



Universidad Nacional Autónoma de México

Instituto de Energías Renovables
Instituto de Ingeniería

Estados de viento para estimación del potencial eólico

TESIS

Que opta por el título de
Ingeniero en Energías Renovables

PRESENTA

Pedro Andrés Sánchez Pérez

DIRECTOR DE TESIS

Dr. Miguel Robles Pérez



Temixco, Morelos a Mayo de 2016



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

Comité tutorial:

Presidente:

Dr. Eduardo Ramos Mora

Secretario:

Dr. Miguel Robles Pérez

Vocal:

Dr. Antonio del Rio Portilla

Vocal:

Dr. Rafael Campos Amezcua

Vocal:

Dr. Oscar Jaramillo Salgado

ÍNDICE GENERAL

Índice de Figuras	iv
Índice de Tablas	v
Nomenclatura	vii
Resumen	ix
Agradecimientos	xi
1. Introducción	1
1.1. Antecedentes	2
1.2. Motivación y metas	2
1.3. Estructura del trabajo	4
2. Conceptos básicos	5
2.1. Fundamentos de energía eólica	6
2.1.1. Recurso eólico	9
2.1.2. Eficiencia de turbinas eólicas	11
2.1.3. Control y Sistema de orientación de <i>Yaw</i> en aerogeneradores	12
2.1.4. Fallas en turbinas eólicas	14
2.2. Procesos estocásticos	15
2.2.1. Cadenas de Markov o procesos Markovianos	17
2.2.2. Cadenas de Markov Continuas	20
2.3. Agrupamiento	22
2.3.1. Algoritmos de agrupamiento	22
2.3.2. k-means	26
2.4. Estados de viento	27
2.5. Conclusiones	31
3. Metodología	33
3.1. Casos de estudio	34
3.2. Desalineamiento	36
3.3. Cálculo de la energía	38
3.4. Cálculo de LCOE	42
4. Resultados	43
4.1. Resultado agrupamiento de estados	44
4.2. Resultado estimación de energía	44
4.3. Resultado de LCOE	44

4.4. Simulador de viento	46
4.5. Discusión de resultados	47
5. Conclusiones generales	51
Bibliografía	55
A. Anexo	59
A.1. Costos en turbinas eólicas	60
A.2. Método de potencia	61
A.3. Series de tiempo	62
A.4. Algoritmos de seguimiento	63
A.5. Distribución normal bivariada para cada estado	64
A.6. Matriz de transición de estados	65
A.7. Tiempos característicos	66
A.8. Matriz q	66
A.9. Simulación de energía y costos	67
A.10. Artículo publicado	70

ÍNDICE DE FIGURAS

2.1. Desglose de componentes principales de un aerogenerador. Tomada parcialmente de [1]	7
2.2. Coeficiente de potencia en función de la relación de velocidades	8
2.3. Curva de potencia típica del aerogenerador.	9
2.4. Histograma de frecuencias de datos de velocidad de viento	10
2.5. Eficiencia de conversión de cada componente de un aerogenerador. Tomada en su totalidad de [2].	11
2.6. Concepto de desalineamiento del aerogenerador (γ) debido al cambio de dirección de viento incidente (U) (a), y diagrama del sistema de orientación (b). Extraído en su totalidad de [3].	13
2.7. Variación del C_p en función de γ y a . Extraído en su totalidad de [3]. . .	14
2.8. Estadísticas de fallas de campos eólicos en USA. Tomada en su totalidad de [4]	15
2.9. Tiempo de espera en un banco	16
2.10. Diagrama de un proceso markoviano	18
2.11. Cadena de <i>Markov</i> a) irreducible y b) aperiódica.	20
2.12. Ejemplo de aplicación de algoritmos de agrupamiento en sepas de pétalos. Base de datos disponible en [5].	23
2.13. Tipos de algoritmos de <i>clustering</i> . Extraído en su totalidad de	23
2.14. Algoritmos de agrupamiento. Base de datos disponible en [5]	25
2.15. Histograma de dirección de los datos obtenidos por la estación meteorológica en la Rumorosa, Baja California.	27
2.16. Rosa de los vientos para la localidad de La ventosa, Oaxaca. Las unidades están en m s^{-1}	28
2.17. Histograma 2-D de dos variables aleatorias X y Y	29
2.18. Histograma 2-D de datos de velocidad y dirección de viento.	30
3.1. Histogramas de frecuencias de velocidad y dirección de los casos de estudios.	35
3.2. Sabana de potencia del aerogenerador.	36
3.3. Plano vectorial v_x - v_y , representando la formación de cúmulos para los casos de estudio. a) Caso de estudio 1, b) caso de estudio 2	37
3.4. Histogramas 3D. a) Caso de estudio 1, b) caso de estudio 2	38
3.5. Curva de potencia y C_p del aerogenerador de 1.8 MW de estudio.	39
4.1. Agrupación en estados de viento del espacio de velocidades para el caso de estudio 1 (a) y 2 (b), utilizando el algoritmo k -means.	45
4.2. Cadena de <i>Markov</i> para casos de estudios	47
4.3. Diagrama de <i>Markov</i> para los casos de estudio	48

4.4. Concepto de apertura de estado α_p para el estado cuatro del caso de estudio dos.	49
4.5. Regiones elípticas donde las componentes normales de cada estado cubre el 90% de los datos asociados a el para el caso de estudio uno (a) y dos (b). Las elipses se colorearon utilizando la potencia relativa de cada estado p_i^* con la escala mostrada en la paleta de colores.	49
A.1. Costo nivelado de Energía (LCOE) de cada tecnología disponible en el mercado. Extraído en su totalidad de [6]	60
A.2. Serie de tiempo de velocidad y dirección para caso de estudio 1	62
A.3. Serie de tiempo de velocidad y dirección para el caso de estudio 2.	63

ÍNDICE DE TABLAS

3.1. Resultados de ajustes para las distribuciones de velocidad de los casos de estudio. Para el caso de la ventosa se utilizó una función <i>Weibull & Weibull</i> [7] y para Baja California una <i>Weibull</i> [8]	35
3.2. Energía anual producida en GWh para las dos localidades utilizando el método de estimación de energía tradicional. Para el caso de Baja California se utilizó una $f(v)$ tipo <i>Weibull</i> y para Oaxaca una <i>Weibull</i> bi-variada [7]	39
3.3. Métodos para la estimación de potencial eólico.	40
4.1. Producción anual de energía (AEP [GWh]) y factor de planta (%) utilizando diferentes métodos.	46
4.2. Costo nivelado de energía utilizando diferentes métodos de estimación de energía	46
A.1. Costos asociados por cada componentes de un aerogenerador de 5 MW con USD del 2009. Adaptado de [9]	61
A.2. Parámetros del ajuste utilizando normales bi-variadas para el caso de estudio 1	64
A.3. Parámetros del ajuste utilizando normales bi-variadas para el caso de estudio 2	64
A.4. Matriz de transición de estados de la cadena de <i>Markov</i> obtenida de la información del caso 1.	65
A.5. Matriz de transición de estados de la cadena de <i>Markov</i> obtenida de la información del caso 2.	65
A.6. Tiempo de estadía total t_r y tiempo medio de vida t_m de las cadenas de <i>Markov</i> para los dos casos de estudio.	66
A.7. Matriz q de la cadena de <i>Markov</i> para los datos del caso 1	66
A.8. Matriz q de la cadena de <i>Markov</i> para los datos del caso 2	66
A.9. Resultados de la simulación de un año para el caso de estudio 1.	68
A.10. Resultados de la simulación de un año para el caso de estudio 2.	69

NOMENCLATURA

Constantes

ρ Densidad del aire

Abreviaciones

AEP Energía producida por año

FP Factor de planta

HAWT Turbina de eje horizontal

LCOE Costo nivelado de la energía

SEN Sistema eléctrico nacional

VAWT Turbina de eje vertical

Variables

γ Ángulo de desalineamiento del rotor

η Eficiencia del aerogenerador

λ Parámetro de forma de la distribución *Weibull*

μ_i Centroide del cluster

σ_{ij} Covarianza

Σ Matriz de covarianza

A Área del aerogenerador

a Coeficiente de inducción axial

F_t Costo del combustible

H_i Tiempo de vida medio de estado

I_t Gastos de capital

k Parametro de escala de la distribución *Weibull*

M_t Gastos de operación y mantenimiento

n Tiempo de vida del proyecto

P Potencia eléctrica

p_{ij}	Matriz de transición de probabilidad
q	Matriz de flujos
r	Costo medio ponderado de capital
u	Vector de distribución de probabilidad
v	Velocidad de viento
V_i	Vector de velocidad
C_p	Coefficiente de potencia

RESUMEN

En este trabajo se desarrolló una herramienta novedosa para el análisis de viento cualquier sitio. Se introduce el concepto de estado de viento para clasificar datos del recurso eólico, como una herramienta que permite asociar características estadísticas o físicas a cúmulos de datos, para poder identificar, por ejemplo, patrones, estados con mayor energía o estados transitorios. Esta herramienta puede ser de utilidad para mejorar el sistema de orientación de un aerogenerador. También se llevó a cabo el análisis del viento de dos casos de estudio con interés eólico. Se realizó una simulación de un aerogenerador bajo condiciones de operación y se implementó un algoritmo de seguimiento de estados para optimizar el uso del sistema de orientación. Se compararon varias metodologías de estimación de energía para resaltar la contribución del concepto de estados de viento. Utilizando la metodología de estados de viento, en los casos de estudio, se logró proponer cambios en el algoritmo de seguimiento que pueden generar un incremento en el FP de 10% y 20% para el caso de estudio uno y dos respectivamente en comparación con el método tradicional.

AGRADECIMIENTOS

Para mí siempre ha sido importante agradecer a las personas que me han ayudado en alguna etapa de mi vida. Hoy, quiero agradecer a todas las personas que estuvieron caminando a mi lado en este trayecto universitario. Algo que parecía muy lejano, llegó a un trabajo de tesis completo.

Durante mi estancia en la universidad he vivido junto a mis compañeros situaciones difíciles. Los primeros, siempre somos el experimento. Hoy, quiero decirles a mis compañeros, a familia y a mis amigos que este experimento fue exitoso y lo seguirá siendo. Se me dio la oportunidad de caminar a hombros de gigantes. Maestros inolvidables y muy sabios. Esto me motiva para superarme como persona y dar siempre lo mejor de mí. Todas las personas que creyeron en mí, quiero decirles que lo logré y nunca hubiera sido posible sin ayuda.

Primero que nada, me gustaría agradecer con todo mi corazón al Dr. Ignacio (Q.E.P.D) por darme la oportunidad de vivir mi carrera en el IER. Sin él, esta meta que cumplo hoy no hubiera sido posible. Gracias a personas como él tuve la oportunidad de conocer a toda la comunidad del IER y a personas increíbles como el Dr. Miguel Robles. Recuerdo la primera vez que lo conocí, hablando de Star Wars y \LaTeX . Me aconsejó como un Jedi aconsejaría a su Padawan y siempre me guio por el camino de la luz. Hoy en día no solo es un modelo a seguir, sino un gran amigo.

Estoy muy agradecido con todas las personas que estuvieron en mi camino en el IER, como el Dr. Mariano López y el Dr. Antonio del Río, por siempre motivarme a superarme en cada momento. También quiero agradecer a todo el grupo de trabajo de eólica del Instituto: Dr. Eduardo Ramos, Dr. Rafael Campos, Dr. Oscar Jaramillo, Dr. Maximiliano Valdez, Dr. Osvaldo Hernández, Dr. Héctor García. Les agradezco por tantas buenas pláticas, consejos y agradables momentos en el IER.

Al Dr. Octavio García, gracias por mostrarse siempre atento a las necesidades de los alumnos y además por darme la oportunidad de participar en sus proyectos de investigación. Un agradecimiento muy especial a Maribel Fernández, por toda la paciencia que tuvo con la primera generación y por siempre estar apoyando incondicionalmente a la LIER. Gracias por brindarnos tu amistad y siempre luchar por nosotros.

También estoy muy agradecido con todos los integrantes de LANEFV, Aarón Sánchez, Dalia Martínez, José Ortega, Rocío Santos, Amílcar Reyes, Carlos Percino, Edgar Solares, Manuel Cervantes, Osvaldo Solorio y los nuevos integrantes. Creo que este grupo está motivado por un objetivo común y cada día crecemos mucho más. Estoy muy honrado de poder pertenecer a este grupo y poder trabajar a su lado. Cada uno de ustedes ha enriquecido mi vida en el IER y espero que siempre se cumplan todos sus sueños.

Este trabajo nunca habría sido posible sin el apoyo incondicional de mis padres Aarón Sánchez y María Azucena. Ellos siempre han sido mi ejemplo a seguir, y hoy más que

nunca, quiero agradecerles por todo lo que han dado por mi. Siempre tendrán mi eterna gratitud y amor. A mis hermanos Mishael Sánchez y Jazmin Sánchez, que aunque en este momento caminamos por sendas diferentes, quiero decirles que siempre estaré para ustedes.

Finalmente, este trabajo nunca hubiera sido posible sin la ayuda de la hermosa Dalia. Gracias por tantos momentos felices a tu lado, por tanto amor que me has brindado. Juntos hemos logrado cosas magníficas, y cuando vuelva a leer este texto estoy seguro que habremos logrado muchas más. Siempre tendrás todo mi ser y amor incondicional. ¡Gracias por compartir esta vida conmigo!

Pedro Andrés Sánchez Pérez

Mayo 2016

A mi familia

1

INTRODUCCIÓN

En el primer capítulo se describe el panorama general de la energía eólica en el mundo y en México. También se incluye la motivación y metas de este trabajo así como la estructura del mismo.

Contenido del capítulo

1.1. Antecedentes	2
1.2. Motivación y metas	2
1.3. Estructura del trabajo	4

1.1 Antecedentes

Actualmente se puede afirmar que hay un consenso mundial en que la mejor opciones para sustituir a los hidrocarburos como fuente primaria de energía, proviene de las fuentes renovables. Muchos países están haciendo inversiones en tecnologías renovables para cumplir las metas y acuerdos para disminuir las emisiones de gases de efecto invernadero (GEI) [10]. Esto ha propiciado un gran potencial de desarrollo tecnológico en varias de las energías renovables. En el año 2015, la contribución mundial de las energías renovables al consumo final de energía fue de 19.1%. De este porcentaje, solo el 10.1% es de las llamadas energías renovables modernas es decir: la solar, eólica, bioenergía, geotermia e hidroenergía [10]. En México en el 2014, se tenía una capacidad instalada de energías renovables de 16 295 MW [11] y esta potencia representa un porcentaje de 7.6% del total de la capacidad instalada, del cual solamente el 1.3% se atribuye a las energías renovables modernas sin incluir la capacidad instalada de gran hidráulica.

La energía eólica es de las tecnologías renovables con mayor contribución a la generación de energía eléctrica en México [11], con un potencial teórico de 40 000 MW, es considerada una energía limpia, porque durante la transformación de energía no existe ningún tipo de emisión contaminante [12]. Se utiliza en diferentes aplicaciones, por ejemplo, producir electricidad a través de turbinas, generar trabajo en los molinos o movimiento para los barcos con velas. Para la generación de energía eléctrica se utilizan aerogeneradores electromagnéticos que aprovechan la energía cinética del viento para transformarla en electricidad. Los aerogeneradores se clasifican de acuerdo a los sitios de instalación en *on-shore* (para las turbinas en tierra) y *off-shore* (para los aerogeneradores en el océano) los cuales funcionan interconectados a la red eléctrica [6]. Existen dos tipos de diseños de aerogenerador, de eje horizontal (HAWT, por sus siglas en inglés) y eje vertical (VAWT, por sus siglas en inglés). Los diseños de cada turbina que existen hoy en día tanto para *on-shore* como *off-shore* son muy eficientes al transformar la energía cinética del viento en energía eléctrica, sin embargo aún se sigue mejorando y probando nueva tecnología [13].

1.2 Motivación y metas

Para poder instalar aerogeneradores en una localidad se requiere conocer previamente el recurso eólico. Utilizando mediciones reales o simulaciones de la velocidad y dirección de viento se puede estimar la cantidad de energía que entregará un aerogenerador. La manera más común de hacer esto es utilizar un ajuste a la distribución de probabilidad del viento y multiplicarlo por la curva de potencia del aerogenerador. Lo anterior da como

resultado una aproximación a la energía que se generará el aerogenerador en un periodo de tiempo. Esta metodología es utilizada comúnmente por desarrolladores de proyectos eólicos sin embargo, omite la variable de dirección del viento, pues asume que los aerogeneradores están automáticamente alineados con la dirección del viento. Sin embargo, con frecuencia pueden existir situaciones en que si no se considera podría llevar a estimaciones erróneas que se reflejen en proyectos que no cumplen con las expectativas de diseño. Para proyectos eólicos que superen 5 MW en capacidad instalada hacer una estimación incorrecta del recurso podría representar pérdidas de dinero importantes. Aunque el aerogenerador cuente con un sistema de seguimiento para capturar una mayor cantidad de energía, el efecto del desalineamiento del aerogenerador no está claro al utilizar la metodología estándar.

Usualmente se considera que la forma de la distribución de probabilidades de la velocidad que se obtiene de los datos de viento no cambia en todo el año [2]. Sin embargo, la distribución de probabilidad de velocidades que vienen de un sector angular fijo si depende de la dirección. Es decir, si se divide la dirección de viento en sectores circulares acotados, la distribución de probabilidad de las velocidades en esa región puede ser igual o totalmente distinta a la distribución de probabilidad de todo el año e incluso puede ser distinta de sector a sector.

Hoy en día, es una necesidad desarrollar herramientas útiles que ayuden a estimar de una manera correcta el recurso eólico en una localidad. Este tipo de herramientas tienen que ser fáciles para poder llevarse a cabo previo al proyecto, pero con información lo suficientemente rica para poder representar condiciones reales en un sitio. Esta necesidad nos motiva al desarrollo de una herramienta novedosa para la estimación del recurso eólico que tome en cuenta la dirección del viento. Para desarrollarla utilizamos diferentes herramientas matemáticas para su aplicación como lo son los modelos estocásticos y algoritmos de agrupamiento. Esta propuesta también incluye la definición del concepto más importante de este trabajo y que será de gran utilidad para simular las condiciones del viento en una localidad. Este concepto es el de *estado de viento* y será explicado a profundidad en el capítulo 2.

Este trabajo tiene como meta principal proporcionar una herramienta que ayude a estimar el recurso eólico de una manera que incluya los cambios de dirección de viento. Se presentan varios casos de las aplicaciones de esta herramienta, las cuales incluyen:

- desarrollar un simulador de datos de viento,
- desarrollar un algoritmo de seguimiento optimizado,
- evaluar el desgaste de las componentes involucradas en el seguimiento,
- clasificar regiones de viento.

1.3 Estructura del trabajo

Este trabajo está constituido por cinco capítulos presentado de la siguiente manera. En el segundo capítulo se profundiza los fundamentos de la energía eólica, así como en los procesos estocásticos y su aplicación en la ingeniería. Se define el concepto de estados de viento y cómo los algoritmos de agrupamiento pueden ser utilizados para discretizar estados. En el tercer capítulo se establece la metodología que se utilizó para el desarrollo del trabajo. También las consideraciones necesarias para la interpretación de los resultados. En el cuarto capítulo se muestran todos los resultados obtenidos de los casos de estudio al aplicar la metodología del capítulo anterior. En el último capítulo se concluye el trabajo y discute los resultados obtenidos así como el trabajo a futuro.

2

CONCEPTOS BÁSICOS

En este capítulo se describen los fundamentos de la energía eólica, así como los procesos estocásticos y su aplicación en la ingeniería. Se define el concepto de estados de viento y cómo los algoritmos de agrupamiento pueden ser utilizados para definirlos y descubrirlos.

Contenido del capítulo

2.1. Fundamentos de energía eólica	6
2.1.1. Recurso eólico	9
2.1.2. Eficiencia de turbinas eólicas	11
2.1.3. Control y Sistema de orientación de <i>Yaw</i> en aerogeneradores	12
2.1.4. Fallas en turbinas eólicas	14
2.2. Procesos estocásticos	15
2.2.1. Cadenas de Markov o procesos Markovianos	17
2.2.2. Cadenas de Markov Continuas	20
2.3. Agrupamiento	22
2.3.1. Algoritmos de agrupamiento	22
2.3.2. k-means	26
2.4. Estados de viento	27
2.5. Conclusiones	31

2.1 Fundamentos de energía eólica

El viento es una fuente de energía renovable disponible en todo el mundo y constituye una alternativa para la generación de energía eléctrica con combustibles fósiles. La energía cinética del viento se puede convertir en mecánica, que a su vez se puede transformar en energía eléctrica. Los dispositivos que realizan esta conversión se llaman aerogeneradores, los cuales se clasifican como se mencionó en el capítulo anterior de acuerdo a su diseño en eje horizontal y eje vertical. Aunque ambas tecnologías se encuentran disponibles comercialmente, la más usada es la de eje horizontal [14] y es la que se enfocará este trabajo.

Los aerogeneradores de eje horizontal se llaman así por la orientación del rotor, el cual es el primer elemento del aerogenerador y es el dispositivo que transforma la energía cinética del viento en energía mecánica a través de las palas o aspas. El rotor se encuentra suspendido en una góndola que contiene los elementos mecánicos y eléctricos para poder realizar la conversión. Cuando el rotor gira por acción del viento —a velocidad variable— el movimiento que genera se acopla directamente al siguiente elemento que es la caja de engranes, esta comienza a girar y aumenta el número de revoluciones por minuto (RPM) para poder acoplarse al generador eléctrico, que transforma movimiento y el torque mecánico en energía eléctrica. Finalmente el generador se conecta a una toma de red eléctrica para poder ser distribuida. La Figura 2.1 muestra un esquema simplificado de los componentes de un aerogenerador de eje horizontal.

Un campo eólico consta de varios aerogeneradores distribuidos en un terreno, líneas de distribución para transportar la energía hacia una central y líneas de transmisión que, para el caso de México inyectan energía eléctrica al Sistema Eléctrico Nacional (SEN). En los campos eólicos de gran capacidad, se instalan aerogeneradores con potencias mínimas de 500 kW, que usualmente tienen una altura de al 30 m sobre nivel de terreno.

Uno de los primeros pasos para poder llevar a cabo la implementación de un campo eólico es diseñar el aerogenerador. El objetivo principal del diseño de aerogeneradores es buscar un aumento en la potencia de salida, y existen cuatro formas principales de hacerlo: 1) colocar el aerogenerador en sitios con vientos mayores, 2) mover el aerogenerador en lugares con mayor altura (montañas) donde los vientos sean mayores que a nivel de suelo, 3) incrementar el área de barrido de los aerogeneradores, 4) incrementar o reducir el tamaño del generador. Todas estas formas de aumentar la potencia dependerán de las condiciones del lugar. A través de los años, se ha observado que la tendencia en el diseño de los aerogeneradores es aumentar el diámetro o área de barrido para incrementar su potencia [6]. Diseñar un aerogenerador abarca desde definir el diámetro del rotor, el número de aspas, la torsión de las mismas, los perfiles aerodinámicos, selección del generador y optimización en el sistema de control, entre muchas otras cosas. Para diseñar aerogeneradores existen múltiples técnicas que utilizan uno o varios modelos de

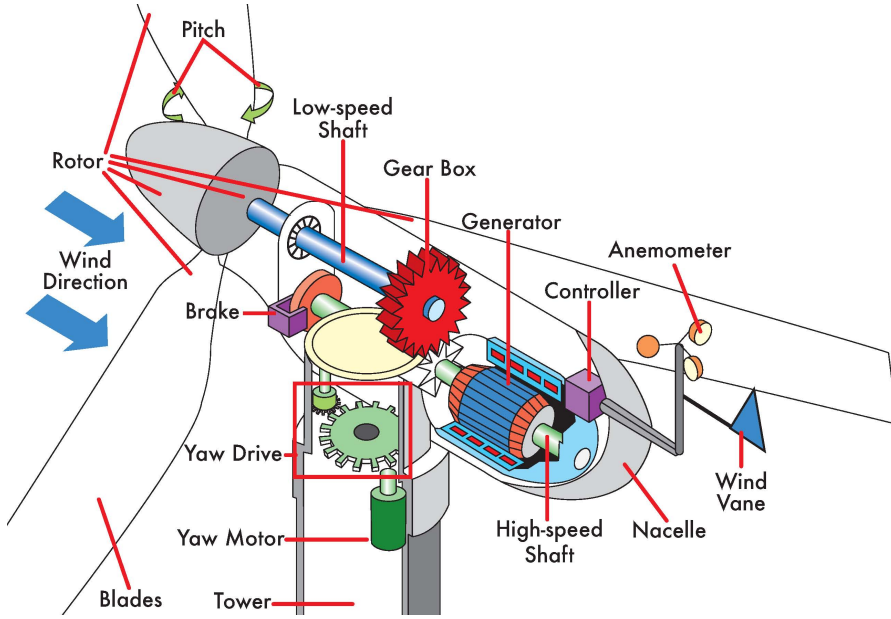


Figura 2.1: Desglose de componentes principales de un aerogenerador. Tomada parcialmente de [1]

optimización. Muchos de los cuales están muy bien descritas en los libros clásicos de aerodinámica y energía eólica como los expuestos en [2, 15, 16].

Para llevar a cabo el diseño del aerogenerador se tienen que considerar varios parámetros los cuales dictaminarán el funcionamiento del aerogenerador en operación. Fingersh et al. [17] mencionan algunos de los parámetros importantes para el diseño, los cuales serán abordados en este trabajo¹. Uno de los primeros conceptos que se deben definir y utilizar para el diseño es el coeficiente de potencia (C_p). El C_p es la razón de proporcionalidad de la cantidad de potencia extraíble del aerogenerador con la potencia cinética del viento. Es un valor adimensional propio de cada aerogenerador. Comúnmente, se deduce el C_p a partir de la aproximación de disco actuador, explicado en [3] y tiene la siguiente expresión:

$$C_p = \frac{P}{P_o} = \frac{\frac{1}{4}\rho A(v_1^2 - v_2^2)(v_1 - v_2)}{\frac{1}{2}\rho A v_1^3}, \quad (2.1)$$

siendo P la potencia mecánica extraíble del disco actuador, P_o la potencia del viento, A el área del disco actuador, v_1 la velocidad del viento antes del disco actuador, v_2 la velocidad después del disco actuador y siendo ρ la densidad del aire (1.205 kg m^{-3} @ 25°C). Reacomodando la ecuación (2.1) se puede expresar el C_p en término de la relación de velocidades como se muestra en la siguiente ecuación:

$$C_p = \frac{P}{P_o} = \frac{1}{2} \left| 1 - \left(\frac{v_2}{v_1} \right)^2 \right| \left| 1 + \frac{v_2}{v_1} \right|. \quad (2.2)$$

¹Referirse a [17, 2, 3] si se requiere conocer a profundidad los parámetros de diseño

La ecuación (2.2) tiene un máximo cuando la relación de velocidades es $\frac{1}{3}$ y para ese valor el $C_{p,max} = 0.593$, el cual se conoce como límite de Betz y se puede observar en la Figura 2.2. La tendencia del diseño es alcanzar el valor máximo de C_p ya que esto implica una conversión de energía mayor.

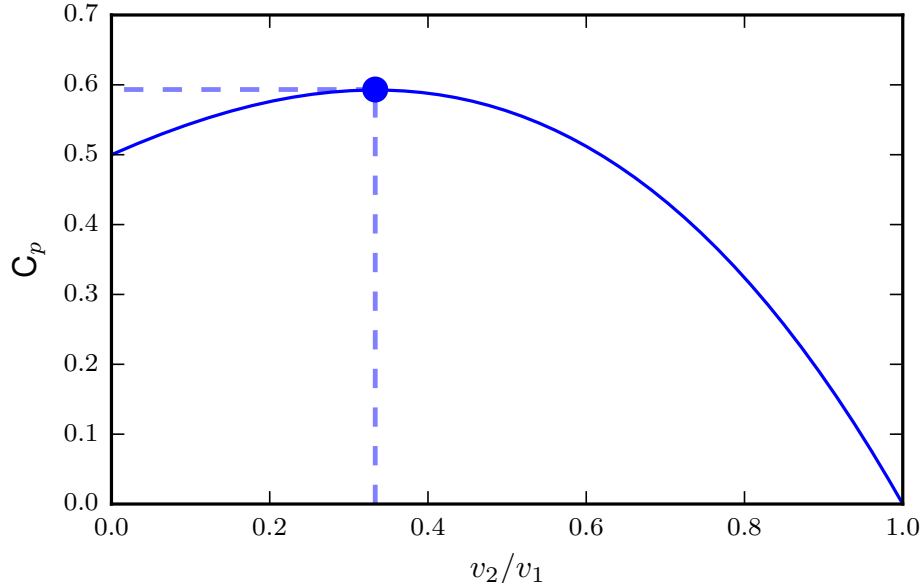


Figura 2.2: Coeficiente de potencia (C_p) en función de la relación de velocidades. Se observa el valor máximo de C_p de 0.593 para turbinas de eje horizontal.

Considerando que $C_{p,max}$ es la cantidad máxima de energía que se puede extraer del viento, la potencia que produce el aerogenerador al valor específico de $C_{p,max}$ será:

$$P = \frac{1}{2} \rho A C_{p,max} v^3. \quad (2.3)$$

La ecuación (2.3) es la relación clave en la energía eólica, porque establece la potencia teórica máxima que produciría un aerogenerador bajo cierta velocidad del viento.

Una vez que se diseñó el aerogenerador para un C_p se obtiene la curva de potencia en función de la velocidad como se muestra en la Figura 2.3. La curva de potencia es característica de cada aerogenerador y se proporciona por el fabricante. Como se observa en la Figura 2.3 existen cuatro regiones en la curva de potencia. La región uno que abarca el intervalo de cero hasta la velocidad de arranque $0 - v_a$, la región dos, la cuál abarca el intervalo de v_a a v_n siendo v_n la velocidad nominal de el aerogenerador, la región tres de producción nominal (P_n) que incluye desde la velocidad nominal hasta la velocidad de corte v_c y la región cuatro que representa las velocidades mayores o iguales a la velocidad de corte, aunque usualmente a velocidades mayores de 25 m s^{-1} el aerogenerador frena por protección y deja de producir potencia.

