1. *Nombre y apellidos del ponente*: *Sergio Arregui Remón*

2. *Cargo:* *Investigador Senior/Responsable de Proyecto*

3. *Empresa:* *CIRCE Centro Tecnológico*

4. *Teléfono y correo electrónico:* *sarregui@fcirce.es*

5. *Título de la ponencia:* *Diagnóstico de pérdidas de rendimiento de aerogeneradores mediante IA explicable*

6. *Selección de la temática preferida*: *Gestión datos, IA y ciberseguridad. La IA para el diagnóstico y predicción del fallo.*

Diagnóstico de pérdidas de rendimiento de aerogeneradores mediante IA explicable

Las infraestructuras modernas de generación eólica generan una cantidad masiva de información procedente de sistemas SCADA y CMS, aunque una parte significativa de estos datos continúa infrautilizada. Este trabajo aborda el diagnóstico de subrendimiento en aerogeneradores mediante la combinación de modelos de predicción basados en aprendizaje automático con técnicas de inteligencia artificial explicable (xAI), con el objetivo de mejorar la toma de decisiones operativas y de mantenimiento, yendo más allá de las curvas de potencia tradicionales.

En la operación estándar de parques eólicos, la producción esperada se modeliza en función del recurso eólico disponible, bien mediante curvas de potencia ajustadas (aproximación clásica) o con modelos avanzados de Machine Learning (ML). Si bien los modelos de ML superan en rendimiento a la aproximación clásica, persiste una variabilidad significativa entre la potencia predicha y la medida, especialmente en rangos intermedios de velocidad del viento. Esta discrepancia se asocia a efectos operativos internos de los aerogeneradores (WTG) que no quedan explicados únicamente por la meteorología. La Ilustración 1 ilustra este fenómeno mostrando, por un lado, la relación entre los residuales de potencia obtenidos con un modelo ML multivariable y la velocidad del viento y, por otro, la distribución estadística de dichos residuales, evidenciando la existencia de patrones no explicados por el modelo meteorológico que motivan la incorporación de variables operativas en el análisis.

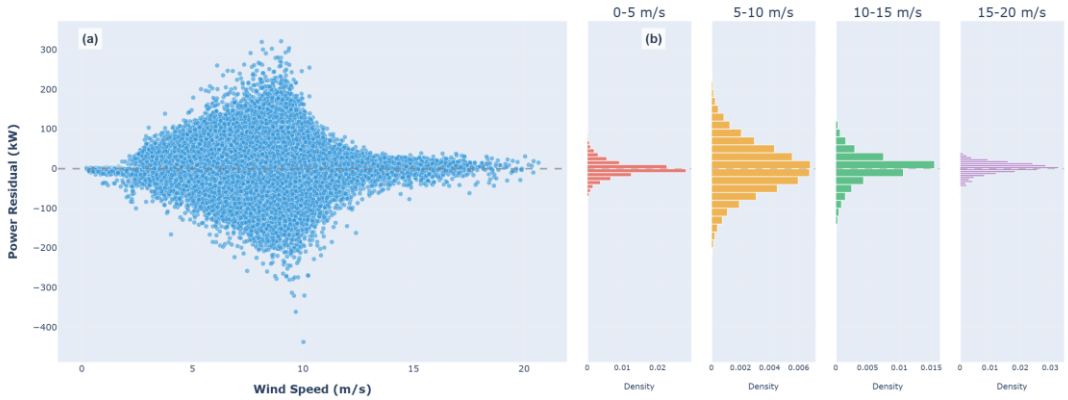


Ilustración 1: Residuales de potencia frente a velocidad del viento (panel a) y distribución estadística de los residuales (panel b)

La metodología propuesta se estructura en cinco etapas. En primer lugar, se realiza un filtrado de datos para seleccionar periodos representativos de operación normal, excluyendo estados de alarma, curtailment y otros regímenes no estacionarios. En segundo lugar, se entrena un modelo meteorológico basado en Histogram-based Gradient Boosting Regression Trees para predecir la producción a partir de variables ambientales, alcanzando una mejora media del 18,5 % en RMSE respecto a enfoques clásicos, en los datasets analizados y para los periodos de testing considerados. En tercer lugar, se lleva a cabo la selección de variables operativas mediante análisis de correlación y clustering jerárquico sobre diecisiete señales, reduciéndolas a un subconjunto de siete características independientes, posteriormente normalizadas mediante Normal Behaviour Models [1], [2]. En cuarto lugar, se entrena un modelo operativo adicional para explicar los residuos del modelo meteorológico a partir de estas variables, logrando incrementos en la capacidad explicativa de entre el 28 % y el 38 % en el parque de Kelmarsh [3] y entre el 15 % y el 32 % en Penmanshiel [4]. La interpretación de estos modelos mediante valores SHAP permite identificar las variables operativas más influyentes en la variabilidad de la producción, proporcionando una base sólida para el análisis posterior. No obstante, el principal valor de la metodología radica en su capacidad para traducir este conocimiento en mejoras operativas cuantificables.

