

Jornadas de Automática

Modelado de la generación fotovoltaica en una comunidad energética

Mitjavila, Berta^a, Oliver, Miquel^{a*}

^a Grupo de investigación Nets, Universitat Pompeu Fabra, Barcelona, España.

To cite this article: Mitjavila, Berta, Oliver, Miquel. 2025. Modelling of electrical consumption and generation in an energy community. *Jornadas de Automática*, 46. <https://doi.org/10.17979/ja-cea.2025.46.12268>

Resumen

Una comunidad energética es una forma de organización colectiva entre la ciudadanía, las empresas y las administraciones locales que permite producir, gestionar y consumir energía de manera conjunta, con el objetivo de obtener beneficios ambientales, sociales y económicos. Este proyecto se centra en el desarrollo de modelos de optimización y predicción de la generación fotovoltaica dentro del marco de una comunidad de este tipo. Se trabaja con datos horarios recogidos entre enero de 2023 y diciembre de 2024, procedentes de dos instalaciones fotovoltaicas ubicadas en Vilassar de Mar (Maresme, Barcelona), así como con datos meteorológicos del mismo municipio. El estudio se estructura en dos líneas principales: (1) la identificación de la distribución estadística que mejor se ajusta a la producción fotovoltaica, y (2) el desarrollo de modelos predictivos basados en técnicas de aprendizaje automático, con el objetivo de estimar la generación futura a partir de predicciones meteorológicas.

Palabras clave: Modelado e integración de sistemas de energías renovables, Modelado de series temporales, Predicción, Aprendizaje automático, Optimización, Modelado.

Modelling of electrical consumption and generation in an energy community

Abstract

An energy community is a form of collective organization between citizens, companies, and local administrations that enables to produce, manage and consume energy jointly, with the aim of achieving environmental, social and economic benefits. This project develops models to optimize and forecast photovoltaic generation within the framework of an energy community. The study uses hourly data collected between January 2023 and December 2024 from two photovoltaic installations located in Vilassar de Mar (Maresme, Barcelona), along with meteorological data from the same municipality. The work is structured around two main lines: (1) identifying the statistical distribution that best fits the photovoltaic production, and (2) developing predictive models based on machine learning techniques to estimate future generation from weather forecasts.

Keywords: Renewable Energy System Modeling and Integration, Forecasting, Time series modelling, Machine Learning, Optimization, Modeling.

1. Introducción

Las comunidades energéticas constituyen formas de organización colectiva integradas por la ciudadanía, pequeñas y medianas empresas y administraciones locales con el objetivo producir y gestionar energía de forma compartida (Institut Català d'Energia, ICAEN, s.f.). Este modelo, basado en la gobernanza democrática y la propiedad colectiva, es una alternativa sostenible al sistema energético actual caracterizado por una estructura altamente centralizada y dependiente de los combustibles fósiles. El auge de estas

iniciativas plantea nuevos retos para la gestión eficiente de la producción y el consumo energético dentro de la comunidad, en un contexto de alta dependencia de condiciones meteorológicas y hábitos de consumo. Este escenario plantea la necesidad de desarrollar modelos que permitan predecir y optimizar el comportamiento energético con el fin de mejorar la gestión del autoconsumo, especialmente con la inclusión de sistemas de almacenamiento que dan más flexibilidad a la inyección de la energía generada a la red.

El presente estudio aborda dos líneas principales de trabajo: por un lado, el ajuste de distribuciones estadísticas para

caracterizar la producción fotovoltaica en función del mes del año (Pasari, S. et al, 2020); por otro, el desarrollo de modelos de regresión que permiten predecir la generación a partir de predicciones meteorológicas y variables temporales (Abad-Alcaraz, V. et al., 2023; J. A. R. Huamani et al., 2023). El análisis se lleva a cabo utilizando datos del municipio costero de Vilassar de Mar en la provincia de Barcelona.

