



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA MECÁNICA

REDUCCIÓN DE DIMENSIONALIDAD PARA ANÁLISIS DE LA GENERACIÓN DE POTENCIA EN UN PARQUE EÓLICO

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO
CIVIL MECÁNICO**

IGNACIO EDUARDO SANHUEZA CONTRERAS

**PROFESOR GUIA:
BENJAMIN HERRMANN PRIESNITZ**

**MIEMBROS DE LA COMISION:
TOMAS JIL BREYTMANN
WILLIAMS R. CALDERÓN MUÑOZ**

**SANTIAGO DE CHILE
2022**

**RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE:** Ingeniero Civil Mecánico
POR: Ignacio Sanhueza Contreras
FECHA: 2022
PROFESOR GUIA: Benjamín Herrmann Priesnitz

REDUCCIÓN DE DIMENSIONALIDAD PARA ANÁLISIS DE LA GENERACIÓN DE POTENCIA EN UN PARQUE EÓLICO

Un parque eólico es un sistema dinámico complejo, donde decenas de aerogeneradores interactúan entre sí a través de estelas turbulentas. Los datos registrados durante la operación de un parque pueden esconder información valiosa que puede ser aprovechada para optimizar la distribución espacial de las turbinas, diseñar estrategias de control en tiempo real, y aprender sobre la fluidodinámica subyacente. El objetivo de este trabajo es extraer y analizar patrones espaciotemporales dominantes en la dinámica de la generación de potencia en un parque eólico a partir de datos de operación reales. Además, se aplica descomposición en valores singulares (SVD), correlación y regresión para realizar la reducción de dimensionalidad de variables y establecer la importancia de cada variable. Finalmente se utiliza la regresión para predecir la potencia total del parque.

Primero se muestran los antecedentes que contienen las bases con las cuales se realiza el trabajo. Luego se presenta la formulación del problema y de los algoritmos utilizados. Finalmente se presentan los resultados para analizar y concluir sobre los resultados obtenidos.

Se estudia un conjunto de datos el cual proviene de una central eólica instalada en Chile. Se trabajará con el lenguaje computacional Python y con librerías de código abierto que contienen implementaciones de los algoritmos a utilizar.

Primero se muestran y analizan los resultados de la distribución espacial de potencia que entrega la SVD que permite tener un acercamiento a la fluido dinámica del sistema. Luego se muestran los gráficos de acumulación de energía y valores singulares, lo que permite realizar la disminución de dimensionalidad como también estudiar las variables importantes. Luego se realiza la comparación entre la distribución espacial de potencia y variables externas, de manera de encontrar que variable es la que describe cada distribución espacial de potencia. Con lo anterior, se realiza una reconstrucción de potencia tomando una SVD de rango 8. Finalmente se comparan los resultados de la predicción de la potencia total del sistema.

Se logra reducir la dimensionalidad del parque de 55 a 8 patrones espaciales de potencia, lo que cual es un gran avance, también, se logra predecir la potencia total del sistema en diferentes ventanas. También es posible predecir la potencia del sistema en su distribución espacial. Mediante la SVD y la correlación se logra captar cuál de las variables describe los primeros 5 modos, lo que permite conocer las variables que describen mejor la dinámica del sistema.

Agradecimientos

Quiero agradecer a mi tutor Benjamín Herrmann, por darme la oportunidad de estudiar un tema que me apasiona, y que con su conocimiento y apoyo me guio en cada una de las etapas de este proyecto para alcanzar los resultados que buscaba.

También agradecer a la Universidad de Chile por entregarme todos los recursos y medios para llevar a cabo el proceso de investigación.

Por último, quiero agradecer a mi familia y amigos, por apoyarme incluso cuando los ánimos estaban bajos. En especial, quiero dar mención especial a mis padres por siempre estar a mi lado, brindándome palabras de apoyo y siempre con la mejor disposición a ayudarme, apoyándome en los momentos difíciles y celebrando las victorias.

Dedicatoria

Le dedico este trabajo a mis padres, que con paciencia me guiaron y acompañaron todo este camino, siempre fueron incondicionales en todo mi proceso de aprendizaje.

A mi hermana, por su apoyo incondicional, y su cariño que me ayudo a seguir adelante.

También a mis abuelos, de los cuales sus palabras y consejos sabios me ayudaron a centrarme en lo importante, me hacen sentir un nieto orgulloso.

Y a mis amigos, por apoyarme y escucharme en los momentos difíciles.

A todos los llevo siempre en mi corazón.

Tabla de contenido

1. Introducción.....	1
2. Antecedentes.....	3
2.1.1 Dinámica de la generación de potencia en un parque eólico.....	3
2.2 Modelamiento de sistemas dinámicos basado en datos.....	5
2.2.1 Correlación.....	6
2.2.1.1 Pearson.....	6
2.2.1.2 Spearman.....	7
2.2.2 Descomposición en valores singulares.....	7
2.2.3 Descomposición en modos dinámicos.....	9
3. Formulación.....	13
3.1 Parque eólico.....	13
3.2 Algoritmo: SVD, DMD, Correlación y Regresión.....	14
3.2.1 SVD.....	14
3.2.2 DMD.....	15
3.2.3 Correlación.....	15
3.2.4 Regresión de modos POD.....	15
4. Metodología.....	17
4.1 Procesado de datos.....	17
4.2 Implementación de algoritmos.....	19
5. Resultados.....	21
5.1 Mes 1: Agosto.....	21
5.1.1 Ventanas.....	22
5.1.2 Matriz de correlación.....	24
5.1.3 SVD.....	27
5.1.3.1 Gráficos valores singulares y energía.....	27
5.1.3.2 Correlación variable y SVD.....	29
5.1.3.3 Gráficos comparativos: variable y SVD temporal.....	31
5.1.3.4 POD 1.....	32
5.1.4 DMD.....	34
5.1.5 Regresión.....	36
6. Conclusión.....	41
7. Bibliografía.....	43

1. Introducción

En la actualidad, el país se ha visto enfrentado a una variedad de problemas y escenarios energéticos complejos, debido a la gran dependencia de recursos energéticos extranjeros como el petróleo y gas natural. Además, nuevas políticas mundiales exigen una generación de energía más limpia y verde, por lo que las termoeléctricas en el país son tecnología obsoleta para los años venideros.

Es por lo anterior que se han implementado una variedad de medidas, tal como la promulgación de la Ley 20.257, que promueve la instalación de centrales basadas en Energías Renovables No Convencionales (ERNC), tal como la energía eólica. También se encuentra el proyecto de cierre total de las termoeléctricas a partir del 2025. La energía eólica es una de las más utilizadas a nivel mundial, y en Chile existe un gran potencial de generación.

Existen desafíos en el proceso de generación de energía en un parque eólico, ya que sus variables son muy susceptibles a cambiar bruscamente y además hay una gran interacción entre todos los generadores instalados. Esto produce mucha incertidumbre, lo que hace difícil la optimización del sistema como también la toma de decisiones. Lo que se busca al optimizar es maximizar la generación de potencia del sistema completo.

Actualmente estos sistemas se estudian mediante simulaciones computacionales que resuelven las ecuaciones que gobiernan la física involucrada las cuales tienen un costo computacional alto que limita su utilización para realizar optimización o tareas en tiempo real. En la actualidad, con el desarrollo de machine learning es posible solo con datos registrados a través del tiempo describir el sistema mediante una ecuación diferencial y de esta manera predecir los valores de potencia que generará el parque, también es posible identificar relaciones entre aerogeneradores y encontrar las variables que más importan dentro del sistema.

De esto nace la motivación y objetivo del trabajo de título, es lograr describir un sistema dinámico complejo, de manera de predecir, identificar relaciones y rastrear las variables más importantes del sistema y de esta manera ayudar en la optimización del parque eólico como también en la toma de decisiones al futuro. A pesar de lo anterior, uno de los limitantes de la predicción radica en el cambio de factores externos, tal como el momento del día en que se predice la potencia eólica u desviaciones extremas en el flujo del viento. Un ejemplo claro de la aplicación de esta tecnología es la optimización de un parque mediante la desviación de la estela de los generadores para así aumentar la generación de potencia total del sistema.

La Figura 1.1 muestra los resultados de un campo de velocidad encontrados por Gebraad y otros. Estos resultados fueron obtenidos mediante simulaciones computacionales y lo que busca es encontrar un ángulo de yaw tal de disminuir la estela que deja cada aerogenerador y de esta manera aumentar la potencia total del sistema. Al lado izquierdo se observa la simulación sin optimizar el ángulo yaw. Al lado derecho se muestra el parque eólico optimizado. El color rojo-naranja indica velocidades altas y el color azul indica velocidades bajas. (Gebraad, y otros, 2016)

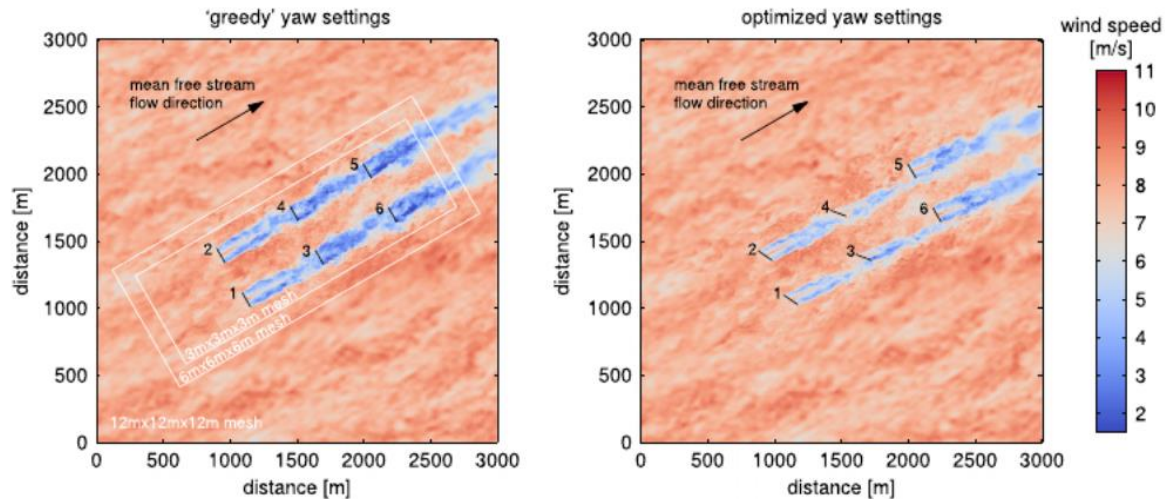


Figura 1.1: Resultados simulación numérica. (Gebraad, y otros, 2016)

Como se observa, en el caso sin optimizar la estela choca directamente con el generador que se encuentra atrás. Con el caso optimizado se observa una clara disminución del color azul en el mapa. Esto implica que el cambio de ángulo reduce de manera significativa la disminución de velocidad de entrada a los generadores, lo que se traduce en un aumento de potencia generada del sistema, que, al fin y al cabo, es la función objetivo a optimizar. (Gebraad, y otros, 2016)

En este caso, el trabajo permite el control de orientación de las turbinas mediante el cambio de ángulo yaw, como también permite la predicción de potencia del sistema a corto plazo. Otros estudios se centran en el posicionamiento de las turbinas en el volumen del control del sistema como es en el caso del paper: “Generalized coupled wake boundary layer model: applications and comparisons with field and LES data for two wind farms” (Stevens, Gayme, & Meneveau, 2016)

Objetivo principal:

- Extraer patrones espacio-temporales dominantes para analizar y predecir la dinámica de la generación de potencia en un parque eólico a partir de datos reales de operación.

Objetivos específicos:

- Procesar datos de operación reales de un parque eólico.
- Identificar un modelo para la dinámica de la generación de potencia en el parque.
- Interpretar los patrones espacio-temporales dominantes en la generación de potencia en términos de la fluidodinámica del sistema.

2. Antecedentes

2.1.1 Dinámica de la generación de potencia en un parque eólico

Se estudia la generación de potencia de un parque eólico ubicado en el norte de Chile. Una de las principales motivaciones son el estudio y predicción del comportamiento de potencia del parque eólico a través del tiempo. La dificultad del estudio se basa en que el sistema es complejo, debido a la gran cantidad de variables que influyen en la generación de potencia, además, es un sistema dinámico, tomando en cuenta que cada generador influye en la producción de potencia de los otros generadores debido a la estela y flujo que deja el generador luego de tener contacto con el aire.

Recientemente, los efectos de interacción de estela se han convertido en un campo de estudio más significativo en la investigación sobre algoritmos de control de turbinas eólicas porque las plantas eólicas han crecido en tamaño y más conocimiento se ha vuelto disponible en la pérdida de eficiencia debido al efecto de interacción con la estela.

Cada aerogenerador de un grupo de aerogeneradores (una planta de energía eólica) puede influir en el rendimiento de otros aerogeneradores a través de la estela que se forma aguas abajo de su rotor. La estela es una estructura de flujo que se caracteriza por una velocidad del viento reducida porque el rotor de la turbina extrae energía cinética del flujo entrante y una turbulencia aumentada porque la turbina obstruye el flujo. Si otra turbina se encuentra en el camino de una estela en un lugar donde el flujo aún no se ha recuperado completamente a las condiciones de flujo libre, la velocidad del viento reducida da como resultado una menor producción de energía eléctrica de esa turbina. La estela puede inducir un aumento de las cargas dinámicas en la turbina. (Gebraad, y otros, 2016)

Los modelos dinámicos tienen una variedad de complicaciones que no son relevantes para los modelos de estado estacionario. Por ejemplo, el efecto temporal de las estructuras turbulentas a gran escala y la no linealidad de la advección afectan los tiempos de viaje entre turbinas. Estas interacciones afectan la producción de energía instantánea de las turbinas individuales dentro de un parque, pero es posible que no cambien la producción de energía promediada en el tiempo. (Shapiro, Starke, Meneveau, & Gayme, 2019)

Existen fenómenos que no se logran observar cuando se estudia un sistema estacionario, tal como, el proporcionar regulación secundaria. La red eléctrica con parques eólicos es más complicada que el caso de una sola turbina debido al acoplamiento aerodinámico entre turbinas que ocurre en escalas de tiempo comparables a las de la señal de referencia reduce su punto de ajuste de potencia. La regulación secundaria es la regulación de potencia activa para el control del equilibrio de generación y demanda instantánea en la red y por ende la regulación de frecuencia.

Es posible que la producción total de energía del parque eólico no disminuye porque las turbinas aguas abajo pueden producir más energía debido a la reducción de los efectos de estela. (Shapiro, Bauweraerts, Meyers, Meneveau, & Gayme, 2017)

El estudio de sistemas complejos y no estacionarios es útil, pero es com. Hay varios métodos para abordar este tipo de problemas, los cuales han ido evolucionando hasta alcanzar algoritmos optimizados y especializados.

El método numérico-físico es el método más utilizado para la resolución de problemas estacionarios y dinámicos con limitación de parámetros. De manera general, este método describe ecuaciones de Navier-Stokes con un modelo de turbulencia y luego de iterar numéricamente se alcanza un resultado. Dos de los modelos más utilizados son: Large Eddy Simulation (LES) y Reynolds averaged Navier-Stokes(RANS). (Stevens, Gayme, & Meneveau, 2016) (Meyers & Meneveau, 2010)

RANS solo puede dar un valor promediado en el tiempo para el campo de velocidad, ya que se basa en promediar el tiempo. De hecho, el campo de velocidad en este método se promedia durante un período de tiempo "t" que es considerablemente más alto que la constante de tiempo de las fluctuaciones de velocidad. Por lo tanto, dentro del período de "t" solo tenemos una velocidad media constante y no podemos monitorear sus variaciones dependientes del tiempo. Por ejemplo, suponga que puede tomar fotografías sucesivas del vector de velocidad en un punto específico de un flujo turbulento durante "t", claramente puede ver que este vector de velocidad varía con el tiempo. Por lo que, si usa RANS, solo puede ver un vector de velocidad constante que es un promedio de las imágenes que ya ha tomado. Por otro lado, LES se basa en filtrar en lugar de promediar. En este método, primero debe elegir un tamaño de filtro, todas las escalas de flujo mayores que el tamaño del filtro especificado se calcularán exactamente y se modelarán las escalas menores que el tamaño del filtro. Ahora si se considera la toma de fotografías de nuevo, si usa LES, puede ver claramente la variación del vector de velocidad en ese punto específico. Al ser menor el tamaño del filtro, más exacta es la resolución de variación de tiempo del vector de velocidad. A medida que el tamaño del filtro llega a cero, los resultados de LES se convierten en DNS (direct numerical simulation). (Martinez, Meneveau, & Stevens, 2016) (Martínez-Tossas, Churchfield, & Meneveau, 2015) (Calaf, Meneveau, & Meyers, 2010)

Una de las similitudes de LES y RANs es que ambos toman un volumen de control para su estudio, la cual tiene que ser definida previamente. En los estudios sobre plantas eólicas, se intenta aumentar las variables a estudiar, por lo que se agregan fórmulas que definen ángulos de yaw y pitch de la turbina y ecuaciones que describen el comportamiento de la estela que deja la turbina.

