

Pronósticos de viento Bayesianos y su impacto en el problema de Pre-despacho

Entrega Final

Samuel Córdova, Christian Meléndez

Santiago, Chile

Abstract

Debido a la naturaleza variable e incierta del viento, la integración a gran escala de generación eólica plantea desafíos significativos para la operación de los sistemas de potencia modernos. Uno de los principales desafíos operacionales es el conocido problema de Pre-despacho bajo incertidumbre, el cual consiste en determinar que generadores encender y apagar en las horas siguientes considerando una serie de posibles pronósticos de generación renovable. Si bien una serie de modelos han sido diseñados para resolver el problema de Pre-despacho, el desempeño de estos depende fuertemente de la calidad de los pronósticos, haciendo necesario la elaboración de pronósticos de la mayor información y calidad posible. Motivado por esto, el presente trabajo apunta a desarrollar un modelo de pronóstico que: i) capture las dependencias espacio-temporales y estacionalidades propias del viento, ii) genere pronósticos puntuales y de intervalo para cuantificar el riesgo en la predicción, iii) incorpore en la construcción de pronósticos la incertidumbre asociada a la estimación de parámetros del mismo modelo. Los pronósticos generados se basan en un modelo Periódico Vectorial Autoregresivo bajo un enfoque Bayesiano, en el cual la incertidumbre de los parámetros es capturada a través de su función de *likelihood*, y las integrales necesarias a estimar para los pronósticos son obtenidas mediante técnicas de Monte Carlo de dos etapas o *two-stage Monte Carlo*.

Keywords: Predicción Bayesiana, Sistemas de Potencia, Pre-despacho, Periódico Vectorial Autogresivo

1. Introducción

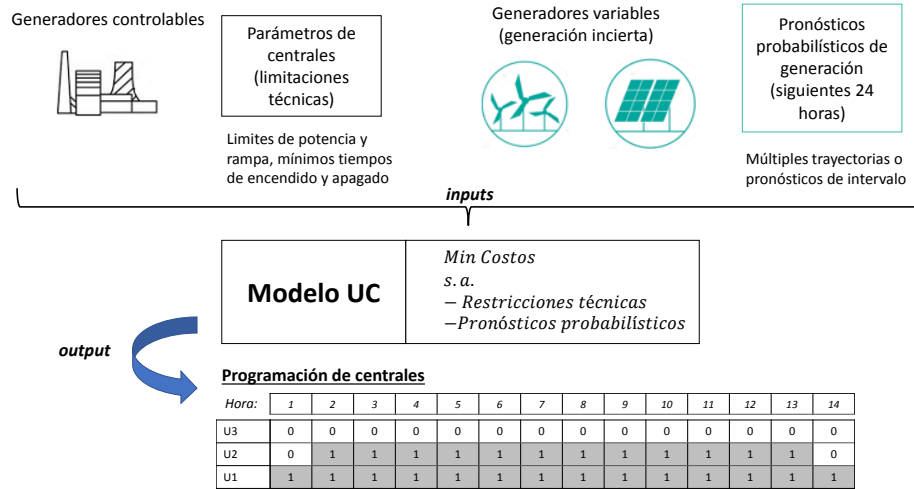
El problema de Pre-despacho (o UC por sus siglas en inglés) es uno de los problemas más desafiantes en la operación de sistemas eléctricos. Este problema consiste en determinar el cronograma de generadores controlables para un cierto horizonte de tiempo de manera de garantizar el suministro de energía al menor costo operacional posible, y teniendo en cuenta las limitaciones impuestas por las características físicas de las unidades generadoras. Para sistemas de potencia tradicionales, caracterizados por una baja o nula participación de generación renovable, el UC ya ha sido extensivamente estudiado [1]. Sin embargo, en los últimos años se ha evidenciado un fuerte crecimiento de las fuentes de generación renovable a nivel mundial, lo cual ha traído nuevos desafíos para la operación de sistemas de potencia [2]. Debido a la incertidumbre y variabilidad de la generación renovable, los sistemas de potencia modernos con participación renovable significativa deben ser capaces de responder a rápidas y fuertes variaciones en la producción de energía renovable, haciendo de la flexibilidad en sistemas eléctricos un tema de estudio relevante en los últimos años. Bajo esta línea, se han desarrollado una serie de modelos de UC para el uso eficiente de recursos flexibles en la operación de sistemas eléctricos altamente renovables, los cuales están basados principalmente en modelos de optimización estocástica y robusta (ver [3] para una revisión exhaustiva). Dichos modelos en general consideran como *input* pronósticos **probabilísticos** de generación eólica, los cuales pueden tomar las siguientes formas [3]:

