
Modelo SARIMAX para la Predicción de Generación Solar en la Planta Sol del Desierto en Chile: Evaluación de Variables y Modelos

¹ Felipe Sebastián Neira Rojas,

² Diego Ignacio Rocha Retamal,

³ Angel Tomas Llanos Herrera*

^{1,2,3} Universidad Católica del Maule,
Avenida San Miguel 3605, Talca, Chile

e-mail: ³ angel.llanos@alu.ucm.cl

Teléfono: ³ +56 9 3838 6768

Abstract -

Este estudio ajusta la dinámica de la generación de energía solar en la planta fotovoltaica Sol del Desierto, mediante la aplicación de modelos estadísticos multivariados. Se evaluó la relación entre variables meteorológicas y la producción energética, priorizando aquellas con mayor significancia correlacional. Inicialmente, se implementó un modelo Vectorial Autoregresivo con 7 retardos para predecir temperatura máxima, mínima, promedio y presión atmosférica, todas evaluadas como series estacionarias. Posteriormente, se construyeron modelos Modelo Autorregresivo Integrado de Media Móvil con Variables Exógenas Estacional considerando variables exógenas (temperatura máxima y presión promedio), evaluando su capacidad predictiva mediante métricas como Error de raíz cuadrada media y Error Absoluto Medio, así como el cumplimiento de supuestos estadísticos. Los resultados indican que la presión promedio mejora el ajuste del modelo Autorregresivo Integrado de Media Móvil con Variables Exógenas Estacional, pero con violaciones de supuestos que reducen la fiabilidad de sus pronósticos numéricos.

This study adjusts the dynamics of solar energy generation at the Sol del Desierto photovoltaic plant by applying multivariate statistical models. The relationship between meteorological variables and energy production was evaluated, prioritizing those with the highest correlational significance. Initially, a Vector Autoregressive model with 7 lags was implemented to predict maximum, minimum, average and atmospheric pressure temperatures, all evaluated as stationary series. Subsequently, Integrated Autoregressive Moving Average Autoregressive Model with Exogenous Seasonal Variables models were built considering exogenous variables (maximum temperature and average pressure), evaluating their predictive capacity by means of metrics such as Root Mean Square Error and Mean Absolute Error, as well as the fulfillment of statistical assumptions. The results indicate that average pressure improves the fit of the Integrated Autoregressive Moving Average with Seasonal Exogenous Variables model, but with violations of assumptions that reduce the reliability of its numerical forecasts.

Palabras clave: energía solar, meteorología, modelos predictivos, planta fotovoltaica, Modelo Autorregresivo Integrado de Media Móvil con Variables Exógenas Estacional

Keywords: solar energy, meteorology, predictive models, photovoltaic plant, Integrated Autoregressive Moving Average Model with Seasonal Exogenous Variables

Introducción

Chile ha avanzado consistentemente hacía convertirse en un referente regional en energía solar, apalancado principalmente por la excepcional radiación del Desierto de Atacama y sus óptimas condiciones climáticas. En el año 2025, el país ha superado los 10 680 MW de capacidad fotovoltaica instalada, con una marcada

concentración de proyectos en las regiones áridas del norte, como Antofagasta, Atacama y Coquimbo. Este despliegue se ve favorecido por la combinación de alta radiación, condiciones de aridez y una fuerte demanda energética proveniente del sector minero.

Este rápido crecimiento plantea desafíos importantes para la comprensión del comportamiento altamente variable de la generación solar, el cual depende directamente de factores climáticos como la radiación solar, la temperatura ambiental y variables temporales (hora del día, estación del año, entre otras). En este contexto, contar con modelos de pronóstico precisos se vuelve fundamental para optimizar la operación del sistema eléctrico nacional, permitiendo mantener la estabilidad de la red y facilitar una mayor integración de energías renovables. Así, los pronósticos con resolución horaria adquieren una relevancia creciente para la planificación operativa de centrales solares.

No obstante, a pesar del notable desarrollo de la energía solar en Chile, aún existen pocas investigaciones que utilicen modelos estadísticos avanzados para analizar datos reales de generación de plantas específicas en el norte del país. Gran parte de los estudios disponibles se concentran en estimaciones sobre el crecimiento de la capacidad instalada o en proyecciones de mediano y largo plazo, sin abordar el comportamiento diario de la producción energética en plantas concretas. En particular, la planta fotovoltaica Sol del Desierto, ubicada en la comuna de María Elena (Región de Antofagasta), carece de estudios publicados que analicen su generación mediante modelos estocásticos ajustados a datos locales.

Esta carencia limita la comprensión del efecto real de las condiciones meteorológicas sobre la producción horaria de energía, dificultando la implementación de herramientas de pronóstico que puedan ser útiles para la operación diaria. En consecuencia, se identifica una brecha relevante en la literatura técnica, donde resulta necesario desarrollar y evaluar metodologías de predicción horaria basadas en modelos estocásticos clásicos, capaces de incorporar observaciones meteorológicas reales del entorno inmediato de la planta.

El uso de modelos estadísticos basados en series de tiempo ha demostrado ser fundamental para el pronóstico de la generación solar, especialmente cuando se dispone de datos históricos con alta frecuencia temporal y registros validados. Entre ellos, los modelos ARIMA (Autorregresivo Integrado de Media Móvil) permiten capturar la estructura temporal de la serie, considerando tanto su tendencia como su autocorrelación. Su extensión estacional, conocida como SARIMA, permite modelar patrones repetitivos, como los ciclos diarios o anuales, mientras que la variante SARIMAX incorpora variables exógenas —típicamente meteorológicas— que enriquecen la capacidad explicativa del modelo.