Un ejemplo claro de aplicación de los métodos de simulación física se encuentra en el paper: "Wind plant power optimization through yaw control using a parametric model for wake effects—a CFD simulation study". Tal como describe su título, se optimiza la potencia generada de un parque eólico mediante el control del ángulo yaw tomando en cuenta el efecto de la estela entre turbinas. (Gebraad, y otros, 2016)

Como no es el tema principal de estudio, en la Figura 2.1 se muestran los elementos generales que componen la optimización de la simulación. En primer lugar, se toman los datos y mediciones actuales y así identificar los parámetros principales del sistema. Luego, los parámetros se ingresan al modelo físico el cual rige la simulación y de esta manera se predice la potencia de

las turbinas. Este último paso, se repite de manera de maximizar la potencia de salida cambiando el ángulo de yaw de las turbinas. Ya optimizado el sistema se configura cada turbina con el ángulo yaw encontrado. Finalmente se repiten todos los procesos. (Gebraad, y otros, 2016)

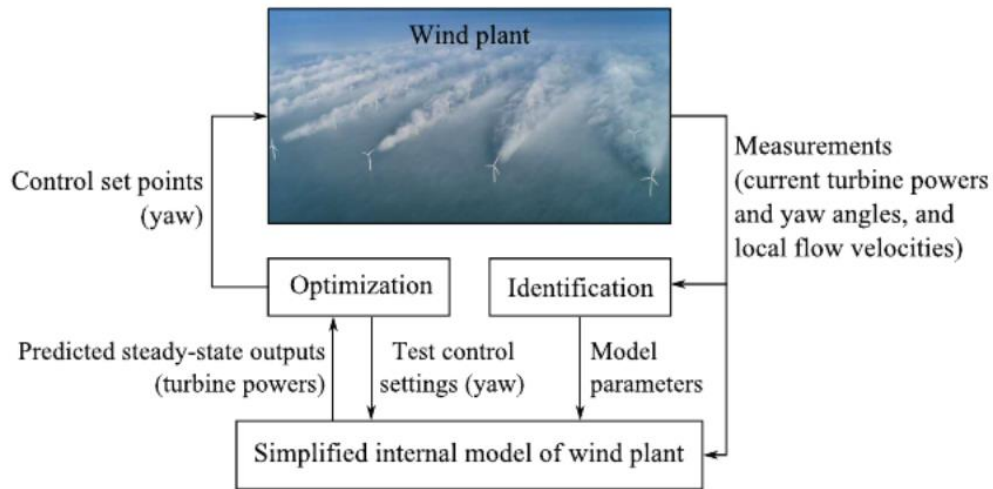


Figura 2.1: Esquema modelación numérica. (Gebraad, y otros, 2016)

Con el paso del tiempo, se han buscado métodos que sean computacionalmente más sencillos y rápidos. El machine learning es un área de la computación que ha ido en aumentando su desarrollo estos últimos años. Una de las ideas de esta tecnología es conseguir relaciones entre las variables del sistema solo utilizando mediciones a través del tiempo, por lo que utiliza solo datos para encontrar resultados. De esta manera la idea principal del machine learning es lograr alcanzar la predicción del sistema a través del tiempo y también encontrar las relaciones reduciendo la dimensionalidad del sistema.

2.2 Modelamiento de sistemas dinámicos basado en datos

Un sistema dinámico es uno cuyo estado cambia a medida que transcurre el tiempo. Este puede ser representado genéricamente de la siguiente manera:

$$\frac{dx}{dt} = f(x) \quad (1)$$

Un sistema dinámico puede ser gobernado usualmente por ecuaciones en diferenciales ordinarias (EDOs). En la práctica, métodos computacionales convierten las EDPs en grandes sistemas de ecuaciones diferenciales ordinarias (EDOs), por ejemplo: mediante la discretización de campos espaciales. Esto permite encontrar soluciones al comportamiento del sistema y lograr estudiarlo.

En la actualidad, se ha logrado describir sistemas dinámicos mediante mediciones de variables a través del tiempo, lo que permite armar y resolver un conjunto de ecuaciones diferenciales. Lo anterior, es lo que realiza el algoritmo DMD, construye y resuelve una matriz de ecuaciones diferenciales ordinarias y de esta manera, entrega resultados del comportamiento del sistema. Como todo algoritmo tiene limitantes, solo es posible predecir un lapsus corto en el tiempo si es que los factores externos varían de manera drástica.

Otro algoritmo que se utilizara es la descomposición en valores singulares (SVD), que de la misma naturaleza del DMD, solo necesita datos para encontrar relaciones. Este algoritmo permite disminuir la cantidad de variables dentro del sistema mediante un sistema de ranking que también permite observar que variable es más importante.

2.2.1 Correlación

La correlación en la estadística es una medida que informa sobre la asociación entre las dos variables. Describe cómo se comporta una variable si hay algún cambio en la otra variable. Si las dos variables aumentan o disminuyen en paralelo, entonces tienen una correlación positiva entre ellas y si una de las variables aumenta y otra disminuye, entonces tienen una correlación negativa entre sí. Si el cambio de una variable no tiene efecto sobre otra variable, entonces tienen una correlación cero entre ellas.

La correlación entrega un coeficiente el cual es una medida estadística que dimensiona la fuerza de la relación entre los movimientos relativos de dos variables. Los valores oscilan entre -1,0 y 1,0. Una correlación de -1,0 muestra una correlación negativa perfecta, mientras que una correlación de 1,0 muestra una correlación positiva perfecta. Una correlación de 0.0 no muestra una relación lineal entre el movimiento de las dos variables. Las correlaciones positivas implican que a medida que x aumenta, también lo hace y. Las correlaciones negativas implican que a medida que x aumenta, y disminuye.

Existen dos tipos de correlación importantes de mencionar: Pearson y Spearman. Ambos métodos comparten la misma idea general de encontrar como se relacionan las variables entre si, con algunas diferencias.

2.2.1.1 Pearson

El coeficiente de correlación de Pearson mide la relación lineal entre dos conjuntos de datos, por lo que cuando una variable aumenta o disminuye en cierta magnitud la otra curva en la misma magnitud.

Para una muestra de datos el coeficiente se calcula de la siguiente manera:

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2(y - m_y)^2}} \quad (2)$$

2.2.1.2 Spearman

Es una medida no paramétrica de la monotonidad¹ de la relación entre dos conjuntos de datos. A diferencia de la correlación de Pearson, la correlación de Spearman no supone que ambos conjuntos de datos estén distribuidos normalmente.

Para una muestra de datos el coeficiente se calcula de la siguiente manera:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (3)$$

n = número de pareja de datos

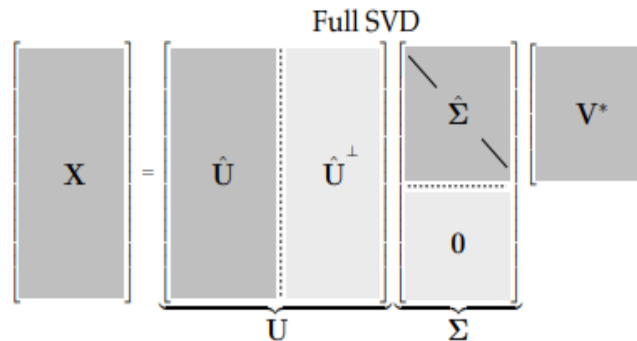
d = diferencia entre par de datos

2.2.2 Descomposición en valores singulares

La descomposición en valores singulares (SVD por sus siglas en inglés) es una factorización matricial ampliamente utilizada para reducir la dimensionalidad para big-data. Este proporciona una forma sistemática de determinar una aproximación de baja dimensión a datos de alta dimensión en términos de patrones dominantes. Esta técnica se basa en datos, ya que los patrones se descubren puramente a partir de datos, sin la adición de conocimiento experto o intuición. La SVD es numéricamente estable y proporciona una representación jerárquica de los datos en términos de un nuevo sistema de coordenadas definido por correlaciones dominantes dentro de los datos. Además, garantiza que la SVD existe para cualquier matriz, a diferencia de la descomposición propia. (Brunton & Kutz, 2019)

La SVD es una descomposición de matriz única que existe para cada matriz de valor complejo $X \in \mathbb{C}^{n \times m}$:

$$X = U \Sigma V^* \quad (4)$$



¹ Una función monótona es aquella que nunca disminuye o nunca aumenta, ya que es un incremento variable independiente

Figura 2.2: Esquema Full SVD (Brunton & Kutz, 2019)

En donde $U \in \mathbb{C}^{n \times n}$ y $V \in \mathbb{C}^{m \times m}$ son matrices unitarias con columnas ortogonales. Y $\Sigma \in \mathbb{R}^{n \times m}$ es una matriz diagonal con valores reales no negativos.

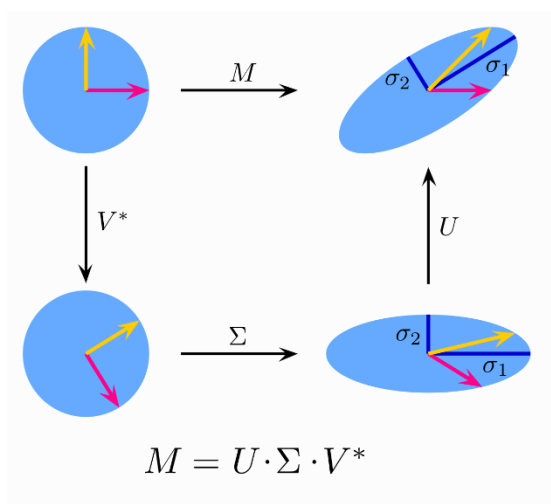


Figura 2.3: Esquema operación SVD.

De manera general e ilustrativa, la Figura 2.2 muestra como la multiplicación de matrices afecta a una matriz M de tamaño 2×2 .

- **Arriba:** La acción de M , indicada por su efecto sobre el disco unitario D y los dos vectores unitarios canónicos e_1 y e_2 .
- **Izquierda:** la acción de V^* , una rotación, en D , e_1 y e_2 .
- **Abajo:** La acción de Σ , una escala por los valores singulares σ_1 horizontalmente y σ_2 verticalmente.
- **Derecha:** La acción de U corresponde a otra rotación.

Explicado gráficamente, la Figura 2.3 muestra las transformaciones del algoritmo SVD en el plano 2D. Luego de realizar las rotaciones y expansiones necesarias se alcanza la elipse de la esquina derecha. Como se mencionó, la SVD tiene como objetivo mostrar un nuevo plano cartesiano que muestra que variable es más importante que la anterior. La elipse de la esquina derecha muestra el nuevo plano cartesiano dado por σ_1 y σ_2 , en donde se ve la importancia de los anteriores vectores según su proyección al nuevo sistema de coordenadas. (Brunton & Kutz, 2019)

Existe una alternativa a la Full SVD, llamada de economy SVD. Esta se puede escribir de la siguiente manera:

$$X = U\Sigma V^* = [\hat{U} \hat{U}^\perp] \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix} V^* = \hat{U} \hat{\Sigma} V^* \quad (5)$$

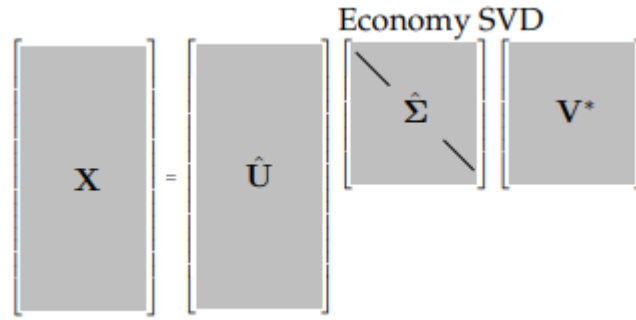


Figura 2.4: Esquema Economy SVD. (Brunton & Kutz, 2019)

Esto puede utilizarse cuando $n \geq m$, la matriz Σ tiene como máximo m elementos distintos de cero en la diagonal, y puede escribirse como $\Sigma = \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix}$.

En ambos tipos de SVD, las columnas $\hat{\mathbf{U}}^\perp$ abarcan un espacio vectorial que es complementario y ortogonal al que abarca $\hat{\mathbf{U}}$. Las columnas de \mathbf{U} se denominan vectores singulares izquierdos de \mathbf{X} y las columnas de \mathbf{V} son vectores singulares derechos. Los elementos diagonales de $\hat{\Sigma} \in \mathcal{C}^{m \times m}$ se denominan valores singulares y están ordenados de mayor a menor. El rango de \mathbf{X} es igual al número de valores singulares distintos de cero. (Brunton & Kutz, 2019)

La aplicación de SVD a ecuaciones diferenciales parciales es conocida como descomposición ortogonal propia (POD por sus siglas en inglés). El éxito de POD se basa en el supuesto que los comportamientos significativos están codificados en patrones de baja dimensionalidad. POD ha encontrado uso principalmente en el rubro de la dinámica de fluidos, aunque también se utiliza en aplicaciones en estructuras, mecánica y análisis vibracional, tecnologías ópticas y MEMS, ciencias atmosféricas, aplicaciones de ingeniería eólica, acústica y neurociencia. (Brunton & Kutz, 2019) (Holmes, Lumley, Berkooz, & Rowley, 2012) (Lumley, 1970)

2.2.3 Descomposición en modos dinámicos

La descomposición en modos dinámicos (DMD por sus siglas en inglés) es una técnica que, a partir de una secuencia temporal de mediciones de un sistema dinámico, simultáneamente encuentra estructuras de baja dimensionalidad y un modelo lineal para describir su evolución. Esta técnica nace en la comunidad de mecánica de fluidos donde se presentan demostraciones del método que consisten en un flujo de canal plano, flujo sobre una cavidad bidimensional, flujo de estela detrás de una membrana flexible y un chorro que pasa entre dos cilindros (Schmid, 2010).

Intenta encontrar estructuras de baja dimensionalidad y de esta manera intentar aproximar el sistema dinámico hacia un sistema dinámico lineal.

$$\frac{dx}{dt} = f(x) \sim \frac{dx}{dt} = Ax \quad (6)$$

Se tiene un sistema con una cantidad grande de variables y todas las mediciones realizadas se colocan en un vector. Luego la ecuación se puede solucionar de la siguiente manera:

$$x = ve^{\lambda t} \quad (7)$$

$$\lambda v = Av \quad (8)$$

Dentro de este sistema se puede conseguir los valores y vectores propios. Estos últimos representan la estructura básica de mi solución.

ϕ_j : vectores propios

λ_j : valores propios

$$x = \sum_{j=1}^n b_j \phi_j e^{\lambda_j t} \quad (9)$$

La solución general de un sistema está dada por la fórmula recién mencionada. La idea de alcanzar un sistema dinámico lineal es siempre llegar a una solución definitiva, ya que de esta manera se logra abarcar todas las soluciones del problema sin tener que dejar algunas fuera y, además, tiene la ventaja de que existe la solución general al problema.

Lo que se busca con el algoritmo DMD es mediante datos alcanzar la solución general de un sistema dinámico lineal.

Se asume que se puede tener medir x_j en cualquier momento del tiempo, tal que: $x_j = x(t_j)$. Con esto se realizará una matriz que contendrá mediciones de variables a través del tiempo:

$$X = [x_1 \dots x_{m-1}] \quad (10)$$

Y otra matriz que comenzará desde x_2 ,

$$X' = [x_2 \dots x_m] \quad (11)$$

Lo que queremos es una forma de describir el futuro, de esta manera se puede escribir,

$$X' = AX \quad (12)$$

Buscar la matriz A tal que, si multiplicamos por X podemos avanzar al futuro, que es la matriz X'.

$$A = \operatorname{argmin} \|X' - AX\|_F = X'X^\dagger \quad (13)$$

Esta última fórmula está definida como el DMD exacto. Donde $\|\cdot\|_F$ corresponde a la norma de Frobenius y \dagger denota la pseudo-inversa.

El problema es cuando se tiene un vector de dimensionalidad muy grande, la matriz A es gigante por lo que no es viable utilizar el sistema propuesto anteriormente. Por lo que se propone el siguiente algoritmo para reducir la dimensionalidad del sistema y trabajar con el DMD.

