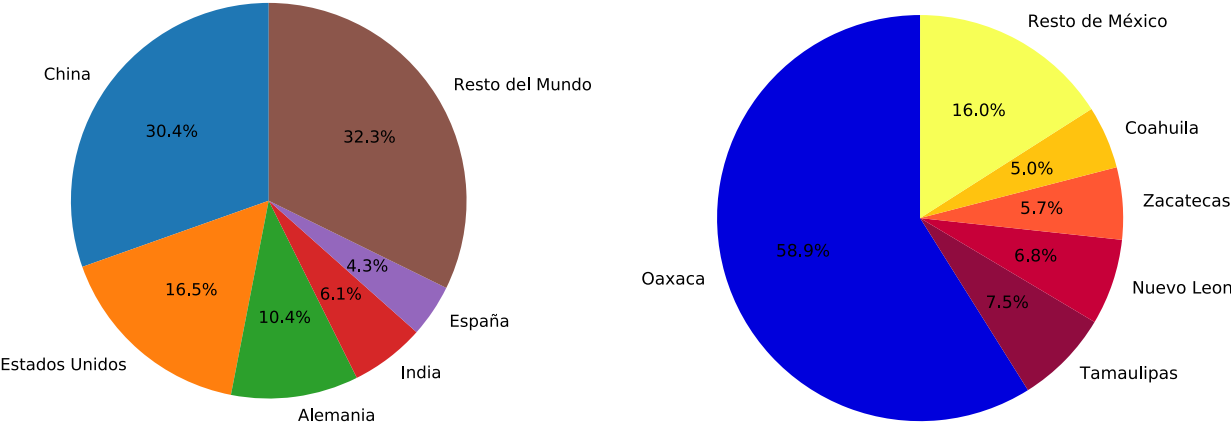


Figura 1.1 Capacidad eólica instalada: (a) a nivel mundial en diciembre de 2017 y (b) en México a finales de 2017.

En el mundo se han realizado estudios de evaluación del recurso eólico a nivel global, por ejemplo, pronóstico del viento en Dakota del Norte (Li et al., 2011), el Noroeste del Pacífico (Slougher et al., 2013), África Oriental (Mughal et al., 2018), entre otros sitios. Mientras tanto en México, el viento ha sido estudiado como fuente de energías renovables en diferentes áreas geográficas. En el Istmo de Tehuantepec se han realizado diversos estudios para el pronóstico del recurso eólico: Cadenas and Rivera (2007) usaron una red neuronal ADALINE en La Venta y Cortés et al. (2012) usaron un modelo neuro-difuso ANFIS en Tehuantepec. Así como el atlas del recurso eólico del estado de Oaxaca producido por NREL en 2004 (Elliott et al., 2004). En otros estados se han realizado diversos estudios acerca del potencial eólico, por ejemplo, posibles emplazamientos para la instalación de parques eólicos en Veracruz (Cancino-Solórzano et al., 2010), pronóstico del potencial eólico en el noroeste de la Península de Yucatán [Soler-Bientz et al. (2010); Figueroa-Espinoza et al. (2014)], en el Noreste de México (Hernández-Escobedo et al., 2014), en Tamaulipas (Carrasco-Díaz et al., 2015) y en Baja California (Cadenas and Rivera, 2010).

A pesar del amplio potencial eólico que tiene México, la energía eólica no ha sido evaluada a detalle en el interior del país, y en específico la región del Bajío [De la Torre-Gea et al. (2011); Alvarado et al. (2018)].

1.2 Justificación

El empleo de las técnicas de inteligencia artificial como las redes bayesianas, se ha convertido en un medio fiable que complementa los métodos tradicionales en la predicción del recurso eólico de una manera más eficiente. Además, en el caso del sector eólico existen varias razones que justifican el uso de una metodología mediante redes bayesianas:

- Son adecuadas para trabajar con conjuntos de datos incompletos (Riggelsen, 2006).
- Se identifican relaciones previamente indeterminadas entre variables, facilitando el aprendizaje sobre las relaciones causales que hay entre ellas (De la Torre-Gea et al., 2011).
- También se proporcionan herramientas para la predicción, modelado, optimización y control de procesos en las energías renovables (Borunda et al., 2016).
- El modelo se genera con el histórico de las mediciones evitando presuponer estructuras previas.

De acuerdo con AMDEE (2019) se tiene previsto que para el año 2024 en el estado de Querétaro se encuentren instalados 60 MW de potencia en energía eólica en la entidad. Por lo que uno de los beneficios esperados es que a través de la predicción del recurso eólico mediante redes bayesianas se obtenga una metodología con el fin de pronosticar la velocidad del viento a largo plazo. En concordancia con Elliott et al. (2004) esto permitirá la identificación áreas aptas para la generación de electricidad mediante proyectos a escala comercial (parques eólicos), energía para poblaciones rurales y energía eólica fuera de la red. De esta forma se podría:

- Promover la investigación y el desarrollo del sector eólico en la entidad, actualmente existen instituciones públicas como la Universidad Autónoma de Querétaro, el CIATEQ y el CIDESI que forman parte del Centro Mexicano de Innovación en Energía Eólica (CEMIE-Eólico, 2019).
- Determinar la áreas de interés para licitación y construcción de parques eólicos, en el estado de Querétaro, ejemplo de ello es el parque Santiago Eólico con capacidad de 105 MW, el cual será construido en el estado de Guanajuato (Iberdrola-México, 2018).
- Obtener energía eléctrica renovable y limpia para el sector empresarial, hogares e instituciones públicas, evitando así la contaminación por medio de gases de efecto invernadero.

1.3 Objetivos

1.3.1 Objetivo general

Desarrollar una metodología mediante redes bayesianas que permita el pronóstico de la velocidad del viento en un horizonte de tiempo a largo plazo para predecir el recurso eólico en el estado de Querétaro.

1.3.2 Objetivo particulares

1. Investigar los métodos para predicción de variables del viento que existen.
2. Obtener los datos del viento de estaciones meteorológicas identificadas en el estado de Querétaro.
3. Diseñar una metodología mediante redes bayesianas para la predicción del recurso eólico en un horizonte de tiempo a largo plazo.
4. Implementar un prototipo como un conjunto de librerías de software.

5. Realizar un diseño experimental para validar la metodología generada.

1.4 Hipótesis

La hipótesis que se ha considerado en esta tesis es:

Se puede desarrollar una metodología mediante redes bayesianas que permita predecir el recurso eólico en el estado de Querétaro, pronosticando la velocidad del viento en un horizonte de tiempo a largo plazo, a partir de la medición local en diferentes sitios.

1.5 Delimitaciones

1. Se consideran las estaciones meteorológicas que se encuentran dentro del área metropolitana de la ciudad de Querétaro.
2. El método para la estimación de parámetros en la red bayesiana es el algoritmo de *Expectation-Maximization*.
3. Se utilizó la variable de la velocidad del viento de ocho sitios durante el periodo comprendido entre Mayo 2017 hasta Abril 2018.

1.6 Contribuciones

Las contribuciones derivadas de este trabajo de tesis se mencionan a continuación:

- Un artículo aceptado para ponencia en el *II Congreso Nacional de Computación y Tecnología Educativa* que se llevo a cabo del 20 al 22 de noviembre de 2018 en Santiago de Querétaro, Qro. El titulo del articulo es: "Análisis de dependencias entre variables de una estación de medición de viento".

- Participación en el *Fondo de Proyectos Especiales de Rectoría* (FOPER) como responsable del proyecto titulado “Uso de redes bayesianas para la predicción de series de tiempo: el caso en la velocidad del viento” durante el periodo marzo - noviembre de 2018.
- Un artículo aceptado y presentado en el *XV Congreso Internacional de Ingeniería* (CONIIN 2019) que se llevó a cabo del 13 al 17 de mayo del presente año en Santiago de Querétaro, Qro. El título del artículo es: “Analysis of geographic dependencies of wind speed through a Bayesian process”.

1.7 Estructura de la tesis

El presente trabajo de tesis desarrollado se encuentra organizado de la siguiente manera: En el Capítulo 2, se presenta una breve descripción de los fundamentos teóricos como la clasificación de escalas de tiempo en la predicción, los métodos de predicción, el algoritmo de Expectation-Maximization y los métodos de predicción de acuerdo a su función de densidad. En el Capítulo 3, se describe a detalle la metodología utilizada: el área de estudio y recolección de los datos, el procesamiento de datos del viento, la estimación de los parámetros y el análisis de correlación. En el Capítulo 4, se hace un análisis de los sitios, se describen los resultados obtenidos de los diferentes escenarios, así como, una comparativa de los trabajos consultados y esta tesis. Por último en el Capítulo 5 se mencionan las conclusiones finales y las sugerencias de trabajos a futuro.

2. FUNDAMENTOS TEÓRICOS

El pronóstico en la velocidad del viento ha sido utilizado en mayor o menor medida en el sector eólico: en la identificación de zonas aptas para instalación de proyectos eólicos [Grothe and Schnieders (2011); Sobolewski and Feijóo (2013)]; en los estudios probabilísticos de flujo de carga en redes de energía (Diaz et al., 2012) y en la oferta-demanda de energía eléctrica (Elberg and Hagspiel, 2015).

2.1 Clasificación de escalas de tiempo en la predicción

De acuerdo con Soman et al. (2010), la predicción del potencial eólico ha sido medido en diferentes horizontes de predicción, el cual va desde unos minutos hasta varios días o más dependiendo del estudio a realizar. Las técnicas de pronóstico son agrupadas en: muy corto, corto, mediano y largo plazo, véase Tabla 2-1. En esta investigación se analiza una predicción en un horizonte a largo plazo.

Tabla 2-1 Escala de tiempo en la predicción del recurso eólico.

Horizonte de tiempo	Rango
Muy corto plazo	De pocos segundos a 30 minutos en adelante
Corto plazo	De 30 minutos a 6 horas en adelante
Mediano plazo	De 6 horas a 1 día en adelante
Largo plazo	De 1 día a 1 semana o más en adelante

2.2 Métodos de predicción del potencial eólico

Dentro de la literatura relacionada [De Giorgi et al. (2011); Foley et al. (2012); Tastu et al. (2013)], existen diferentes métodos para el pronóstico del potencial eólico. Generalmente se dividen en métodos físicos, métodos estadísticos tradicionales o de “caja negra”, estos también incluyen a los llamados métodos de

aprendizaje basados en inteligencia artificial o de “caja gris”. Los métodos híbridos incluyen la combinación de alguna de las características de los métodos anteriores.

2.2.1 Métodos físicos

Los métodos físicos son aquellos que consideran las predicciones meteorológicas numéricas y se basan en modelos dinámicos de los flujos atmosféricos, como los modelos mesoescala [Figuroa-Espinoza et al. (2014); Gross and Magar (2015)]. Estos consisten en varios submodelos que en su conjunto entregan el pronóstico eólico en un determinado punto de la red (Monteiro et al., 2009). El punto clave de los métodos físicos es mejorar las predicciones utilizando consideraciones físicas sobre el terreno como son: la rugosidad, la orografía, los obstáculos y el perfil del viento (De Giorgi et al., 2014).

2.2.2 Métodos estadísticos

Los métodos estadísticos se analizan una gran cantidad de datos y los procesos meteorológicos no se encuentran representados notoriamente. Estos métodos son basados en uno o varios modelos que permiten encontrar una relación entre las variables medidas por una estación de viento. Por lo general, se emplean series de tiempo, regresiones o técnicas de inteligencia artificial (De Giorgi et al., 2011).

Se han desarrollado muchas investigaciones para predecir el potencial eólico utilizando métodos estadísticos. Los modelos autorregresivos han sido utilizados para estimar dependencias temporales en una sola localización (Koivisto et al., 2016). Carta and Velázquez (2011), realizó un pronóstico a largo plazo de la velocidad del viento mediante el método *Measure Correlate Predict*. El uso de modelos de copula para el análisis de dependencias geográficas en el sector eólico ha sido utilizado por [Grothe and Schnieders (2011); Elberg and Hagspiel (2015)].

Las técnicas de inteligencia artificial han sido empleadas en el pronóstico de la velocidad del viento, como son las redes neuronales (Li et al., 2011), las máquinas

de soporte vectorial (Santamaría-Bonfil et al., 2016), los algoritmos genéticos (Liu et al., 2011) y las redes bayesianas (De la Torre-Gea et al., 2011).