En general, estos enfoques son especialmente adecuados cuando se dispone de observaciones continuas y confiables, y cuando se busca un balance entre capacidad predictiva, interpretabilidad y eficiencia computacional. La literatura ha documentado su eficacia en diversos contextos, especialmente en aquellos donde la relación entre variables meteorológicas y la generación solar es conocida o estimable.

El desarrollo histórico de modelos aplicados a la generación fotovoltaica revela una evolución desde enfoques univariantes hacia modelos que integran información climática local. En un estudio inicial, Fara et al. (2021) mostraron que un modelo ARIMA, basado exclusivamente en la serie histórica de generación, podía igualar o incluso superar en precisión a redes neuronales más complejas. Este resultado refuerza la solidez de los enfoques estocásticos clásicos, especialmente cuando se cuenta con datos de alta frecuencia y series estacionarias. [1]

Posteriormente, Benítez et al. (2023) mejoraron este enfoque mediante la incorporación de variables exógenas —como la irradiancia satelital y la temperatura ambiente— dentro de un modelo SARIMAX. Este ajuste redujo significativamente el error de predicción a 24 horas, destacando la importancia de considerar las condiciones climáticas locales como elementos centrales en la modelación de la generación solar. [2]

En paralelo, estudios desde el enfoque de la minería de datos también han contribuido al entendimiento del fenómeno. Fabara et al. (2019) desarrollaron un modelo Random Forest para una planta de 3.6 MW en Ecuador, utilizando irradiancia, temperatura y humedad como predictores. Aunque no se trató de un análisis estocás-

tico, sus resultados respaldan la relevancia de las variables meteorológicas en la predicción de la producción fotovoltaica.[3]

De forma más reciente, Herrera et al. (2023) exploraron el uso de redes neuronales profundas —específicamente LSTM y CNN— aplicadas a series horarias de irradiancia y generación en 23 plantas solares de España.[4] Estos modelos capturan con alto detalle la variabilidad estocástica diaria, aunque su aplicación práctica puede verse limitada por su menor interpretabilidad y mayores requerimientos computacionales. En contraste, Soria et al. (2022) utilizaron un enfoque clásico de regresión lineal múltiple para modelar la irradiancia en Lima, Perú, a partir de variables como temperatura, humedad y viento. Sus resultados evidenciaron altos niveles de significancia estadística, lo que reafirma que los modelos tradicionales siguen siendo herramientas útiles cuando se integran adecuadamente con datos meteorológicos locales.[5]

En conjunto, estos estudios evidencian que tanto los modelos clásicos (ARIMA, regresión) como los enfoques de aprendizaje automático (LSTM, CNN, Random Forest) poseen potencial para predecir la generación solar. No obstante, un desafío recurrente en todos ellos es la incertidumbre inherente a los datos climáticos y la complejidad de su comportamiento. La generación fotovoltaica es altamente sensible a factores como la nubosidad, la radiación solar, la temperatura o el viento, lo que exige modelos robustos capaces de adaptarse a esta variabilidad.

Pese al abundante desarrollo metodológico, la literatura muestra una escasez de estudios centrados en la zona norte de Chile, a pesar de que esta región concentra algunos de los mayores niveles de irradiancia solar del mundo. Existe, por tanto, una brecha investigativa importante, tanto en la calibración de modelos que consideren patrones diarios y estacionales locales, como en la validación empírica con datos reales de plantas solares específicas. La mayoría de trabajos utiliza datos diarios o mensuales, lo que limita la captura de la variabilidad intra-diaria (como el ciclo día/noche) y reduce la utilidad operativa de los pronósticos.

De este modo, se reconoce la necesidad de avanzar en la aplicación de modelos clásicos de series de tiempo ajustados a la escala horaria, que integren variables climáticas locales y permitan mejorar la precisión y aplicabilidad de los pronósticos para plantas fotovoltaicas en regiones como Antofagasta.

El objetivo general de este estudio es ajustar y evaluar si variables meteorológicas pertenecientes a la zona de la planta son importantes para poder pronosticar la generación solar de la planta PV Sol del Desierto, utilizando datos diarios reales. Para ello, se consideran los siguientes objetivos específicos:

- Estimar variables meteorológicas con modelos vectoriales autorregresivos (VAR).
- Identificar las variables climáticas más relevantes para la predicción, evaluando su poder explicativo sobre la variabilidad diaria de la generación solar.
- Evaluar el desempeño predictivo de los modelos ajustados mediante métricas como MAE y RMSE, así como la confiabilidad de los intervalos de pronóstico.

Los resultados esperados de esta investigación contribuirán a mejorar la planificación operativa en sistemas eléctricos con alta participación de energías renovables, aportando evidencia basada en datos reales para una de las zonas con mayor potencial solar del mundo.

Material y Método

Para el desarrollo del análisis predictivo se utilizaron dos fuentes principales de información. La primera corresponde a los datos reales de generación eléctrica de la planta fotovoltaica Sol del Desierto, con una frecuencia horaria desde el 1 de enero de 2024 hasta el 2 de julio de 2025. La segunda fuente corresponde a registros meteorológicos de la zona de María Elena, recopilados desde el 1 de enero de 2019 hasta el 1 de julio de 2025. Cada base de datos contiene variables distintas relevantes para el análisis, como se resume en la Tabla 1.