Paso 1: Aplicar SVD a los datos

$$X = U_r \Sigma_r V_r^*, \text{rank} - r \text{ truncation} \quad (14)$$

Paso 2:

$$A_{n \times n} = X'X^\dagger = X'V\Sigma_r^{-1}U^* \quad (15)$$

Como no se quiere trabajar con una matriz A muy grande, se trabaja con una matriz A' rxr ya que trabaja solo con el espacio truncado. Por ejemplo, la matriz A es 10e9 x 10e9, el otro es 200x200.

$$A'_{r \times r} = U^*AU = U^*X'V\Sigma^{-1} \quad (16)$$

Paso 3: Se encuentran los vectores y valores propios.

$$\hat{A}W = W\Lambda, \quad W = [\text{vectores propios}] \quad \Lambda = [\lambda_1 \cdots 0 \vdots \vdots 0 \cdots \lambda_r] \quad (17)$$

Paso 4: Ahora se vuelve a nuestro sistema original de alta dimensión, para mostrar los resultados.

$$\Phi = X'V\Sigma^{-1}W \quad (18)$$

Lo que culmina en la descripción de los DMD modes, los cuales no son ortogonales. DMD no está restringido a modos ortogonales lo que puede ser una ventaja sobre otros algoritmos.

$$x = \Phi e^{\Omega t} b = \sum_{k=1}^r \phi_k e^{w_k t} b_k \quad (19)$$

$$x(0) = \Phi b \rightarrow b = \Phi \backslash x(0) \quad (20)$$

Este algoritmo funciona especialmente para la predicción a corto plazo. Ya que a largo plazo las relaciones pueden cambiar drásticamente debido al cambio de las variables.

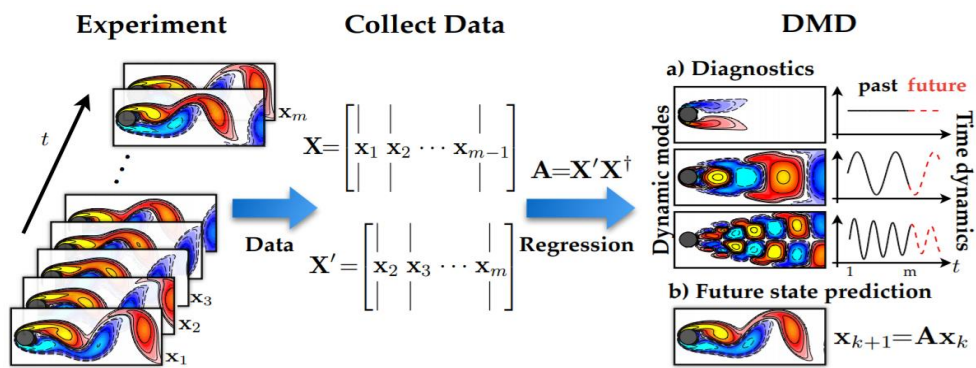


Figura 2.5: Esquema operaciones DMD. (Brunton & Kutz, 2019)

La Figura 2.5 muestra la aplicación del algoritmo DMD a la estela sobre un cilindro. Como se observa, en primer lugar, se tienen mediciones de la estela a través del tiempo. Luego se arman las matrices X y X' para realizar la regresión. Luego de la regresión, se obtiene la descomposición del sistema en modos dinámicos. Y finalmente, se realiza la predicción de los estados. (Brunton & Kutz, 2019) (Hemati, Williams, & Rowley, 2014) (Schmid, 2010)

En la actualidad, existen variaciones del algoritmo DMD que permite acortar las limitantes del algoritmo original. Un ejemplo claro se encuentra el presente trabajo de título, en donde será necesario aplicar DMD en diferentes horas del día debido a que los factores externos cambian radicalmente a través del día. Una opción para solucionar esta limitación es aplicar el algoritmo DMD con control, que toma en cuenta los factores externos dentro del propio algoritmo. (Brunton & Kutz, 2019)

3. Formulación

En la siguiente sección se menciona los lineamientos generales de donde y como se obtuvieron los datos, tal como la orientación y ubicación de los generadores eólicos, la orientación y ubicación de los mástiles de medición y las simulaciones que se realizaron para estimar la presión y temperatura del lugar.

3.1 Parque eólico

El parque está conformado por 55 generadores. Se tiene una base de datos que toma muestra de velocidad del viento en 3 alturas diferentes, ángulos cada 10 minutos, y en relación con los datos anteriores, se tiene una base de datos de producción de potencia de cada generador cada 10 minutos. Además, la empresa genero una simulación de la que nos entregaron los datos de presión y temperatura en las cercanías de parque eólico cada 10 min.

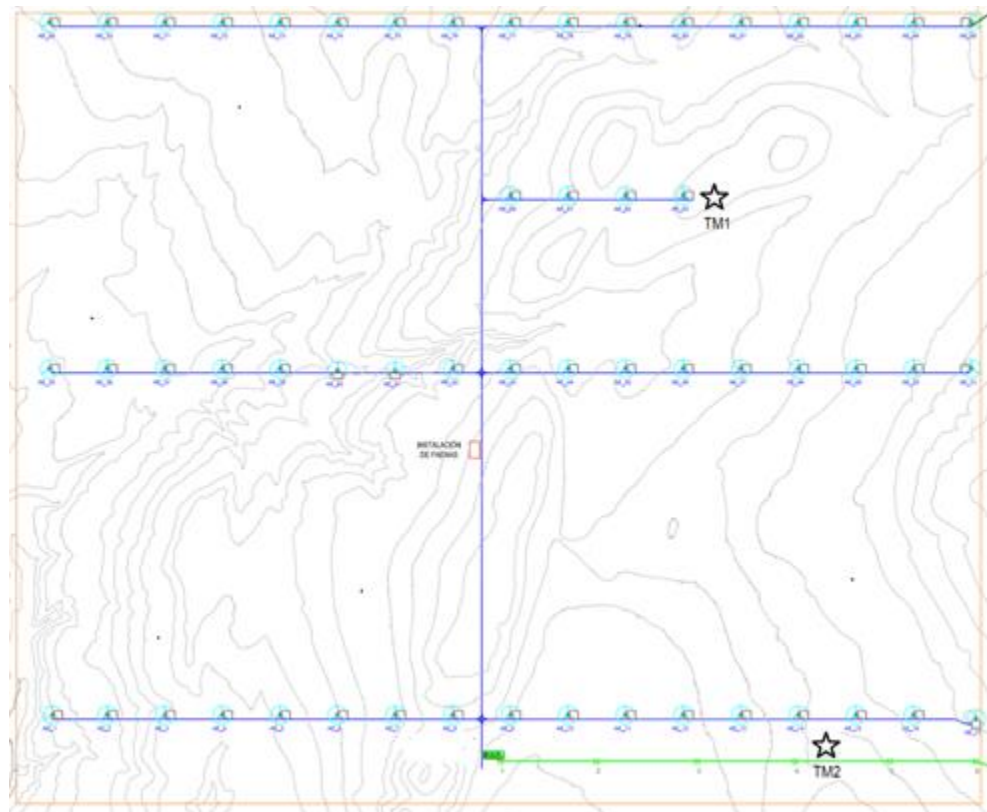


Figura 3.1 Ubicación espacial de generadores eólicos y mástiles de medición TM1 Y TM2.

En la Figura 3.1 se muestra la distribución espacial de los generadores eólicos y los mástiles de medición. Cada generador muestra la velocidad del viento y la potencia generada. Y los mástiles TM1 y TM2 entregan la dirección del viento en grados y la velocidad del viento a 38m, 58m y 78 m de altura.

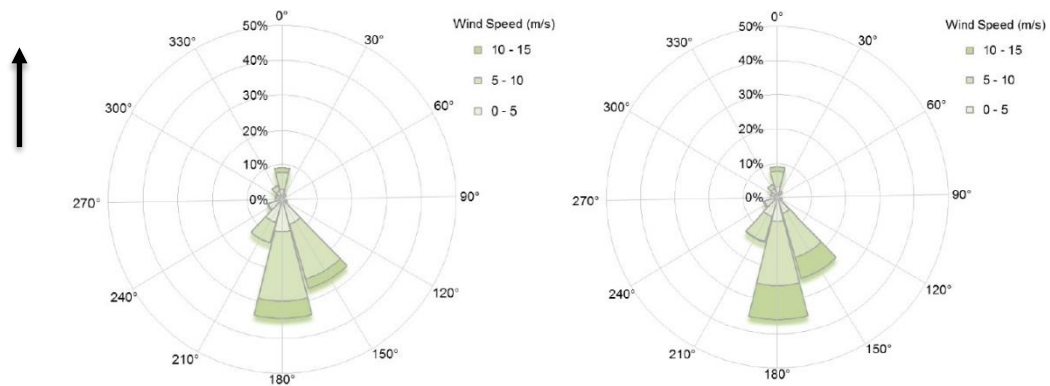


Figura 3.2: Dirección y magnitud del viento.TM1 al lado izquierdo y TM2 al lado derecho.

La Figura 3.2 muestra la rosa de los vientos para los puntos de medición TM1 y TM2, donde se indica la dirección del viento con respecto al ángulo 0° y la magnitud según la escala de colores. El sistema de medición tiene poca sensibilidad, o lo que se ilustra en la imagen son rangos muy grandes de velocidad. En la base de datos la velocidad tiene una sensibilidad de mucha más precisión.

Para los análisis futuros hay que tener claro cómo se superpone la rosa de los vientos por sobre el mapa de los generadores. En este caso es simple la relación, en la Figura 3.1 el lado superior de la imagen corresponde a norte, y el ángulo 0° tiene que está situado en dirección al norte, por lo que se puede visualizar la Figura 3.1 y Figura 3.2 una superpuesta directamente sobre la otra.

Los últimos datos que serán utilizados corresponden a presión y temperatura del ambiente. Cada simulación se realizó en 4 puntos diferentes del mapa. La simulación y los puntos no se pueden mostrar por confidencialidad, pero cabe destacar que los datos simulados no son exactos con respecto a mediciones experimentales, pero de igual manera sirven para estudiar cual es la incidencia de las variables dentro del parque eólico.

3.2 Algoritmo: SVD, DMD, Correlación y Regresión

Cabe destacar que para aplicar el algoritmo será necesario utilizar ventanas de tiempo en donde las variables externas, tales como: viento, temperatura, ángulo, sean constantes, esto debido a los límites del algoritmo DMD.

3.2.1 SVD

El algoritmo SVD se importa de la librería Scipy. Este método de la librería permite obtener las matrices y los modos POD de manera directa. La función SVD acepta el hiperparametro para escoger si se computa las matrices completas del algoritmo. En el caso de estudio, no son matrices extremadamente grandes por lo que no será necesario utilizar la versión económica del SVD.

3.2.2 DMD

Para aplicar el algoritmo DMD en Python se realizó el código de manera personal, ya que se quiere aplicar el método más simple y básico del algoritmo. Se siguieron los pasos mencionados en antecedentes.

```
def dmd(X, Y, truncate=None):
1     U2, Sig2, Vh2 = svd(X, False) # SVD de la matriz de entrada
2     r = len(Sig2) if truncate is None else truncate # Truncar
3     U = U2[:, :r]
4     Sig = diag(Sig2)[:r, :r]
5     V = Vh2.conj().T[:, :r]
6     Atil = dot(dot(dot(U.conj().T, Y), V), inv(Sig)) # Armar A
7 tilde
8     mu, W = eig(Atil)
9     Phi = dot(dot(dot(Y, V), inv(Sig)), W) # Armar modos
10 dinamicos
    return mu, Phi
```

Figura 3.3: Formulación de código Python para algoritmo DMD.

3.2.3 Correlación

La formulación para utilizar el algoritmo de correlación fue importada por la librería Scipy.

Este método permite encontrar la correlación entre dos vectores. En específico se utilizó la correlación Spearman, debido a que la Pearson tiene el supuesto en buscar la relación lineal entre ambos y puede que esto no ocurra en nuestros datos, además, como no se conoce por cual distribución se rigen los datos es de utilidad utilizar la correlación de Spearman ya que es no paramétrica, mientras que Pearson tiene como supuesto que los datos este distribuidos normalmente.

3.2.4 Regresión de modos POD

Para realizar la regresión, se busca un modelo que indique el aporte de cada modo en función de las variables externas.

En primer lugar, se encuentran los modos POD mediante SVD, tal que:

$$u : \text{modo POD}$$
$$U : \text{matriz que contiene modos dominantes}$$

De esta manera se obtiene la SVD:

$$x(t) = U \cdot \Sigma \cdot v^T \quad (21)$$

$$a(t) = \Sigma \cdot v^T \quad (22)$$

Se toma $a(t)_{55 \times t}$ que corresponde a las amplitudes de los modos y se despeja $\gamma_{5 \times 8}$ tal que, multiplicada por las variables de control, $e_{t \times 55}$, me entregue $a(t)_{55 \times t}$.

$$a(t) = e \cdot \gamma \quad (23)$$

$$\gamma = E \cdot \Sigma \cdot v^T \quad (24)$$

Donde E corresponde a datos de e en una ventana de tiempo t. Ya con gamma despejado se procede a reemplazar en otras ventanas y observar que tal se ajusta a nuevas variables externas, manteniendo los modos U de la ventana de entrenamiento.

$$x(t) = U \cdot \gamma \cdot v^T \quad (25)$$

De esta manera se asume que la proyección sobre los mismos modos sigue la misma función, por lo que los modos de entrenamiento son representativos para todas las ventanas. Es por eso, que se enfoca en modelar sólo la proyección sobre estos modos U.

4. Metodología

En la presente sección se mencionan los pasos necesarios para el desarrollo del trabajo de título que se separó en dos secciones principales. La primera corresponde a la construcción de la base de datos que será analizada. La segunda corresponde a la aplicación de los algoritmos DMD y SVD a la base de datos, para obtener los resultados finales, que corresponden a encontrar las variables más importantes dentro del parque eólico, predecir y encontrar cómo se comporta la potencia a través del tiempo. En la sección anterior nos muestran los pilares de cómo se obtienen los datos, los cuales están crudos y sin ajustar, por lo que es una parte muy importante del trabajo de título ajustar la base de datos.

Antes de ingresar los datos a cualquier algoritmo es necesario realizar un preprocesamiento de datos que se realizará mediante excel y python. Esto corresponde a limpieza de outliers, limpieza de mediciones mal realizadas, reemplazo de celdas vacías, entre otras aplicaciones que sean pertinentes. De esta manera, se asegura que los resultados sean fiables ya que la data que se utilizara es correcta.

4.1 Procesado de datos

Los datos corresponden a datos de mediciones en tiempo real y mediciones en simulaciones. Los datos se exportan de Excel en donde se tienen que adquirir las columnas que se van a incluir en la base de datos final, que corresponde a la potencia por generador, la velocidad del viento en la posición de cada generador, la dirección del viento, la velocidad del viento en los mástiles de medición y la presión y temperatura de las simulaciones.

Luego en Python, en primer lugar, se rellena todo el mes ya que hay momentos en que el sistema de adquisición fallaba por lo que hay lapsos de tiempo que no existen. Luego, al tener todo el mes separado en lapsos de 10 minutos hay que rellenar las celdas vacías, se integra el valor de la potencia y velocidad por turbina, si es que existen celdas vacías se realiza una interpolación entre los extremos, de esta manera se realiza una aproximación aceptable. De igual manera, ya en el año 2019 no era común que existieran fallas y, además, este tipo de fallas era de corta duración, por lo que los datos son fiables.

Luego se agregan a la base de datos el valor de la velocidad de los mástiles de dirección y también la dirección del viento. Estos son del mes por lo que no son por turbina. Y finalmente, se agregan los datos las simulaciones sobre la temperatura y presión del parque eólico.

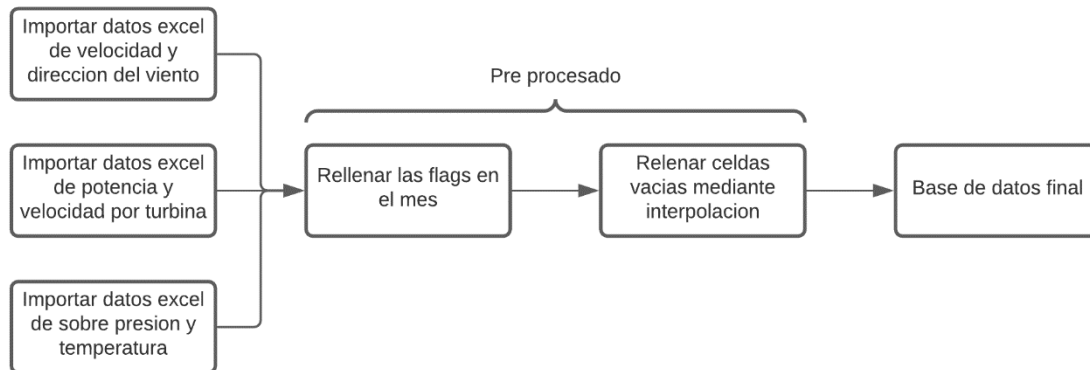


Figura 4.1: Esquema de creación de base de datos.

