

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO - CHILE



**PARTICLE SWARM OPTIMIZATION PARA EL AJUSTE DE
MODELOS PROBABILÍSTICOS A DATOS DEL VIENTO EN
VALPARAÍSO**

ALONSO JAVIER SANDOVAL ACEVEDO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

PROFESOR GUÍA : MARÍA CRISTINA RIFF.
PROFESOR CORREFERENTE : X. X.

JULIO 2016

Índice general

| | |
|--|-----------|
| 1. Introducción | 1 |
| 1.1. Identificación del problema | 1 |
| 1.2. Objetivos | 3 |
| 1.2.1. Objetivos específicos | 3 |
| 2. Estado del arte | 4 |
| 2.1. Particle Swarm Optimization | 4 |
| 2.2. Velocidad del viento | 6 |
| 2.2.1. Distribución de Weibull | 6 |
| 2.2.2. Métodos numéricos | 7 |
| 2.2.3. Particle Swarm Optimization | 9 |
| 2.3. Dirección del viento | 11 |
| 3. Desarrollo de la solución | 14 |
| 3.1. Velocidad del viento | 14 |
| 3.1.1. Modelo Matemático | 14 |
| 3.1.2. Representación | 15 |
| 3.1.3. Descripción del algoritmo | 16 |
| 3.1.4. Experimentos | 16 |
| 3.2. Dirección del viento | 18 |
| 3.2.1. Modelo Matemático | 18 |
| 3.2.2. Representación | 21 |
| 3.2.3. Descripción del algoritmo | 21 |
| 3.2.4. Experimentos | 23 |
| 4. Análisis y conclusiones | 25 |
| 4.1. Resultados y conclusiones | 25 |
| 4.1.1. Análisis de los resultados - velocidad del viento | 25 |
| 4.1.2. Análisis de los resultados - dirección del viento | 30 |
| 4.1.3. Conclusiones | 31 |
| Bibliography | 35 |

Índice de figuras

| | |
|--|----|
| 2.1. Movimiento de una partícula | 5 |
| 2.2. Función de distribución de probabilidad de Weibull | 7 |
| 2.3. Distribución de Weibull con histograma - Maracanaú | 10 |
| 2.4. Modelo de ajuste FVMV para suroeste y noreste en la estación Mersing . | 12 |
| 2.5. Ajuste dirección del viento, aeropuerto St. John | 13 |
| 3.1. Ejemplo colección de datos Enero Valparaíso 2015 | 17 |
| 3.2. Esquema de uso del algoritmo | 24 |
| 4.1. Superficie de datos viento de Valparaíso. Fuente: Elaboración Propia. . . | 26 |
| 4.2. Ajuste con PSO a datos del viento de Valparaíso. Fuente: Elaboración Propia. | 27 |
| 4.3. Ajuste con PSO a datos Valparaíso 2015, 2014 y 2013, baja y buena calidad. Fuente: Elaboración Propia. | 27 |
| 4.4. Ajuste con PSO a registros del viento en Valparaíso (Con todos los datos). Fuente: Elaboración Propia. | 28 |
| 4.5. Ajuste con PSO a datos Valparaíso 2015, por rango de meses. Fuente: Elaboración Propia. | 29 |
| 4.6. Ajuste con PSO a datos (cifras puras) Valparaíso 2015, 2014, 2013 | 29 |
| 4.7. Graficos de ajustes anuales | 31 |
| 4.8. Gráficos de ajuste de MVM por meses. Fuente: Elaboración propia. . . . | 33 |
| 4.9. Gráficos de ajuste de MVM por meses. Fuente: Elaboración propia. . . . | 34 |

Capítulo 1

Introducción

1.1. Identificación del problema

El viento es uno de los fenómenos meteorológicos más comunes de las zonas costeras de Valparaíso. Su presencia incide en distintos aspectos del medio ambiente, como las condiciones del clima, la sensación térmica, algunos desastres naturales, entre otros. Esto atrae el interés de investigadores, lo que ha fomentado a lo largo de la historia el estudio de este fenómeno con el fin de poder controlar las variables que condicionan su comportamiento, de manera de poder predecir los fenómenos subyacentes a este.

Entre los diversos efectos del viento, uno de los más simples y útiles es su capacidad de mover objetos, lo cual ha permitido al hombre aprovechar esta propiedad para convertir la energía cinética del viento en energía eléctrica. Dicha fuente de energía es conocida como energía eólica, ampliamente calificada como renovable y poco contaminante.

Acorde a esto último, en los últimos años, tanto el gobierno de Chile como la ciudadanía, han mostrado un creciente interés en el uso de las fuentes de energías renovables y con poco impacto ambiental, por lo que distintos proyectos en la materia han sido llevados a cabo, desde estudios de factibilidad y recopilación de datos hasta el emplazamiento de las primeras centrales de fuentes de energía limpia. El año 2014, el ministerio de energía publicó un extenso reporte acerca de la situación actual del país en materias de energías renovables, en donde se pueden ver distintas proyecciones y estimaciones de implementación de posibles plantas de generación a lo largo del país. [1]

Cualquiera sea el motivo de estudio de las características del viento (en particular el tema energético), existen dos variables fundamentales ha considerar: la velocidad y la dirección

del viento. Por lo tanto, es necesario en primera instancia, recopilar mediciones del viento de la zona a estudiar. Actualmente estos datos están disponibles y son obtenidos por diversos centros meteorológicos a lo largo del país. Para este proyecto, el servicio meteorológico de la Armada de Chile, que cuenta con el equipo necesario para registrar el comportamiento del viento a lo largo de las distintas épocas del año y en diferentes zonas de Chile, ha dispuesto de los registros del viento en la comuna de Valparaíso de los años 2013, 2014 y 2015.

Con los datos a disposición, es necesario obtener modelos que expliquen la distribución de estos, de manera de poder obtener fácilmente información precisa que permita por ejemplo, evaluar el potencial eléctrico de cierta zona [2].

En la literatura, son ampliamente aceptadas la distribución de *Weibull* para modelar el conjunto datos de velocidad y la distribución de Von Mises para el modelo de la dirección del viento. Ambas distribuciones probabilistas requieren de la determinación de parámetros para que el modelo se ajuste a los datos obtenidos. La elección del método para poder encontrar los parámetros de ajustes definirá la calidad de los modelos, por ello, diversas fórmulas han sido presentadas en la literatura, las cuales tienen mayor o menor precisión dependiendo de las características de los datos, la calidad del método o la estrategia utilizada. Comúnmente, se utilizan métodos numéricos para la determinación de estos parámetros, sin embargo, estudios recientes han abordado nuevas estrategias utilizando métodos heurísticos, en particular la meta-heurística *Particle Swarm Optimization* (PSO), con la cual se ha logrado mejorar la calidad del modelo para la velocidad del viento, como se expone en un caso de estudio en Brasil [3], mientras que para la dirección del viento se ha propuesto una forma más sencilla para poder obtener los parámetros de ajuste a través del uso de PSO. La ventaja del uso de PSO es que es un algoritmo de uso general, de fácil implementación, buen tiempo de ejecución y con resultados cercanos al óptimo a encontrar. Por lo anterior, el trabajo a realizar en este proyecto se basa en validar el uso de PSO para encontrar los parámetros de ajuste de la distribución de *Weibull* con datos de la zona de la comuna de Valparaíso y abordar la estrategia propuesta por Heckenbergerova et al. [4] para el ajuste del modelo de dirección del viento.

1.2. Objetivos

Aplicar métodos actuales de optimización basados en la meta-heurística *Particle Swarm Optimization* para el ajuste de modelos de dirección y velocidad del viento a los datos recopilados del viento en Valparaíso, con el fin de presentar resultados que permitan inferir información precisa acerca de las condiciones de la región para la generación de energía eólica y otras potenciales aplicaciones.

1.2.1. Objetivos específicos

1. Implementar un algoritmo basado en *Particle Optimization Swarm* para **encontrar** los parámetros de un modelo probabilístico que se ajusten a los datos de la velocidad del viento en Valparaíso.
2. Implementar un algoritmo basado en *Particle Optimization Swarm* para **optimizar** los parámetros de un modelo probabilístico que se ajusten a los datos de la dirección del viento en Valparaíso.
3. Evaluar los modelos sobre los datos del viento para validar la propuesta realizada.

Capítulo 2

Estado del arte

2.1. Particle Swarm Optimization

Como se introduce en el artículo de Kaveh [5], el algoritmo *Particle Swarm Optimization* es una meta-heurística inspirada en las observaciones de la naturaleza acerca del comportamiento social de poblaciones de enjambres. Esta abstracción está basada, por ejemplo, en las gaviotas, las cuales suelen moverse en grupos, conocidos como bandadas, cerca del mar en búsqueda de zonas donde hayan alimento (peces). El método simula la conducta de los individuos a través de partículas que se mueven dentro de un espacio (de búsqueda), siendo estas afectadas por factores individuales (conocimiento propio) y colectivos (conocimiento del enjambre), los cuales dirigen el movimiento de estos grupos a ciertas zonas las que son determinadas por una función objetivo (*fitness function*).

Para cada partícula su vector posición \vec{x} representa una solución candidata, la cual varía dentro del espacio de búsqueda a velocidad \vec{v} . Después de varias iteraciones, el enjambre o conjunto de partículas, se irá concentrando en aquellas zonas donde la posición obtenga mejores valores para la función objetivo.

El modelo clásico presentado por Kennedy y Eberhart [6], describe la variación de la velocidad y de la posición de las partículas como se presenta a continuación:

$$v_{i,j}^{k+1} = v_{i,j}^k + c_1 r_1 (x_{best_{i,j}}^k - x_{i,j}^k) + c_2 r_2 (x_{gbest_j}^k - x_{i,j}^k) \quad (2.1)$$

$$x_{i,j}^{k+1} = x_{i,j}^k + v_{i,j}^{k+1} \quad (2.2)$$

Como se explica en Kaveh [5] $x_{i,j}^k$ y $v_{i,j}^k$ son la j -ésima componente de la posición y la velocidad de la partícula i respectivamente en la iteración o tiempo k , r_1 y r_2 son número aleatorios uniformes que varían de 0 a 1, $xbest_i$ y $xgbest$ representan las mejores soluciones alcanzadas por la partícula y por el enjambre respectivamente, c_1 y c_2 son parámetros que representan la confianza en la solución individual de la partícula (parámetro cognitivo) y la incidencia del aspecto colectivo o solución global (parámetro social), respectivamente. Un esquema de la interacción de estos componentes se aprecia en la figura 2.1

El modelo clásico presentado tiene ciertas complicaciones en la forma en que se actualiza

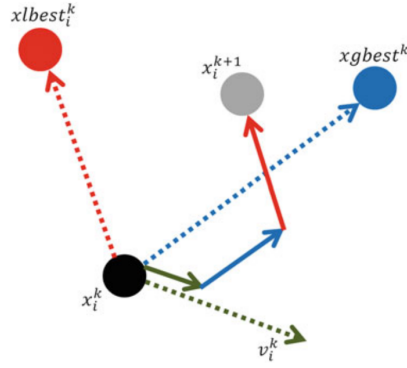


Figura 2.1: Movimiento de una partícula
Creado por Kaveh [5].

la velocidad, una de ellas es la incidencia de la velocidad previa en una partícula, por lo que a modo de balancear esta variable, se añade un factor que escala esta velocidad, dado que como se explica en Kaveh [5], si la velocidad previa se elimina, las partículas quedan atrapadas en una región local, pero si se le da demasiado peso, converge rápidamente a un óptimo conocido. Por esto, la forma del PSO base actual, tiene un parámetro w , que representa la incidencia de la velocidad previa (factor de inercia). Por lo tanto, ahora se tiene que la partícula actualiza su velocidad de la siguiente forma:

$$v_{i,j}^{k+1} = wv_{i,j}^k + c_1r_1(xbest_{i,j}^k - x_{i,j}^k) + c_2r_2(xgbest_j^k - x_{i,j}^k) \quad (2.3)$$

$$x_{i,j}^{k+1} = x_{i,j}^k + v_{i,j}^{k+1} \quad (2.4)$$

Donde w es el factor conocido como “inercia” de la partícula, y regula la incidencia de la velocidad previa en la actual.