Para que un aerogenerador produzca potencia, la velocidad del viento siempre tiene que ser mayor a v_a . Por lo tanto, en un campo eólico es deseable que la velocidad promedio

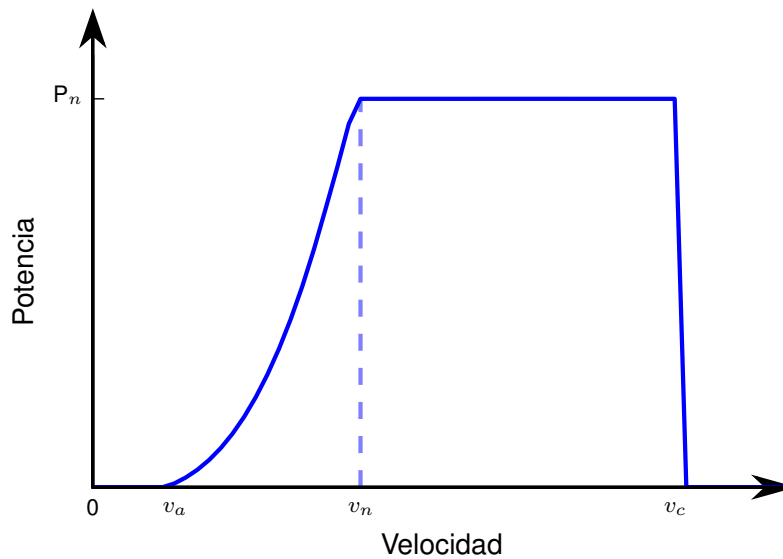


Figura 2.3: Curva de potencia típica del aerogenerador.

del viento en el sitio sea mayor a v_a . Dado que los aerogeneradores se diseñan a un C_p el diseño del aerogenerador se vuelve un tanto específico en vez de genérico. Esto nos lleva a la conclusión que no existe ningún aerogenerador que funcione en operación nominal en cualquier sitio.

2.1.1 Recurso eólico

La evaluación del recurso es tan importante como lo es el diseño, ya que estimar la producción anual de energía (AEP, por sus siglas en inglés) del aerogenerador es clave para calcular la viabilidad de proyectos eólicos. También da un panorama general de la disponibilidad del recurso en un área específica. Existen muchos artículos que hablan acerca de métodos para estimar el recurso como los expuestos en [18]. Algunos de ellos son medición en sitio, predicción a través de una medición, bases de datos globales², métodos numéricos, etc.

Una forma de estimar el recurso eólico es con mediciones en campo. Utilizando sensores de viento se obtiene la serie temporal de la velocidad en el periodo de tiempo que se lleven acabo las mediciones. Con los datos de la serie temporal se puede analizar el comportamiento de la velocidad.

Para comparar el recurso eólico en diferentes sitios, surgió la necesidad de estandarizar las mediciones que se hacen en campo. Las mediciones que se realizan tienen que seguir la norma internacional IEC 61400-12-1 [21]. La cual establece los parámetros mínimos que debe cumplir las mediciones de todas las variables meteorológicas (en particular velocidad y dirección) para hacer la estimación del recurso. La norma establece que los valores de

²Algunos mapas eólicos disponibles son los expuestos en [19, 20].

velocidad de viento deberán ser guardados en promedios de 10 min con una frecuencia de muestreo de 1 Hz [22]. Usualmente, este formato es el que se encuentra disponible en bases de datos, y con base en estas se puede estimar el recurso y diseñar el aerogenerador para el sitio.

Utilizando la base de datos promediadas se puede graficar un histograma de frecuencias de velocidad como se observa en la Figura 2.4. El histograma dará como resultado la frecuencia o probabilidad de aparición de una velocidad en particular y es una herramienta utilizada en la estimación de energía. En muchos sitios, se ha observado que el histograma de velocidad se puede describir utilizando una distribución de probabilidad tipo Weibull [23] y es lo más utilizado hoy en día. La distribución Weibull se define de la siguiente manera:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp -\left(x/\lambda\right)^k & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad (2.4)$$

siendo k el parámetro de forma y λ el parámetro de escala. Utilizando k y λ se puede optimizar una turbina para que trabaje bajo un viento característico en un año y estimar la producción de energía como lo sugiere la norma IEC 61400-12 [21].

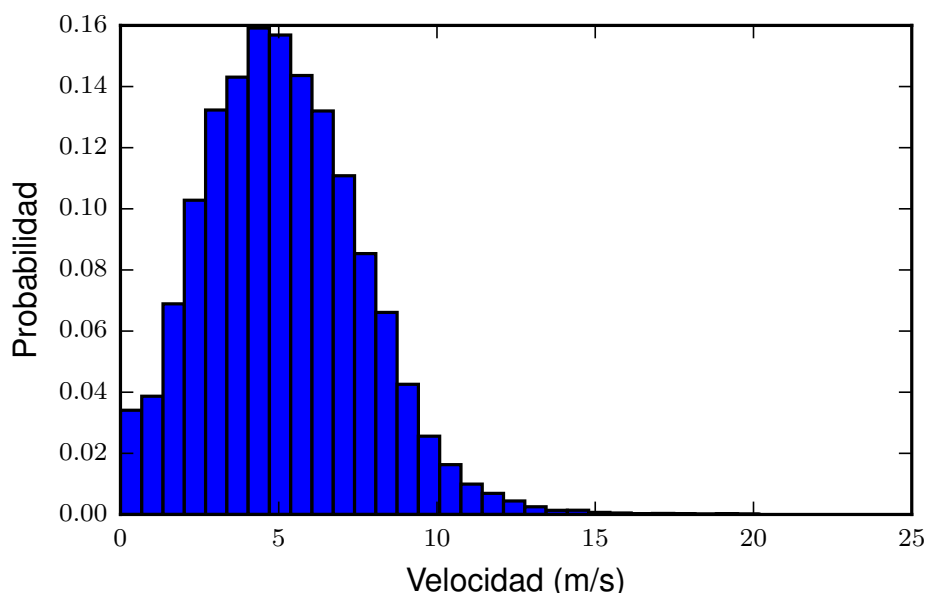


Figura 2.4: Histograma de frecuencias de datos de velocidad de viento

2.1.2 Eficiencia de turbinas eólicas

Para evaluar la energía anual producida, se tienen que incluir todas las pérdidas asociadas a la transformación de los componentes del aerogenerador. Cada componente tiene una eficiencia asociada. En la Figura 2.5 se observa un ejemplo gráfico la eficiencia desglosada de cada elemento que conforman un aerogenerador. Para conocer la eficiencia total de conversión del aerogenerador mostrado en la Figura 2.5 basta con hacer el cociente de la energía de entrada (3045 kW) entre la energía de salida (1176 kW) dando un valor de 38%. Este valor es el orden de magnitud común en todos los aerogeneradores de eje horizontal.

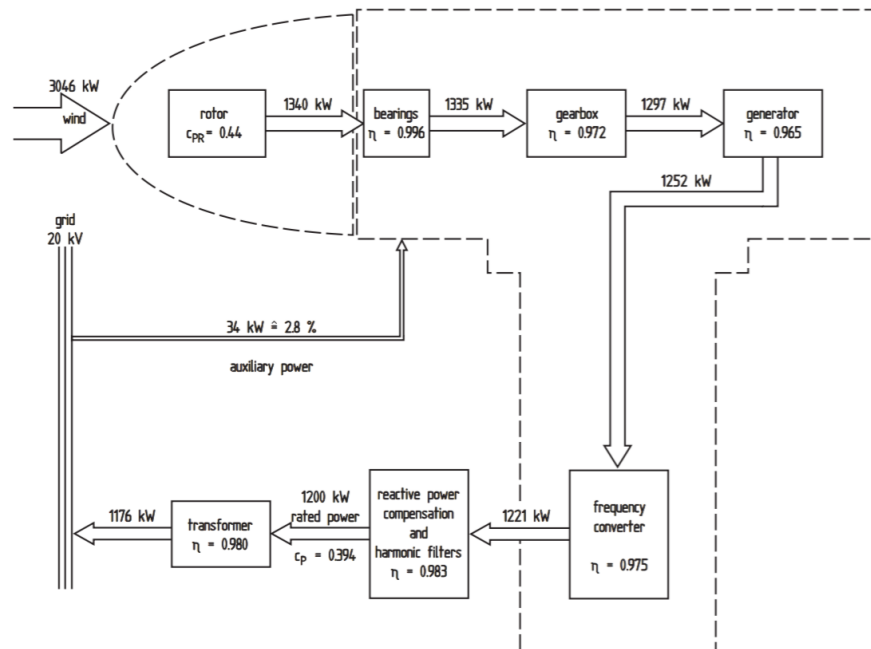


Figura 2.5: Eficiencia de conversión de cada componente de un aerogenerador. Tomada en su totalidad de [2].

Otro elemento importante para estimar la energía anual producida es el Factor de planta (FP), el cual es un indicador de desempeño del aerogenerador. Para determinar el FP de un campo eólico se requiere conocer la cantidad de energía que el campo entregó en un periodo de tiempo y dividirlo entre la energía que generaría el campo si estuviera trabajando el mismo intervalo de tiempo a su potencia nominal:

$$FP = \frac{\text{Energía generada en un tiempo (MWh)}}{\text{Potencia nominal instalada (MW)} \times \text{Tiempo (h)}}. \quad (2.5)$$

El FP a diferencia del límite de Betz puede tener valores mayores a 59.3%, debido a que es únicamente la energía generada entre la energía que puede generar el aerogenerador.

Mientras más cercano a la unidad sea el FP mayor será el desempeño del campo eólico. Algunos valores típicos para granjas eólicas abarcan el intervalo de 15-52% dependiendo de la disponibilidad del recurso.

Si se cuenta con el factor de planta se puede obtener AEP utilizando la siguiente ecuación

$$AEP = FP \times 8760 \text{ h} \times P_n \times \eta \quad (\text{MW h}), \quad (2.6)$$

donde 8760 representa el número de horas al año, P_n la potencia nominal del aerogenerador y η la eficiencia de conversión del aerogenerador.

2.1.3 Control y Sistema de orientación de *Yaw* en aerogeneradores

Una información relevante para el desarrollo de este trabajo es el funcionamiento del sistema de control en las turbinas. El cual se encarga de los aspectos de protección en un campo eólica y la activación o desactivación de algunos de los componentes del aerogenerador. La componente en la que se hará más énfasis para la propuesta de este trabajo es el sistema de orientación o sistema *Yaw* el cual, es un mecanismo que se utiliza para capturar una mayor cantidad de energía a través del seguimiento automático del viento.

Imaginemos un aerogenerador posicionado hacia una orientación fija. Si el vector de viento (U) llega de manera normal hacia la góndola se dice que está orientado. Es decir, el viento le llega de frente. Por otro lado, si el vector U tiene una diferencia se dice que está desalineado. Esto se puede ver gráficamente en la Figura 2.6a. El desalineamiento se refiere al ángulo (γ) que se forma entre el vector de viento U con respecto al frente del aerogenerador.

El funcionamiento del sistema de orientación es simple. Orienta el aerogenerador para que γ sea lo más pequeño posible. La torre del aerogenerador que sostiene a la góndola gira en un eje de rotación como se observa en la Figura 2.6b. Se encuentra soportado en unos rodamientos de gran tamaño que le permiten girar sobre un eje. Cuenta con un sistema de frenado, un motor principal que mueve el aerogenerador y el sistema de engranes. En los campos eólicos, los aerogeneradores cuentan con un algoritmo de seguimiento que orienta automáticamente el rotor si existe una variación en el ángulo de incidencia del viento.

El sistema de orientación es uno de los sistemas más complicados en aerogeneradores de gran capacidad [24]. El algoritmo de seguimiento debe orientar el aerogenerador y esto implica mover un peso considerable lo cual somete a los rodamientos principales a un desgaste continuo.

Como se mencionó anteriormente, el aerogenerador se orienta para capturar más energía, porque si está desalineado el C_p disminuye. Esta variación del C_p en función de γ se

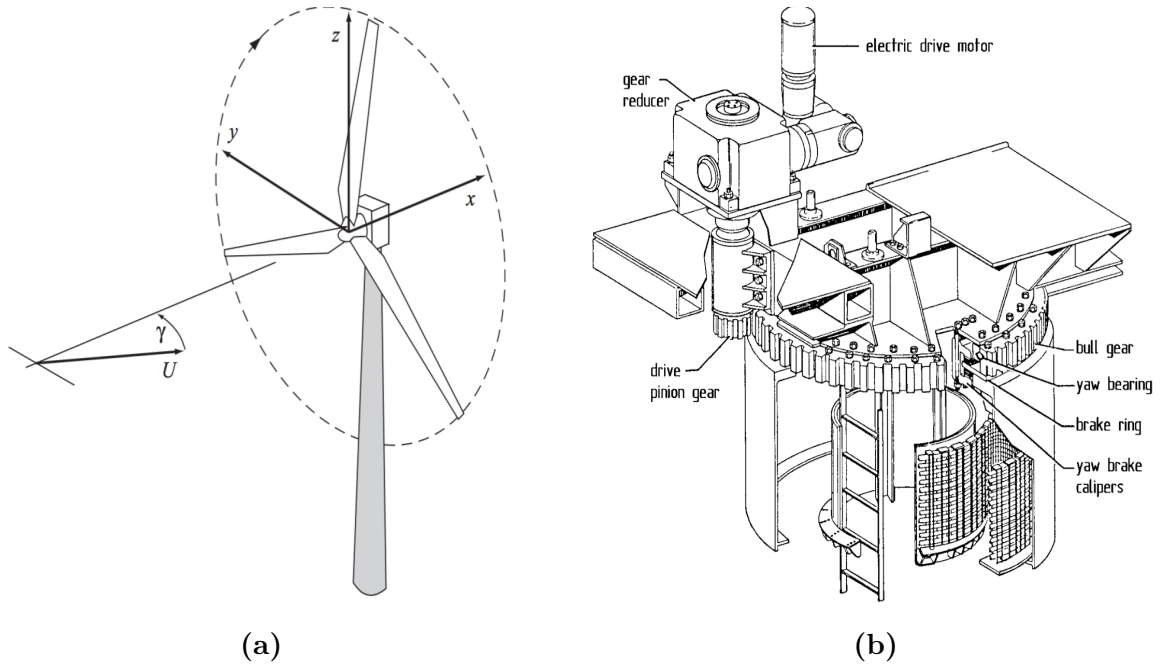


Figura 2.6: Concepto de desalineamiento del aerogenerador (γ) debido al cambio de dirección de viento incidente (U) (a), y diagrama del sistema de orientación (b). Extraído en su totalidad de [3].

puede determinar con la ecuación [2]:

$$C_p = 4a(\cos \gamma - a)^2 \quad (2.7)$$

en donde $a = 1 - \frac{v_2}{v_1}$ es el coeficiente de inducción axial utilizado en la teoría de elemento de aspa (*BEM*, explicada en [2]). Graficando la ecuación (2.7) se obtiene el comportamiento del C_p en función del γ como se muestra en la Figura 2.7. Se puede observar, mientras mayor sea el desalineamiento menor será el C_p . Adicionalmente, si un aerogenerador está desalineado se somete a cargas dinámicas mucho mayores [25].

Se puede utilizar la ecuación (2.7) para aproximar de una manera mejor la potencia del aerogenerador, porque esta ecuación considera las pérdidas por desalineamiento.

El sistema de seguimiento³ implica un gasto energético importante ponerlo en operación todo el año. Considerando que la tendencia de los aerogeneradores es a crecer en tamaño de rotor, se puede hacer la suposición que todas las componentes crecerán proporcionalmente a medida que el aerogenerador crece. Esto implica un incremento en los costo operación y mantenimiento. Ashuri [9] advierte que en un futuro el sistema de seguimiento será uno de los factores que impactarán más en el costo final de un aerogenerador. Por lo anterior, establecer una metodología para que el algoritmo de seguimiento disminuya y optimice el uso es ya una necesidad.

En la siguiente sección se introducen las estadísticas de falla de las componentes

³Ver TablaA.1 para ver el desglose por componente

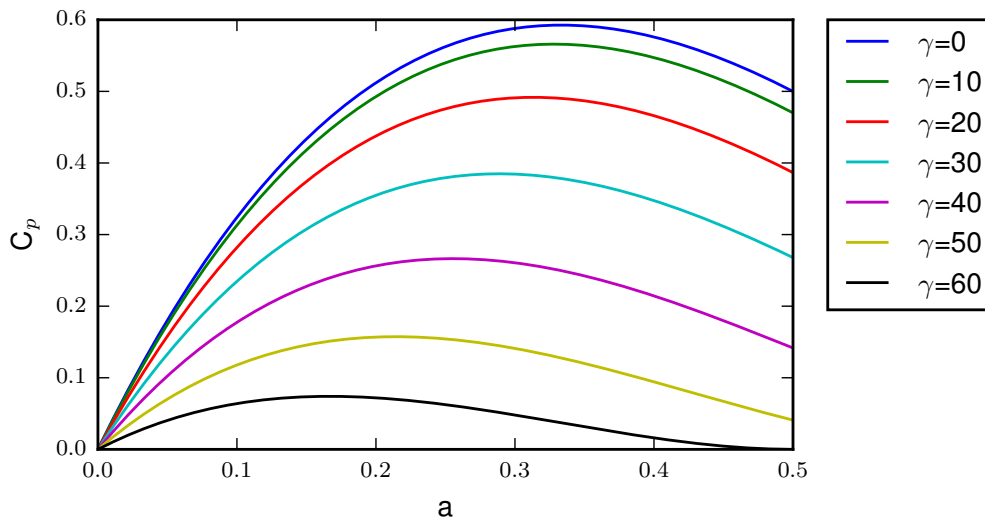


Figura 2.7: Variación del C_p en función de γ y a . Extraído en su totalidad de [3].

principales de un aerogenerador. Esta información será útil para llevar a cabo una simulación aproximada de las condiciones reales, que a su vez complementará la metodología propuesta en este trabajo.

2.1.4 Fallas en turbinas eólicas

Hoy en día, existen varios estudios en EUA [26, 4] donde se han realizado estadísticas de fallas en las grandes plantas de generación eólica. Utilizando miles de aerogeneradores en operación se han podido determinar los principales mecanismos de falla, cuántos eventos al año hay y cuánto tiempo se queda el sistema fuera de operación por esta falla (ver Figura 2.8). Como se observa en la Figura 2.8 las principales componentes que fallan —según las estadísticas de [4, 26]— son el rotor, generador eléctrico y el sistema de control (*Yaw*). Particularmente el sistema de orientación puede llegar a fallar un total de 3 a 4 veces por turbina por año. Aunque la eventualidad es poca, si falla el sistema de orientación repercute en tiempos largos fuera de operación [4]. En la cual se puede observar que aunque una falla del sistema de orientación se puede presentar 3 a 4 veces por año es de las fallas que representan mayor tiempo fuera de operación. Considerando que la operación del aerogenerador es de 20 años, la cantidad de fallas es considerable.

Una falla en un campo eólico es pérdida de dinero. Una manera de aumentar la vida útil de los materiales es optimizar el uso, y se puede llevar a cabo desde el nivel de control de la planta.

En la siguiente sección se mostrarán las herramientas matemáticas necesarias para llevar a cabo una simulación que, utilizando las estadísticas de fallas y el desalineamiento, complemente una simulación de operación de un aerogenerador.

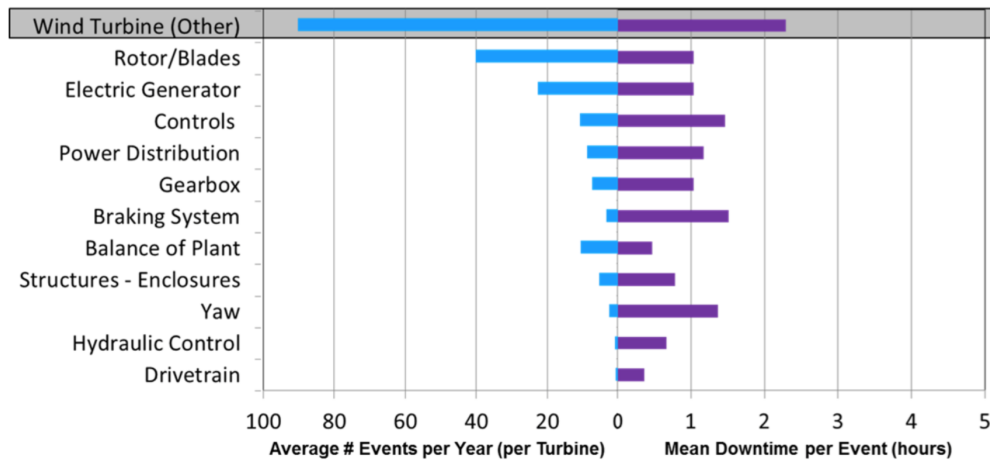


Figura 2.8: Estadísticas de fallas de campos eólicos en USA. Tomada en su totalidad de [4]

2.2 Procesos estocásticos

Imaginemos un problema simple de un proceso que involucra una variable aleatoria dependiente del tiempo: Una persona llega a un banco y se forma en la fila para ser atendido. Esta misma persona se tarda 50 minutos para ser atendido por el cajero. Al día siguiente, la misma persona, vuelve a ir al banco, pero esta vez se tarda 20 minutos, y al día siguiente se tarda 5 minutos y al siguiente 40. Como es de esperarse, el tiempo de espera para este ejemplo cambia dependiendo del día. Utilizando estadística se puede calcular un promedio y una desviación estándar, y si se registran los valores día a día se tendrá una serie de tiempo (ver Figura 2.9a) que a su vez tiene asociada un histograma de frecuencias (ver Figura 2.9b). Visualizando el tiempo de espera en el histograma de frecuencias, se observa que existe una cierta aleatoriedad y se puede expresar en una distribución de probabilidad. Esto lleva a la conclusión que el tiempo de espera puede ser un proceso aleatorio.

Siendo rigurosos, se puede establecer las variables deterministas para simular este evento, sin embargo el planteamiento puede ser de difícil a muy complejo. Una manera sencilla de observar las fluctuaciones temporales del evento es asumir que el tiempo de espera es aleatorio o casi aleatorio. Planteando lo anterior una pregunta interesante de responder sería: ¿cuánto tiempo se tarda una persona en el banco un día n ?

En la naturaleza existen muchos procesos que se pueden recrear, aproximar y predecir utilizando modelos para variable(s) aleatoria(s) en el tiempo o un estado finito. Estos eventos contienen variables que se les asocia una distribución de probabilidad. La cual permite conocer la probabilidad de que aparezca un evento.

En este trabajo se utilizará la siguiente definición de variable aleatoria:

Definición 1 Se llama variable aleatoria a un “evento” que puede ser definido con una

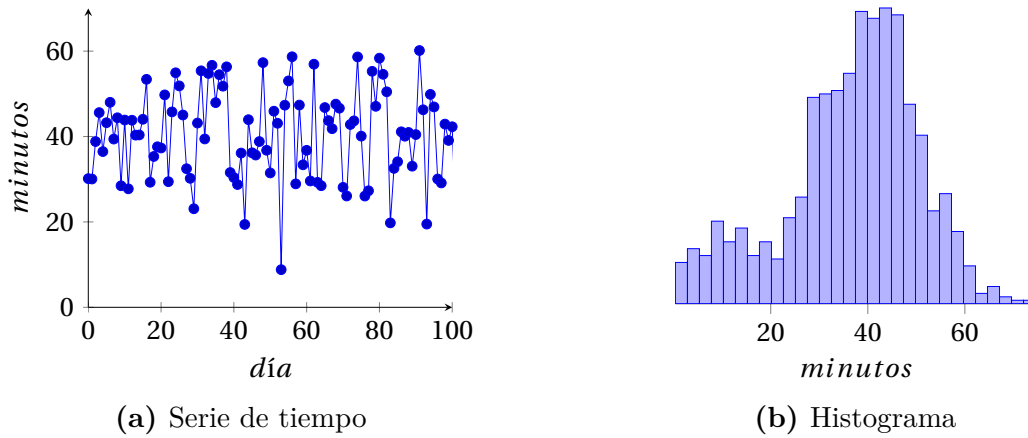


Figura 2.9: Tiempo de espera en un banco

cierta probabilidad.

Se pueden asociar modelos de variables aleatorias a muchos eventos, algunos ejemplos son: caminantes al azar, crecimiento de bacterias, valores de la bolsa, ruido transistor, etc. Si alguno de estos modelos se asocia con una serie temporal se denomina *modelo estocástico* y al proceso completo como el desarrollo continuo del modelo en el tiempo se llama proceso estocástico [27]. Con lo anterior se llega a la siguiente definición:

Definición 2 *Un proceso estocástico es un conjunto de variables aleatorias dependientes del tiempo.*

Hoy en día existen grandes avances sobre los procesos estocásticos aplicados a muchas ramas de investigación. El objetivo de este trabajo no es explicar a fondo la matemática detrás de los procesos sino su acoplamiento con la energía eólica y la simulación de viento ⁴.

En general, existen dos variantes principales de los procesos estocásticos: discretos y continuos. Supongamos un proceso donde existe una variable aleatoria $X(t), t \in S$, donde S es un conjunto finito de “eventos” y contable. Como los eventos contenidos en S son contables y definidos, se dice que es un proceso con tiempo discreto (DT, por sus siglas en inglés). Explicando un poco más a detalle, un proceso DT no es un continuo de eventos, es decir los valores del espacio S están bien definidos o parametrizados. Ahora supongamos otro proceso en un tiempo $t \in \mathbb{R}$. Por el hecho de pertenecer a un conjunto de números mucho mayor que uno contable se denomina de tiempo continuo (CT, por sus siglas en inglés). Ejemplos de este proceso se pueden encontrar en un salón de clases donde se medirá la temperatura $T(t)$ para cada $t \in [0, \infty]$, medir el voltaje de salida de un transistor, calcular el tiempo de espera en un banco, entre otros.

Ambas variantes de los procesos estocásticos tienen sus aplicaciones en la ingeniería. Sin embargo, un proceso, desde el punto de vista de un ingeniero es CT, porque involucra

⁴si se requiere conocer mayor información sobre la matemática de los procesos estocásticos se puede consultar [28].

eventos relacionados usualmente con fechas (tiempo). Desde esta perspectiva, un proceso estocástico CT ayudará a representar un problema que involucre una serie de tiempo.

Una suposición muy importante de los procesos estocásticos es que cada aparición de la variable aleatoria X es independiente de cada aparición anterior o futura. Imaginando el comportamiento de la temperatura en un salón de clases, estrictamente hablando, si se mide la temperatura en un tiempo en donde el salón está lleno, es razonable pensar que la temperatura sea alta. También, es razonable que al siguiente instante de tiempo la temperatura anterior involucró de alguna manera el resultado de la temperatura actual. Este proceso es más realista que uno totalmente aleatorio.

Utilizando el razonamiento anterior se llega a una gama nueva de modelos estocásticos que su principal objetivo es, por ejemplo, saber cuál será el valor de las acciones del mercado en un tiempo t_{i+1} si se conoce t_i . Este tipo de modelos se denominan cadenas de *Markov* o procesos markovianos.

2.2.1 Cadenas de Markov o procesos Markovianos

Las cadenas de *Markov*⁵ son procesos estocásticos de transición entre eventos (estados) definidos. Se utilizan para poder contestar preguntas como: ¿cuál es la probabilidad de que llueva mañana si el día de hoy llovió?? ¿cuánto tardaré en ganar en BlackJack si llevo tres juegos ganados? ¿cuánto tomará al viento en cambiar de dirección si conozco la dirección actual? Estas preguntas pueden ser respondidas si se utilizan cadenas de *Markov*. Usualmente lo que se conoce comúnmente cadena de *Markov* son los procesos estocástico que cumplen el siguiente enunciado: “el resultado de un experimento específico puede afectar el resultado del siguiente experimento”. La propiedad más importante de estos procesos es en la probabilidad solo con el evento anterior, si tenemos una secuencia de estados en el tiempo se puede predecir un siguiente estado si se conoce el estado anterior; esto solo es posible porque se asume que los eventos son dependientes entre si. Para entender mejor la importancia de las cadenas de *Markov* y sus aplicaciones se explicará sus variantes, propiedades y ejemplos de uso.

Supongamos un conjunto de estados $S = A, B, \dots, N$. Un proceso *Markoviano* inicia en el estado A , y en el siguiente paso de tiempo se mueve al estado B con una probabilidad 0.3 y esta probabilidad no depende de en qué estado estaba la cadena antes de llegar al estado actual [29]. Una vez en el estado B , la cadena se mueve al estado C con una probabilidad de 0.5, y así sucesivamente se irá moviendo la cadena entre los estados. Una manera de representarlo gráficamente el proceso es utilizando un diagrama de *Markov* como el mostrado en la Figura 2.10.

Se puede observar en la Figura 2.10 que existen tres eventos principales A , B y C , adicionalmente están dibujadas unas flechas que apuntan entre estados. Estas flechas las podemos interpretar como una “comunicación” entre estado y estado. Cada una de estas

⁵La primera aparición de este concepto fue en 1906 por el ruso Andrei Andreyevich Markov

flechas tiene asociado un número que será la probabilidad de que ocurra esa comunicación. Si todos los estados tienen comunicación con los demás estado, se dice que todo el sistema está comunicado ⁶.

