

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

<b>1. Introducción</b>	<b>1</b>
1.1. Identificación del problema . . . . .	1
1.2. Objetivos . . . . .	3
1.2.1. Objetivos específicos . . . . .	3
<b>2. Estado del arte</b>	<b>4</b>
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 . . . . .	10
2.3. Dirección del viento . . . . .	11
<b>3. Desarrollo de la solución</b>	<b>15</b>
3.1. Velocidad del viento . . . . .	15
3.1.1. Modelo Matemático . . . . .	15
3.1.2. Representación . . . . .	16
3.1.3. Descripción del algoritmo . . . . .	17
3.1.4. Experimentos . . . . .	17
3.2. Dirección del viento . . . . .	19
3.2.1. Modelo Matemático . . . . .	19
3.2.2. Representación . . . . .	22
3.2.3. Descripción del algoritmo . . . . .	22
3.2.4. Experimentos . . . . .	24
3.3. Aplicación de los algoritmos propuestos . . . . .	25
<b>4. Análisis y conclusiones</b>	<b>28</b>
4.1. Resultados y conclusiones . . . . .	28
4.1.1. Análisis de los resultados - velocidad del viento . . . . .	28
4.1.2. Análisis de los resultados - dirección del viento . . . . .	33
4.1.3. Aplicaciones . . . . .	37
4.1.4. Conclusiones . . . . .	38
<b>Bibliography</b>	<b>45</b>

## 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 . . . . .	11
2.4. Modelo de ajuste FVMV para suroeste y noreste en la estación Mersing .	13
2.5. Ajuste dirección del viento, aeropuerto St. John . . . . .	14
3.1. Ejemplo colección de datos Enero Valparaíso 2015 . . . . .	18
3.2. Esquema de uso del algoritmo . . . . .	27
4.1. Superficie de datos viento de Valparaíso. Fuente: Elaboración Propia. . .	29
4.2. Ajuste con PSO a datos del viento de Valparaíso. Fuente: Elaboración Propia.	30
4.3. Ajuste con PSO a datos Valparaíso 2015, 2014 y 2013, baja y buena calidad. Fuente: Elaboración Propia. . . . .	30
4.4. Ajuste con PSO a registros del viento en Valparaíso (Con todos los datos). Fuente: Elaboración Propia. . . . .	31
4.5. Ajuste con PSO a datos Valparaíso 2015, por rango de meses. Fuente: Elaboración Propia. . . . .	32
4.6. Ajuste con PSO a datos (cifras puras) Valparaíso 2015, 2014, 2013 . . . .	32
4.7. Pruebas iniciales . . . . .	33
4.8. Graficos de ajustes anuales . . . . .	34
4.9. Gráficos de ajuste de MVM por meses. Fuente: Elaboración propia. . . .	39
4.10. Gráficos de ajuste de MVM por meses. Fuente: Elaboración propia. . . .	40
4.11. Gráficos de ajuste por meses en coordenadas polares. Fuente: Elaboración propia. . . . .	41
4.12. Gráficos de ajuste por meses en coordenadas polares. Fuente: Elaboración propia. . . . .	42
4.13. Comparación de variaciones en el PSO. Fuente: Elaboración propia. . . .	43
4.14. Comparación de variaciones en el PSO. Fuente: Elaboración propia. . . .	44

# Índice de tablas

4.1. Tabla de tests estadísticos . . . . .	33
4.2. Tabla de tests estadísticos . . . . .	35

# 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. Ejemplo de ello es el estudio realizado sobre el potencial eléctrico en la provincia de Bushehr, Iran por Dabbaghiyan et al. [2], o el trabajo de Chang en Taiwan [3].

En primera instancia, es necesario 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 estas mediciones, de manera de poder obtener información relevante respecto al comportamiento del viento en función del estudio a realizar.

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* (en adelante también referido como PSO), con la cual se ha logrado mejorar la calidad del modelo para la velocidad del viento, como se expone en el caso de estudio en el noreste de Brasil por Carneiro et. al. [4], 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. [5] 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 [6], 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 en la interacción en grupo de seres vivos, por ejemplo, las gaviotas, las cuales suelen moverse en bandadas, vuelan en conjunto cerca del mar en búsqueda de zonas donde hayan alimento (peces), moviéndose individualmente, pero siendo influenciadas o guiadas por otras. El método simula la conducta de los individuos a través de partículas que se mueven dentro de un espacio (rango de posibles valores conocido como espacio de búsqueda), siendo estas afectadas por factores individuales (conocimiento propio del lugar donde estoy) y factores colectivos (conocimiento del enjambre del mejor lugar encontrado), los cuales dirigen el movimiento de estos grupos a zonas las cuales son escogidas o 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 las posiciones obtengan mejores puntajes al ser evaluadas por la función objetivo.

El modelo clásico presentado por Kennedy y Eberhart [7], describe la variación de la



velocidad y de la posición de las partículas como se presenta a continuación:

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

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

Como se explica en Kaveh [6]  $x_{i,j}^k$  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,  $x_{best_i}$  y  $x_{gbest}$  representan las mejores soluciones alcanzadas por la partícula  $i$  y por el enjambre en conjunto 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 deficiencias, en particular la forma en que se

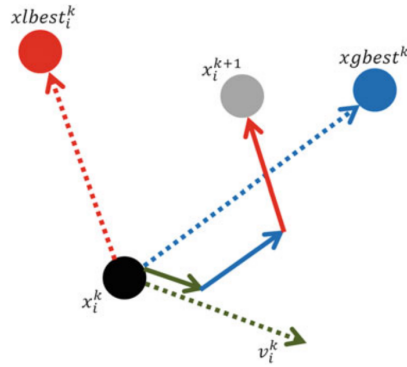


Figura 2.1: Movimiento de una partícula  
Creado por Kaveh [6].

actualiza la velocidad no considera una ponderación para el aporte de la velocidad previa en una partícula. Para corregir esto, simplemente se añade un factor que escala la velocidad previa. Como se explica en Kaveh [6], este peso que se añade a la velocidad previa permite tener un control más apropiado respecto al comportamiento del enjambre. En términos extremos, si la velocidad previa se elimina (peso nulo), las partículas quedan atrapadas en una región local, pero si se le da demasiado peso, converge rápidamente a un óptimo local. Es por esto que la forma del PSO actual tiene un parámetro  $w$  que representa la incidencia de la velocidad previa en la velocidad actual de la partícula (parámetro llamado factor de

inercia). Nuevamente, 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, tal como podría esperarse de una partícula en aceleración.

A modo de complemento, en el trabajo inicialmente citado, se puede observar una revisión completa del estado del arte del método *Particle Swarm Optimization* en términos de diseño o arquitectura del algoritmo, donde se exponen las distintas modificaciones y alternativas existentes en 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 ajustar un modelo probabilístico (distribución de densidad de probabilidad) que explique el comportamiento de las velocidades del viento medido. 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. [4], Kongnam et al. [8], Dabbaghiyan et al. [2], Fadare [9], Weisser [10] y Chang [3]. En el trabajo realizado por Carneiro et al. [4], 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 (o amplitud) de la distribución respectivamente, y  $v$  es el valor de la velocidad del viento a la que el modelo asociará una determinada densidad. Un ejemplo de como varia la forma de esta distribución se aprecia en la figura 2.2, en donde se ven distintas curvas de Weibull, con diferentes parámetros  $k$  manteniendo  $c$  constante.

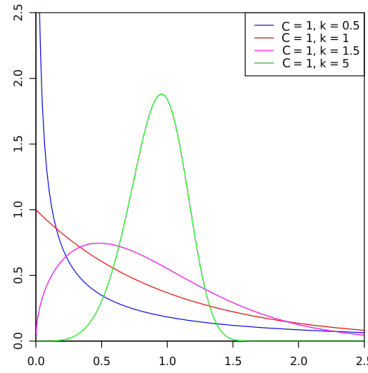


Figura 2.2: Función de distribución de probabilidad de Weibull  
Adaptación propia desde [11].

