



BUAP

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

Facultad de Ingeniería

Secretaría de Investigación y Estudios de Posgrado

**MODELO PREDICTIVO DE GENERACIÓN
SOLAR FOTOVOLTAICA DE CORTO PLAZO**

TESIS

Que para obtener el grado de

**MAESTRO EN INGENIERÍA CON OPCIÓN TERMINAL EN
SISTEMAS ELÉCTRICOS DE POTENCIA**

Presenta:

JOSÉ GUADALUPE BAHENA VÁZQUEZ

Director de tesis:

M.I Marco Antonio Correa Nieto

Co director de tesis:

Dr. José Isrrael Rodríguez Mora

Índice

Capítulo 1	8
1.1 Introducción.....	8
1.2 Planteamiento del problema.....	8
1.3 Justificación.....	9
1.4 Objetivos	10
1.5 Alcances y metas	10
Capítulo 2 Marco Referencial	11
2.1 Sistema Eléctrico de Potencia.....	11
2.2 Capacidad instalada de las centrales eléctricas del MEM	13
2.3 Principales centrales eléctricas del MEM	16
2.4 Generalidades de pronóstico de generación solar	17
2.5 Supuestos de la regresión lineal	20
2.5.1 Linealidad	20
2.5.2 Media de los residuos.....	20
2.5.3 Homocedasticidad	20
2.5.4 Normalidad	21
2.5.5 Autocorrelación de residuos	21
2.5.6 Multicolinealidad	21
2.6 Correlación de variables	22
2.7 Modelo de regresión lineal simple	22
2.8 Estimación de los parámetros por Mínimos Cuadrados	24
2.8 Error cuadrático medio MSE	26
2.9 Coeficiente de determinación R²	26
2.10 Tratamiento de datos.....	27
2.11 Tipos de análisis de datos	29
2.12 Python	29
2.12.1 Colab	30
2.12.2 Librería PANDAS	31
2.12.3 Librería Matplotlib	31
2.12.4 Librería NUMPY	32

2.12.5	Librería Seaborn	32
2.13	Variables meteorológicas de influencia	32
Capítulo 3 Desarrollo		36
3.1	Obtener datos de entrada	37
3.2	Tratamiento de datos de entrada	38
3.3	Desarrollo de software para obtener supuestos de regresión lineal	39
3.4	Obtener y analizar resultados de los supuestos.....	40
3.5	Formación de grupos con los datos de entrada	42
3.6	Desarrollo de software utilizando regresión lineal	45
3.7	Obtener y analizar modelos de pronóstico de generación	46
Capítulo 4 Resultados		48
4.1	Resultados de linealidad	48
4.2	Resultado de la media de los residuos	48
4.3	Resultado de homocedasticidad	48
4.4	Distribución de los datos	49
4.5	Autocorrelación de los residuos	50
4.6	Correlación de las variables	50
4.7	Regresión lineal para el conjunto de datos hora-anual	51
4.8	Resultados de la regresión lineal grupo horas-anual	53
4.9	Regresión lineal para el conjunto de datos hora-mensual	55
4.10	Resultados de la regresión lineal grupo horas-mensual	69
4.11	Regresión lineal para el conjunto de datos hora-estación.....	76
4.12	Resultados de la regresión lineal grupo horas-estación.....	80
4.13	Aplicación del modelo	83
Capítulo 5 Conclusiones.....		86
Capítulo 6 Recomendaciones.....		88
Bibliografía.....		89