2. Obtención y tratamiento de datos

Esta sección describe las características de los datos empleados, su origen, su localización, y el tratamiento aplicado para garantizar su calidad, consistencia y adecuación al análisis.

2.1. Localización geográfica

Todos los datos utilizados en este estudio corresponden al municipio de Vilassar de Mar, situado en la comarca del Maresme, provincia de Barcelona. Se trata de un municipio costero con un clima mediterráneo caracterizado por inviernos suaves y húmedos, y veranos cálidos y secos. Las principales precipitaciones se registran durante la primavera y el otoño.



Figura 1: Ubicación del municipio de estudio

Los datos de generación y los registros meteorológicos históricos se obtienen de las ubicaciones exactas de las instalaciones, mientras que las predicciones meteorológicas utilizadas son generales para el municipio, ya que no se especifica su punto de origen exacto.

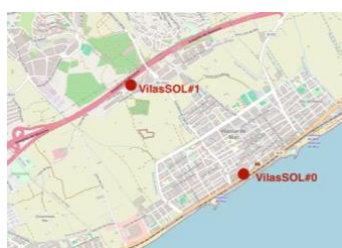


Figura 2: Ubicación de las instalaciones de estudio

La localización condiciona las características climáticas, las cuales influyen directamente en la producción fotovoltaica.

La Figura 1 y la Figura 2 han sido obtenidas de OpenStreetMap (<https://www.openstreetmap.org>) y sus datos están disponibles bajo la licencia *Open Data Commons Open Database License* (ODbL).

2.2. Descripción de las instalaciones fotovoltaicas

El análisis se basa en datos procedentes de dos instalaciones fotovoltaicas ubicadas en Vilassar de Mar. La primera, denominada VilasSOL#0, se compone de sólo 10 placas de generación fotovoltaica y ha servido de referencia

para la instalación industrial posterior. Se encuentra ubicada en el núcleo de población de Vilassar de Mar a pocos metros de la segunda instalación. La segunda instalación, denominada VilasSOL#1 y con 130 placas fotovoltaicas, está situada en la calle Camí del Míg, dentro del polígono industrial Els Garrofers. Ambas instalaciones cuentan con sistemas de monitorización que registran datos horarios de generación eléctrica de manera constante.

Tabla 1: Características técnicas de las instalaciones

	VilasSOL#0	VilasSOL#1
Módulos	10 módulos Jinko JKM420N-54HL4R-B, monocristalinos, 420 Wp	168 módulos Jinko Tiger Neo N-type 54HcL4R, monocristalinos, 425 Wp (420–440 W)
Inversores	1 inversor Huawei SUN2000, monofásico	2 inversores Huawei trifásicos: SUN2000-50KTL-M3 (50 kW) y SUN2000-20KTL-M5 (20 kW)
Baterías	No dispone	2 baterías SonnenBatterie 10 (22 kWh, 4,6 kW, LFP, 10.000 ciclos/10 años, IP30, con powermeter y toroidales)



a) Instalación VilasSOL#0



b) Instalación VilasSOL#1

Figura 3: Vista de los paneles fotovoltaicos

2.3. Fuentes de datos

Los datos de generación fotovoltaica se obtienen directamente de las aplicaciones de monitorización de los inversores instalados en VilasSOL#1 y VilasSOL#0, W-Manager (<https://www.w-manager.com>) y FusionSolar (<https://fusionsolar.huawei.com>) respectivamente. Estos, se utilizan bajo un acuerdo de confidencialidad aceptado por los socios de las comunidades. Dicho acuerdo, titulado "*Acord de Confidencialitat: Projecte d'optimització i predicció dels consums d'energia elèctrica en una Comunitat Energètica*", regula la cesión y uso de estos datos para este estudio.