### 2.2.2. Métodos numéricos

Tradicionalmente, se utilizan métodos numéricos para estimar los parámetros de la distribución de Weibull. En el artículo de Chang [12], 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  $\sigma$  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:

$$k = \left(\frac{\sigma}{\bar{v}}\right)^{-1,086} \quad (2.8)$$

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

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 método, 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:

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

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, usando una simulación basada en el método de Montecarlo. El análisis de los datos del viento fue desarrollado bajo 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 al momento de ajustar la distribución de Weibull a los datos de prueba.

En la búsqueda de nuevas alternativas surge una propuesta para mejorar el ajuste de la función de distribución de probabilidad a los datos, a través del uso de la meta-heurística *Particle Swarm Optimization*, propuesta que mejora la calidad de los resultados en comparación con los métodos numéricos presentados.

### 2.2.3. Particle Swarm Optimization

En Carneiro et al. [4], 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. Tal y como se adelanta anteriormente, el modelo utilizado para este propósito es la distribución de Weibull. Para poder utilizar dicha distribución, es necesario ajustar los parámetros del modelo a los datos recolectados.

En el estudio mencionado, se propone un *particle swarm optimization* para encontrar los parámetros  $k$  y  $c$  de la distribución de Weibull y a su vez corroborar que la calidad de la solución encontrada por el PSO, comparada con los métodos numéricos tradicionales, es de mejor calidad.

Así, se define la función de aptitud para el PSO como:

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

Donde  $\epsilon$ , es el error cuadrático a minimizar entre los valores del histograma de los datos recolectados 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 controlar la convergencia del método, se establece que estos parámetros varíen durante la ejecución del algoritmo dentro de un rango definido ( $w \in \{0,4, 0,9\}$ ,  $c1$  y  $c2 \in \{0, 2,5\}$ ), aumentando los factores de inercia y cognitivo, y disminuyendo el factor social a medida que se avanza en las iteraciones. Esto se traduce en que se explotan las soluciones iniciales, para posteriormente en iteraciones finales el algoritmo explore en una vecindad local al óptimo encontrado hasta el momento.

Finalmente, para evaluar los resultados de la propuesta, se compara el PSO con cinco de los seis métodos numéricos mencionados en la sección 2.2.2 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 medir la calidad de cada método, 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 demuestran que mediante la meta-heurística *particle swarm optimization* se pueden obtener los mejores parámetros de ajuste para la distribución de Weibull. Un ejemplo de esto es expuesto en la figura 2.3. En Kon-

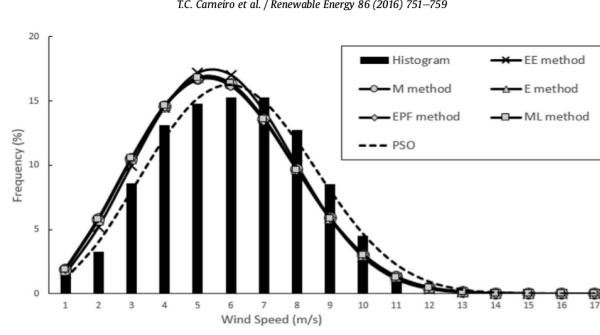


Fig. 5. Weibull distribution curves and histogram - Maracanaú (February 2012–January 2013).

Figura 2.3: Distribución de Weibull con histograma - Maracanaú  
Creado por Carneiro et al. [4].

gnam et al. [8], el PSO es utilizado para el problema del control 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

En el trabajo acerca del modelado del comportamiento de la dirección del viento en Malasia [13], se propone que para identificar la(s) dirección(es) dominante del viento se utilice la función de densidad de probabilidad *finite von Mises-Fisher* (FVMF) ajustada a las mediciones obtenidas. Estos datos acerca del viento fueron obtenidos desde cinco estaciones meteorológicas ubicadas en distintas zonas en la península de Malasia.

La FVMF, o la función de densidad de probabilidad de von Mises como también se le llamará más adelante, de forma genérica, está definida de la siguiente forma:

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

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),  $\mu_h$  es el parámetro de dirección predominante (análogo a la media  $\mu$  en la distribución normal),  $k_h$  es el parámetro de concentración (análogo al recíproco de la dispersión  $\sigma^2$ ), 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 FVFM, 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. [14].

Por último, los resultados de este trabajo muestran que FVFM 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. [5], se utiliza una estrategia diferente a la 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), o simplemente *mixture of von Mises distribution*, 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



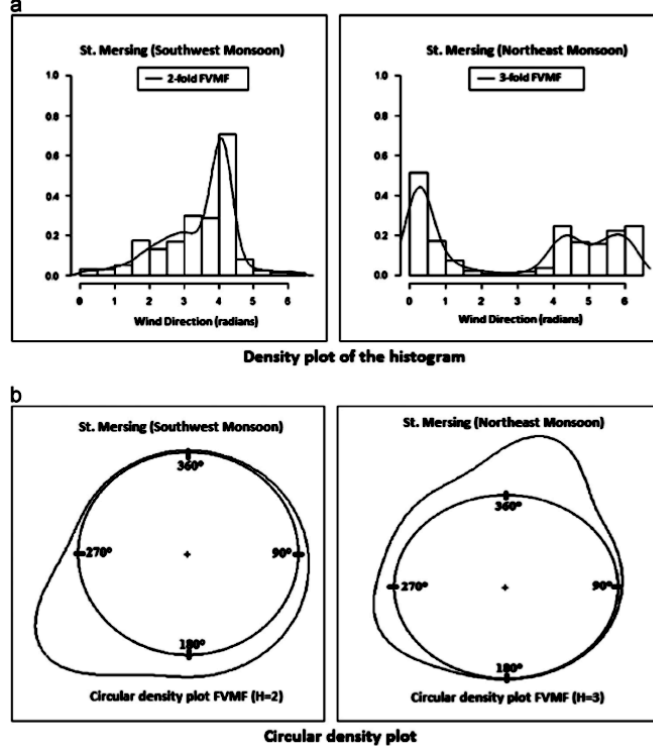


Figura 2.4: Modelo de ajuste FVMV para suroeste y noreste en la estación Mersing  
Creado por [13].

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  $\mu_j$  y  $k_j$ ,  $\theta$  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. [5]. Luego optimiza estos parámetros mediante el uso de la meta-heurística *particle swarm optimization*, con parámetros fijos, 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, comparando estos resultados con otra propuesta similar la cual se expone en un trabajo previo de los mismos autores [15], en donde se utiliza la misma estrategia pero reemplazando el PSO con algoritmos genéticos. Sin embargo, estos resultados no consiguen pasar el test estadístico impuesto por ellos mismos, 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.

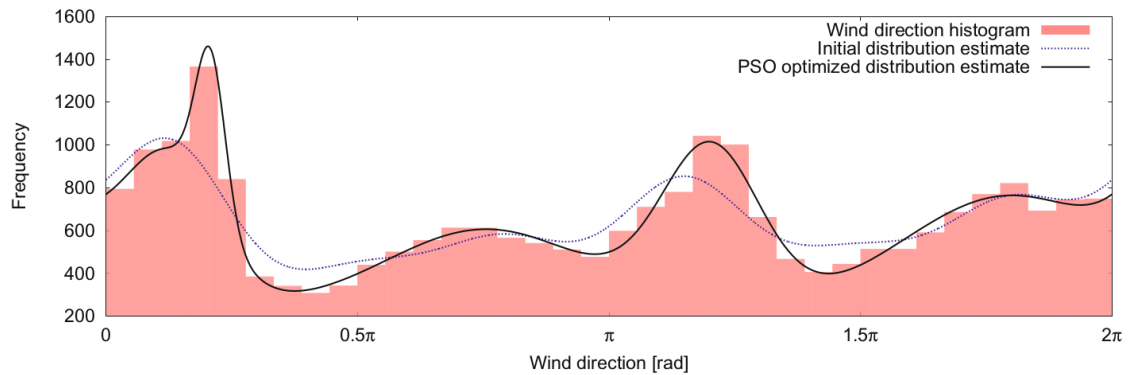


Figura 2.5: Ajuste dirección del viento, aeropuerto St. John  
Creado por Heckenbergerova et al. [5].