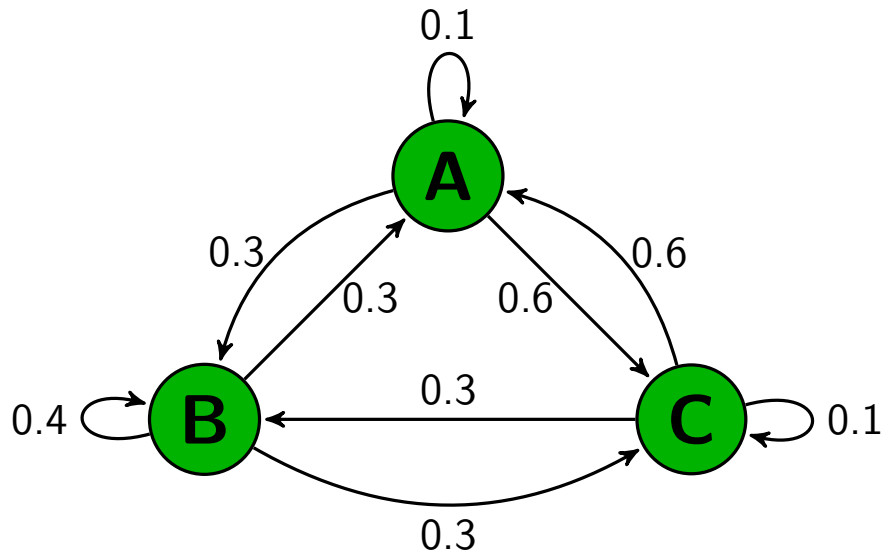


Figura 2.10: Diagrama de un proceso markoviano

También se puede representar la cadena de *Markov* de la Figura 2.10 como una matriz. En el cual, cada elemento de la matriz representa la probabilidad de transición entre estados,

$$\begin{array}{c} A \quad B \quad C \\ A \begin{pmatrix} 0.1 & 0.3 & 0.6 \end{pmatrix} \\ B \begin{pmatrix} 0.3 & 0.4 & 0.3 \end{pmatrix} \\ C \begin{pmatrix} 0.6 & 0.1 & 0.3 \end{pmatrix} \end{array}.$$

Esta matriz se denomina *matriz de transición de probabilidad*, la cual es una matriz estocástica en la que la suma de todos los elementos de una fila es igual a uno. Usualmente se define como p_{ij} , siendo i el estado actual y j el estado siguiente. La matriz proporciona la información suficiente acerca de la dinámica del movimiento entre estados.

Para ver la relevancia de la matriz de probabilidad se usará un ejemplo aplicado a meteorología: Se tiene una matriz de probabilidad de transición \mathbf{P} que tiene los siguientes elementos

$$\mathbf{P} = \begin{array}{c} L \quad S \quad N \\ L \begin{pmatrix} 3/4 & 1/4 & 0 \end{pmatrix} \\ S \begin{pmatrix} 1/2 & 0 & 1/2 \end{pmatrix} \\ N \begin{pmatrix} 1/2 & 1/4 & 1/4 \end{pmatrix} \end{array},$$

⁶En este trabajo, utilizaremos únicamente una cadena de *Markov* totalmente comunicada

donde L significaría la probabilidad de día lluvioso, S día soleado y N con nieve. Supongamos que el día de hoy nevó, eso nos coloca en la fila 3 de la matriz anterior, la probabilidad de que llueva mañana es de $1/2$, que esté soleado $1/4$ y que caiga nieve otra vez de $1/4$. Si al día siguiente, por probabilidad, llueva, estaríamos en la fila 1 y aquí sucede algo interesante. Una vez ubicado en la fila 1 tengo $3/4$ de probabilidad de que llueva y $1/4$ de que esté soleado. Sin embargo, la probabilidad de que caiga nieve es cero, es decir, no existe algún evento en el cual si llovió el día de ayer, nieve al día siguiente. Esto significa que no hay conexión entre estados.

Utilizando el mismo ejemplo, se puede preguntar cuál es la probabilidad que llueva a un cierto tiempo n . Esto se puede determinar utilizando la matriz de transición P . Ya que se conoce cuál es la probabilidad de cambio de estados, se puede llegar al siguiente teorema matemático para conocer la probabilidad de transición en un tiempo definido

$$p_{ij}^{(n)} = \sum_{k=i}^r p_{ik} p_{kj}, \quad (2.8)$$

siendo r el número de estados y n el paso que se desea extender la cadena de *Markov*.

Ya se habló de cómo se puede conocer la probabilidad a un tiempo n , sin embargo no se ha mencionado que requerimientos se necesita cumplir para poder extender la cadena. Para esto se requiere un vector inicial. Este será el que determina la distribución de probabilidad de la cadena, por ejemplo, si tengo el problema de meteorología, este vector inicial contendría las probabilidad iniciales de que llueva, salga el sol o que nieve. Conociendo el vector inicial (\mathbf{u}_0) se puede conocer la distribución de probabilidad a un tiempo dado ($\mathbf{u}^{(n)}$) si también se cuenta con la matriz de transición de probabilidad, es decir,

$$\mathbf{u}^{(n)} = \mathbf{u}_0 P^n. \quad (2.9)$$

La ecuación (2.9) permite conocer una distribución de probabilidad propagada por la cadena a un tiempo específico. Esto es de gran utilidad para muchas aplicaciones de la ingeniería y ciencia en general. Por ejemplo, si se requiere saber cuál será la distribución de probabilidad del viento a un cierto tiempo, cómo se comportará la tasa de mortalidad a un cierto año, etc. Sin embargo, para poder determinar esta distribución de probabilidad se tienen que cumplir dos características en la cadena de *Markov*: aperiódica e irreducible.

Las propiedades son las siguientes:

- a) Irreducible es la propiedad que no importa en que estado se encuentre actualmente la cadena podrá llegar a cualquier otro estado en un tiempo finito (espacio de estados comunicados, ver Figura 2.11a).
- b) Aperiódica es lo contrario a periódico y quiere decir que caigo al mismo estado en el mismo tiempo, ver Figura 2.11b.

Con estas dos propiedades se garantiza que si se tiene un espacio de estados finitos la distribución de probabilidad se acerca a una distribución estacionaria conforme crece n .

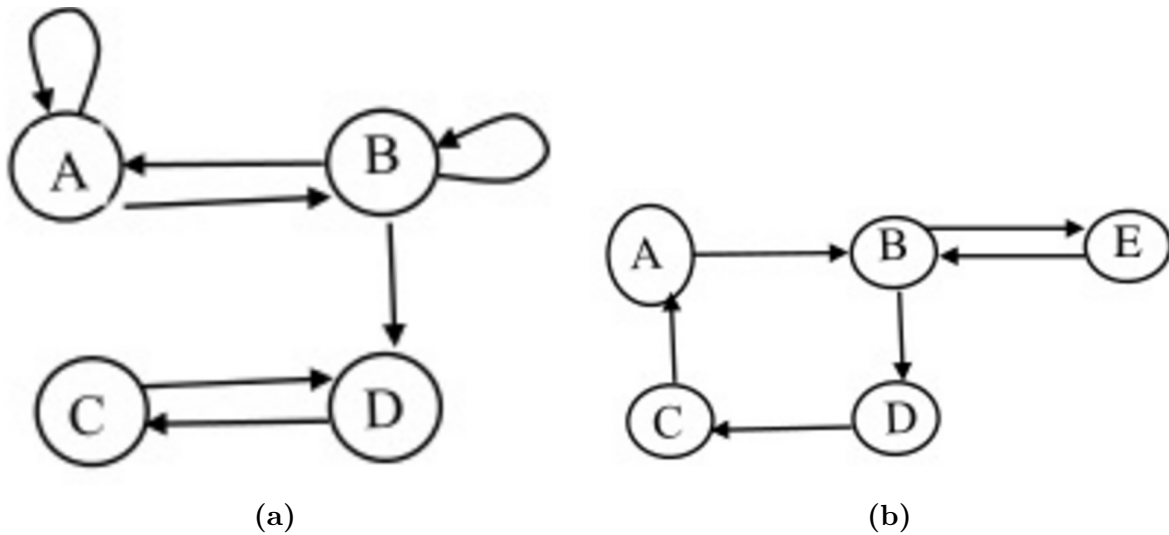


Figura 2.11: Cadena de *Markov* a) irreducible y b) aperiódica.

Las cadenas de *Markov* discretas son útiles y sencillas para muchas aplicaciones. Pueden tener infinitud de estados definidos, y muchas variantes en la comunicación entre estados. Sin embargo, no proporcionan una idea de la dinámica de transición para un tiempo continuo, porque considera que las transiciones de probabilidad no cambian en el tiempo. Como primera aproximación, se puede suponer que un sistema conserva su matriz de probabilidad de transición, pero cómo garantizar esto. Por lo anterior, se describirán las cadenas de *Markov* continuas que consideran una dependencia temporal de la matriz de transición de probabilidad en la siguiente sección.

2.2.2 Cadenas de Markov Continuas

Existen muchos procesos o eventos cotidianos que suceden en tiempo continuo, conducir un coche en la carretera, jugar un partido de fútbol, resolver un examen o realizar un reporte. Sería muy difícil tratar de recrearlos pensando en tiempos discretos e incluso podría traer malos resultados. Sin embargo, si consideramos ejemplos como los anteriores como procesos continuos sería más fácil representarlos y observarlos.

Una aplicación de modelos estocásticos son las cadenas de *Markov* de tiempo continuo (CTMC, por sus siglas en inglés) que tienen una aplicación esencial para series de tiempo. A diferencia de las cadenas de *Markov* discretas, son modelos de tiempo continuo, es decir, el resultado es una serie o lista de eventos en el tiempo. Un ejemplo de un proceso continuo sería la trayectoria de una rata en un laberinto, que para fines prácticos es de mayor utilidad seguir los pasos de la rata paso a paso $t \geq 0$ (serie de tiempo) que saber dónde está la rata a un tiempo n [30].

De igual forma a que las cadenas de *Markov* discretas, las cadenas continuas son modelos donde existen estados finitos $X(t)$ $t \geq 0$, pueden *predecir el futuro* dado el

presente y esto se denomina la propiedad de *Markov*.

Definición 3 Un proceso estocástico $X(t) : t \geq 0$ con estados discretos S es llamado cadena de Markov de tiempo continuo (CTMC) si para toda $t \geq 0, s \geq 0, i \in S, j \in S$,

$$P(X(s+t) = j | X(s) = i, \{X(u) : 0 \leq u \leq s\}) = P(X(s+t) = j | X(s) = i) = P_{ij}(t),$$

donde P_{ij} se puede interpretar como una matriz estocástica ⁷. Representa la transición de eventos, de un estado i (presente) a un estado j (futuro). Para cada tiempo $t \geq 0$ existe una matriz de transición

$$P(t) = (P_{ij}(t)). \quad (2.10)$$

Adicionalmente este modelo permite determinar el tiempo de permanencia en un estado, es decir, cuánto tiempo (continuo) llovió recordando el ejemplo del clima. Esta información se puede calcular determinando la matriz de flujos^(q) entre estados

$$q = \begin{pmatrix} -a_0 & a_0 P_{0,1} & a_0 P_{0,2} \\ a_1 P_{1,0} & -a_1 & a_1 P_{1,2} \\ a_2 P_{2,0} & a_2 P_{2,1} & -a_2 \end{pmatrix},$$

donde $a_i P_{i,j}$ se interpreta como el flujo de transición de un estado i a un estado j dado que la cadena está en el estado i y donde

$$a_i = \sum_{j \neq i} q_{ij} = -\frac{1}{H_i},$$

siendo H_i el tiempo medio de vida media del estado.

Ahora, con las ecuaciones anteriores falta discutir cómo se crea una CTMC. Uno de los pasos más importantes es determinar la matriz de transición de probabilidad \mathbf{P} . Para el caso discreto, calcular la matriz se resuelve utilizando la ecuación (2.8). Sin embargo, para el caso continuo, se complica el asunto, debido a que ahora, la matriz de transición de probabilidad $P_{ij}(t)$ requiere una serie de ecuaciones diferenciales para resolverse.

Las CTMC se pueden aplicar a infinidad de temas como contaminación antropogénica [31], aplicaciones en dinámica de vegetaciones [32], economía [33, 34], meteorología [35, 36] y aplicaciones eólicas [37, 38, 39, 40, 41, 42, 43, 44].

Para cumplir con los objetivos de este trabajo, la cadena de *Markov* servirán como herramienta en conjunto con el agrupamiento de estados para simular la dinámica del viento de los casos de estudio. Se planteó realizar un simulador de viento en el cual se determinen un número inicial de estados utilizando los algoritmos de agrupación y calcular la matriz de probabilidad de transición para poder recrear la distribución en un paso de tiempo n o recrear la dinámica de los vectores de velocidad del viento.

⁷Una matriz estocástica es aquella que la suma de sus filas es igual a cero.

2.3 Agrupamiento

Dentro del análisis de datos existe un área particular y útil al momento de clasificar grandes cantidades de información. Esta rama se le conoce como algoritmos de agrupamiento o *clustering*. La cual busca separar un conjunto de datos en cúmulos o *cluster* que comparten alguna relación entre sí. Se puede aplicar en muchas las áreas como bioinformática, inteligencia artificial, teoría de la información y energías renovables ya que existen muchos datos que pueden estar relacionados con algún tipo de fenómeno.

Para ejemplificar un algoritmo de agrupamiento imaginemos un problema en el cual se tiene un conjunto de bacterias. Se miden las propiedades de cada una de ellas y se grafican dos propiedades cualquiera como se observa en la Figura 2.12. Cuando se grafican dos propiedades de los genes existen zonas donde hay una concentración importante de información. Estas son las zonas de interés por que la información que reflejan es un grupo o clasificación. Al caer dentro de la misma zona, las diferentes bacterias comparten una propiedad en esta región específica y podemos clasificarlas. La pregunta que enseguida sale a luz es ¿por qué clasificar? Los seres humanos clasificamos todo. Desde la comida, especies de animales, religiones, lenguaje, la tecnología, entre muchas otras cosas. Es de gran utilidad poder asociar dos o más cosas de una manera gráfica podemos ver a través de su co-dependencias. Es aquí la gran importancia de los algoritmos de agrupamiento. No importa que tipo de datos sean, se pueden clasificar y esto puede indicar propiedades compartidas así como alguna relación numérica, física, estadística o química. En el área científica poder asociar a un grupo una propiedad de estado es crucial ya que así, se puede predecir o conocer el comportamiento de dicho grupo.

2.3.1 Algoritmos de agrupamiento

Para poder separar la información por cúmulos se requiere aplicar algoritmos computacionales que realicen el trabajo de clasificar cada elemento de un conjunto de datos. Gracias al avance tecnológico hoy en día existen varios algoritmos para agrupar un conjunto de datos bajo muchos criterios y siguen desarrollándose más. Ésto complica un poco la definición del resultado obtenido de una agrupación. No existe una definición formal de la palabra cúmulo ya que no puede ser precisado su definición debido a sus múltiples interpretaciones [45] dependientes del área de aplicación. Sin embargo, lo que si es claro es que si existe una correlación entre datos se pueden clasificar o separar de los demás datos.

Los algoritmos computacionales para agrupamiento también están divididos en dos grupos principales: los jerárquicos y los particionados. Ambos se utilizan para separar un conjunto de datos en muchos campos, sin embargo, existe una diferencia en el concepto de

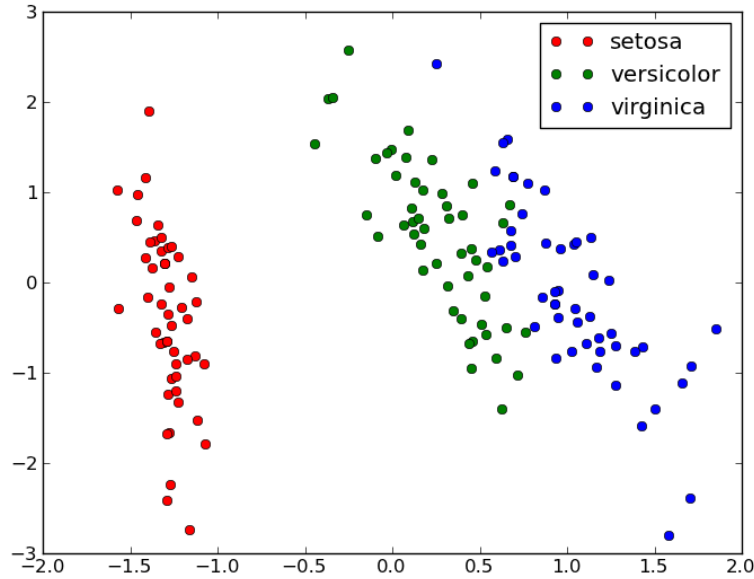


Figura 2.12: Ejemplo de aplicación de algoritmos de agrupamiento en sepas de pétalos. Base de datos disponible en [5].

cúmulo. El primer ejemplo serían los jerárquicos (ver Figura 2.13b) los cuales consideran que existen cúmulos dentro de cúmulos, es decir, sub-cúmulos o cúmulos de segundo nivel. Por otro lado, los particionados (ver Figura 2.13a) consideran que no existen cúmulos sobrepuestos, es decir, cada dato solo tiene asociado un cúmulo único.

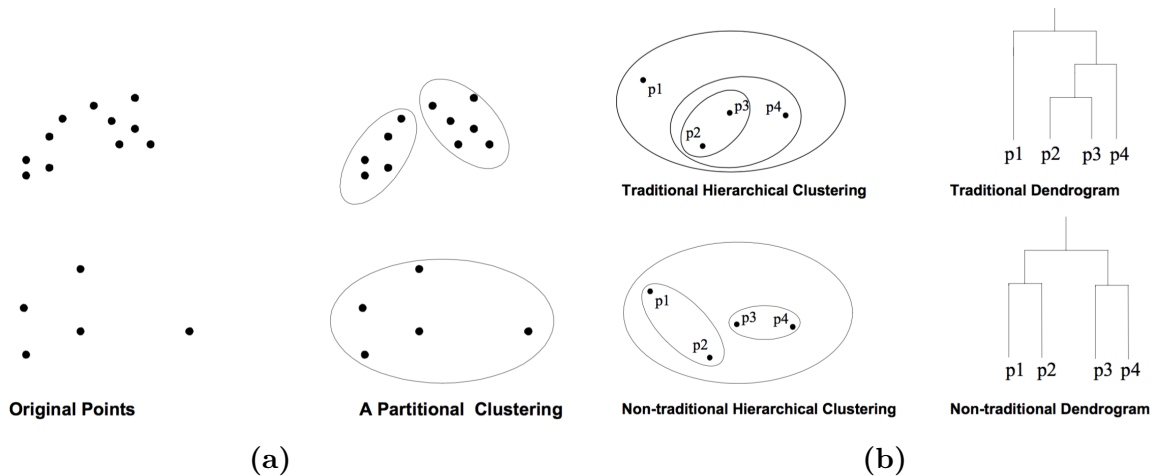


Figura 2.13: Tipos de algoritmos de *clustering*. Extraído en su totalidad de

Ambos grupos pueden ser utilizados para el mismo trabajo. Sin embargo, es importante tener clara la diferencia a la hora de determinar la naturaleza de los datos. Si estamos hablando de datos que solo pueden tener un único valor, no es tan apropiado

pensar que existen sub-clasificaciones. Para ver esto más claro se coloca las gráficas de la Figura 2.13 en la cual podemos observar que si tenemos un conjunto de datos como los que se muestran en la Figura 2.13a podemos separarlos de dos formas particional o jerárquica⁸.

Existen muchas maneras de aplicar estos dos tipos de agrupación y la manera de hacer es con algoritmos. Algunos de los algoritmos más utilizados son:

- k -means,
- Density-based spatial *clustering* of applications with noise (DBSCAN),
- Biclustering,
- Distribution *clustering*
- Spectral *clustering*,
- Entropy based *clustering*,
- Density Based,
- CLARA,
- BIRCH,

entre muchos otros. Cada día surgen más, ya que la única complicación de crear uno propio es establecer el criterio por el cual se va a clasificar⁹.

Los algoritmos separan la información utilizando diferentes criterios de clasificación y/o buscando minimizar diferentes propiedades. Para el caso de DBSCAN, se busca encontrar las zonas con mayor densidad de datos, el algoritmo k -means busca reducir la distancia euclidiana entre punto hasta que se converge en un centro y la *distribution clustering* busca encontrar todos los puntos pertenecientes a una distribución de probabilidad. Ya que cada uno busca una cosa diferente, las agrupaciones son diferentes en cada uno, como se observa en la Figura 2.14. Existe una gama alta de posibilidades de particiones y una de las preguntas que se ha tratado de resolver es ¿cuál es la mejor?, para responder esta pregunta es necesario saber qué es lo que se busca. En la Figura 2.14 se observan tres conjuntos de datos en cada fila, en la fila superior se tienen dos círculos, en este caso se esperaría que la clasificación de datos diera como resultados dos círculos de diferentes datos, sin embargo, al no buscar la misma propiedad, solo dos de ellos clasifican los datos en dos círculos diferentes (DBSCAN y SpectralClustering). La misma situación se tiene para los datos de la fila de en medio, en donde nuevamente los algoritmos DBSCAN y SpectralClustering originan una separación clara de los datos. En la última fila

⁸Una manera gráfica de observar los *cluster* jerárquicos es con el dendograma el cual pone las sub-clasificaciones o los *clusters raiz*

⁹Es muy importante el tiempo de cómputo que toma cada uno. No basta con clasificar si no buscar la manera más eficiente de buscarlos

se tiene una otra serie de datos con una separación distinta, en donde se observa que solamente dos algoritmos separan de manera más clara los datos (DBSCAN y MeanShift). Se observa también que de los algoritmos que fueron funcionales para separar la serie de datos de la fila 1 y 2, solamente uno de ellos, separa de manera ordenada los datos de la fila 3 (DBSCAN).

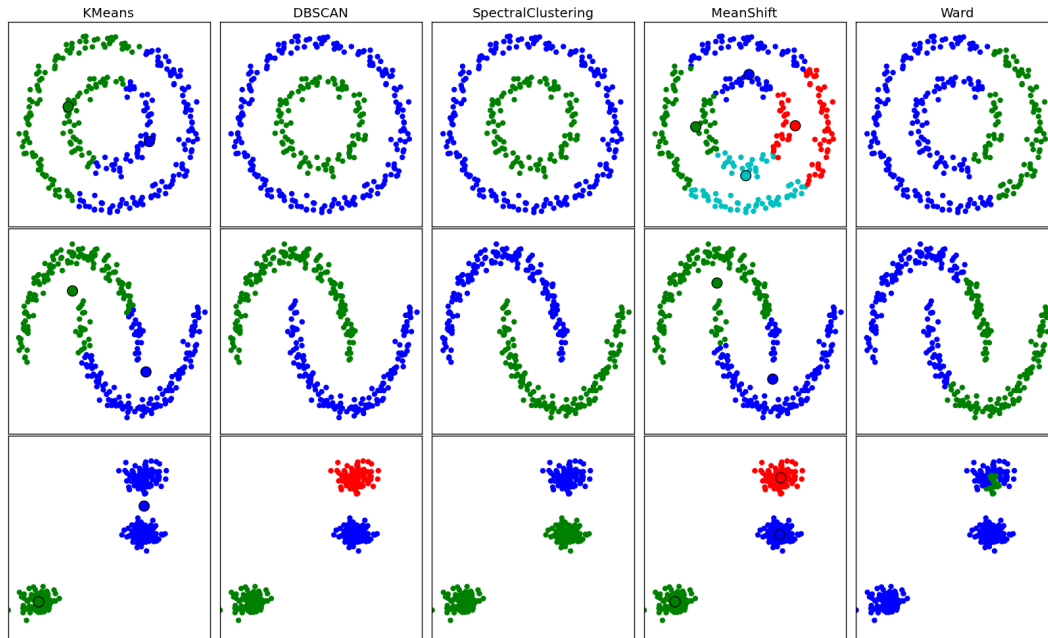


Figura 2.14: Algoritmos de agrupamiento. Base de datos disponible en [5]

Como se puede observar en la Figura 2.14 cada algoritmo de agrupamiento regresa una separación diferente de cúmulos, es por eso que se debe de tener cuidado al aplicar cualquier, porque pueden no entregar información relevante referente a los datos. ¿Cuál es el algoritmo de agrupamiento apropiado? Para responder la pregunta anterior se requiere conocer la naturaleza de los datos. Si se conoce, se puede saber el posible resultado y así se buscaría el algoritmo de agrupamiento que regrese el resultado esperado. Sin embargo, cuando se desconoce la naturaleza de los datos es difícil asegurar el funcionamiento de un algoritmo de agrupamiento específico y más aun cuando el número de cúmulos no es evidente cuál es el número apropiado de cúmulos. Existen muchos trabajos que buscan el número apropiado de cúmulos con estadística [46, 47]; buscarlo no es el fin de este trabajo.

Un primer paso para determinar cuál es el mejor algoritmo de agrupamiento es empezar con los más sencillos e ir aumentando la complejidad dependiendo de los resultados. Tal vez el algoritmo más sencillo sea el k -means en cual se hablará a continuación.

2.3.2 k-means

Los algoritmos de agrupamiento pueden demandar mucho cálculo computacional en grandes bases de datos. También dependiendo del número de dimensiones el cálculo puede elevarse considerablemente. Para las aplicaciones de energía eólica, los datos de velocidad del viento suelen ser extensas dependiendo del número de años y la resolución, por lo que escoger un algoritmo de agrupamiento no óptimo puede llevar a mucho tiempo computacional y errores en la interpretación. Existen algoritmos computacionales que son simples y no requieren mucho tiempo computacional, y es por eso que para el desarrollo de este trabajo y fines cualitativos se utilizó el algoritmo k -means.

El algoritmo k -means consiste en lo siguiente:

1. Inicializar el centro del *cluster*

$$\mu_i = \theta_i$$

2. Asignar el *cluster* más cercano para cada punto

$$c_i = \{j : d(x_j, \mu_i) \leq d(x_j, \mu_i)\}, i$$

3. Elegir el centro del *cluster* con la media de los datos

$$\mu_i = \frac{1}{|c_i|} \sum_{j \in c_i} x_j, \forall_i$$

4. Repetir pasos 1 y 2 hasta converger.

Utilizando una simple función para minimizar la distancia se encuentra el centro de lo que será el cúmulo y, con base en la distancia de los datos se aplicará el criterio de pertenencia a un cúmulo o otro. Este algoritmo utiliza un criterio geométrico, por tanto, su complejidad en programación es mínima y también es ampliamente utilizado como primera aproximación para separar datos. Será de gran utilidad para la primera aproximación de unos datos desconocidos. Sin embargo, este algoritmo tiene una limitación: el número de k , el cual representa el número de cúmulos en el conjunto de datos. Este valor se tiene que especificar antes de iniciar el algoritmo porque así inicializará los centros. Saber el número de k sin conocer los datos puede ser difícil, pero existe una manera sencilla de atacar este problema. Si se tiene una gráfica en la cual se vean representados los cúmulos de datos se puede saber el número aproximado de cúmulo. Aunque es un método gráfico y para aplicaciones de automatización puede ser complejo, es un primer paso a dar.

Por lo anterior, para datos los cuales se desconocen su naturaleza (bases de datos de viento) se utilizará el algoritmo k -means para separar como primera aproximación y se utilizarán los máximos locales para determinar el número de k . Esto con la intención de

separar un concepto que es de gran utilidad para entender la dinámica de viento y se definirá en este trabajo: estados de viento.

2.4 Estados de viento

Cuando se obtiene una base de datos de una región se puede graficar el histograma de frecuencias para la velocidad y si es posible de dirección. Si se realiza un ajuste al histograma se obtiene una función de probabilidad la cual tiene información sobre el parámetro de tendencia central y dispersión de los datos. Es decir, una región donde caen todas las mediciones de velocidad. Este mismo concepto se puede aplicar para la dirección. En la Figura 2.15 se muestra un histograma de dirección con los datos obtenidos por la estación meteorológica en la Rumorosa, Baja California. Se observa que existe acumulación de puntos.. También indica que existen ráfagas de viento proveniente de otras direcciones que no se presentan muy frecuente en el año.

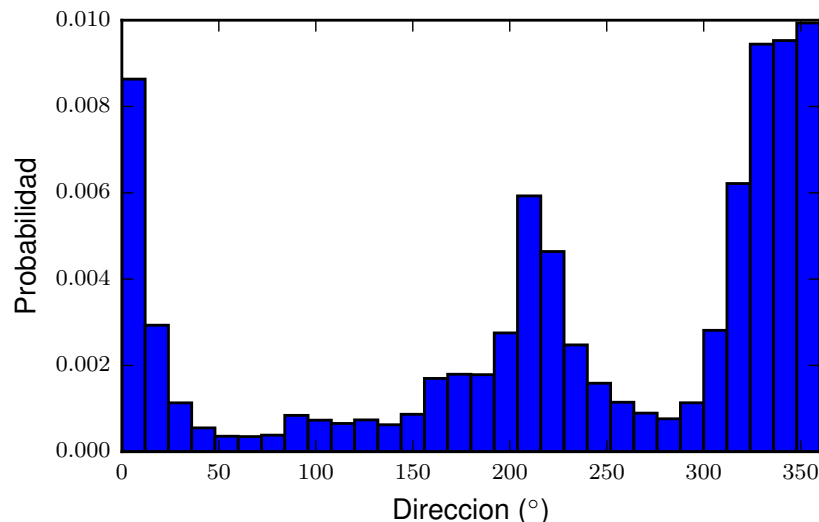


Figura 2.15: Histograma de dirección de los datos obtenidos por la estación meteorológica en la Rumorosa, Baja California.

En la literatura, ya se ha establecido que existen regiones de viento donde caen la mayor parte de las mediciones y a estas zonas se le conocen usualmente como dirección preferencial o túnel de viento, como por ejemplo lo encontrado en los artículos [2, 48, 49, 50]. Estas regiones son de gran importancia para en la generación de energía eléctrica. Cuando se instala un gran aerogenerador, el sistema de orientación se orienta hacia donde sople más el viento. Cuando existen ráfagas de viento, el sistema de orientación, en principio, las sigue si duran más de un tiempo determinado por el sistema de control. Sin embargo, las ráfagas pueden ser totalmente aleatorias y esto implicaría movimiento

innecesario que aumentaría los costos de mantenimiento y el desgaste de la componente, por otro lado las ráfagas pueden estar siempre dentro de una región bien definida. Si esto último es el caso, observaríamos una tendencia de acumulación de puntos en la dirección donde se presenten las ráfagas indicando una zona preferencial.