Índice de Figuras

<i>Figura 1 Sistema Eléctrico Nacional.</i>	11
<i>Figura 2 Conformación del Sistema Eléctrico Nacional.</i>	12
<i>Figura 3 Capacidad instalada interconectada de la CFE y del resto de los permisionarios (MW).</i>	14
<i>Figura 4 Adiciones de Capacidad instalada en 2023 (MW).</i>	14
<i>Figura 5 Capacidad de las centrales en pruebas de la CFE y del resto de los permisionarios (MW).</i>	15
<i>Figura 6 Porcentaje de la capacidad instalada por tipo de tecnología.</i>	15
<i>Figura 7 Principales centrales eléctricas de la CFE y PIE.</i>	16
<i>Figura 8 Principales Centrales Eléctricas de la CFE y PIE.</i>	16
<i>Figura 9 Categorías para el horizonte de pronóstico.</i>	17
<i>Figura 10 Diagrama a bloques del modelo de persistencia.</i>	18
<i>Figura 11 Diagrama a bloques del pronóstico análogo.</i>	18
<i>Figura 12 Diagrama a bloques sobre el proceso de pronóstico estadístico.</i>	19
<i>Figura 13 Diagrama a bloques sobre el proceso de pronóstico físico.</i>	19
<i>Figura 14. Correlación de rangos de Sperman's.</i>	34
<i>Figura 15 Comparación de modelos de regresión para pronóstico.</i>	35
<i>Figura 16 Diagrama de flujo del desarrollo del proyecto.</i>	36
<i>Figura 17 Código fuente para importar datos en Colab.</i>	39
<i>Figura 18 Código fuente para identificar si los datos presentan linealidad.</i>	40
<i>Figura 19 Código fuente para determinar la media de los residuos.</i>	41
<i>Figura 20 Código fuente para determinar la homocedasticidad de los datos.</i>	41
<i>Figura 21Código fuente para determinar si los residuos presentan una distribución normal.</i>	41
<i>Figura 22 Código fuente para determinar si los residuos presentan autocorrelación.</i>	42
<i>Figura 23 Código fuente para determinar la correlación de las variables.</i>	42
<i>Figura 24 Formación de grupos para su análisis.</i>	42
<i>Figura 25 Código fuente para la división de datos e inicio de la regresión lineal.</i>	46
<i>Figura 26 Código fuente para el cálculo de predicciones y métricas del conjunto de datos.</i>	46
<i>Figura 27 Código fuente para la creación de una tabla de resultados y sus gráficos.</i>	47
<i>Figura 28 Código fuente graficar el resultado del pronóstico por hora por día o de corto plazo.</i>	47
<i>Figura 29 Resultado de la prueba de linealidad.</i>	48
<i>Figura 30 Resultado de la prueba de media de los residuos.</i>	48
<i>Figura 31 Resultado de la prueba de homocedasticidad.</i>	49
<i>Figura 32 Resultado de la distribución de los residuos.</i>	49
<i>Figura 33 Resultado de autocorrelación.</i>	50