## 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  $\geq 0$ . 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 [4] 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  $\alpha, \beta, \gamma$  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. [4]. 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

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

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

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. [4]):

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  $\mu_j$ ,  $k_j$  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. [16] y se describe a continuación.

Para la construcción del histograma se divide el rango de datos que va de 0 a  $2\pi$  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  $\mu_j$  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. [14], 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. [5] 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. [4] descrita previamente en 3.1.

La función objetivo para el PSO es el test estadístico  $\chi^2$  descrito en [5] 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  $\mu$  está representado en el rango  $i \in \{1, \dots, k\}$  y para ser decodificado debe ser escalado por  $2\pi$ .

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. [5].

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  $\mu_j$ , 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  $\mu_j$ ,  $k_j$  y  $w_j$ , para cada  $j \in \{1, \dots, k\}$ .

```
sol = inicializarVectorSolución(3*K)
for  $j = 0$  to  $K$  do
  datos $j$  = datosEnRango( $j * D/K$ )
  s $j$  = obtenerSenoPromedio(datos $j$ )
  c $j$  = obtenerCosenoPromedio(datos $j$ )
  u $j$  = obtenerDirecciónPredominante(s $j$ , c $j$ )
  k $j$  = obtenerConcentración(s $j$ , c $j$ )
  w $j$  = obtenerPeso( $j * (T/K)$ ,  $(j + 1) * (T/K)$ )
  addToSolution(sol, u $j$ , k $j$ , w $j$ )
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. [5]. 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.

### 3.2.4. Experimentos

Similar a lo descrito en , los datos de dirección del viento son tratados para rescatar las mediciones pertinentes al trabajo aquí expuesto. Estos se encuentran inicialmente en un formato como el que se puede apreciar en la figura 3.1.

Nuevamente, 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.

Para evaluar la calidad de la solución, se utiliza el test *Chi square goodness fit* [17], con lo cual se evalúa que tan bien representa el modelo propuesto a los datos medidos. Para ello, la hipótesis nula  $H_0$  es que los datos de dirección del viento se distribuyen según la función *mixture of von Mises distribution* y la hipótesis alternativa  $H_1$  niega dicha afirmación. Se rechaza  $H_0$  si el valor de la función objetivo del PSO para la solución final encontrada

excede el valor crítico de  $\Xi^2$  para un nivel de significancia de  $\alpha = 0,05$  y 13 grados de libertad, es decir **22.362**, el cual se puede encontrar en la tabla de la distribución  $\Xi^2$  [18]. Los grados de libertad son definidos a partir de la cantidad de clases de frecuencia definidas para el estudio, en este caso, se dividió el rango de valores de  $[0, 2\pi]$  en 14 tramos iguales, por lo quedan  $(n - 1)$  grados de libertad, 13 en este caso.