2.2.3 Métodos híbridos

Los modelos híbridos se benefician de las predicciones de los métodos estadísticos en un horizonte a corto plazo, así como, de los pronósticos de los métodos físicos en un horizonte a largo plazo, teniendo en cuenta las características del terreno (Monteiro et al., 2009).

2.3 Algoritmo de Expectation-Maximization

La pérdida de datos es relativamente frecuente en muchos problemas con datos reales, en sensores para la medición de viento puede ocurrir principalmente por cortes durante la transmisión o fallos durante el proceso de adquisición de los datos (Alcaraz-Jiménez, 2006). Sin embargo, Little and Rubin (2019) describen diversos métodos para tratar la pérdida de datos, uno de estos es la estimación de la media y la matriz de covarianzas de forma iterativa permitiendo sustituir los valores perdidos mediante estimaciones razonables.

El algoritmo de Expectation-Maximization (EM) es un método iterativo para la estimación de la máxima verosimilitud, útil para trabajar con datos incompletos [Dempster et al. (1977); Lauritzen (1995); Riggelsen (2006)]. Anteriormente el algoritmo EM había sido propuesto por diversos autores [Hartley (1958); Woodbury (1970); Sundberg (1974)] en distintas circunstancias, pero no fue hasta que Dempster et al. (1977), generalizó el método y desarrolló las bases del mismo. Después de esto han surgido muchas variantes y modificaciones del algoritmo original [Lauritzen (1995); Friedman (1998)].

El algoritmo EM se utiliza frecuentemente como un algoritmo de agrupamiento en el área de *Machine Learning* como el método de cadenas de Markov

Monte Carlo (Chai et al., 2017) y en mixturas de Gaussianas (Hernández-Díaz et al., 2015), utilizadas en procesos de clasificación. Otra aplicación del algoritmo EM es la estimación de parámetros para el aprendizaje en una red neuronal (Li et al., 2011) y en una red Bayesiana [Riggelsen (2006); Reed and Mengshoel (2014)].

En cuanto al uso del algoritmo EM en la predicción del recurso eólico solo se ha utilizado en dos trabajos consultados. En Li et al. (2011), se aplican tres modelos de red neuronal en el que se tiene los pronósticos de la velocidad del viento, posteriormente utilizan el método EM para promediar la varianza de cada uno de los pronósticos obtenidos, teniendo un mejor rendimiento en el pronóstico final. Por su parte, Baran (2014) emplea el algoritmo EM para estimar los parámetros de la función normal truncada de la distribución del viento. Este enfoque Bayesiano lo ponen a prueba con dos grandes conjuntos de datos obtenido mejores rendimientos y siendo más eficiente en tiempos de cómputo.

2.4 Métodos de predicción de acuerdo a su función de densidad

Dentro de las metodologías existentes para predecir la función de densidad del potencial eólico, la literatura menciona los métodos: paramétricos y no paramétricos, ambas técnicas han sido utilizadas [Feijóo et al. (2011); Puth et al. (2015)]. Los métodos paramétricos se basan en la suposición de que se conoce la forma de las funciones de densidad de la variables y pueden ser descritas por una función de densidad de probabilidad (Tastu et al., 2013). Los métodos no paramétricos suelen requerir suposiciones menos restrictivas en cuanto a la medición de los datos y menos suposiciones acerca de la forma de las distribución de probabilidad. Esta característica permite que estos métodos sean más aplicables y fáciles de usar a los conjuntos de datos donde su función de distribución no puede ser definida a priori (Weaver et al., 2017).

2.4.1 Métodos no paramétricos

Este trabajo se ha centrado en los métodos no paramétricos debido a una menor restricción en la medición de los datos, así como a una menor suposición acerca de la distribución de las mediciones en el viento. Algunos de los métodos no paramétricos que se conocen son: la prueba Kruskal-Wallis, la prueba de los rangos con signo de Wilcoxon, la prueba Kolmogórov-Smirnov y el coeficiente de correlación de Spearman. A continuación se explicará el método no paramétrico de coeficiente de correlación de Spearman el cual fue utilizado en esta investigación.

2.4.2 Correlación

De acuerdo con Weaver et al. (2017), para determinar si una variable varía con otra, se usa la correlación para representar el nivel de asociación o relación entre dos variables estadísticas. La evaluación de la correlación entre dos variables permite establecer la fuerza de esta relación; el sentido entre la variación de las variables, si es positiva, negativa o inexistente; así como la forma de la línea que mejor ajuste: la línea recta, la curva monótonica y la curva no monótonica.

El análisis de la correlación proporciona un valor y una dirección cuantificable, pero no permite determinar la causa y efecto entre la relación de dos variables. Los valores cuantificables son el valor de la probabilidad (p) que indica la importancia de la relación y el coeficiente de correlación (ρ) que indica la fuerza de la relación.

Para Mackridge and Rowe (2018), el valor de probabilidad p se analiza con respecto al valor crítico (0.05 u otro de acuerdo a la elección personal) para determinar el significado, de esta manera cuando se evalúa la relación entre dos variables, el valor de probabilidad indica si existe una asociación entre las variables. El valor crítico de $p \leq 0.05$ es considerado una relación estadísticamente significativa, menor a esto no es estadísticamente significativo.

2.4.3 Coeficiente de correlación de Spearman

El coeficiente de correlación de Spearman es una técnica no paramétrica que se aplica en aquellos estudios en los que no se conoce la forma de su distribución, para los cuales se cuentan con datos de intervalo o de razón y su forma es monótonica. Este coeficiente varía de -1.0 a +1.0 y se interpreta de la siguiente manera:

- Valores cercanos a +1.0 indican que existe una fuerte relación entre las variables, es decir, a medida que aumenta una variable también se incrementa la otra.
- Valores cercanos a -1.0 señalan que hay una fuerte relación negativa entre las variables, mientras que una aumenta la otra disminuye.
- Cuando el valor es 0.0 no existe una correlación entre las variables.

La tabla con los valores críticos del coeficiente de correlación de rangos de Spearman utilizada en esta tesis se tomó del trabajo de Ramsey (1989) y es mostrada en la Tabla 2-2.

2.5 Comentarios

En esta tesis, se presenta un proceso estocástico para obtener una metodología para la predicción del recurso eólico analizando las dependencias geográficas, utilizando el algoritmo EM para obtener una matriz de relaciones de las series de tiempo de la velocidad del viento registradas en 8 sitios ubicadas en Querétaro. Finalmente, se utiliza el coeficiente de correlación de Spearman para determinar el grado de asociación entre los sitios en diferentes escenarios.

Tabla 2-2 Valores críticos del coeficiente de correlación de rangos de Spearman.

Número de casos	Valor crítico para α :			
	$P = 0.10$	$P = 0.05$	$P = 0.02$	$P = 0.01$
5	0.900	–	–	–
6	0.829	0.886	0.943	–
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.678	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.626	0.679
15	0.446	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570
21	0.370	0.435	0.508	0.556
22	0.361	0.425	0.496	0.544
23	0.353	0.415	0.486	0.532
24	0.344	0.406	0.476	0.521
25	0.337	0.398	0.466	0.511

3. METODOLOGÍA

En este Capítulo se describen los procedimientos realizados para desarrollar la predicción del recurso eólico mediante un proceso Bayesiano en Querétaro y dar cumplimiento a lo objetivos planteados en esta tesis. La metodología utilizada se muestra en la Figura 3.1. Posteriormente se detalla cada fase de la metodología planteada.

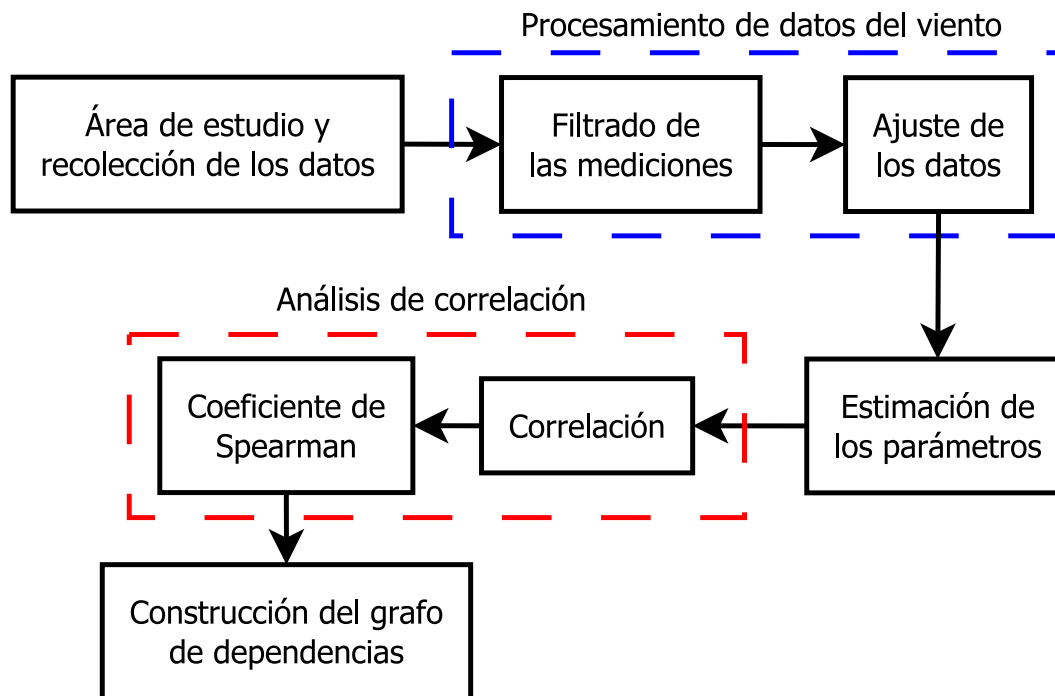


Figura 3.1 Esquema de la metodología utilizada.

3.1 Área de estudio y recolección de los datos

3.1.1 Descripción de la región de estudio

El estado de Querétaro es una de las 32 entidades federativas que integran a México. Colinda al norte con Guanajuato y San Luis Potosí, al sur con Hidalgo y el Estado de México y al oeste con Michoacán. La zona metropolitana de la ciudad de Querétaro tiene dos municipios, El Marqués y Querétaro, que en su conjunto ocupan

una superficie de 1,477 Km² que representa el 12.6% del territorio estatal. Actualmente estos dos municipios tiene una población de 1 millón 35 mil 206 habitantes, lo que representa el 50.8% de la población estatal (INEGI, 2016). Su localización geográfica se aprecia en la Figura 3.2.

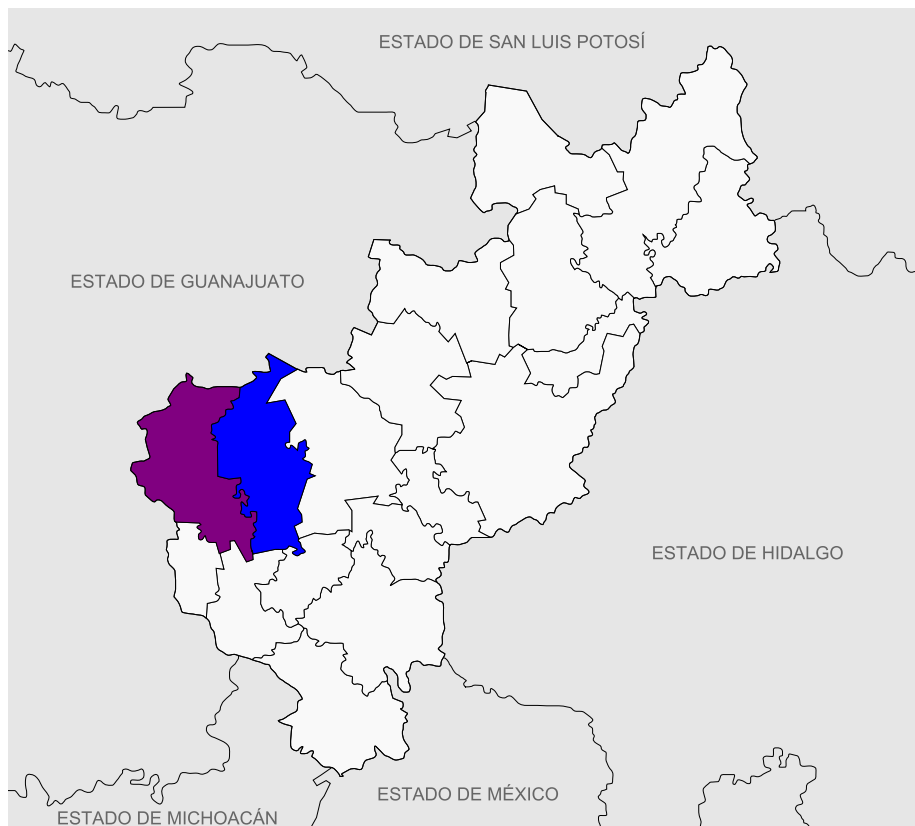


Figura 3.2 Localización geográfica de los municipios de Querétaro (morado) y El Marqués (azul).