1. **Múltiples trayectorias:** Se generan múltiples posibles pronósticos de generación eólica a partir de los cuales se determina la programación de centrales en que minimiza el costo esperado de operación. De esta forma, el modelo de UC resuelve el problema a través del método de Sample Average Approximation (SAA).
2. **Pronósticos de intervalo:** A diferencia del caso anterior, la información de los pronósticos es condensada y entregada como un intervalo para el cual se espera que estará contenida la generación renovable. De esta forma, la programación de centrales obtenida será válida para cualquier realización de viento contenida dentro del intervalo generado. Este tipo de modelos entra en la categoría de modelos de Optimización Robusta y Optimización de Intervalos.

En la Figura 1 se ilustra el problema de UC tradicional en donde se utilizan los pronósticos probabilísticos para determinar la programación de

centrales para las próximas 14 horas. Se puede observar que en base a los parámetros de generadores y pronósticos entregados, el cronograma de las unidades U1, U2, y U3 es determinado para los horas $t = 1, \dots, 14$. En la misma Tabla de la Figura 1 se puede observar el detalle de dicho cronograma: la unidad U1 es encendida durante las siguientes 14 horas, la unidad U2 es encendida para $t = 2, \dots, 13$, y la unidad U3 nunca es encendida debido a posibles restricciones operacionales o a un alto costo de generación.

Figure 1: Ilustración del problema UC para la determinación de la programación de centrales para los próximas 14 horas. El modelo de UC determina la programación óptima en base a los pronósticos probabilísticos entregados y los parámetros de las centrales controlables



Un aspecto importante a destacar, es que dado que los modelos de UC se limitan a recibir como input los pronósticos probabilísticos, en la práctica su desempeño final dependerá de la calidad de los pronósticos [4]. De esta forma, el desarrollo de modelos de pronóstico de viento de alta calidad se vuelve sumamente relevante. En esta línea una serie modelos de pronóstico pueden ser encontrados en la literatura, incluidos modelos físicos y estadísticos (ver [5] para una revisión exhaustiva). En esta investigación el foco estará en modelos estadísticos debido a su simplicidad y eficiencia computacional [6], junto a su amplia aplicación en sistemas de potencia, para la cual ya se ha evidenciado sus beneficios en forma de reducción de costos operacionales [7, 8].

El resto del informe esta organizado de la siguiente manera. La sección 2 presenta la metodología propuesta para desarrollar un modelo de pronósticos adecuado, así como los pasos necesarios para la incorporación de dichos pronósticos en la planificación de las centrales. En la sección 3 distintas métricas para evaluar el desempeño de pronósticos eólicos son presentadas para su posterior uso, y la base de datos utilizada en la investigación es descrita. Finalmente, en la sección 4 se indica los avances realizados hasta la fecha y el trabajo futuro a realizar para la entrega final.

2. Metodología propuesta

Para capturar la interacción entre los pronósticos de viento y la planificación de centrales una serie de componentes tienen que ser modeladas. Primero, es necesario construir el modelo de pronósticos de viento a utilizar, el cual es descrito en la sección 2.1. Segundo, es necesario convertir los pronósticos de viento en un solo pronóstico de generación eólica total, a partir de la cual el sistema será posteriormente planificado. El método utilizado para realizar dicha conversión es presentado en la sección 2.2. Tercero, dado que en la práctica trabajar con un número indefinido de escenarios de generación eólica se puede volver inmanejable computacionalmente, es necesario condensar la información, para lo cual desarrollamos en la sección 2.3 un método para la construcción de intervalos en los cuales se estima que la generación eólica estará contenida. Finalmente, a partir de los intervalos de generación es posible determinar un cronograma de centrales que con cierto nivel de seguridad garantice la operación del sistema eléctrico a costo mínimo. El modelo de Pre-despacho utilizado en la determinación del cronograma es descrito en la sección 2.4.