	Turbina	Fecha	Potencia	Velocidad	Velocidad_viento	Direccion_viento	Temperatura	Presion
Turbina 1	WTG-01	-	-	-	-	-	-	-
	WTG-01	-	-	-	-	-	-	-
	WTG-01	-	-	-	-	-	-	-
	WTG-01	-	-	-	-	-	-	-
	WTG-01	-	-	-	-	-	-	-
Turbina 2	WTG-02	-	-	-	-	-	-	-
	WTG-02	-	-	-	-	-	-	-
	WTG-02	-	-	-	-	-	-	-
	WTG-02	-	-	-	-	-	-	-
	WTG-02	-	-	-	-	-	-	-
Turbina 55	WTG-55	-	-	-	-	-	-	-
	WTG-55	-	-	-	-	-	-	-
	WTG-55	-	-	-	-	-	-	-
	WTG-55	-	-	-	-	-	-	-
	WTG-55	-	-	-	-	-	-	-

Figura 4.2: Esquema organización base de datos.

También se exportan los Excel que contienen los datos de las coordenadas de cada turbina eólica. Esta está distribuida en cada fila es una turbina de 1 hasta 55 y dos columnas que corresponden a la componente X e Y.

Es importante mencionar que se seleccionó pequeñas ventanas de datos en donde la dirección del viento se comportaba de manera cíclica, este proceso se repite a través del mes seleccionado.

Para aplicar los algoritmos de SVD y DMD es necesario crear una matriz con valores de la potencia de cada turbina en distintos instantes de tiempo de potencia y turbinas a través del tiempo. Esta matriz se extrae de la base de datos anteriormente mencionada. De esta manera se construye la matriz, espacio-temporal que contiene en sus filas las turbinas y las columnas la potencia en cada lapso de tiempo.



Figura 4.3: Esquema matriz de potencia.

4.2 Implementación de algoritmos

En primer lugar, se aplica el algoritmo SVD sobre la matriz de potencia espacio-temporal. En este caso se ingresan los datos al algoritmo y se observan los resultados de las matrices entregadas. Este proceso se repite para cada ventana de tiempo creada.

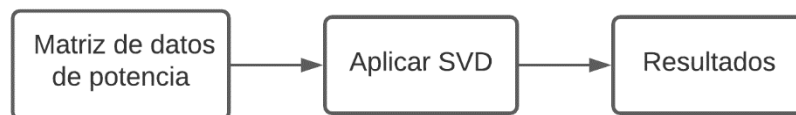


Figura 4.4: Esquema aplicación SVD.

Luego se aplica el algoritmo DMD sobre la matriz de potencia espacio-temporal. En este caso se ingresan los datos al algoritmo y se observan los resultados de las matrices entregadas. Este proceso se repite para cada ventana de tiempo creada.



Figura 4.5: Esquema aplicación DMD.

Para aplicar la correlación se toman dos vectores con la misma cantidad de datos. Se ingresan ambos datos al algoritmo y se guarda el coeficiente de correlación.

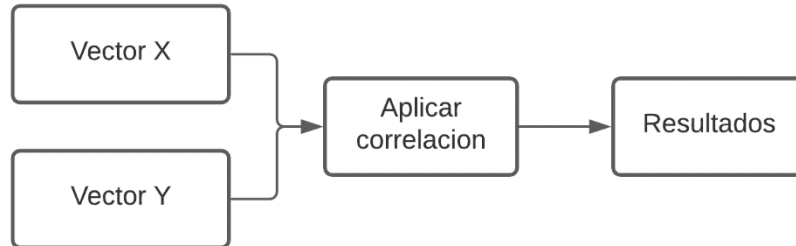


Figura 4.6: Esquema aplicación correlación.

Luego se aplica la regresión sobre la matriz de potencia espacio-temporal. En este caso se ingresan los datos al algoritmo y se observan los resultados de las matrices entregadas. Este proceso se repite para cada ventana de tiempo creada.



Figura 4.7: Esquema aplicación regresión.

5. Resultados

En esta sección se muestran los resultados principales entregados por el SVD y DMD. Todos los resultados se entregarán según el mes que se estudió y además comentando las ventanas del mes seleccionado.

5.1 Mes 1: Agosto

Para esta sección se entregan los resultados para el mes de agosto del 2019.

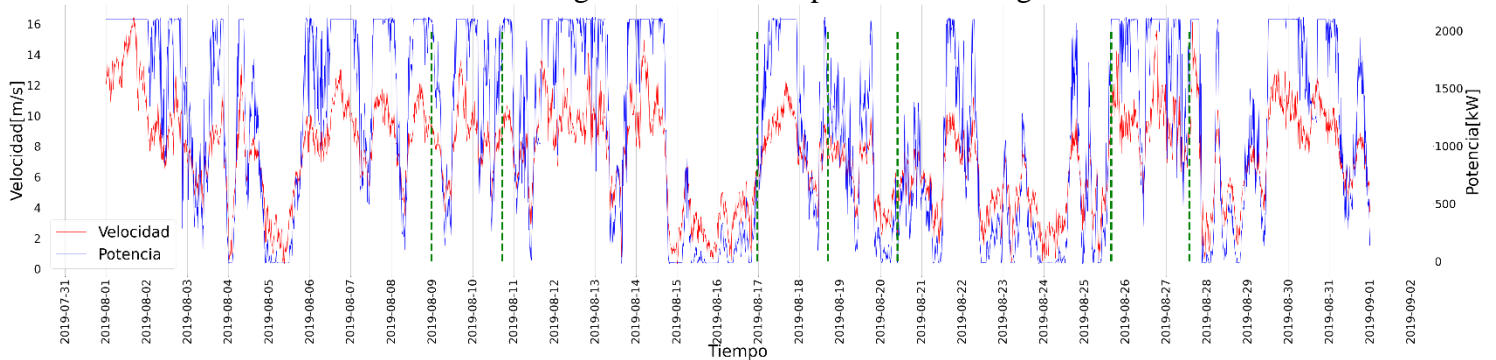


Figura 5.1: Gráfico de velocidad y potencia de un aerogenerador.

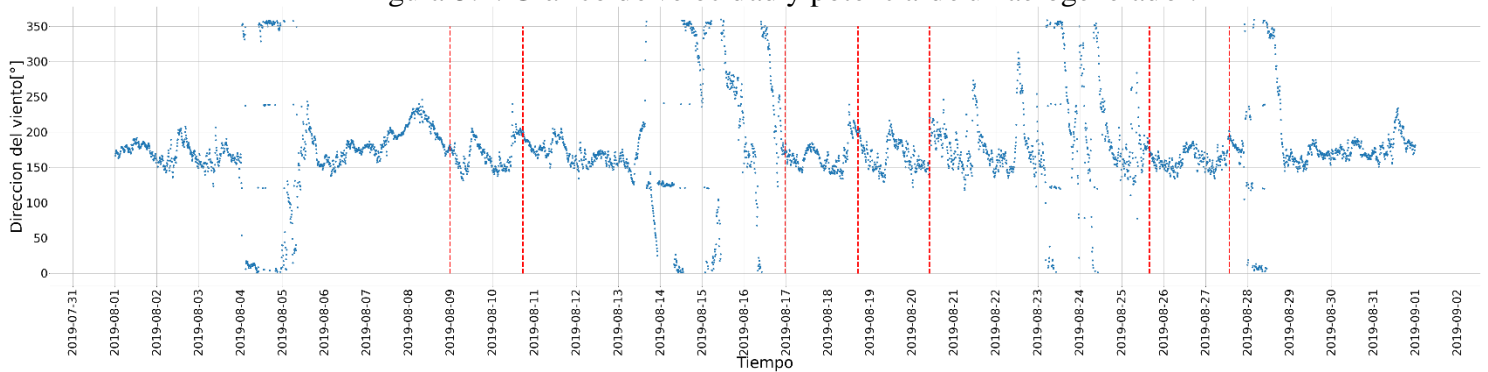


Figura 5.2: Gráfico dirección del viento de un aerogenerador.

Como se observa en la Figura 5.1 y Figura 5.2 se escogieron 4 ventanas en donde el criterio de elección es intentar que la dirección del viento tenga fluctuaciones pequeñas a través del tiempo, de esta manera se intenta cumplir con el supuesto para la aplicación de DMD.

5.1.1 Ventanas

Esta ventana corresponde al mes de agosto del año 2019 entre las fechas 9/8 y 10/8.

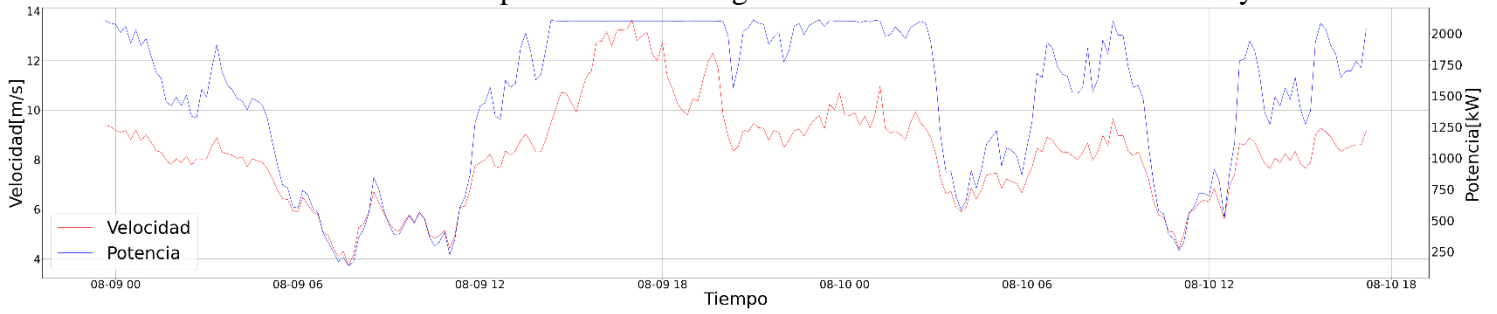


Figura 5.3: Gráfico de velocidad y potencia para la ventana 1.

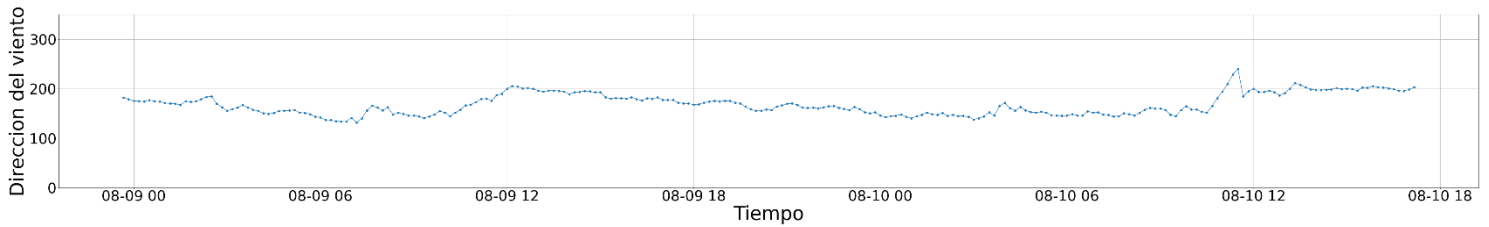


Figura 5.4: Gráfico de dirección del viento ventana 1.

Esta ventana corresponde al mes de agosto del año 2019 entre las fechas 17/8 y 18/8

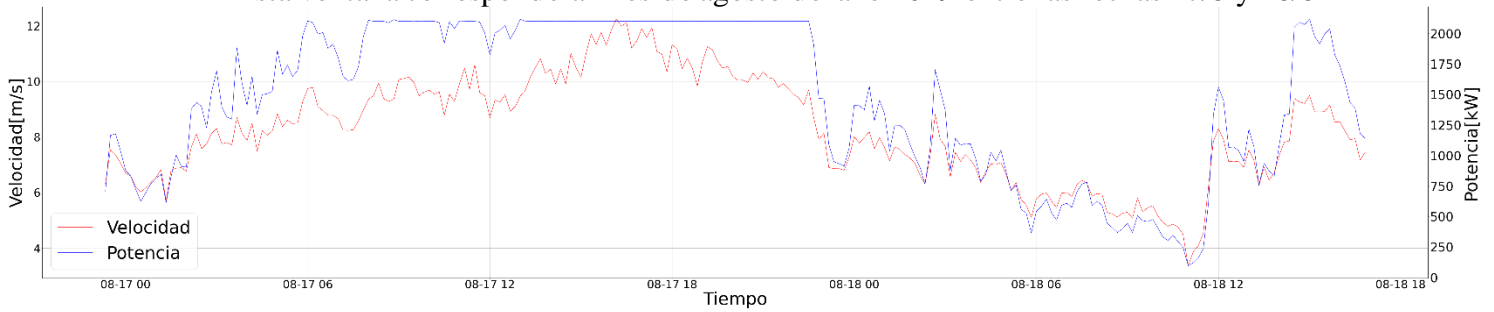


Figura 5.5: Gráfico de velocidad y potencia para la ventana 2.

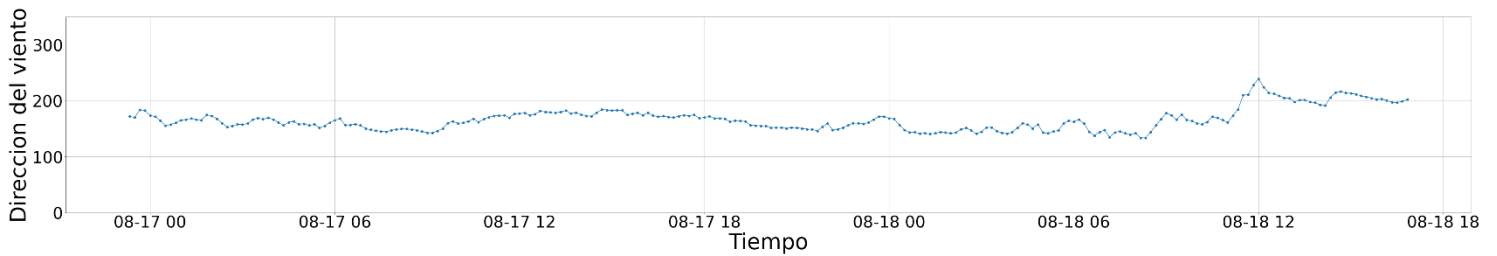


Figura 5.6: Gráfico de dirección del viento ventana 2.

Esta ventana corresponde al mes de agosto del año 2019 entre las fechas 18/8 y 20/8

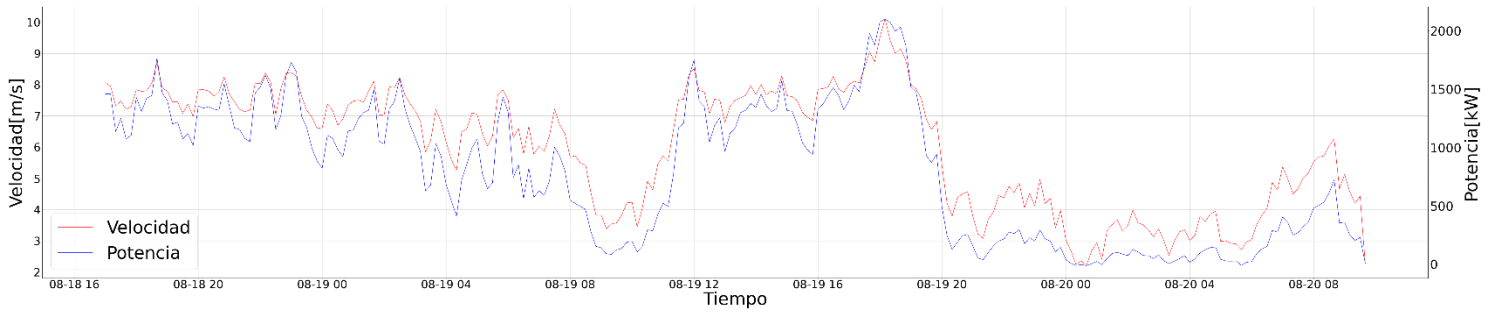


Figura 5.7: Gráfico de velocidad y potencia para la ventana 3.

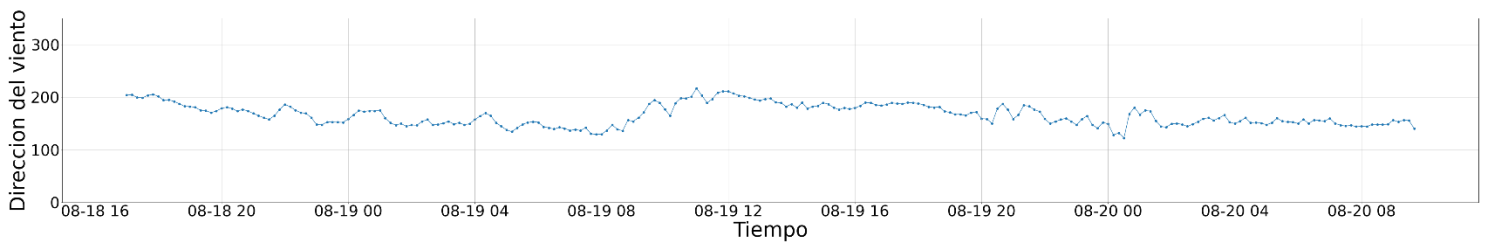


Figura 5.8: Gráfico de dirección del viento ventana 3.