Usualmente, la técnica para visualizar el recurso eólico de una entidad es la rosa de los vientos como la que se observa en la Figura 2.16. En una rosa de los vientos se puede observar en un plano polar, la intensidad de puntos en dirección con su velocidad asociada. Mientras mayor sea la intensidad se puede suponer que es una dirección preferencial como por ejemplo la Figura 2.16: para este caso la dirección preferencial serían vientos provenientes del norte o nor-noroeste. La rosa de vientos es una herramienta muy útil para observar el comportamiento de la intensidad en dirección y velocidad. Sin embargo, mantiene oculta la dinámica interna de los datos así como la dispersión.

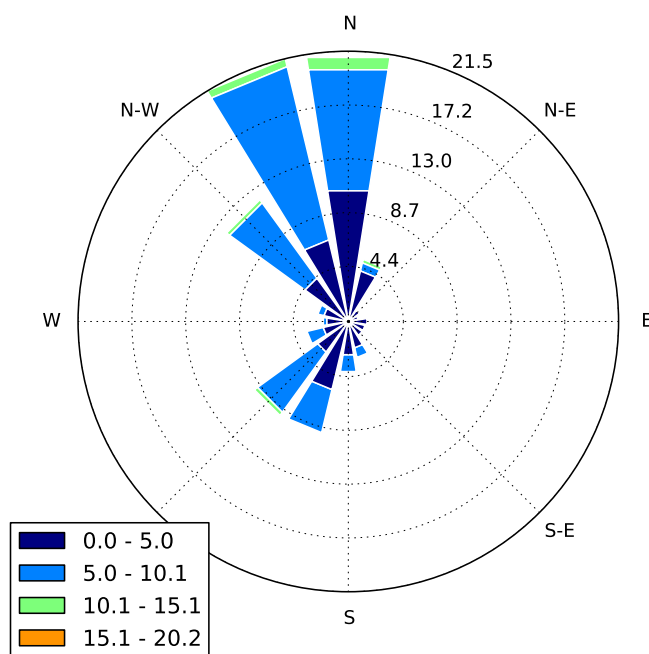


Figura 2.16: Rosa de los vientos para la localidad de La ventosa, Oaxaca. Las unidades están en m s^{-1}

Por lo anterior, se propone en este trabajo complementar la rosa de los vientos con una visualización en un campo vectorial. Dado que la dirección y velocidad se miden en el mismo instante, se puede visualizar el campo vectorial [51]. Un ejemplo de esta tipo de visualización se puede observar en la Figura 2.17. En la cual se puede observa dos conjuntos de datos aleatorios, independientes entre sí, que sobrepuestos generan regiones donde se acumulan valores. Para hacer esto con datos de velocidad y dirección es descomponer el vector de velocidad en coordenadas cartesianas. Considerando el vector de

velocidad V_i , la descomposición de vectores sigue la siguiente ecuación:

$$V_i = v_i(\cos \theta_i, \sin \theta_i),$$

en donde v_i y θ_i son los datos velocidad y dirección del viento en un paso de tiempo i . Con la ecuación anterior, la serie de tiempo de velocidad y dirección se transforma en una nueva colección de datos asociados. Graficando el resultado en un plano $v_x - v_y$ se obtienen la Figura 2.18. En la cual se observan cúmulos donde se encuentran la mayor cantidad de datos. Estos cúmulos son regiones donde el viento pasa un tiempo definido y son de gran interés para este trabajo.

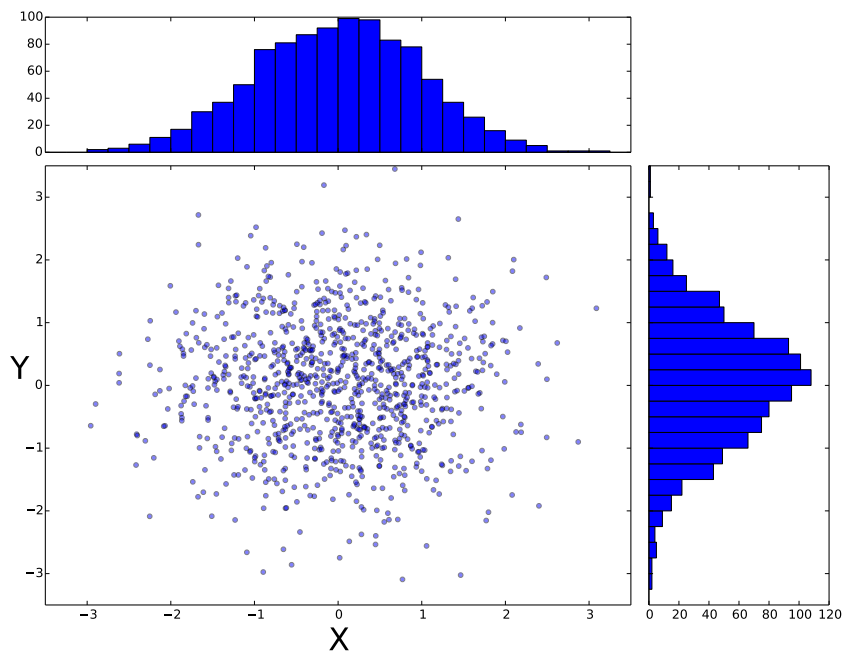


Figura 2.17: Histograma 2-D de dos variables aleatorias X y Y .

Por lo anterior, se encontró que la dinámica de viento puede ser aproximado utilizando estas regiones que denominaremos *estados de viento*. Llegando a la siguientes dos definiciones:

Definición 4 *El conjunto de vectores que se agrupan en una región del espacio de velocidades siguiendo una distribución de probabilidades Normal bi-variada se denominará estado de viento. La distribución de probabilidades se puede definir como la función de estado.*

Utilizando la definición anteriores se puede aproximar la dinámica de viento en estados discretos. Esto facilitará las simulaciones del desempeño de un aerogenerador en un sitio específico. Ya que, generará un espacio donde existan un número finito de estados y la dinámica del viento sea cambiar de estado de viento. También, permitirá conocer el

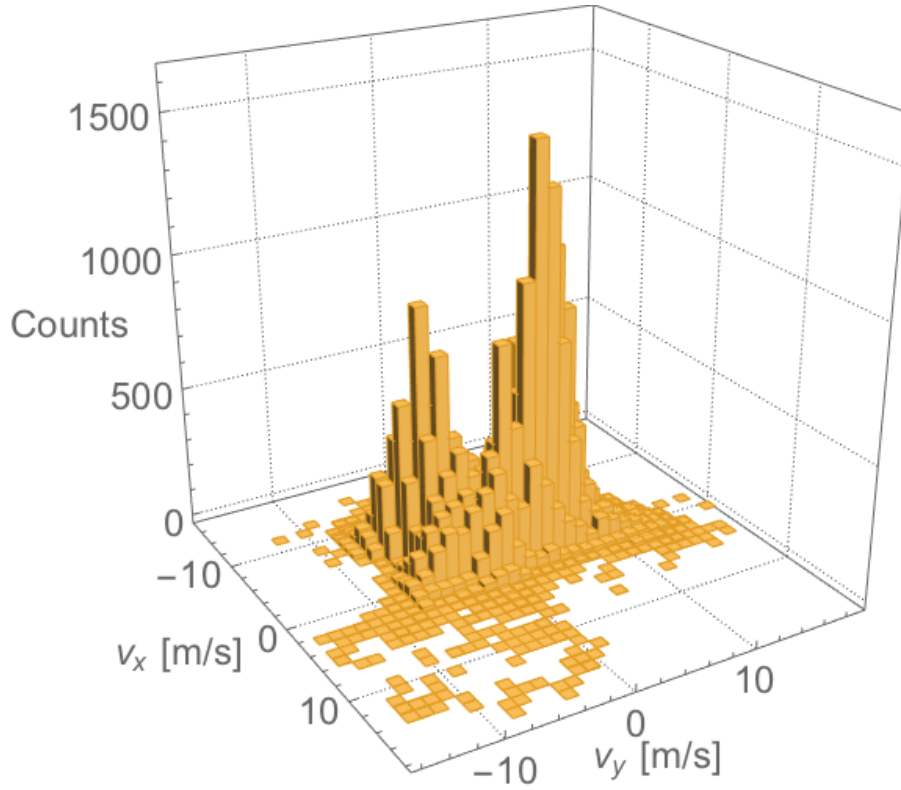


Figura 2.18: Histograma 2-D de datos de velocidad y dirección de viento.

tiempo que pasa el viento en una región específica y podrá capturar la componente de viento oculta en el método de estimación de energía tradicional.

La gran pregunta de esta definición es cómo clasificar los estados de viento. Una manera rápida y sencilla es utilizar los algoritmos de agrupamiento previamente explicados. De ser así, se podría utilizar como un método de clasificación de estados. Una vez clasificados, se puede recrear la serie temporal utilizando las cadenas de Markov continuas explicadas en la sección 2.2.2. Sin embargo, para hacer esto se tiene que hacer una suposición importante y es que los datos de cada cúmulo siguen una función de distribución normal bi-variada:

$$f(\vec{r}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{r} - \vec{\mu})^T \Sigma^{-1} (\vec{r} - \vec{\mu})\right), \quad (2.11)$$

donde $\vec{\mu}$ es el centroide de la distribución y Σ la matriz de covariancia, en la cual sus componentes tienen el valor de :

$$\Sigma_{i,j} = E[(r_i - \mu_i)(x_j - \mu_j)]. \quad (2.12)$$

El superíndice T indica el vector transpuesto y Σ^{-1} la matriz inversa de Σ . Para el caso

bi-dimensional, donde $\mu = (\mu_x, \mu_y)$ y la matriz Σ

$$\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}.$$

Aplicando lo anterior a datos de velocidad se puede encontrar las distribuciones de probabilidad de cada estado de viento.

2.5 Conclusiones

Los estados de viento pueden ser una herramienta para la estimación correcta del desempeño de un aerogenerador. A través de algoritmos de agrupamiento, se puede clasificar los estados de viento. Utilizando un modelo de normales bi-variadas de se puede realizar un ajuste de probabilidad a cada estado. La ventaja de implementar el concepto de estados de viento es poder desarrollar un algoritmo de seguimiento de estados, así reduciendo el porcentaje de uso de esta componente, reduciendo su desgaste, y los gastos de operación y mantenimiento.

Con el resultado de los ajustes, se puede recrear la serie temporal utilizando cadenas de *Markov* continuas en estados discretos. Esto ayudará al cómputo del desempeño general del aerogenerador. determinarán qué estados de viento aportan más energía y cuales de los estados de viento son transitorios.

En el siguiente capítulo, se utilizarán todos los conceptos explicados para proponer una metodología a seguir. La cual tendrá como resultado la clasificación de estados, la creación de un algoritmo de seguimiento parcial de estados y la generación de estados de viento.

3

METODOLOGÍA

En el tercer capítulo se establece la metodología que se utilizó para el desarrollo del trabajo. También las consideraciones necesarias para la interpretación de los resultados.

Contenido del capítulo

3.1. Casos de estudio	34
3.2. Desalineamiento	36
3.3. Cálculo de la energía	38
3.4. Cálculo de LCOE	42

Para llevar a cabo el cumplimiento de los objetivos de este trabajo se utilizará un modelo estocástico junto con algoritmos de agrupación para determinar los estados de viento. Haciendo posible la estimación del potencial eólico y el factor de planta para dos casos de estudio. Para cumplir esto se llevarán a cabo los siguientes pasos:

1. Se obtendrán los datos históricos de velocidad y dirección de viento de dos regiones de interés eólico,
2. se convertirán los datos en un campo vectorial para que a través de un algoritmo de agrupación poder clasificarlo por estados,
3. se aproximará cada estado con una distribución de probabilidad *Gaussiana* bi-variada y se obtendrán los centroides y la matriz Σ .
4. se propondrá un algoritmo de seguimiento del centroide de los estados de viento para reducir el uso del sistema de orientación y obtener una disminución en el costo nivelado de energía (LCOE, por sus siglas en inglés),
5. con los parámetros de ajuste para cada estado, se creará una serie de tiempo utilizando una cadena de *Markov* continua con estados discretos,
6. se determinará la matriz de flujos (q_{ij}) para conocer el tiempo de vida de cada estado.

Todos los puntos anteriores tienen como intención complementar la estimación de energía de una región. Adicionalmente esto puede ayudar a disminuir el LCOE optimizando una componente y tener una herramienta para calcular el desempeño en una región en particular. Aplicando la metodología se obtendrá los estados de viento, un algoritmo de seguimiento parcial, cuáles son los de mayor energía y su tiempo de vida asociado, una serie temporal de estados y la estimación del LCOE para diferentes modos de operación del aerogenerador.

Para este trabajo se utilizarán dos bases de datos de dos regiones con potencial eólico en México. Ambas sitios han estado en la mira de empresas generadoras de energía para colocar campos eólicos. Las características estadísticas del viento en cada caso de estudio son totalmente distintas una de otra como se presentará a continuación.

3.1 Casos de estudio

Para hacer la estimación del recurso eólica de una región, primero, se deben de conocer los datos históricos de la localidad. En este trabajo se utilizaron datos de la región del Istmo de Tehuantepec o La Ventosa y Baja California Sur. Aunque los lugares se encuentran a gran distancia uno de otro, ambos tienen un potencial eólico evidente. Por

Tabla 3.1: Resultados de ajustes para las distribuciones de velocidad de los casos de estudio. Para el caso de la ventosa se utilizó una función *Weibull & Weibull* [7] y para Baja California una *Weibull* [8]

Ubicación	Ajuste
Ventosa	$k_1 = 1.674, c_1 = 4.034$ $k_2 = 5.232, c_2 = 16.097$
B. Calif	$k = 1.34, c = 4.70$

ejemplo, en la Ventosa entidades como el National Renewable Energy Laboratory (NREL, por sus siglas en inglés) han determinado su gran potencial y se han hecho mapas eólicos de la región [19].

Se ha encontrado que en Baja California se pueden obtener FP cercanos a 25 % [8] en comparación con la Ventosa casi de 50 % [7]. En las dos regiones existen aerogeneradores instalados, pero el potencial eólico de Baja California es mucho menor —energéticamente hablando. Está relacionado con la velocidad máxima que alcanza los vientos de la región y los cambios de dirección. A continuación se muestran en la Figura 3.1 los histogramas de frecuencias de velocidad y dirección de cada región ^{1 2}.

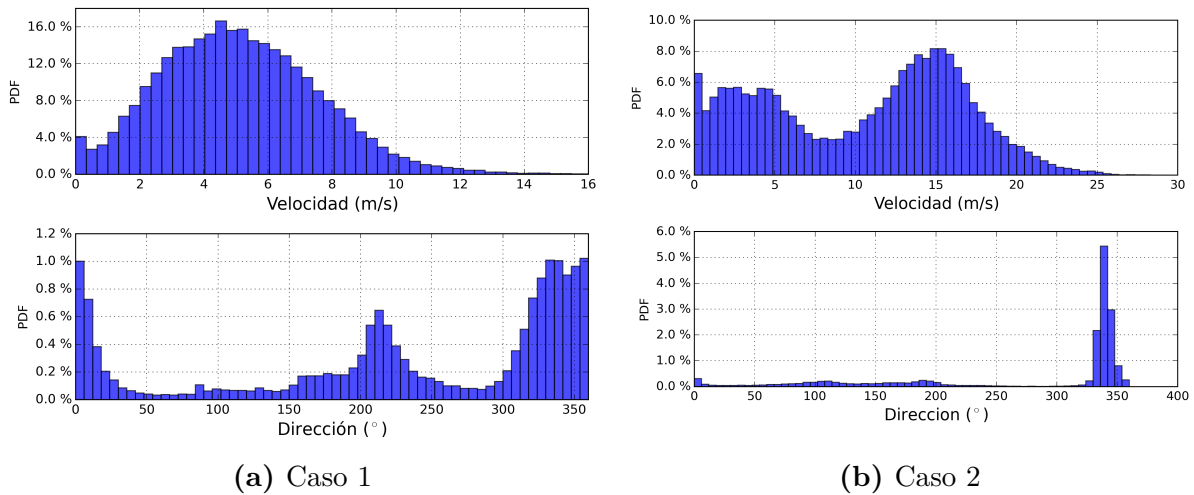


Figura 3.1: Histogramas de frecuencias de velocidad y dirección de los casos de estudios.

Como se puede apreciar en la Figura 3.1, los lugares tienen propiedades estadísticas diferentes. El histograma de la Figura 3.1a el viento sigue una distribución *Weibull* y para el caso de la Ventosa una suma de dos distribuciones *Weibull* o *Weibull & Weibull*. En la Tabla 3.1 se muestran los resultados de los ajustes al histograma respectivos.

En ambos casos la dirección juega un papel importante. Como se puede observar en

¹Ver Anexo A.3 para las series de tiempo completas

²La serie de tiempo fue extrapolada a la altura de interés para el caso 1, 80 m y para el caso 2, 60 m utilizando el método de potencias ver Anexo A.2

el histograma de dirección de la Figura 3.1b la mayor parte del viento proviene de una región específica entre $315 - 340^\circ$. Esto indica la existencia de un estado preferencial. Por otro lado se encuentra la Figura 3.1a en donde se puede observar que existen por lo menos dos regiones con probabilidad alta de aparecer. Así mismo, ésta región tiene mayor cantidad de cambios en dirección (Ver Figura A.3b).

Para determinar los estados de viento se utilizará el método k -means explicado en la sección 2.3. Una vez clasificados, se ajustará la ecuación (2.11)

3.2 Desalineamiento

La estimación de energía con el método tradicional supone que el viento siempre llega de una manera normal al aerogenerador. Para las localidades en donde existe una dirección preferencial es buena aproximación sin seguimiento, sin embargo, para localidades donde el viento cambia mucho de dirección utilizar este método puede llevar a un error sustancial y una mala estimación del recurso. Para poder incluir la dependencia angular se requiere utilizar la ecuación (2.7) la cual nos da como resultado una curva en función del ángulo de desalineamiento del rotor. Si se realiza este cálculo para un intervalo de ángulos se obtiene una sabana de potencia en donde el eje x sera velocidad, el eje y la dirección y z la potencia como se puede observar en la Figura 3.2.

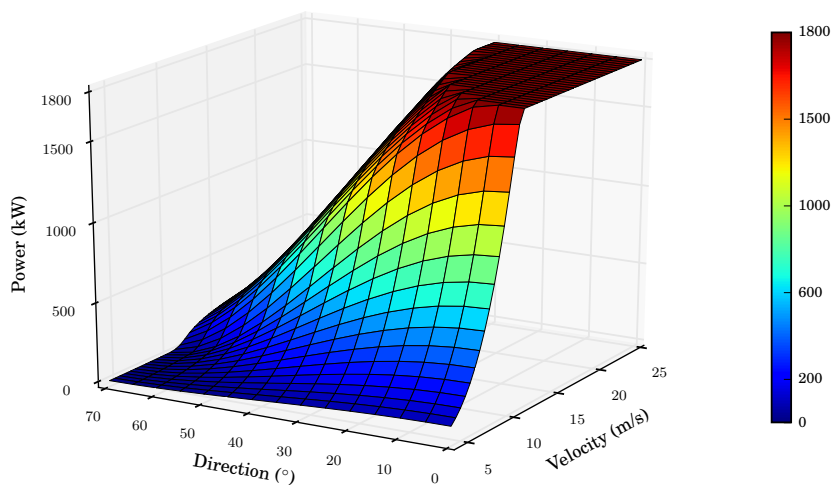


Figura 3.2: Sabana de potencia del aerogenerador.

Si observamos el ángulo cero de la Figura 3.2 obtenemos la curva ilustrada en la Figura 3.5. Esta sabana representa el desempeño real del aerogenerador. Utilizar esta sabana para estimar el potencial tiene más sentido si se va a tomar en cuenta el desalineamiento.

De igual manera que la curva de potencia, la distribución de probabilidad no brinda información del comportamiento del ángulo para los datos. Observando detenidamente la serie de tiempo de dirección de viento (ver Anexo A.3) resalta el cambio temporal de la dirección en ambos lugares y esto nos da una dinámica temporal de cambios en dirección y da un indicio a una dinámica de estados.

Si el viento cambia de dirección el aerogenerador se tiene que re-orientar de tal forma que la incidencia sea normal. Esto implica un desgaste en el sistema de engranes del aerogenerador, un gasto energético y una mayor probabilidad de falla como se explicó en la sección 2.1.4.

Por tanto, es pertinente incluir el ángulo a la hora de calcular la cantidad de energía y en los algoritmos de seguimiento de los grandes aerogeneradores. Para poder observar la dinámica temporal del ángulo y velocidad basta con graficar los datos de velocidad contra los datos de dirección en el plano vectorial $v_x - v_y$.

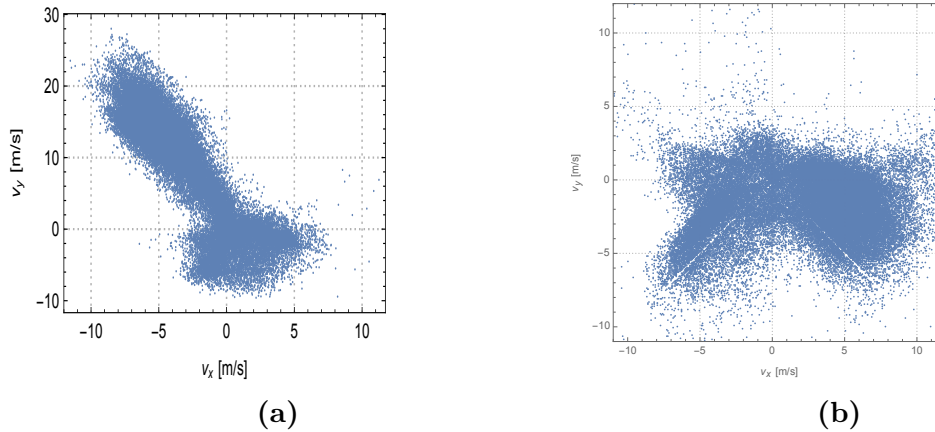


Figura 3.3: Plano vectorial v_x-v_y , representando la formación de cúmulos para los casos de estudio. a) Caso de estudio 1, b) caso de estudio 2

Ambos casos de estudios presentan cúmulos. Dependiendo de las propiedades estadísticas del cúmulo su potencial energético será mayor o menor. Esto servirá como indicativo para determinar los estados que produzcan una mayor cantidad de energía. Encontrar el estado con mayor energía puede indicar la región de viento óptima para generar energía, sin embargo esto se tiene que empatar con histograma de frecuencias del cúmulo de la Figura 3.4 la cual servirá para determinar el número de cúmulos.

Una vez que cada dato de la serie tenga un cúmulo asociado se puede ajustar una función bi-variada a cada cúmulo. Utilizando la ecuación (2.11) se puede aproximar la distribución de datos dentro del cúmulo. Cuando se tenga la discretización por cúmulos se puede recrear la serie temporal discreta utilizando las ecuaciones para cadenas de *Markov* continuas de la sección 2.2.2. Debido a que es una serie de tiempo, utilizar cadenas de *Markov* continuas que involucran un serie de temporal es lógico. Sin embargo, el resultado de número de estados es discreto.

Para recrear la serie se requiere la matriz de transición anual a partir de todos los

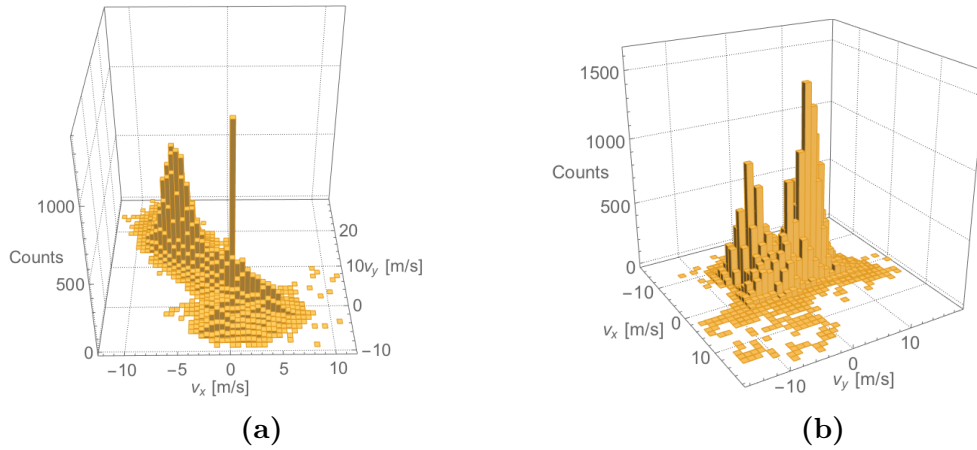


Figura 3.4: Histogramas 3D. a) Caso de estudio 1, b) caso de estudio 2

datos y como se definió en la ecuación (2.8) se requiere contar los cambios de estados y normalizarlos por el total de una fila. Con la matriz de transición se calcula la matriz q con la ecuación (2.2.2) para determinar los flujos entre estados y así generar la cadena de *Markov*. Es importante recordar que los dos últimos objetivos la matriz de transición de probabilidad y la matriz q son claves.

3.3 Cálculo de la energía

Para estimar la energía generada de manera tradicional utilizando las series de tiempo, se requiere la curva de potencia de un aerogenerador. Actualmente en el mercado existe una gama amplia de empresas que fabrican aerogeneradores de gran capacidad y la tendencia actual es a mayores capacidades. Una curva de potencia típica, usada para su estudio en este trabajo se observa en la Figura 3.5. Con la gráfica de potencia del aerogenerador también se obtiene la de C_p . Estas dos curvas sirven para entender el comportamiento del C_p y su evolución cuando incrementa la velocidad. Esta gráfica como se mencionó anteriormente no considera desalineamiento, es decir, $\gamma = 0$.

Como ya se habló en la sección 2.1 la curva del aerogenerador representa la dependencia de la potencia con respecto a la velocidad. Para determinar la energía utilizando la distribución de probabilidad se requiere resolver la siguiente integral:

$$E = \int_{t=0}^{t=T} \int_{v_a}^{v_c} P(v)f(v)dvdt, \quad (3.1)$$

siendo $P(v)$ la curva de potencia del aerogenerador, $f(v)$ la distribución de probabilidad, v la velocidad y t el tiempo. Si se resuelve la ecuación 3.1 para el intervalo de tiempo de un año y para el rango de velocidades $[v_a, v_c]$ se obtiene energía generada en el año. Para

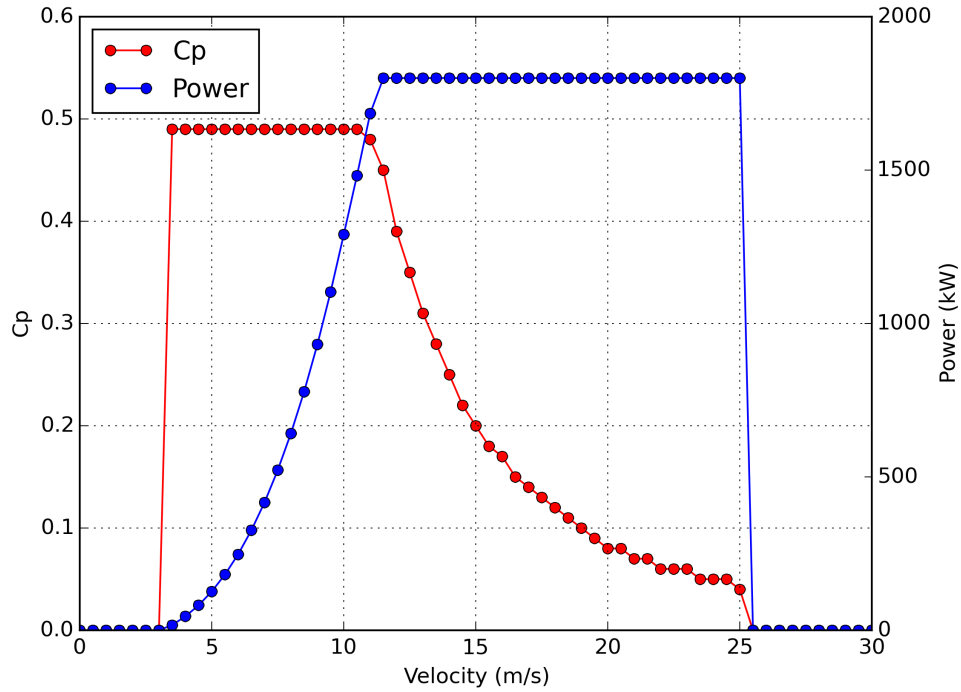


Figura 3.5: Curva de potencia y C_p del aerogenerador de 1.8 MW de estudio.

ambos casos de estudio, el resultado de esta multiplicación da los valores de la Tabla 3.2.

Tabla 3.2: Energía anual producida en GW h para las dos localidades utilizando el método de estimación de energía tradicional. Para el caso de Baja California se utilizó una $f(v)$ tipo *Weibull* y para Oaxaca una *Weibull* bi-variada [7]

B. California [GW h]	Oaxaca [GW h]
2.62	8.842

Este método tiene una suposición muy importante que el ángulo de desalineamiento es cero y esto no se cumple en todo el año. Analizando la curva de potencia del aerogenerador, resalta que no existe ninguna dependencia angular. Esto implica necesariamente que γ es cero, es decir, que no existe desalineamiento (caso ideal).

Con fines de comparación de la metodología propuesta en este trabajo se calcularán cuatro energías con diferentes metodologías que se nombrarán como: ideal, fija, real, parcial y *Weibull* como se muestra en la Tabla 3.3. Cada método requiere la serie de velocidad e información adicional dependiendo la complejidad del método. Para el método ideal, la suposición principal es que la energía que produce el aerogenerador es igual a la evaluación de la función de potencia utilizando el punto de velocidad en el momento i ,

Tabla 3.3: Métodos para la estimación de potencial eólico.