Figura 34 Resultado de correlación.	51
Figura 35 Resultado de la regresión de la hora 7 a la 19.	52
Figura 36 Gráfico de pronóstico para el grupo de datos horas-anual con datos de prueba.	54
Figura 37 Gráfico de R^2 para datos de entrenamiento y prueba del conjunto horas-anual.	54
Figura 38 Gráfico de MSE para datos de entrenamiento y prueba del conjunto horas-anual.	55
Figura 39 Resultado de la regresión de la hora 8 a la 18 para el mes de enero.	56
Figura 40 Resultado de la regresión de la hora 8 a la 19 para el mes de febrero.	57
Figura 41 Resultado de la regresión de la hora 8 a la 19 para el mes de marzo.	58
Figura 42 Resultado de la regresión de la hora 7 a la 19 para el mes de abril.	59
Figura 43 Resultado de la regresión de la hora 7 a la 19 para el mes de mayo.	61
Figura 44 Resultado de la regresión de la hora 7 a la 19 para el mes de junio.	62
Figura 45 Resultado de la regresión de la hora 7 a la 19 para el mes de julio.	63
Figura 46 Resultado de la regresión de la hora 7 a la 19 para el mes de agosto.	64
Figura 47 Resultado de la regresión de la hora 7 a la 19 para el mes de septiembre.	66
Figura 48 Resultado de la regresión de la hora 7 a la 18 para el mes de octubre.	67
Figura 49 Resultado de la regresión de la hora 8 a la 18 para el mes de noviembre.	68
Figura 50 Resultado de la regresión de la hora 8 a la 18 para el mes de diciembre.	69
Figura 51 Gráfico de pronóstico para el grupo de datos horas-mensual con datos de prueba.	71
Figura 52 Gráfico de pronóstico para el grupo de datos horas-mensual con datos de prueba julio a diciembre.	72
Figura 53 Gráfico de R^2 para datos de entrenamiento y prueba del conjunto horas-mensual.	73
Figura 54 Gráfico de R^2 para datos de entrenamiento y prueba del conjunto horas-mensual (continuación).	74
Figura 55 Gráfico de MSE para datos de entrenamiento y prueba del conjunto horas-anual.	75
Figura 56 Gráfico de MSE para datos de entrenamiento y prueba del conjunto horas-anual (continuación).	75
Figura 57 Resultado de la regresión de la hora 7 a la 19 para la estación primavera.	77
Figura 58 Resultado de la regresión de la hora 7 a la 19 para la estación verano.	78
Figura 59 Resultado de la regresión de la hora 7 a la 18 para la estación otoño.	79
Figura 60 Resultado de la regresión de la hora 8 a la 19 para la estación invierno.	80
Figura 61 Gráfico de pronóstico para el grupo de datos horas-mensual con datos de prueba.	82
Figura 62 Gráfico de R^2 para datos de entrenamiento y prueba del conjunto horas-mensual.	82
Figura 63 Gráfico de MSE para datos de entrenamiento y prueba del conjunto horas-anual.	83
Figura 64. Pronóstico de generación: 6 de febrero 2024 parque 1.	84
Figura 65 Pronóstico de generación: 17 de abril 2024 parque 1.	84
Figura 66 Pronóstico de generación: 24 de febrero 2024 parque 2.	84
Figura 67 Pronóstico de generación: 26 de abril 2024 parque 4.	85
Figura 68 Pronóstico de generación: 6 de enero 2024 parque 5.	85

Índice de Tablas

<i>Tabla 1. Resultados de la publicación.....</i>	<i>33</i>
<i>Tabla 2. Coeficiente de correlación de los rangos de Sperman's.</i>	<i>34</i>
<i>Tabla 3. Resultados de la correlación de Pearson y Spearman's.</i>	<i>35</i>
<i>Tabla 4. Variables meteorológicas de la ubicación de un parque de generación solar fotovoltaico.</i>	<i>37</i>
<i>Tabla 5. Datos de generación eléctrica de parque solar fotovoltaico.</i>	<i>38</i>
<i>Tabla 6. Variables meteorológicas y generación eléctrica.</i>	<i>39</i>
<i>Tabla 7. Grupo horas-Anual para la hora 8 de cada día del mes de enero 2023.....</i>	<i>43</i>
<i>Tabla 8. Grupo horas-Mensual para la hora 12 de cada día del mes de junio 2023.</i>	<i>44</i>
<i>Tabla 9. Grupo horas-Estación para primavera 2023 (extracto) hora uno.....</i>	<i>45</i>
<i>Tabla 10. Tabla de resultados con las ecuaciones para cada hora en el grupo de datos hora-anual.</i>	<i>53</i>
<i>Tabla 11. Tabla de resultados con las ecuaciones para cada hora en el grupo de datos hora-mes.....</i>	<i>69</i>
<i>Tabla 12. Tabla de resultados con las ecuaciones para cada hora en el grupo de datos hora-mes continuación.</i>	<i>70</i>
<i>Tabla 13. Resultados con las ecuaciones para cada hora en el grupo de datos hora-estación.....</i>	<i>81</i>

Glosario de términos

CENACE	Centro Nacional de Control de Energía.
GCROR	Gerencia de Control Regional Oriental.
IEEE	Instituto de Ingenieros Eléctricos y Electrónicos.
LIE	Ley de Infraestructura Eléctrica.
PIE	Productores Independientes de Energía.
PND	Plan Nacional de Desarrollo.
PRODESEN	Programa de Desarrollo del Sistema Eléctrico Nacional.
RGD	Red General de Distribución.
RNT	Red Nacional de Transmisión.
SEN	Sistema Eléctrico Nacional.
SENER	Secretaría de Energía.