2.1. Pronósticos de viento y enfoque Bayesiano

Actualmente existe una vasta literatura enfocada en caracterizar estadísticamente el recurso eólico. En dichos estudios, se ha observado que en general el viento puede ser modelado como una serie de tiempo con las dos principales siguientes características [4]:

- **Dependencia espacio-temporal:** Fuertes correlaciones tanto en el aspecto temporal como geográfico han sido observados para el recurso eólico. Debido a esto, se hace necesario entonces la utilización de modelos que incorporen correlaciones-cruzadas tal como el modelo Vectorial Autoregresivo (VAR).

- **Estacionalidades:** Es posible identificar patrones diarios, así como anuales (estaciones del año) en las series de tiempo de viento. En general, dichos patrones son modelados a través de una componente determinística comúnmente llamada *tendencia determinística periódica*.

Una de las más recientes propuestas aplicadas a sistemas de potencia consiste en [4], en donde se plantea la utilización de un modelo *Periódico Vectorial Autoregresivo* (PVAR) para la construcción de pronósticos probabilísticos. A modo general, un modelo PVAR corresponde a un modelo VAR con coeficientes variantes en el tiempo de forma periódica [9]:

$$\begin{aligned}\varepsilon(t) &= A_1(t)\varepsilon(t-1) + \cdots + A_k(t)\varepsilon(t-p) + \omega(t) \\ \omega(t) &\sim N(0, \Sigma(t)),\end{aligned}$$

en donde el vector $\varepsilon(t)$ representa la velocidad del viento en el instante t para los distintos puntos geográficos donde están instaladas las centrales. Del modelo, el proceso $\varepsilon(t)$ es resultado de una combinación lineal de sus k valores previos, de acuerdo a las matrices de coeficientes $A_j(t)$, con $j = 1 \dots k$, y un ruido blanco normal $\omega(t)$ con covarianza $\Sigma(t)$. Nótese que $\varepsilon(t) = r(t) - T(t)$ resulta de restar la tendencia determinística periódica $T(t)$ del recurso eólico $r(t)$. Al igual que en [4], consideraremos que $T(t)$ depende de si el instante t está asociado a la hora h del día y mes m del año. De forma similar, las matrices $A_j(t)$ y $\Sigma(t)$ dependerán si t corresponde al mes m del año.

La principal ventaja del modelo anterior está en el hecho de que este captura las estacionalidades y la dependencia espacio-temporal del recurso eólico de manera simple. Sin embargo, surge la pregunta de como estimar correctamente los parámetros $A_j(t)$ y $\Sigma(t)$. Originalmente, en [4] la estimación de dichos parámetros fue realizada mediante el método de *maximum likelihood estimation* (MLE). Si bien dicho método mostró resultados favorables en la reducción de costos operacionales, la estimación mediante MLE puede resultar limitante, dado que no reconoce la incertidumbre en los parámetros estimados.

En la presente investigación, proponemos extender el modelo PVAR de [4], tomando ahora un enfoque Bayesiano en el sentido de reconocer la incertidumbre de los parámetros estimados. De esta forma, la propuesta radica en construir una función de likelihood para los parámetros $A_j(t)$ y $\Sigma(t)$ del modelo PVAR y construir pronósticos de mayor precisión gracias a esta extensión.