Método	Ecuación	Datos	Información adicional
Ideal (E_0)	$\Delta T \int_0^\infty P(v)f(v)dv$	v	Ninguna
Real (E_r)	$\Delta T \int_0^\infty P(v, \gamma)f(v)dvd\gamma$	v, γ	Algoritmo de seguimiento
Parcial (E_p)	$n\Delta T \int_0^\infty P(v, \gamma)f(v)dvd\gamma$	v, γ, n	Estados de viento
<i>Weibull</i> (E_w)	$\Delta T \int_0^\infty P(v)f(v)dv$	v	Parámetros de ajuste

es decir,

$$E_0 = \sum_{i=0}^n P_i(v)\Delta T.$$

Esta energía se denominará ideal (E_0) y será el punto de comparación, ya que se considera que la producción de energía solo es función de la velocidad instantánea, recordando la sabana de potencia ilustrada en la Figura 3.2, equivale a la curva cuando $\gamma = 0$ o la curva de potencia del aerogenerador.

La segunda energía que se calculará es la fija E_f . Consiste en colocar el aerogenerador apuntando hacia la dirección preferencial (γ_o) de la zona. No se considera ningún sistema de seguimiento, es decir,

$$E_f = \sum_{i=0}^n P_i(v, \gamma_o)\Delta T.$$

La tercera energía que se consideró fue la Real (E_r). Consiste principalmente en la generación de un aerogenerador “real” bajo condiciones de operación, utilizando un algoritmo de seguimiento que se explicará posteriormente, pierde potencia por el desalineamiento y se orienta automáticamente para incrementar la potencia. De igual forma que E_0 , la energía será la suma de la función de potencia dependiente del ángulo y velocidad en un intervalo i , es decir,

$$E_r = \sum_{i=0}^n P_i(v, \gamma)\Delta T.$$

La cuarta energía considerada y la propuesta de este trabajo es la parcial (E_p). Consiste en optimizar el seguimiento de la energía real para capturar la mayor cantidad de viento posible sin girar demasiado. Para esto, se requiere separar el campo de velocidad y dirección en estados de viento, y el cálculo será el mismo que la energía, pero sumando la energía de todos los estados, es decir,

$$E_p = \sum_{i=0}^N \sum_{j=0}^n P_{ij}(v, \gamma)\Delta T,$$

siendo N el número total de estados de viento y P_{ij} la sabana de potencia del aerogenerador.

Para fines de comparación, se utilizará una quinta energía determinada por los parámetros de *Weibull* (E_w). Utilizando los parámetros de ajuste obtenidos por Jaramillo and Borja [7], Jaramillo et al. [8] se calculará la energía,

$$E_w = \int_0^{\infty} P(v) f_w, \quad (3.2)$$

como se menciona en la sección 2.1.4 el sistema de orientación puede fallar 3 a 4 veces por año.

Para complementar la simulación de la estimación de energía, se utilizará un modelo de falla reparables de *Poisson* homogéneo [52]. Este es un modelo muy general que establece las siguientes ecuaciones:

$$P\{N(t) = k\} = \frac{(\lambda T)^k e^{-\lambda T}}{k!}, \quad (3.2)$$

$$m(t) = \lambda, \quad (3.3)$$

donde k será el número de fallas que presenta el aerogenerador en un intervalo de tiempo T bajo una tasa de fallas $m(t)$. Para completar se usará la tasa de fallas descrita en la sección 2.1.4.

Para poder concluir el cálculo de la energía se explicará las suposiciones de los algoritmos de seguimiento³ utilizados. Los algoritmos de seguimiento de un aerogenerador no están disponibles a detalle en la literatura, sin embargo se pueden encontrar las consideraciones generales y casos típicos. Hau and Von Renouard [2] mencionan que el sistema de orientación es casi instantáneo. Para un caso típico si se detecta un cambio de dirección por más de dos minutos el sistema de orientación se activa. Adicionalmente menciona las consideraciones que debe tomar los límites de operación del aerogenerador:

1. Tiempo de ráfaga,
2. Velocidad de corte (v_{max}),
3. Velocidad angular mínima del aerogenerador ($v_{\theta,min}$),
4. Desalineamiento máximo permitido (γ_{max}).

Para su análisis en este trabajo se consideran los siguientes valores:

- $\gamma_{max} = 8^\circ$.
- $v_{max} = 25 \text{ m s}^{-1}$.
- $v_{\theta,min} = 4 \text{ m s}^{-1}$

³Algunos ejemplos de criterios generales de seguimiento [2],

Un algoritmo de seguimiento estándar que se usará en este trabajo es el siguiente: para cada valor de velocidad y dirección en la base de datos y se determinará el desalineamiento γ , si la velocidad y dirección están dentro de los límites de operación, se asumirá que el aerogenerador entrega potencia, del caso contrario la potencia es igual a cero y el aerogenerador se tiene que re-orientar para el siguiente instante de tiempo. De igual forma, el algoritmo de seguimiento parcial tendrá los mismos límites de operación. Lo diferente de esta propuesta es que la orientación será directamente al centroide del estado. Por lo que el desalineamiento se calculará con la dirección actual y el valor del centroide (ver Algoritmo en el Anexo A.4).

3.4 Cálculo de LCOE

Para poder determinar el LCOE se tienen que hacer suposiciones (Ver Anexo A.1 para definición del costo nivelado de la energía). La primera suposición es que el aerogenerador del caso de estudio tendrá una vida útil de 20 años. En este tiempo producirá una cantidad de energía dependiendo del recurso y número de fallas por el sistema de orientación. Existen muchas publicaciones sobre los valores de capital y operación y mantenimiento como [53, 54, 55]. Con base en los parámetros propuestos por Ragnarsson et al. [55], se harán las siguientes consideraciones. El Costo de capital (CAPEX) será asumido como un valor constante \$2229 USD/kWh, los gastos de operación (OPEX) son \$25 USD/kWh con una inflación de 2.20 % cada año. Para los OPEX variables, se consideró \$0.0034 USD/kWh. También se utilizará una tasa de descuento fija de 10 %.

En el siguiente capítulo se muestran los resultados de los valores obtenidos de FP y LCOE para cada una de las energías estimadas. También se mostrará la nueva serie temporal de estados utilizado las cadenas de *Markov* y las matrices de flujo q resultantes.

4

RESULTADOS

En el cuarto capítulo se muestran todos los resultados obtenidos utilizando la metodología del capítulo anterior.

Contenido del capítulo

4.1. Resultado agrupamiento de estados	44
4.2. Resultado estimación de energía	44
4.3. Resultado de LCOE	44
4.4. Simulador de viento	46
4.5. Discusión de resultados	47

4.1 Resultado agrupamiento de estados

Con los datos mostrados en las gráficas de Figura 3.4 se agrupó la información en estados de viento. Por simplicidad, se utilizó el algoritmo k -means (discutido en la sección 2.3). Considerando un espacio de estados continuo se usó el algoritmo de agrupamiento dos veces para cada caso. La primera separación fue utilizando un número de $k = 2$ y después $k = 3$ dando un total de $k = 5$ ¹. Lo anterior dio como resultado las siguientes particiones ilustradas en las gráficas de la Figura 4.1a. Utilizando un número de cúmulos igual a cinco, se logró obtener la separación de los picos importantes detectados en los histogramas 2D de la Figura 3.4.

4.2 Resultado estimación de energía

Se utilizaron cinco métodos para estimar la energía en cada caso de estudios. Se multiplicó el resultado por una eficiencia general del aerogenerador de 35 % tomada de [3], así como los parámetros de disponibilidad (K_a) y factor de desempeño (K_p). Para la estimación de energía con el método *Weibull* se utilizaron los parámetros de escala obtenidos de los ajustes a las distribuciones de probabilidad de las respectivas referencias (ver sección 3). Adicionalmente, se incluyó un factor de falla para el sistema de orientación utilizando el modelo de *Poisson* homogéneo [52]. Debido a falta de información, se consideró que el intervalo de tiempo entre la falla y la reparación es despreciable. Con base en las estadísticas de [26], se consideró una tasa de fallas por sistema de orientación de cuatro eventos por año. Cada falla repercute en dos horas fuera de operación dentro de la simulación.

Utilizando la energía de cada método se determinó el factor de planta con la ecuación (2.5). Los resultados de la estimación de energía con su respectivo factor de planta se observan en la Tabla 4.1.

4.3 Resultado de LCOE

Utilizando las consideraciones hechas en la sección 3.4, se obtuvieron valores económicos para cada método. Los resultados obtenidos para LCOE se muestran en la Tabla 4.2. El valor de LCOE fue calculado con la serie temporal de los parámetros de la ecua-

¹Si se utiliza $k = 5$ en la primera separación el resultado no separaba los picos importantes encontrados gráficamente. Esta es una característica del algoritmo k -means

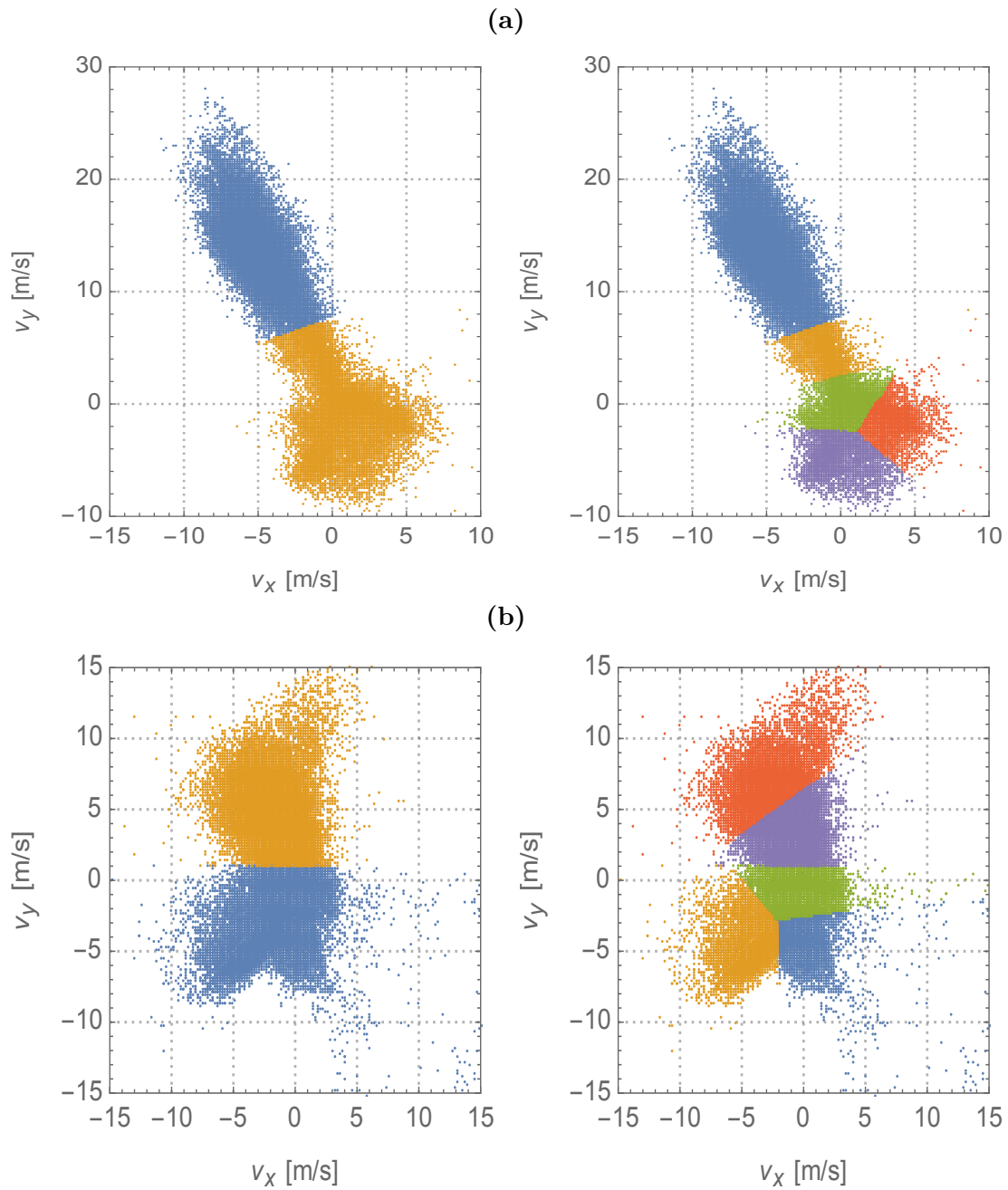


Figura 4.1: Agrupación en estados de viento del espacio de velocidades para el caso de estudio 1 (a) y 2 (b), utilizando el algoritmo k -means.

Tabla 4.1: Producción anual de energía (AEP [GWh]) y factor de planta (%) utilizando diferentes métodos.

Método	Caso 1		Caso 2	
	AEP	FP	AEP	FP
Weibull	8.79	51-58 ^a	2.62	16.6
Ideal	10.19	58.3	11.74	25.8
Real	9.85	56.4	4.28	24.4
Parcial	9.90	57.0	3.84	22.0
Fijo	9.68	55.3	2.13	12.2

^a El FP para el caso de estudio 2 depende de la distribución de probabilidad que se utilice *Weibull* o *Weibull & Weibull* [7].

ción (A.1) para el tiempo de vida del aerogenerador (20 años). Ver Tabla A.9 para la simulación completa.

Tabla 4.2: Costo nivelado de energía utilizando diferentes métodos de estimación de energía

Método	LCOE (¢USD (2012)/kWh)	
	Caso 1	Caso 2
Real	4.60	26.97
Parcial	4.51	55.88
Fijo	8.32	172.72

4.4 Simulador de viento

Con los estados de viento se puede calcular la matriz de transición entre estados utilizando la ecuación (2.8) y determinar la matrices q con la ecuación (2.2.2). En ambos casos de estudio, la matriz de probabilidad de transición y la matriz de flujos q es de 5x5 estados —ver anexo A.8 para resultados de matrices. Con los valores de la matriz q se creo la cadena de *Markov* continua para los dos casos de estudios como se muestra en la Figura 4.2

Utilizando la matriz q se realizó un diagrama de *Markov* para ambos casos de estudios como se ilustra en la Figura 4.3. En la Figura 4.3 se observan gráficamente los flujos de la matriz q , mientras más intensa sea la fecha, mayor es el flujo. Adicionalmente, se calculó el tiempo de vida por estado (Ver Anexo A.6)

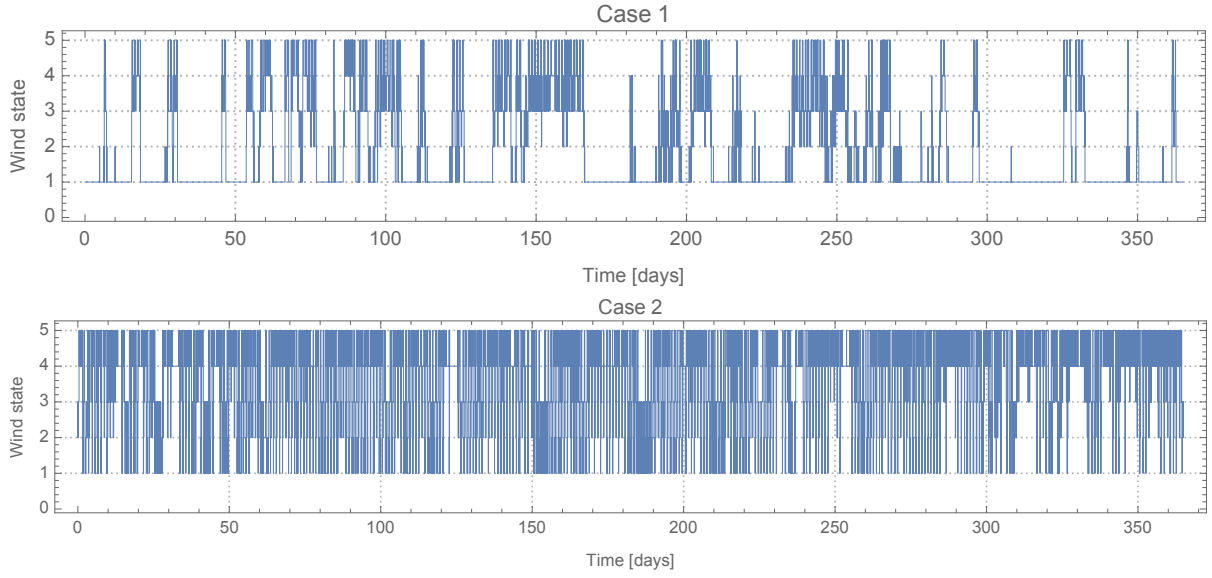


Figura 4.2: Cadena de *Markov* para casos de estudios

4.5 Discusión de resultados

La contribución principal de los estados de viento es para mejorar la manera tradicional de estimar el recurso eólico. Evidentemente, la definición de estados que se presentó en este trabajo es únicamente basada en el hecho que el conjunto de datos de velocidad siguen una distribución *Gaussiana* a partir de los cúmulos. Esto no representa ninguna relación física de los estados en sí. Sin embargo, puede ser una herramienta útil para determinar la energía de cada estado. Suponiendo que el estado i sigue una función *Gaussiana* bi-variada $f_i(v_x, v_y)$, como se mencionó en la sección 2.4, la potencia por unidad de área de cada estado se puede calcular utilizando:

$$\frac{P_i}{A} = \frac{1}{2}\rho \int_0^\infty \int_0^\infty |\vec{v}|^3 f_i(v_x, v_y) dv_x dv_y, \quad (4.1)$$

donde $|\vec{v}| = (v_x^2 + v_y^2)^{1/2}$ y ρ es la densidad del aire con un valor promedio de 1.21 kg/m^3 , a nivel del mar.

Si m es el número total de estados, la potencia total del sitio es simplemente la suma de todos los estados $P = \sum_{i=1}^m P_i$ y la potencia relativa por estado puede aproximarse como $p_i^* = P_i/P$. La potencia relativa es importante debido a que brinda información del peso del estado en todo el sitio. Es el indicador que ayuda a diferenciar los estados más energéticos.

Otra información relevante con este método es la dispersión angular. Un estado ideal sería aquel que su dispersión angular sea pequeña, es decir toda la energía se concentra

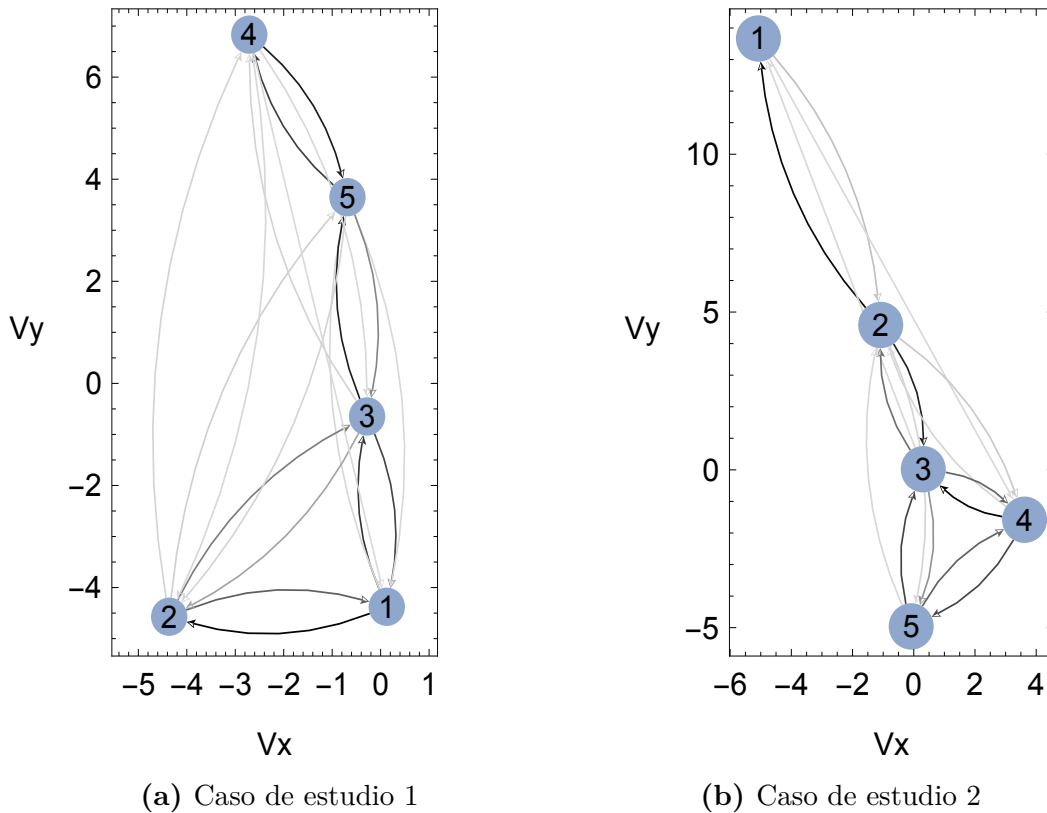


Figura 4.3: Diagrama de *Markov* para los casos de estudio

en una región angosta de dirección.

Podemos definir la *apertura del estado* (α) como el ángulo que se forma con la elipse de la i -ésima componente *Gaussiana* del estado que abarca el 90% de los datos del estado de viento, y representa la región angular donde el viento sopla con mayor probabilidad, como se muestra en la Figura 4.4.

Esta información complementa el método tradicional, porque se puede obtener una gráfica en donde se represente la dispersión del estado y su potencia relativa, como se observa en la Figura 4.5. Las gráficas de la Figura 4.5 son de suma importancia. Ya que indican los estados con su potencia relativa y la dispersión de los estados. Esta información puede servir para optimizar el algoritmo de seguimiento de viento. Como se mencionó en la sección 3.3, un algoritmo estándar sigue al viento cada dos minutos aproximadamente. Si existe una gran dispersión en el estado, puede producir muchos movimientos innecesarios que desgastarán los rodamientos del aerogenerador e inclusive puede ser que se esté siguiendo un estado de viento de poca energía.

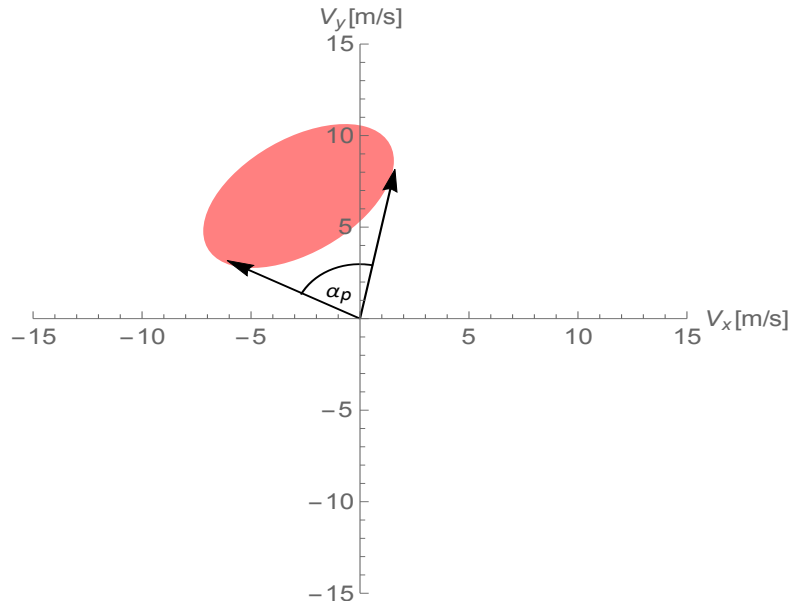


Figura 4.4: Concepto de apertura de estado α_p para el estado cuatro del caso de estudio dos.

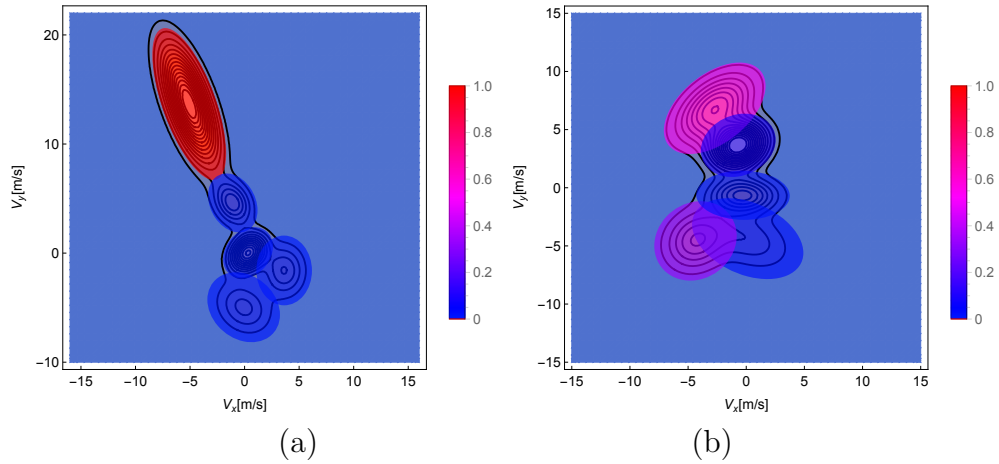


Figura 4.5: Regiones elípticas donde las componentes normales de cada estado cubre el 90 % de los datos asociados a el para el caso de estudio uno (a) y dos (b). Las elipses se colorearon utilizando la potencia relativa de cada estado p_i^* con la escala mostrada en la paleta de colores.

Con la información anterior, se programó un algoritmo de seguimiento para que buscara los centroides de los estados de viento e identificara los estados con mayor potencia relativa y poca dispersión angular.

Utilizando la información del espacio de velocidad se puede recrear la generación de energía utilizando cuatro métodos de estimación propuestos en la sección 3.3. Como se observa en la Tabla 4.1, el método de producción ideal es la mayor energía. Lo anterior, debido a que no considera ningún tipo de pérdidas por dirección, se calcula la potencia

punto a punto para todo el conjunto de datos. Para fines prácticos, este método se puede utilizar como referencia y punto de comparación de varios métodos de producción. Utilizar el método *Weibull* tiene sus ventajas y desventajas. La primera ventaja es que es método más rápido de hacer ya que no requiere mucha información detallada del sitio. Sin embargo, la desventaja es que no se considera pérdidas por desalineamiento. Utilizando el método de *Weibull* o tradicional se obtuvo una sub-estimación del recurso eólico en comparación con el método para ambos casos de estudios. que para fines económicos puede ser crucial para el desarrollo de un campo eólico

Los siguientes métodos simulan las condiciones reales de operación de un aerogenerador. Es evidente, que si no se tiene un estado con poca apertura, el fijo es un mal método de captura de energía. Para el caso de estudio 2, aunque se orientó al estado con mayor potencia relativa, no capta la mayor parte la energía ocasionando un bajo factor de planta y poca generación. En el caso de estudio 1, como se tiene un estado de alta velocidad promedio y poca dispersión angular, y al orientarlo hacia su centroide, la generación es comparable con los otros dos métodos.

Los métodos parcial y real son similares. Sin embargo, el seguimiento parcial parece mejorar la captación de energía en regiones donde la dispersión de estados es pequeña como el caso 1, y fallar contra el real en regiones donde la dispersión angular por estado es mayor. Para el caso de estudio 2, los factores de planta fueron mucho mayores a los estimados utilizando el método *Weibull*. Lo anterior se debe a que las características estadísticas del sitio, no logran reproducir la dinámica completa de la región. Por otro lado, en el caso de estudio 1 se tiene una sobre estimación del factor de planta utilizando el método *Weibull & Weibull* y sub-estimación con *Weibull*.

Para probar la eficacia del seguimiento parcial, se calculó el LCOE. Los resultados de la Tabla 4.2 muestran que existe una reducción en el costo de la energía para el caso de estudio 1. Sin embargo, existe un aumento para el caso de estudio 2 debido a la gran diferencia entre el factor de potencia. Es importante mencionar, que el algoritmo parcial tiene ventajas adicionales sobre el real. Por ejemplo, al reducirse el número de movimientos del sistema de orientación al año, los costos por operación y mantenimiento reducen. Si existen menos movimientos, la probabilidad de fallas del sistema de seguimiento disminuye, aumentando el tiempo de generación del aerogenerador. Ragnarsson et al. [55] menciona que los factores que más afecta al LCOE son la disponibilidad del aerogenerador, si reducimos la contribución de pérdidas por del sistema de seguimiento, el LCOE será mayor. Así mismo, la reducción de costos de mantenimiento y operación ayudarán a reducir aún más el LCOE.

5

CONCLUSIONES GENERALES

En el último capítulo se presentan las conclusiones generales del trabajo y se discute los resultados obtenidos.

La metodología convencional para estimar la energía que produce un aerogenerador se basa en el uso de la función de distribución de *Weibull* del sitio y la curva de potencia del aerogenerador. Utilizando los parámetros de ajuste se puede estimar la energía, sin embargo éste método no considera los efectos angulares ni pérdidas por desalineamiento. Se demostró que dependiendo del sitio y el método de estimación de energía se puede obtener una sub- o sobre- valoración del recurso. Con la finalidad de tener una metodología de evaluación del recurso eólico que considere la dirección se presentó en este trabajo el concepto de estados de viento.

El concepto de estados de viento es una herramienta de gran utilidad para el análisis del viento. Se acopla con herramientas matemáticas como los algoritmos de agrupamiento, distribuciones de probabilidad para simular los datos en un estado de viento, cadenas de *Markov* para simular los datos de viento en un espacio discreto, entre otras posibles aplicaciones. También se puede utilizar los estados de viento para descubrir corrientes de viento principales. Determinando la apertura del estado de viento se puede estimar los estados con mayor potencial eólico. Así mismo, se pueden localizar las regiones angulares en donde no es recomendable orientar el aerogenerador.

El concepto de estados de viento como se definió en este trabajo es puramente geométrico debido a la implementación de algoritmo de agrupamiento *k-means*. Cada uno de los algoritmos de agrupamiento –en principio– puede ser aplicado al concepto de estados de viento, pero como se mencionó en el capítulo 2.3 depende del resultado esperado. Si bien, el algoritmo *k-means* es utilizado en problemas diversos por su simplicidad, para el caso de energía eólica es necesario buscar un algoritmo de agrupamiento que relacione las variables físicas o matemáticas involucradas en el sistema. Una primera aproximación sería incluir la velocidad y dirección de viento, adicionalmente se podría incluir más variables meteorológicas, con el fin de darle un origen a la formación de estados de viento.