Capítulo 1

1.1 Introducción

El presente proyecto aborda el pronóstico de generación de energía eléctrica para parques solares fotovoltaicos el ámbito de la Gerencia de Control Regional Oriental (GCROR) de México, a través del diseño de un modelo predictivo u algoritmo que involucra el comportamiento de variables meteorológicas.

Los datos de entrada para el modelo se obtienen en colaboración con CENACE (Centro Nacional de Control de Energía), estas variables son la temperatura ambiente, sensación térmica, índice de calor, precipitación, humedad relativa, velocidad del viento, nubosidad, irradiancia solar, así como la generación en unidades MWh/h para el año 2023.

Se implementa un modelo de correlación de variables para identificar asociación respecto a la variable generación, posteriormente el modelo se analiza con datos de entrenamiento y prueba con el método estadístico de regresión lineal.

Al final del proyecto se espera que el modelo de predicción u algoritmo sugerido se implemente en las actividades diarias dentro de la Gerencia de Control Oriental de CENACE, así como la posible publicación de resultados.

1.2 Planteamiento del problema

En México, dentro de la planeación del sector eléctrico se debe cumplir con los compromisos internacionales en materia ambiental para la reducción de emisiones y mitigación del cambio climático, por lo que en el PRODESEN contempla el incremento ordenado de la generación eléctrica con energías limpias y renovables para el cumplimiento de las metas establecidas en materia de Energías Limpias, el futuro nos demanda que hagamos hoy un uso más racional y sostenible de todos nuestros recursos energéticos y de todas las tecnologías disponibles para el desarrollo nacional [1].

La energía alternativa se puede definir como cualquier fuente de energía que representa una alternativa a los combustibles fósiles. La energía renovable se puede definir como una fuente que es regenerada o restablecida en forma natural. La mayoría de las fuentes de energía alternativa se pueden definir también como renovables, por ejemplo, el sol, el viento y el agua [2].

A pesar de contar con un gran número de ventajas, el principal problema consiste en la intermitencia de los recursos para generar electricidad, la generación fotovoltaica presenta un mayor riesgo en este sentido, ya que depende de la irradiancia solar, la cual como sabemos, no está disponible durante todo el día, así como de las condiciones meteorológicas como la nubosidad, temperatura ambiente entre otras, lo que provoca grandes fluctuaciones en cuanto a la generación, requiere también de grandes superficies para la instalación de los paneles y un control o almacenamiento de energía para cuando se requiera.

El pronóstico de generación de parques fotovoltaicos impacta en el despacho de generación en cuanto a la operación en tiempo real y en aspectos del Mercado Eléctrico Mayorista en México.

1.3 Justificación

La evolución estimada de la capacidad instalada distribuida fotovoltaica según el Programa de Desarrollo del Sistema Eléctrico Nacional PRODESEN (2024-2038), para el escenario de planeación para el 2023 se estima un valor de 8,426 MW, lo que representa un crecimiento de 152.2 % respecto al cierre del 2023, así como un crecimiento promedio anual de 760 MW, por lo que es indispensable contar con modelos de pronóstico actualizados con las variables meteorológicas de la región que sirvan como herramientas para el predespacho de generación dentro del Mercado Eléctrico Mayorista en México [1].

El resultado de esta investigación apoyará a pronosticar la generación de energía eléctrica de parques solares fotovoltaicos en el ámbito de la Gerencia de Control Regional Oriental (GCROR) de México y su posible utilización por parte de CENACE en sus procesos de pronóstico.

1.4 Objetivos

4.12.2 Objetivo general

Desarrollar un modelo predictivo de generación solar fotovoltaica de corto plazo con históricos de variables meteorológicas para su utilización en los procesos de pronóstico de CENACE-GCROR.