Datos meteorológicos	Generación de energía
Condición de cielo visible	Generación de energía solar (MWh)
Presión a nivel del mar (hPa)	
Humedad relativa del aire (%)	
Rango óptico meteorológico del instante (m)	
Temperatura del aire seco (°C)	
Temperatura máxima en 1 minuto (°C)	
Intensidad del viento al momento de la medición (kt)	
Visibilidad promedio del último minuto (m)	

Cuadro 1: Variables consideradas en cada base de datos

En la base de datos de correspondiente a la Planta SV Sol del Desierto se identificaron diarios en dónde la planta fotovoltaica no genera energía, por lo que se filtró la base de datos desde las horas 09:00 hasta las 20:00, promediando los diarios para utilizar una base de datos diaria debido a los recursos computacionales disponibles.

Dado que las fuentes poseen coberturas temporales distintas, la cantidad de observaciones disponibles varía entre ambas, como se muestra en la Tabla 2.

Fuente de datos	Número de observaciones
Variables meteorológicas	2 374
Generación de energía eléctrica	549

Cuadro 2: Cantidad de observaciones por fuente

Tal como se indicó en la formulación del problema, se busca predecir los valores futuros de la generación solar en la planta, en función de las condiciones meteorológicas locales. Considerando que las variables climáticas pueden estar correlacionadas entre sí (por ejemplo, temperatura y presión), se justifica el uso de modelos multivariantes.

De acuerdo con Peña (2005), cuando varias series temporales presentan correlaciones cruzadas, resulta metodológicamente adecuado modelar su comportamiento conjunto.[7] En este contexto, se emplea un modelo VAR(p), definido por la siguiente expresión:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t \quad (0.1)$$

Donde:

- c : Vector de constantes (interceptos) de dimensión $k \times 1$.
- y_t : Vector de $k \times 1$ de variables endógenas en el tiempo t .

- A_i : Matrices de coeficientes autorregresivos de dimensión $k \times k$ para cada rezago i .
- ε_t : Vector de errores aleatorios con media cero y covarianza constante.

Complementariamente, se considera también el uso de modelos SARIMAX, en base a lo mencionado en literatura, que permiten incorporar tanto estructura estacional como variables exógenas, para la predicción mediante la siguiente formulación:

$$\Phi(B^s) \phi(B) (1 - B)^d (1 - B^s)^D y_t = \Theta(B^s) \theta(B) \varepsilon_t + \beta(B) X_t \quad (0.2)$$

Donde:

- y_t : Serie endógena observada en el tiempo t .
- X_t : Vector de variables exógenas conocidas en t o en rezagos anteriores.
- $\beta(B)$: Polinomio de coeficientes asociado a X_t , del tipo $\beta_0 + \beta_1 B + \dots + \beta_k B^k$.
- B : Operador de rezago, tal que $B y_t = y_{t-1}$.
- d : Orden de diferenciación regular.
- D : Orden de diferenciación estacional.
- s : Periodicidad estacional.
- $\phi(B)$: Polinomio autorregresivo no estacional.
- $\Phi(B^s)$: Polinomio autorregresivo estacional.
- $\theta(B)$: Polinomio de medias móviles no estacional.
- $\Theta(B^s)$: Polinomio de medias móviles estacional.
- ε_t : Término de error aleatorio o ruido blanco.

Ambos enfoques permiten capturar distintos aspectos del comportamiento temporal de la serie de generación y su relación con las condiciones meteorológicas, facilitando el desarrollo de pronósticos con enfoque técnico-operativo.

Para garantizar la validez estadística de los modelos ajustados, se consideraron los supuestos fundamentales asociados a los modelos VAR y SARIMAX. En primer lugar, se evaluó la estacionariedad de las series mediante el test de Dickey–Fuller aumentado (ADF), condición necesaria para la estimación en modelos VAR. Posteriormente, se validaron los supuestos sobre los residuos, tales como la normalidad (utilizando el test de Jarque–Bera), la homocedasticidad (mediante el test ARCH–LM) y la ausencia de autocorrelación serial (con el test de Ljung–Box). Estas evaluaciones permiten verificar la calidad del ajuste y la confiabilidad de las inferencias obtenidas a partir de los modelos. Los resultados de estas pruebas se reportan en la sección de resultados, junto con las métricas de error que se presentan a continuación.

La evaluación del desempeño de modelos de predicción en contextos como la generación de energía solar requiere el uso de métricas que permitan cuantificar con precisión los errores entre los valores observados y los valores estimados. En este estudio se consideran dos métricas reconocidas en la literatura: el Error Absoluto Medio (MAE) y la Raíz del Error Cuadrático Medio (RMSE). Ambas métricas capturan distintas propiedades del error: mientras que el MAE mide el promedio absoluto de las desviaciones, siendo robusto frente a valores atípicos y fácil de interpretar en términos físicos, el RMSE penaliza con mayor intensidad los errores grandes,

lo que lo hace particularmente sensible a observaciones extremas. Por tanto, la combinación de ambas métricas permite una evaluación más completa del desempeño, capturando tanto la precisión promedio (MAE) como la estabilidad frente a errores extremos (RMSE).

Para el cálculo del RMSE se considera la ecuación 3.3.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (0.3)$$

Por otra parte, para el cálculo del Error Absoluto Medio se considera la ecuación 3.4.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (0.4)$$

Según Martínez et al. (2009), el Criterio de Información de Akaike (AIC) constituye una herramienta fundamental en la selección de modelos estadísticos, ya que permite balancear la calidad del ajuste con la complejidad del modelo. Este criterio penaliza la inclusión excesiva de parámetros, favoreciendo aquellos modelos que logran una mayor eficiencia explicativa con menor sobreajuste.[6] Su formulación matemática se encuentra en la ecuación 2.5.