Finalmente, en el trabajo inicialmente citado, también se puede ver una revisión completa del estado del arte del método *Particle Swarm Optimization* en términos de diseño, donde se

expone las distintas modificaciones y alternativas propuestas por la literatura que pretenden mejorar aspectos como:

1. Configuración de parámetros (inercia, cognitivo, social, aleatorios).
2. Problemas asociados a la convergencia prematura.
3. Estructura de algoritmo o topologías que modifican la comunicación entre partículas (o la incidencia de las soluciones globales y particulares).
4. Sesgos en la búsqueda por la forma de la región o por la interacción de las partículas (operadores de combinación como el promedio, que tienden a centrar la búsqueda en determinada región).
5. Algoritmos híbridos con PSO.
6. Versión discreta del PSO.

2.2. Velocidad del viento

2.2.1. Distribución de Weibull

Dado un conjunto de datos de velocidad obtenidos de la medición del viento, se puede crear un histograma que represente la frecuencia de estos datos. A partir de esto, es posible generar un modelo probabilístico que explique el comportamiento de las velocidades del viento medido, ajustándose a los datos recolectados. Dicho modelo comúnmente se basa en la distribución de Weibull, la cual es ampliamente aceptada por la comunidad dedicada al estudio meteorológico, tal y como se menciona en el trabajo de Carneiro et al. [3], Kongnam et al. [7] y Dabbaghiyan et al. [2]. En particular, en el trabajo realizado por Carneiro et al. [3], se describe la distribución de Weibull como:

$$f_{weibull}(v) = \frac{k}{c} \cdot \left(\frac{v}{c}\right)^{k-1} \cdot e^{-\left(\frac{v}{c}\right)^k} \quad (2.5)$$

Donde k y c son los parámetros de ajuste que representan la forma y la escala de la distribución respectivamente, y v es el valor de la velocidad del viento a la que el modelo asociará una determinada frecuencia. Un ejemplo de como se transforma esta distribución

se aprecia en la figura 2.2, en donde se ven distintas curvas de Weibull, con diferentes parámetros k , y c constante.

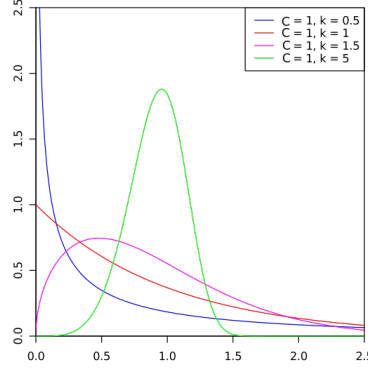


Figura 2.2: Función de distribución de probabilidad de Weibull
Adaptación propia desde [8].

2.2.2. Métodos numéricos

Tradicionalmente, se utilizan métodos numéricos para estimar los parámetros de la distribución de Weibull. En el artículo de Chang [9], se realiza una comparación de seis métodos numéricos comúnmente utilizados para la obtención de k y c . A continuación, se describen brevemente estos métodos:

1. **The Moment:** Se basa en la iteración numérica de las siguientes dos ecuaciones:

$$\bar{v} = c\Gamma\left(1 + \frac{1}{k}\right) \quad (2.6)$$

$$\sigma = c\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right]^{\frac{1}{2}} \quad (2.7)$$

Donde \bar{v} es el promedio y σ la desviación estándar de los datos de velocidad del viento.

2. **Empirical:** Considerado un caso especial del método del momento. Los parámetros son calculados de la siguiente forma:

$$k = \left(\frac{\sigma}{\bar{v}}\right)^{-1.086} \quad (2.8)$$

$$c = \frac{\bar{v}}{\Gamma\left(1 + \frac{1}{k}\right)} \quad (2.9)$$

3. **Graphical:** Se ajustan rectas a los datos de velocidad del viento usando mínimos cuadrados. Con una doble transformación logarítmica, la función de distribución acumulativa queda:

$$\ln\{-\ln[1 - F(v)]\} = k \ln(v) - k \ln(c) \quad (2.10)$$

Realizando un gráfico para $\ln(v)$ en vez de $\ln(-\ln(1 - F(v)))$, la pendiente de la recta que se ajusta mejor a los pares de datos es el parámetro de la forma de la distribución de Weibull. El parámetro de escala se obtiene por la intersección con la coordenada y .

4. **Maximum likelihood:** En este métodos, son necesarias muchas iteraciones. Los parámetros de Weibull están dado por:

$$k = \left[\frac{\sum_{i=1}^n v_i^k \ln(v_i)}{\sum_{i=1}^n v_i^k} - \frac{\sum_{i=1}^n \ln(v_i)}{n} \right]^{-1} \quad (2.11)$$

$$c = \left(\frac{1}{n} \sum_{i=1}^n v_i^k \right)^{\frac{1}{k}} \quad (2.12)$$

Donde v_i es la velocidad del viento en el paso i y n es el número de puntos de datos distintos de cero.

5. **Modified maximum likelihood:** Este método es utilizado si es que se tiene disponible los datos de velocidad del viento en una distribución de frecuencias. Los parámetros de Weibull son calculados como:

$$k = \left[\frac{\sum_{i=1}^n v_i^k \ln(v_i) f(v_i)}{\sum_{i=1}^n v_i^k f(v_i)} - \frac{\sum_{i=1}^n \ln(v_i) f(v_i)}{f(v \geq 0)} \right]^{-1} \quad (2.13)$$

$$c = \left[\frac{1}{f(v \geq 0)} \sum_{i=1}^n v_i^k f(v_i) \right]^{1/k} \quad (2.14)$$

Donde v_i es la velocidad del viento central al intervalo i , n es el número de intervalos. $f(v_i)$ es la frecuencia de la velocidad del viento dentro del intervalo i y $f(v \geq 0)$ la probabilidad de que la velocidad del viento sea mayor o igual a cero.

6. **Energy pattern factor method:** El factor del patrón de energía es definido como:

$$E_{pf} = \frac{\bar{v}^3}{\bar{v}^3} \quad (2.15)$$

Donde \bar{v}^3 es el promedio de las velocidades del viento cúbicas. Los parámetros de Weibull pueden ser calculados como:

$$k = 1 + \frac{3,69}{E_{pf}^2} \quad (2.16)$$

$$c = \frac{\bar{v}}{\Gamma(1 + \frac{1}{k})} \quad (2.17)$$

Estos métodos fueron comparados a través de pruebas de desempeño, con una simulación de Monte Carlo para este caso, y el análisis de los datos del viento con criterios tales como el test Kolmogorov-Smirnov, *parameter error*, *root mean square error* y el error de energía del viento. De ello, bajo distintas condiciones ciertos métodos se comportan mejor que otros ajustando la distribución de Weibull a los datos de prueba. Sin embargo, como se verá a continuación, una propuesta realizada para mejorar el ajuste a través del uso de la meta-heurística *Particle Swarm Optimization*, mejora la calidad de los resultados comparado con estos métodos numéricos presentados.

2.2.3. Particle Swarm Optimization

En Carneiro et al. [3], se realiza un caso de estudio de las características del viento en las zonas costeras de Parnaíba y Maracanaú, y en una zona interior, Petrolina, en Brasil. Allí se explica la necesidad de obtener un modelo para el comportamiento estocástico del viento, de manera de poder evaluar el potencial energético de aquellas regiones. Como se menciona anteriormente, el modelo utilizado es la distribución de Weibull. Para ello, es preciso ajustar el modelo a los datos recolectados. Por esto, en el estudio mencionado, se propone un PSO para encontrar los parámetros k y c de la distribución de Weibull y a su vez mejorar la calidad de la solución comparada con los métodos numéricos tradicionales. Así, para lograr el objetivo, la función de aptitud se define como:

$$\epsilon(v_i) = \frac{1}{2} \sum_{i=0}^n (f_{real}(v_i) - f_{weibull}(v_i))^2 \quad (2.18)$$

Donde ϵ , es el error cuadrático a minimizar entre los valores del histograma de datos y la función de distribución de Weibull.

El PSO utilizado es el modelo clásico presentado en la sección anterior, considerando los parámetros w , $c1$ y $c2$, sin embargo, para abolir la convergencia prematura, se establece que estos parámetros varíen durante la ejecución del algoritmo dentro de un rango definido ($w \in \{0,4, 0,9\}$, $c1$ y $c2 \in \{0, 2,5\}$), en donde se privilegia la exploración en el inicio de las iteraciones, para posteriormente favorecer la explotación al final de la ejecución.

Finalmente, para evaluar los resultados de la propuesta, se compara con el PSO con cinco de los seis métodos numéricos mencionados anteriormente utilizados para la estimación de los parámetros de Weibull: *Moment Method* (M), *Energy Method* (E), *Energy Pattern Factor Method* (EPF), *Energy Equivalent Method* (EE) y *Maximum Likelihood* (ML). Además, para evaluar la eficiencia de los métodos, se utilizan tres *test* estadísticos: *correlation* (r), *relative bias* (RB) y *root mean square error* (RMSE).