4.12.2 Objetivos particulares

1. Recopilar datos históricos de generación de parques solares fotovoltaicos y de variables meteorológicas donde se encuentran ubicados.
2. Analizar los datos meteorológicos y de generación de los parques.
3. Realizar un estudio de correlación de variables.
4. Elaborar un modelo predictivo u algoritmo de generación solar fotovoltaica.
5. Evaluar el modelo de predicción con datos de entrenamiento, prueba y reales.
6. Posible publicación de resultados.

1.5 Alcances y metas

Al término de la investigación se espera cubrir los siguientes alcances:

- Un análisis de correlación entre variables meteorológicas y de generación eléctrica.
- Un modelo u algoritmo predictivo de generación solar para parques fotovoltaicos en el ámbito de la Gerencia de Control Regional Oriental de México.
- Una tesis de la investigación que integre los resultados obtenidos.
- Posible redacción de un artículo científico.

Capítulo 2 Marco Referencial

2.1 Sistema Eléctrico de Potencia

De acuerdo con el Instituto de Ingenieros Eléctricos y Electrónicos IEEE un sistema eléctrico de potencia es una red formada por unidades generadoras eléctricas, cargas y/o líneas de transmisión de potencia, incluyendo el equipo asociado, conectado eléctrica y mecánicamente a la red [3].

Por otra parte, el diccionario de término eléctricos y electrónicos de la IEEE define el sistema de potencia como las fuentes de potencia eléctrica, conductores y equipos requeridos para suplir la potencia eléctrica [3].

La Figura 1 muestra el Sistema Eléctrico Nacional, el cual está conformado por ocho Gerencias de Control Regionales donde la península de Baja California está compuesta por tres Sistemas Eléctricos que operan aislados entre sí, como se muestra en la Figura 2 [4].



Figura 1 Sistema Eléctrico Nacional.

La operación de estas regiones está bajo la responsabilidad de nueve Centros de Control Regional ubicados en las Ciudades de México, Puebla, Guadalajara, Mérida, Hermosillo, Gómez Palacio, Monterrey, Mexicali, La Paz y un pequeño centro de control en Santa Rosalía Baja California Sur, para el Sistema Interconectado Mulegé (SIM). El Centro Nacional en la Ciudad de México en conjunto con el Centro Nacional Alterno, ubicado en la Ciudad de Puebla coordinan el MEM y la operación segura y confiable del SEN [4]

El Sistema Interconectado nacional está integrado por las siete Gerencias de Control Regional: Central, Oriental, Occidental, Noroeste, Norte, Noreste y Peninsular. En ellas se comparten los recursos y reservas de capacidad ante la diversidad de demandas y situaciones operativas; esto permite el intercambio de energía para lograr un funcionamiento más económico y confiable en su conjunto [4].

El Sistema Interconectado Baja California (SIBC), opera interconectado a la Red Eléctrica de la región Oeste de EE. UU. (Western Electricity Coordinating Council, WECC) por medio de dos líneas de transmisión conectadas a un nivel de tensión de 230 kV en corriente alterna, mientras que los Sistemas Interconectados Baja California Sur (SIBCS) y SIM están eléctricamente aislados entre sí, así como del SIN y SIBC. [4].



Figura 2 Conformación del Sistema Eléctrico Nacional.

El SEN está constituido por Redes Eléctricas en diferentes niveles de tensión:

- Red Nacional de Transmisión (RNT): Sistema integrado por las Redes Eléctricas que se utilizan para transportar energía a las RGD y a los Usuarios Finales que por las características de sus instalaciones lo requieran, así como las interconexiones a los sistemas eléctricos extranjeros que determine la SENER. Incluye las tensiones iguales o mayores a 69 kV [4].
- Redes Generales de Distribución (RGD): Redes Eléctricas que se utilizan para distribuir energía eléctrica al público en general; están integradas por las Redes Eléctricas en media tensión, cuyo Suministro Eléctrico ocurre a niveles mayores a 1 kV y menores a 69 kV, así como las Redes Eléctricas en baja tensión en las cuales el Suministro Eléctrico es igual o menor a 1 kV [4].