Estos modelos de series temporales, junto con las métricas, permitirán evaluar la capacidad predictiva de los modelos generados, buscando encontrar el de mejor desempeño en cuanto a AIC.

$$\text{AIC} = -2\log(\mathcal{L}(\hat{\theta})) + 2K \quad (0.5)$$

donde $\log(\mathcal{L}(\hat{\theta}))$ representa el logaritmo de la máxima verosimilitud del modelo ajustado y K corresponde al número de parámetros libres. En este contexto, un menor valor de AIC indica un modelo estadísticamente más adecuado entre un conjunto de alternativas comparables.

Resultados y discusión

Se considera un análisis de las variables a trabajar, tanto de la generación de energía de la planta, como los valores meteorológicos. Se logró visualizar los siguientes valores diarios para la energía generada en la Plata SV Sol del Desierto

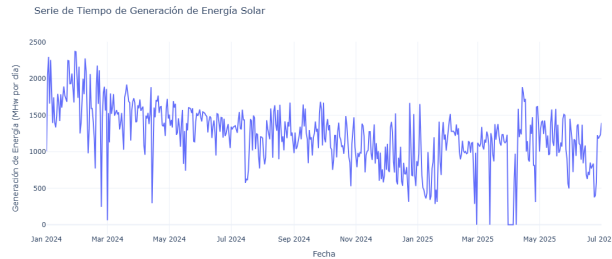


Figura 1: Valores Reales de Generación de Energía de la Planta Sol del Desierto

Se puede notar una gran volatilidad en la generación de energía de esta planta, la cual en las fechas ya mencionadas, considera un rango entre 0 y 2375 Mw por día, considerando días con una generación de 0 Mw por día

los que pueden corresponder a días de mantenimiento o fallas de la planta. La generación de energía solar en la planta de paneles fotovoltaicos Sol del Desierto en Mw por día se logra ver que de forma habitual se encontrará entre 244.7 Mw y 2179.5 Mw, considerando en total 17 días que contó con valores fuera de lo común.

Para generar un modelo de predicción de la generación de energía en base a las variable meteorológicas se estudio cuales de las variables predictoras poseen una mayor relación con la generación de energía. Como resultados generales hubo una correlación muy baja entre la variable dependiente y las independientes, obteniendo una correlación más alta entre la generación de energía y la temperatura máxima, siendo esta una correlación significativa de -0,16. Aportando conocimiento para considerar las variables meteorológicas como variables exógenas para la predicción de la generación de energía, planteando un modelo ARIMAX/SARIMAX a realizar para predecir la productividad de la planta.

Por otra parte se evaluó la relación entre las variables predictoras a considerar en el modelo, mediante un mapa de calor, que considera el índice de correlación de Pearson, en conjunto con la significancia de cada variable.

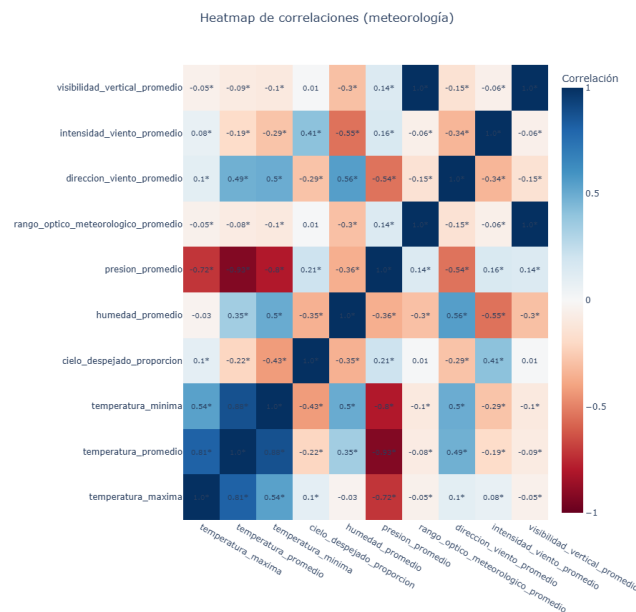


Figura 2: Índices de Correlación y significancia para variables meteorológicas

Como observación importante notar una relación negativa significativa fuerte entre las variables presión a nivel del mar en conjunto con las temperaturas (máximas, mínimas y promedios). Por otra parte, se encuentran mayoritariamente correlaciones significativas dentro del análisis exploratorio, en dónde las más altas se encuentran entre las variables dirección de viento-humedad, intensidad de viento-humedad, dirección de viento-presión a nivel del mar. Estas relaciones resultan ser significativas mayores a 0.5.

Debido a esto se plantea un modelo Vectorial Autorregresivo (VAR) el cual permite predecir variables endógenas, en un mismo vector de variables. Para este modelo se consideraron solamente las variables que tienen una mayor correlación entre ellas, siendo estas significativas. Por lo tanto, primero se evaluó si la serie es estacionaria, para validar la realización del modelo. Para esto se utilizó el test de Dickey Fuller, cuyos resultados se encuentran en el cuadro 3.

Al evaluar la estacionariedad de la serie para cada variable, en base al Test de Dickey Fuller, se rechaza H_0 , indicando que las 4 variables meteorológicas no tienen raíz unitaria, siendo estacionaria. Esto nos permite seguir con el planteamiento del modelo.