Esta ventana corresponde al mes de agosto del año 2019 entre las fechas 25/8 y 27/8

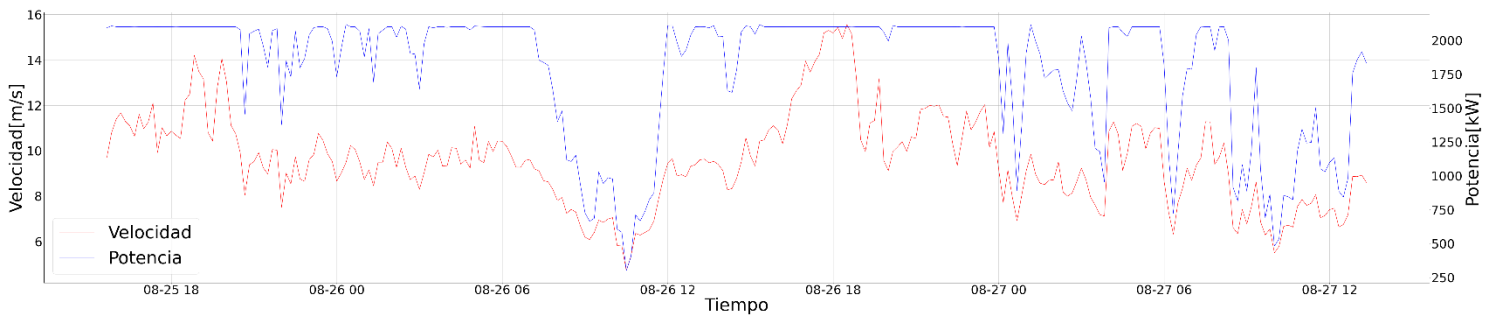


Figura 5.9: Gráfico de velocidad y potencia para la ventana 4.

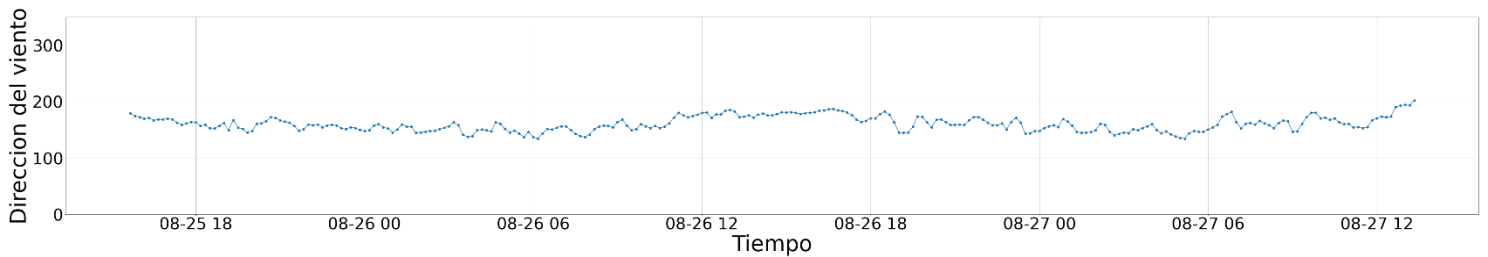


Figura 5.10: Gráfico de dirección del viento ventana 4.

Se muestra en la Figura 5.4, Figura 5.6, Figura 5.8 y Figura 5.10 se observa que en comparación a los extremos a los que puede llegar la dirección del viento, esta se mantiene constante, que es uno de los supuestos para aplicar los algoritmos.

Por otro lado, en la Figura 5.3, Figura 5.5, Figura 5.7, Figura 5.9 se observa la velocidad del viento y la generación de potencia. La velocidad fluctúa entre 2 m/s hasta 16 m/s y la potencia alcanza muchas veces el máximo de 2000 kW. Esto último produce saturación de potencia que ocurre a una velocidad de 10 m/s, por lo que velocidades más altas no son aprovechadas por el generador eólico.

5.1.2 Matriz de correlación

Para tener un acercamiento a las variables se realiza una matriz de correlación como se observa en las Figura 5.11, Figura 5.12, Figura 5.13 , Figura 5.14, mediante Spearman. De esta manera se puede conocer cuál de las variables están completamente relacionadas entre sí y así comenzar a descartar variables.

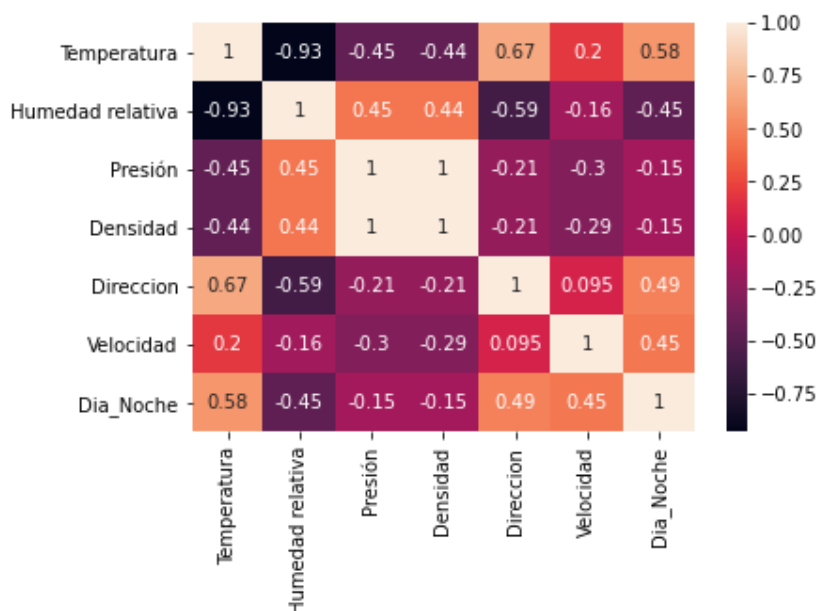


Figura 5.11: Matriz de correlación ventana 1.

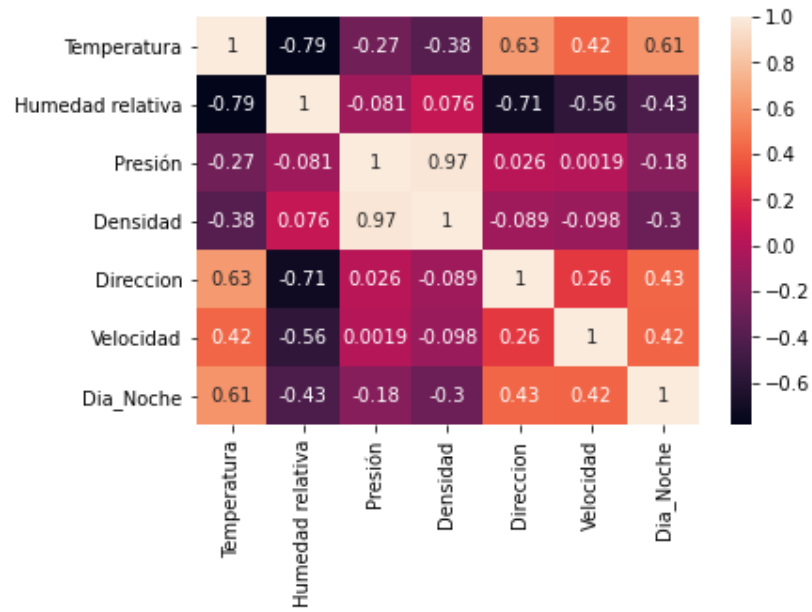


Figura 5.12: Matriz de correlación ventana 2.

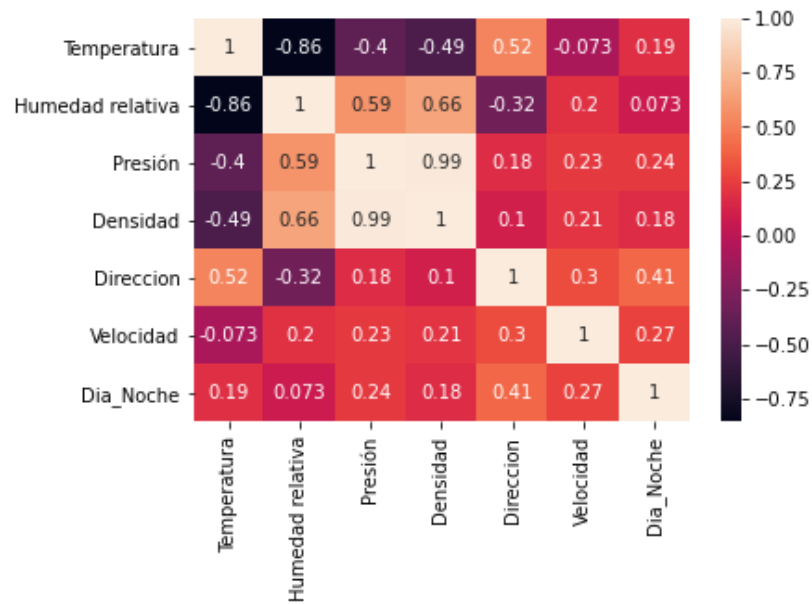


Figura 5.13: Matriz de correlación ventana 3.

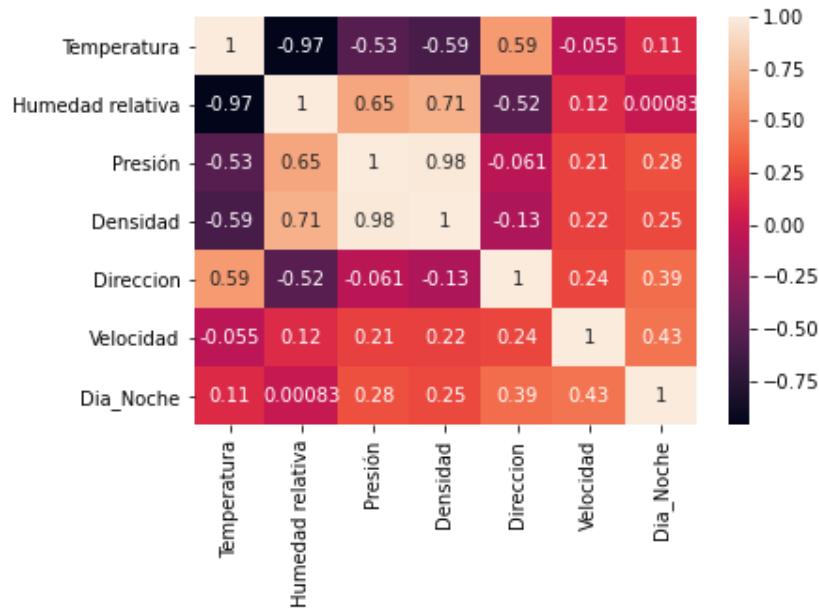


Figura 5.14: Matriz de correlación ventana 4.

En la sección de resultados en primer lugar se observan los gráficos de correlación entre variables, de lo cual se pueden obtener un estudio preliminar del comportamiento entre las variables y realizar un análisis de reducción de dimensionalidad. En la Figura 5.11, Figura 5.12, Figura 5.13, Figura 5.14 se observan dos correlaciones muy fuertes, con magnitud superior a 0.8. La primera corresponde a la temperatura y la humedad relativa con una correlación negativa de alta magnitud, lo que indica que si la temperatura aumenta la humedad relativa disminuye, lo que tiene una explicación lógica con respecto al medio en el que se encuentran. A pesar de lo anterior, ambas de las variables se mantienen en estudios futuros, al tener correlación negativa no es tan claro el dejar una u otra variable afuera. Por el contrario, existe otra correlación fuerte pero positiva entre las variables de densidad y presión, con un valor cercano a 1, lo que nos indica claramente que ambas variables describen el mismo fenómeno. Con lo anterior es posible dejar una de las variables fuera para estudio de SVD o predicción, ya que solo es necesaria una variable para describir un fenómeno.

También existen correlaciones de media-fuerte intensidad, tal como la dirección del viento/temperatura y hora del día/temperatura. La más clara de explicar es la última, ya que se conoce que dependiendo de la hora la temperatura puede oscilar entre temperaturas negativas en la noche y en el día alcanzar máximas de 30°. La más compleja de explicar es cómo puede variar la temperatura según la dirección del viento, aunque puede ser posible que cuando el viento proviene de la costa es más frío que cuando proviene del interior. De igual manera, todas estas variables se toman en cuenta para la aplicación de SVD y regresión, ya que la correlación no es suficientemente fuerte.

5.1.3 SVD

5.1.3.1 Gráficos valores singulares y energía

Luego se aplica el algoritmo SVD a los datos de potencia, donde mediante estos resultados se espera truncar los modos POD más importantes para captar la mayor cantidad de información o energía sin tener que utilizar todas las variables

Se muestran los resultados principales de SVD que corresponden al gráfico de valores singulares y el gráfico de energía, mediante estos es posible tomar una decisión para cual será el rango de truncado para logra captar la mayor cantidad de energía disminuyendo la cantidad de variables.

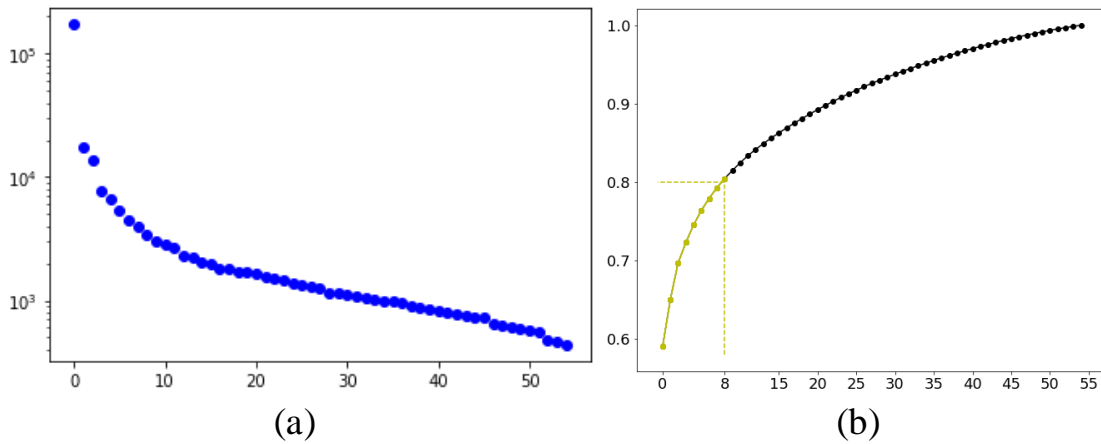


Figura 5.15: Gráfico de valores singulares SVD ventana 1.

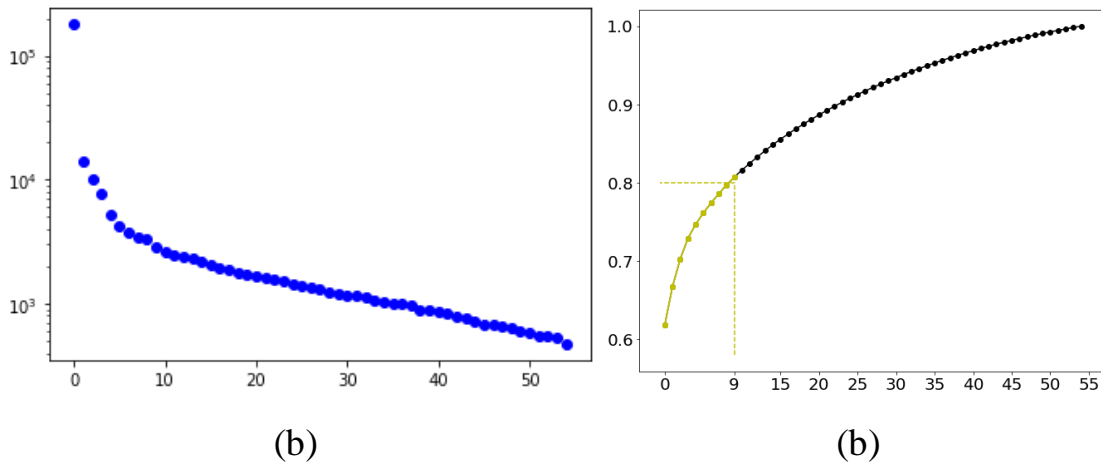


Figura 5.16: Gráfico de valores singulares SVD ventana 2.