2.2. Conversión a pronósticos de generación eólica total

Luego de definir el modelo para caracterizar la velocidad del viento $\varepsilon(t)$, es necesario calcular la generación eólica total que recibe el sistema eléctrico. Para esto, es necesario aplicar el concepto de *curva de potencia*, la cual se definen como una función $f(\cdot)$ tal que:

$$p(t) = f(\varepsilon(t))$$

en donde el vector $p(t)$ indica la generación eólica para la velocidad del viento $\varepsilon(t)$. En base a lo anterior, podemos calcular la generación eólica total $W(t)$ como:

$$W(t) = \sum_{g \in G_w} p_g(t) = \sum_{g \in G_w} f(\varepsilon_g(t))$$

donde G_w corresponde al conjunto de generadores eólicos. De esta forma, conociendo $f(\cdot)$, y caracterizando $\varepsilon(t)$ (modelo PVAR), podemos realizar pronósticos para la variable aleatoria $W(t)$. Para nuestro estudio, utilizaremos un modelo de 5 parámetros logístico (5PL) para $f(\cdot)$, el cual ha sido ampliamente utilizado en otros estudios de generación eólica [10].

$$f(\varepsilon) = d + \frac{a - d}{\left(1 + \left(\frac{u}{c}\right)^b\right)^g}$$

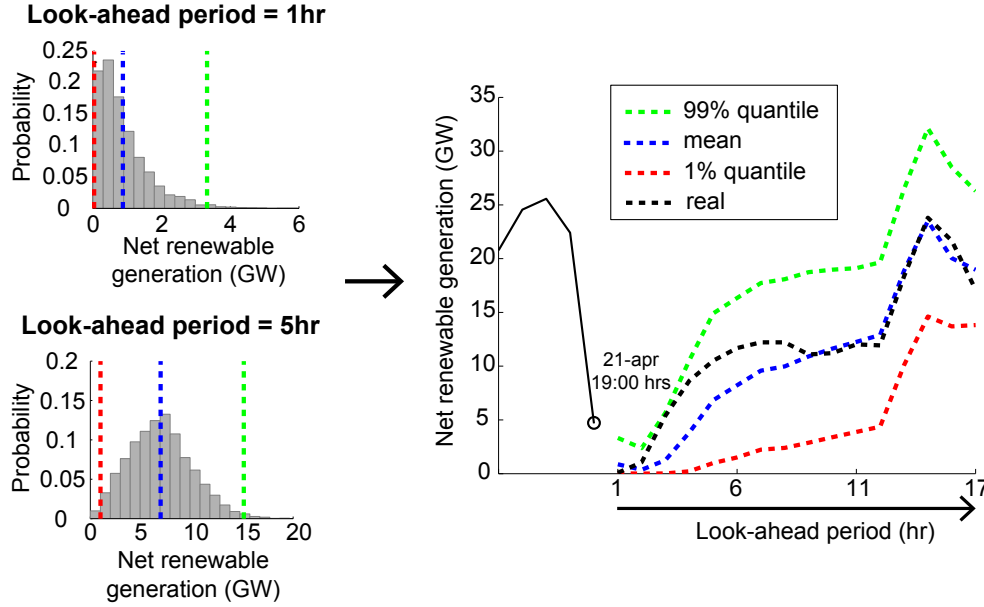
donde a, b, c, d, g son los parámetros de la 5PL. Dichos parámetros pueden ser obtenidos a través de un ajuste de curva 5PL a los datos entregados por la hoja técnica del fabricante. Durante nuestra investigación, consideramos la hoja técnica de la turbina eólica VestasV90 de 2MW disponible en <https://en.wind-turbine-models.com/turbines/16-vestas-v90>.

2.3. Construcción de pronósticos puntuales y de intervalo de generación eólica

Luego de haber definido la metodología para simular series de tiempo de velocidad de viento y convertirlas a generación total eólica, es posible generar una serie de trayectorias simuladas para resolver el problema de Pre-despacho. Si bien en teoría dichas trayectorias podrían ser directamente incorporadas en un modelo de optimización de Pre-despacho a través del método de Sample Average Approximation (SAA), esto puede fácilmente transformarse en un problema inmanejable computacionalmente [3]. Ante eso, es necesario implementar métodos alternativos de representar la incertidumbre simulada, tales como métodos de reducción de escenarios o construcción de intervalos. Para nuestra investigación, consideramos el enfoque

de [4], en donde intervalos de generación eólica son construidos en base a cierto nivel de seguridad deseado. De esta forma, el problema de Pre-despacho es resuelto para un cierto *intervalo de credibilidad* en donde se estima que las trayectorias de generación eólica estarán contenidas. En la Figura 2, se ilustra el método con el cual los intervalos son construidos.