Los experimentos consistieron en el ajuste de varios subconjunto de datos provenientes de las mediciones obtenidas para la dirección del viento en los años 2013, 2014 y 2015. Así, se prueba la utilidad de la propuesta realizada independiente del rango de tiempo a modelar. Es importante considerar que no se incluyen los días en los que no hubo viento por la evidente imposibilidad de registrar la dirección. Los subconjuntos definidos fueron los siguientes:

1. **Anual:** Se consideran los datos de todo el año elegido.
2. **Meses acumulados:** Se considera una agrupación mensual pero reuniendo los datos de tres años consecutivos (2013, 2014, 2015). Es decir, para el mes de Enero, se ajusta el modelo a los datos de Enero-2013, Enero-2014 y Enero-2015 en conjunto.
3. **Meses:** Se escogen algunos meses para ser comparados consigo mismos durante los tres años escogidos. Por ejemplo, Enero-2013, Enero-2014 y Enero-2015 por separado.

Para el funcionamiento del PSO, se estableció un límite de 50500 iteraciones, se utilizaron 100 partículas, y se usó un límite de parada si es que el valor en la función objetivo de la mejor solución encontrada al momento era menor a 22.362 (criterio basado en la estrategia de cumplir el test *Chi square goodness fit*).

### 3.3. Aplicación de los algoritmos propuestos

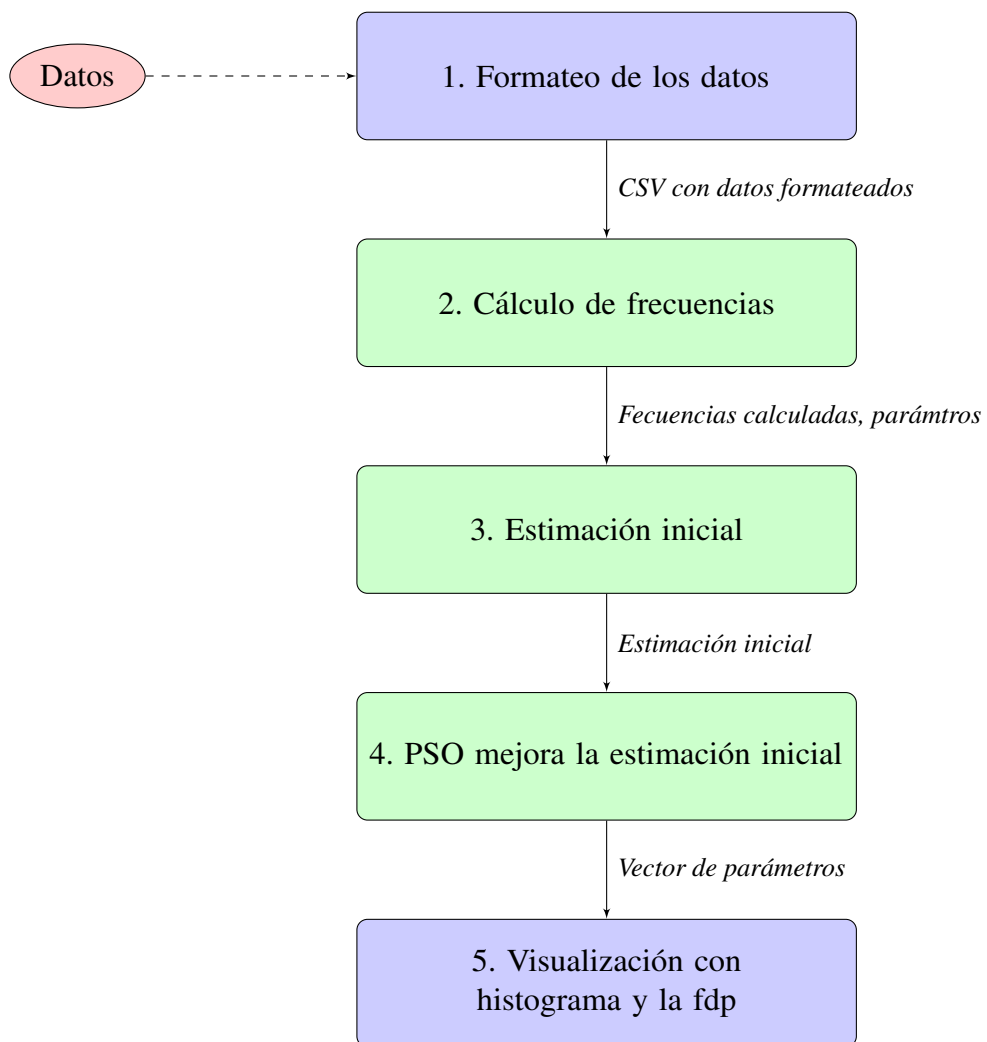
El esquema 3.2 resume el funcionamiento del algoritmo para el ajuste del modelo probabilístico a los datos del viento. A continuación se detallarán los pasos a seguir en el proceso de obtener el ajuste.

1. Lo primero que se realiza es el formateo de los datos, es decir, la conversión desde un set de datos externos a el formato utilizado por el programa desarrollado, CSV. El *script* utilizado dependerá del formato externo en el cual se proveen los datos.

2. Una vez obtenido los datos en el formato deseado, se procede a calcular las frecuencias de los registros obtenidos, para posteriormente ser comparadas con el modelo teórico.
3. Antes de comenzar con el ajuste, se calcula una solución inicial ya sea de forma aleatoria (como es el caso para el ajuste de la velocidad del viento) o mediante algún método conveniente (como la aproximación numérica para el caso del ajuste de la dirección del viento). La elección de esta última implica que se quiere guiar la solución a encontrar.
4. Definiendo la cantidad de partículas y el número de iteraciones, se ejecuta el algoritmo *Particle Swarm Optimization* para que mejore la estimación inicial de acuerdo a una función objetivo previamente definida.
5. Finalmente, una vez que el algoritmo termina, entrega los valores de los parámetros para la función de densidad de probabilidad (fdp). Con ello, es posible elaborar un histograma y graficar la fdp para evaluar cualitativamente el ajuste obtenido.

Una vez finalizado el proceso, se tiene un modelo matemático con el cual es posible trabajar para, por ejemplo, alguna de las aplicaciones que se exponen en la siguiente sección de análisis.

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. [4], 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.



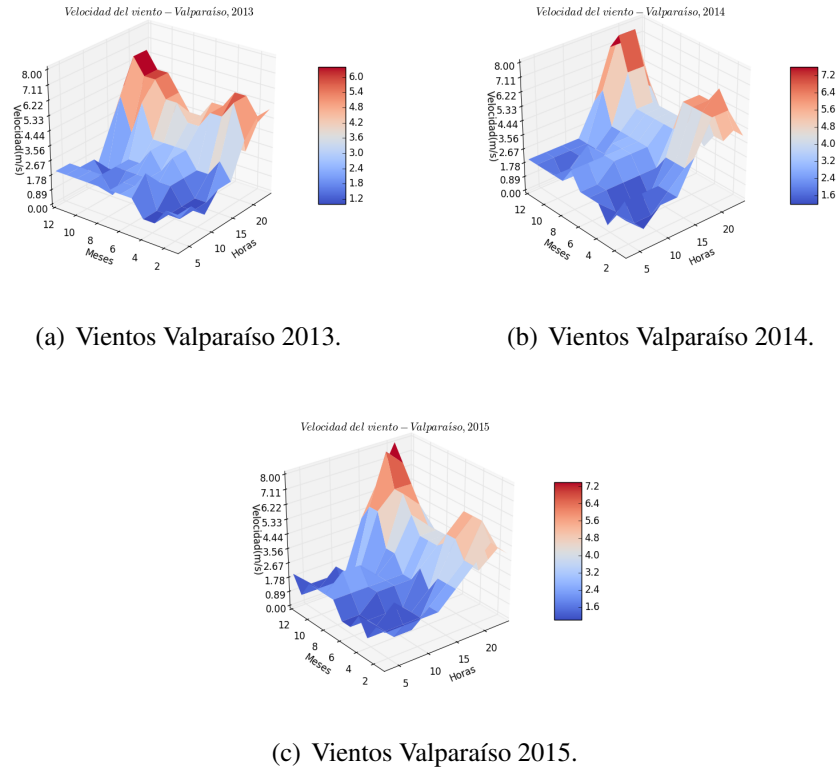


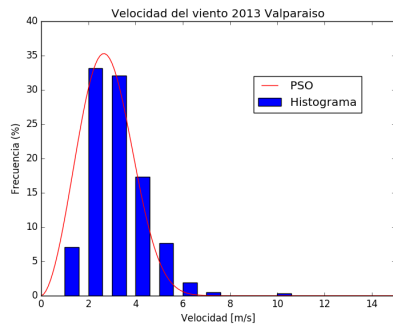
Figura 4.1: Superficie de datos viento de Valparaíso. Fuente: Elaboración Propia.

## Experimento 2, datos de tres años y promedios diarios

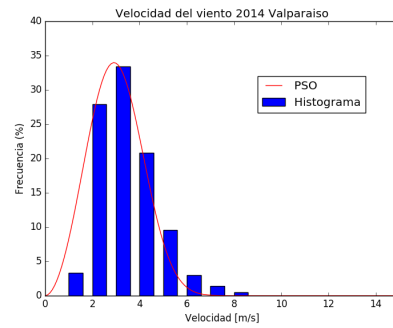