3.1.2 Descripción de los sitios

La localización geográfica de las estaciones meteorológicas se muestran en la Figura 3.3. Los datos de la velocidad del viento fueron proporcionados en todos los sitios por la Red Centro de Investigación de Aguas de Querétaro (Red-CIAQ, 2018), excepto la estación Juriquilla por la Red Universitaria de Observatorios Atmosféricos (RUOA, 2018).

La información geográfica de los 8 sitios y las características de la velocidad del viento durante el periodo del 1 mayo de 2017 al 30 abril de 2018 se encuentran resumidos en la Tabla 3-1.

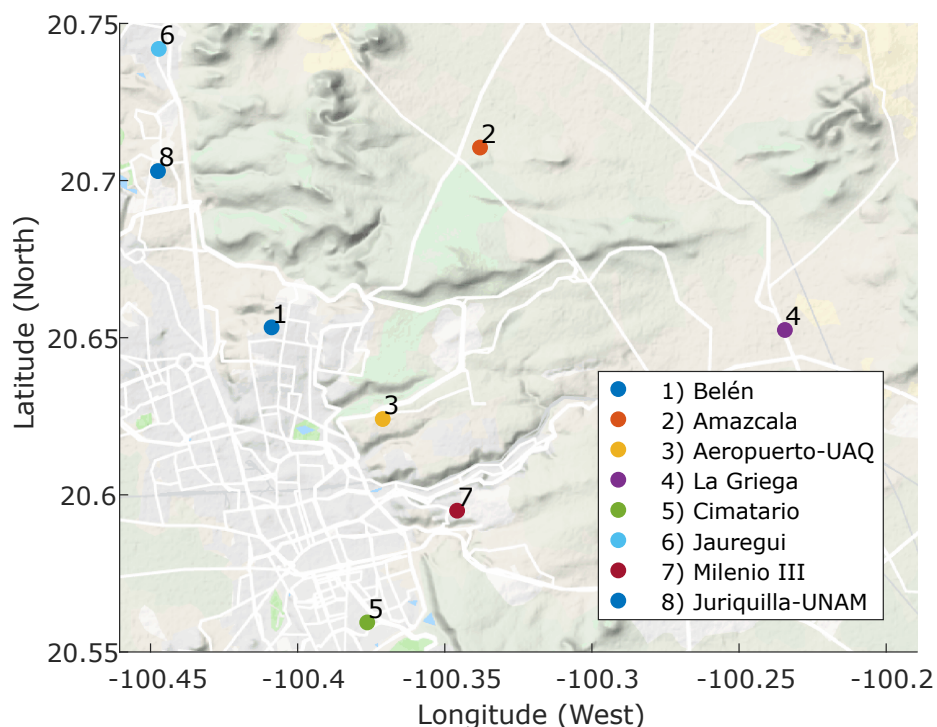


Figura 3.3 Estaciones meteorológicas en Querétaro y El Marqués.

Tabla 3-1 Información geográfica de los ocho sitios y sus velocidades del viento.

Sitios	Latitud (Norte)	Longitud (Oeste)	Elevación (m)	Velocidad del viento (m/s)	
				<i>Media</i>	<i>Desv. estándar</i>
Belén	20°39'11"	100°24'31"	1964	0.857	0.989
Amazcala	20°42'37"	100°20'16"	1970	0.438	1.044
Aeropuerto	20°37'26"	100°22'15"	1973	4.323	2.187
La Griega	20°39'08"	100°14'03"	1909	1.060	1.449
Cimatarío	20°33'33"	100°22'34"	1924	1.366	1.255
Jauregui	20°44'30"	100°26'49"	1968	1.308	1.207
Milenio III	20°35'41"	100°20'44"	1966	1.408	1.082
Juriquilla	20°42'10"	100°26'50"	1943	2.853	1.434

3.2 Procesamiento de datos del viento

3.2.1 Filtrado de las mediciones

Las mediciones fueron tomadas cada minuto a 10 metros sobre el nivel del suelo para los 8 sitios. En el proceso de recopilación de los datos, las mediciones de la velocidad del viento se registraron continuamente y los datos se tomaron

promediados cada 10 minutos, lo cual cumple con lo establecido en la norma IEC 61400-12-1 (IEC, 2017). Así mismo, los datos fueron convertidos de Km/h a m/s, la cual es la manera más común de manejar los datos del viento en la predicción del recurso eólico.

3.2.2 Ajuste de los datos

Para el ajuste de los datos se utilizó la normalización estadística conocida como característica *MinMax Scaling* esta suele ajustar todos los valores en el rango de [0,1], también conocida como normalización basada en unidad, y es definida en la Ecuación 3.1.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.1)$$

3.3 Estimación de los parámetros

Una vez procesados los datos de la velocidad del viento en los ocho sitios en cuestión, estos se representan como $M = \{(\mu_1, \sigma_1), \dots, (\mu_k, \sigma_k)\}$, donde cada (μ_1, σ_1) , representa los parámetros de una Gaussiana o un sitio en particular. Para estimar los parámetros de la Gaussiana, se considera una serie de tiempo de la velocidad del viento de un sitio particular que corresponden a un intervalo de tiempo determinado, en este caso un año.

De acuerdo con Jiménez-Hernández et al. (2010), los parámetros Gaussianos se estiman por medio del algoritmo EM de forma recursiva, y es definida en la Ecuación 3.2 y 3.3, para el caso monovariado monomodal.

$$\mu_{k+1} = \rho\mu_k + (1 - \rho)x_k \quad (3.2)$$

$$\sigma_{k+1}^2 = \rho\sigma_k^2 + (1 - \rho)(x_k - \mu_{k+1})^2 \quad (3.3)$$

donde $\rho \in [0, 1]$ el cual representa el factor de aprendizaje.

La estimación de los parámetros utilizando el algoritmo EM permite obtener la variabilidad de los datos y la asociación que hay entre las variables, esto se resume en la matriz de covarianzas. Esta matriz es cuadrada y simétrica de orden k , donde los términos diagonales son las varianzas y lo no diagonales, las covarianzas entre las variables, la Ecuación 3.2 y 3.3 se reescriben en forma vectorial como:

$$\mu_{k+1} = \rho\mu_k + (1 - \rho)\mathbf{x}_k \quad (3.4)$$

$$\Sigma_{k+1}^2 = \rho\Sigma_k^2 + (1 - \rho)(\mathbf{x}_k - \mu_{k+1})(\mathbf{x}_k - \mu_{k+1})^T \quad (3.5)$$

3.4 Análisis de correlación

Teniendo la matriz de covarianzas se puede obtener la matriz de correlación aplicando la Ecuación 3.6. Los valores de la correlación se ubican entre -1 y +1. La matriz de correlación es cuadrada y simétrica que tiene unos en su diagonal y fuera de ella los coeficientes de correlación entre las variables, véase Ecuación 3.7.

$$\rho(A, B) = \frac{cov(A, B)}{\sigma_A\sigma_B} \quad (3.6)$$

$$R = \begin{pmatrix} 1 & \rho(1, 2) & \cdots & \rho(1, n) \\ \rho(2, 1) & 1 & \cdots & \rho(2, n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(n, 1) & \rho(n, 2) & \cdots & 1 \end{pmatrix} \quad (3.7)$$

Posteriormente, se aplica el coeficiente de correlación de Spearman utilizando como valor crítico ρ de 0.05, el cual representa un nivel de confianza de 95 %, recordando que (Mackridge and Rowe, 2018) indica que esta valor crítico es estadísticamente significativo.

3.5 Comentarios

El algoritmo EM y el coeficiente de correlación de Spearman son métodos estadísticos que permiten trabajar con variables continuas como la velocidad del viento. Ambos métodos se han utilizado por separado en diferentes trabajos en la predicción del potencial eólico por lo que en esta tesis se propone el uso de ambos en un proceso Bayesiano.

Los resultados son discutidos en tres escenarios. Como primer escenario se analiza la estación primavera-verano, en el periodo de mayo a octubre. En el segundo escenario, se examina la estación otoño-invierno durante el periodo de noviembre a abril. En el tercer escenario, se analiza un año de mediciones. Finalmente, se discuten y comentan los resultados.

4. RESULTADOS

La siguiente investigación fue implementada en Matlab R2017a, los mapas fueron consultados desde la API Google Static Map (Google, 2019) usando el paquete de mapeo de Bar-Yehuda (2018) y se adaptó el paquete de Tuszynski (2017) para mostrar las curvas de nivel del área de estudio. Los resultados obtenidos mediante el proceso estocástico generado permiten realizar un análisis de dependencias geográficas de la velocidad del viento, en ocho sitios diferentes. A continuación, se discutirá cada uno de los escenarios analizados.

4.1 Análisis de los sitios

La velocidad del viento está en constante cambio, por lo que la predicción del recurso eólico varía continuamente. Estas fluctuaciones dependen de las condiciones climáticas y de la superficie local como las áreas arboladas y el relieve del entorno. En la Figura 4.1, se muestran las series de tiempo de los ocho sitios evaluados durante el mes de diciembre en 2017, en intervalos de una hora.

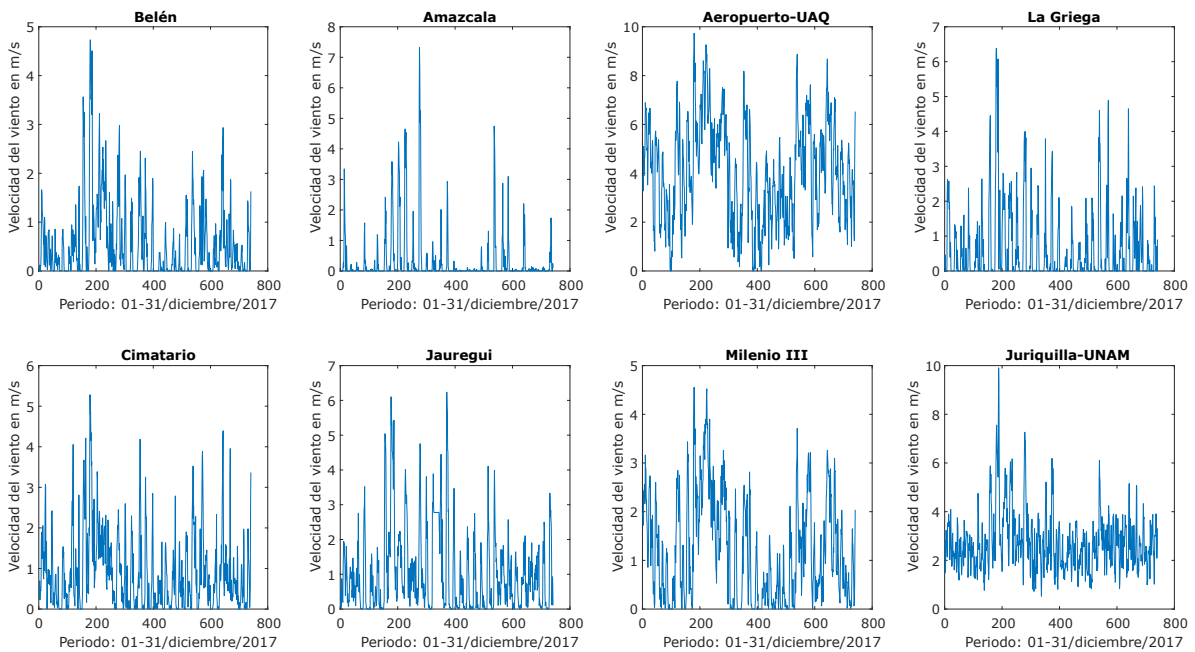


Figura 4.1 Series de tiempo de la velocidad del viento en m/s en diciembre 2017.