Para demostrar una aplicación posible del concepto de estados de viento se analizaron dos casos de estudio con características estadísticas de viento diferentes. En el primero se encontró que existe un estado principal que contribuye con el 90 % de la energía disponible del sitio. Adicionalmente este estado de viento tiene poca apertura angular. En el segundo caso existen dos estados que aportan más del 80 % de la energía disponible, y se observó que la apertura promedio de los estados de viento del caso dos fue de 60°. Mientras mayor sea la apertura del estado de viento es más probable que el aerogenerador se desoriente con mayor frecuencia. Otra aplicación de los estados de viento es un algoritmo optimizado para el sistema de orientación.

Es importante mencionar que los valores de LCOE para el caso de estudio 2 son muy altos porque el sitio tiene velocidades promedio menores que el caso de estudio 1. Por lo que el aerogenerador la mayor parte de tiempo trabaja debajo del P_n . Para llevar a cabo una mejor comparación se puede utilizar un aerogenerador de menor potencia para el caso de estudio 2. Sin embargo, la información es útil para demostrar la aplicación de estados de viento.

Usar el sistema de orientación en estados de viento en donde existe un apertura

angular considerable puede propiciar desgaste innecesario en los rodamientos principales. Lo anterior implicaría un aumento considerable en la contribución al OpEx del sistema de orientación y, dentro de un futuro en el CapEx. Utilizando el algoritmo de seguimiento de estados se logró reducir un 71 % el uso del sistema de orientación en una simulación de 20 años de operación.

Para poder simular los datos de viento utilizando un estado de viento es necesario asociar variables estadísticas al estado. Se utilizaron funciones normales bi-variadas para aproximar la medida de tendencia central. Al ser geométrica la separación de los estados de viento, el valor de tendencia central de las normales bi-variadas ajusta con el centroide del cúmulo. Sin embargo, un mejor ajuste por alguna otra distribución puede ser encontrado.

Aplicar las cadenas de *Markov* a estados de viento ayudó a recrear la dinámica de transición entre estados. Dentro de las cadenas de *Markov* las matrices q juegan un papel muy importante. Es la matriz que da información sobre el flujo entre estados.

Las matrices q de datos de viento parecen ser una buena huella de las características del viento y podría ser utilizada para tomar decisiones de proyectos eólicos. Proporcionan la información de flujo o cambio de estados. Podrían dar información relacionadas con las ráfagas de viento y estados transitorios.

Regresando al problema de la estimación correcta del recurso, utilizar esta aproximación podría lograr identificar regiones donde existan múltiples estados con mayor potencia relativa. Se puede utilizar una matriz q como patrón para buscar en bases de datos lugares con matrices q similares que representarán ubicaciones con potencial eólico similar a la región de donde se calculó la matriz q . Todo lo anterior abre múltiples líneas de investigación que se abordaran en trabajos futuros.

BIBLIOGRAFÍA

-
- [1] The inside of a wind turbine, 2014. URL <http://goo.gl/ArVUNv>. Fecha de consulta: 03-02-2016.
 - [2] E. Hau and H. Von Renouard. *Wind Turbine Costs*. Springer, 2006.
 - [3] T. Burton, D. Sharpe, N. Jenkins, and E. Bossanyi. *Wind energy handbook*. John Wiley & Sons, 2001.
 - [4] V. A. P. Hines, A. Ogilvie, and C. Bond. Crew database wind plant reliability benchmark (WEU summit 2013). Technical report, Sandia National Laboratories, 2013.
 - [5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, Mathieu Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.
 - [6] R. Wiser, Z. Yang, M. Hand, O. Hohmeyer, D. Infield, P. H. Jensen, V. Nikolaev, M. O'Malley, G. Sinden, and A. Zervos. 2011: Wind energy. in IPCC special report on renewable energy sources and climate change mitigation. Technical report, Intergovernmental Panel on CLimate Change, 2011. URL <http://goo.gl/ngr1Mo>. Fecha de consulta: 3-03-2016.
 - [7] O. A Jaramillo and M. A Borja. Wind speed analysis in La Ventosa, México: a bimodal probability distribution case. *Renewable Energy*, 29(10):1613–1630, 2004.
 - [8] O. A Jaramillo, R Saldaña, and U Miranda. Wind power potential of Baja California Sur, México. *Renewable Energy*, 29(13):2087–2100, October 2004.
 - [9] T. Ashuri. *Beyond classical upscaling: integrated aeroservoelastic design and optimization of large offshore wind turbines*. PhD thesis, TU Delft, Delft University of Technology, 2012.
 - [10] Renewable Energy Policy Network. Renewables 2005 global status report. *Washington, DC: World-watch Institute*, 2005.
 - [11] Secretaría de Energía. Balance nacional de energía, 2014. URL <http://goo.gl/JM03EY>. Fecha de consulta: 03-04-2016.
 - [12] Asociación Mexicana de Energía Eólica. El potencial eólico mexicano, 2013. URL <http://goo.gl/r0hxjq>. Fecha de consulta: 03-03-2016.
 - [13] G. T Bellarmine and J. Urquhart. Wind energy for the 1990s and beyond. *Energy Conversion and Management*, 37(12):1741–1752, 1996.
 - [14] S. Eriksson, H. Bernhoff, and M. Leijon. Evaluation of different turbine concepts for wind power. *Renewable and Sustainable Energy Reviews*, 12(5):1419–1434, 2008.
 - [15] R. Lanzafame and M. Messina. Fluid dynamics wind turbine design: Critical analysis, optimization and application of bem theory. *Renewable energy*, 32(14):2291–2305, 2007.
 - [16] J. F. Manwell, J. G. McGowan, and A. L. Rogers. Wind turbine design. *Wind Energy Explained: Theory, Design and Application*, 2002.
 - [17] L. J. Fingersh, M. M. Hand, and A. S. Laxson. Wind turbine design cost and scaling model. Technical report, National Renewable Energy Laboratory, 2006.

-
- [18] L. Landberg, L. Myllerup, O. Rathmann, E. L. Petersen, B. H. Jørgensen, HJ. Badger, and N. G. Mortensen. Wind resource estimation—an overview. *Wind Energy*, 6(3):261–271, 2003.
- [19] D. Elliott, M. Schwartz, G. Scott, S. Haymes, D. Heimiller, and R. George. Wind energy resource atlas of oaxaca. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO., 2003.
- [20] I.E.L.P Troen and E. L Petersen. European wind atlas. Technical report, 1989.
- [21] IEC-61400-12-1. Wind turbines - Part 12-1: Power performance measurements of electricity producing wind turbines. Standard, International Electrotechnical Commission, 2000.
- [22] MEASNET Procedure. Evaluation of site specific wind conditions, 2009.
- [23] W. Weibull. Wide applicability. *Journal of applied mechanics*, 103, 1951.
- [24] F. D. Bianchi, H. De Battista, and R. J. Mantz. Wind turbine control systems: principles, modelling and gain scheduling design, 2006.
- [25] M. G. Kim and P. H. Dalhoff. Yaw systems for wind turbines—overview of concepts, current challenges and design methods. In *Journal of Physics: Conference Series*, volume 524, page 012086. IOP Publishing, 2014.
- [26] S. Sheng. Report on wind turbine subsystem reliability—a survey of various databases. *National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/PR-5000-59111*, 2013.
- [27] N. T. J. Bailey. *The elements of stochastic processes with applications to the natural sciences*, volume 25. John Wiley & Sons, 1990.
- [28] J. L. Doob. *Stochastic processes*, volume 101. New York Wiley, 1953.
- [29] C. M. Grinstead and J. L. Snell. *Introduction to probability*. American Mathematical Soc., 2012.
- [30] L. J.S. Allen. *An introduction to stochastic processes with applications to biology*. CRC Press, 2010.
- [31] D. Lusseau. Effects of tour boats on the behavior of bottlenose dolphins: Using markov chains to model anthropogenic impacts. *Conservation Biology*, 17(6):1785–1793, 2003.
- [32] H. Balzter. Markov chain models for vegetation dynamics. *Ecological Modelling*, 126(2-3):139–154, February 2000.
- [33] G. Tauchen. Finite state markov-chain approximations to univariate and vector autoregressions. *Economics letters*, 20(2):177–181, 1986.
- [34] D. Quah. Empirical cross-section dynamics in economic growth. *European Economic Review*, 37(2):426–434, 1993.
- [35] J. E. Caskey. A markov chain model for the probability of precipitation occurrence in intervals of various length. *monthly Weather Review*, 91(6):298–301, 1963.
- [36] K. R. Gabriel and J. Neumann. A markov chain model for daily rainfall occurrence at tel aviv. *Quarterly journal of the Royal Meteorological Society*, 88(375):90–95, 1962.
- [37] A. D. Sahin and Z. Sen. First-order markov chain approach to wind speed modelling. *Journal of Wind Engineering and Industrial Aerodynamics*, 89(3):263–269, 2001.

- [38] F. Castino, R. Festa, and C. F. Ratto. Stochastic modelling of wind velocities time series. *journal of Wind Engineering and industrial aerodynamics*, 74:141–151, 1998.
- [39] F. Y. Ettoumi, H. Sauvageot, and A. H. E Adane. Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution. *Renewable Energy*, 28(11):1787–1802, 2003.
- [40] H. Aksoy, Z. F. Toprak, A. Aytok, and N. E. Ünal. Stochastic generation of hourly mean wind speed data. *Renewable Energy*, 29(14):2111–2131, 2004.
- [41] A. Shamshad, M. Bawadi, W. Wanhussin, T. Majid, and S. Sanusi. First and second order Markov chain models for synthetic generation of wind speed time series. *Energy*, 30(5):693–708, April 2005.
- [42] G. DÁmico, F. Petroni, and F. Prattico. Wind speed modeled as an indexed semi-markov process. *Environmetrics*, 24(6):367–376, 2013.
- [43] G. DÁmico, F. Petroni, and F. Prattico. Wind speed and energy forecasting at different time scales: A nonparametric approach. *Physica A: Statistical Mechanics and its Applications*, 406:59–66, 2014.
- [44] Z. Song, Y. Jiang, and Z. Zhang. Short-term wind speed forecasting with markov-switching model. *Applied Energy*, 130:103–112, 2014.
- [45] V. Estivill-Castro. Why so many clustering algorithms: A position paper. *SIGKDD Explor. Newsl.*, 4(1):65–75, 2002.
- [46] R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2): 411–423, 2001.
- [47] M. Yan. Methods of determining the number of clusters in a data set and a new clustering criterion. 2005.
- [48] F. Castino, L. Rusca, and G. Solari. Wind climate micro-zoning: a pilot application to liguria region (north western italy). *Journal of wind engineering and industrial aerodynamics*, 91(11):1353–1375, 2003.
- [49] A. Clifton and J. K. Lundquist. Data clustering reveals climate impacts on local wind phenomena. *Journal of Applied Meteorology and Climatology*, 51(8):1547–1557, 2012.
- [50] L. G. Glascoe, R. E. Glaser, H. N Chin, and G. A. Loosmore. Regional-scale wind field classification employing cluster analysis. In *13th Conference on the Applications of Air Pollution Meteorology with the Air and Waste Management Association, Session*, 2004.
- [51] M. F. Al-Roby and A. M. El-Halees. Data mining techniques for wind speed analysis. *Journal of Computer Eng*, 2(1), 2011.
- [52] M. Natrella. Nist/sematech e-handbook of statistical methods. 2010.
- [53] A. C. Moné, B. M. Smith, and M. Hand. 20143 cost of wind energy review. Technical report, National Renewable Energy laboratory, 2015.
- [54] D. Gielen. Renewable energy technologies: cost analysis series – wind power. Irena working paper, International Renewable Energy Agency (IRENA), 2012.
- [55] B. F. Ragnarsson, G. V. Oddsson, R. Unnthorsson, and B. Hrafnkelsson. Levelized cost of energy analysis of a wind power generation system at búrfell in iceland. *Energies*, 8(9):9464–9485, 2015.

A

ANEXO

Contenido del capítulo

A.1. Costos en turbinas eólicas	60
A.2. Método de potencia	61
A.3. Series de tiempo	62
A.4. Algoritmos de seguimiento	63
A.5. Distribución normal bivariada para cada estado	64
A.6. Matriz de transición de estados	65
A.7. Tiempos característicos	66
A.8. Matriz q	66
A.9. Simulación de energía y costos	67
A.10. Artículo publicado	70

A.1 Costos en turbinas eólicas

Cada componente de la turbina tiene un costo, y esto es importante a la hora de analizar un proyecto eólico. Saber cuánto costará generar energía eléctrica considerando todos los gastos que implica colocar y mantener en funcionamiento una turbina eólica es crucial. Por tanto, se requiere conocer la relación energía/costo de una turbina eólica para poder compararla con otras tecnologías.

Un indicador de referencia para evaluar la amortización del costo de una planta de generación eléctrica es el costo nivelado de energía (LCOE, por sus siglas en inglés). El LCOE es el precio de la electricidad que amortiza el costo. En otras palabras, da una idea global de cuánto cuesta generar electricidad con cualquier tecnología. Económicamente hablando, un LCOE alto en una tecnología indicaría un tiempo mayor del retorno de capital, mientras un valor menor indica una tasa menor de retorno. El LCOE varía de tecnología como se muestra en la Figura A.1.

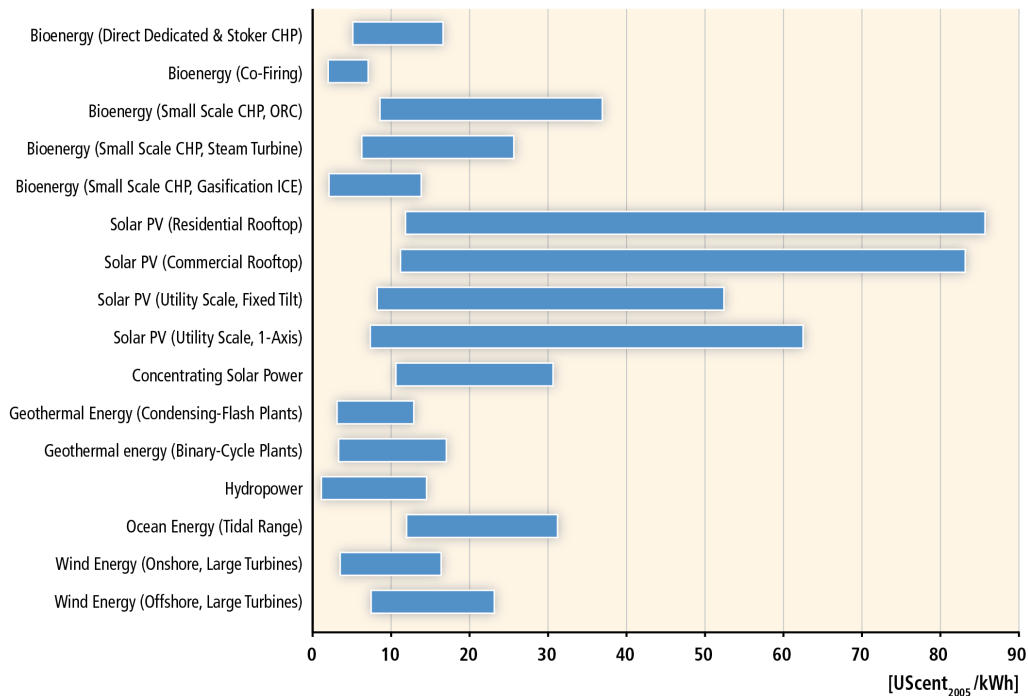


Figura A.1: Costo nivelado de Energía (LCOE) de cada tecnología disponible en el mercado. Extraído en su totalidad de [6]

La definición matemática del LCOE puede variar dependiendo de las variables que se consideren. Una forma generalizada de la ecuación está disponible en [54], la cual se

usará en este trabajo y establece:

$$\text{LCOE} = \frac{\sum_{t=1}^n \frac{I_t + M_t + F_t}{(1+r)^t}}{\sum_{t=1}^n \frac{E_t}{(1+r)^t}} \quad (\text{A.1})$$

donde I_t son los gastos de capital (CAPEX, por sus siglas en inglés), M_t los gastos de operación y mantenimiento (OPEX, por sus siglas en inglés), F_t el costo de combustible, n el tiempo de vida del proyecto, E_t es AEP y r el costo medio ponderado de capital (WACC, por sus siglas en inglés). Como se observa en la Figura A.1, la tecnología eólica *onshore* tiene un rango de LCOE de 5-10 US¢/kWh y para *offshore* de 10-20 US¢/kWh.

Para estimar el LCOE se necesita conocer los costos involucrados en toda la instalación del aerogenerador. Por ejemplo, las turbinas eólicas mayores a 1 MW poseen componentes grandes y de gran peso que tienen gastos de capital y operativos considerables. En la Tabla A.1 de [9] se puede observar el desglose de costos aproximados y su peso por componente para aerogeneradores.

Tabla A.1: Costos asociados por cada componentes de un aerogenerador de 5 MW con USD del 2009. Adaptado de [9]

Componentes	Costo (1000 USD)	Masa (kg)
Aspas	1062.3	17480.0
Cabeza (Hub)	130.2	60540.0
<i>Pitch</i> y rodamientos	242.0	28878.0
Cono (Nose cone)	13.6	1810.0
Eje de baja velocidad (Low speed shaft)	166.8	16256.0
Rodamientos principales (Main bearings)	64.4	5400.0
Caja de engranes (Gearbox)	877.2	39688.0
Freno mecánico y acoplamiento (Mechanical brake and coupling)	11.0	1053.0
Generador (Generator)	398.0	17623.0
Electrónica de potencia (Power electronics)	393.2	-
<i>Yaw</i> y rodamientos	146.3	13152.0
Tablero principal (Main frame)	162.7	31773.0
Conexiones eléctricas (Electrical connections)	308.8	-
Sistema hidráulico y enfriamiento (Hydraulic and cooling system)	77.2	423.7
Sistema de control y seguridad (Control, Safety and conditions monitoring)	65.3	-
Torre	1491.3	596520.0
<i>Marinization</i>	939.1	-
Total	4722.2	830,0596.7

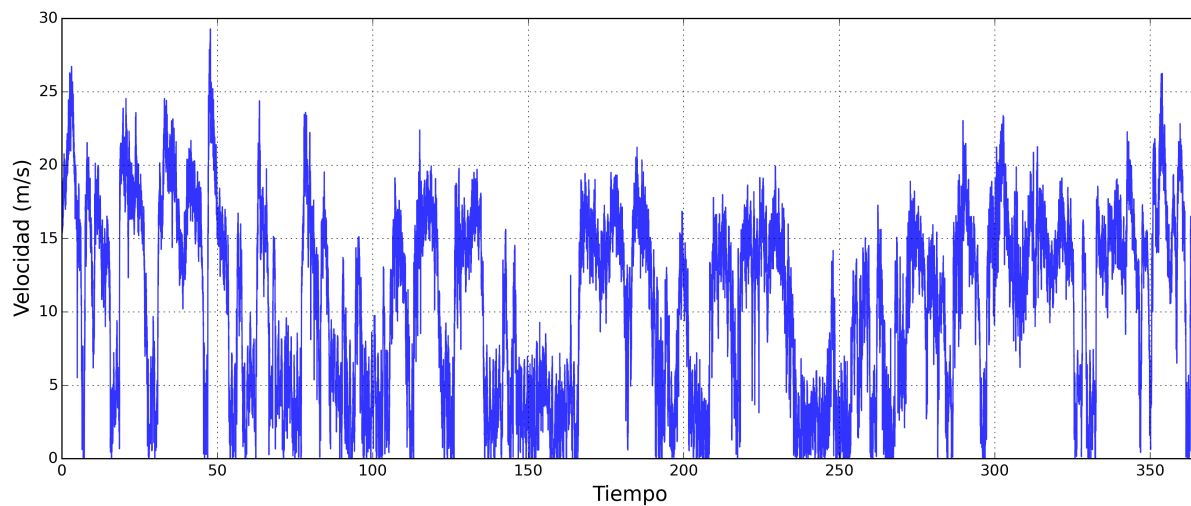
A.2 Método de potencia

Si se tienen datos de velocidad de viento a una altura de referencia z_1 y se quiere conocer los valores a una altura z_2 , una buena aproximación es utilizarla ley de potencias [9]:

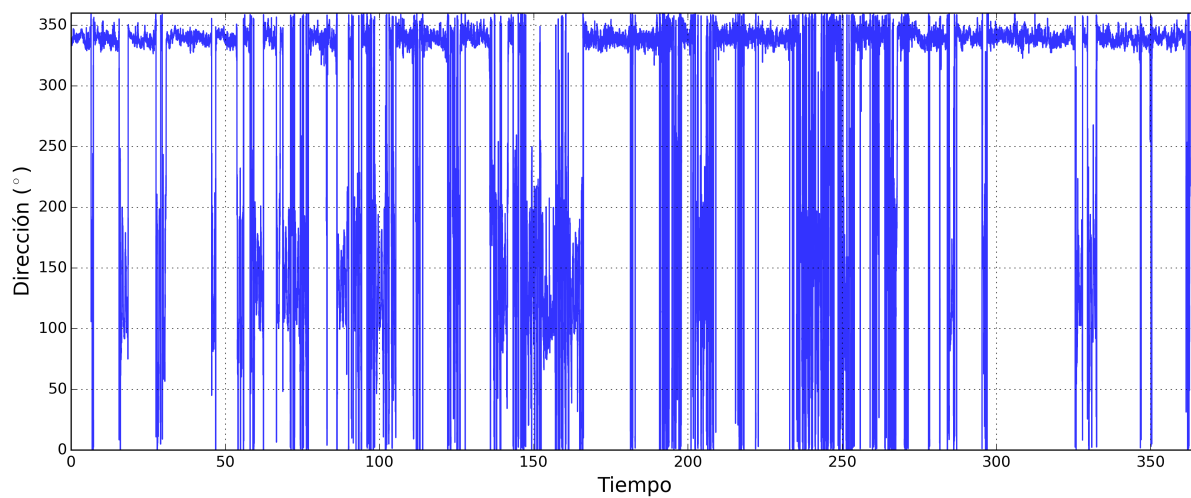
$$v_2 = v_1 \left(\frac{z_2}{z_1} \right)^\alpha, \quad (\text{A.2})$$

donde v_1 es velocidad en la altura z_1 , v_2 velocidad en la altura z_2 , z_1 altura 1 (baja), z_2 altura 2 (alta) y α coeficiente de capa límite valor dependiente del sitio. Valores α típicos son de 0.1 a 0.2. Para lugares donde la rugosidad es pequeña, se puede aproximar el coeficiente de capa límite como $\frac{1}{7}$ [7].

A.3 Series de tiempo

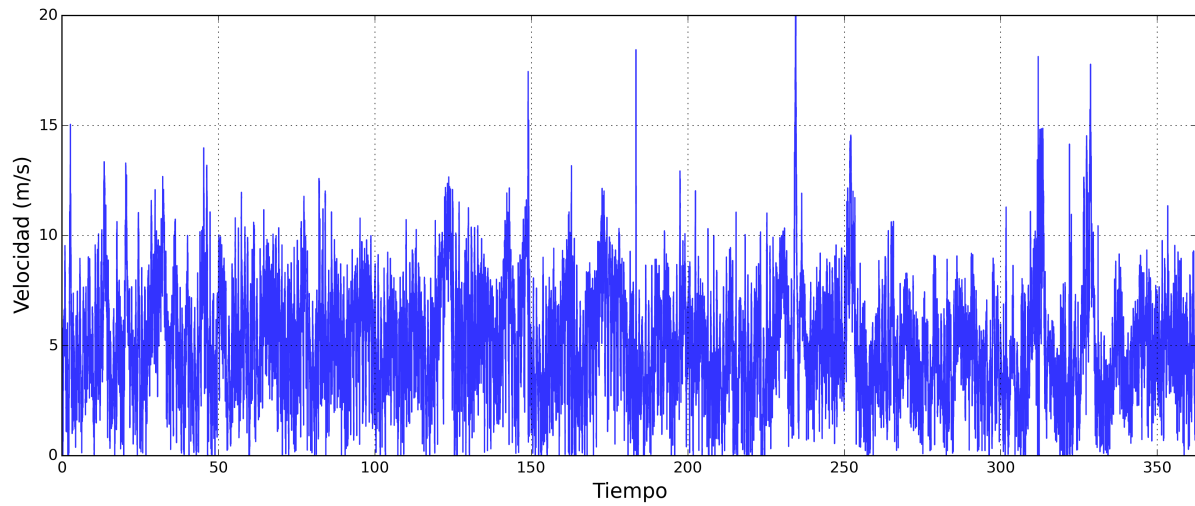


(a) Velocidad

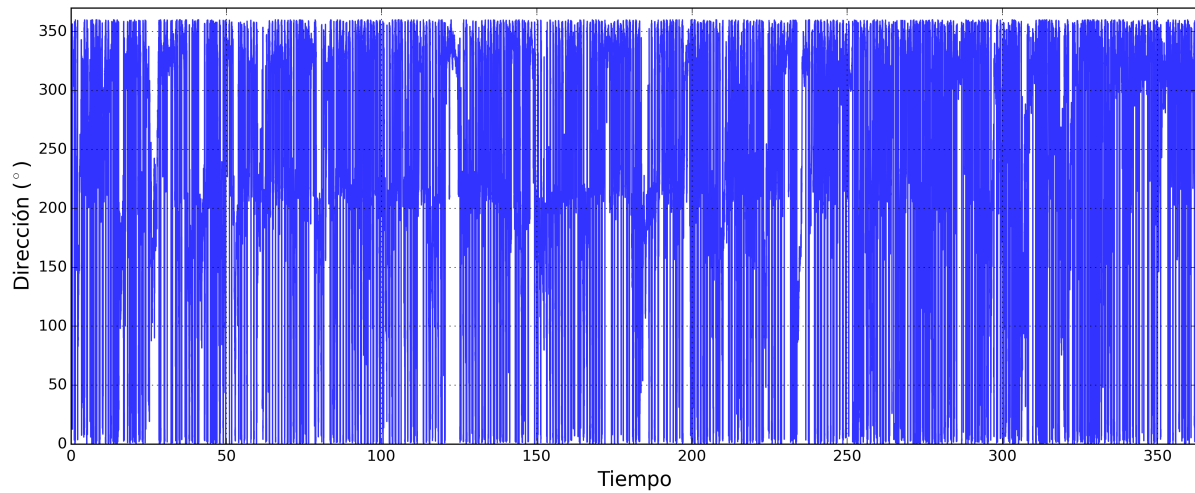


(b) Dirección

Figura A.2: Serie de tiempo de velocidad y dirección para caso de estudio 1



(a) Velocidad



(b) Dirección

Figura A.3: Serie de tiempo de velocidad y dirección para el caso de estudio 2.

A.4 Algoritmos de seguimiento

Una aplicación adicional de los estados de viento fue un algoritmo de seguimiento parcial el cual considera lo siguiente:

Algorithm 1 Algoritmo de seguimiento parcial

```

1:  $\phi_0 =$  Estado inicial
2: for  $i \leftarrow 1, n$  do
3:    $\gamma \leftarrow |\theta_i - \theta(\phi_o)|$ 
4:   if conditions then
5:      $P_i \leftarrow f(v_i, \gamma)$ 
6:   else
7:      $P_i \leftarrow 0$ 
8:      $\phi_o = \phi_{i+1}$ 
9:   end if
10: end for

```

A.5 Distribución normal bivariada para cada estado**Tabla A.2:** Parámetros del ajuste utilizando normales bi-variadas para el caso de estudio 1

Estado	$\mu_x[m/s]$	$\mu_y[m/s]$	$\Sigma_{xx}[(m/s)^2]$	$\Sigma_{xy}[(m/s)^2]$	$\Sigma_{yx}[(m/s)^2]$	$\Sigma_{yy}[(m/s)^2]$
1	-5.0746	13.6675	2.4915	-3.3270	-3.3270	10.5284
2	-1.0804	4.5985	1.1161	-0.3334	-0.3334	1.5873
3	0.3163	0.0234	1.1334	0.2858	0.2858	1.2095
4	3.6258	-1.5895	1.4108	-0.0093	-0.0093	2.2184
5	-0.0843	-4.9619	2.4047	-0.5016	-0.5016	2.2301

Tabla A.3: Parámetros del ajuste utilizando normales bi-variadas para el caso de estudio 2

Estado	$\mu_x[m/s]$	$\mu_y[m/s]$	$\Sigma_{xx}[(m/s)^2]$	$\Sigma_{xy}[(m/s)^2]$	$\Sigma_{yx}[(m/s)^2]$	$\Sigma_{yy}[(m/s)^2]$
1	0.1269	-4.3762	4.9710	-1.2755	-1.2755	2.6258
2	-4.3630	-4.5714	2.7266	0.3326	0.3326	2.5311
3	-0.2794	-0.6473	3.5139	-0.0438	-0.0438	0.9712
4	-2.7106	6.8333	4.1543	1.8232	1.8232	3.3550
5	-0.6847	3.6476	2.2533	0.3284	0.3284	1.6114

A.6 Matriz de transición de estados

Tabla A.4: Matriz de transición de estados de la cadena de *Markov* obtenida de la información del caso 1.

Estado	1	2	3	4	5
1	0.9921	0.0079	0.0000	0.0001	0.0000
2	0.0680	0.8665	0.0590	0.0049	0.0015
3	0.0001	0.0341	0.9056	0.0376	0.0226
4	0.0000	0.0010	0.0701	0.8815	0.0473
5	0.0000	0.0002	0.0485	0.0386	0.9127

Tabla A.5: Matriz de transición de estados de la cadena de *Markov* obtenida de la información del caso 2.