Finalmente, en la quinta etapa, se aplican técnicas de interpretabilidad y optimización contrafactual para cuantificar y atribuir las desviaciones observadas. Para ello, ejecutamos una optimización contrafactual mediante un algoritmo de evolución diferencial [5] que actúa sobre las variables operativas más relevantes en cada intervalo temporal de 10 minutos. Este procedimiento busca maximizar la producción estimada por el modelo, generando escenarios operativos alternativos que permiten cuantificar el margen de mejora disponible, y distribuirlo entre las

diferentes variables mediante una asignación basada en el cálculo de valores SHAP. La Ilustración 2 muestra un ejemplo de optimización sobre un punto 10-minutal concreto.

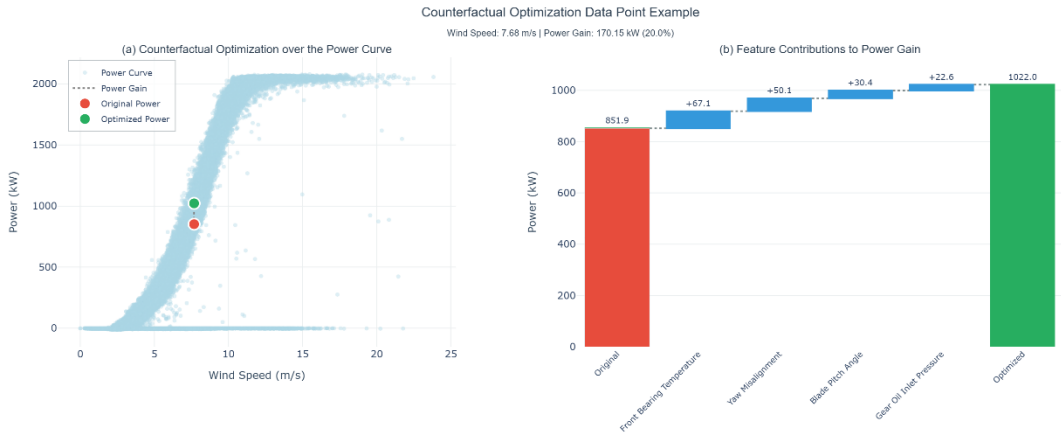


Ilustración 2: Resultados de la metodología de optimización contrafactual y atribución de variables para un intervalo SCADA representativo de 10 minutos. En el panel (a) se muestra la curva de potencia completa como referencia (correspondiente al aerogenerador Kelmarsh 3), sobre la que se identifica el punto original analizado (círculo rojo). El círculo verde indica la producción optimizada obtenida tras la aplicación del algoritmo de optimización contrafactual. En el panel (b) se presenta la descomposición de la ganancia potencial de potencia, atribuida a las cuatro variables operativas consideradas en el proceso de optimización.

Este análisis permite no solo estimar incrementos teóricos de producción, sino también atribuir de manera explícita dichas mejoras a subsistemas concretos del aerogenerador. Los resultados obtenidos, validados sobre datos reales de los parques de Kelmarsh y Penmanshiel, demuestran que la combinación de modelos de Machine Learning con técnicas de xAI permite no solo caracterizar la variabilidad no explicada en la producción, sino también transformarla en acciones concretas mediante la cuantificación del potencial de mejora. Esta aproximación habilita estrategias de mantenimiento y operación más focalizadas y basadas en datos, alineadas con la maximización del rendimiento de los activos.

[1] A. Meyer, 'Multi-target normal behaviour models for wind farm condition monitoring', *Appl. Energy*, vol. 300, p. 117342, Oct. 2021, doi: 10.1016/j.apenergy.2021.117342.

[2] X. Chesterman, T. Verstraeten, P.-J. Daems, A. Nowé, and J. Helsen, 'Overview of normal behavior modeling approaches for SCADA-based wind turbine condition monitoring demonstrated on data from operational wind farms', *Wind Energy Sci.*, vol. 8, no. 6, pp. 893–924, Jun. 2023, doi: 10.5194/wes-8-893-2023.

[3] 'Kelmarsh'. [Online]. Available: https://zenodo.org/records/16807551

[4] ‘Penmanshiel’. [Online]. Available: https://zenodo.org/records/16807304
[5] R. Storn and K. Price, ‘Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces’, *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, Dec. 1997, doi: 10.1023/A:1008202821328.