Series	p-value
Temperatura Máxima	0.0078
Presión Promedio	0.0057
Temperatura Mínima	0.0026
Temperatura Promedio	0.0287

Cuadro 3: p-valor del ADF Test para variables meteorológicas

Para la elección de la cantidad de retardos p a utilizar en el modelo, se consideró el menor valor del Criterio de Información de Akaike, el cual al comparar con los demás modelos obtuvo una cantidad de 7 retardos ($p=7$).

Por lo tanto, se planteo un modelo VAR(7), que podemos observar en la ecuación 3.1.

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + A_3 y_{t-3} + A_4 y_{t-4} + A_5 y_{t-5} + A_6 y_{t-6} + A_7 y_{t-7} + \varepsilon_t \quad (0.6)$$

Donde:

- c : Vector de constantes (interceptos) de dimensión 4×1 , correspondiente a:

$$c = \begin{bmatrix} -156,68 \\ 259,06 \\ 95,81 \\ 36,67 \end{bmatrix}$$

- y_t : Vector de variables endógenas en el tiempo t , de dimensión 4×1 , dado por:

$$y_t = \begin{bmatrix} \text{temperatura máxima}_t \\ \text{presion promedio}_t \\ \text{temperatura mínima}_t \\ \text{temperatura promedio}_t \end{bmatrix}$$

- A_i : Matrices de coeficientes autorregresivos de dimensión 4×4 para cada rezago $i = 1, \dots, 7$. Por ejemplo, para el rezago 1:

$$A_1 = \begin{bmatrix} 0,6689 & -0,3791 & -0,0038 & 0,4446 \\ 0,3694 & 0,7047 & -0,0303 & 0,3683 \\ 0,1229 & 0,1898 & 0,2956 & 0,5937 \\ 0,3102 & 0,1716 & 0,0365 & 0,6390 \end{bmatrix}$$

- ε_t : Vector de errores aleatorios con media cero y covarianza constante, de dimensión 4×1 .

Mediante este modelo se obtuvieron las siguientes predicciones a estas 4 variables a estudiar.

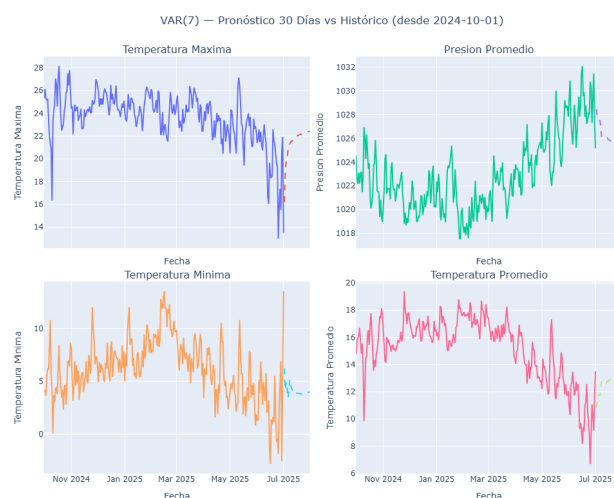


Figura 3: Predicciones de 30 días para Presión a Nivel del Mar y Temperaturas

En la Figura 3 se muestra las predicciones generadas mediante un modelo VAR(7) para las variables meteorológicas de temperatura máxima, mínima, promedio y presión atmosférica promedio, proyectadas a 30 días a partir del 2 de julio de 2025. En general, las series históricas exhiben patrones estacionales y variaciones consistentes con el comportamiento climático típico de zonas áridas, mientras que las proyecciones resultan coherentes en magnitud y dirección. La temperatura máxima, tras un descenso pronunciado, muestra una recuperación moderada hacia valores entre 22 y 24°C. La presión promedio, que había alcanzado un máximo superior a 1030 hPa, presenta una leve reversión en la tendencia, estabilizándose en torno a 1027 hPa. Por su parte, la temperatura mínima revela una alta variabilidad histórica, y el modelo proyecta un ascenso suave hacia rangos de 6 a 8°C, mientras que la temperatura promedio muestra un repunte tras una caída sostenida, alcanzando valores cercanos a 13–14°C. Estas predicciones reflejan una transición climática posterior al invierno, con ajustes suaves en las condiciones térmicas y barométricas, y demuestran que el modelo VAR logra capturar relaciones cruzadas entre variables con una dinámica temporal coherente.

Las métricas correspondientes al modelo, ya sea RMSE y MAE se pueden encontrar en el cuadro 4.

Variable	RMSE	MAE
Temperatura Máxima	3.540	2.484
Presión Promedio	2.421	1.982
Temperatura Mínima	3.424	2.673
Temperatura Promedio	2.156	1.815

Cuadro 4: Métricas de error para el modelo VAR(7) en el pronóstico de 30 días (RMSE y MAE).

Los resultados obtenidos para las métricas de error RMSE y MAE permiten evaluar la precisión del modelo VAR(7) en el pronóstico de las variables meteorológicas durante el periodo del 2 de julio al 31 de Julio de 2025. En general, se observa que la variable con mejor ajuste es la temperatura promedio, con un RMSE de 2.156 y un MAE de 1.815, lo que indica errores de predicción relativamente bajos y estables. Le sigue la presión promedio, con un RMSE de 2.421, reflejando un desempeño aceptable dado el rango observado de esta variable. En contraste, las mayores discrepancias se registran en la temperatura máxima y la temperatura mínima, con RMSE superiores a 3.4, lo que sugiere una mayor variabilidad en estas series y una sensibilidad más alta a eventos extremos recientes. La diferencia entre RMSE y MAE en todas las variables indica la presencia de algunos errores puntualmente más altos (outliers), aunque sin afectar de forma crítica la estabilidad.