Para las variables meteorológicas, se utilizan las siguientes fuentes, que proporcionan datos históricos y predicciones meteorológicas para el municipio: NASA POWER (<https://power.larc.nasa.gov>), Agencia Estatal de Meteorología (AEMET) (<https://www.aemet.es>) y TEMIS (<https://www.temis.nl>). Estos datos son de uso abierto y están disponibles públicamente.

2.4. Marco temporal del conjunto de datos

Los dos conjuntos de datos presentados anteriormente cuentan con una resolución horaria, es decir, registran un valor por cada hora del día (24 registros diarios). En el caso de la instalación VilasSOL#1, se dispone de información desde enero de 2023 hasta diciembre de 2024, lo que supone dos años completos de datos. Esto equivale a un total de 17.544 registros, distribuidos en 8.760 registros correspondientes al año 2023 y 8.784 al año 2024, dado que este último fue bisiesto.

Los datos meteorológicos presentan la misma cobertura temporal y, además, incluyen registros correspondientes a mayo de 2025, que han sido utilizados para la realización de predicciones. Los datos meteorológicos tienen una cadencia diaria.

El conjunto de datos de generación correspondiente a la instalación VilasSOL#0, únicamente contiene registros del año 2024, al tratarse del primer año completo de funcionamiento desde la puesta en marcha del sistema. Por lo tanto, y dado que el año fue bisiesto, se dispone de 8.784 registros para esta instalación.

2.5. Preprocesamiento de los datos

El preprocesamiento tiene en cuenta el correcto formado de los datos en función de su tipología y se han realizado los cambios pertinentes para asegurar su calidad y consistencia.

Además, hemos convertido todos los datos a Tiempo Universal Coordinado (UTC), restando 1 hora en invierno y 2 horas en verano respecto al horario oficial. Esto se ha hecho, por un lado, para evitar distorsiones por el cambio horario, y, porque, aunque España utiliza UTC+1 en invierno y UTC+2 en verano, geográficamente se encuentra en el huso horario UTC+0. El horario oficial no coincide con la posición real del sol. Usar UTC permite que las horas reflejen mejor la radiación solar real, lo que mejora la precisión del análisis de producción fotovoltaica.

Respecto los valores nulos de las variables de producción fotovoltaica, se ha determinado que el conjunto de datos de VilasSOL#1 contiene un 35,7% de valores nulos, concentrados mayoritariamente entre las 1:00 y las 7:00 h y de 19:00 a 23:00 h, coincidiendo con franjas de baja producción solar, por lo que se han imputado con ceros. En VilasSOL#01, el 5,2% de nulos se concentra entre el 22/06 y el 11/07 de 2024 por un fallo de conexión; se imputan mediante interpolación temporal horaria.

Finalmente, se han conservado los valores atípicos para poder modelar de manera más precisa la distribución de los datos.

3. Análisis estadístico de la generación

El objetivo de esta parte del estudio es identificar la distribución de probabilidad que mejor se ajusta a los datos de generación fotovoltaica. Para ello se analizan dos variables clave: a Producción Fotovoltaica (kWh) registrada en la instalación VilasSOL#1 y el Rendimiento Fotovoltaico (kWh) de la instalación VilasSOL#0. Ambas variables provienen de las aplicaciones de monitorización respectivas y representan la cantidad total de energía eléctrica generada durante el periodo analizado. En este contexto, los términos Producción y Rendimiento Fotovoltaico se consideran equivalentes, ya que ambos hacen referencia a la misma magnitud física: la generación bruta de energía eléctrica del sistema fotovoltaico. Para el análisis, se consideran únicamente los valores positivos de generación (Pasari, S. et al, 2020).

El estudio se realiza de forma mensual con el objetivo de capturar mejor la variabilidad climática propia de cada mes, lo que permite un ajuste más preciso de las distribuciones y de sus parámetros.

3.1. Distribuciones y ajuste de los parámetros

Se han seleccionado las siguientes distribuciones de probabilidad para modelar los datos de producción energética: exponencial, gamma, normal, log-normal, logística, log-logística, Rayleigh y Weibull, distribuciones comunes en el modelado de datos energéticos o ambientales (Pasari, S. et al, 2020).