Figure 2: Ilustración del método de construcción de intervalos de credibilidad en base a las simulación resultantes. Imagen extraída de [4]



De la Figura, es posible observar el método implementado. Primero, una distribución de probabilidad empírica es construida a partir de las trayectorias de generación eólica simuladas. Luego, a partir de dichas distribuciones de probabilidad, el valor esperado y los quantiles correspondientes son identificados, donde estos últimos dependerán del nivel de seguridad deseado (e.g., 95%). El valor esperado resultante corresponde al pronóstico puntual a luego ser utilizado en el modelo de Pre-despacho, mientras que los quantiles definen el rango del intervalo. Destacamos que a diferencia de [4], nuestros intervalos de credibilidad son construidos considerando la incertidumbre de los parámetros del modelo VAR, mientras que en el estudio original estos corresponden a valores únicos obtenidos a través de MLE.

Otro aspecto relevante a destacar consiste en el hecho de que de acuerdo a las ecuaciones definidas en la Sección 2.2, determinar de forma analítica los

Table 1: My caption

<i>Central</i>	<i>Costo variable</i>	<i>Restricción técnica</i>
Carbón	40 USD/MWh	Encendido/apagado lento. Su operación (on/off) debe ser determinada con anticipación y no puede ser luego ajustada
Diesel	100 USD/MWh	Altamente flexible. Puede ser ajustada (on/off) fácilmente en tiempo real
Eólica	<1 USD/MWh	No puede ser controlada (depende de la velocidad del viento)

valores esperados y quantiles de la generación eólica total $W(t)$ requeriría la resolución de integrales sumamente complejas, lo que justifica la necesidad de utilizar el método de Monte Carlo para nuestro caso.

2.4. Modelo de Pre-despacho

Tal como mencionamos anteriormente, los pronósticos probabilísticos resultante de nuestra metodología corresponde a *inputs* de un modelo de Pre-despacho, en el cual se determina la programación de centrales. Lo anterior, tiene como consecuencia que la calidad de los pronósticos influya directamente en el desempeño final (medido en costos operacionales y seguridad del sistema) del cronograma propuesto [4]. Para ilustrar esta idea, describimos uno de los modelos de Pre-despacho más básicos: *lista de mérito*.

Bajo este enfoque, tenemos que determinar la operación de las centrales con un cierto periodo de anticipación, el cual comúnmente corresponde a 1 día. De esta forma, al principio de cada día se deben determinar que unidades encender en las 24 horas siguientes en base a los pronósticos de viento, la demanda eléctrica futura (la cual asumimos conocida), las restricciones técnicas de las centrales y sus costos variables de generación. En la Tabla 1, ilustramos centrales 'tipo' características junto a sus restricciones técnicas y costos variables.

Notamos que para una cierta demanda eléctrica, las centrales Eólicas siempre serán despachadas primero dado su bajo costo variable. De esta forma, la demanda neta resultante (luego de restar la demanda original de