Además, se consideró la validación de supuesto del modelo VAR, la cual se visualiza en el siguiente cuadro 5.

Test	p-valor	Cumple
Normalidad (temperatura maxima)	0.000	No
Homocedasticidad(temperatura maxima)	0.081	Sí
Normalidad (presion promedio)	0.056	Sí
Homocedasticidad (presion promedio)	0.091	Sí
Normalidad (temperatura minima)	0.121	Sí
Homocedasticidad (temperatura minima)	0.310	Sí
Normalidad (temperatura promedio)	0.612	Sí
Homocedasticidad (temperatura promedio)	0.055	Sí

Cuadro 5: Supuestos del modelo VAR (agrupados por variable)

La tabla anterior presenta los resultados de los test estadísticos aplicados para verificar dos supuestos fundamentales del modelo VAR: normalidad y homocedasticidad de los residuos, agrupados por variable endógena. En cuanto al supuesto de normalidad, solo los residuos asociados a la variable temperatura máxima presentan un p-valor inferior a 0.05, por lo que se rechaza la hipótesis nula de normalidad en ese caso. En cambio, para las variables presión promedio, temperatura mínima y temperatura promedio, los p-valores son superiores al umbral común de significancia, por lo que no se rechaza la hipótesis nula, indicando que los residuos se distribuyen normalmente. Respecto a la homocedasticidad, es decir, la constancia de la varianza de los errores, todas las variables cumplen con este supuesto, ya que los p-valores obtenidos en cada caso superan el nivel crítico de 0.05. Estos resultados sugieren que el modelo VAR ajustado se comporta adecuadamente en términos de varianza constante, aunque presenta una ligera desviación respecto a la normalidad en una de las variables, lo cual debe ser considerado al interpretar las inferencias.

A partir de lo anterior, para poder predecir la generación de energía de la Planta SV Sol del Desierto se consideró un modelo SARIMAX, incorporando las variables meteorológicas como covariables exógenas, dado que no presentan una relación con la variable dependiente. La selección del mejor modelo se realizó mediante la comparación de los valores AIC para cada combinación de varios posibles parámetros, tanto de autocorrelación (p), diferenciación (d) y media móvil (q).

Como resultado al mejor modelo se considero un SARIMAX(1,1,2)(0,1,1,7), para los modelos que consideraban como variable predictora temperaturas máxima o presión a nivel del mar, los cuales según las métricas del RMSE y MAE tienden a un mejor modelo predictivo.

Por lo tanto, para la predicción de generación de energía según la temperatura máxima se considera la ecuación 3.2

$$(1 - 0,5908B)(1 - B)(1 - B^7)y_t = (1 + 1,2075B + 0,2273B^2)(1 - 0,9415B^7)\varepsilon_t + 12,1618 \cdot x_t \quad (0.7)$$

Donde:

- y_t : Generación de energía solar (Mw) en el tiempo t .
- x_t : Temperatura máxima.
- B : Operador de rezago, tal que $By_t = y_{t-1}$.
- $(1 - B)$: Diferenciación regular de orden $d = 1$, para eliminar tendencia.
- $(1 - B^7)$: Diferenciación estacional de orden $D = 1$, con periodicidad semanal ($s = 7$).
- $(1 - 0,5573B)$: Componente autorregresiva (AR) de orden 1.

- $(1 - 0,8539B - 0,0438B^2)$: Componente de medias móviles (MA) de orden 2.
- $(1 - 0,0141B^7)$: Componente estacional de medias móviles (MA estacional) de orden 1.
- ε_t : Término de error aleatorio (ruido blanco).

Para evaluar el rendimiento del modelo se consideró la evaluación de supuestos en el cuadro 6

Test	p-valor	Cumple
Autocorrelación	0.660	Sí
Normalidad	$2,71 \times 10^{-46}$	No
Homocedasticidad	$3,24 \times 10^{-7}$	No

Cuadro 6: Supuestos del modelo SARIMAX para la generación de energía solar

Los resultados presentados en la tabla anterior evidencian que el modelo SARIMAX ajustado para la predicción de la generación de energía solar cumple con el supuesto de ausencia de autocorrelación en los residuos, según el test de Ljung–Box (p-valor = 0.660). No obstante, no se satisfacen los supuestos de normalidad ni de homocedasticidad, ya que los p-valores obtenidos en los test de Normalidad ($2,71 \times 10^{-46}$) y Homocedasticidad ($3,24 \times 10^{-7}$) son significativamente menores al nivel de significancia del 5 %, indicando una distribución asimétrica de los residuos y presencia de heterocedasticidad.

Por otra parte, el modelo con mejores métricas consiste en la predicción de generación de energía solar en base a la presión a nivel del mar, el cual podemos visualizar en la ecuación 3.3.

$$(1 - 0,5908B)(1 - B)(1 - B^7)y_t = (1 + 1,2075B + 0,2273B^2)(1 - 0,9415B^7)\varepsilon_t + 12,1618 \cdot x_t \quad (0.8)$$

Donde:

- y_t : Generación de energía solar (Mw) en el tiempo t .
- B : Operador de rezago, tal que $By_t = y_{t-1}$.
- $(1 - B)$: Diferenciación regular de orden $d = 1$, aplicada para remover tendencia.
- $(1 - B^7)$: Diferenciación estacional de orden $D = 1$, con periodicidad semanal ($s = 7$).
- $(1 - 0,5908B)$: Componente autorregresiva (AR) de orden 1.
- $(1 + 1,2075B + 0,2273B^2)$: Componente de medias móviles (MA) de orden 2.
- $(1 - 0,9415B^7)$: Componente estacional de medias móviles (MA estacional) de orden 1.
- ε_t : Término de error aleatorio (ruido blanco).
- x_t : Variable exógena en el tiempo t , correspondiente a la presión promedio.
- 12,1618: Coeficiente estimado que mide el impacto inmediato de la presión promedio sobre la generación de energía solar.