Los resultados que se exponen en el trabajo citado son alentadores, demostrando que el PSO obtiene los mejores resultados de ajuste a los datos. Un ejemplo de esto es expuesto en la figura 2.3. En Kongnam et al. [7], el PSO es utilizado para el problema del control

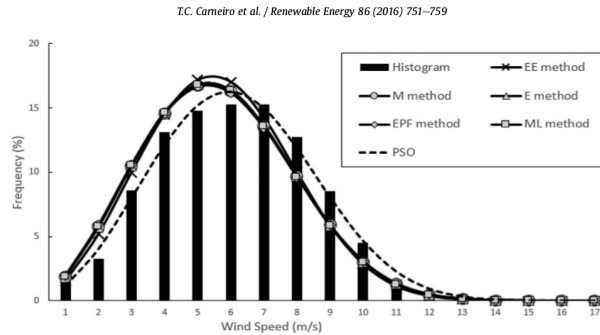


Figura 2.3: Distribución de Weibull con histograma - Maracanaú
Creado por Carneiro et al. [3].

de la velocidad de las turbinas de viento para maximizar la generación de energía. En este trabajo, se utiliza la distribución de Weibull para el modelado de la velocidad del viento. La construcción del PSO es llevada a cabo considerando el problema de la convergencia prematura, por lo que se desarrollan funciones que varían estos parámetros a lo largo de la ejecución.

2.3. Dirección del viento

La dirección del viento es información esencial para la investigación acerca de la energía eólica, dado que con esta, por ejemplo, se pueden ubicar de forma estratégica las turbinas que capturan la energía. En el resumen acerca de las energías renovables y sustentables [10], se explica que para identificar la dirección dominante del viento la función de densidad *finite von Mises-Fisher* (FVMF) es utilizada para ajustarse a los datos. Para las pruebas, estos datos fueron obtenidos de cinco estaciones ubicadas en distintas zonas en la península de Malasia. La FVMF, de forma genérica, está definida de la siguiente forma:

$$f(x; \mu_h, k_h) = \sum_{h=1}^H (w_h) \frac{k^{\frac{d}{2}-1}}{(2\pi)^{\frac{d}{2}} I_{\frac{d}{2}-1}(k)} e^{(k_h \mu_h^T x)} \quad (2.19)$$

Donde $x = [\cos(\theta_i), \sin(\theta_i)]$, $\frac{k^{\frac{d}{2}-1}}{(2\pi)^{\frac{d}{2}} I_{\frac{d}{2}-1}(k)}$ es una constante de normalización, d es la dimensión del vector aleatorio x ($d = 2$, para este caso), μ_h es el parámetro de dirección predominante (análogo a la media μ en la distribución normal), k_h es el parámetro de concentración (análogo al recíproco de la dispersión σ^2), estos dos últimos para cada $h = 1, 2, \dots, H$ componente del FVFM y w_h es el parámetro de mezcla o peso de las funciones de von Mises (*mixture parameter*).

Además, el parámetro de mezcla del FVMF está sujeto a la siguiente restricción:

$$0 \leq w_h \leq 1 \text{ y } \sum_{h=1}^H w_h = 1 \text{ para } (h = 1, 2, \dots, H) \quad (2.20)$$

Para estimar los parámetros del FVMF, se sugiere utilizar el método *expectation maximization*, debido a que los métodos regulares son incapaces de manejar la complejidad del modelo, consideraciones que se mencionan en el trabajo de Banerjee et al. [11].

Por último, los resultados de este trabajo muestran que FVMF provee un razonable ajuste a diferentes conjunto de datos, obteniendo un modelo que explica más del 90 % de la variación de los datos, en este caso, obtenidos de estaciones ubicadas en la península de Malasia. En la figura 2.4 se aprecia el ajuste del modelo a los datos, tanto la comparación con el histograma, como en su versión circular.

En el trabajo de Heckenbergerova et al. [4], se utiliza una estrategia diferente a la

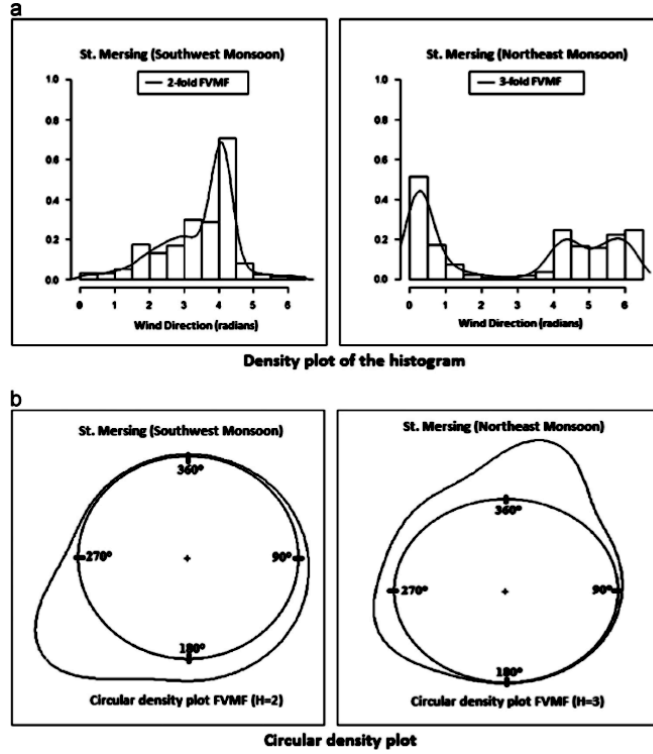


Figura 2.4: Modelo de ajuste FVMV para suroeste y noreste en la estación Mersing
Creado por [10].

anteriormente mencionada. Basados en la ya mencionada meta-heurística inspirada en la biología, *Particle Swarm Optimization*, proponen una forma distinta para encontrar un modelo de ajuste, utilizando la distribución estadística *finite mixture of circular normal von Mises* (MvM), similar a la mencionada previamente.

En este caso, se define la *simple von Mises distribution* (SvM) como:

$$f(\theta; \mu, k) = \frac{1}{2\pi I_0(k)} e^{k \cos(\theta - \mu)} \quad (2.21)$$

Donde $k \geq 0$, $0 \leq \mu \leq 2\pi$, $0 \leq \theta \leq 2\pi$ y $I_0(k)$ representa la versión modificada de la función de Bessel de primera clase y orden cero:

$$I_0(k) = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} e^{k \cos(\theta)} d\theta = \sum_{k=0}^{\infty} \frac{1}{(k!)^2} \left(\frac{k}{2}\right)^{2k} \quad (2.22)$$

Para $k = 0$, la distribución SvM se vuelve uniforme alrededor de un círculo con todas las direcciones equi-probables. Cuando una colección de datos tiene más de una dirección predominante, es necesario utilizar una mezcla (*mixture*) de distribuciones. Así, la función

de densidad de probabilidad *finite mixture of simple von Mises* (MvM-pdf) queda como:

$$\phi(\theta; v) = \sum_{j=1}^k w_j \cdot f_j(\theta; \mu_j, k_j) \quad (2.23)$$

Donde k es el número de funciones de la mezcla, j es el índice de una particular SvM-pdf con parámetros μ_j y k_j , θ es una variable angular ($0 \leq \theta \leq 2\pi$), y v es un vector parámetro de la forma:

$$v = (\mu, k, w) = (\mu_1, \dots, \mu_k, k_1, \dots, k_k, w_1, \dots, w_k) \quad (2.24)$$

Para lograr el objetivo, se obtiene en primer lugar una aproximación numérica de los parámetros del MvM a partir de los datos recolectados de la dirección del viento, estrategia nombrada como estimación analítica en el trabajo de Heckenbergerova et al., para luego optimizarlos mediante el uso de un PSO, en su versión modificada, para evitar la convergencia prematura, en donde la solución está representada por una codificación del vector \vec{v} mencionado anteriormente 2.24.

Como test estadístico, es utilizado el *Pearson's chi-squared goodness-off-fit*. Los resultados muestran la mejora que se logra a la estimación inicial, comparándose además con algoritmos genéticos. Sin embargo, estos no consiguen pasar el test estadístico impuesto, por lo que existe trabajo futuro a realizar para mejorar la propuesta y lograr la precisión deseada. Los resultados obtenidos para los datos recolectados en el aeropuerto de St John localizado en Newfoundland, Canadá, son apreciables en la figura 2.5.

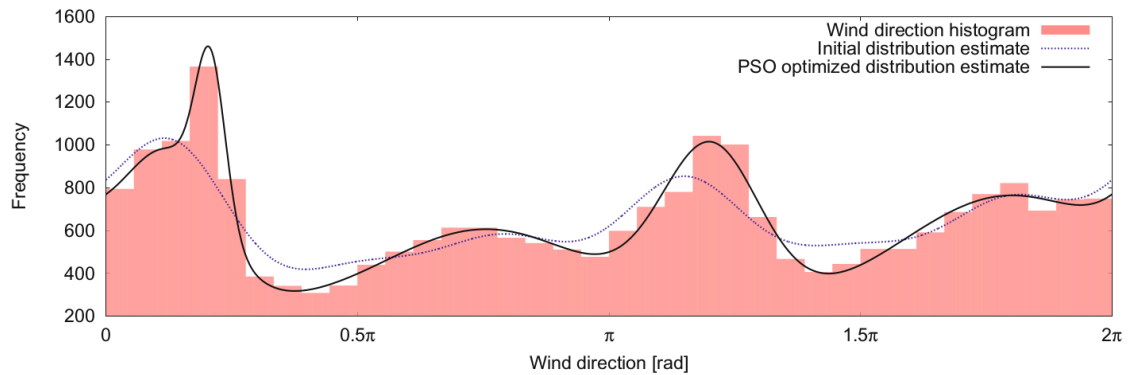


Figura 2.5: Ajuste dirección del viento, aeropuerto St. John
Creado por Heckenbergerova et al. [4].

Capítulo 3

Desarrollo de la solución

3.1. Velocidad del viento

3.1.1. Modelo Matemático

Como se adelanta anteriormente, para encontrar los parámetros de la distribución de Weibull que se ajusten a los datos de prueba se utilizará el *Particle Swarm Optimization*. La función de distribución de Weibull está definida en la ecuación 2.5. El PSO a utilizar está representado por la ecuación 2.3. La función objetivo se describe con la fórmula 2.18 y es aquella con la que se busca minimizar el error cuadrático entre la frecuencia real de los datos y la estimada por la distribución de Weibull. Los parámetros a encontrar k y c deben ser ≥ 0 . Por último, a modo de favorecer la exploración al comienzo y la explotación al final de las iteraciones del PSO, se utilizará la recomendación de [3] para la variación de parámetros del enjambre:

$$w(j) = (1 - \frac{j}{iter_{max}})^{\alpha}(w_{max} - w_{min} + w_{min}) \quad (3.1)$$

$$c_1(j) = (1 - \frac{j}{iter_{max}})^{\beta}(c_{1max} - c_{1min}) + c_{1min} \quad (3.2)$$

$$c_2(j) = (1 - \frac{j}{iter_{max}})^{\gamma}(c_{2min} - c_{2max}) + c_{2max} \quad (3.3)$$

Donde $w_{max} = 0,9$ y $w_{min} = 0,4$, c_{1max} , c_{2max} y c_{1min} , c_{2min} son 2.5 y 0 respectivamente. Los parámetros α, β, γ son definidos como 0.5, 1.5 y 1.0 respectivamente. $iter_{max}$ es el máximo número de iteraciones.