En este experimento se realizó el ajuste considerando los promedios diarios y un intervalo de tres años consecutivos. El gráfico 4.3(b), muestra el resultado del ajuste con PSO y la configuración estándar de los demás experimentos, es decir, 100 iteraciones y 50 partículas. En este gráfico se aprecia 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

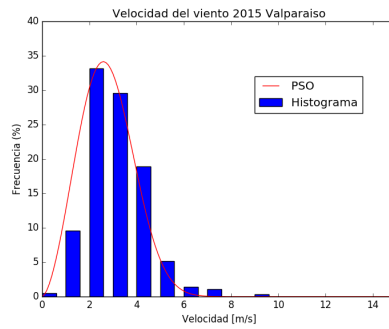
Los gráficos 4.4(a), 4.4(b) y 4.6 son ajustes de Weibull con los parámetros obtenidos en el experimento anterior. Es decir, la idea es evaluar el modelo general de los tres años



(a) PSO Valparaíso 2013.

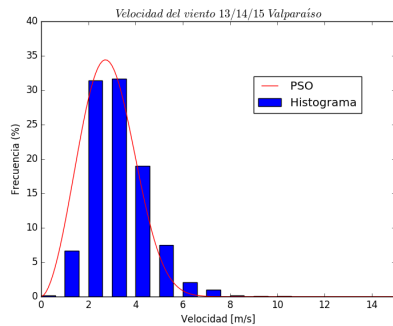


(b) PSO Valparaíso 2014.

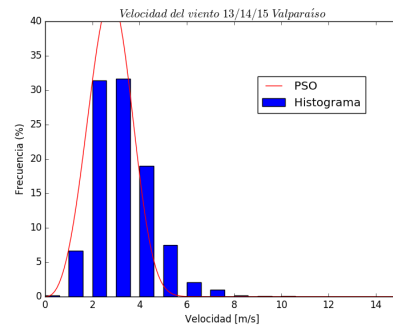


(c) PSO Valparaíso 2015.

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



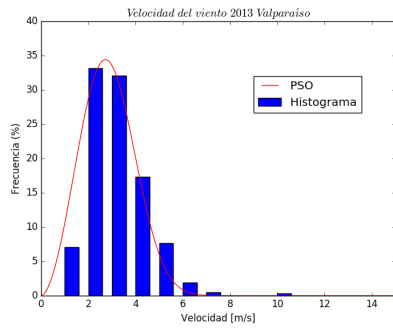
(a) Buen ajuste PSO Valparaíso 2013.



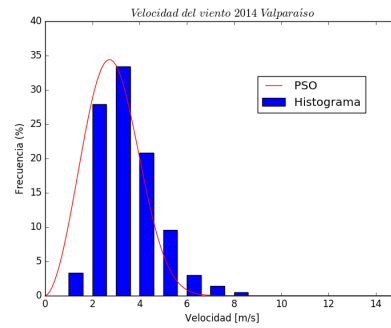
(b) Mal ajuste PSO Valparaíso 2013.

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

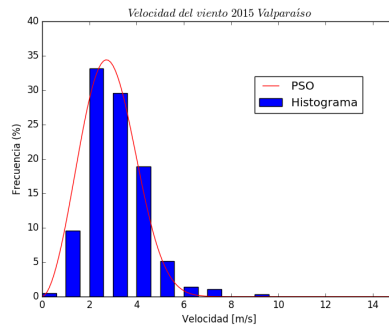
versus el histograma de datos de cada año en particular. El ajuste desde los resultados estadísticos (tabla 4.1), es levemente menos preciso que el modelo ajustado a cada año en particular, pero sigue siendo aceptable como posible opción a considerar.



(a) Velocidad viento Valparaíso 2013.



(b) Velocidad viento Valparaíso 2014.



(c) Velocidad viento Valparaíso 2015.

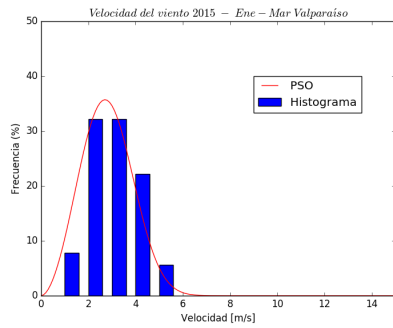
Figura 4.4: Ajuste con PSO a registros del viento en Valparaíso (Con todos los datos).  
Fuente: Elaboración Propia.

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

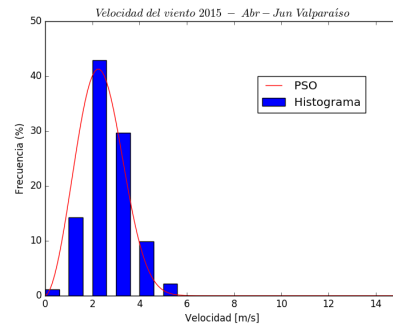
Es posible que se requiera un análisis más acotado, por ello los gráficos 4.5(a), 4.5(b), 4.5(c), 4.5(d), muestran un ajuste considerando un lapso de 3 meses para el año 2015, con el que se demuestra que es posible definir cualquier intervalo (manteniendo como unidad de dato el promedio diario de velocidad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

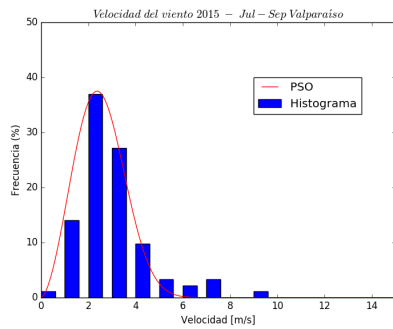
La razón de por qué se utiliza el promedio diario de los datos del viento para ajustar Weibull y no las mediciones puras (las mediciones tomadas cada 3 horas diariamente) es expuesta en el gráfico 4.6. La distribución de Weibull no se ajusta a una distribución de datos con más de un máximo, por lo que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



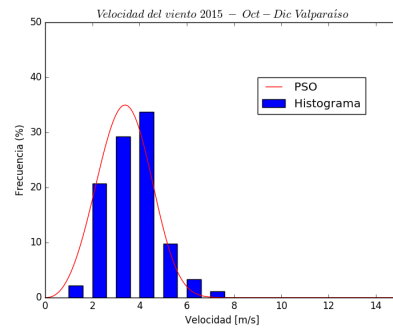
(a) Enero - Marzo.



(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.

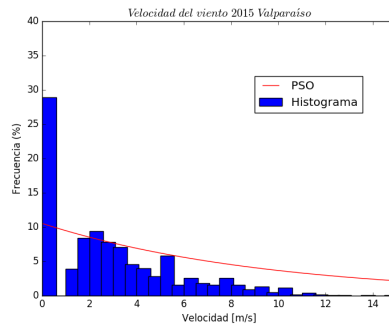


Figura 4.6: Ajuste con PSO a datos (cifras puras) Valparaíso 2015, 2014, 2013  
Fuente: Elaboración Propia.

## 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

#### Experimento 1, pruebas iniciales

En la figura 4.7(a) se puede observar un ajuste bastante distorsionado respecto al histograma de datos. Esto se debe a que la función objetivo del PSO evalúa la diferencia entre las frecuencias experimentales y las obtenidas teóricamente, por lo tanto, diferentes curvas pueden tener igual magnitud del área bajo la curva y por ende, la misma probabilidad con la que se obtiene la frecuencia teórica.