Para estimar los parámetros de cada distribución se aplica el método de máxima verosimilitud (*Maximum Likelihood Estimation*, MLE) (Myung, 2003).

3.2. Ajuste paramétrico y evaluación de la bondad del ajuste

Una vez seleccionadas las distribuciones y estimados sus parámetros mediante el método de máxima verosimilitud, se realiza el ajuste de cada una a los datos empíricos de generación fotovoltaica.

Posteriormente, se evalúa la calidad de dicho ajuste mediante una prueba de bondad de ajuste. En este estudio se ha empleado la prueba de *Kolmogórov-Smirnov* (KS) (Frank J. Massey Jr., 2012), que permite comparar la función de distribución empírica con la función teórica correspondiente, evaluando la mayor diferencia absoluta entre ambas (estadístico D). Esta prueba es especialmente adecuada para distribuciones continuas y muestras de tamaño medio, como en el caso del presente análisis.

A continuación, se presentan los valores del estadístico D obtenidos en la prueba KS para ambas instalaciones fotovoltaicas analizadas.

Tabla 2: Resultados de la prueba KS para VilasSOL#0

	Exponencial	Gamma	Normal	Log-normal	Logística	Log-logística	Rayleigh	Weibull
Enero	0.0824	0.2490	0.1250	0.0929	0.1329	0.1315	0.1185	0.2325
Febrero	0.1268	0.1373	0.1078	0.4129	0.1067	0.1026	0.1173	0.2499
Marzo	0.1303	0.1098	0.1226	0.1284	0.1130	0.1358	0.1154	0.1477

Abril	0.1296	0.1429	0.1223	0.1322	0.1204	0.1533	0.1216	0.2641
Mayo	0.1176	0.1020	0.1023	0.1120	0.0980	0.1038	0.1076	0.1179
Junio	0.1223	0.1658	0.1192	0.1269	0.1231	0.1473	0.1367	0.2736
Julio	0.1472	0.1019	0.1018	0.4288	0.0972	0.0997	0.1172	0.1721
Agosto	0.1404	0.1040	0.1048	0.1049	0.0984	0.1050	0.1049	0.1331
Septiembre	0.1102	0.1348	0.1242	0.1166	0.1273	0.1564	0.1415	0.1532
Octubre	0.1095	0.1483	0.1039	0.1015	0.1111	0.0985	0.1080	0.1025
Noviembre	0.1351	0.2593	0.1476	0.3806	0.1544	0.1219	0.1487	0.2608
Diciembre	0.1298	0.2560	0.1131	0.4108	0.1115	0.1079	0.1267	0.2903

Tabla 3: Resultados de la prueba KS para VilasSOL#1

	Exponencial	Gamma	Normal	Log-normal	Logística	Log-logística	Rayleigh	Weibull
Enero	0.1076	0.1619	0.1172	0.1216	0.1282	0.1379	0.1263	0.1037
Febrero	0.1255	0.1295	0.1318	0.1223	0.1300	0.1868	0.1492	0.1680
Marzo	0.1204	0.1554	0.1257	0.1312	0.1235	0.1816	0.1244	0.1167
Abril	0.1204	0.1853	0.1175	0.1325	0.1218	0.1888	0.1308	0.1570
Mayo	0.1052	0.1440	0.1295	0.1197	0.1306	0.2505	0.1435	0.1957
Junio	0.1209	0.1482	0.1301	0.1294	0.1293	0.1746	0.1402	0.1483
Julio	0.1340	0.1128	0.1186	0.1247	0.1141	0.2025	0.1235	0.1427
Agosto	0.1311	0.1222	0.1235	0.1278	0.1166	0.3290	0.1229	0.1354
Septiembre	0.1223	0.1822	0.1251	0.1274	0.1275	0.2231	0.1392	0.1276
Octubre	0.1101	0.1572	0.1117	0.1235	0.1291	0.1734	0.1361	0.1040
Noviembre	0.1200	0.1180	0.1223	0.1208	0.1200	0.1301	0.1245	0.1055
Diciembre	0.1282	0.1316	0.1131	0.1223	0.1196	0.1669	0.1185	0.1346