2.2 Capacidad instalada de las centrales eléctricas del MEM

A continuación, se presentan la capacidad instalada interconectada a la RNT y las RGD de las Centrales Eléctricas pertenecientes a la CFE, Productores Independientes de Energía Eléctrica (PIE), Autoabastecimiento (AU), Cogeneración (COG), Pequeña Producción (PP), Importación (IMP), Exportación (EXP) y Usos Propios Continuos (UPC) al 31 de diciembre 2023. Se considera la infraestructura que fue construida al amparo de la anterior Ley del Servicio Público de Energía Eléctrica (LSPEE); así como, la Capacidad Instalada bajo el amparo de la LIE, relacionada con las Centrales Eléctricas de los Generadores Participantes del Mercado [4].

Al cierre del 2022 la Capacidad Instalada por parte de la CFE, los PIE y del resto de los permisionarios fue de 87,130 MW, sin considerar las Centrales Eléctricas en pruebas, mientras que al cierre de diciembre de 2023 se incrementó hasta 89,008 MW, lo cual refleja un incremento de 2.15 % con respecto al 2022. Este incremento se debe principalmente, a adiciones como Ciclo Combinado (766 MW), Fotovoltaicas (934 MW) y Eólicas (134 MW). Mientras que para las Centrales Eléctricas en pruebas se tiene 1,436 MW destacando las Centrales de Ciclo Combinado (394 MW), Centrales FV (426 MW),

las Centrales Eólicas (516 MW) y Centrales con otras tecnologías (100 MW) [4].

En el Figura 3 se presenta la capacidad instalada en el SEN por tecnología para el 2020, 2021, 2022 y 2023. En el año 2022, la Capacidad Instalada de las Centrales Eléctricas de Energía Limpia tales como Hidroeléctrica, Geotermoeléctrica, Eoloeléctrica, Fotovoltaica, Bioenergía, Nuclear y Cogeneración Eficiente fue de 31,369 MW, y al cierre de diciembre de 2023 se tiene 32,449 MW; un incremento del 3.4% con respecto al 2022, debido principalmente al aumento de las Centrales Eléctricas FV y EO [4].

TECNOLOGÍA	2020 ^{1/}	2021 ^{6/}	2022 ^{7/}	2023 ^{8/}
Hidroeléctrica	12,612	12,614	12,613	12,612
Geotermoeléctrica	951	976	976	976
Eoloeléctrica	6,504	6,977	6,921	7,055
Fotovoltaica*	5,149	5,955	6,535	7,469
Bioenergía ^{2/}	378	378	408	407
Suma limpia renovable	25,594	26,899	27,453	28,519
Nucleoeléctrica	1,608	1,608	1,608	1,608
Cogeneración Eficiente ^{3/}	2,305	2,305	2,308	2,322
Suma limpia no renovable	3,913	3,913	3,916	3,930
Total capacidad de energía limpia	29,506	30,812	31,369	32,449
Porcentaje	35.50	35.8	36.00	36.46
Ciclo combinado	31,948	33,640	34,413	35,178
Térmica convencional ^{4/}	11,809	11,793	11,343	11,300
Turbogás ^{5/}	3,545	3,744	3,815	3,888
Combustión interna	850	701	728	729
Carboeléctrica	5,463	5,463	5,463	5,463
TOTAL	83,121	86,153	87,130	89,008

Figura 3 Capacidad instalada interconectada de la CFE y del resto de los permisionarios (MW).

En la Figura 4 se presentan las adiciones de capacidad instalada en 2023, de las centrales eléctricas en operación [4].