Estado	1	2	3	4	5
1	0.8600	0.0794	0.0600	0.0000	0.0006
2	0.0446	0.9158	0.0338	0.0021	0.0037
3	0.0549	0.0191	0.8548	0.0021	0.0690
4	0.0001	0.0008	0.0004	0.9287	0.0701
5	0.0002	0.0016	0.0306	0.0569	0.9107

A.7 Tiempos característicos

Tabla A.6: Tiempo de estadía total t_r y tiempo medio de vida t_m de las cadenas de *Markov* para los dos casos de estudio.

Estado	Caso 1		Caso 2	
	t_r (días)	t_m (días)	t_r (días)	t_m (días)
1	232.417	0.868973	37.3542	0.0496071
2	26.9444	0.0520163	50.7361	0.0825203
3	50.9583	0.0735329	53.3403	0.0477641
4	26.5486	0.0586062	100.861	0.0973563
5	28.0694	0.0795168	122.701	0.0777575

A.8 Matriz q

Tabla A.7: Matriz q de la cadena de *Markov* para los datos del caso 1

Estado	1	2	3	4	5
1	-1.1508	1.1421	0.0000	0.0087	0.0000
2	9.7979	-19.2247	8.4990	0.7052	0.2227
3	0.0196	4.9060	-13.5993	5.4162	3.2576
4	0.0000	0.1507	10.0947	-17.0630	6.8177
5	0.0000	0.0356	6.9827	5.5576	-12.5760

Tabla A.8: Matriz q de la cadena de *Markov* para los datos del caso 2

Estado	1	2	3	4	5
1	-20.1584	11.4311	8.6470	0.0000	0.0803
2	6.4236	-12.1182	4.8670	0.2956	0.5320
3	7.9238	2.7602	-20.9362	0.3004	9.9518
4	0.0099	0.1091	0.0595	-10.2716	10.0931
5	0.0326	0.2363	4.4009	8.1906	-12.8605

A.9 Simulación de energía y costos

Tabla A.9: Resultados de la simulación de un año para el caso de estudio 1.

Year	Real	Parcial	Fijo	C _{p,r}	C _{p,p}	C _{p,f}	OPEX Fijo _r	OPEX Fijo _p	OPEX VAR _r	OPEX VAR _p	CAPEX	LCOE _{r,1}	LCOE _{r,2}	LCOE _{p,1}	LCOE _{p,2}	LCOE _{f,1}	LCOE _{f,2}
	GWh	GWh	GWh	%	%	%	USD/kW	USD/kW	\$/kW	\$/kW	eUSD/kW	eUSD/kW	eUSD/kW	eUSD/kW	eUSD/kW	eUSD/kW	e USD/kW
1	9.854	9.889	9.677	0.564	0.566	0.554	45	36	0.003	0.002	2229	2063.524	5043.482	2055.356	5079.15	3895.215	4863.338
2	9.856	9.892	9.677	0.564	0.566	0.554	45.99	36.792	0.003	0.002	0	37.873	4578.496	30.299	4611.396	31.19	4413.193
3	9.856	9.891	9.677	0.564	0.566	0.554	47.002	37.601	0.003	0.002	0	35.124	4154.511	28.1	4184.198	26.248	4004.712
4	9.861	9.896	9.677	0.564	0.566	0.554	48.036	38.429	0.003	0.002	0	32.574	3774.05	26.06	3800.831	22.09	3634.04
5	9.859	9.894	9.677	0.564	0.566	0.554	49.093	39.274	0.003	0.002	0	30.209	3423.089	24.168	3447.667	18.59	3297.677
6	9.861	9.896	9.677	0.564	0.566	0.554	50.173	40.138	0.003	0.002	0	28.016	3107.705	22.413	3129.775	15.645	2992.447
7	9.86	9.896	9.677	0.564	0.566	0.554	51.276	41.021	0.003	0.002	0	25.982	2819.62	20.786	2839.784	13.166	2715.469
8	9.861	9.896	9.677	0.564	0.566	0.554	52.405	41.924	0.003	0.002	0	24.096	2558.902	19.277	2577.135	11.08	2464.128
9	9.861	9.896	9.677	0.564	0.566	0.554	53.557	42.846	0.003	0.002	0	22.347	2321.868	17.878	2338.497	9.325	2236.051
10	9.862	9.896	9.677	0.564	0.566	0.554	54.736	43.789	0.003	0.002	0	20.724	2107.385	16.58	2122.342	7.848	2029.085
11	9.861	9.896	9.677	0.564	0.566	0.554	55.94	44.752	0.003	0.002	0	19.22	1912.037	15.376	1925.618	6.604	1841.275
12	9.857	9.892	9.677	0.564	0.566	0.554	57.171	45.736	0.003	0.002	0	17.825	1733.593	14.26	1745.976	5.558	1670.848
13	9.851	9.885	9.677	0.564	0.566	0.554	58.428	46.743	0.003	0.002	0	16.531	1571.22	13.225	1582.329	4.678	1516.196
14	9.858	9.893	9.677	0.564	0.566	0.554	59.714	47.771	0.003	0.002	0	15.331	1427.846	12.265	1437.965	3.937	1375.858
15	9.854	9.889	9.677	0.564	0.566	0.554	61.027	48.822	0.003	0.002	0	14.218	1294.58	11.374	1303.799	3.313	1248.51
16	9.857	9.892	9.677	0.564	0.566	0.554	62.37	49.896	0.003	0.002	0	13.185	1175.685	10.548	1184.016	2.788	1132.95
17	9.854	9.889	9.677	0.564	0.566	0.554	63.742	50.994	0.003	0.002	0	12.228	1066.119	9.783	1073.693	2.346	1028.085
18	9.851	9.885	9.677	0.564	0.566	0.554	65.144	52.116	0.003	0.002	0	11.34	966.791	9.073	973.616	1.975	932.926
19	9.847	9.883	9.677	0.564	0.566	0.554	66.578	53.262	0.003	0.002	0	10.517	876.713	8.414	883.028	1.662	846.576
20	9.855	9.89	9.677	0.564	0.566	0.554	68.042	54.434	0.003	0.002	0	9.754	796.757	7.803	802.483	1.399	768.217
21	9.846	9.881	9.677	0.564	0.566	0.554	69.539	55.631	0.003	0.002	0	9.046	721.697	7.237	726.848	1.177	697.112
22	9.86	9.895	9.677	0.564	0.566	0.554	71.069	56.855	0.003	0.002	0	8.389	656.814	6.711	661.485	0.991	632.588
23	9.854	9.889	9.677	0.564	0.566	0.554	72.633	58.106	0.003	0.002	0	7.78	595.306	6.224	599.516	0.834	574.036
24	9.865	9.9	9.677	0.565	0.567	0.554	74.231	59.384	0.003	0.002	0	7.215	541.351	5.772	545.208	0.702	520.904
25	9.861	9.896	9.677	0.564	0.566	0.554	75.864	60.691	0.003	0.002	0	6.691	490.897	5.353	494.385	0.59	472.69

Tabla A.10: Resultados de la simulación de un año para el caso de estudio 2.

Year	Real	Parcial	Fijo	C _{p,r}	C _{p,p}	C _{p,f}	OPEX Fijo _r	OPEX Fijo _p	OPEX VAR _r	OPEX VAR _p	CAPEX	LCOE _{r,1}	LCOE _{r,2}	LCOE _{p,1}	LCOE _{p,2}	LCOE _{f,1}	LCOE _{f,2}
	GWh	GWh	GWh	%	%	%	USD/kW	USD/kW	\$/kW	\$/kW	eUSD/kW	eUSD/kW	eUSD/kW	eUSD/kW	eUSD/kW	eUSD/kW	e USD/kW
1	4.278	3.840	2.136	0.245	0.220	0.122	45.000	36.000	0.003	2229.000	2067.276	952.302	2059.094	426.761	3905.708	237.329	4863.338
2	4.283	3.845	2.136	0.245	0.220	0.122	45.990	36.792	0.003	0.000	38.011	867.891	30.409	388.470	31.417	215.754	4413.193
3	4.281	3.842	2.136	0.245	0.220	0.122	46.000	36.800	0.003	0.000	34.563	787.920	27.651	352.878	25.970	196.140	4004.712
4	4.282	3.843	2.136	0.245	0.220	0.122	47.012	37.610	0.003	0.000	32.112	716.670	25.690	320.890	21.935	178.309	3634.040
5	4.277	3.838	2.136	0.245	0.220	0.122	47.000	37.600	0.003	0.000	29.185	650.024	23.349	291.329	18.124	162.099	3297.677
6	4.273	3.835	2.136	0.245	0.220	0.122	48.034	38.427	0.003	0.000	27.116	590.029	21.693	264.622	15.308	147.363	2992.447
7	4.284	3.846	2.136	0.245	0.220	0.122	48.000	38.400	0.003	0.000	24.633	539.022	19.707	241.222	12.643	133.966	2715.469
8	4.282	3.844	2.136	0.245	0.220	0.122	49.056	39.245	0.003	0.000	22.887	489.487	18.310	219.176	10.678	121.787	2464.128
9	4.281	3.843	2.136	0.245	0.220	0.122	49.000	39.200	0.003	0.000	20.782	444.848	16.626	199.204	8.815	110.716	2236.051
10	4.277	3.839	2.136	0.245	0.220	0.122	50.078	40.062	0.003	0.000	19.309	403.636	15.447	180.902	7.446	100.651	2029.085
11	4.279	3.841	2.136	0.245	0.220	0.122	50.000	40.000	0.003	0.000	17.526	367.252	14.021	164.547	6.144	91.501	1841.275
12	4.280	3.842	2.136	0.245	0.220	0.122	51.100	40.880	0.003	0.000	16.283	334.123	13.027	149.638	5.189	83.182	1670.848
13	4.279	3.841	2.136	0.245	0.220	0.122	51.000	40.800	0.003	0.000	14.774	303.579	11.819	136.000	4.280	75.620	1516.196
14	4.278	3.841	2.136	0.245	0.220	0.122	52.122	41.698	0.003	0.000	13.726	275.866	10.981	123.623	3.615	68.746	1375.858
15	4.276	3.839	2.136	0.245	0.220	0.122	52.000	41.600	0.003	0.000	12.449	250.479	9.960	112.336	2.981	62.496	1248.510
16	4.284	3.846	2.136	0.245	0.220	0.122	53.144	42.515	0.003	0.000	11.566	228.600	9.253	102.306	2.518	56.815	1132.950
17	4.281	3.844	2.136	0.245	0.220	0.122	53.000	42.400	0.003	0.000	10.486	207.560	8.389	92.959	2.075	51.650	1028.085
18	4.279	3.841	2.136	0.245	0.220	0.122	54.166	43.333	0.003	0.000	9.743	188.459	7.794	84.447	1.753	46.954	932.926
19	4.279	3.841	2.136	0.245	0.220	0.122	54.000	43.200	0.003	0.000	8.830	171.383	7.064	76.774	1.444	42.686	846.576
20	4.279	3.841	2.136	0.245	0.220	0.122	55.188	44.150	0.003	0.000	8.204	155.799	6.563	69.794	1.220	38.805	768.217
21	4.281	3.843	2.136	0.245	0.220	0.122	55.000	44.000	0.003	0.000	7.433	141.765	5.946	63.485	1.005	35.277	697.112
22	4.282	3.844	2.136	0.245	0.220	0.122	56.210	44.968	0.003	0.000	6.906	128.933	5.525	57.725	0.849	32.070	632.588
23	4.282	3.844	2.136	0.245	0.220	0.122	56.000	44.800	0.003	0.000	6.254	117.187	5.004	52.474	0.699	29.155	574.036
24	4.284	3.846	2.136	0.245	0.220	0.122	57.232	45.786	0.003	0.000	5.811	106.651	4.649	47.727	0.590	26.504	520.904
25	4.279	3.842	2.136	0.245	0.220	0.122	57.000	45.600	0.003	0.000	5.261	96.709	4.209	43.345	0.486	24.095	472.690

A.10 Artículo publicado

Parte de los resultados de esta tesis se encuentran publicados en la referencia: Sánchez-Pérez, P. A., Robles, M., & Jaramillo, O. A. (2016). Real time Markov chains: Wind states in anemometric data. *Journal of Renewable and Sustainable Energy*, 8(2), 023304. En este trabajo se incluye la versión preimpresa del artículo publicado en www.arxiv.org

Real time Markov chains: Wind states in anemometric data.

P. A. Sánchez-Pérez,^{1, a)} M. Robles,^{2, b)} and O. A. Jaramillo^{2, c)}

¹⁾*Licenciatura en Ingeniería en Energías Renovables, Universidad Nacional Autónoma de México*

²⁾*Instituto de Energías Renovables, Universidad Nacional Autónoma de México Privada Xochicalco SN 62580, Temixco, Mor., México.*

(Dated: 8 March 2016)

The description of wind phenomena is frequently based on data obtained from anemometers, which usually report the wind speed and direction only in a horizontal plane. Such measurements are commonly used either to develop wind generation farms or to forecast weather conditions in a geographical region. Beyond these standard applications, the information contained in the data may be richer than expected and may lead to a better understanding of the wind dynamics in a geographical area. In this work we propose a statistical analysis based on the wind velocity vectors, which we propose may be grouped in wind states associated to binormal distribution functions. We found that the velocity plane defined by the anemometric velocity data may be used as a phase space, where a finite number of states may be found and sorted using standard clustering methods. The main result is a discretization technique useful to model the wind with Markov chains. We applied such ideas in anemometric data for two different sites in Mexico where the wind resource is considered reliable. The approximated Markov chains of both places give a set of values for transition probabilities and residence times that may be regarded as a signature of the dynamics of the site.

PACS numbers: 92.60.Gn, 02.50.Ga

^{a)}pesap@ier.unam.mx

^{b)}mrp@ier.unam.mx

^{c)}ojs@ier.unam.mx

I. INTRODUCTION

It is well known that Markov chains are the simplest mathematical models for random phenomena. However, at the same time they are a class of stochastic process rich enough to be applied at many different problems and considered the theoretical base of many other models¹. A Markov chain can be understood as an stochastic process without memory, therefore the future state of a system only depends on the present situation and not in its history. These models assume that the system under study can only visit a finite number of states, with probabilities that does not depend on time; then, the system is moving between states with a given probability.

Markov chain approaches are frequently present in wind speed forecasting. Modern probabilistic models based on autoregressive (AR) and moving average (MA) models, frequently uses Markov chains in order to estimate some of the multiple parameters needed². Markov chains are the base of forecasting methods³, but they are also useful to generate synthetic wind speed time series needed in simulations approaches⁴⁻⁷.

In order to apply a Markov chain model to climatological data, wind speed have to be converted into discrete states. The most common way to do this is splitting the data associating wind speed values between two limits for a given state^{4,5,7}. Although this separation is practical and give useful results, those states do not represent any physical property of the system beyond a discrete speed. Also it is important to mention that in many studies the analysis focus only in modeling the wind speed, perhaps because is the main quantity needed to forecast the power generation or to estimate energy production. Some other studies introduce changes in wind direction, using Markov chains and defining some goeographical channels, which classify of the angles in discrete values⁸; these channels treated as Markov states allow to calculate probabilities of transition between them.

The wind direction seems to be a very important parameter for being taken in account in the wind assessment when the statistical analysis report that the speed frequencies do not follow a standard Weibull distributions, and then one may suspect the influence of many physical wind modes related with different directions⁹. In such cases forecasting become difficult and more complex methods are needed¹⁰.

This work is addressed to present and discuss a strategy to analyze multi-modal wind data and generate physical states that can be used in Markov chain models of wind. We

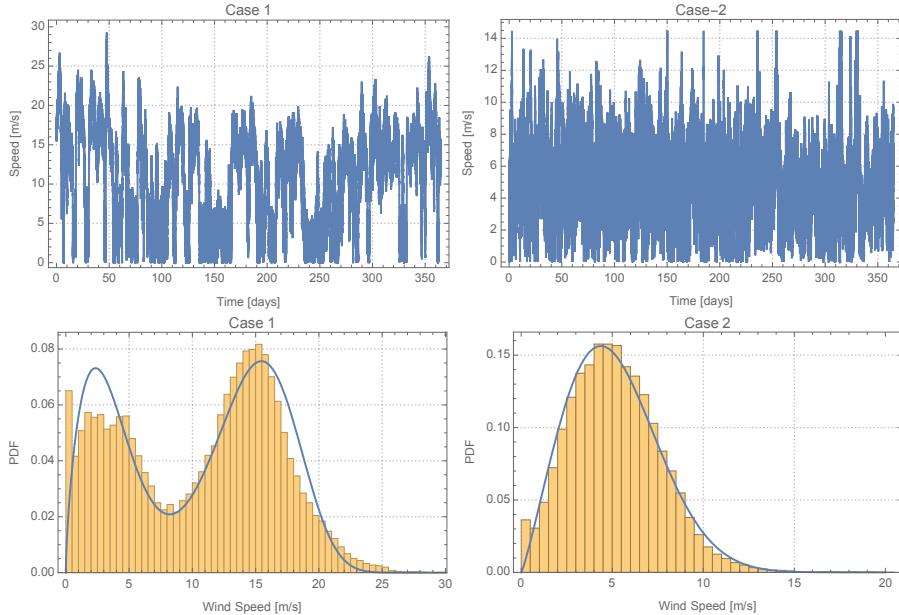


FIG. 1. Probability distribution functions for the two time series chosen as case of study. The histograms were computed directly from the data set and the continuous lines are the best fit functions taken obtained in^{9,11}.

will base the discussion on two data sets containing one year of anemometric data at two different locations in Mexico. The structure of this work is as follows: first we will present the study cases to introduce the need to consider the velocity vector time series and its statistical properties. This will lead us to establish the definition of “wind states” and to propose the use of clustering methods to classify them. Then we will show the results of building Markov Chains for the data and illustrate the procedure to get the simplest stochastic model. A further discussion and conclusions will be addressed at the end of the work.

II. DEFINITION FOR “WIND STATES”.

We choose two study cases. In the first one we use wind data analyzed by Jaramillo and Borja in⁹. These data correspond to one year of measurement, with 10 min averaged measure at the wind farms located in la Ventosa, Oaxaca at 32 m above the ground level. Case 2 consists in one year of measurement with the same time resolution in a wind station at Baja California, Mexico¹¹ at 10 m above the ground level. These two cases were chosen because of their different behavior on their statistical properties. Case 1 was reported to be-

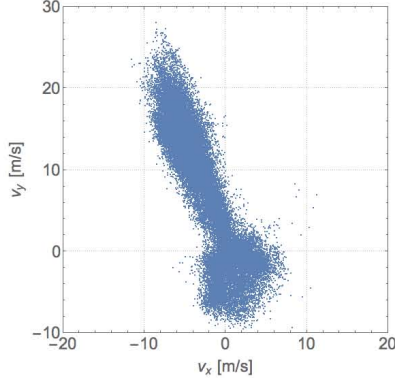


FIG. 2. Scatter plot for the velocity vectors for one year wind speed data on case 1.

have as a multi-modal function in which wind speed distribution can be approximated as a superposition of two Weibull distributions. This behavior is associated to the changes on wind direction. In contrast, case 2 is a mono-modal and wind speed probability distribution function (PDF), fits with a single Weibull distribution. The PDFs of both data sets are shown in Figure 1. For case 1 fit parameters for a bi-Weibull distribution can be found in reference⁹ and for case 2 in reference¹¹.

Let's start with case 1 and instead of using a wind rose, we will analyze the velocity components with a scatter plot of velocity vectors. Using the wind speed and direction, usually given in weather data, one can construct an average vectorial velocity for each measurement, by simply doing $\vec{V}_i = v_i(\cos \theta_i, \sin \theta_i)$, where v_i and θ_i are the speed and direction data in the time step i . This is a new collection from two time series, one per each velocity and his associated direction. Now, if we plot all points in the plane v_x - v_y instead of visualizing the time series separately, we would obtain the plot shown in Figure 2.

It is important to point out that figure 2 resembles a phase space in the statistical mechanics sense. The wind velocity has vast regions that can not be visited ever, and the points accumulate around well defined regions that may be regarded as "states". If we recall that the angle from the data set is measured from the north direction, the first direct observation of this plot is that a preferred direction exist. In other words, there is a preferred "state" in the NNW direction and this state is where the higher speeds were found. The same result was found previously in⁹.

To drive the analysis deeper we can plot a three dimensional histogram with his smooth density shown in Figure 3(a) and (b) respectively. Figure 3(a) counts the number of velocity

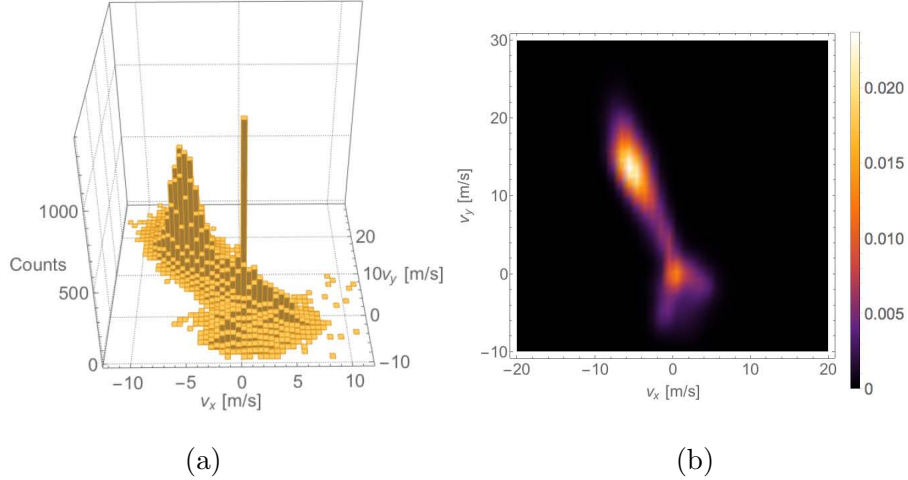


FIG. 3. Three dimensional frequency histogram (a) and smooth density plot (b), for the velocity vectors in the phase plane, for the study case 1.

vectors that lie in a specific squared bin, Figure 3(b) is a colored numerical interpolation for the PDF. If the two dimensional velocity space were an “ensemble” of all possible wind “micro-states”, we could conclude that for long time evolution, there are stable “macro-states” represented by regions where the system is more probably to be found. In this case, it is possible to say that, there are two main states, the first located around the origin, which means the absence of wind or where the system is in repose. The second state appears in the third quadrant (the NNW direction) and the wind speeds are comprehended in the interval from 2 to 25 m s^{-1} always in the same direction. These two main states could not be the only ones, there could be more “states” with lower probabilities that may be also relevant for specific purposes. In order to look after these states, we propose a numerical strategy that we will discuss in section III.

The same analysis can be applied to case 2 and it gives the results in Figure 4 (a), (b) and (c). First, clearly there are more than one macro-states compatible with the figures. It is possible to tell at a glance that at least four relevant cumulus of points can be located. As in the first case, the repose state is also present and probably would be in any case. Obviously, the worst situation for power generation is when the repose state is visited more than any other state. On the opposite side, the best situation for power generation could be close to case 1, because one dominant state exists with good speeds in a narrow direction. Therefore, case 2 is in the middle, is a place with three main states that can generate some wind power,

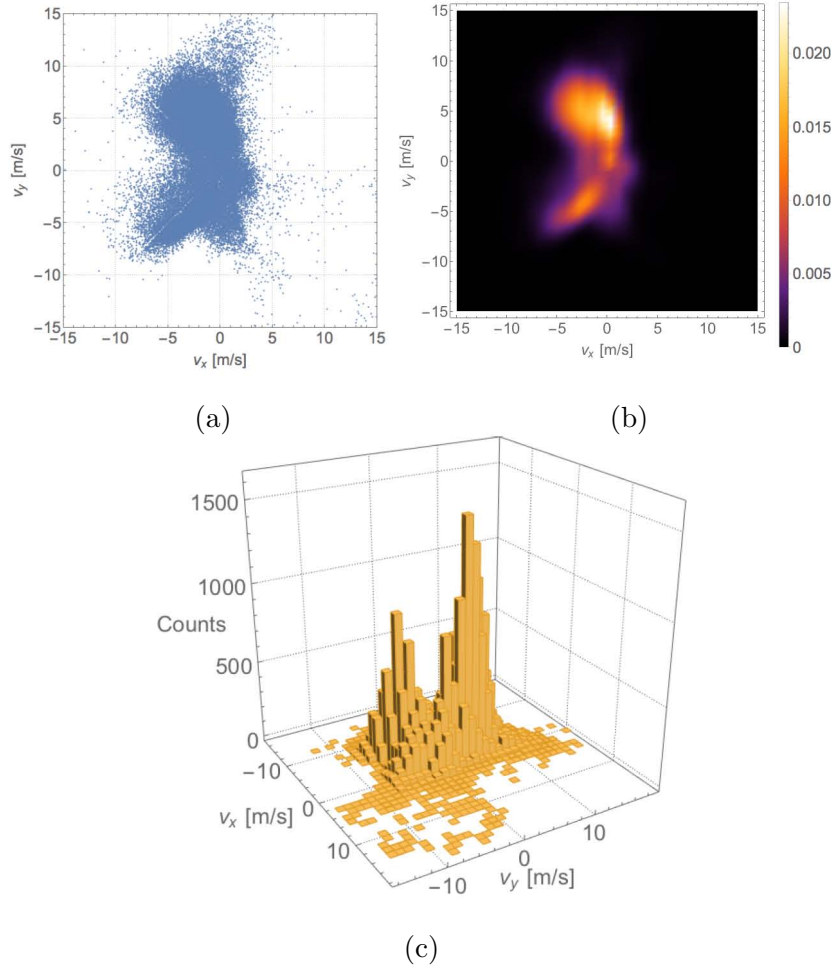


FIG. 4. Scatter plot (a), smooth densiyt plot (b) and the three dimensional histogram (c) for the velocity vectors in phase plane, for the study case 2.

the speed on each is moderated and is located in three different orientations; it is possible to say that an specific wind generator could be designed to take advantage of this situation.

Before to conclude this section, based on the last qualitative analysis, we will introduce a more precisely definition of what we understand as a wind state. It should be referred to a region in the v_x - v_y plane where the wind repeatedly arrives. Physically they are not the same as a thermodynamic equilibrium state or an out of equilibrium steady states. They have an internal dynamics that changes all the time, but they are a property of the long time series and are determined by the geographical and climatic conditions in the site. A wind state is a region in the velocity phase space that contain the accessible wind velocities having a common probability distribution function that characterize them as a cluster. The PDFs for

all the states are like a landscape where the velocity vectors move. The transition dynamics between different wind states is highly stochastic and probably depends on the seasonal cycles or many other climate changes. The physics behind is certainly very complex and therefore beyond the scope of this work. Nevertheless, it is possible to quantify the previous analysis and try to establish a link with stochastic models. The first step is try classify the wind data in the states without know much about the nature of them. To do so, in the next section we use a clustering method, which have become and option to classify time series data efficiently¹²⁻¹⁴ .

III. CLUSTERING WIND STATES FROM WIND SPEED DATA

The only information to establish the landscape of the wind velocity phase space are the meteorological data of the site. In previous section the qualitative analysis let us know that the velocity points tend to get together in cumulus. Therefore a direct way to try to sort them in the different wind states is applying a clustering algorithm. Those methods are usual in different computational science problems including, data mining, machine learning, pattern recognition or bio-informatics.

Currently, there are several effort to apply clustering methods to time series and wind data such the works reported in the literature¹²⁻¹⁵, one of the most used is the k -means clustering algorithm¹⁶. The method is used to partition a set of n -points in k -clusters, classifying with a criterion of the nearest mean. Although it is a difficult problem, modern heuristic algorithms can find rapidly an optimized solution. To apply the k -means method, is necessary propose the optimum number of clusters present in the set (k), since this is a whole different problem in computational science, we decided to use as a first approximation the higher count of peaks in the histograms obtained in previous section. It worth to mention that the idea of applying k -means algorithm is not new, can be found previously for example in¹⁷, to reduce the number of speed states in the time series. For case 1, the smooth density histogram of the Figure 3(b) indicates the presence of two main states. This will help us to choose the number of cluster for k -means algorithm. Applying using $k = 2$ one may easily find the initial partition shown in Figure 5(a). To discover the internal structure close to the repose state, a second application of k -means using $k = 4$, but only for this partition, we obtain Figure 5(b). It is important to notice that k -means clustering do not give an

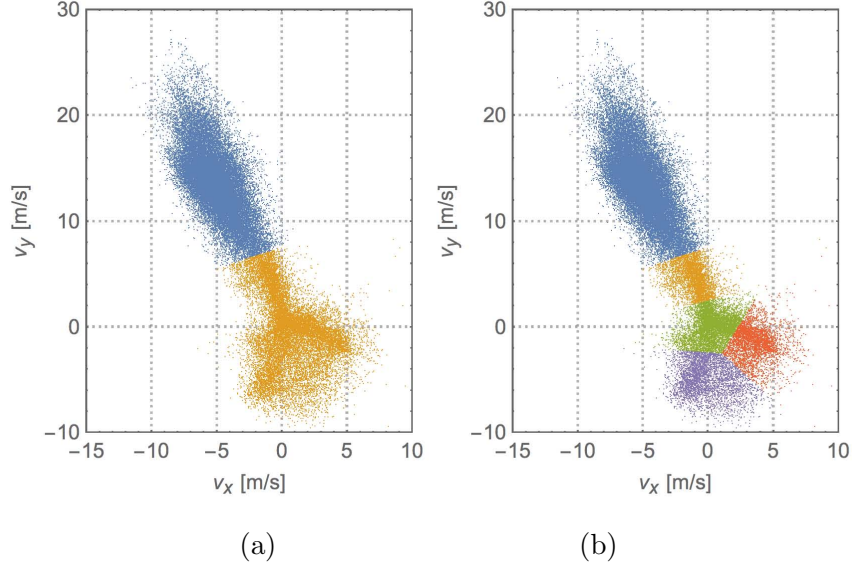


FIG. 5. Two steps k -means clustering applied to case 1 data. In the first step (a) the algorithm was applied to the whole data requesting two partitions. For the second step (b) the algorithm was applied only to the data in the lower partition using $k = 4$.

unique solution, so we need to be careful with the partition. If initially we request $k = 5$ to the entire time series we will obtain a different configuration of clustering. Depending on the centroid initialization clustering will vary because the algorithm can converge to a local minimum that may not be representative. The last partition from this procedure reveals a structure closer to the expected. However, this is a limitation to the method because it is needed to see internal structure of the time series before clustering. A similar procedure was applied for case 2. A graphical resume is depicted in Figure 6. A final partition having five states seems to be the compatible with the explored histograms.