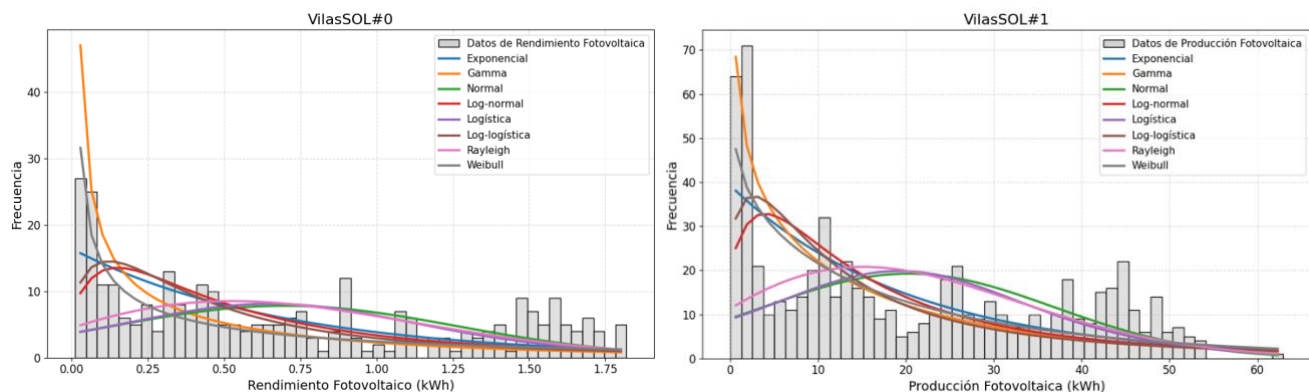


Figura 4: Comparativa gráfica del ajuste de las distribuciones para enero

A modo ilustrativo, en la Figura 4 se representa el ajuste de las distintas distribuciones sobre los datos de generación fotovoltaica para el mes de enero. Esta permite apreciar visualmente el grado de adecuación de cada distribución, complementando la evaluación numérica presentada en las Tablas 2 y 3.

3.3. Discusión de los resultados

Según los resultados presentados en la Tabla 2, las distribuciones que mejor se ajustan a la producción fotovoltaica para VilasSOL#0 son las siguientes. para enero, la distribución exponencial; febrero, log-logística; marzo, gamma; abril, logística; mayo, logística; junio, logística; julio, logística; agosto, logística; septiembre, exponencial; octubre, log-logística; noviembre y diciembre, log-logística.

A nivel estacional, no se observa un patrón uniforme durante los meses de invierno. Durante la primavera y el

verano, la distribución logística es la que se ajusta mejor, lo que indica una producción energética concentrada alrededor de la media y con menor asimetría en comparación con otros períodos. En otoño, la distribución que predomina es la log-logística, caracterizada por una asimetría y variabilidad, lo que sugiere una presencia más significativa de valores extremos o fluctuaciones en la producción fotovoltaica.

Como se observa en la Tabla 3, las distribuciones que mejor se ajustan a la producción fotovoltaica para VilasSOL#1 son las siguientes: para enero, Weibull; febrero, log-normal; marzo, Weibull; abril, normal; mayo, exponencial; junio, exponencial; julio, exponencial; agosto, logística; septiembre, exponencial; octubre, Weibull; noviembre, Weibull; y diciembre, normal.

En este caso, durante los meses de otoño e invierno predomina la distribución Weibull, indicando una distribución con sesgos moderados y cierta concentración

alrededor de valores centrales. Por otro lado, en primavera y verano predomina la distribución exponencial, lo que sugiere una mayor presencia de eventos con baja producción energética y una cola larga que representa valores altos ocasionales.