3.1.2. Representación

Cada vector posición de las partículas del enjambre representa una solución candidata la cual varía dentro de cierto espacio de búsqueda definido por los límites de las componentes. Así, para el caso de los parámetros de la distribución de Weibull, la posición de las partículas está representada por los parámetros k y c quedando de la forma:

$$x = (k, c) \quad (3.4)$$

Para ambos parámetros, se establecen los límites entre $0 \leq (k, c) \leq 20$, criterio que se basa en el trabajo de Carneiro et al. [3]. De esta forma, las partículas se moverán dentro de ese rango, manteniéndose en los lugares que minimizan la función objetivo, la cual representa el error de la predicción de Weibull versus los datos reales.

Así, para cada partícula se define una estructura que posee las siguientes propiedades:

1. Posición: vector de números flotantes de largo dos, los cuales representan la ubicación de la partícula dentro del espacio de búsqueda y sus componentes a los parámetros k y c .
2. Velocidad: vector de números flotantes que representan el cambio de valor de cada componente de la posición de la partícula en determinada iteración. Se actualiza en base a la ecuación de velocidad del PSO.
3. Mejor resultado personal: vector flotante que guarda la mejor posición conseguida por la partícula durante las iteraciones transcurridas.

Mientras que el enjambre, siendo esencialmente una estructura que posee referencia a todas las partículas, queda representado de la siguiente forma:

1. Partículas: Arreglo de referencias a las estructuras de partículas creadas.
2. Mejor posición global: De todos los mejores resultados particulares a cada partícula, se almacena la mejor posición de todas. La que persiste al final del ciclo de iteraciones, es la solución final.
3. W, C1 y C2: Son los parámetros de inercia, cognitivo y social respectivamente.

3.1.3. Descripción del algoritmo

La lógica del algoritmo 1 se basa en mover las partículas dentro del rango definido para los componentes de la solución hasta que todas las partículas se concentren en alguna zona que represente una buena solución al problema, no necesariamente el óptimo. Lo importante en cada iteración es actualizar o mover el enjambre, revisar y guardar las mejores soluciones y actualizar los parámetros de inercia, cognitivo y social que definen las velocidades. Las

Algorithm 1 PSO para el ajuste de los parámetros de la distribución de Weibull

Require: Datos de frecuencias de velocidades del viento.

Ensure: Valores para los parámetros k y c .

```
enjambre = inicializar(w,c1,c2)
for  $i = 1$  to  $Iter_{max}$  do
  for Each partículas en enjambre do
    actualizarVelocidadPartícula(partícula)
    actualizarPosiciónPartícula(partícula)
    revisarLímitesPosición(partícula)
    guardarMejorResultadoPartícula(partícula)
  end for
  guardarMejorResultadoGlobal(enjambre)
  actualizarParámetros(enjambre)
end for
retornarMejorResultadoGlobal(enjambre).
```

iteraciones fueron probadas hasta un máximo de 1000 y 50 partículas, (A excepción del experimento donde se consideraron todos los promedios diarios, 2013, 2014 y 2015, en el cual, se utilizaron 200 partículas). Los parámetros de w , $c1$ y $c2$ fueron definidos tal y como explica en el modelo matemático, en la sección 3.1.1.

3.1.4. Experimentos

Los experimentos fueron realizados con datos del viento obtenidos por la Armada de Chile para la región de Valparaíso en los años 2013, 2014 y 2015. Estos fueron tratados mediante *scripts* desarrollados en python para obtener las frecuencias de las distintas velocidades del viento registradas a lo largo del año. Los datos se organizaban de la siguiente forma: Por cada año, se tiene una tabla en un archivo excel de cada mes, en donde se registra por cada fila los resultados de la medición de cada día. Las mediciones son

registradas en un intervalo de tres horas, es decir, se tienen registros diarios para las 3:00, 6:00, 9:00, 12:00, 15:00, 18:00, 21:00 y 00:00 horas.

Un ejemplo es la tabla mostrada en la figura 3.1. El ajuste de la distribución de Weibull

| | | ESTACIÓN VALPARAÍSO | | | | MES / AÑO ENERO 2015 | | | | | | | | | | | | | | | |
|-----|-----------|---------------------|-----------|------|-----------|----------------------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|------|------|------|------|------|------|
| Dia | 03:00 UTC | | 06:00 UTC | | 09:00 UTC | | 12:00 UTC | | 15:00 UTC | | 18:00 UTC | | 21:00 UTC | | 24:00 UTC | | | | MAX. | | MED. |
| | Dir. | Int. | Dir. | Int. | Dir. | Int. | Dir. | Int. | Dir. | Int. | Dir. | Int. | Dir. | Int. | Dir. | Int. | Dir. | Int. | Dir. | Int. | |
| 1 | 220 | 10 | 200 | 03 | 000 | 00 | 280 | 05 | 230 | 10 | 230 | 15 | 220 | 18 | 210 | 10 | 220 | 18 | 220 | 18 | 10 |
| 2 | 000 | 00 | 250 | 02 | 080 | 04 | 000 | 00 | 270 | 06 | 240 | 13 | 240 | 12 | 200 | 04 | 240 | 13 | 240 | 13 | 04 |
| 3 | 000 | 00 | 090 | 03 | 030 | 04 | 050 | 09 | 080 | 04 | 350 | 06 | 310 | 04 | 250 | 05 | 050 | 09 | 050 | 09 | 04 |
| 4 | 260 | 06 | 260 | 05 | 040 | 04 | 010 | 04 | 010 | 03 | 270 | 03 | 270 | 09 | 270 | 02 | 270 | 09 | 270 | 09 | 04 |
| 5 | 000 | 00 | 250 | 05 | 280 | 04 | 010 | 03 | 040 | 03 | 030 | 04 | 310 | 06 | 240 | 03 | 310 | 06 | 310 | 06 | 04 |
| 6 | 170 | 03 | 210 | 06 | 200 | 05 | 230 | 05 | 310 | 02 | 260 | 04 | 270 | 03 | 000 | 00 | 210 | 06 | 210 | 06 | 04 |
| 7 | 000 | 00 | 050 | 08 | 010 | 09 | 010 | 10 | 020 | 09 | 320 | 05 | 240 | 07 | 240 | 10 | 010 | 10 | 010 | 10 | 09 |
| 8 | 220 | 04 | 000 | 00 | 100 | 07 | 230 | 04 | 230 | 10 | 230 | 15 | 230 | 14 | 240 | 12 | 230 | 15 | 230 | 15 | 09 |
| 9 | 160 | 02 | 090 | 02 | 100 | 03 | 060 | 03 | 030 | 03 | 020 | 05 | 310 | 04 | 250 | 06 | 250 | 06 | 250 | 06 | 03 |
| 10 | 230 | 08 | 020 | 05 | 000 | 00 | 020 | 03 | 300 | 04 | 240 | 06 | 240 | 15 | 240 | 09 | 240 | 15 | 240 | 15 | 06 |
| 11 | 240 | 07 | 000 | 00 | 110 | 04 | 300 | 04 | 000 | 00 | 310 | 05 | 290 | 06 | 240 | 06 | 240 | 07 | 240 | 07 | 05 |
| 12 | 210 | 02 | 110 | 05 | 000 | 00 | 100 | 02 | 290 | 03 | 300 | 06 | 240 | 07 | 140 | 04 | 240 | 07 | 240 | 07 | 04 |
| 13 | 120 | 08 | 120 | 04 | 200 | 04 | 260 | 03 | 330 | 03 | 270 | 09 | 240 | 10 | 230 | 13 | 230 | 13 | 230 | 13 | 06 |
| 14 | 220 | 06 | 000 | 00 | 120 | 03 | 090 | 02 | 350 | 08 | 260 | 06 | 240 | 09 | 260 | 05 | 240 | 09 | 240 | 09 | 06 |
| 15 | 280 | 05 | 240 | 08 | 000 | 00 | 000 | 00 | 270 | 03 | 270 | 07 | 270 | 09 | 230 | 09 | 270 | 09 | 270 | 09 | 06 |
| 16 | 240 | 04 | 260 | 03 | 020 | 02 | 310 | 03 | 310 | 02 | 290 | 04 | 250 | 07 | 250 | 08 | 250 | 08 | 250 | 08 | 04 |
| 17 | 270 | 02 | 250 | 03 | 020 | 02 | 020 | 07 | 010 | 06 | 350 | 05 | 300 | 02 | 220 | 07 | 020 | 07 | 020 | 07 | 04 |
| 18 | 220 | 02 | 280 | 04 | 270 | 02 | 010 | 04 | 010 | 13 | 010 | 08 | 310 | 04 | 270 | 03 | 010 | 13 | 010 | 13 | 04 |
| 19 | 240 | 04 | 000 | 00 | 020 | 03 | 350 | 10 | 020 | 03 | 350 | 02 | 270 | 03 | 250 | 09 | 350 | 10 | 350 | 10 | 03 |
| 20 | 250 | 06 | 050 | 03 | 030 | 06 | 000 | 00 | 340 | 02 | 230 | 15 | 220 | 12 | 210 | 12 | 230 | 15 | 230 | 15 | 06 |
| 21 | 240 | 03 | 000 | 00 | 240 | 09 | 240 | 05 | 260 | 05 | 240 | 13 | 230 | 12 | 240 | 11 | 240 | 13 | 240 | 13 | 07 |
| 22 | 170 | 04 | 200 | 08 | 190 | 06 | 000 | 00 | 230 | 06 | 230 | 17 | 230 | 13 | 240 | 06 | 230 | 17 | 230 | 17 | 06 |
| 23 | 240 | 06 | 240 | 04 | 070 | 06 | 080 | 03 | 040 | 06 | 360 | 02 | 240 | 16 | 210 | 05 | 240 | 16 | 240 | 16 | 06 |
| 24 | 240 | 07 | 000 | 00 | 130 | 02 | 140 | 02 | 280 | 05 | 230 | 13 | 240 | 22 | 270 | 05 | 240 | 22 | 240 | 22 | 05 |
| 25 | 000 | 00 | 250 | 02 | 150 | 04 | 090 | 03 | 050 | 04 | 230 | 10 | 230 | 13 | 230 | 12 | 230 | 13 | 230 | 13 | 04 |
| 26 | 200 | 07 | 200 | 06 | 000 | 00 | 310 | 02 | 250 | 08 | 250 | 15 | 060 | 04 | 220 | 10 | 250 | 15 | 250 | 15 | 07 |
| 27 | 210 | 06 | 180 | 10 | 210 | 05 | 250 | 02 | 010 | 06 | 040 | 03 | 270 | 07 | 250 | 12 | 250 | 12 | 250 | 12 | 06 |
| 28 | 240 | 09 | 000 | 00 | 000 | 00 | 300 | 02 | 200 | 07 | 270 | 07 | 250 | 10 | 000 | 00 | 250 | 10 | 250 | 10 | 05 |
| 29 | 000 | 00 | 000 | 00 | 210 | 03 | 110 | 02 | 090 | 03 | 230 | 16 | 230 | 15 | 220 | 11 | 230 | 16 | 230 | 16 | 03 |
| 30 | 180 | 02 | 220 | 11 | 230 | 06 | 260 | 08 | 240 | 07 | 250 | 18 | 230 | 20 | 230 | 12 | 230 | 20 | 230 | 20 | 10 |
| 31 | 230 | 06 | 240 | 03 | 240 | 05 | 000 | 00 | 260 | 04 | 250 | 15 | 230 | 11 | 210 | 07 | 250 | 15 | 250 | 15 | 06 |

PROMEDIOS MENSUALES

12

05

MÁXIMA INTENSIDAD MENSUAL

22

Figura 3.1: Ejemplo colección de datos Enero Valparaíso 2015
Obtenido desde el Instituto Meteorológico de la Armada de Chile.

a estos datos de velocidad del viento se hizo considerando las siguientes configuraciones para el cálculo del histograma de frecuencias:

1. **Todos los años.** Se considera el promedio diario de velocidad del viento como dato unitario para el cálculo de las frecuencias, considerando todos los días en el intervalo de Enero del 2013 hasta Diciembre del 2015.
2. **Anual.** Se considera el promedio diario de velocidad del viento como dato unitario para el cálculo de las frecuencias en un lapso anual (2013, 2014 y 2015).
3. **Por temporada.** Se considera el promedio diario de velocidad del viento como dato unitario para el cálculo de las frecuencias en un lapso de tres meses (Enero - Marzo ; Abril - Junio; Julio - Septiembre; Octubre - Diciembre).