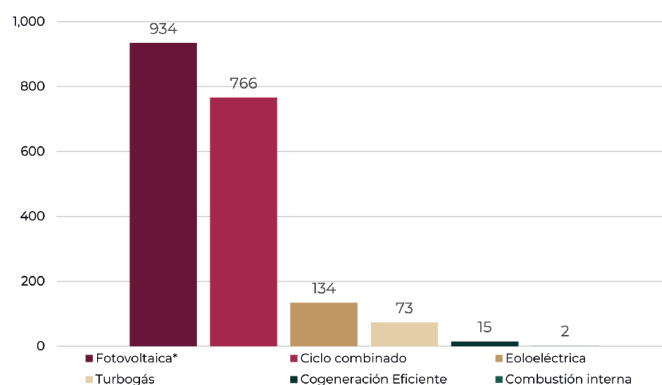


Figura 4 Adiciones de Capacidad instalada en 2023 (MW).

En la Figura 5 se presenta la capacidad de las centrales eléctricas en pruebas en el SEN por tipo de tecnología y Gerencia de Control Regional (GCR) al 31 de diciembre de 2023 [4].

TECNOLOGÍA	CEN	ORI	OCC	NOR	NTE	NES	PEN	SIMUL	TOTAL
Hidroeléctrica			12	3					15
Geotermoeléctrica									0
Eoloeléctrica						516			516
Fotovoltaica			306	120					426
Bioenergía ^v						2			2
Suma de Energía Limpia Renovable	0	0	318	123	0	518	0	0	959
Nucleoeléctrica									0
Cogeneración Eficiente ^v			6	12					18
Suma de Energía Limpia No Renovable	0	0	6	12	0	0	0	0	18
Total Capacidad de Energía Limpia	0	0	324	135	0	518	0	0	977
Porcentaje	0.00	0.00	47.86	100.00	0.00	94.63	0.00	0.00	67.93
Ciclo Combinado			344		50				394
Térmica Convencional ^{2/}									0
Turbogás							19		19
Combustión Interna	3		9			29		8	49
Carboeléctrica									0
TOTAL	3	0	677	135	50	547	19	8	1,439

Figura 5 Capacidad de las centrales en pruebas de la CFE y del resto de los permisionarios (MW).

La Figura 6 muestra el porcentaje la capacidad instalada por tipo de tecnología al 31 de diciembre del 2023 [4].

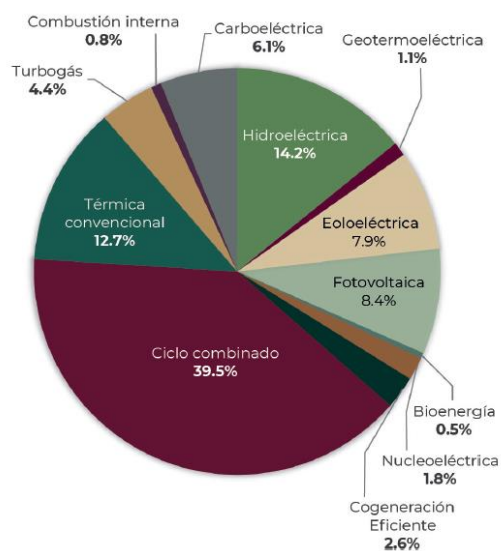


Figura 6 Porcentaje de la capacidad instalada por tipo de tecnología.

2.3 Principales centrales eléctricas del MEM

La Figura 7 muestra la ubicación de las Centrales Eléctricas de la CFE y los PIE que destacan por su tecnología o importancia regional al 31 de diciembre de 2023 [4].