De acuerdo con la Tabla 3-1, la media y desviación estándar en cada sitio varía ampliamente, esto demuestra que la función de densidad de probabilidad en cada sitio es diferente, ya que la velocidad del viento no es estable a largo plazo, y por lo tanto es un proceso no estacionario. Así mismo, en cada sitio las variaciones en la velocidad del viento en intervalos mensuales, permite determinar una tendencia que crece o decrece a largo plazo, por lo que la serie no oscila alrededor de un valor constante, véase Figura 4.2.

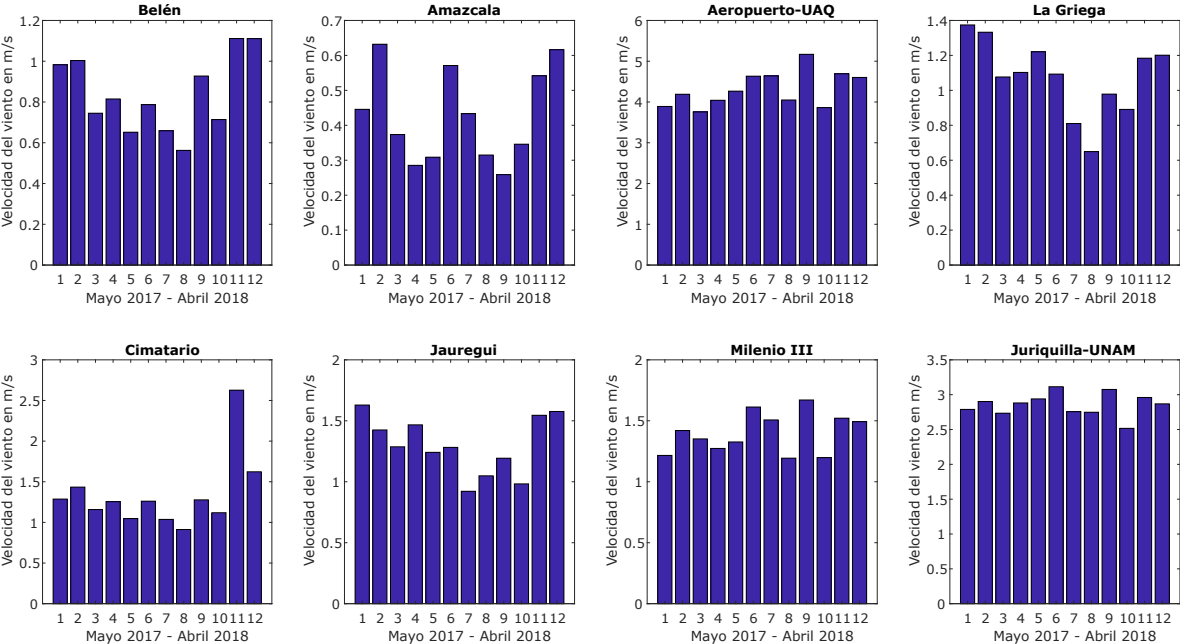


Figura 4.2 Variaciones en la velocidad del viento mensualmente.

4.2 Resultados de los escenarios

4.2.1 Escenario Primavera-Verano

En este escenario se evaluó el periodo de mayo 2017 a octubre 2018. En la Figura 4.3 se muestra como los diferentes sitios se encuentran relacionados. Las correlaciones durante este periodo son generalmente altas entre los sitios, sin embargo, las estaciones más lejanas (por ejemplo, Jauregui y La Griega) tienen correlaciones más bajas. En la Figura 4.4, se muestran las curvas de nivel de las dependencias geográficas durante este periodo. En la Tabla 4-1, se muestran las

correlaciones detalladas de la velocidad el viento en los ocho sitios durante este periodo. Por ultimo, para mayor claridad en la Figura 4.5 se muestra el grafo de la dependencias que se obtuvo en este escenario.

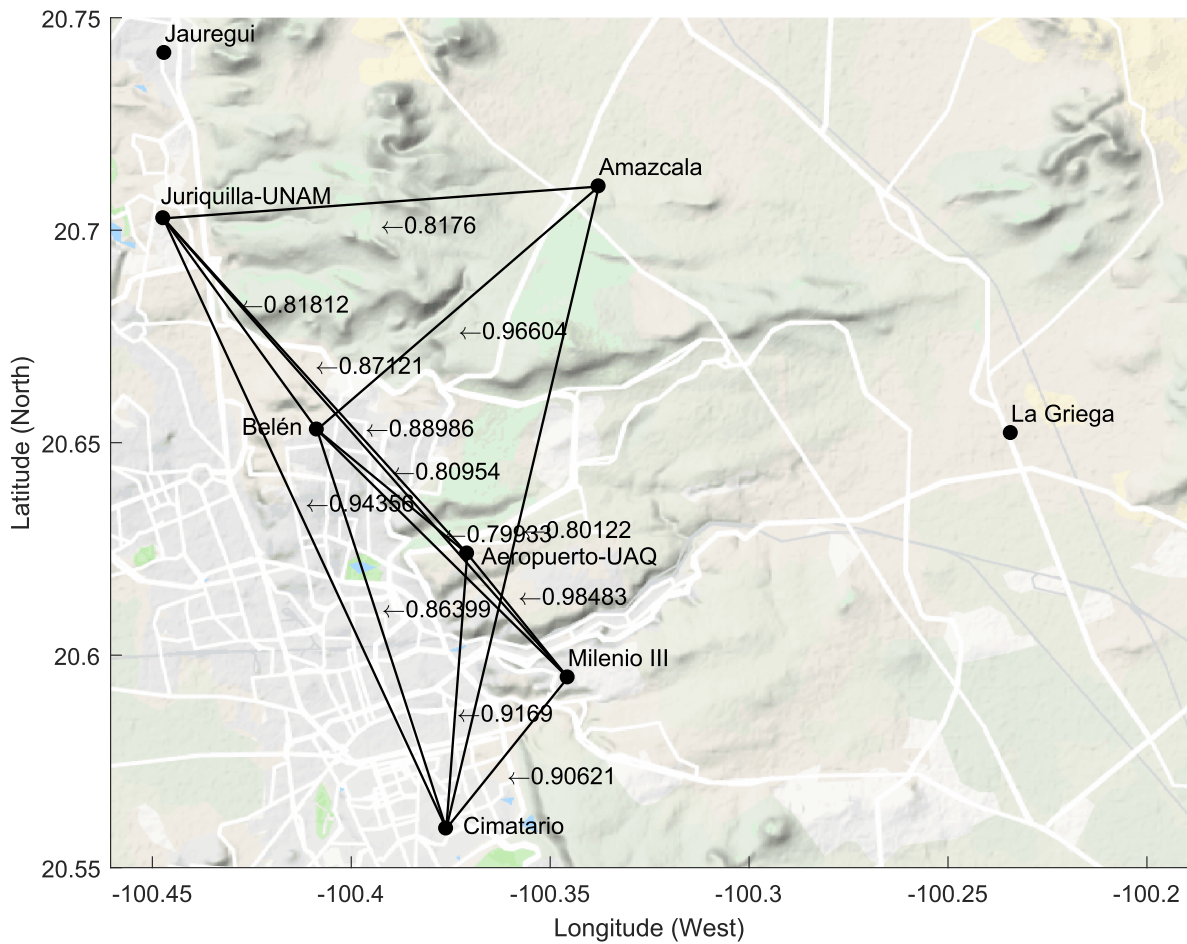


Figura 4.3 Ubicación de las dependencias geográficas durante el periodo mayo 2017 - octubre 2017.

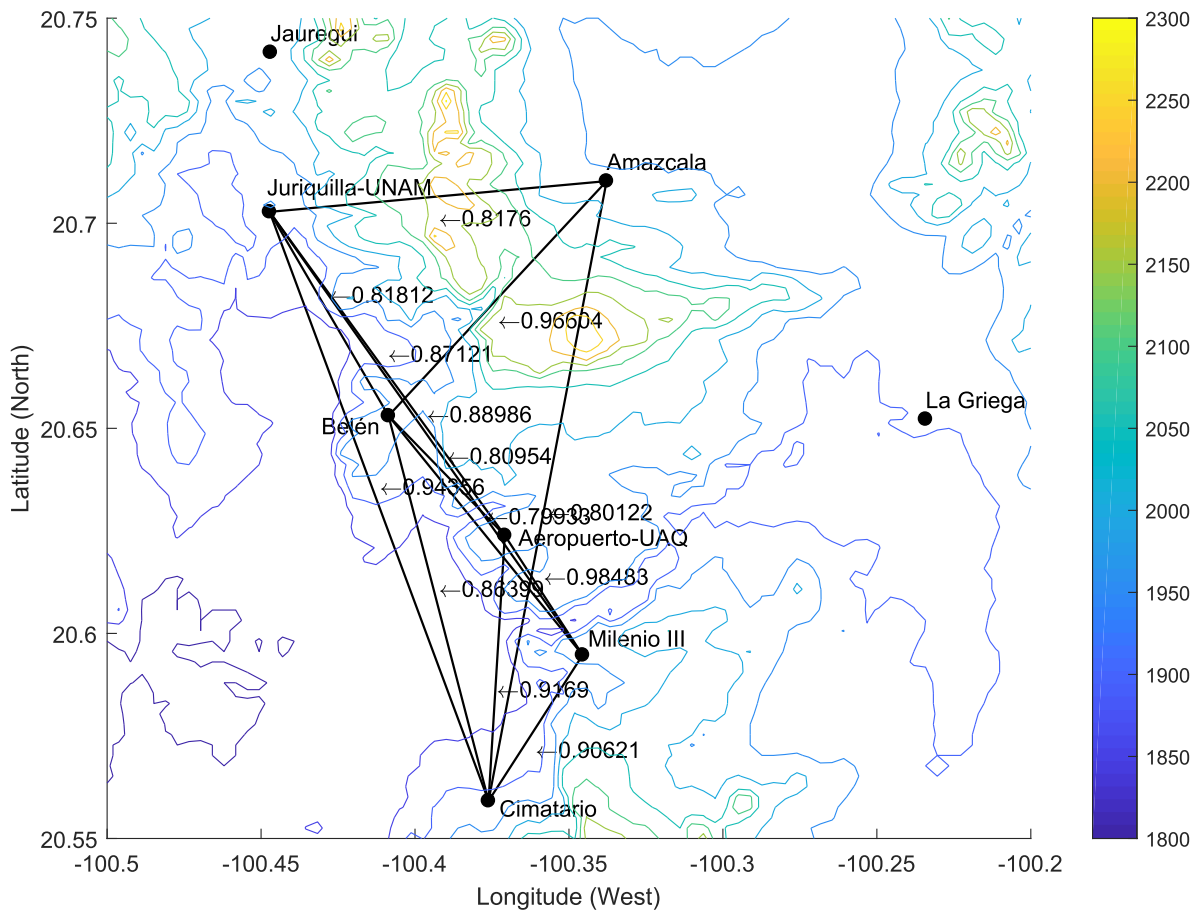


Figura 4.4 Curvas de nivel de las dependencias geográficas durante el periodo mayo 2017 - octubre 2017.

Tabla 4-1 Coeficientes de correlación entre las velocidades del viento de los diferentes sitios para el periodo mayo 2017 a octubre 2017.

Sitios ^a	BN	AZ	AP	LG	CT	JR	M3	JQ
BN	1	0.966	0.809	-0.539	0.864	0.317	0.799	0.818
AZ		1	0.689	-0.425	0.801	0.251	0.695	0.818
AP			1	-0.607	0.917	0.614	0.985	0.871
LG				1	-0.601	-0.084	-0.572	-0.431
CT					1	0.537	0.906	0.944
JR						1	0.590	0.550
M3							1	0.889
JQ								1

^a Acrónimo de sitios: BN = Belen, AZ = Amazcala, AP = Aeropuerto, LG = La Griega, CT = Cimatario, JR = Jauregui, M3 = Milenio III, JQ = Juriquilla.