4. **Datos brutos.** Se considera cada medición realizada (8 por día) como dato unitario, en un lapso de un año.

Una vez obtenido los datos de frecuencias, se procede a aplicar el algoritmo PSO obteniendo los parámetros de ajuste k y c . De esta manera, se evalúa la calidad del modelo generado (distribución de Weibull), para las distintas configuraciones mediante gráficos y los siguientes test estadísticos (utilizados en el trabajo de Carneiro et al. [3]):

1. *Root Mean Square Error*

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (X_i - Y_i)^2}{N}} \quad (3.5)$$

2. *Correlation*

$$r = \frac{\sum_{i=1}^N (X_i - X_{med}) \cdot (Y_i - Y_{med})}{\sqrt{\sum_{i=1}^N (X_i - X_{med})^2} \cdot \sqrt{\sum_{i=1}^N (Y_i - Y_{med})^2}} \quad (3.6)$$

3. *Relative Bias*

$$RB = \frac{X_{med} - Y_{med}}{Y_{med}} \quad (3.7)$$

Donde N es el número de datos, Y_i la frecuencia de dichos datos, X_i la frecuencia entregada por la distribución de Weibull, X_{med} la media de X_i e Y_{med} la media de Y_i .

Las pruebas fueron realizadas en un computador con sistema operativo Ubuntu 16.04 64-bit, 3.8 GB de memoria y procesador doble núcleo Intel Pentium 2.60 GHz.

3.2. Dirección del viento

3.2.1. Modelo Matemático

Como se comenta anteriormente, la distribución de densidad de probabilidad que se utilizará para describir la distribución de datos de dirección del viento es la *mixture von mises distribution* descrita en 2.23, la cual consiste básicamente en una combinación lineal de la *simple von Mises distribution* descrita en 2.21.

De forma preliminar, los datos se ordenan en un histograma de densidad con el cual se obtiene un esqueleto de la distribución de densidad de probabilidad. Posteriormente se requieren encontrar los parámetros de ajuste μ_j , k_j y w_j para cada j -ésima *simple von Mises distribution*. La forma en que se realiza esto último en este trabajo está basado en José A. Carta et al. [12] y se describe a continuación.

Para la construcción del histograma se divide el rango de datos que va de 0 a 2π en T clases con frecuencia O_i la cual representa la suma de las observaciones en el rango de la clase T . Posteriormente se definen k sectores del mismo largo desde las T clases, relacionados al número de direcciones de viento predominantes (o con mayor frecuencia). Esto define el número de funciones de von Mises a utilizar. La estimación de k se realiza mediante la observación del histograma generado, un análisis cualitativo de las direcciones predominantes en los datos.

Para la aproximación inicial de los parámetros de la *mixture von mises distribution* se utiliza una estimación numérica basada en los datos recolectados acerca de la dirección del viento. Sea $j \in \{1...k\}$ el subíndice del sector representado por la j -ésima función de von Mises. La dirección del viento predominante μ_j se estima de la siguiente forma:

$$\mu_j = \begin{cases} \arctan(\frac{s_j}{c_j}) & s_j \geq 0, c_j > 0 \\ \frac{\pi}{2} & s_j > 0, c_j = 0 \\ \pi + \arctan(\frac{s_j}{c_j}) & c_j < 0 \\ \pi & s_j > 0, c_j = -1 \\ 2\pi + \arctan(\frac{s_j}{c_j}) & s_j < 0, c_j > 0 \\ 3\frac{\pi}{2} & s_j < 0, c_j = 0 \end{cases} \quad (3.8)$$

En donde s_j y c_j representan el seno y coseno promedio del sector j .

Tradicionalmente, se estima el parámetro de concentración k_j con la ecuación:

$$\frac{I_1(k_j)}{I_0(k_j)} = \sqrt{s_j^2 + c_j^2} \quad (3.9)$$

Donde $I_1(k_j)$ es la función modificada de Bessel de primera clase y orden 1. Como se explica en Banerjee et al. [11], debido a la falta de una solución analítica a la ecuación 3.9, no es posible estimar directamente los valores de k . Se podrían utilizar métodos para ecuaciones no lineales, pero para datos de altas dimensiones, problemas de desbordamiento

(*overflow*) o inestabilidad numérica se vuelven concurrentes. Por tanto, se utiliza la propuesta realizada en el trabajo de Heckenbergerova et al. [4] con lo cual el parámetro k_j puede ser aproximado por:

$$|k_j| = \{23,29041409 - 16,8617370\sqrt{s_j^2 + c_j^2}\} \quad (3.10)$$

Los pesos iniciales w_j son aproximados como:

$$w_j = \frac{\sum_{i=J_l}^{J_u} O_i}{\sum_{i=1}^T O_i} \quad (3.11)$$

Donde J_l y J_u son los índices de los bordes del sector j .

Esta estimación inicial de los parámetros de la *mixture von mises distribution* es mejorada mediante la meta-heurística *Particle Swarm Optimization*. Para el PSO se utiliza la representación descrita en 2.3 y la modificación a este sugerida en Carneiro et al. [3] descrita previamente en 3.1.

La función objetivo para el PSO es el test estadístico χ^2 descrito en [4] como sigue a continuación:

$$\chi^2 = \sum_{i=1}^T \frac{(O_i - np_i)^2}{np_i} \quad (3.12)$$

Donde T es el número de clases de frecuencia definido para construir el histograma, n es la suma de las frecuencias observadas O_i y p_i es la probabilidad teórica de cada clase de frecuencia predicha por el modelo ajustado.

Para el cálculo del p_i se utiliza:

$$p_i = \int_{l_i}^{u_i} f(x)dx \quad (3.13)$$

Donde u_i y l_i son los bordes de la i -ésima clase de frecuencia.

La forma de la solución a encontrar es descrita en 2.24. Esta es restringida por la condición para los pesos de la *mixture von mises distribution*, la cual obliga a que se deba cumplir que la suma de los pesos sea igual a 1, como se describe en 2.20.

3.2.2. Representación

La representación del PSO es similar al utilizado para el ajuste de la distribución de datos de velocidad del viento. Las partículas y el enjambre están representados por 3.1.2 y 3.1.2 respectivamente.

La solución para el PSO que mejora la estimación inicial de los parámetros para la *mixture von Mises distribution* está representado por un vector v en el cual se encuentran los valores para todos los parámetros de cada *simple von Mises distribution*. Estos valores están codificados para que el algoritmo se mueva en el rango desde 0 a 1.

El vector solución tiene la forma:

$$v = (\overbrace{v_1, \dots, v_k}^{\mu}, \overbrace{v_{k+1}, \dots, v_{2k}}^k, \overbrace{v_{2k+1}, \dots, v_n}^w). \quad (3.14)$$

El parámetro μ está representado en el rango $i \in \{1, \dots, k\}$ y para ser decodificado debe ser escalado por 2π .

El parámetro k está representado en el rango $i \in \{k + 1, \dots, 2k\}$ y para ser decodificado debe ser escalado por $[0, 700]$.

El parámetro w_j está representado en el rango $i \in \{2k + 1, \dots, n\}$ cuyos valores van en el rango $[0, 1]$.

3.2.3. Descripción del algoritmo

El algoritmo para el ajuste de la función de densidad de probabilidad para la dirección del viento se basa en la propuesta de Heckenbergerova et al. [4].

Como se ha ido vislumbrado, consiste en dos fases. La primera, una aproximación basada en la estimación numérica de los parámetros requeridos para la *mixture of von Mises distribution* a través de operaciones simples con los datos recolectados, y la segunda, una mejora de la solución inicial obtenida en la fase anterior mediante el uso de la meta-heurística *Particle Swarm Optimization*.

El algoritmo para la aproximación inicial de la solución se describe en 2. En donde la estimación de los parámetros se realiza como se describe en 3.8 para los μ_j , 3.9 para los k_j y 3.11 para los pesos w_j . Una vez obtenida la aproximación inicial se procede a mejorar esta mediante el uso del PSO. El algoritmo es bastante similar en estructura al desarrollado para la velocidad del viento 1. Sin embargo, existen diferencias importantes, relevantes al

Algorithm 2 Aproximación inicial de los parámetros de la *mixture von Mises distribution*

Require: Datos de frecuencias de la dirección del viento.

Require: K, Cantidad de *simple von Mises distribution*.

Require: T, clases de frecuencias.

Require: D, Total de datos.

Ensure: Valores para los parámetros μ_j , k_j y w_j , para cada $j \in \{1, \dots, k\}$.

```
sol = inicializarVectorSolución(3*K)
for  $j = 0$  to  $K$  do
  datosj = datosEnRango( $j * D/K$ )
  sj = obtenerSenoPromedio(datosj)
  cj = obtenerCosenoPromedio(datosj)
  uj = obtenerDirecciónPredominante(sj, cj)
  kj = obtenerConcentración(sj, cj)
  wj = obtenerPeso( $j * (T/K)$ ,  $(j + 1) * (T/K)$ )
  addToSolution(sol, uj, kj, wj)
end for
retornarSoluciónInicial(sol).
```

Algorithm 3 PSO para la mejora de la aproximación de los parámetros de la *mixture von Mises distribution*

Require: Datos de la dirección del viento.

Require: Solución inicial para el ajuste de la *mixture von Mises distribution*.