Para evaluar el modelo, se evaluaron los supuestos correspondientes, cuyos resultados se encuentran en el cuadro 7.

Test	p-valor	Cumple
Autocorrelación	$4,24 \times 10^{-22}$	No
Normalidad	0	No
Homocedasticidad	$3,97 \times 10^{-14}$	No

Cuadro 7: Supuestos del modelo SARIMAX para la generación de energía solar

Los resultados de la tabla anterior muestran que el modelo SARIMAX ajustado para la generación de energía solar no cumple con ninguno de los supuestos estadísticos evaluados. El test de Ljung–Box reporta un p-valor extremadamente bajo ($4,24 \times 10^{-22}$), lo que indica la presencia significativa de autocorrelación en los residuos. Asimismo, el test de normalidad retorna un p-valor igual a cero, sugiriendo que los residuos no se distribuyen normalmente. Por último, el test de homocedasticidad ARCH–LM evidencia heterocedasticidad en los errores (p-valor $3,97 \times 10^{-14}$), indicando que la varianza de los residuos no es constante en el tiempo. En conjunto, estos resultados sugieren que, si bien el modelo puede tener valor predictivo en términos numéricos, sus residuos violan supuestos fundamentales, lo cual limita la validez de inferencias estadísticas.

Por última como comparativa para analizar si ambos modelo generan una buena predicción sobre la variable generación de energía solar, al momento de cambiar la variable predictora, se utilizó el RMSE y MAE a visualizar en el cuadro 8.

Variable	RMSE	MAE
Temperatura Máxima	3.540	2.484
Presión Promedio	2.421	1.98

Cuadro 8: Métricas de error para el modelo VAR(7) en el pronóstico de 30 días (RMSE y MAE).

Los resultados presentados en la anterior tabla muestran que el modelo VAR(7) logra un desempeño aceptable en el pronóstico de las variables meteorológicas durante el periodo de 30 días proyectado. La presión promedio obtuvo los menores valores de error, con un RMSE de 2.421 y un MAE de 1.98, lo que indica un ajuste robusto y consistente, con errores promedio relativamente bajos. Por su parte, la temperatura máxima presenta errores levemente mayores (RMSE = 3.540), lo que sugiere una mayor variabilidad o dificultad para capturar su dinámica a través del modelo VAR. En conjunto, estas métricas reflejan que el modelo es capaz de representar adecuadamente el comportamiento de corto plazo de las variables climáticas, siendo más preciso en aquellas con menor variabilidad estructural. Sin embargo, en ambos casos no hubo cumplimiento de supuesto, lo que indica que es un modelo predictivo no significativo para el caso.

Conclusión

El presente estudio permitió modelar y pronosticar tanto variables meteorológicas como la generación de energía solar en la planta fotovoltaica Sol del Desierto, utilizando dos enfoques principales: modelos VAR para variables climáticas interrelacionadas, y modelos SARIMAX para incorporar efectos exógenos en la predicción de generación.

Primero, el modelo VAR(7) demostró capacidad para capturar la dinámica conjunta de las variables meteorológicas más relevantes, en particular la temperatura promedio y la presión atmosférica. Este modelo se ajustó tras confirmar la estacionariedad de las series mediante el test de Dickey-Fuller, y la selección óptima del rezago se basó en el criterio AIC. Se destaca un mejor ajuste en las variables menos volátiles como la temperatura promedio al evaluar RMSE y MAE de los ajustes. Además, los supuestos del modelo se cumplieron en su mayoría, salvo en la variable temperatura máxima, cuyo residuo no presentó distribución normal. Esto refuerza la validez general del modelo VAR como herramienta de pronóstico meteorológico multivariado.

En cuanto a la predicción de generación de energía, se exploraron modelos SARIMAX con dos variables exógenas distintas: temperatura máxima y presión promedio. Ambos modelos tuvieron distinta estructura estacional y autorregresiva, y presentaron diferencias relevantes tanto en desempeño como en validez estadística. El modelo con presión promedio mostró métricas de error inferiores, indicando mayor precisión en el ajuste; sin embargo, no cumplió con los supuestos de normalidad, homocedasticidad ni ausencia de autocorrelación, lo que limita su aplicabilidad en análisis inferenciales, indicando que sus resultados no son significativos para la variable a predecir. En cambio, el modelo con temperatura máxima presentó supuestos más aceptables, aunque con un desempeño predictivo ligeramente inferior.

En resumen, los resultados sugieren que:

- El modelo VAR(7) es adecuado para pronosticar condiciones meteorológicas locales en entornos áridos, y sus predicciones pueden ser utilizadas como insumo para modelos de predicción energética.
- El modelo SARIMAX con presión promedio como covariable posee mejor capacidad predictiva, pero presenta problemas en los supuestos del modelo, especialmente en la estructura de los residuos.
- El modelo SARIMAX con temperatura máxima, si bien menos preciso, entrega resultados más consistentes desde un punto de vista estadístico.