Se observa que, en la mayoría de los casos y con algunas excepciones, las diferencias entre las distribuciones ajustadas no son especialmente significativas. Asimismo, ninguna distribución resulta ser la óptima en todos los meses, lo que pone de manifiesto la necesidad de realizar análisis mensuales individualizados, en línea con la marcada estacionalidad de la producción energética.

En estudios basados en datos naturales, se considera que un ajuste es adecuado cuando el valor del estadístico D de Kolmogórov-Smirnov (KS) es inferior a 0.15 (Razavi, S et al., 2012). En el presente análisis, la distribución con el menor valor de D en cada mes para ambas instalaciones cumple consistentemente con este criterio. Además, la mayoría de los valores del estadístico D son inferiores al umbral de 0.15.

En conjunto, los resultados permiten concluir que los ajustes son estadísticamente sólidos en todos los meses analizados para ambas instalaciones.

4. Modelado predictivo de la generación

En este apartado se desarrolla un modelo predictivo de regresión basado en técnicas de aprendizaje automático supervisado, cuyo objetivo es estimar la generación fotovoltaica diaria a partir de variables meteorológicas y temporales. Debido a las limitaciones en la disponibilidad de datos de predicción proporcionados por AEMET, así como a la accesibilidad de los datos históricos públicos y al período de funcionamiento de las instalaciones analizadas, el modelado se ha realizado exclusivamente con los datos del año 2023 correspondientes a la instalación VilasSOL#1.

4.1. Variables predictivas

Se utilizan variables meteorológicas como el índice ultravioleta máximo, la temperatura máxima y mínima, la velocidad del viento y la humedad relativa, junto con variables temporales (pendientes de especificar), para predecir la generación fotovoltaica. En términos de correlación, el índice ultravioleta máximo muestra una relación positiva alta con la generación (0.87), seguido por la temperatura máxima (0.62) y mínima (0.55). En cambio, la velocidad del viento (-0.22) y la humedad relativa (-0.22) presentan correlaciones negativas débiles.

A nivel temporal se incluyen el mes y la semana del año, dado que, al entrenar el modelo con datos de un solo año, incluir el día exacto provoca que los modelos no generalicen bien y los resultados no sean satisfactorios. Por ello, se seleccionan las variables mes y semana, a las que se aplica una codificación cíclica mediante funciones seno y coseno para representar correctamente la continuidad temporal. Así, los modelos se entrenan con cuatro variables relacionadas con el tiempo: el valor seno y coseno del mes, y el valor seno y coseno de la semana.

Adicionalmente, con el objetivo de ampliar la aplicabilidad del modelo al año 2024 y, en consecuencia, también, a la instalación de VilasSOL#0, se exploró una estrategia basada

en el uso del índice ultravioleta acumulado diario, para el cual sí se dispone de registros históricos. Dado que las predicciones meteorológicas de AEMET proporcionan únicamente el valor máximo del índice UV diario, se intentó estimar el valor acumulado a partir de este máximo, mediante un modelo auxiliar de aproximación. Sin embargo, esta aproximación introdujo un alto grado de error y redujo notablemente el rendimiento predictivo del modelo principal, por lo que esta estrategia fue descartada.

4.2. Modelos de regresión y metodología entrenamiento

Se plantean modelos de regresión, incluyendo regresión lineal simple y versiones con regularización *Ridge* y *Lasso*. Además, se consideran modelos basados en *Decision Tree*, *Support Vector Machine*, *Random Forest* y redes neuronales artificiales entrenadas con los algoritmos *Adam* y *Root Mean Square Propagation* (Abad-Alcaraz, V et al., 2023).