Ensure: Solución inicial mejorada.

```
enjambre = inicializar(w, c1, c2)
for  $i = 1$  to  $Iter_{max}$  do
  for Each partículas en enjambre do
    actualizarVelocidadPartícula(partícula)
    actualizarPosiciónPartícula(partícula)
    revisarLímitesPosición(partícula)
    guardarMejorResultadoPartícula(partícula)
  end for
  guardarMejorResultadoGlobal(enjambre)
  actualizarParámetros(enjambre)
end for
retornarMejorResultadoGlobal(enjambre).
```

problema actual que se destacarán a continuación.

Para la inicialización de las partículas, se realizaron pequeñas perturbaciones a la solución inicial tal y como se sugiere en Heckenbergerova et al. [4]. Esto evita que la solución escape a zonas que tengan un buen resultado en la función objetivo, pero que la forma escape a la del histograma. Debido a que la función objetivo definida 3.12 mide las diferencias de frecuencias entre los datos reales y los teóricos, es decir, las áreas de las barras del

histograma de densidad versus el área bajo la curva de la distribución de probabilidad en algún intervalo, más de una forma de la curva podría parecer una buena solución. (referencia imagen) Por ello, la idea es mantener la forma inicial encontrada, mejorándola sin deformarla. Así, las perturbaciones iniciales a los valores entre 0 y 1 de las posiciones de las partículas eran del orden de 10^{-3} .

La forma en que se cuidaron las condiciones de borde consistieron en limitar el avance de las partículas a los bordes 0 y 1 manteniéndolos en dichos valores si es que se excedían a ellos.

Para cuidar la restricción de pesos se normalizaran los valores determinados en cada iteración, es decir, se suman todos los valores w_j y se ponderan dichos valores por el recíproco de la suma obtenida.

Debido a que la función objetivo implica determinar la frecuencia teórica, es necesario determinar la probabilidad de cierto rango de direcciones mediante el cálculo del área bajo la curva de la distribución de densidad de probabilidad para luego ser multiplicada por la suma del total de datos y así obtener el valor requerido. Por ende, para el cálculo de la integral se utilizaron sumas de Riemann con una partición conveniente al desempeño del algoritmo y la precisión requerida.

Finalmente, la solución obtenida es decodificada tal y como se explica en la sección anterior 3.2.2.

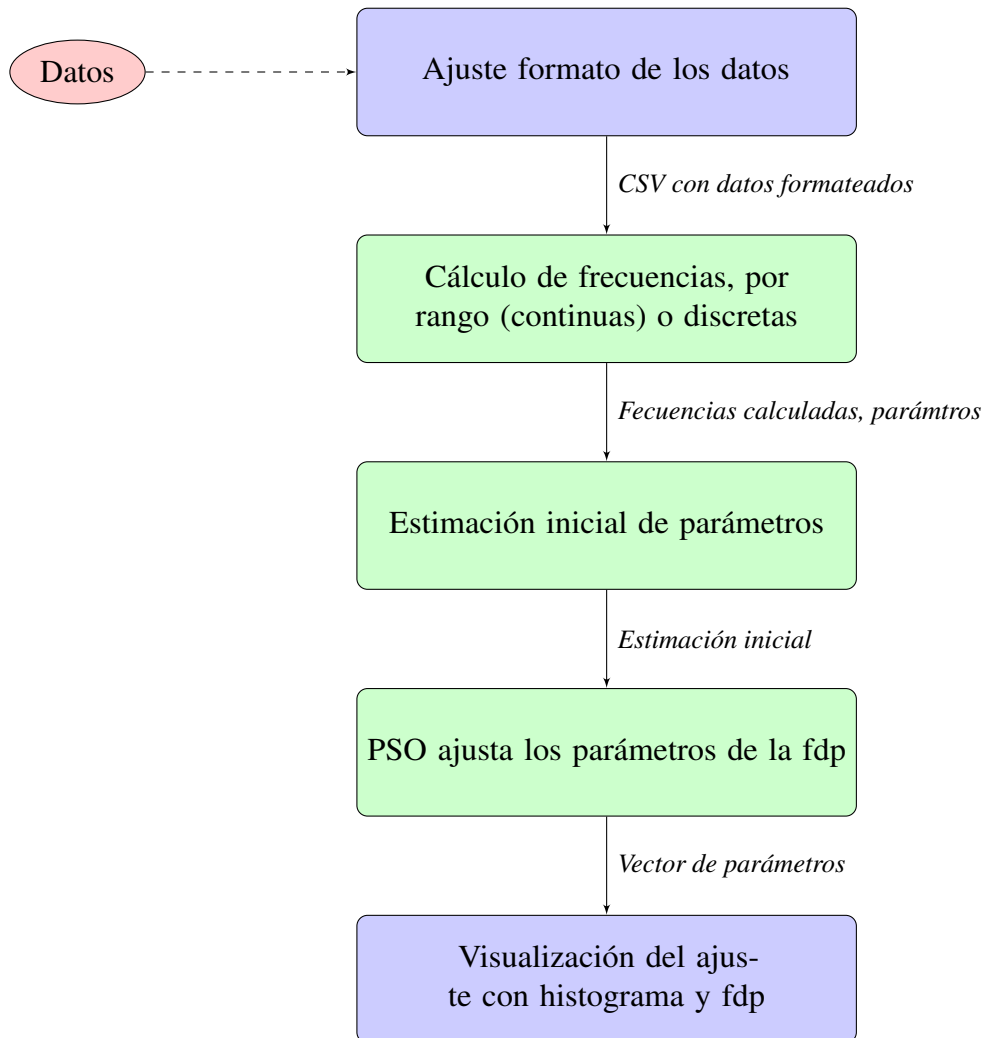
El esquema 3.2 resume el funcionamiento del algoritmo.

3.2.4. Experimentos

ads [13]

sad [14]

Figura 3.2: Esquema de uso del algoritmo



Capítulo 4

Análisis y conclusiones

4.1. Resultados y conclusiones

4.1.1. Análisis de los resultados - velocidad del viento

Visualización de los datos

Los gráficos 4.1(a), 4.1(b) y 4.1(c) muestran la distribución de datos de velocidad del viento en Valparaíso a lo largo de los meses del año y las horas del día, lo cual permite visualizar la naturaleza de la intensidad del tiempo de forma cualitativa. Por ejemplo, se logra apreciar que las máximas velocidades son obtenidas en los meses finales de primavera y comienzos de verano.

Experimento 1, datos anuales y promedios diarios

Las figuras 4.2(a), 4.2(b) y 4.2(c) muestran el ajuste de la distribución de Weibull a los histogramas de datos del viento (promedios diarios), con los parámetros k y c que se muestran en las primeras tres filas de la tabla 4.1 determinados por el PSO. El ajuste tiene buena forma, lo cual es corroborado por los datos estadísticos obtenidos con los test previamente mencionados (RMSE, r , RB), expuestos en la tabla 4.1. Si se compara con la precisión conseguida en el trabajo de Carneiro et al. [3], se aprecia que el ajuste conseguido es levemente más impreciso, sobre todo en lo relativo al test RB. Esto podría deberse a la naturaleza de los datos trabajados.

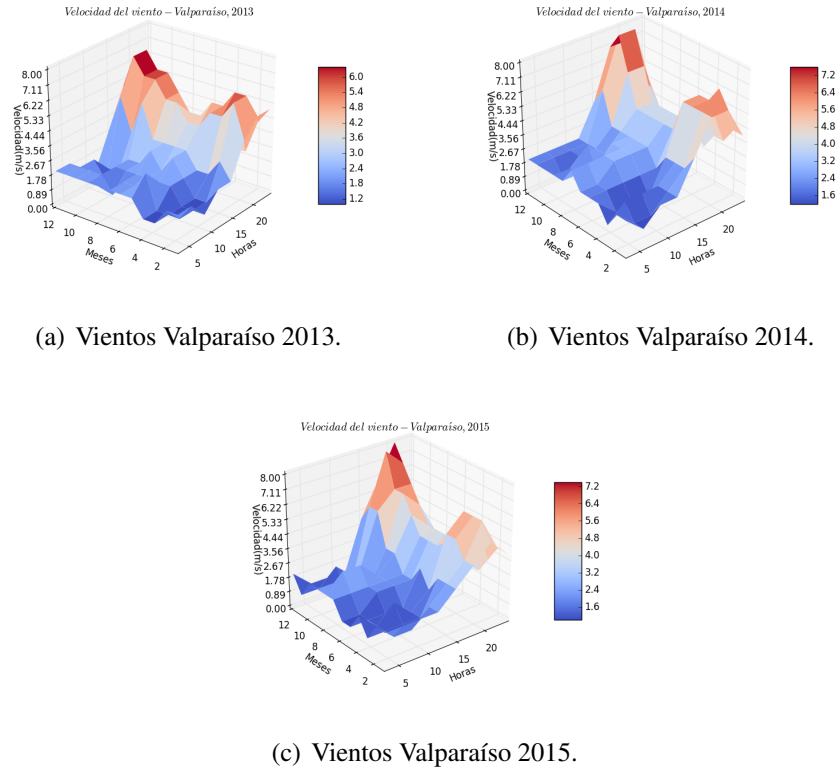


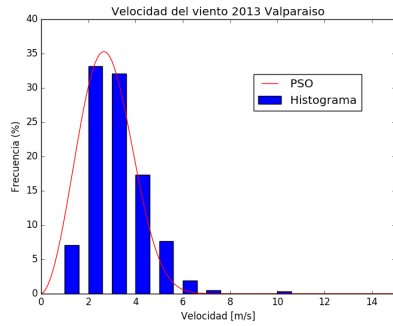
Figura 4.1: Superficie de datos viento de Valparaíso. Fuente: Elaboración Propia.

Experimento 2, datos de tres años y promedios diarios

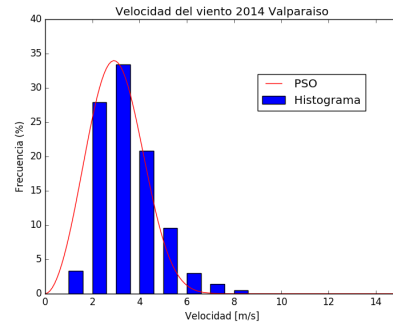
En este experimento se realizó el ajuste considerando los promedios diarios y un intervalo de tres años consecutivos. El gráfico 4.3(b), muestra el resultado del ajuste con PSO y la configuración estándar de los demás experimentos, es decir, 100 iteraciones y 50 partículas. En este gráfico se aprecia que el ajuste no es bueno, a pesar de las cifras en la tabla 4.1, fila 4: PSO (Intento 1), dado que oscila bastante alrededor de las barras del histograma, por lo que se repite el experimento aumentando el número de partículas a 200 obteniendo el gráfico 4.3(a), con el cual se obtiene un ajuste más adecuado, además de mejorar los resultados de los test estadísticos (tabla 4.1, fila 5: PSO (Intento 2)).