4.2.2 Escenario Otoño-Invierno

En este escenario se evaluó el periodo de noviembre 2017 a abril 2018. En la Figura 4.6 se muestran las correlaciones entre los diferentes sitios. Las co-

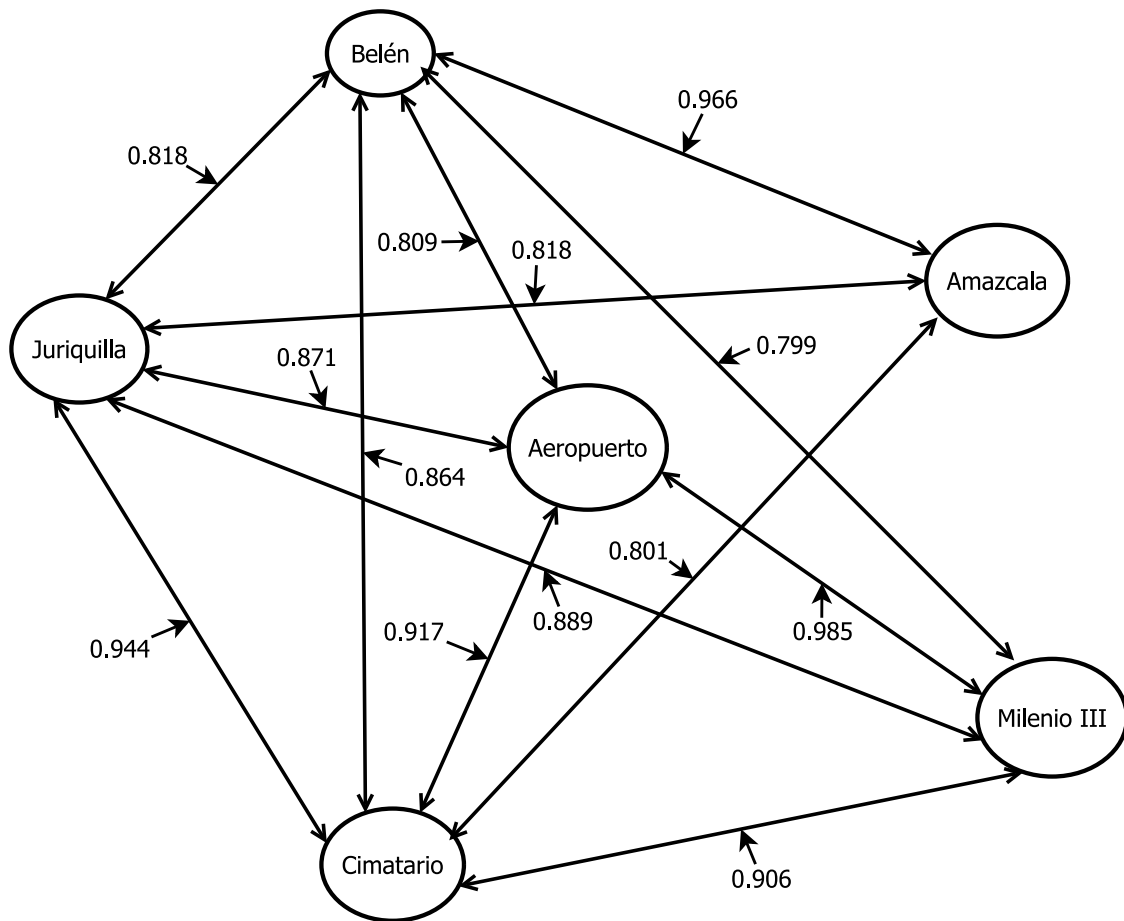


Figura 4.5 Grafos de dependencias de los sitios durante el periodo mayo 2017 - octubre 2017.

relaciones durante este periodo son altas positivamente entre los sitios, Belén – Aeropuerto, Belén – Milenio III y Aeropuerto – Milenio III, por lo que existe una fuerte asociación entre estos sitios. Por otro lado, existe una alta correlación negativa entre los sitios: Belén – Juriquilla y Cimatario – Juriquilla, por lo que cuando en un sitio incrementa la velocidad del viento en el otro disminuye. En la Figura 4.7, se muestran las curvas de nivel de las dependencias geográficas durante este periodo. En la Tabla 4-2, se muestran las correlaciones detalladas de la velocidad el viento en los ocho sitios durante este periodo. Por ultimo, para una mayor claridad en la Figura 4.8 se muestra el grafo de la dependencias que se obtuvo en este escenario.

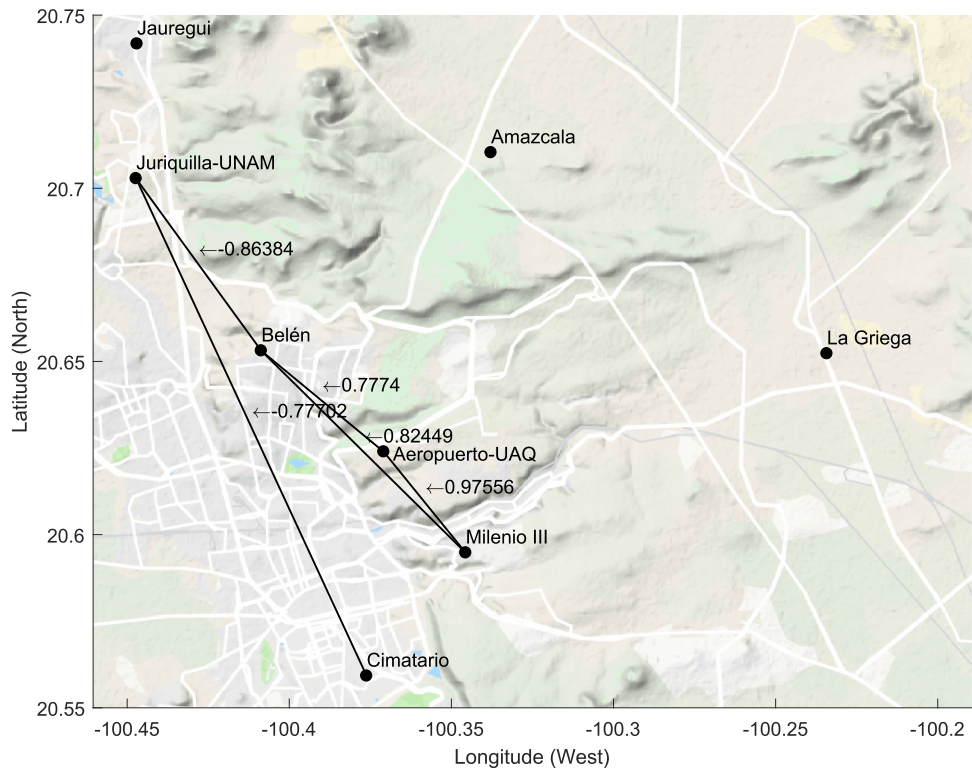


Figura 4.6 Ubicación de las dependencias geográficas durante el periodo noviembre 2017 - abril 2018.

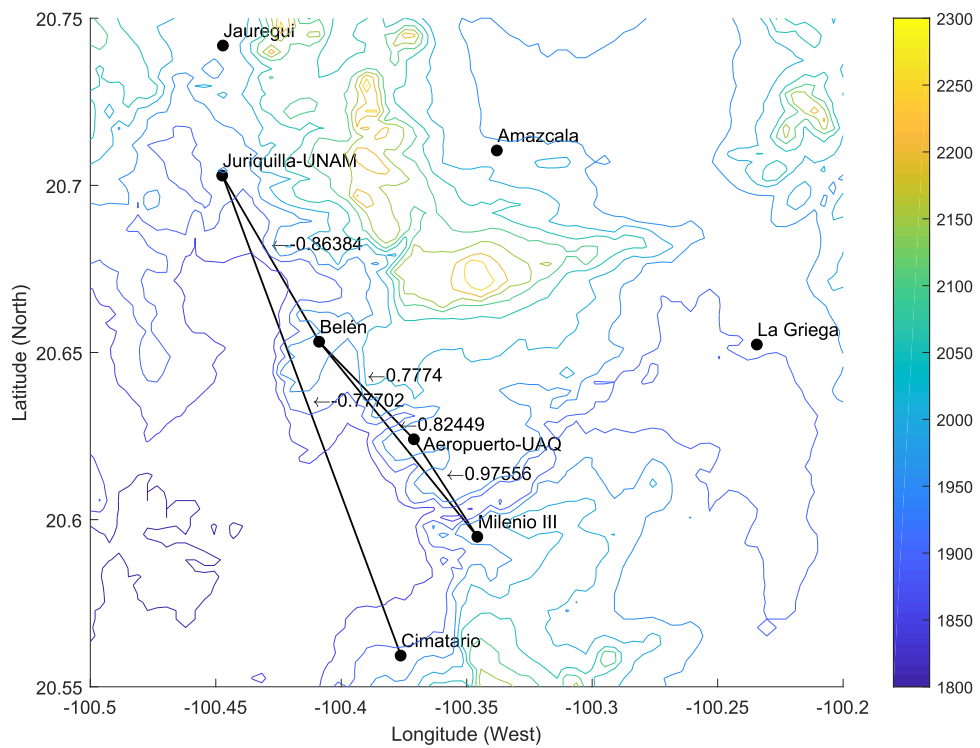


Figura 4.7 Curvas de nivel de las dependencias geográficas durante el periodo noviembre 2017 - abril 2018.

Tabla 4-2 Coeficientes de correlación entre las velocidades del viento de los diferentes sitios para el periodo noviembre 2017 a abril 2018.

Sitios ^b	BN	AZ	AP	LG	CT	JR	M3	JQ
BN	1	0.088	0.777	-0.393	0.592	0.682	0.824	-0.864
AZ		1	-0.250	-0.663	-0.023	-0.137	-0.189	-0.029
AP			1	0.095	0.106	0.502	0.976	-0.499
LG				1	-0.609	-0.186	0.027	0.362
CT					1	0.436	0.189	-0.777
JR						1	0.457	-0.664
M3							1	-0.592
JQ								1

^b Acrónimo de sitios: BN = Belén, AZ = Amazcala, AP = Aeropuerto, LG = La Griega, CT = Cimatario, JR = Jauregui, M3 = Milenio III, JQ = Juriquilla.

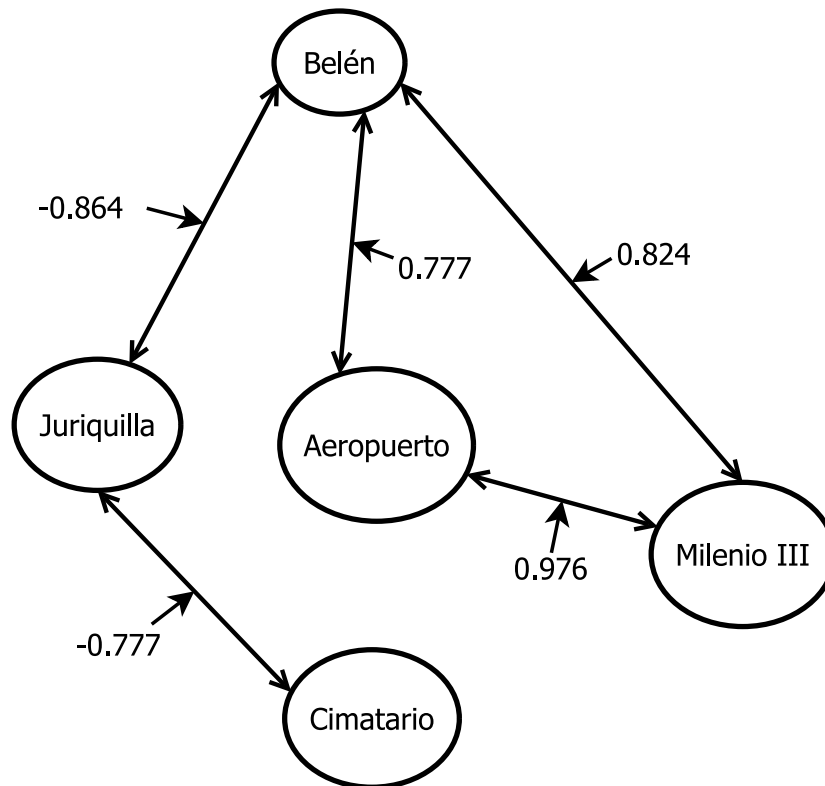


Figura 4.8 Grafos de dependencias de los sitios durante el periodo noviembre 2017 - abril 2018.