La selección de hiperparámetros se lleva a cabo mediante el método de *Grid Search*, que permite evaluar de forma sistemática múltiples combinaciones de parámetros para encontrar la configuración que optimiza el rendimiento del modelo. Este enfoque es ampliamente utilizado en el ajuste fino de modelos de aprendizaje automático por su eficacia y simplicidad (Pedregosa, F. et al., 2011).

La red neuronal artificial implementada presenta la siguiente configuración: se utilizan cuatro variables de entrada (características), una única capa oculta con n neuronas y función de activación *ReLU* (Nair, V. and Hinton, G.E., 2010), y una capa de salida compuesta por una sola neurona con activación lineal, adecuada para tareas de regresión. La función de pérdida empleada es el error cuadrático medio (MSE), y como algoritmos de optimización se consideran *Adam* y *Root Mean Square Propagation* (RMSprop) (Tieleman, T. and Hinton, G., 2012). El modelo se entrena durante 100 épocas, con un tamaño de lote (*batch size*) de 32, reservando el 10% de los datos para validación.

El número de neuronas en la capa oculta (J. A. R. Huamani et al., 2023) se determina según el teorema de Kolmogorov (Kolmogorov, A.N., 1957), el cual establece que el número máximo de neuronas necesarias en la capa oculta de una red de retropropagación puede estimarse mediante la expresión $2n + 1$, donde n representa el número de variables de entrada. En el presente caso, con $n = 9$, se determina que el número de neuronas en la capa oculta debe satisfacer $N \leq 19$.

Adicionalmente, según la fórmula empírica Ge y Zu (Ge, Z., & Sun, Z., 2007), el número de neuronas de la capa oculta debe satisfacer la siguiente desigualdad:

$$\sum_{i=0}^n C_N^i > K$$

Donde K representa el número de muestras del conjunto de datos disponibles para entrenar la red neuronal, en el presente estudio 365. El lado izquierdo de la desigualdad corresponde a la suma de combinaciones.

Resolviendo la expresión, se determina que el número de neuronas de la capa oculta debe cumplir obtenemos que $9 \leq N \leq 19$. Se entrenaron y validaron redes neuronales para todos los valores posibles dentro de este rango, evaluando el

rendimiento mediante el MSE. Los resultados indican que el valor óptimo de N es 19 para el modelo optimizado con Adam y 15 para el modelo optimizado con RMSprop.

4.3. Evaluación del modelo y discusión de los resultados

Los modelos se evaluaron comparando predicciones diarias con la generación real de energía fotovoltaica en VilasSOL#1 durante diez días, del 21 al 31 de mayo de 2025, aprovechando datos consistentes de ese periodo.

La métrica principal empleada ha sido el coeficiente de determinación R^2 , que permite evaluar la capacidad explicativa del modelo respecto a la variabilidad observada en los datos reales (J. A. R. Huamani et al., 2023). La tabla que se muestra a continuación resume los parámetros utilizados y el R^2 obtenidos para los dos modelos con mejor rendimiento: el *Decision Tree* y el *Random Forest*.

Tabla 4: Parámetros y R^2 de los modelos con mejor rendimiento

Modelo	R^2	Parámetro	Valor
Decision Tree	0.3411	Criterio de división	Friedman MSE
		Profundidad máxima	10
		Tamaño mínimo de hoja	4
		Tamaño mínimo de división	10
Random Forest	0.3567	Número de árboles	200
		Tamaño mínimo de división	5

Los resultados de la Tabla 4 muestra que ambos son insuficientes para un ajuste adecuado del problema. Los demás modelos presentan valores negativos de R^2 , lo que indica que no logran explicar la variabilidad de los datos y tienen un rendimiento muy limitado dada la limitación temporal de los datos y del tiempo de vida de las instalaciones.

En resumen, aunque Decision Tree y Random Forest son los más prometedores dentro del conjunto, su capacidad predictiva aún es limitada y sería necesario explorar mejoras o modelos alternativos para obtener resultados más robustos.