Table 2: Métricas propuestas para evaluar el desempeño de los pronósticos

Pronostico	Metrica
Puntual (valor esperado)	Error cuadrático medio
Intervalo (quantiles)	Cobertura condicional[11]

la generación renovable) debe ser satisfechas por las centrales térmicas de Carbón y Diesel. Acá, es donde las restricciones técnicas entran en juego, dado que naturalmente las centrales a Carbón serían las despachas a continuación, dado que su costo variable es menor a las de las centrales Diesel. Sin embargo, una central a Carbón es en general *poco flexible*, por lo que en caso de programar su operación (on/off) para las siguientes horas, esta no podrá ser luego ajustada. En contraste una Central Diesel, si puede ser ajustada en tiempo real, pero con el inconveniente de entregar energía al mayor costo variable. De esta forma, bajo el método de la lista de mérito la idea sería la siguiente: utilizar los pronósticos de intervalo para la generación eólica y determinar *cuántas* centrales a Carbón son necesarias para asegurar cierto nivel de seguridad. En caso que la generación eólica sea menor a la definida por el intervalo, se recurrirá a la generación costosa de las centrales Diesel. Resulta entonces claro, que para este ejemplo, la calidad de los pronósticos juega un rol clave, puesto que pronósticos de intervalo sobredimensionados llevarán a un encendido excesivo de centrales de carbón, mientras que intervalos subdimensionados llevarán a un exceso de uso de generación diesel de alto costo.

3. Métricas para la evaluación de desempeño y base de datos

Actualmente, existen una serie de métricas en la literatura para evaluar el desempeño de los pronósticos de viento. En la presente investigación, consideramos las métricas utilizadas en [4], las cuales son descritas en la Tabla 2. El error cuadrático medio será calculado a través de la media del cuadrado de la diferencia entre la potencia real generada y la potencia media estimada por nuestro modelo. En el caso de la Cobertura condicional, dado un pronóstico, calcularemos la proporción de puntos que quedan fuera del intervalo de confianza con respecto al total de puntos pronosticados. Esto nos dará una buena idea de qué tan buenos resultan los pronósticos realizados.

Respecto a la base de datos utilizada, esta fue obtenida a partir de la Campaña de Medición de Recurso Solar y Eólico del Ministerio de Energía

¹. Los datos corresponden a mediciones de la velocidad del viento a 20 y 10 metros de altura para distintas ubicaciones geográficas (ver Figura 2). Estas mediciones están hechas, en general, cada 10 minutos, sin embargo, existen espacios entre datos en que el intervalo de tiempo es mayor. Estas faltas de datos se manejarán mediante interpolación entre los tiempos que sí tienen datos.

Figure 3: Extracto de la base de datos

	Fecha Hora (YYYY-MM-DD HH:MM)	Velocidad de viento en 20.0 metros [mean,m/s]	Velocidad de viento en 10.0 metros [mean,m/s]
0	2009-07-12 00:10	10.7	10.1
1	2009-07-12 00:20	10.7	10.4
2	2009-07-12 00:30	10.2	9.9
3	2009-07-12 00:40	10.0	9.8
4	2009-07-12 00:50	10.2	10.2

Las bases de datos que consideraremos cuentan, en promedio, con 8 años de mediciones en fechas similares (todas entre 2010 y 2018). Esto equivale a alrededor de 500.000 filas por cada planta eólica.

Hasta el momento hemos desarrollado el código de tal manera que la granularidad de los datos es variable, es decir, es posible elegir un intervalo de mediciones de x minutos o x horas, adaptando la base de datos de acuerdo a este parámetro: se interpola cuando se requiere rellenar fechas sin información o se saca el promedio cuando se quieren intervalos de tiempo más grandes. Los experimentos que realizaremos más adelante incluirán variaciones de esta granularidad para determinar la mejor en cuanto a calidad de predicción y costo computacional.

A pesar de que las bases de datos cuentan con velocidades a 20 metros y a 10 metros de altura, estos valores se extrapolaron a 85 metros de altura (debido a que es la altura típica que tienen las turbinas eólicas) utilizando *Wind Profile Power Law*, la cual establece una relación entre las velocidades de los vientos y las alturas. Específicamente, la relación entre ambos fenómenos está dada por:

$$\frac{u}{u_r} = \left(\frac{z}{z_r} \right)^\alpha$$

En donde u es la velocidad del viento (en metros por segundos) a la

¹<http://walker.dgf.uchile.cl/Mediciones>

altura z (análogamente se definen u_r y z_r). α es un coeficiente que puede ser determinado empíricamente. De hecho, como las bases de datos cuentan con valores de velocidades a distintas alturas, es posible determinar α para cada central eléctrica. Para hacerlo, se tomó el promedio de velocidades a 20 metros de altura y el promedio a 10 metros de altura de cada central. Luego, con estos valores y ocupando la relación anterior, fue posible determinar un α para cada central, con el cual se pudo estimar la velocidad del viento a 85 metros de altura.

