

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

Índice de figuras

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. [?]

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 [?].

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 [?], 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. [?] 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 [?], 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 [?], 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 [?] $x_{i,j}^k$ 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 $gbest$ 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 ??

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

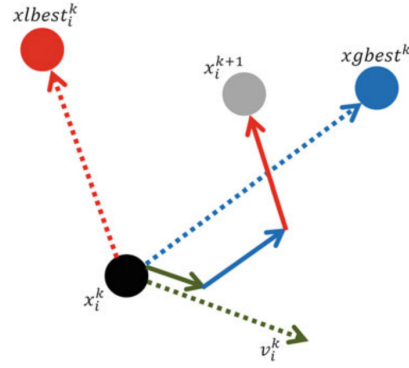


Figura 2.1: Movimiento de una partícula
Creado por Kaveh [?].

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 [?], 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. [?], Kongnam et al. [?] y Dabbaghiyan et al. [?]. En particular, en el trabajo realizado por Carneiro et al. [?], 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 ??, en donde se ven distintas curvas de Weibull, con diferentes parámetros k , y c constante.

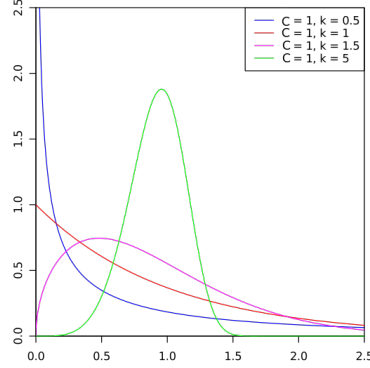


Figura 2.2: Función de distribución de probabilidad de Weibull
Adaptación propia desde [?].

2.2.2. Métodos numéricos

Tradicionalmente, se utilizan métodos numéricos para estimar los parámetros de la distribución de Weibull. En el artículo de Chang [?], 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. [?], 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 ?? En Kongnam et al. [?], el PSO es utilizado para el problema del control

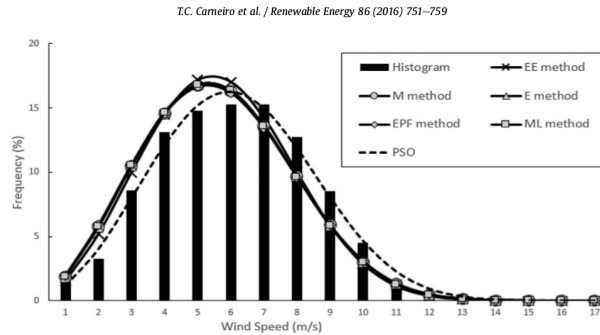


Figura 2.3: Distribución de Weibull con histograma - Maracanaú
Creado por Carneiro et al. [?].

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 [?], 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. [?].

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 ?? 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. [?], se utiliza una estrategia diferente a la

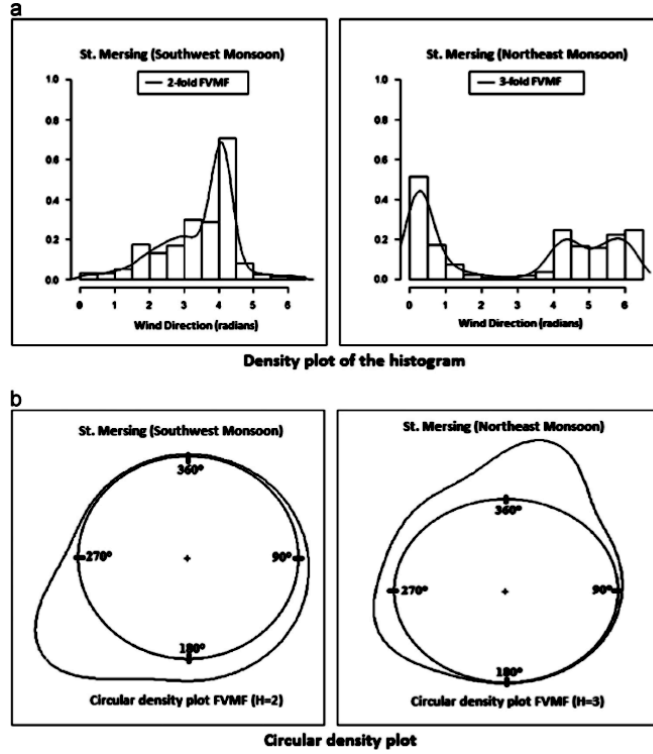


Figura 2.4: Modelo de ajuste FVMV para suroeste y noreste en la estación Mersing
Creado por [?].

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??.

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 ??.

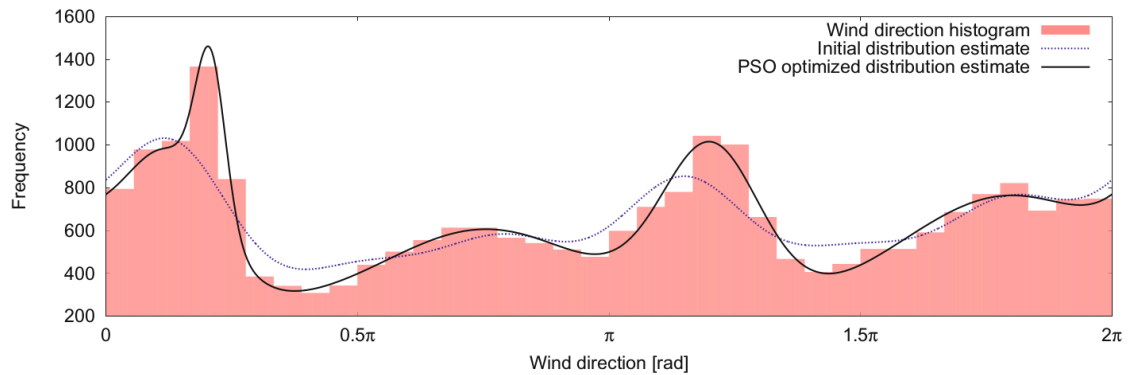


Figura 2.5: Ajuste dirección del viento, aeropuerto St. John
Creado por Heckenbergerova et al. [?].

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 ???. El PSO a utilizar está representado por la ecuación ???. La función objetivo se describe con la fórmula ?? 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 [?] 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. [?]. 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 ?? 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 iteraciones fueron probadas hasta un máximo de 1000 y 50 partículas,

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).
```

(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 ??.

3.1.4. 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 ???. El ajuste de la distribución de Weibull a

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

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. [?]):

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 ??, la cual consiste básicamente en una combinación lineal de la *simple von Mises distribution* descrita en ??.

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. [?] 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. [?], debido a la falta de una solución analítica a la ecuación ??, 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. [?] 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 ?? y la modificación a este sugerida en Carneiro et al. [?] descrita previamente en ??.

La función objetivo para el PSO es el test estadístico χ^2 descrito en [?] 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 ??. 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 ??.