Los valores que se obtiene del R^2 en los modelos ponen en manifiesto las limitaciones de estos, atribuibles a la escasez de datos históricos y a la dependencia de modelos de predicción meteorológica para realizar las estimaciones.

5. Conclusiones

El ajuste de las distribuciones presenta resultados que nos demuestran que la modalización estadística puede ser aplicable a entornos reales y como es el caso del presente estudio a instalaciones con diferentes magnitudes. El resultado es bueno para todos los meses, sin hacer una distinción en función de las tendencias climatológicas en las diferentes estaciones del año. En este caso, la metodología propuesta es aplicable a otras comunidades energéticas con características similares, facilitando la transición hacia sistemas energéticos más descentralizados y renovables.

Respecto al modelo de predicción, aunque los resultados no sean aplicables, debido a la alta correlación entre el índice ultravioleta máximo diario y la generación fotovoltaica, se considera interesante profundizar en esta línea, ya que podría aportar información relevante para mejorar la precisión y utilidad de las predicciones. Asimismo, a un nivel más general, se destaca la necesidad de incluir variables temporales y meteorológicas, dado el impacto de la variabilidad estacional y climática en la generación fotovoltaica. En estudios futuros, sería interesante avanzar en esta dirección.

Finalmente, se concluye que la combinación del análisis estadístico y un modelo de regresión mejorado, que pueda ser aplicado de manera efectiva, permitiría optimizar el autoconsumo y la integración de sistemas de almacenamiento, contribuyendo así a una gestión energética, sostenible y descarbonizada. Este estudio pretende poner de manifiesto la necesidad de esta transición energética y ofrecer herramientas que faciliten su implementación.

Agradecimientos

Este trabajo ha sido realizado gracias al apoyo de la Universidad Pompeu Fabra, en particular del Grupo de Investigación NeTS, así como de los socios de la Comunidad Energética de Vilassar de Mar, VilasSOL.

Referencias

- Abad-Alcaraz, V., et al., 2023. Desarrollo de modelos de predicción de radiación solar mediante técnicas de machine learning. *XLIV Jornadas de Automática*, 382-387. DOI: <https://doi.org/10.17979/spudc.9788497498609.382>
- CAEN, s.f. [Website]. *Catalan Institute of Energy*. Available at: <https://icaen.gencat.cat>
- Frank J. Massey Jr., 2012. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 68-18. DOI: <https://doi.org/10.2307/2280095>
- Ge, Z., Sun, Z., 2007. Neural Network Theory and MATLAB R2007 Application; 108-122.
- J. A. R. Huamani et al., 2023. Prediction of hourly solar radiation using an Artificial Neural Network model in a district of Peru. 18th Iberian Conference on Information Systems and Technologies (CISTI), 1-4. DOI: 10.23919/CISTI58278.2023.10211948
- Kolmogorov, A.N., 1957. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk. Russ. Acad. Sci.* 1957, 114, 953-956.
- Myung, I.J., 2003. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* 47, 90-100. DOI: [https://doi.org/10.1016/S0022-2496\(02\)00028-7](https://doi.org/10.1016/S0022-2496(02)00028-7)
- Pasari, S. et al, 2020. Statistical Modeling of Solar Energy. In: K. S. Sangwan and C. Herrmann (Eds.), *Sustainable Production, Life Cycle Engineering and Management*, India, pp. 157-165.
- Razavi, S. et al., 2012. Review of surrogate modeling in water resources. *Water Resources Research*, 48 (7). DOI: 10.1029/2011WR011527
- Pedregosa, F. et al., 2011. Scikit-learn: Machine Learning in Python. The *Journal of Machine Learning Research*, Volume 12, pp 2825 - 2830.
- Tieleman, T. and Hinton, G., 2012. Lecture 6.5 - RMSProp: Divide the gradient by a running average of its recent magnitude. *Coursera: Neural Networks for Machine Learning*.
- Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 807-814