Por esto, es importante que el PSO busque mejorar la solución inicial en una vecindad cercana a esta, de manera de obtener una evolución como la que se aprecia en la figura ??.

Por ello, el algoritmo inicializa las partículas en la posición de la solución inicial encontrada más o menos una pequeña perturbación.

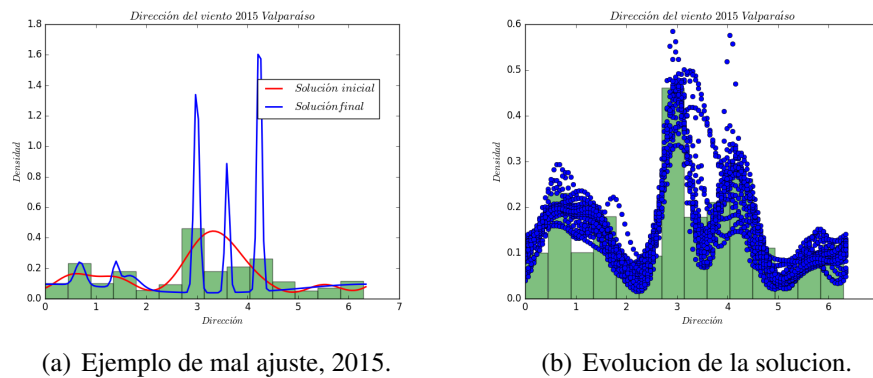


Figura 4.7: Pruebas iniciales

## Experimento 2, Ajustes anuales

En este experimento se ajustan los vientos anualmente, en los años de los cuales se obtuvieron datos para este estudio, resultado la figura 4.8(c) para el 2013, la figura 4.8(b) para el 2014 y la figura 4.8(a) para el 2015. Tal como se realiza en otros estudios [5] [13], el ajuste de los datos de dirección del viento en un formato de largo plazo, permite observar el comportamiento global de los vientos pudiendo evaluar la norma o generalidad en los datos registrados. En las figuras referenciadas previamente se observa la evolución de la solución desde la aproximación inicial hasta la mejorada por el PSO.

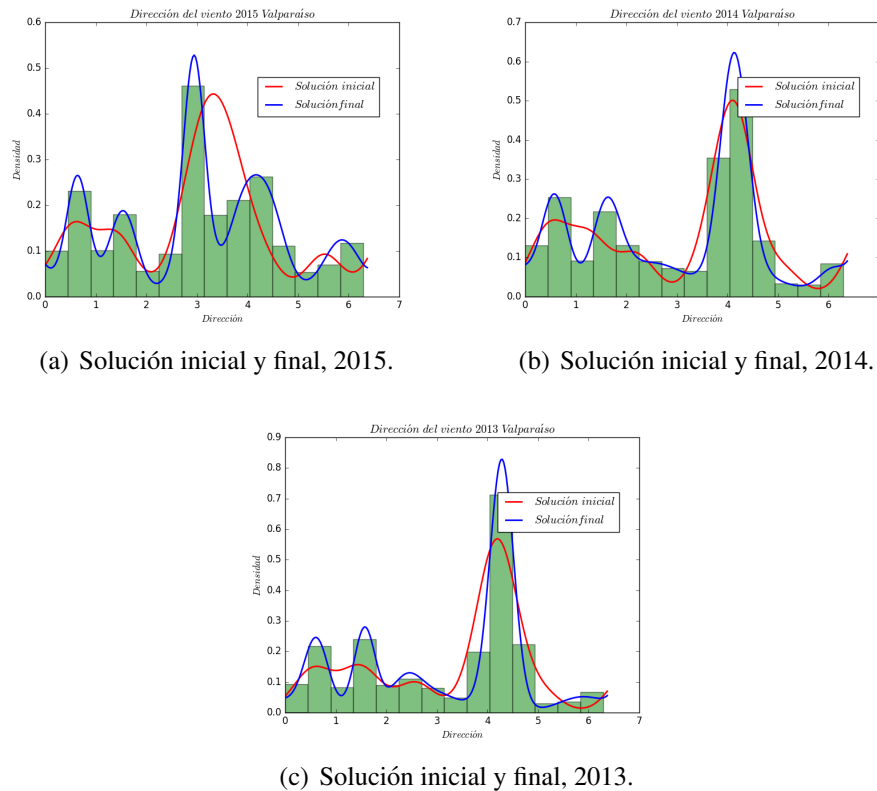


Figura 4.8: Graficos de ajustes anuales

## Experimento 3, Ajustes de meses acumulados

Similar al ejercicio anterior, los gráficos agrupados en 4.9 y 4.10, permiten visualizar el ajuste de los datos del viento por meses. En este caso se agruparon los datos de los tres años escogidos por cada mes, es decir, la figura 4.9(a), por ejemplo, contiene los datos del mes de enero de los años 2013, 2014 y 2015.

Se puede observar que algunos gráficos muestran un mejor ajuste que otros, a pesar del correcto valor obtenido en la función objetivo como se expresa posteriormente en la tabla ???. Esto se debe a que probablemente la solución se escapa más de los deseado de la solución inicial, asunto que será retomado más adelante.

#### Experimento 4, Ajuste por meses, visualización en coordenadas polares

En las figuras agrupadas 4.12 y ?? se aprecia una visualización más interesante. Allí se puede ver de forma más intuitiva las direcciones dominantes del viento, relativas al sistema de referencia utilizado en meteorología conocido comúnmente como la rosa de los vientos [19].

#### Resumen de los experimentos

Tabla 4.2: Tabla de tests estadísticos

Periodo de tiempo	PSI	PSF1	PSF2	Tiempo 1	Tiempo 2	Iteraciones 1	Iteraciones 2	Cantidad datos
2015	461.834005	251.376863	-	38m46.168s	-	50500	-	2076
2015	461.834005	75.165259	21.580166	32m47.136s	8m5.748s	50500	9005	2076
2014	301.12834005	41.314831	19.869808	33m23.256s	20m37.844s	50500	29494	2499
2013	424.247097	48.983099	20.76838	30m23.724s	11m41.128s	50500	16778	2346
Enero	144.420285	8.273903	20.890711	32m37.239s	0m3.332s	50500	86	633
Febrero	130.397909	28.736218	19.450824	36m5.152s	0m12.260s	50500	288	567
Marzo	104.544257	6.042533	18.610565	32m37.384s	0m3.124s	50500	49	564
Abril	74.974500	7.511217	21.135048	32m58.456s	0m12.564s	50500	192	559
Mayo	243.656241	16.017976	21.650182	31m59.612s	1m6.924s	50500	1237	589
Junio	286.308891	48.086027	21.018451	33m52.584s	1m42.996s	50500	1989	554
Julio	177.338315	20.136877	22.064056	31m48.368s	0m17.888s	50500	315	604
Agosto	79.579614	21.30693	21.940013	34m26.344s	0m12.940s	50500	202	578
Septiembre	92.775196	12.381647	21.313293	31m37.732s	0m1.776s	50500	27	541
Octubre	188.149963	34.749368	21.810323	29m56.496s	0m45.356s	50500	825	564
Noviembre	363.902142	96.347418	26.25178	32m53.668s	42m21.400s	50500	50500	582