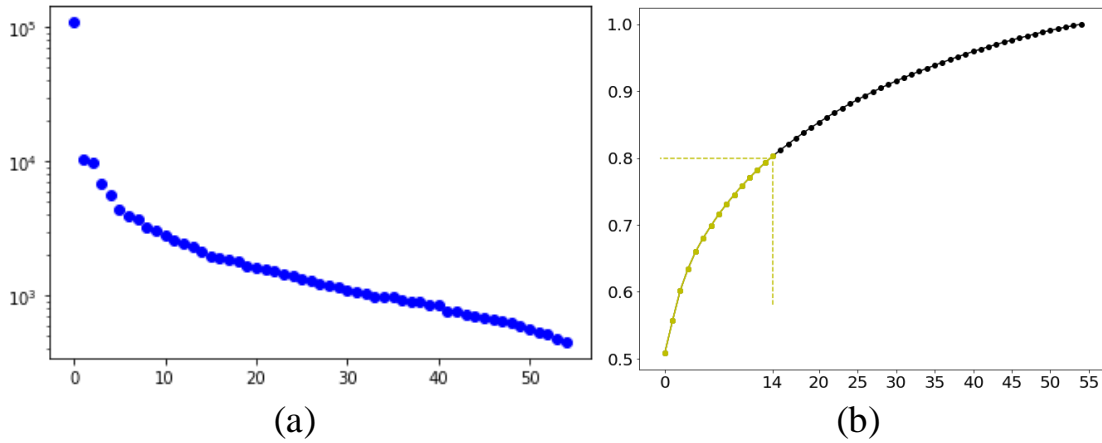


Figura 5.17: Gráfico de valores singulares SVD ventana 3.

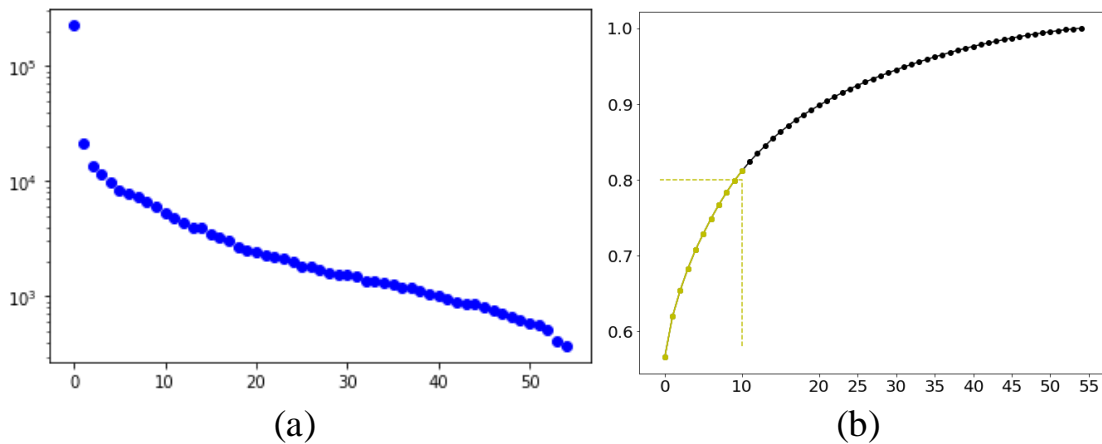


Figura 5.18: Gráfico de valores singulares SVD ventana 4.

Luego se aplica la SVD y se muestran los graficos en la Figura 5.15b, Figura 5.16b, Figura 5.17b, Figura 5.18b que muestran acumulacion de energia el cual muestra la cantidad de informacion porcentual es almacenada según el valor singular de truncado. De manera explicativa, muestran los graficos en la Figura 5.15a, Figura 5.16a, Figura 5.17a, Figura 5.18a la curva azul muestra en el eje Y la importancia de valores singulares.

De manera general, en primer lugar se logra inferir que claramente el valor singular 1 es significativamente mas importante que los demas, esto indica que el modo 1 tiene la mayor informacion sobre el sistema y, por lo tanto, que es este modo el cual rige el parque eolico. En segundo lugar, para realizar la reduccion de dimensionalidad se toma como limite almacenar aunque sea un 80% de informacion, para esto se trazan las lineas en cada grafico de energia y se infiere que son necesarios entre 8 y 14 valores singulares, por lo que se tomara el minimo de este rango, ya que no tenemos una cantidad grande para explicar mas de 8 modos.

5.1.3.2 Correlación variable y SVD

Se muestran los resultados de la correlación entre los modos del SVD temporales con las variables en estudio. De esta manera podemos tener otro acercamiento de cuál de las variables será necesario tomar en cuenta.

	Variable	Correlación con SVD	Columna (modo)
Ventana 1	Velocidad	0.93	1
	Dirección del viento	0.62	4
	Temperatura	0.64	6
	Componente de temperatura sobre velocidad	0.55	3
	Componente de dirección sobre velocidad	0.75	3
	Densidad de aire	-0.60	11
	Presión	-0.35	11
Ventana 2	Velocidad	0.94	1
	Presión	0.59	2
	Componente de presión sobre velocidad	0.63	3
	Componente de presión sobre velocidad	-0.93	1
Ventana 3	Velocidad	0.88	1
	Dirección del viento	0.61	4
	Temperatura	0.52	3
	Presión	0.36	2
Ventana 4	Velocidad	0.84	1
	Dirección del viento	-0.38	4
	Humedad relativa	0.75	2
	Presión	0.76	2
	Densidad	0.76	2
	Componente de temperatura sobre velocidad	0.49	5
	Componente de dirección sobre velocidad	0.47	5

Tabla 5.1: Recapitulación de los resultados importantes.

Ahora en un estudio más exhaustivo de esta comparación, si se analiza la Tabla 5.1 se observan varios fenómenos, en primer lugar, se corrobora que en todas las ventanas la velocidad corresponde al modo 1 de la SVD. En segundo lugar, se observa como en algunos casos existen variables que describen el mismo modo, esto indica que es posible que otros modos puedan ser descritos por combinaciones entre más de 2 variables. En tercer lugar, se observa que hay modos que se repiten entre 1 y 4, por lo que las variables que se están comparando muestra que ese es el rango que permiten describir, ya para modos sobre 5 no tenemos en el poder variables que logren describir esos modos, por lo que, o son variables que no han sido medidas y por lo tanto no se encuentran en la base de datos o son combinaciones u son proyecciones entre variables. Finalmente, no se observa una cantidad grande de magnitudes negativas en la correlación, por lo que es posible desechar la idea de que hay existencia de correlación negativas dentro de nuestro sistema entre variables y modos.

Variables	Correlación	Modo
Velocidad	0,9	1
Presión	0,6	2
Temperatura	0,6	3
Dirección del viento	0,62	4

Tabla 5.2 : Tabla resumen de resultados entre modos y variables.

Luego se muestran los resultados que indica a que variable describe cada modo de la SVD. Esto es importante ya que de esta manera podemos tener una idea más elaborada de cuáles son las variables que rigen el sistema y son las que verdaderamente importan. En la Tabla 5.2 se muestra un resumen de las variables asignadas a su respectivo modo mediante correlación, en donde hay una correlación muy alta con la velocidad del viento en relación con el modo 1. Esto es interesante ya que el modo 1 nos indica el promedio de potencia generada dentro del sistema, y al ser un parque eólico, es lógico que la variable más importante sea la velocidad del viento lo que se corrobora con los resultados encontrados en la comparación entre SVD y variables. Ya para los siguientes modos la correlación baja sustancialmente a un promedio de 0.63, lo que corresponde a una magnitud media-fuerte de correlación, lo que según la bibliografía sigue siendo importante y se puede traducir en que cada variable describe el comportamiento de cada modo.

5.1.3.3 Gráficos comparativos: variable y SVD temporal

En la parte inferior se muestran los gráficos comparando una variable de estudio con el modo correspondiente al SVD, de esta manera se puede observar de manera más visual como varían ambas curvas a través del tiempo y la relación entre ellas.

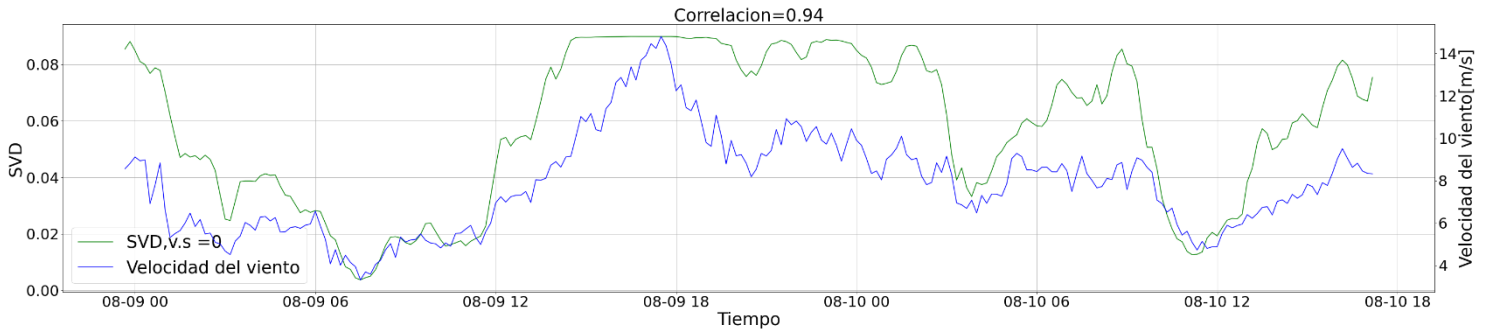


Figura 5.19: Gráfico comparación SVD y velocidad del viento. Ventana 1.

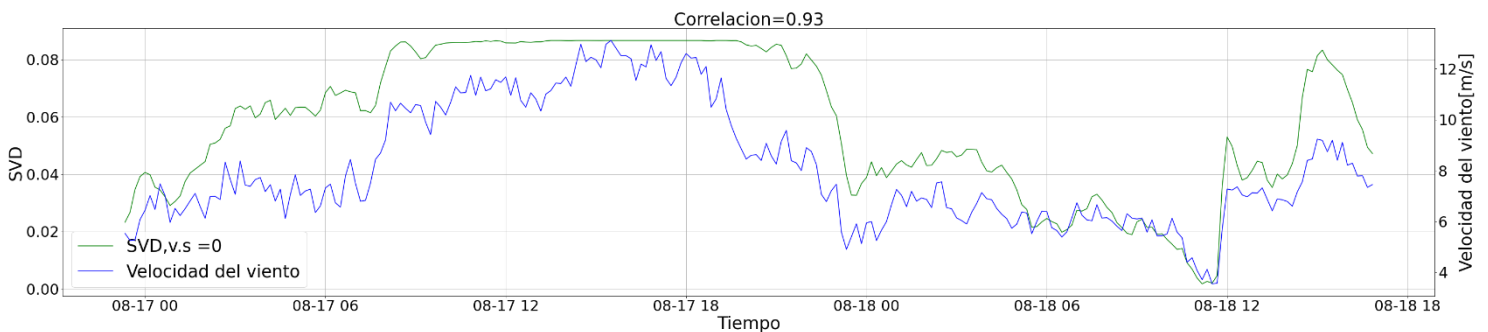


Figura 5.20: Gráfico comparación SVD y velocidad del viento. Ventana 2.

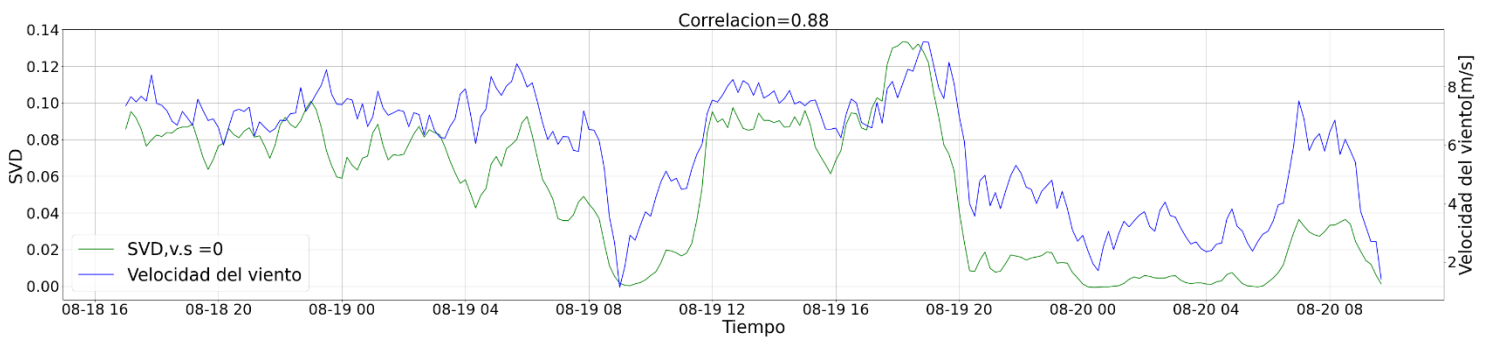


Figura 5.21: Gráfico comparación SVD y velocidad del viento. Ventana 3.

Luego, en la Figura 5.19, Figura 5.20, Figura 5.21 se muestra los resultados de comparación entre el modo 1 y la variable de velocidad, donde claramente se observa la similitud entre las curvas. A pesar de no ser totalmente idénticas en el marco general se observa una clara similitud. Esto se ve reflejado en el alto grado de correlación de las curvas.

5.1.3.4 POD 1

Se muestra el modo 1 el cual corresponde al promedio de generación de energía de modo de estudio para futuro análisis de cuál es el comportamiento de las turbinas con respecto a su generación de energía.

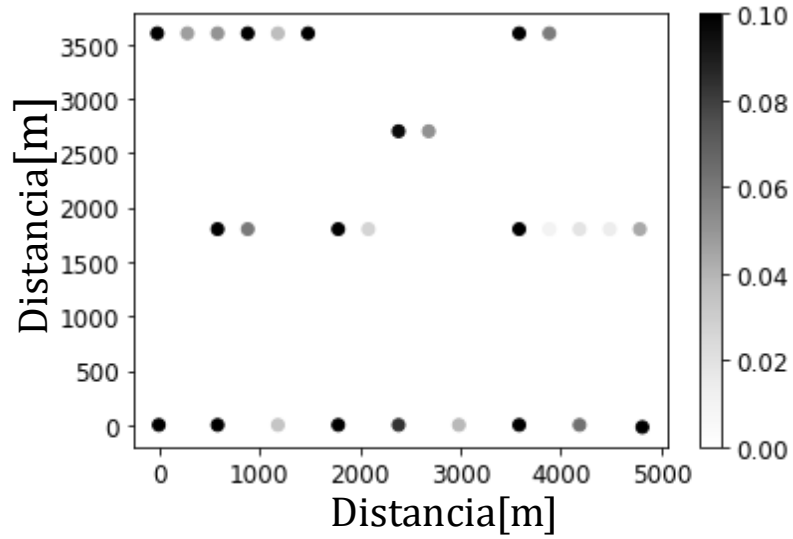


Figura 5.22: POD 1 espacial ventana 1, correspondiente al aporte de potencia de cada aerogenerador a al sistema.

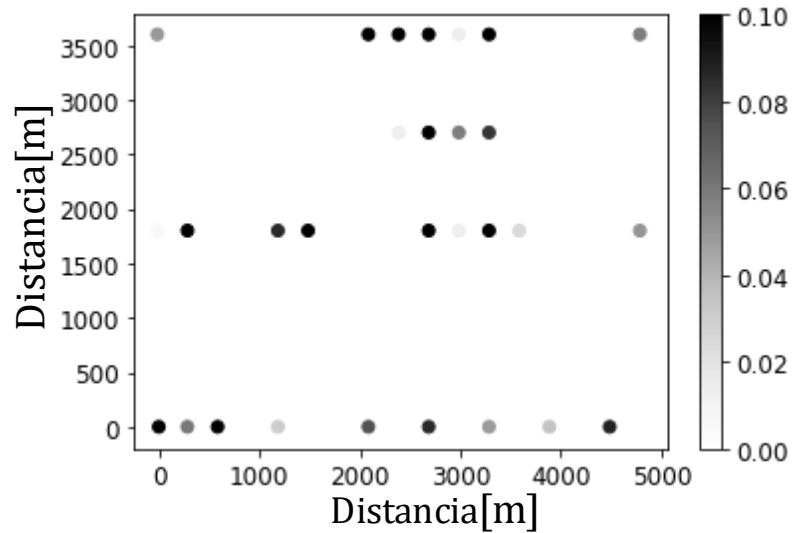


Figura 5.23: POD 1 espacial ventana 2, correspondiente al aporte de potencia de cada aerogenerador a al sistema.