It is important to note that each partition obtained from the three dimensional histogram can be associated with one wind state. Therefore the clustering procedure illustrated may be assumed as a classification method. Once we classified the states we can build a continuous Markov Chain using the time series of the wind states as discussed in the next section. To do this, one more assumption has to be done: the points in a given state distribute with Gaussian bivariate function. Also, it is possible to find the local PDF that can be assumed as a Gaussian bivariate function by simplicity in order to correlate the wind speed and direction.

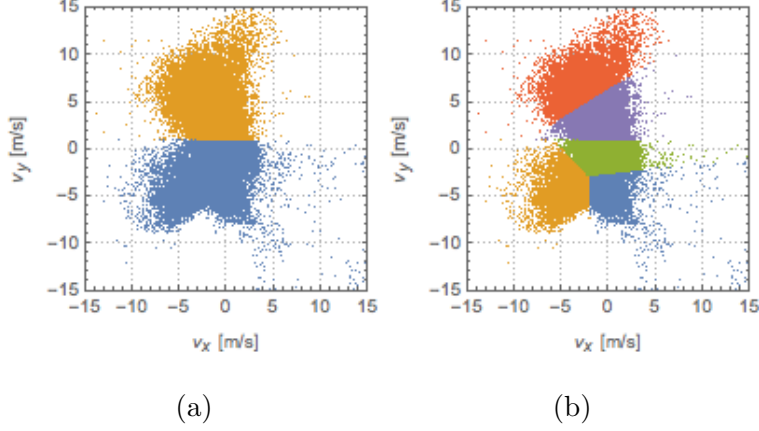


FIG. 6. Two steps k -means clustering applied to case 2 data. In the first step (a) the algorithm was applied to the whole data requesting two partitions. For the second step (b) the algorithm was applied again to the two new partitions requesting $k = 3$ and $k = 2$ respectively.

A Gaussian multivariate function for a vectorial space \vec{r} , with dimension d , is defined as:

$$f(\vec{r}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{r} - \vec{\mu})^T \Sigma^{-1} (\vec{r} - \vec{\mu})\right), \quad (1)$$

where $\vec{\mu}$ is the centroid of the distribution and Σ the covariance matrix, whose components follow the expected value:

$$\Sigma_{i,j} = E[(r_i - \mu_i)(r_j - \mu_j)]. \quad (2)$$

The super index T indicates the transpose of the vector and Σ^{-1} is the inverse of the matrix Σ . For two dimensional case, where $\mu = (\mu_x, \mu_y)$ and the Σ matrix is

$$\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}.$$

Lets apply the following statistical procedure to each partition of the data in case 1:

1. Compute the centroid and covariance matrix.
2. Calculate the Gaussian bi-variate function for each partition.
3. Compose the total distribution function as a superposition of all previous Gaussian, re-normalizing with their respective weights in proportion to the total number of points.

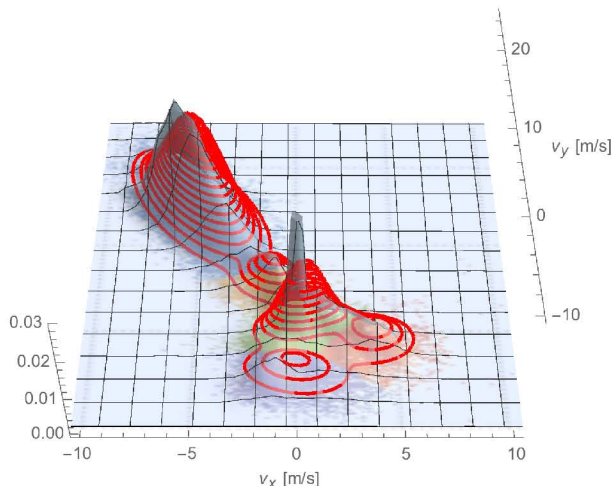


FIG. 7. Probability distribution function obtained from the superposition of Gaussian bi-variate functions (contour lines in red) compared with the numerical distribution obtained directly from real data (light blue surface) in case 1. As a guide the velocity points were included in the background v_x - v_y plane.

The Figure 7 contain the total probability distribution function obtained for case 1 data, compared with a direct numerical computation from the real data. It is important to point out that the Gaussian bi-variate functions can be applied without considering a fitting method. Repeating the same steps on case 2 data it is possible to depict the Figure 8. These data have a more complex structure and do not agree as well compared with the case 1. However, applying this methodology to case 2 gives some information about the landscape; The numerical values of $\vec{\mu}$ and the matrix Σ for both cases can be found in the appendix.

Beyond the precision of the method, there are two concepts that worth to remark. Firstly, the notion of state must be associated with a PDF and this could be described, at least on a first approximation, by considering Gaussian distributions. In some way resembles the state functions in quantum mechanics and statistical mechanics, but yet without a master equation known. The second concept is more computational, the use of clustering methods, adapted to this problem, together with the availability of wide historical data bases, may substitute the lack of dynamical equations in this complex problem, allowing to build up efficient numerical methods reliable for practical applications.

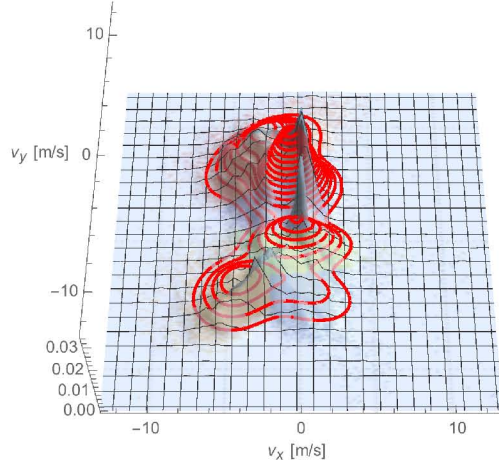


FIG. 8. Probability distribution function obtained from the superposition of Gaussian bi-variate functions (contour lines in red) compared with the numerical distribution obtained directly from real data (light blue surface) of case 2. As a guide the velocity points were included in the background v_x - v_y plane.

To conclude this section we can say that clustering methods may be one key to study the wind states and determine their state functions. The advantage of use them is that can be easily connected with data mining algorithm to apply it systematically to all geographical areas where data is available. In the next section we will introduce the dynamics between states in both study cases when we interpret the dynamics as Markov chains.

IV. MARKOV CHAINS FOR THE WIND STATES.

After the clustering process, the velocity points are sorted in one and only one partition set. In other words, they are associated with one wind state. Each wind state can now be labeled arbitrary, then we decided to use the numbers given in the column “states” in tables II and III, as can be seen in Appendix A.

A real time Markov chain is developed from the original wind data time series. The chain is assembled by considering the association of each vector with its state to which it belongs. The Figure 9 presents the real time Markov chains for both study cases. The first information coming from figures, at glance, is that case 2 is a lot more complex than the case 1 since it looks more unstable. For the case 1 the state 1 is preferred over the rest.

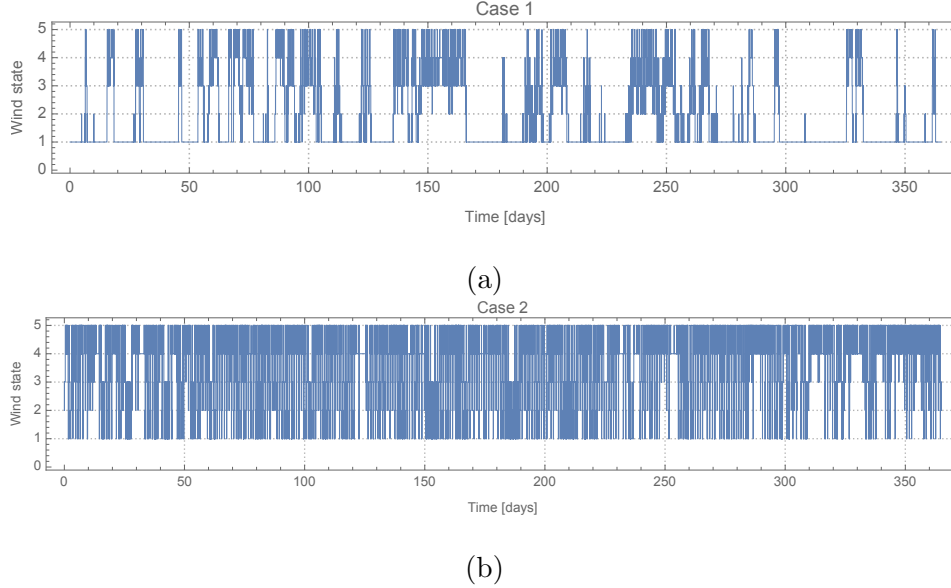


FIG. 9. Time Markov chains associated with both case of study. Case 1(a) and case 2 (b).

To quantify this appreciations we computed the transition probability matrix between the states for both cases, assuming that the sequence as a discrete Markov chain. The results are reported in Tables IV and V. In Table IV it is important to point out the zero entries correspond for transitions between state 1 to 3 and 5, from state 4 to 1 and from 5 to 1. The existence of these unreachable or forbidden transitions must be considered because they can have direct technological applications in the design of wind generators and control algorithms.

To go forth, using the real Markov chains we can measure the total time t_r and the mean life time t_m for the wind in each state. A summary of these values are presented in Table VI, where it is important to highlight the 232.417 days that the wind in case 1 stays in the state 1, together with a mean lifetime close to 0.9 days of the time series. This state is close to ideal to produce wind power, is a state with speeds around 14 m s^{-1} with directions in a narrow sector visited many days in a year and each time is reached it stays in average almost one day. It looks possible to locate a wind turbine without yaw system and therefore does not have a significant reduction on the electric generation instead of a wind turbine with yaw system included. In contrast none of the states, in case 2, stays in a state for more than 2.4 hours, in average although that some of them reach 122 days in total. This imply that a generator should be always tracking the different states, which may imply a higher

maintenance investment, due to the increase of costs in the yaw operation system. Clearly, the information contained in the Markov chain is relevant for technology purposes.

The knowledge of mean life time for each state allows to compute the q-matrix that generates the chains, if were assumed as continuous Markov processes. The off diagonal elements of the q-matrix, represent the time rate at which the systems leave the state i and arrive to state j , q_{ij} . The diagonal elements q_{ii} are the negative average rate at which the system leaves the state i , therefore if the mean life time of states i is t_{mi} then $q_{ii} = -1/t_{mi}^1$. The computation then is simple, for a state i find the transition rate to other state j and then multiply by the rate of leaving state i , $-q_{ii}$. For case 1 and 2 the resulting q-matrices are reported in tables VII and VIII, also in the Appendix A.

For case 1, the state 1, which would be the best candidate for energy production has at most 1.15 transitions per day, the rest do up to 19 transitions. In case 2 is not possible to find a state with less that 10 transitions per day, which can be regarded as a disadvantage.

Beyond the last interpretations, the Q -matrices contain the information needed to create a synthetic Markov chain with close physical properties to the real system. It is a the simplest model and may not produce an accurate forecast, however, it seems a powerful quantity to characterize the states of wind in a geographical situation and could be the base to create machine learning algorithms to automatic search good sites for wind energy production.

The Q -matrices can be represented as graphs. The normal procedure to build them is placing the states in nodes or vertexes linked by arrows that indicate the transitions allowed. With this idea in mind we represented the q-matrices obtained in a special form. First we placed the vertexes of the graph, representing the states, right in the plane of velocities, just in the center of the bi-normal distribution which represent the associated state function. The links between states are drawn as gray arrows using a scale associated with the value of the q_{ij} entry in the q-matrix; black color was set to the maximum of all the entries, excluding the diagonals, and the minimum to a gray value that yet allow to distinguish their presence in the graph. The results are figures 10(a) and (b). These graphs are a visual description of the mechanics of wind in the analyzed sites.

For the case 1, the graph 10(a) immediately address the issue on whether is really useful a tracking method for an optimal generator. Graphically we realize that state 1 is the best in terms of wind speed, and the changes in direction when a transition takes place from it, are

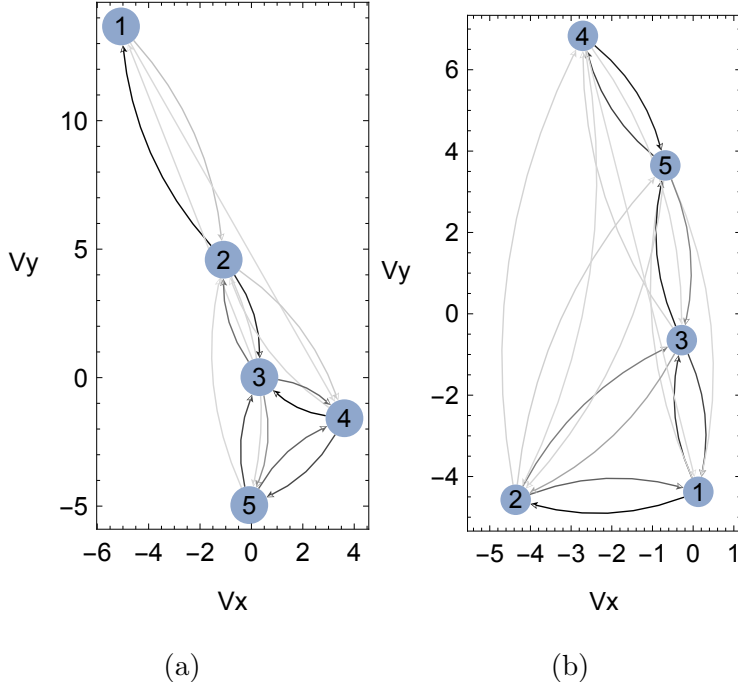


FIG. 10. Graphs associated with the Q -matrices for case 1(a) and case 2 (b).

always to states with lower velocity. Moreover, the fluxes indicate that there is a pressure to return either to state 1 or 2. In contrast case 2 (figure 10(b)) is more stochastic. There are at least three states with good speed potential (states 1, 4 and 5) but, all of them point to different directions and have important fluxes to the other states. This interpretation may change the evaluation of the wind resource done by traditional methods.

To finish this section we must remark the fact that, although the wind states seems to exists independently of the time, the time series for wind velocities always show seasonal effects, for that reason Markov chains are not the best representation. However, the transition and q -matrices are a numerical properties that may represent a signature of the mechanics of wind that may be useful beyond the physical understanding.

V. DISCUSSION

At this point we reach the question on how this results may contribute either to the understanding of wind phenomena or to improve traditional wind assessment. Clearly, the definition of wind state given here is only based on the assumption that wind velocities are described by two dimensional Gaussian modes that may be obtained from the cluster

partition. It is important to point out that the physical causes for each state are indeed a complex problem beyond the scope of this work. However, it is possible to propose how to characterize physically the states and try to find out how useful are to estimate the power generation. If the state i has a Gaussian component $f_i(v_x, v_y)$ as given in equation (1), the wind power per unit area for the state may be calculated by:

$$\frac{P_i}{A} = \frac{1}{2}\rho \int_0^\infty \int_0^\infty |\vec{v}|^3 f_i(v_x, v_y) dv_x dv_y, \quad (3)$$

where $|\vec{v}| = (v_x^2 + v_y^2)^{1/2}$ and ρ is the air density that in average may be considered as 1.21 kg/m^3 , at sea level.

If m is the number of wind states the total power of the site is simply $P = \sum_{i=1}^m P_i$ and the relative power of a state may be adopted as $p_i^* = P_i/P$. This relative power give us the information about how the state may weight in the total energy production and then, it can be used as a figure of merit.

Another important effect is the angular dispersion of wind states. An ideal state is such that concentrate the total power in a very narrow angular sector. We can define the “aperture” (α_p) of a state as the angle subtended by the ellipse formed with the isoline of the i -esime Gaussian component that covers the 90% of data in the state, and represent the angular region where the wind blows with more probability, as is shown in figure 11. Note that this new definition may enhance the information given by the standard methods, such wind rose graphs, where the frequency of wind directions are grouped in discrete circular sectors, with the same angle, and colored with the wind speed. This statement may be observed in Figure 12, where the ellipses used for the compute α_p are presented, for both study cases, colored with the relative power. These graphs show clearly the regions where wind produces most of the energy and therefore, together with the dynamics shown in Figure 10, could assist, for example in the optimization of tracking methods, defining the regions where generator could not produce more energy than the consumed by yaw system, and then could be completely avoided, balancing the tracking algorithms to maximize production.

The values of α_p obtained for both study cases are in Table I, with the relative power p_i^* and the direction θ of center of Gaussian component. All the angles are measured in degrees and θ values are relative to north direction (v_y). We can remark some observations: the numbers immediately say that state 1 on the study case 1 is the best of all, because it maximizes the relative power and minimizes the aperture. That was already known, but

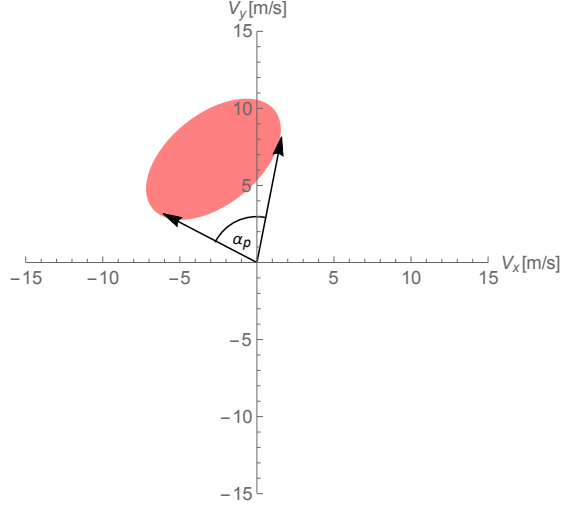


FIG. 11. Graphical construction of α_p in the state 4 for the study case 2.

now this conclusion may be achieved through an automatic quantitative procedure easy to implement. Note that, since on both cases the state 3 defines regions that contain the origin, the aperture must be fixed on 360° , *i.e.* the maximum possible value. In contrast the p_3^* is very low, therefore this state maximizes aperture and minimizes power, and it is the worst situation for energy generation. That was also known, but again the procedure

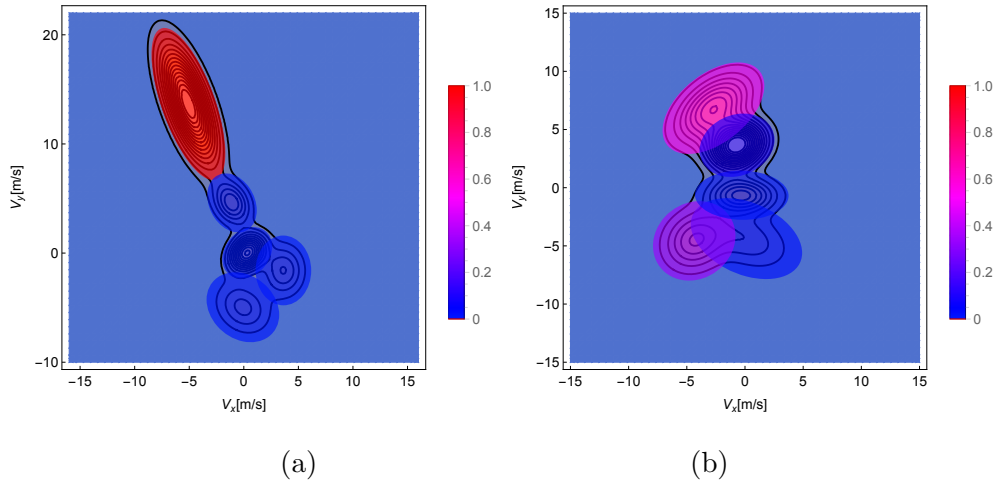


FIG. 12. Elliptical regions where Gaussian components of each state covers 90% of the data associated with it for the study case 1 (a) and 2 (b). Ellipses were colored using relative power p_i^* with the color code shown in legends.

	Case 1			Case 2		
State	p_i^* [%]	α_p [°]	θ [°]	p_i^* [%]	α_p [°]	θ [°]
1	98.68	21.95	20.37	1.92	119.90	181.70
2	0.43	58.88	13.22	34.75	64.81	136.30
3	0.03	360.00	274.24	0.85	360.00	156.70
4	0.29	93.25	246.30	54.16	73.60	21.64
5	0.57	82.74	0.97	8.33	103.10	10.63

TABLE I: Relative power, aperture and average direction for all the states in study cases 1 and 2.

may detect such states. Summarizing, the Gaussian components assumed as state functions may allow to define quantitative numerical strategies to analyze the energy production in a specific location.

VI. CONCLUSIONS

We have presented a numerical analysis of wind velocity vectors obtained from anemometric time series. The analysis, based on the k -means clustering method was applied to two study cases. The data, accumulated for one year, sort in clusters well delimited in the 2D velocity space, and may be associated with bivariate Gaussian distributions. These relations allow us to introduce the concept of “wind state”. The geographical and meteorological characteristics of the site should determine the formation of those states in a complex manner, however, their characteristics may be approximated through the computational analysis shown. It is reasonable to think that the bigger the wind data base is in the site, more detailed states can be found, therefore the wind states should be time independent in analogy with thermodynamic states, hence the method may be applied the same for different time-stepping discretization instead of 10 min averaged measurements. The local dynamics may be described in a rough manner by the transition probabilities between them. The instate dynamics is subject of further investigation. It seems that the wind states introduce a “natural” discretization of the velocity that may convert the time series in discrete Chains, that in first approximation can be assumed as Markovian process. With this assumption the transition matrices and Q-matrices can be obtained for any set

of data, regardless the time resolution or how long the history of data is, and represent in a simple way a first approximation to their stochastic dynamic behavior. In certain way these matrices are a footprint of the dynamics in a given site, and it is possible to be unique, like a signature. We are convinced that the method illustrated here is an alternative analysis to the traditional evaluation methods, and provides new information about the wind. Therefore, it can be useful to understand the properties of wind, to improve the wind resource assessment and drive new control methods in order to enhance energy production. Finally, it is worth to pointing out that physical properties added to the wind states, such relative power and aperture angle, allow to compare different wind regions, and to choose the most convenience for energy production.

ACKNOWLEDGMENT

We would like to acknowledge Dr. Rafael Campos and Dr. Mishael Sánchez-Pérez for the useful discussions and suggestions. Also we acknowledge Dr. Maximiliano Valdez González for the computational assistance.

Appendix A: Tables.

In this appendix we include the tables with the numerical calculation associated with the normal bivariate state functions, transition matrix, characteristic times and q matrices for the two study cases presented in the text. All the values were computed from the time series, sorting the points in the state through the k-means algorithm described.

1. Normal bivariate distribution parameters for each state.

State	$\mu_x[m/s]$	$\mu_y[m/s]$	$\Sigma_{xx}[(m/s)^2]$	$\Sigma_{xy}[(m/s)^2]$	$\Sigma_{yx}[(m/s)^2]$	$\Sigma_{yy}[(m/s)^2]$
1	-5.0746	13.6675	2.4915	-3.3270	-3.3270	10.5284
2	-1.0804	4.5985	1.1161	-0.3334	-0.3334	1.5873
3	0.3163	0.0234	1.1334	0.2858	0.2858	1.2095
4	3.6258	-1.5895	1.4108	-0.0093	-0.0093	2.2184
5	-0.0843	-4.9619	2.4047	-0.5016	-0.5016	2.2301

TABLE II: Gaussian parameters for wind states for the study case 1.

State	$\mu_x[m/s]$	$\mu_y[m/s]$	$\Sigma_{xx}[(m/s)^2]$	$\Sigma_{xy}[(m/s)^2]$	$\Sigma_{yx}[(m/s)^2]$	$\Sigma_{yy}[(m/s)^2]$
1	0.1269	-4.3762	4.9710	-1.2755	-1.2755	2.6258
2	-4.3630	-4.5714	2.7266	0.3326	0.3326	2.5311
3	-0.2794	-0.6473	3.5139	-0.0438	-0.0438	0.9712
4	-2.7106	6.8333	4.1543	1.8232	1.8232	3.3550
5	-0.6847	3.6476	2.2533	0.3284	0.3284	1.6114

TABLE III: Gaussian parameters for wind states for the study case 2.

2. Transition matrix between states.

State	1	2	3	4	5
1	0.9921	0.0079	0.0000	0.0001	0.0000
2	0.0680	0.8665	0.0590	0.0049	0.0015
3	0.0001	0.0341	0.9056	0.0376	0.0226
4	0.0000	0.0010	0.0701	0.8815	0.0473
5	0.0000	0.0002	0.0485	0.0386	0.9127

TABLE IV: Transition matrix for the real time Markov chain obtained from the case 1 data.

State	1	2	3	4	5
1	0.8600	0.0794	0.0600	0.0000	0.0006
2	0.0446	0.9158	0.0338	0.0021	0.0037
3	0.0549	0.0191	0.8548	0.0021	0.0690
4	0.0001	0.0008	0.0004	0.9287	0.0701
5	0.0002	0.0016	0.0306	0.0569	0.9107

TABLE V: Transition matrix for the real time Markov chain obtained from the case 2 data.

3. Characteristic times.

States	Case 1		Case 2	
	t_r (days)	t_m (days)	t_r (days)	t_m (days)
1	232.417	0.868973	37.3542	0.0496071
2	26.9444	0.0520163	50.7361	0.0825203
3	50.9583	0.0735329	53.3403	0.0477641
4	26.5486	0.0586062	100.861	0.0973563
5	28.0694	0.0795168	122.701	0.0777575

TABLE VI: Total residence time t_r and mean life time t_m from the Markov chains for the two study cases.

4. Q-Matrices.

State	1	2	3	4	5
1	-1.1508	1.1421	0.0000	0.0087	0.0000
2	9.7979	-19.2247	8.4990	0.7052	0.2227
3	0.0196	4.9060	-13.5993	5.4162	3.2576
4	0.0000	0.1507	10.0947	-17.0630	6.8177
5	0.0000	0.0356	6.9827	5.5576	-12.5760

TABLE VII: Q-matrix for the real time Markov chain obtained from the case 1 data.

State	1	2	3	4	5
1	-20.1584	11.4311	8.6470	0.0000	0.0803
2	6.4236	-12.1182	4.8670	0.2956	0.5320
3	7.9238	2.7602	-20.9362	0.3004	9.9518
4	0.0099	0.1091	0.0595	-10.2716	10.0931
5	0.0326	0.2363	4.4009	8.1906	-12.8605

TABLE VIII: Q-matrix for the real time Markov chain obtained from the case 2 data.

REFERENCES

- ¹J. R. Norris, *Markov Chains* (Cambridge University Press, 1997).
- ²J. Jung and R. P. Broadwater, “Current status and future advances for wind speed and power forecasting,” *Renewable and Sustainable Energy Reviews* **31**, 762–777 (2014).
- ³G. D’Amico, F. Petroni, and F. Prattico, “Wind speed and energy forecasting at different time scales: A nonparametric approach,” *Physica A: Statistical Mechanics and its Applications* **406**, 59–66 (2014).
- ⁴A. Shamshad, M. Bawadi, W. Wanhussin, T. Majid, and S. Sanusi, “First and second order Markov chain models for synthetic generation of wind speed time series,” *Energy* **30**, 693–708 (2005).
- ⁵A. D. Sahin and Z. Sen, “First-order Markov chain approach to wind speed modelling,” *Journal of Wind Engineering and Industrial Aerodynamics* **89**, 263–269 (2001).
- ⁶H. Aksoy, Z. Fuat Toprak, A. Aytek, and N. Erdem Ünal, “Stochastic generation of hourly mean wind speed data,” *Renewable Energy* **29**, 2111–2131 (2004).
- ⁷G. D’Amico, F. Petroni, and F. Prattico, “Wind speed modeled as an indexed semi-Markov process,” *Environmetrics*, n/a–n/a (2013).
- ⁸F. Castino, R. Festa, and C. Ratto, “Stochastic modelling of wind velocities time series,” *Journal of Wind Engineering and industrial aerodynamics* **74**, 141–151 (1998).
- ⁹O. A. Jaramillo and M. Borja, “Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case,” *Renewable Energy* **29**, 1613–1630 (2004).
- ¹⁰E. Erdem and J. Shi, “Arma based approaches for forecasting the tuple of wind speed and direction,” *Applied Energy* **88**, 1405–1414 (2011).

- ¹¹O. A. Jaramillo, R. Saldaña, and U. Miranda, “Wind power potential of Baja California Sur, México,” *Renewable Energy* **29**, 2087–2100 (2004).
- ¹²S. Rani and G. Sikka, “Recent techniques of clustering of time series data: A survey,” *Int. J. Comput. Appl* **52**, 1–9 (2012).
- ¹³A. Clifton and J. K. Lundquist, “Data Clustering Reveals Climate Impacts on Local Wind Phenomena,” *Journal of Applied Meteorology and Climatology* **51**, 1547–1557 (2012).
- ¹⁴U. Lund, “Cluster analysis for directional data,” *Communications in Statistics - Simulation and Computation* **28**, 1001–1009 (1999), <http://dx.doi.org/10.1080/03610919908813589>.
- ¹⁵A. D. Piazza, M. Carmela, D. Piazza, M. Ieee, A. Ragusa, M. Ieee, G. Vitale, and M. Ieee, “Statistical Processing of Wind Speed Data for Energy Forecast and Planning,” , 4–9 (2010).
- ¹⁶J. MacQueen *et al.*, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1 (Oakland, CA, USA., 1967) pp. 281–297.
- ¹⁷A. P. Leite, C. L. T. Borges, and D. M. Falcão, “Probabilistic Wind Farms Generation Model for Reliability Studies Applied to Brazilian Sites,” *IEE Transactions on Power Systems* **21**, 1493–1501 (2006).