Diciembre	335.286306	16.101387	22.300924	34m52.916s	9m56.148s	50500	11640	586
Enero 2015	42.941029	2.762456	16.340722	36m18.196s	0m4.108s	50500	61	220
Enero 2014	68.102447	26.631838	20.683871	32m4.328s	0m1.872s	50500	29	206
Enero 2013	194.988635	10.285949	19.319721	35m29.220s	0m0.448s	50500	6	207
Mayo 2015	60.843802	7.525581	17.998473	34m53.804s	0m2.300s	50500	39	173
Mayo 2014	142.856664	37.046558	20.943375	29m33.576s	0m4.832s	50500	82	213
Mayo 2013	135.687704	20.536476	20.8897	30m50.044s	0m5.312s	50500	97	203
Septiembre 2015	77.411834	24.247726	21.293589	29m27.732s	0m8.420s	50500	137	140
Septiembre 2014	19.563136	1.99932	18.247167	33m42.336s	0m0.084s	50500	1	200
Septiembre 2013	68.904049	20.665142	16.617326	46m6.980s	0m0.388s	50500	5	201

La tabla 4.2 muestra un resumen de los resultados obtenidos al ajustar la distribución de probabilidad *von Mises distribution* en diferentes subconjuntos de tiempo de los datos colectados en los años 2013, 2014 y 2015. La tabla está organizada como sigue a continuación.

1. *Periodo de tiempo*: Rango de tiempo donde se consideraron los datos para el ajuste.

2. *PSI*: Puntaje solución inicial, es decir, el valor obtenido al evaluar la solución inicial en la función objetivo.
3. *PSF1*: Puntaje solución final 1, se refiere a la solución obtenida por el PSO al utilizar la estrategia explicada en ??.
4. *PSF2*: Puntaje solución final 2, se refiere a la solución obtenida por el PSO al utilizar la estrategia explicada en ??.
5. *Tiempo 1*: Tiempo empleado por el PSO con la estrategia 1 ??.
6. *Tiempo 2*: Tiempo empleado por el PSO con la estrategia 2 ??.
7. *Iteraciones 1*: Iteraciones empleadas por el PSO con la estrategia 1 ??.
8. *Iteraciones 2*: Iteraciones empleadas por el PSO con la estrategia 2 ??.
9. *Cantidad datos*: Cantidad de datos utilizados en el ajuste.

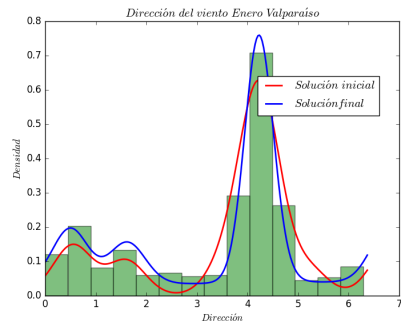


### **4.1.3. Aplicaciones**

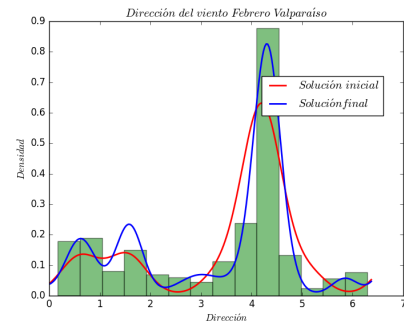
#### 4.1.4. Conclusiones

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.

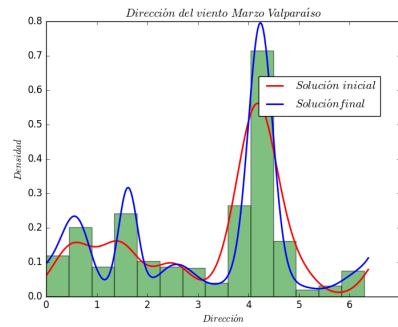
En este trabajo se presentó un caso de estudio para los datos del viento en Valparaíso, en donde los resultados muestran que la distribución de Weibull se ajusta a la distribución de datos de promedios diarios de velocidad del viento. Esto quiere decir, que si estimamos la velocidad más probable, por 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.



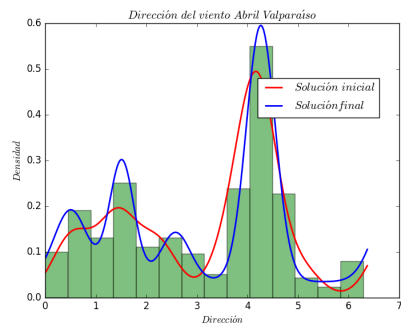
(a) Solución inicial y final, Enero.



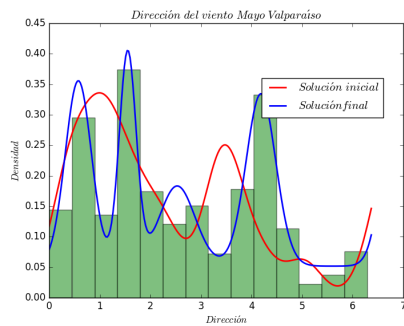
(b) Solución inicial y final, Febrero.



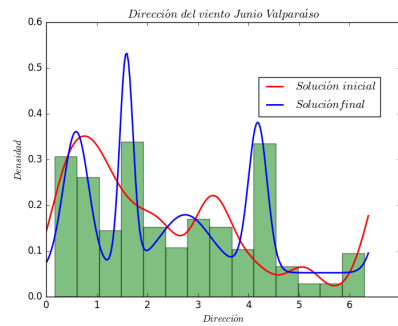
(c) Solución inicial y final, Marzo.



(d) Solución inicial y final, Abril.

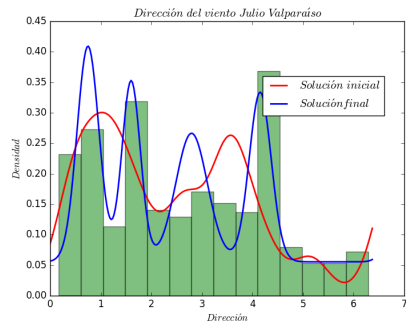


(e) Solución inicial y final, Mayo.

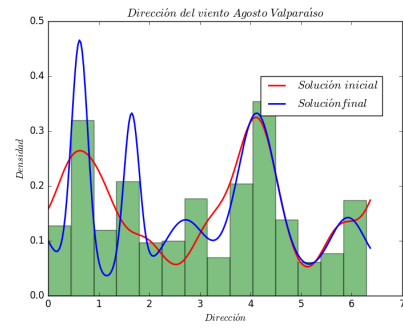


(f) Solución inicial y final, Junio.

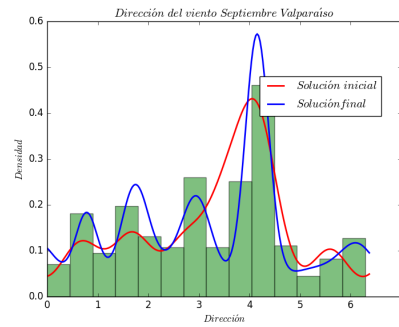
Figura 4.9: Gráficos de ajuste de MVM por meses. Fuente: Elaboración propia.



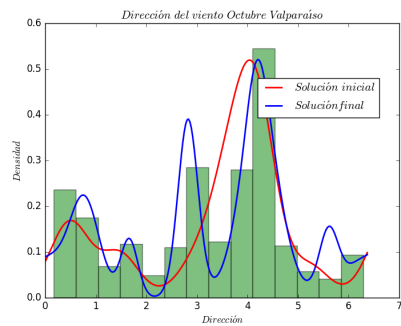
(a) Solución inicial y final, Julio.



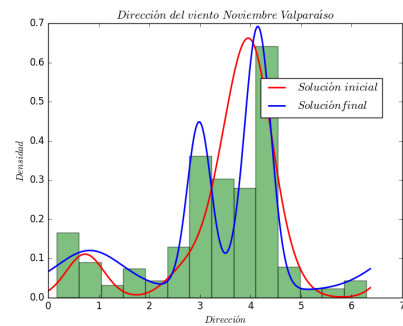
(b) Solución inicial y final, Agosto.



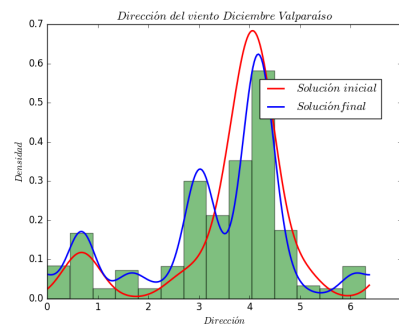
(c) Solución inicial y final, Septiembre.



(d) Solución inicial y final, Octubre.

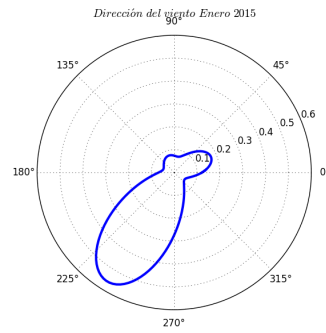