4.2.3 Escenario Anual

En este escenario se evaluó el periodo de mayo 2017 a abril 2018. En la Figura 4.9 se muestran las correlaciones entre los diferentes sitios, las correlaciones durante este periodo se comportan de la misma forma que en el escenario otoño-

invierno. En la Figura 4.10, se muestran las curvas de nivel de las dependencias geográficas durante este periodo. En la Tabla 4-3, se muestran las correlaciones detalladas de la velocidad el viento en los ocho sitios durante este periodo. Por ultimo, para mayor claridad en la Figura 4.11 se muestra el grafo de la dependencias que se obtuvo en este escenario.

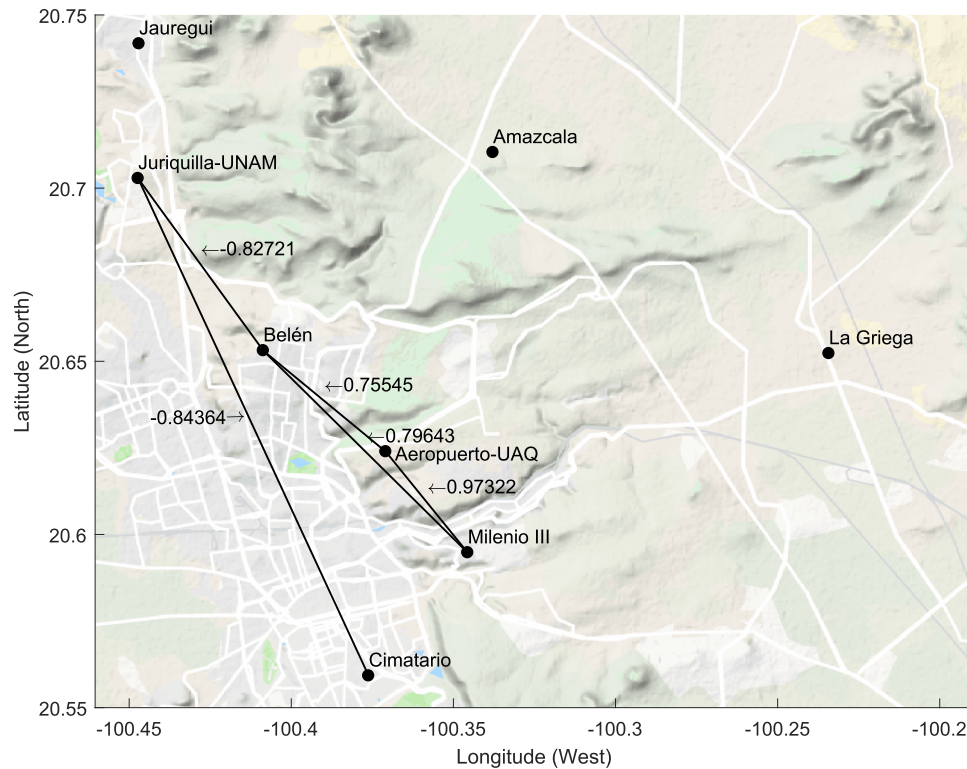


Figura 4.9 Ubicación de las dependencias geográficas durante el periodo mayo 2017 - abril 2018.

Tabla 4-3 Coeficientes de correlación entre las velocidades del viento de los diferentes sitios para el periodo mayo 2017 a abril 2018.

Sitios ^c	BN	AZ	AP	LG	CT	JR	M3	JQ
BN	1	-0.006	0.755	-0.412	0.595	0.694	0.796	-0.827
AZ		1	-0.298	-0.606	-0.162	-0.134	-0.271	0.105
AP			1	0.092	0.110	0.412	0.973	-0.413
LG				1	-0.578	-0.377	0.036	0.379
CT					1	0.558	0.213	-0.843
JR						1	0.345	-0.653
M3							1	-0.522
JQ								1

^c Acrónimo de sitios: BN = Belen, AZ = Amazcala, AP = Aeropuerto, LG = La Griega, CT = Cimatario, JR = Jauregui, M3 = Milenio III, JQ = Juriquilla.

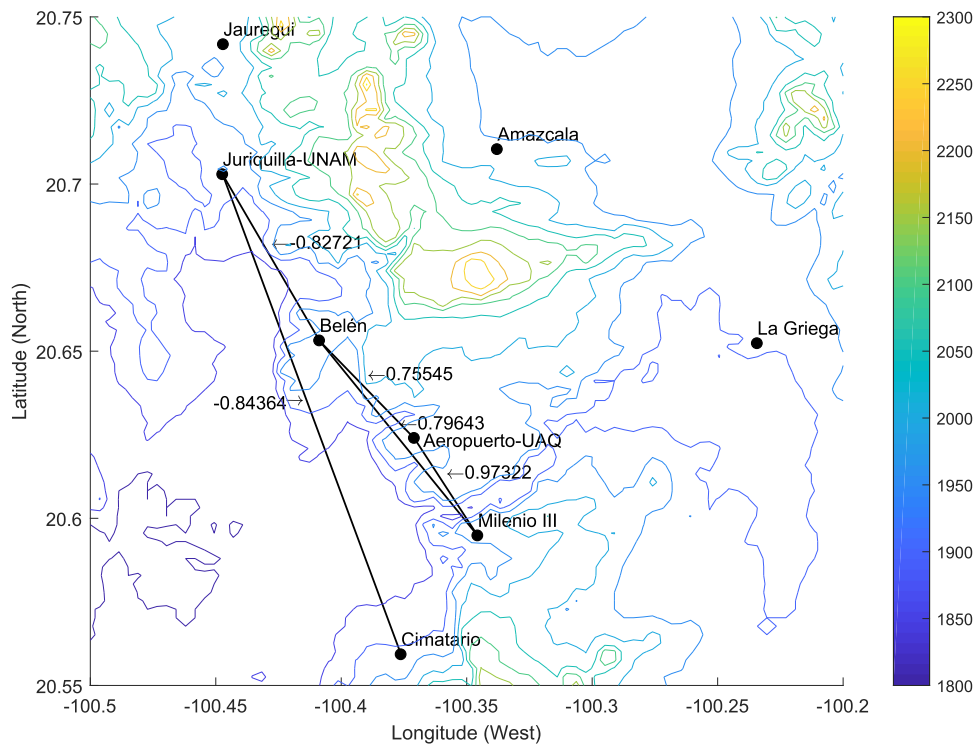


Figura 4.10 Curvas de nivel de las dependencias geográficas durante el periodo mayo 2017 - abril 2018.

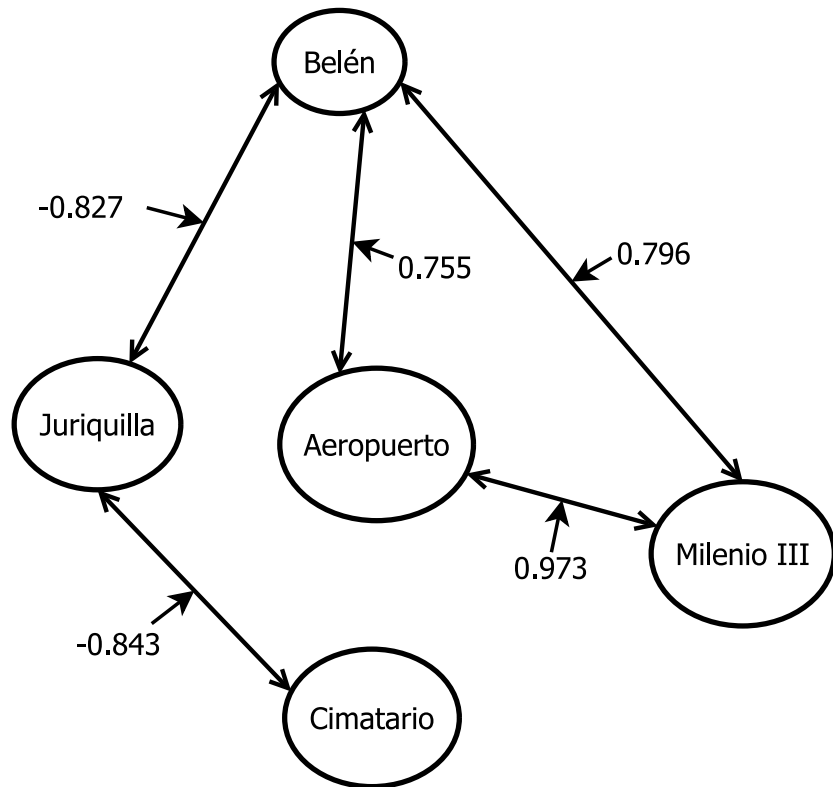


Figura 4.11 Grafos de dependencias de los sitios durante el periodo mayo 2017 - abril 2018.

4.3 Comentarios de los escenarios

Se puede observar en los tres escenarios que los sitios de Jauregui, Amazcala y La Griega, no se encuentran correlacionados con otro sitio, con excepción de Amazcala con respecto a los sitios de Juriquilla, Belén y Cimatario durante el periodo primavera-verano. Esto se entiende geográficamente debido a que el sitio de Jauregui se encuentra entre los límites del valle de Querétaro y el valle de Buenavista (CONCYTEQ, 2003), mientras que los sitios Amazcala y La Griega pertenecen al valle de Amazcala (CONAGUA, 2018a), los sitios restantes pertenecen al valle de Querétaro (CONAGUA, 2018b) y por lo tanto su correlación es muy fuerte, véase Figura 4.12.

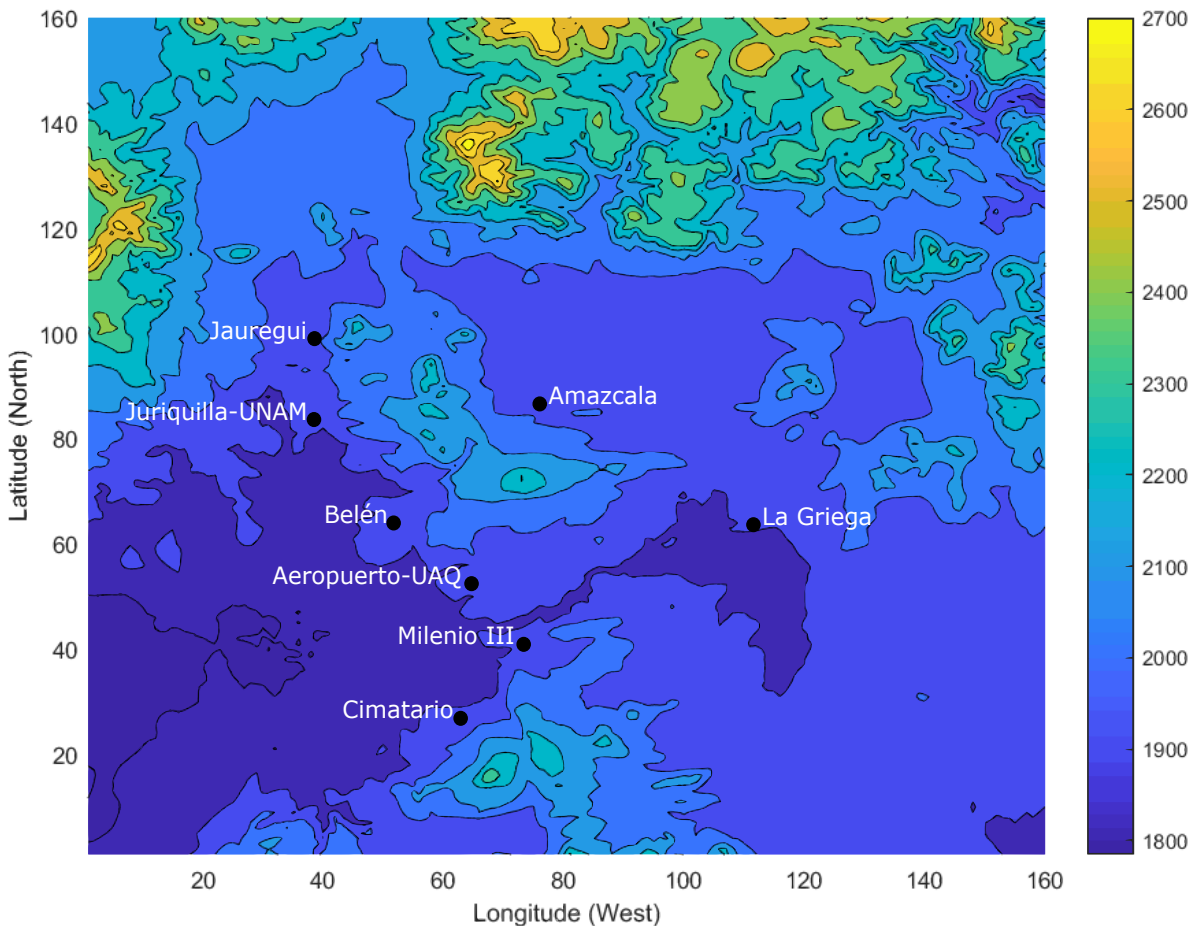


Figura 4.12 Curvas de nivel rellenas que representan los valles de Querétaro, Amazcala y Buenavista.