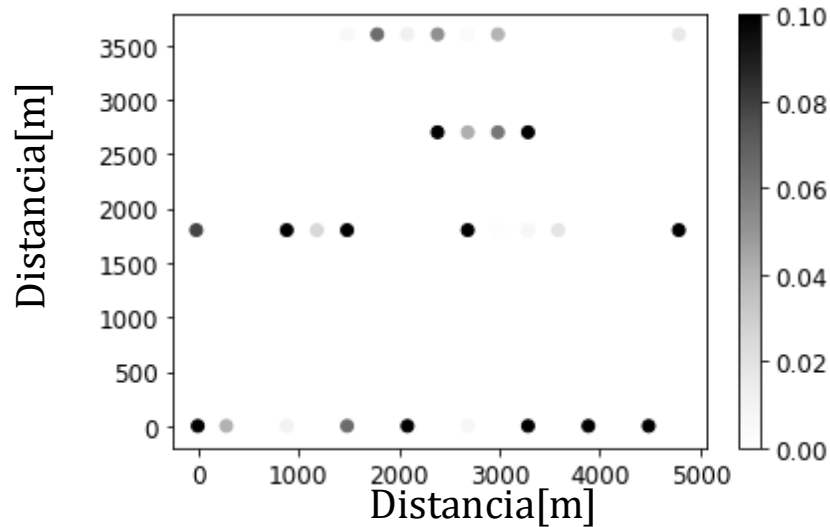


Figura 5.24: POD 1 espacial ventana 3, correspondiente al aporte de potencia de cada aerogenerador a al sistema.

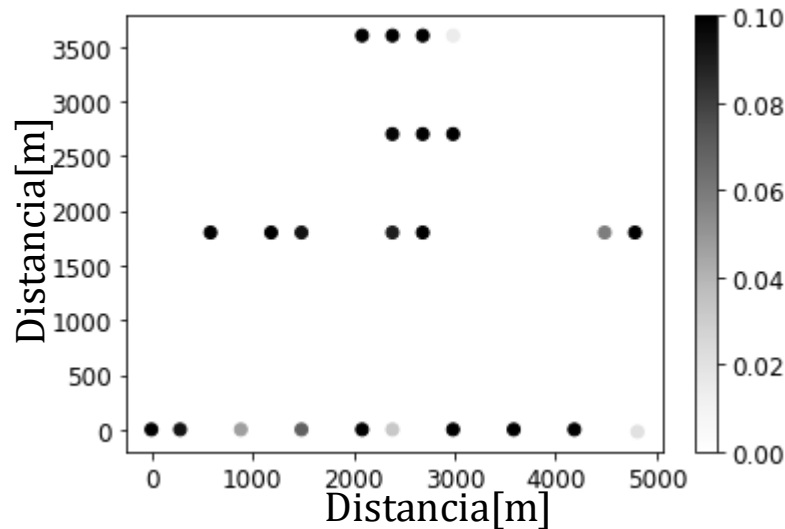


Figura 5.25: POD 1 espacial ventana 4, correspondiente al aporte de potencia de cada aerogenerador a al sistema.

Luego, se muestran en la Figura 5.22, Figura 5.23, Figura 5.24, Figura 5.25 el POD 1 para las 4 ventanas, que corresponde al modo 1 de la SVD de la generación de potencia del parque eólico y cada uno de estos modos muestra la distribución espacial de potencia del parque eólico y cuál es el aporte de potencia de cada generador al sistema completo. Al analizar esto se quiere encontrar algún tipo de patrón similar entre los 4 modos, ya que esto indicaría un comportamiento similar a pesar de las condiciones externas. En primer lugar, se observa que en la primera fila donde $y=0$ los generadores se activan de manera intercalada, por lo que uno genera y luego el de su lado no y así sucesivamente, ocurre en todas las ventanas, aunque cabe destacar que algunas ventanas este fenómeno es invertido.

5.1.4 DMD

Luego se aplica el algoritmo de DMD a la matriz de potencia.

Se muestra el primer modo 1 corresponde al promedio de potencia el cuál es el más importante para permite analizar la generación de potencia por turbina. En conjunto al modo 1 se muestra la variación temporal del modo.

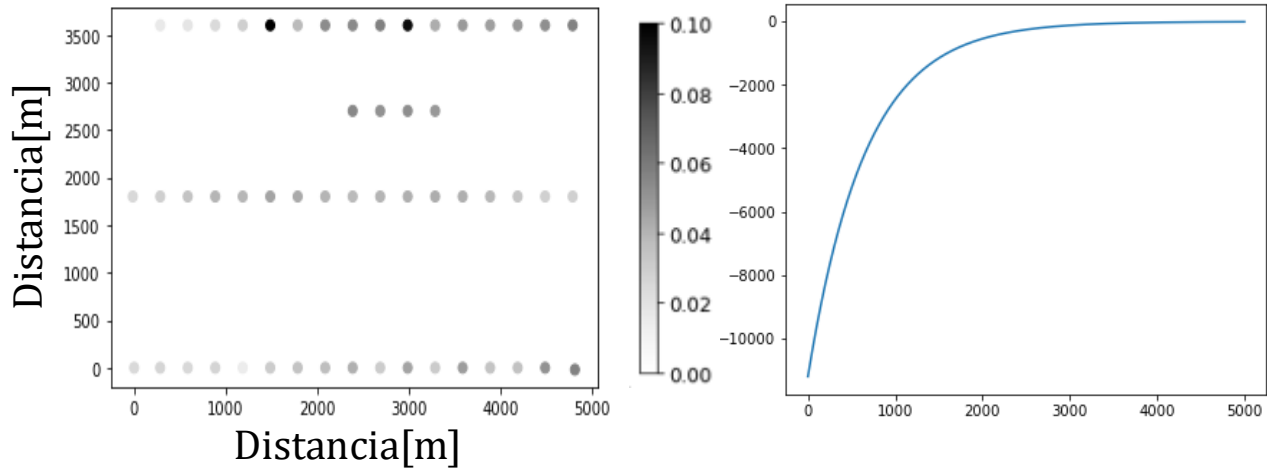


Figura 5.26: Grafico espacial y temporal para el modo 1. Ventana 1.

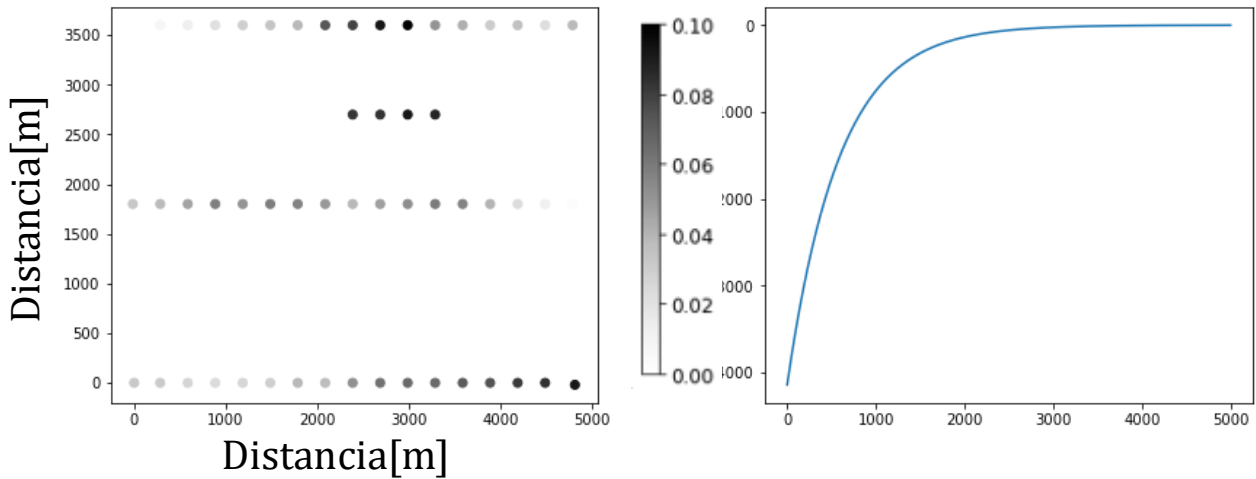


Figura 5.27: Grafico espacial y temporal para el modo 1. Ventana 2.

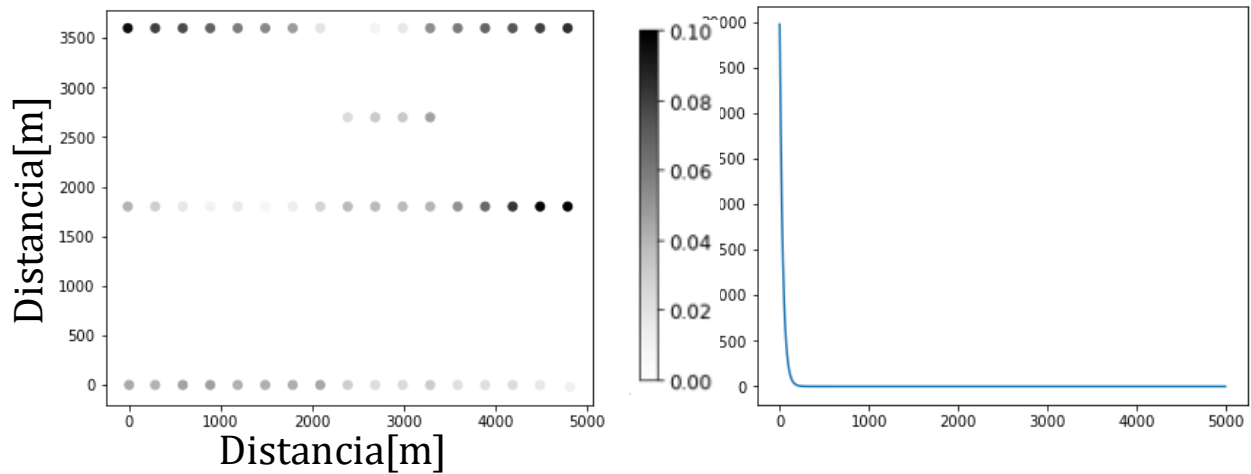


Figura 5.28: Grafico espacial y temporal para el modo 1. Ventana 3.

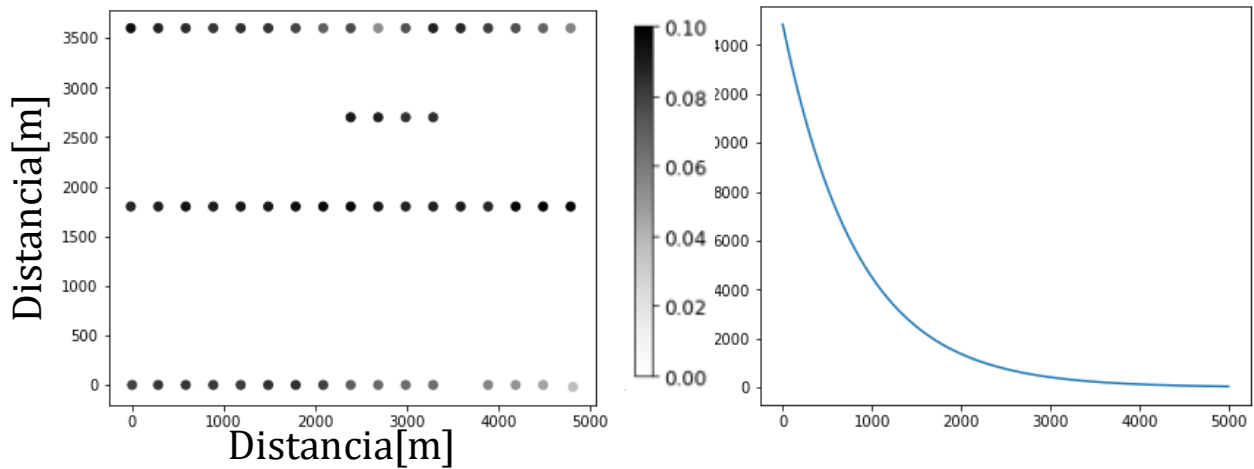


Figura 5.29: Grafico espacial y temporal para el modo 1. Ventana 4.

Luego se observan los resultados de la aplicación de DMD a la matriz de potencia, en donde se muestra el grafico temporal (curva azul) de la predicción del DMD y se muestra la distribución espacial para el modo 1. Con DMD se esperaba encontrar algún tipo de patrón de baja dimensionalidad y así predecir. Los resultados fueron los esperados, ya que no se toma en cuenta el efecto de las variables externas e, lo que conlleva a que la el grafico temporal termine estabilizándose. Una alternativa para mejorar resultados es utilizar DMDc (DMD con control), que permite tener una retroalimentación de las variables externas.

5.1.5 Regresión

Se muestran los resultados de la regresión, donde se muestra el aporte de cada variable a los 8 modos truncados. Estos son los coeficientes gammas encontrados en la ecuación (24) representados en forma de tabla. Luego se observa la calidad de la predicción entre ventanas mediante la medición de errores.

Ventana 1	Modo							
	1	2	3	4	5	6	7	8
Velocidad	26407,7	1374,0	2643,6	632,0	410,6	64,6	181,6	71,5
Dirección	11481,0	6767,7	4435,3	4828,9	60,0	517,4	467,3	186,0
Temperatura	6215,7	3247,2	304,0	1448,4	318,1	1288,9	100,1	230,4
Presión	7976,5	1836,6	1773,8	2810,5	89,4	556,6	362,5	66,3
Hora del día	155,7	215,8	511,7	16,3	161,3	73,1	82,7	47,5

Tabla 5.3 : Tabla resumen de resultados entre modos y variables, para la ventana 1.

Ventana 2	Modo							
	1	2	3	4	5	6	7	8
Velocidad	24441,0	1221,6	2076,4	280,3	420,3	153,7	20,0	7,4
Dirección	8221,4	5899,9	99,2	2077,6	1305,7	1351,2	847,6	1037,7
Temperatura	11593,3	4911,2	759,1	527,0	503,0	80,2	591,4	397,5
Presión	1417,8	1178,2	1927,6	1689,3	807,0	779,2	145,2	403,0
Hora del día	1213,3	307,7	101,2	225,3	124,8	145,2	8,8	16,4

Tabla 5.4 : Tabla resumen de resultados entre modos y variables, para la ventana 2.

Ventana 3	Modo							
	1	2	3	4	5	6	7	8
Velocidad	16926,3	412,9	844,3	488,2	108,1	109,9	90,6	155,9
Dirección	6849,4	1463,3	3869,9	498,5	107,1	426,5	318,6	1104,3
Temperatura	1243,5	845,9	1181,9	1943,0	251,5	611,9	199,3	1238,0
Presión	9167,1	729,0	3389,1	1389,0	210,5	37,2	40,3	191,4
Hora del día	274,4	343,5	19,5	245,1	89,3	31,9	29,4	28,2

Tabla 5.5 : Tabla resumen de resultados entre modos y variables, para la ventana 3.

Ventana 4	Modo							
	1	2	3	4	5	6	7	8
Velocidad	10062,0	1140,4	2826,2	1569,3	2125,5	1172,6	123,2	62,4
Dirección	2544,9	5,2	1108,1	2302,4	1,9	795,3	355,6	765,2
Temperatura	4813,9	6662,7	1292,8	206,8	1454,9	61,1	168,7	451,4
Presión	8232,2	5514,8	3741,3	2733,6	409,8	111,3	320,8	348,0
Hora del día	183,7	189,9	187,6	25,8	77,7	125,9	114,6	24,5

Tabla 5.6 : Tabla resumen de resultados entre modos y variables, para la ventana 4.

Al analizar los resultados de la regresión de todas las ventanas se confirma nuevamente que la velocidad es la variable que más peso tienen sobre el modo 1, por lo que, se reafirma que es la variable de mayor relevancia para el estudio de la potencia del parque eólico. El segundo modo es más incierto, indicando que la dirección del viento es la variable que más aporta. Para el tercer modo nos indica presión o dirección del viento como variables importantes. Ya para el cuarto modo nos indica la dirección del viento o la potencia. A pesar de, que la regresión no indica concordancia con los modos anteriormente estudiados mediante correlación en la Tabla 5.2, las variables importantes si son las 4 mencionadas: Velocidad, Dirección, Temperatura y Presión, en donde la hora del día no afecta en gran magnitud a los modos.



Figura 5.30: Comparación curvas predicción de potencia total y real entre el set entrenamiento ventana 1 y el set verificación ventana 2.

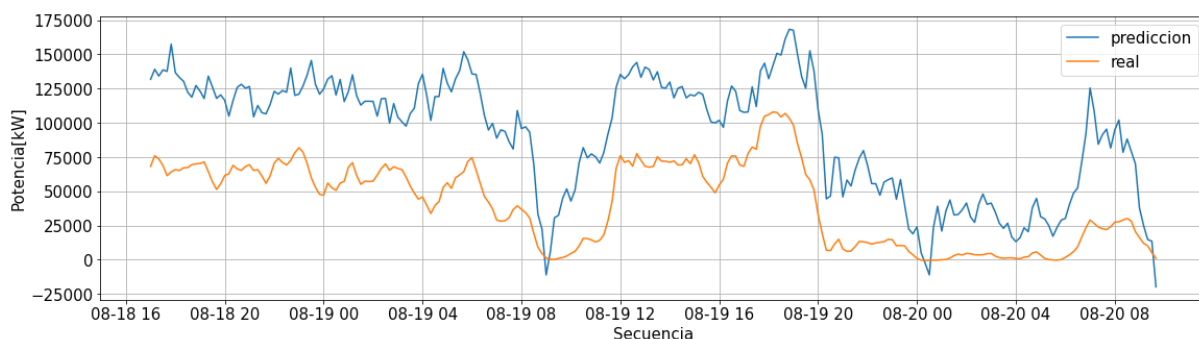


Figura 5.31: Comparación curvas predicción de potencia total y real entre el set entrenamiento ventana 1 y el set verificación ventana 3.