El desarrollo del presente trabajo enfrentó diversas limitaciones que restringieron el alcance de los modelos implementados. En primer lugar, no fue posible trabajar con series temporales de alta resolución (datos horarios), lo cual habría permitido capturar con mayor precisión la variabilidad intradía de las variables meteorológicas y de la generación de energía. Esta limitación se debió a restricciones de tipo computacional, tanto en capacidad de procesamiento como en almacenamiento, que imposibilitaron el manejo eficiente de bases de datos más voluminosas.

En segundo lugar, una variable meteorológica clave como la radiación solar no pudo ser incorporada en el modelamiento, debido a la ausencia de datos actualizados y completos para el periodo de análisis. Si bien su inclusión habría sido deseable dado su vínculo directo con la generación fotovoltaica, la opción de predecirla previamente y usarla como covariable fue descartada, dado que ello excedía la capacidad computacional y el diseño metodológico definido para el presente estudio.

Estas limitaciones implican que, si bien los modelos construidos ofrecen resultados interpretables y razonables en función de las variables disponibles, existen factores relevantes no considerados que podrían mejorar la precisión y robustez de las predicciones en investigaciones futuras.

Por tanto, para próximos estudios se sugiere incorporar nuevas variables predictoras para el estudio, siendo recomendable complementar los modelos SARIMAX con modelos alternativos (e.g., redes neuronales recurrentes o TCN) que sean capaces de captar una mayor volatilidad de los datos, para obtener una predicción más precisa.

Referencias

- [1] Fara, J., Moya, Y., Mancha, A., & García, M. (2021). Forecasting of energy production for photovoltaic systems based on ARIMA and artificial neural network advanced models. *International Journal of Renewable Energy Development*, 10(4), 883–893.
- [2] Benítez, I. B., Ibañez, J. A., Lumabad, C. I. D., Cañete, J. M., & Principe, J. A. (2023). Day-Ahead Hourly Solar Photovoltaic Output Forecasting Using SARIMAX, Long Short-Term Memory, and Extreme Gradient Boosting: Case of the Philippines. *Energies*, 16(23), 7823. <https://doi.org/10.3390/en16237823>
- [3] Fabara, J. R., Tenreiro Oeiras, R. A., Palomo Cobas, H., Conde Long, H., Gómez Delgado, C. G., &

Rodríguez Patiño, J. P. (2019). Predicción de la generación para un sistema fotovoltaico mediante la aplicación de técnicas de minería de datos. En *Actas de las XIX Jornadas de Ingeniería de Sistemas* (pp. 916–920).

[4] Herrera Fernández, F. B., Limonte Ruiz, A. A., Álvarez Morales, M., & García Tamayo, J. G. (2023). Sistema para predicción de la generación en bloques de plantas fotovoltaicas. *Ingeniería Energética*, 44(3), 73–79.

[5] Soria, J. J., Poma, O., Sumire, D. A., Rojas, J. H. F., Chipa, S. M. R. (2022). Multiple Linear Regression Model of Environmental Variables, Predictors of Global Solar Radiation in the Area of East Lima, Peru. *IOP Conference Series: Earth and Environmental Science*, 1006(1), 012009. <https://doi.org/10.1088/1755-1315/1006/1/012009>

[6] Martínez, D. R., Albín, J., Cabaleiro, J., Peña, T., Rivera, F., Blanco, V. (2009). El Criterio de Información de Akaike en la Obtención de Modelos Estadísticos de Rendimiento. En *XX Jornadas de Paralelismo*.

[7] Peña, T., Ibañez, J. A., Cabrera, F., & González, A. I. (2022). Self-consumption analysis of different photovoltaic technologies under real conditions: A case study. *IOP Conference Series: Earth and Environmental Science*, 1006(1), 012009. <https://doi.org/10.1088/1755-1315/1006/1/012009>

Amarilla, C. D. (2025). Modelos de regresión simple para variables radiométricas en horas bajo cielo despejado en la ciudad de San Lorenzo – Paraguay. *Reportes Científicos de la FACEN*, 15(2), 3–19. <https://doi.org/10.18004/rcfacen.2024.15.2.003>

Espinosa-Ramírez, B. H. A., Garrido-Hernández, A., García-Domínguez, G., Vargas-León, E. A., & Castillo-Minjarez, J. M. A. (2023). Efecto de la temperatura en la eficiencia de paneles fotovoltaicos. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, 11(Especial 5). <https://doi.org/10.29057/icbi.v11iEspecial5.11841>

Khouili, O. (2025). Evaluating the impact of deep learning approaches on solar and photovoltaic power forecasting: A systematic review. *Energy Strategy Reviews*, 59, Art. 101735. <https://doi.org/10.1016/j.esr.2025.101735>

Palomo, W., Quinatoa, C., Mullo, M., & Castillo, J. (2025). Predicción de radiación solar en sistemas fotovoltaicos utilizando técnicas de aprendizaje automático. *CONECTIVIDAD*, 6(1), 338–355. <https://doi.org/10.37431/conectividad.v6i1.196>

Song, Z., Xiao, F., Chen, Z., & Madsen, H. (2025). Probabilistic ultra-short-term solar photovoltaic power forecasting using natural gradient boosting with attention-enhanced neural networks. *Energy and AI*, 20, Article 100496. <https://doi.org/10.1016/j.egyai.2025.100496>

Zhao, M., Yu, R., Chang, C., Bao, D., Mei, A., Liu, Y., & Wang, N. (2023). Effect of sand and dust shading on the output characteristics of solar photovoltaic modules in desertification areas. *Energies*, 16(23), 7910. <https://doi.org/10.3390/en16237910>