Experimento 3, ajuste a datos anuales con resultados del experimento 2

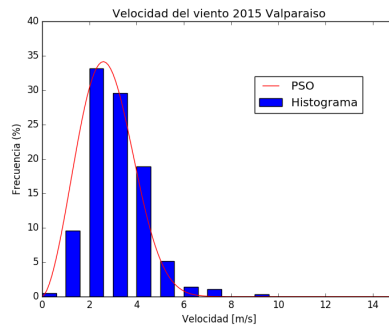
Los gráficos 4.4(a), 4.4(b) y 4.6 son ajustes de Weibull con los parámetros obtenidos en el experimento anterior. Es decir, la idea es evaluar el modelo general de los tres años



(a) PSO Valparaíso 2013.

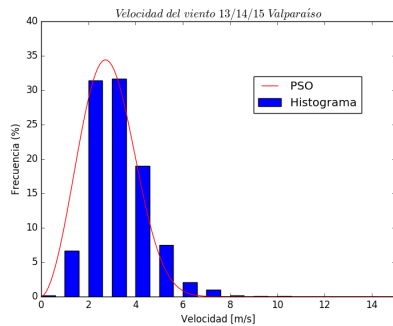


(b) PSO Valparaíso 2014.

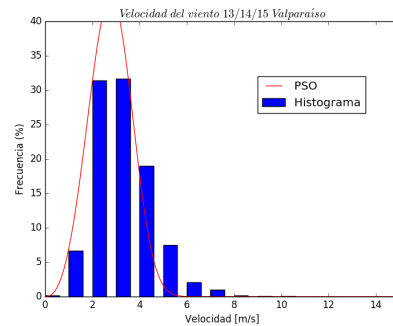


(c) PSO Valparaíso 2015.

Figura 4.2: Ajuste con PSO a datos del viento de Valparaíso. Fuente: Elaboración Propia.



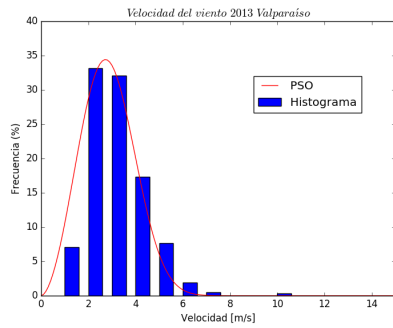
(a) Buen ajuste PSO Valparaíso 2013.



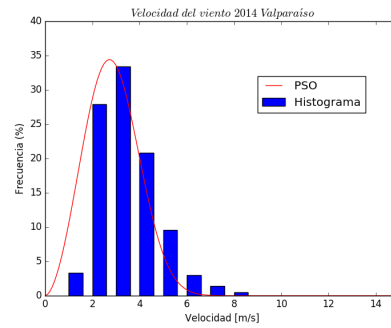
(b) Mal ajuste PSO Valparaíso 2013.

Figura 4.3: Ajuste con PSO a datos Valparaíso 2015, 2014 y 2013, baja y buena calidad. Fuente: Elaboración Propia.

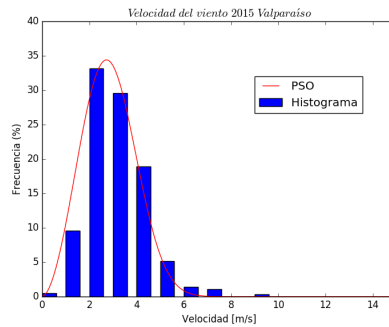
versus el histograma de datos de cada año en particular. El ajuste desde los resultados estadísticos (tabla 4.1), es levemente menos preciso que el modelo ajustado a cada año en particular, pero sigue siendo aceptable como posible opción a considerar.



(a) Velocidad viento Valparaíso 2013.



(b) Velocidad viento Valparaíso 2014.



(c) Velocidad viento Valparaíso 2015.

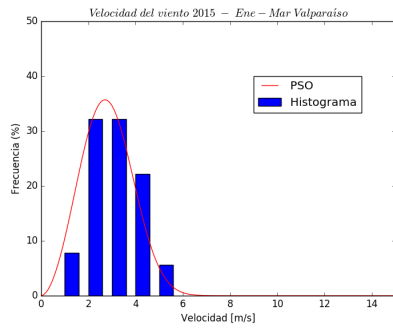
Figura 4.4: Ajuste con PSO a registros del viento en Valparaíso (Con todos los datos).
Fuente: Elaboración Propia.

Experimento 4, ajuste a datos de tres meses y promedios diarios

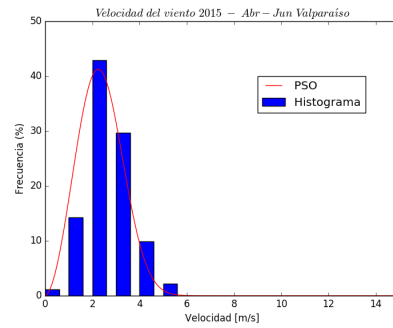
Es posible que se requiera un análisis más acotado, por ello los gráficos 4.5(a), 4.5(b), 4.5(c), 4.5(d), muestran un ajuste considerando un lapso de 3 meses para el año 2015, con el que se demuestra que es posible definir cualquier intervalo (manteniendo como unidad de dato el promedio diario de velocidad del viento) y obtener un ajuste adecuado de los datos mediante la distribución de Weibull.

Experimento 4, ajuste a datos año 2015 y datos brutos

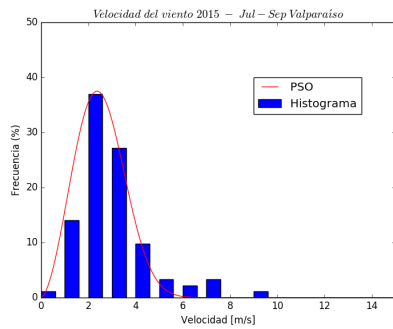
La razón de por qué se utiliza el promedio diario de los datos del viento para ajustar Weibull y no las mediciones puras (las mediciones tomadas cada 3 horas diariamente) es expuesta en el gráfico 4.6. La distribución de Weibull no se ajusta a una distribución de datos con más de un máximo, por lo que de requerirse un modelo para este caso se debe buscar otra distribución o modificar Weibull.



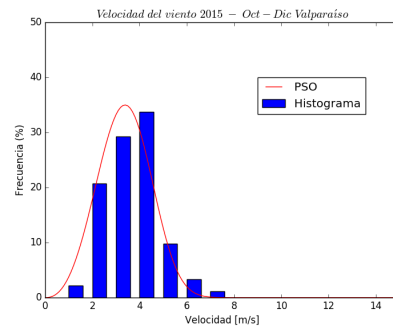
(a) Enero - Marzo.



(b) Abril - Junio.



(c) Julio - Septiembre.



(d) Octubre - Diciembre.

Figura 4.5: Ajuste con PSO a datos Valparaíso 2015, por rango de meses. Fuente: Elaboración Propia.

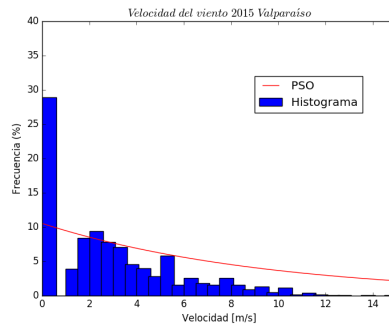


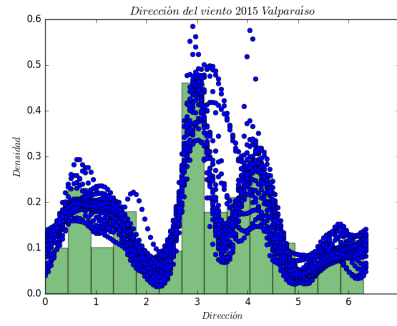
Figura 4.6: Ajuste con PSO a datos (cifras puras) Valparaíso 2015, 2014, 2013
Fuente: Elaboración Propia.

Resumen de los experimentos

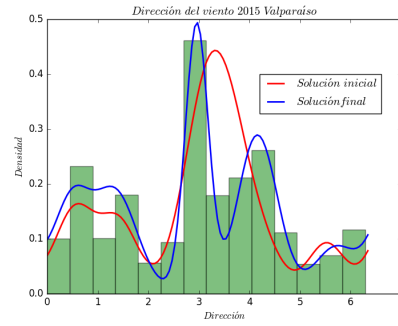
Tabla 4.1: Tabla de tests estadísticos

| # | Método | Período | k | c | RMSE | r | RB |
|----|--------------------|------------|------|------|-----------------|----------------|-------------------|
| 1 | PSO | 2013 | 2.78 | 3.12 | 0.0226585230791 | 0.984353070415 | 0.00197971468299 |
| 2 | PSO | 2014 | 2.91 | 3.37 | 0.0232779965263 | 0.982087745069 | 0.000754465101398 |
| 3 | PSO | 2015 | 2.65 | 3.10 | 0.0164721412159 | 0.992323803649 | 0.00302918178445 |
| 4 | PSO (Intento 1) | 2015-14-13 | 3.47 | 3.07 | 0.0360794587206 | 0.975240385258 | 0.000411212628513 |
| 5 | PSO (Intento 2) | 2015-14-13 | 2.78 | 3.20 | 0.016175531561 | 0.994989105807 | 0.00190916669626 |
| 6 | PSO (Intento 2) | 2013 | 2.78 | 3.20 | 0.0240448436122 | 0.981963054492 | 0.00192186034284 |
| 7 | PSO (Intento 2) | 2014 | 2.78 | 3.20 | 0.0301463089474 | 0.970662237238 | 8.89024791609e-05 |
| 8 | PSO (Intento 2) | 2015 | 2.78 | 3.20 | 0.0202342934641 | 0.98662798667 | 0.00192175053173 |
| 9 | PSO | Ene-Mar | 2.85 | 3.15 | 0.0230380400157 | 0.982158006469 | 0.00641888742608 |
| 10 | PSO | Abr-Jun | 2.76 | 2.65 | 0.0204300909755 | 0.993857185938 | 0.00303620481316 |
| 11 | PSO | Jul-Sep | 2.66 | 2.83 | 0.0251002816356 | 0.985858767021 | 0.00443453471038 |
| 12 | PSO | Oct-Dic | 3.40 | 3.75 | 0.0260278634297 | 0.978479679326 | 0.000716653529598 |
| 13 | PSO (datos brutos) | 2015 | 1.00 | 9.49 | 0.0451794472583 | 0.751732944794 | 0.676094670465 |

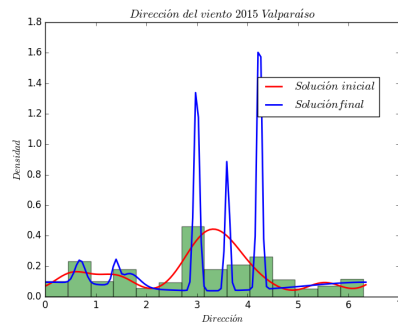
4.1.2. Análisis de los resultados - dirección del viento



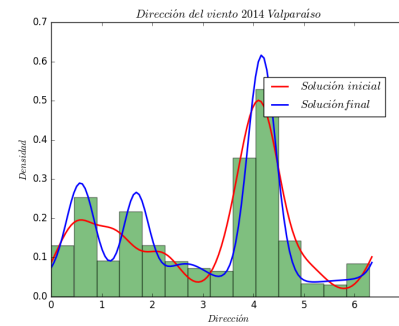
(a) Evolucion de la solucion.