En la Figura 4.13 se muestran los histogramas de los ocho sitios. Los datos de todos los sitios están muy sesgados. Los valores medios de los sitios varían

desde 0.43 m/s hasta más de 4.32 m/s. el sitio de Aeropuerto, presenta valores promedio durante todo el año por encima de los 4.3 m/s, alcanzando valores máximos de 14.2 m/s. Otro sitio prometedor es Juriquilla, donde se cuenta con una media de durante todo el año de 2.9 m/s, alcanzando valores máximos de 10.7 m/s. En ambos sitios existen posibilidades de montar aerogeneradores de baja potencia.

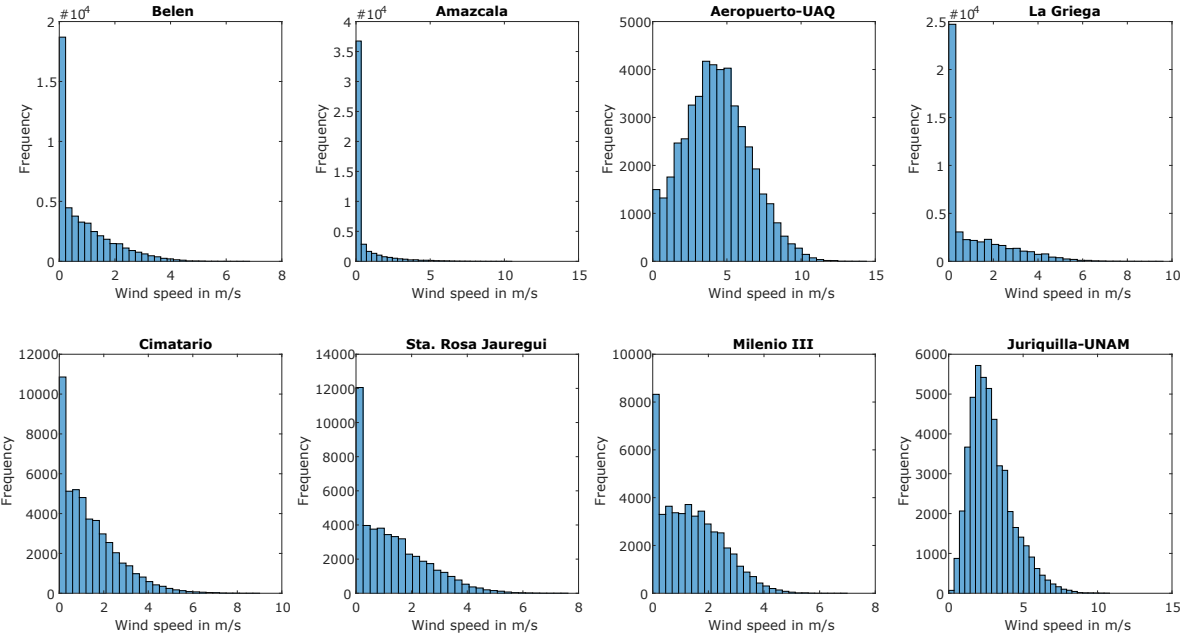


Figura 4.13 Histogramas de la velocidad del viento en m/s para el periodo mayo 2017 a abril 2018.

4.4 Comparación con los trabajos consultados

La presente propuesta de tesis ha sido comparada con trabajos relacionados en la predicción del recurso eólico. En la Tabla 4-4 se muestran las claves de referencia de los trabajos consultados.

Para realizar la comparación se consideraron cuatro características: modelo empleado, método de predicción, horizonte de tiempo de la predicción, localización geográfica y el uso de un método no paramétrico. En la Tabla 4-5 se muestran la comparación entre los trabajos consultados y la propuesta de tesis.

Se puede concluir de acuerdo al análisis de los trabajos consultados lo siguiente:

Tabla 4-4 Claves de referencia de los trabajos consultados.

Clave	Referencia del trabajo
T1	Bayesian adaptive combination of short-term wind speed forecasts from neural network models (Li et al., 2011).
T2	Wind characterization analysis incorporating genetic algorithm: A case study in Taiwan Strait (Liu et al., 2011).
T3	Short-term wind power forecasting: probabilistic and space-time aspects (Tastu et al., 2013).
T4	Estimation of wind farms aggregated power output distributions (Sobolewski and Feijóo, 2013).
T5	Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components (Baran, 2014).
T6	On the wind power potential in the northwest of the Yucatan Peninsula in Mexico (Figueroa-Espinoza et al., 2014).
T7	Spatial dependencies of wind power and interrelations with spot price dynamics (Elberg and Hagspiel, 2015).
T8	Offshore wind energy potential estimation using UPSCALE climate data (Gross and Magar, 2015).
T9	Wind speed modeling using a vector autoregressive process with a time-dependent intercept term (Koivisto et al., 2016).
T10	Wind speed forecasting for wind farms: A method based on support vector regression (Santamaría-Bonfil et al., 2016).
P1	Propuesta de tesis (Cantón, 2019).

Tabla 4-5 Comparación de los trabajos consultados.

Clave	Modelo empleado	Método de predicción	Escala de tiempo (plazo)	Localización geográfica	Método no paramétrico
T1	Bayesiano+RNA	Híbrido	Corto	Dakota del Norte	No aplica
T2	Genético	Estadístico	Corto	Taiwan	Kolmogorov-Smirnov
T3	Copulas	Estadístico	Corto	Dinamarca	No aplica
T4	Estimador Kernel	Estadístico	Largo	No indica	Coef. Spearman
T5	Bayesiano	Estadístico	Muy corto	Hungría	Kolmogorov-Smirnov
T6	Mesoescala	Físico	Corto	Yucatán	No aplica
T7	Copulas	Estadístico	Corto	Alemania	No indica
T8	Mesoescala	Físico	Mediano	Costas de México	No aplica
T9	VAR ¹	Estadístico	Corto	Finlandia	Kolmogorov-Smirnov
T10	SVM	Híbrido	Mediano	Oaxaca	NN-Chaos-Test
P1	Bayesiano	Estadístico	Largo	Querétaro	Coef. Spearman

1. Los modelos empleados para la predicción del recurso eólico son muy diversos, la mayoría de ellos son los basados en técnicas de inteligencia artificial.
2. En los métodos de pronóstico del potencial eólico predominan los métodos estadísticos, seguido de los métodos físicos de tipo mesoescala, y los métodos híbridos donde se combinan características de los métodos anteriores.
3. En el horizonte de predicción más de dos tercios de los trabajos pertenecen a un plazo: muy corto y corto; la escala de tiempo de predicción depende de la aplicación que se le da a la investigación (Soman et al., 2010).
4. La mayoría de los trabajos tienen como localización geográfica lugares cercanos a costas e islas, en donde por su ubicación permite un mejor aprovechamiento del recurso eólico; solo los trabajos de [Li et al. (2011); Baran (2014)] y esta propuesta de tesis buscan predecir el recurso eólico en el interior de un continente donde las cadenas montañosas influyen en el comportamiento del viento, véase el caso del Paso de Chivela en Fernandez et al. (2018).
5. La mitad de los trabajos emplea un método no paramétrico principalmente la prueba Kolmogorov–Smirnov, seguido del coeficiente de correlación de Spearman y la prueba de red neuronal caótica; la otra mitad de los trabajos considera la función de la distribución del viento como paramétrica.

5. CONCLUSIONES Y TRABAJO FUTURO

En este capítulo se hace mención de las conclusiones finales durante el desarrollo de la presente tesis, además se describen algunos trabajos a futuro.

5.1 Conclusiones

Esta investigación presenta una metodología Bayesiana que permite analizar las dependencias geográficas en ocho sitios del estado de Querétaro mediante un proceso estocástico. Esta propuesta emplea el bajo coste computacional del algoritmo EM, así como su capacidad para trabajar con datos incompletos, para estimar la velocidad del viento.

Con base en los grafos de dependencias obtenidos se observa la influencia del relieve en las dependencias entre los sitios estudiados. Por ejemplo, la influencia de las cadenas montañosas en la nula relación entre los sitios que se encuentran en diferentes valles geográficos. Además, el análisis de las series de tiempo permite caracterizar el recurso eólico en el área metropolitana de Querétaro para encontrar los sitios de interés para un estudio más profundo.

El presente trabajo de tesis titulado *Desarrollo de una metodología para predicción del recurso eólico mediante redes Bayesianas* cumple con los objetivos establecidos en este documento, ya que se desarrolló una metodología Bayesiana para predecir el recurso eólico en Querétaro que permite pronosticar en un horizonte de tiempo a largo plazo.

De acuerdo con la revisión del estado del arte este sería el único trabajo que utiliza el algoritmo EM para estimar los parámetros y el coeficiente de Spearman para ver el grado de asociación entre los sitios.

Además, se logró comprobar la hipótesis planteada en este trabajo de tesis,

ya que fue posible analizar las dependencias geográficas a partir de la medición local en diferentes puntos. Cabe mencionar que la metodología planteada en este trabajo puede aplicarse a cualquier problema de muestreo, tal como, análisis de riesgos financieros, entre otros casos.

5.2 Trabajo futuro

El haber desarrollado una metodología para predecir el recurso eólico mediante un proceso Bayesiano, se sugieren los siguientes puntos para realizar trabajos a futuro:

- Extender el análisis de dependencias geográficas a todo el estado de Querétaro, esto permitiría analizar un área más amplia para encontrar zonas que puedan ser explotadas a escala comercial.
- Incluir otras variables meteorológicas como la dirección del viento, temperatura y presión atmosférica; para obtener un mejor análisis del potencial eólico.
- Analizar ciertas características del terreno como la elevación para obtener un análisis de dependencias geográficas mediante un método de predicción híbrido.
- Estudiar la temporalidad mensualmente con la finalidad de distinguir la temporada de mayor o menor aprovechamiento del recurso eólico en un sitio particular.

6. REFERENCIAS

- Alcaraz-Jiménez, J. J. (2006). *Combinación del aprendizaje multiarea y del algoritmo en problemas de clasificación con datos incompletos*. Tesis de Licenciatura, Universidad Politécnica de Cartagena, Cartagena, España. (Citado en página 9.)
- Alvarado, J. Á., Ríos-Moreno, G., Herrera-Ruíz, G., Ventura-Ramos, E., and Trejo-Perea, M. (2018). Statistical analysis for the evaluation of solar and wind resources, measured in querétaro. In *2018 XIV International Engineering Congress (CO-NIIN)*, pages 1–8. IEEE. (Citado en página 3.)
- AMDEE (2019). Asociación mexicana de energía eólica AMDEE. Recuperado 10 abril 2019, de <https://www.amdee.org/mapas-eolicos>. (Citado en páginas 2 y 3.)
- Bar-Yehuda, Z. (2018). Plot google map. Recuperado 10 febrero 2019, de https://la.mathworks.com/matlabcentral/fileexchange/27627-zoharbyplot_google_map. (Citado en página 20.)
- Baran, S. (2014). Probabilistic wind speed forecasting using bayesian model averaging with truncated normal components. *Computational Statistics & Data Analysis*, 75:227–238. (Citado en páginas 10, 31 y 32.)
- Borunda, M., Jaramillo, O., Reyes, A., and Ibargüengoytia, P. H. (2016). Bayesian networks in renewable energy systems: A bibliographical survey. *Renewable and Sustainable Energy Reviews*, 62:32–45. (Citado en página 3.)
- Cadenas, E. and Rivera, W. (2007). Wind speed forecasting in the south coast of Oaxaca, Mexico. *Renewable Energy*, 32(12):2116–2128. (Citado en página 2.)
- Cadenas, E. and Rivera, W. (2010). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model. *Renewable Energy*, 35(12):2732–2738. (Citado en página 2.)