Una vez transformadas todas las velocidades a 85 metros de altura, se normalizaron los datos y esta base de datos resultante es la ocupada para realizar las predicciones, con el respectivo intervalo de tiempo que se requiera.

4. Estado actual y trabajo futuro

Los desarrollos realizados para la entrega final están presentes en los archivos **entrega_final.ipynb**, **forecast.py**, **data_preprocessing.ipynb**, **parameters_generation.py** y **utils.py**. En **data_preprocessing.ipynb** se realizó el preprocesamiento de los datos explicados en 3. En **utils.py** se encuentran implementados los siguientes métodos:

1. Cálculo de función de *likelihood* de un modelo VAR para el set de datos.
2. Simulación de los parámetros del modelo VAR a partir de la función de *likelihood* previamente calculada. La simulación es realizada mediante Gibbs Sampling en conjunto con Simulated Annealing. No se ocupó solo Metropolis Hastings porque esto no permitía obtener parámetros variados (la mayoría eran rechazados) y porque queríamos lograr obtener un sampleo más representativo de la real distribución.
3. Simulación de series de tiempo para la velocidad del viento para un parámetro de VAR fijo (*two-stage Monte Carlo*).
4. Conversión de velocidad de viento para distintas centrales en generación eólica total del sistema.

En **forecast.py** se puede encontrar el código utilizado para generar las trayectorias de generación eólica y, a partir de estas, los pronósticos puntuales y de intervalo.

En **parameters_generation.py** está el código utilizado solo para generar las muestras.

Finalmente en *entrega_final.ipynb* se hace una demostración de todos los métodos implementados, incluyendo una evaluación de los pronósticos usando MSE y Cobertura condicional comparando con datos reales. En este archivo se explica con mayor detalle la metodología

References

- [1] N. P. Padhy, Unit commitment - a bibliographical survey, *IEEE Transactions on Power Systems* 19 (2004) 1196–1205.
- [2] Y. V. Makarov, C. Loutan, J. Ma, P. de Mello, Operational impacts of wind generation on california power systems, *IEEE Transactions on Power Systems* 24 (2009) 1039–1050.
- [3] Q. P. Zheng, J. Wang, A. L. Liu, Stochastic optimization for unit commitment - a review, *IEEE Transactions on Power Systems* 30 (2015) 1913–1924.
- [4] S. Cordova, H. Rudnick, A. Lorca, V. J. Martinez, An efficient forecasting-optimization scheme for the intra-day unit commitment process under significant wind and solar power, *IEEE Transactions on Sustainable Energy* (2018) 1–1.
- [5] S. S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power and wind speed forecasting methods with different time horizons, in: *North American Power Symposium 2010*, pp. 1–8.
- [6] L. Xie, Y. Gu, X. Zhu, M. G. Genton, Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch, *IEEE Transactions on Smart Grid* 5 (2014) 511–520.
- [7] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Huang, K. Subbarao, Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures, *IEEE Trans. on Sustainable Energy* 2 (2011) 433–442.
- [8] C. Lowery, M. O’Malley, Impact of wind forecast error statistics upon unit commitment, *IEEE Transactions on Sustainable Energy* 3 (2012) 760–768.
- [9] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer Berlin Heidelberg.
- [10] M. Lydia, S. S. Kumar, A. I. Selvakumar, G. E. P. Kumar, A comprehensive review on wind turbine power curve modeling techniques, *Renew. and Sust. Energy Reviews* 30 (2014) 452 – 460.

- [11] P. Christoffersen, Evaluating interval forecasts, *International Economic Review* 39 (1998) 841–62.