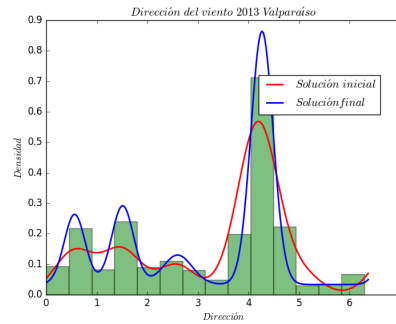
(b) Comparacion solución inicial y final, 2015.



(c) Ejemplo de mal ajuste, 2015.



(d) Solucion inicial y final, 2014.



(e) Solucion inicial y final, 2013.

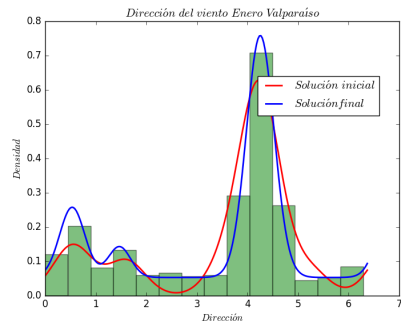
Figura 4.7: Graficos de ajustes anuales

4.1.3. Conclusiones

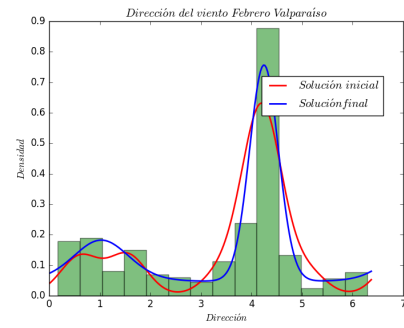
Para obtener información para alguna investigación relacionada a la velocidad del viento en una zona es necesario obtener un modelo que permita explicar las mediciones que se obtienen. Para ello, la distribución de Weibull es una de las funciones más utilizadas para el ajuste de los datos. Distintos casos de estudio alrededor del mundo demuestran la utilidad de la distribución, utilizando diversos métodos para encontrar los parámetros de

ajuste. En este punto, la meta-heurística *Particle Swarm Optimization* ha demostrado ser una alternativa eficiente para este problema, otorgando soluciones de alta calidad.

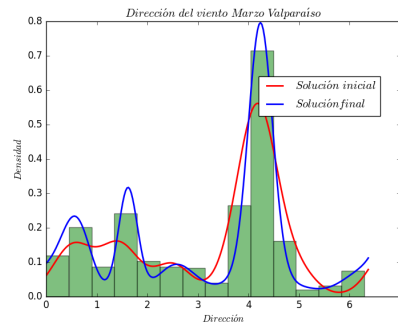
En este trabajo se presentó un caso de estudio para los datos del viento en Valparaíso, en donde los resultados muestran que la distribución de Weibull se ajusta a la distribución de datos de promedios diarios de velocidad del viento. Esto quiere decir, que si estimamos la velocidad más probable, por ejemplo, esta se referirá al promedio más probable que se dé cierto día. Teniendo esto en cuenta, es posible obtener modelos para distintos intervalos de tiempo, teniendo en cuenta que es posible modificar la calidad del modelos obtenido, mediante el ajuste de los parámetros de la distribución de Weibull.



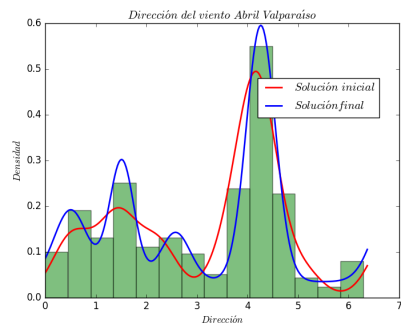
(a) Solución inicial y final, Enero.



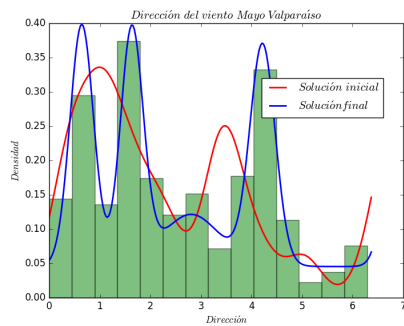
(b) Solución inicial y final, Febrero.



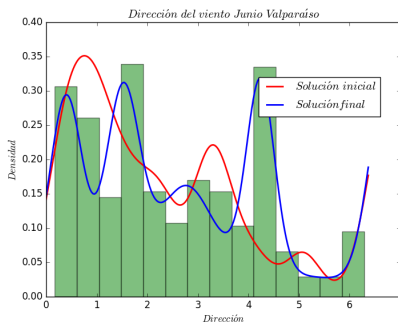
(c) Solución inicial y final, Marzo.



(d) Solución inicial y final, Abril.

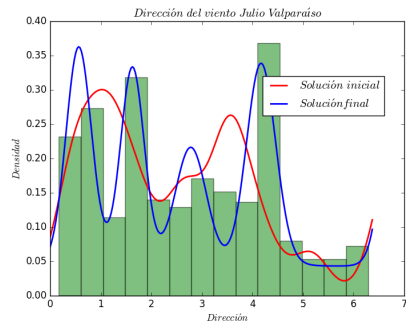


(e) Solución inicial y final, Mayo.

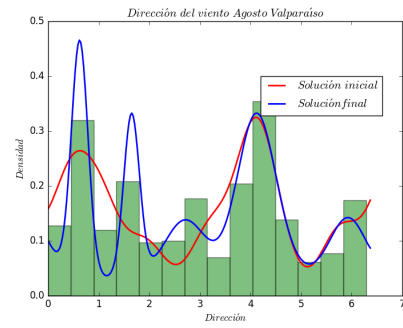


(f) Solución inicial y final, Junio.

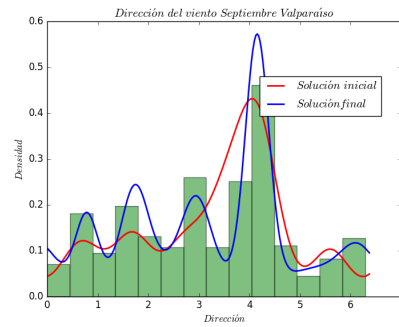
Figura 4.8: Gráficos de ajuste de MVM por meses. Fuente: Elaboración propia.



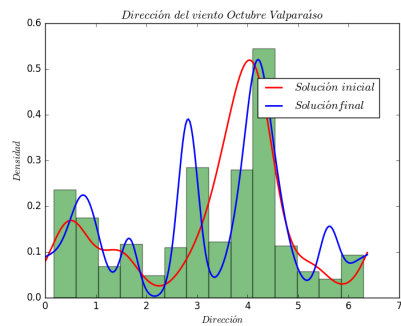
(a) Solución inicial y final, Julio.



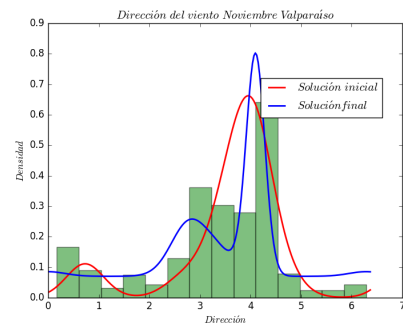
(b) Solución inicial y final, Agosto.



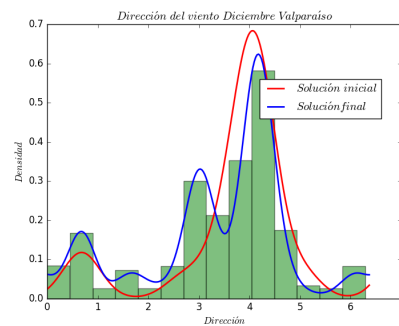
(c) Solución inicial y final, Septiembre.



(d) Solución inicial y final, Octubre.



(e) Solución inicial y final, Noviembre.



(f) Solución inicial y final, Diciembre.

Figura 4.9: Gráficos de ajuste de MVM por meses. Fuente: Elaboración propia.

Bibliografía

- [1] S. C. O., “Energías renovables en Chile. el potencial eólico, solar e hidroeléctrico de Arica a Chiloé,” tech. rep., MINENERGIA GIZ, 2014. Proyecto Estrategia de Expansión de las Energías Renovables en los Sistemas Eléctricos Interconectados.
- [2] A. Dabbaghiyan, F. Fazelpour, M. D. Abnavim, and M. a. Rosen, “Evaluation of wind energy potential in province of Bushehr, Iran,” 2015.
- [3] T. C. Carneiro, S. P. Melo, P. C. M. Carvalho, and A. P. d. S. Braga, “Particle swarm optimization method for estimation of Weibull parameters: A case study for the Brazilian northeast region,” 2015.
- [4] J. Heckenbergerova, P. Musilek, and P. Krömer, “Optimization of wind direction distribution parameters using particle swarm optimization,” 2015.
- [5] A. Kaveh, “Particle swarm optimization,” pp. 9–40, 2014. Chapter of Advances in Metaheuristic Algorithms for Optimal Design of Structures.
- [6] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of IEEE international conference on neural networks*, pp. 1942 – 1948, 1995.
- [7] C. Kongnam and S. Nuchprayoon, “A particle swarm optimization for wind energy control problem,” 2015.
- [8] Calimo, “Weibull probability density function.” URL https://commons.wikimedia.org/wiki/File:Weibull_PDF.svg, 2010. [Revisado el 19/06/2016].
- [9] T. P. Chang, “Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application,” 2010.

- [10] L. to the Editor, “Modeling the wind direction behaviors during the monsoon seasons in peninsular malaysia,” 2015.
- [11] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra, “Clustering on the unit hypersphere using von mises-fisher distributions,” *Journal of Machine Learning Research*, 2005.
- [12] J. A. Carta, C. Bueno, and P. Ramírez, “Statistical modelling of directional wind speeds using mixtures of von mises distribution: Case study,” 2007.
- [13] Y. U. Department of statistics, “Chi-square goodness of fit test.” URL <https://www.stat.yale.edu/Courses/1997-98/101/chigf.htm>, 1998. [Revisado el 31/10/2016].
- [14] E. C. o. S. Department of statistics, “Chi-square distribution table.” URL <http://sites.stat.psu.edu/~mga/401/tables/Chi-square-table.pdf>, 2006. [Revisado el 31/10/2016].