- Cancino-Solórzano, Y., Gutiérrez-Trashorras, A. J., and Xiberta-Bernat, J. (2010). Analytical methods for wind persistence: Their application in assessing the best site for a wind farm in the state of Veracruz, Mexico. *Renewable Energy*, 35(12):2844–2852. (Citado en página 2.)
- Carrasco-Díaz, M., Rivas, D., Orozco-Contreras, M., and Sánchez-Montante, O. (2015). An assessment of wind power potential along the coast of Tamaulipas, northeastern Mexico. *Renewable Energy*, 78:295–305. (Citado en página 2.)
- Carta, J. A. and Velázquez, S. (2011). A new probabilistic method to estimate the long-term wind speed characteristics at a potential wind energy conversion site. *Energy*, 36(5):2671–2685. (Citado en página 8.)
- CEMIE-Eólico (2019). Centro Mexicano de Innovación en Energía Eólica CEMIEE. Recuperado 11 abril 2019, de <http://www.cemieeolico.org.mx/>. (Citado en página 4.)
- Chai, H., Lei, J., and Fang, M. (2017). Estimating Bayesian networks parameters using EM and Gibbs sampling. *Procedia Computer Science*, 111:160–166. (Citado en página 10.)
- CONAGUA (2018a). Actualización de la disponibilidad media anual de agua en el acuífero Valle de Amazcala. Technical report, Comisión Nacional del Agua. Ciudad de México, México. (Citado en página 29.)
- CONAGUA (2018b). Actualización de la disponibilidad media anual de agua en el acuífero Valle de Querétaro. Technical report, Comisión Nacional del Agua. Ciudad de México, México. (Citado en página 29.)
- CONCYTEQ (2003). El sistema de agua en la región de Querétaro de la cuenca Lerma-Chapala. Technical report, Consejo de Ciencia y Tecnología del Estado de Querétaro. Santiago de Querétaro, Qro., México. (Citado en página 29.)
- Cortés, E., Algreto-Badillo, I., and García, V. (2012). Performance analysis of AN-SIS in short term wind speed prediction. *IJCSI International Journal of Computer Science Issues*, 9(5):94–102. (Citado en página 2.)

- De Giorgi, M., Campilongo, S., Ficarella, A., and Congedo, P. (2014). Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (ls-svm) and artificial neural network (ann). *Energies*, 7(8):5251–5272. (Citado en página 8.)
- De Giorgi, M. G., Ficarella, A., and Tarantino, M. (2011). Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. *Energy*, 36(7):3968–3978. (Citado en páginas 7 y 8.)
- De la Torre-Gea, G., Soto-Zarazúa, G. M., Guevara-González, R. G., and Rico-García, E. (2011). Bayesian networks for defining relationships among climate factors. *International Journal of Physical Sciences*, 6(18):4412–4418. (Citado en páginas 3 y 9.)
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22. (Citado en página 9.)
- Diaz, G., Abd-el Motaleb, A. M., and Mier, V. (2012). On the capacity factor of distributed wind generation in droop-regulated microgrids. *IEEE Transactions on power systems*, 28(2):1738–1746. (Citado en página 7.)
- Elberg, C. and Hagspiel, S. (2015). Spatial dependencies of wind power and interrelations with spot price dynamics. *European Journal of Operational Research*, 241(1):260–272. (Citado en páginas 7, 8 y 31.)
- Elliott, D., Schwartz, M., Scott, G., Haymes, S., Heimiller, D., and George, R. (2004). *Atlas de recursos eó del estado de Oaxaca*. National Renewable Energy Laboratory NREL. (Citado en páginas 2 y 3.)
- Feijóo, A., Villanueva, D., Pazos, J. L., and Sobolewski, R. (2011). Simulation of correlated wind speeds: A review. *Renewable and Sustainable Energy Reviews*, 15(6):2826–2832. (Citado en página 10.)
- Fernandez, P., Sosa Tinoco, I., Otero Casal, C., and Miguez-Macho, G. (2018).

- Downslope windstorm study on the isthmus of Tehuantepec using WRF high-resolution simulations. In *AGU Fall Meeting Abstracts*. (Citado en página 32.)
- Figuroa-Espinoza, B., Salles, P., and Zavala-Hidalgo, J. (2014). On the wind power potential in the northwest of the Yucatan Peninsula in Mexico. *Atmósfera*, 27(1):77–89. (Citado en páginas 2, 8 y 31.)
- Foley, A. M., Leahy, P. G., Marvuglia, A., and McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37(1):1–8. (Citado en página 7.)
- Friedman, N. (1998). The Bayesian structural EM algorithm. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 129–138. Morgan Kaufmann Publishers Inc. (Citado en página 9.)
- Google (2019). Google Maps platform. Recuperado 09 febrero 2019, de <https://cloud.google.com/maps-platform/>. (Citado en página 20.)
- Gross, M. S. and Magar, V. (2015). Offshore wind energy potential estimation using upscale climate data. *Energy Science & Engineering*, 3(4):342–359. (Citado en páginas 8 y 31.)
- Grothe, O. and Schnieders, J. (2011). Spatial dependence in wind and optimal wind power allocation: A copula-based analysis. *Energy Policy*, 39(9):4742–4754. (Citado en páginas 7 y 8.)
- Hartley, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14(2):174–194. (Citado en página 9.)
- Hernández-Díaz, T., Vázquez-Cervantes, A., González-Barboza, J.-J., Barriga-Rodríguez, L., Herrera-Navarro, A. M., Baldenegro-Pérez, L.-A., and Jiménez-Hernández, H. (2015). Detecting background and foreground with a laser array system. *Measurement*, 63:195–206. (Citado en página 10.)
- Hernández-Escobedo, Q., Saldaña-Flores, R., Rodríguez-García, E., and Manzano-Agugliaro, F. (2014). Wind energy resource in northern Mexico. *Renewable and Sustainable Energy Reviews*, 32:890–914. (Citado en páginas 1 y 2.)

Iberdrola-México (2018). Iberdrola en México. Recuperado 16 abril 2019, de <https://www.iberdrolarenovablesmexico.com/parque-eolico-santiago-eolico/>. (Citado en página 4.)

IEA (2018). World energy outlook 2018. Technical report, International Energy Agency. (Citado en página 1.)

IEC (2017). *Wind energy generation systems - Part 12-1: Power performance measurements of electricity producing wind turbines*. International Electrotechnical Commission. (Citado en página 17.)

INEGI (2016). Panorama sociodemográfico de querétaro 2015. Technical report, Instituto Nacional de Estadística Geografía e Informática, Aguascalientes, México. (Citado en página 15.)

Jiménez-Hernández, H., González-Barbosa, J.-J., and Garcia-Ramírez, T. (2010). Detecting abnormal vehicular dynamics at intersections based on an unsupervised learning approach and a stochastic model. *Sensors*, 10(8):7576–7601. (Citado en página 17.)

Koivisto, M., Seppänen, J., Mellin, I., Ekström, J., Millar, J., Mammarella, I., Kompula, M., and Lehtonen, M. (2016). Wind speed modeling using a vector autoregressive process with a time-dependent intercept term. *International Journal of Electrical Power & Energy Systems*, 77:91–99. (Citado en páginas 8 y 31.)

Lauritzen, S. L. (1995). The em algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, 19(2):191–201. (Citado en página 9.)

Li, G., Shi, J., and Zhou, J. (2011). Bayesian adaptive combination of short-term wind speed forecasts from neural network models. *Renewable Energy*, 36(1):352–359. (Citado en páginas 2, 8, 10, 31 y 32.)

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*, volume 793. Wiley. (Citado en página 9.)

- Liu, F.-J., Chen, P.-H., Kuo, S.-S., Su, D.-C., Chang, T.-P., Yu, Y.-H., and Lin, T.-C. (2011). Wind characterization analysis incorporating genetic algorithm: A case study in taiwan strait. *Energy*, 36(5):2611–2619. (Citado en páginas 9 y 31.)
- Mackridge, A. and Rowe, P. (2018). *A Practical Approach to Using Statistics in Health Research: From Planning to Reporting*, chapter Spearman Correlation. John Wiley & Sons. (Citado en páginas 11 y 19.)
- Mikati, M., Santos, M., and Armenta, C. (2012). Modelado y simulación de un sistema conjunto de energía solar y eólica para analizar su dependencia de la red eléctrica. *Revista Iberoamericana de Automática e Informática industrial*, 9(3):267–281. (Citado en página 1.)
- Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., Conzelmann, G., et al. (2009). Wind power forecasting: State-of-the-art 2009. Technical report, Argonne National Lab.(ANL), Argonne, IL (United States). (Citado en páginas 8 y 9.)
- Mughal, M. O., Lynch, M., Yu, F., and Sutton, J. (2018). Forecasting and verification of winds in an east african complex terrain using coupled mesoscale-and micro-scale models. *Journal of Wind Engineering and Industrial Aerodynamics*, 176:13–20. (Citado en página 2.)
- Puth, M.-T., Neuhäuser, M., and Ruxton, G. D. (2015). Effective use of spearman's and kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, 102:77–84. (Citado en página 10.)
- Ramsey, P. H. (1989). Critical values for spearman's rank order correlation. *Journal of educational statistics*, 14(3):245–253. (Citado en página 12.)
- Red-CIAQ (2018). Hidroinformática. Recuperado 28 febrero 2019, de <https://hidroinformatica.uaq.mx/>. (Citado en página 15.)
- Reed, E. and Mengshoel, O. J. (2014). Bayesian network parameter learning using em with parameter sharing. In *BMA@ UAI*, pages 48–59. (Citado en página 10.)

- Riggelsen, C. (2006). Learning bayesian networks from incomplete data: An efficient method for generating approximate predictive distributions. In *Proceedings of the 2006 SIAM International Conference on Data Mining*, pages 130–140. SIAM. (Citado en páginas 3, 9 y 10.)
- RUOA (2018). Red universitaria de observaciones atmosféricas. Recuperado 28 febrero 2019, de <https://www.ruoa.unam.mx/>. (Citado en página 15.)
- Santamaría-Bonfil, G., Reyes-Ballesteros, A., and Gershenson, C. (2016). Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy*, 85:790–809. (Citado en páginas 9 y 31.)
- Sawin, J. L., Martinot, E., Sonntag-O'Brien, V., McCrone, A., Roussell, J., Barnes, D., Flavin, C., Mastny, L., Kraft, D., Wang, S., et al. (2018). Renewables 2018 global status report. (Citado en página 1.)
- SENER (2018). *Programas de desarrollo del sistema eléctrico nacional*. Gobierno de México. (Citado en página 2.)
- Sloughter, J. M., Gneiting, T., and Raftery, A. E. (2013). Probabilistic wind vector forecasting using ensembles and bayesian model averaging. *Monthly Weather Review*, 141(6):2107–2119. (Citado en página 2.)
- Sobolewski, R. and Feijóo, A. (2013). Estimation of wind farms aggregated power output distributions. *International Journal of Electrical Power & Energy Systems*, 46:241–249. (Citado en páginas 7 y 31.)
- Soler-Bientz, R., Watson, S., and Infield, D. (2010). Wind characteristics on the yucatán peninsula based on short term data from meteorological stations. *Energy Conversion and Management*, 51(4):754–764. (Citado en página 2.)
- Soman, S. S., Zareipour, H., Malik, O., and Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium 2010*, pages 1–8. IEEE. (Citado en páginas 7 y 32.)

- Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, pages 49–58. (Citado en página 9.)
- Tastu, J., Madsen, H., and Pinson, P. (2013). *Short-term wind power forecasting: probabilistic and space-time aspects*. PhD thesis, DTU, Lyngby, Denmark. (Citado en páginas 7, 10 y 31.)
- Tuszynski, J. (2017). `getelevations` function. Recuperado 10 febrero 2019, de <https://www.mathworks.com/matlabcentral/fileexchange/50704-getelevations-latitude-longitude-varargin>. (Citado en página 20.)
- Weaver, K. F., Morales, V. C., Dunn, S. L., Godde, K., and Weaver, P. F. (2017). *An introduction to statistical analysis in research: with applications in the biological and life sciences*, chapter Pearson's and Spearman's Correlation. John Wiley & Sons. (Citado en páginas 10 y 11.)
- Woodbury, M. A. (1970). A missing information principle: theory and applications. Technical report, Duke University Medical Center Durham United States. (Citado en página 9.)