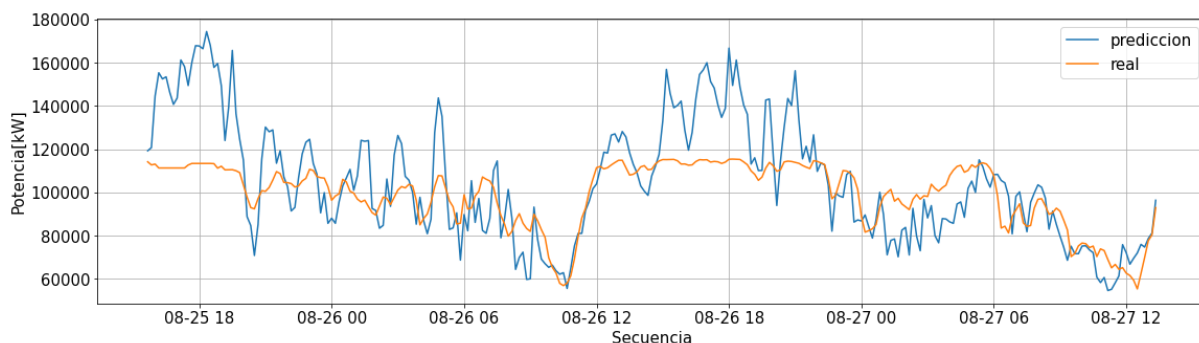


Figura 5.32: Comparación curvas predicción de potencia total y real entre el set entrenamiento ventana 1 y el set verificación ventana 4.

Al analizar los resultados de las curvas de comparación entre datos reales y de predicción, tomando en cuenta que solo se están utilizando 8 modos de los 55, se da cuenta que la regresión logra predecir la dinámica temporal de la potencia a grandes rasgos, esto quiere decir que en el largo plazo la regresión predice la variación de potencial, como se observa en las Figura 5.30, Figura 5.31, Figura 5.32. Por otro lado, la regresión no logra captar dos factores importantes, en primer lugar hay grandes desfases en magnitud entre los datos reales y los predichos como se observa en la Figura 5.31, por lo que el error es alto al ser magnitudes grandes, en segundo lugar, en el corto plazo, la regresión no percibe con exactitud los cambios de potencia, como se observa en la Figura 5.30, hay un lapso en donde la potencia se mantiene relativamente estable para la curva normal, pero en la predicción, los cambios se mantienen en ese promedio con un comportamiento brusco, esto se puede traducir a que la desviación estándar es muy alta, pero logra predecir el promedio y mantenerse.

Una de las causas de porque la predicción de potencia total del parque eólico no fue tan exacta puede ser lo básico del algoritmo y lo complejo del sistema. Además, puede ser que al ser tan complejo el sistema, uno de los modos no se reconstruye de manera correcta, lo que afecta la predicción del sistema. De todas maneras, la predicción mejora cuando se toma la potencia total del parque en vez de estimar la potencia individualmente de cada aerogenerador.

Las siguientes figuras muestran la comparación de la predicción para el modo 1 de la SVD, el cual corresponde al promedio de velocidad del parque eólico.

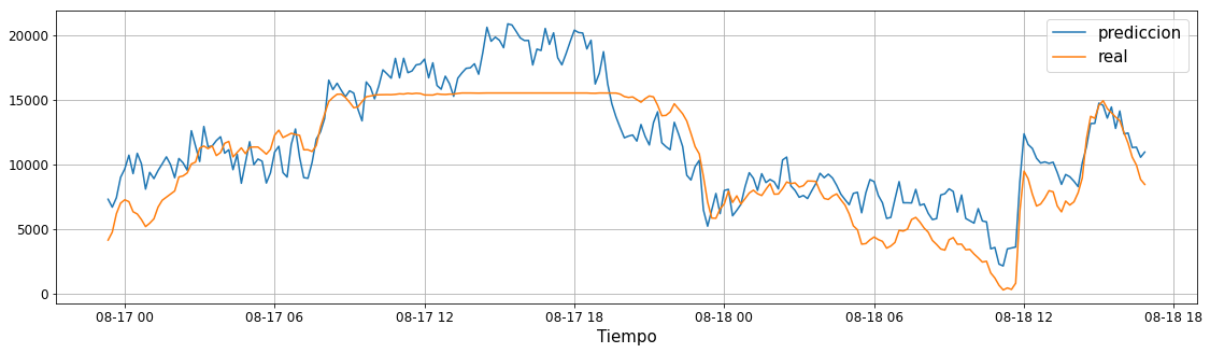


Figura 5.33: Comparación de predicción para el modo 1 entre la ventana 1 y la ventana 2.

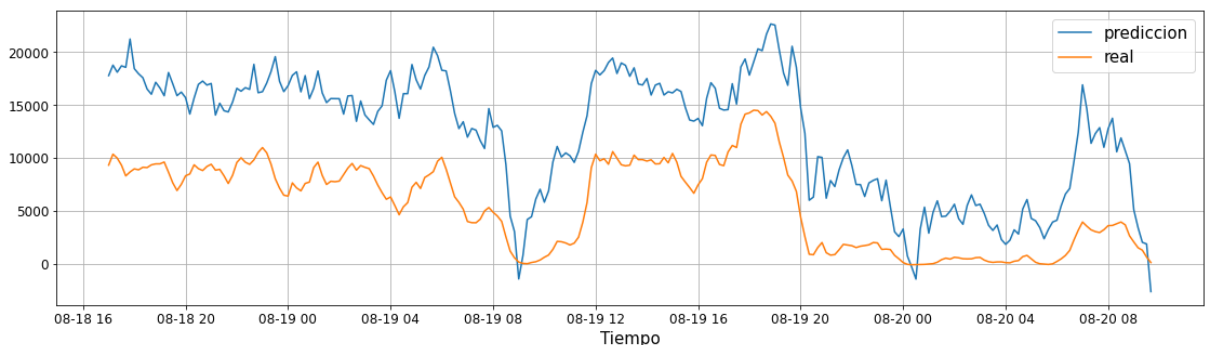


Figura 5.34: Comparación de predicción para el modo 1 entre la ventana 1 y la ventana 3.

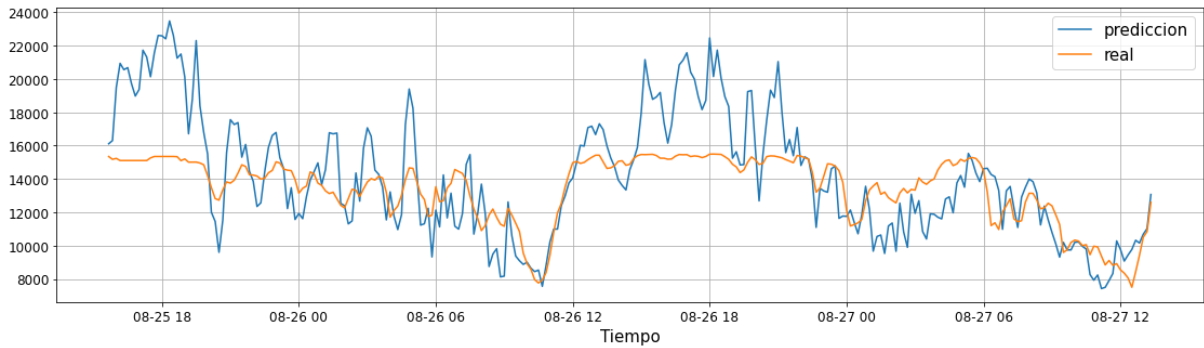


Figura 5.35: Comparación de predicción para el modo 1 entre la ventana 1 y la ventana 4.

Al comparar las Figura 5.33, Figura 5.34, Figura 5.35 con las Figura 5.30, Figura 5.31, Figura 5.32 se observa una gran similitud entre los modos 1 de predicción y los gráficos de predicción que contienen todos los modos, lo que nos entrega una idea mucho más clara que el modo 1 es efectivamente la variable más importante para la generación de potencia.

La Tabla 5.7 muestra los errores porcentuales entre la ventana de entrenamiento y la ventana de verificación. Cabe recordar que la ventana 1 corresponde a la ventana de entrenamiento y las ventanas 2, 3 y 4 a la ventana de verificación.

	Error Porcentual		
	Modo 1	Modo 2	Modo 3
Ventana 2	36.5	415.9	318.6
Ventana 3	1224	1099	993
Ventana 4	14.4	167.9	297.4

Tabla 5.7: Errores porcentuales entre ventana 1 y las ventana 2, ventana 3 y ventana 4.

En primer lugar, cabe recalcar que la ventana 3 es un caso especial, en donde falla el supuesto que los modos serian similares entre las ventanas, ya que claramente el resultado se encuentra alterado, es por esto, por lo que no se tomara en cuenta para compararlo con otras ventanas.

La predicción en el modo 1 para la ventana 2 y ventana 4 es de menor error, teniendo en cuenta que el este modo es el cual lleva la mayor cantidad de aporte para la predicción de potencia, se puede inferir que la predicción para ambas ventanas será de mejor calidad.

A medida que vamos avanzando en los modos el error se dispara hasta 10 veces el valor del modo 1, lo que indica que no hay suficiente información para representar de manera correcta los modos que no sean el modo 1. Esto puede deberse a que los modos sobrepasados el modo 1 se vuelven más complejos por lo que la cantidad de variables no es suficientemente alta para reconstruirlo. Otra causa puede ser que, a medida que vamos avanzando en los modos, este se

vuelve más representativo solo para la ventana en la cual se está estudiando, por lo que el modo 2 empieza a describir solo ese espacio de tiempo y el modo 3 lo mismo, pero aún más específico, por lo que es difícil que ocurra el mismo comportamiento en dos ventanas diferentes.

Una forma de reparar esto es aumentando la cantidad de variables medidas en el parque eólico, de esta manera puede ser que sea posible describir una mayor cantidad de modos. Otra solución también puede ser aumentar el lapso de tiempo estudiado y realizar un clustering, de esta manera los modos 2 y 3 podrían ser representativos para un sistema más general, ya que no se estarían tomando una densidad tan alta de puntos, si no que se podría buscar un equilibrio para encontrar la cantidad de puntos necesaria para encontrar la esencia del comportamiento del parque eólico.

6. Conclusión

Es posible concluir, que en primera instancia el objetivo principal se logró, ya que se logra estudiar fenómenos fluidodinámicos del sistema a gran escala, como, por ejemplo, la variación de potencia de noche y día. Por otro lado, fenómenos a menor escala no lograron ser captados por ninguno de los algoritmos. También, fenómenos fluidodinámicos tal como la interferencia de estela de un aerogenerador sobre otro no es tan clara.

A continuación del punto anterior, se logra predecir la potencia del sistema a gran escala, tal como se observa en el análisis, por lo que se logra predecir la potencia total del sistema, pero el algoritmo falla para fluctuaciones a menor escala.

Cabe destacar que, se logran completar los objetivos específicos, que de todas maneras nos entregan gran información sobre el sistema. Mediante la aplicación de SVD realizando la disminución de dimensionalidad, nos ayuda a captar cuales son las variables que afectan de manera incidente al parque eólico. También nos entrega un esbozo de cómo se comporta la dinámica temporal y espacial del parque, tal como que existen fluctuaciones día noche en generación de potencia pero que no afecta en gran magnitud la predicción. Por otro lado, la aplicación de correlación nos permite esbozar fenómenos fluidodinámicos del sistema, tal como la alta correlación entre temperatura/humedad relativa y presión/densidad, lo que nos indica que, entre esos pares de variables, si es que una varía la otra variable también se ve afectada. Finalmente, la aplicación de la regresión permite obtener la predicción de potencia del parque eólico.

Además, se logra realizar un modelo para el parque eólico mediante la aplicación de la regresión, correlación y SVD, ya que, se descubren las variables importantes del parque, para luego obtener los modos y así, descubrir un modelo que se adapte a los modos mediante regresión.

También se cumple de manera parcial el análisis de fenómenos fluidodinámicos del sistema mediante la aplicación de SVD y DMD, donde se logran analizar a gran escala unos pocos fenómenos fluidodinámicos, tal como la estela que deja un generador en los modos de SVD. También, una especie de fenómeno donde los generadores estaban intercalados en la generación de potencia.

También es importante mencionar que se encontraron desafíos al procesar datos reales de un parque eólico. Tal como, la existencia de ruido en las mediciones lo que puede afectar los resultados. También, existían errores en la captación de datos, por lo que existían espacios vacíos en momentos del día. Lo anterior se manejó para trabajar con datos más limpios y así conseguir los mejores resultados.

Como se mencionó en un inicio, el parque eólico es un sistema dinámico y complejo, es por eso por lo que siempre se ha buscado los métodos más completos para estudiar la dinámica de estos sistemas. Los algoritmos utilizados en este trabajo corresponden a su versión más básica, por lo que queda espacio para que futuras generaciones tomen este informe como guía para aplicar una

versión más avanzada de DMD y SVD, tal como, DMDc (Brunton & Kutz, 2019) (con control) y la variante robusta de la SVD (Brunton & Kutz, 2019) para descartar outliers, y de esta manera encontrar fenómenos fluidodinámicos más específicos dentro del parque.

Es posible que exista sesgo en los datos debido a la alta velocidad del viento, lo que generaba que la potencia muchas veces se saturaba. Por lo que, otra recomendación es realizar estudios para otras ventanas donde el viento tenga menor magnitud, preferiblemente ente 3m/s a 8m/s, para captar mejor la fluidodinámica del sistema. Con este ajuste se espera disminuir un poco el impacto de la velocidad como variable y así permitir salir a flote las otras variables. Se espera que en la SVD las variables más importantes estén más cercanas, además que, en el estudio de los modos, la velocidad no sea la variable más importante por una diferencia muy alta.

Sumado a eso, se propone variar el tamaño de las ventanas en conjunto con realizar un clustering en los datos, ya que de esta manera se puede estudiar la fluidodinámica para ventanas más largas o cortas de tiempo y el clustering para disminuir el ruido de los datos.

7. Bibliografía

- Brunton, S. L., & Kutz, J. N. (2019). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.
- Calaf, M., Meneveau, C., & Meyers, J. (2010). Large eddy simulation study of fully developed wind-turbine array boundary layers. *Physics of fluids*, *22*, 015110.
- Gebraad, P. M., Teeuwisse, F. W., Van Wingerden, J. W., Fleming, P. A., Ruben, S. D., Marden, J. R., & Pao, L. Y. (2016). Wind plant power optimization through yaw control using a parametric model for wake effects—a CFD simulation study. *Wind Energy*, *19*, 95–114.
- Hemati, M. S., Williams, M. O., & Rowley, C. W. (2014). Dynamic mode decomposition for large and streaming datasets. *Physics of Fluids*, *26*, 111701.
- Holmes, P., Lumley, J. L., Berkooz, G., & Rowley, C. W. (2012). *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press.
- Lumley, J. L. (1970). *Stochastic Tools in Turbulence*, Academic Press, New York.
- Martinez, L. A., Meneveau, C., & Stevens, R. (2016). Wind farm large-eddy simulations on very coarse grid resolutions using an actuator line model. *34th Wind Energy Symposium*, (pág. 1261).
- Martínez-Tossas, L. A., Churchfield, M. J., & Meneveau, C. (2015). Large eddy simulation of wind turbine wakes: detailed comparisons of two codes focusing on effects of numerics and subgrid modeling. *Journal of Physics: Conference Series*, *625*, pág. 012024.
- Meyers, J., & Meneveau, C. (2010). Large eddy simulations of large wind-turbine arrays in the atmospheric boundary layer. *48th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition*, (pág. 827).
- Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, *656*, 5–28.
- Shapiro, C. R., Bauweraerts, P., Meyers, J., Meneveau, C., & Gayme, D. F. (2017). Model-based receding horizon control of wind farms for secondary frequency regulation. *Wind Energy*, *20*, 1261–1275.
- Shapiro, C. R., Starke, G. M., Meneveau, C., & Gayme, D. F. (2019). A wake modeling paradigm for wind farm design and control. *Energies*, *12*, 2956.
- Stevens, R. J., Gayme, D. F., & Meneveau, C. (2016). Generalized coupled wake boundary layer model: applications and comparisons with field and LES data for two wind farms. *Wind Energy*, *19*, 2023–2040.