(e) Solución inicial y final, Noviembre.

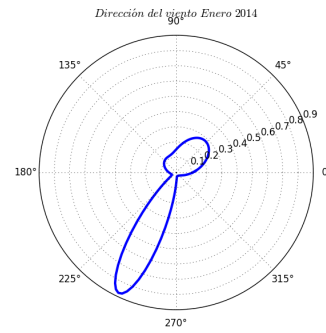


(f) Solución inicial y final, Diciembre.

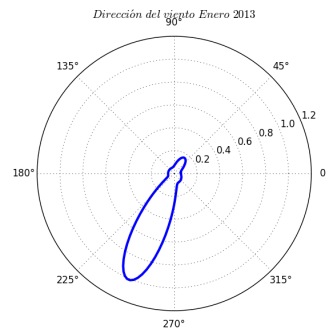
Figura 4.10: Gráficos de ajuste de MVM por meses. Fuente: Elaboración propia.



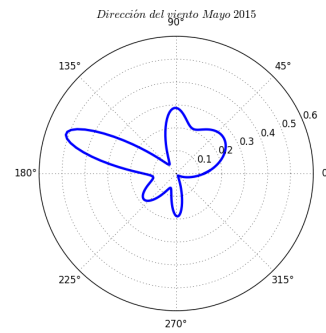
(a) Dirección enero 2015.



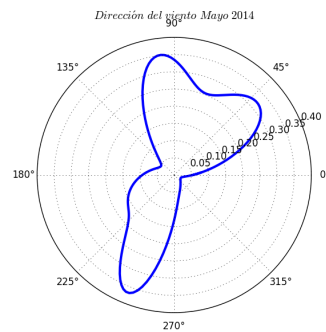
(b) Dirección enero 2014.



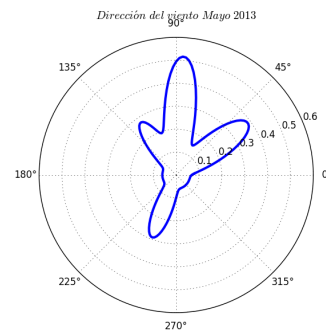
(c) Dirección enero 2013.



(d) Dirección mayo 2015.

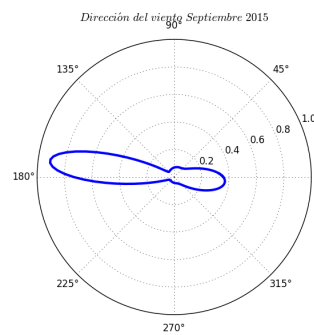


(e) Dirección mayo 2014.

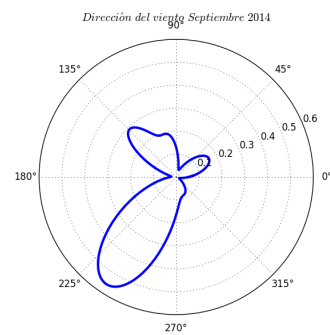


(f) Dirección mayo 2013.

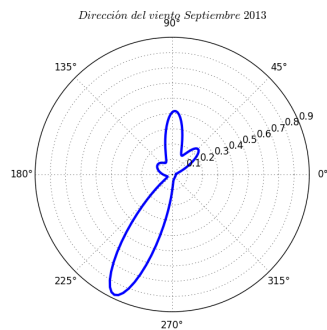
Figura 4.11: Gráficos de ajuste por meses en coordenadas polares. Fuente: Elaboración propia.



(a) Dirección septiembre 2015.



(b) Dirección septiembre 2014.



(c) Dirección septiembre 2013.

Figura 4.12: Gráficos de ajuste por meses en coordenadas polares. Fuente: Elaboración propia.

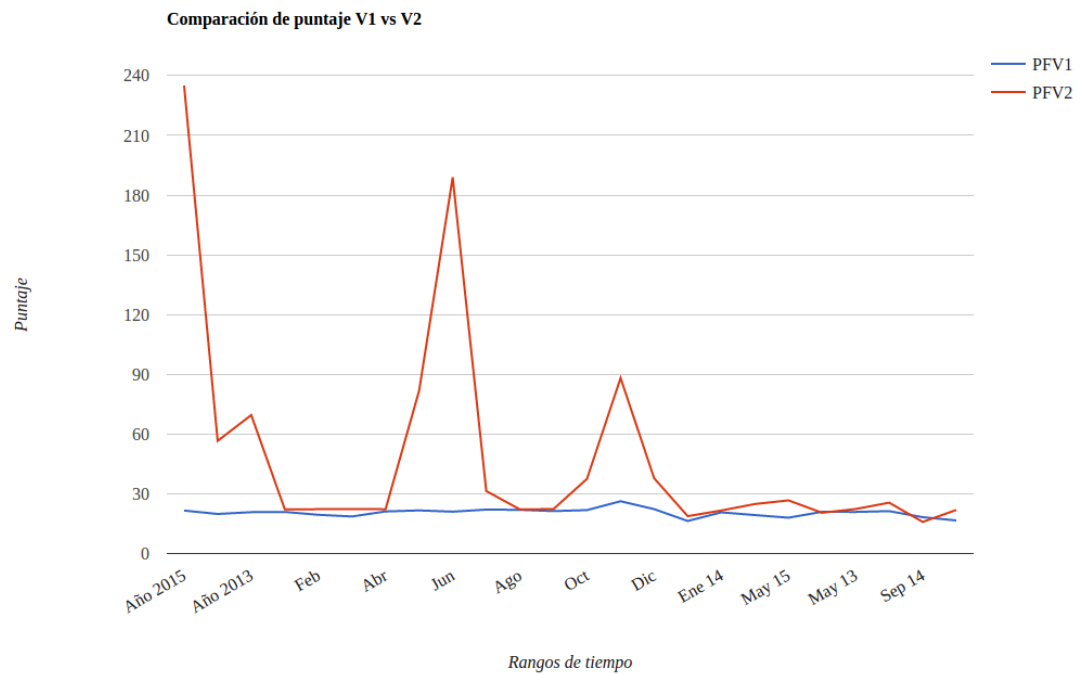
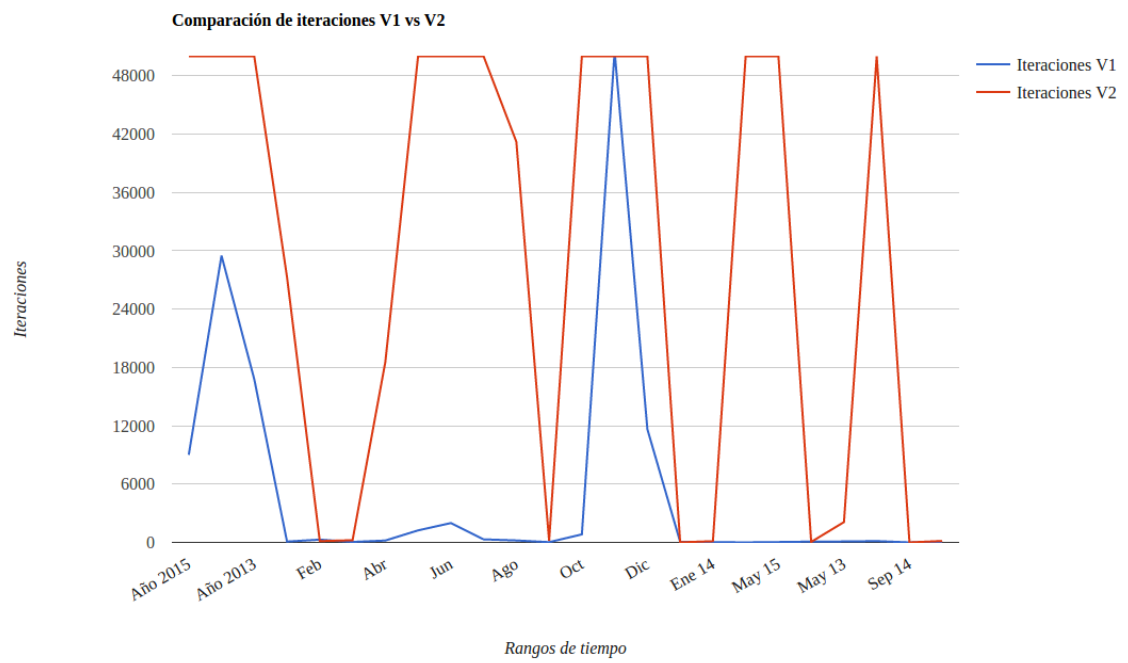


Figura 4.13: Comparación de variaciones en el PSO. Fuente: Elaboración propia.

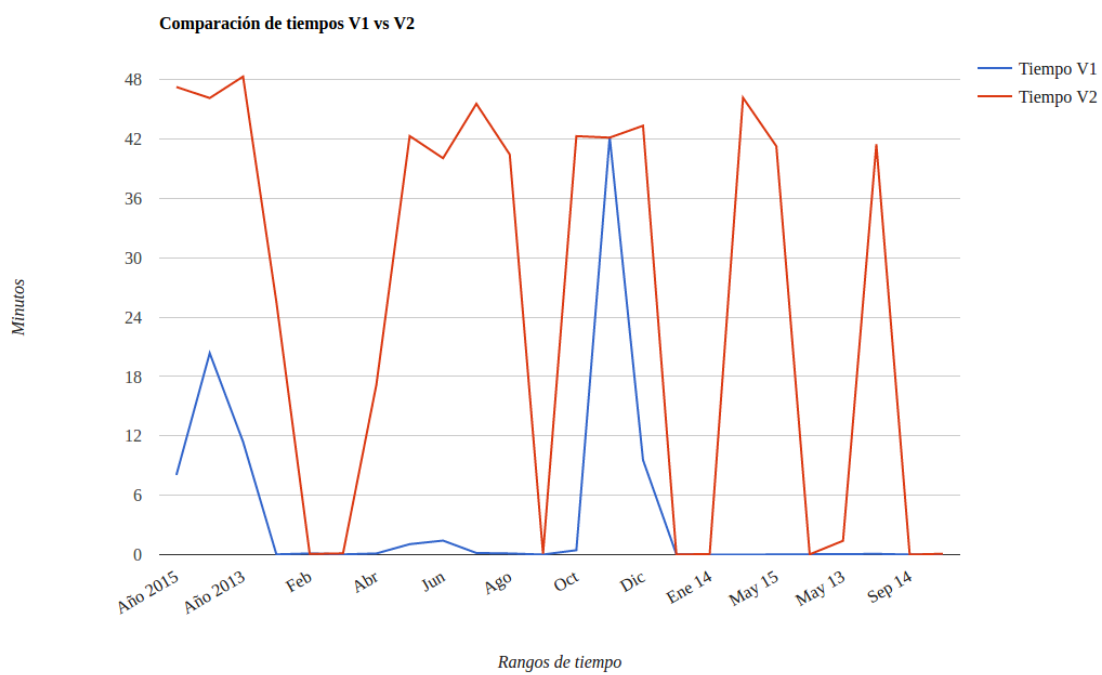


Figura 4.14: Comparación de variaciones en el PSO. Fuente: Elaboración propia.



## 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. P. Chang, “Wind energy assessment incorporating particle swarm optimization method,” 2010.
- [4] 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.
- [5] J. Heckenbergerova, P. Musilek, and P. Krömer, “Optimization of wind direction distribution parameters using particle swarm optimization,” 2015.
- [6] A. Kaveh, “Particle swarm optimization,” pp. 9–40, 2014. Chapter of Advances in Metaheuristic Algorithms for Optimal Design of Structures.
- [7] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of IEEE international conference on neural networks*, pp. 1942 – 1948, 1995.
- [8] C. Kongnam and S. Nuchprayoon, “A particle swarm optimization for wind energy control problem,” 2015.
- [9] D. Fadare, “A statistical analysis of wind energy potential in ibadan, nigeria, based on weibull distribution function,” 2008.

- [10] D. Weisser, “A wind energy analysis if grenada: an estimation using the weibull density function,” 2002.
- [11] Calimo, “Weibull probability density function.” URL [https://commons.wikimedia.org/wiki/File:Weibull\\_PDF.svg](https://commons.wikimedia.org/wiki/File:Weibull_PDF.svg), 2010. [Revisado el 19/06/2016].
- [12] T. P. Chang, “Performance comparison of six numerical methods in estimating weibull parameters for wind energy application,” 2010.
- [13] N. Masseran and A. M. Razali, “Modeling the wind direction behaviors during the monsoon seasons in peninsular malaysia,” 2015.
- [14] 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.
- [15] J. Heckenbergerova, P. Musilek, J. Mejznar, and M. Vancura, “Estimation of wind direction distribution with genetic algorithms,” 2013.
- [16] J. A. Carta, C. Bueno, and P. Ramírez, “Statistical modelling of directional wind speeds using mixtures of von mises distribution: Case study,” 2007.
- [17] 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].
- [18] 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].
- [19] E. L. Ramírez and S. F. M. Vivanco, “Análisis de parámetros metereológicos utilizando el modelo de la rosa de los vientos,” 2007.