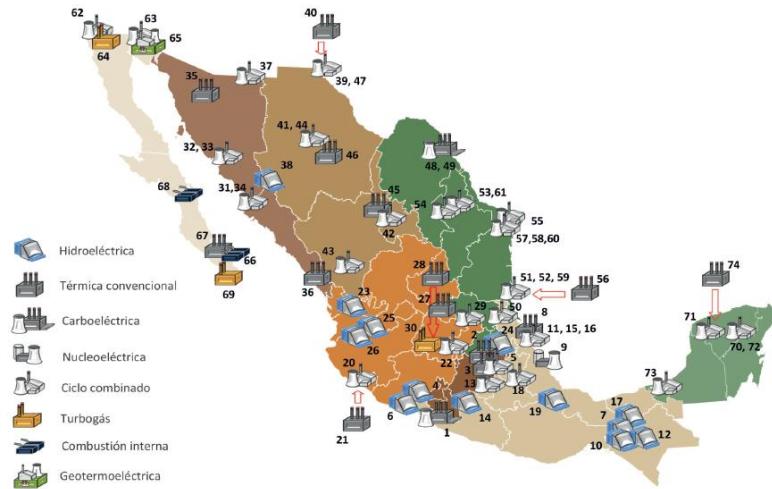


Figura 7 Principales centrales eléctricas de la CFE y PIE.

La Figura 8 muestra la ubicación de las principales centrales privadas al 31 de diciembre del 2023 [4].

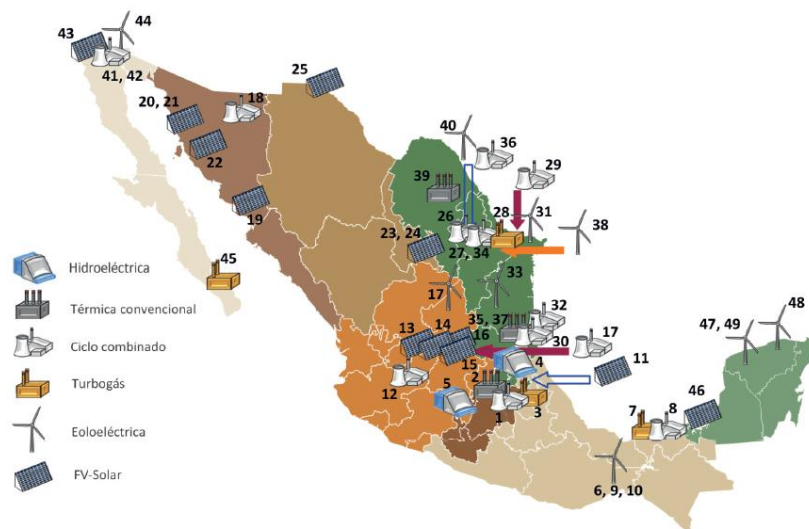


Figura 8 Principales Centrales Eléctricas de la CFE y PIE.

2.4 Generalidades de pronóstico de generación solar

La generación de las centrales fotovoltaicas depende de variables meteorológicas, por lo que es intrínsecamente variable. A lo largo del tiempo se han desarrollado distintos tipos de pronósticos que se clasifican de acuerdo con distintos criterios, como el horizonte de pronóstico, la metodología usada y/o el agente que los genera. De acuerdo con el horizonte, los pronósticos se pueden dividir en las siguientes categorías (ver Figura 9) [5].

PMP	Pronóstico de mediano plazo (PMP), el cual puede tener un horizonte desde 2 y hasta 20 días.
PCP	Pronóstico de corto plazo (PCP), el más común es el de día en adelante (en inglés day-ahead), con horizontes que pueden fluctuar entre 6 y 48 horas.
PMCP	Pronóstico de muy corto plazo (PMCP), o intradiario (en inglés intraday), que es un pronóstico con un horizonte que puede fluctuar desde minutos hasta 6 horas.

Figura 9 Categorías para el horizonte de pronóstico.

Los distintos tipos de pronósticos son usados por diferentes usuarios y para distintos fines. Los PMP usualmente son usados por los generadores, para decidir aspectos relacionados con la seguridad de suministro y la manutención de la red. El PCP es usado principalmente por el organismo encargado de la regulación o el operador de red, para prevenir eventuales congestiones del sistema, programar la generación y el seguimiento de carga. Los principales usuarios de los PMCP son los operadores de la red, por ejemplo, para la programación de generación de las unidades, la operación en tiempo real y las restricciones de transmisión. Además de lo anterior, existen otros usos para los pronósticos, especialmente de los PCP y PMCP. En aquellos casos en los que se cuenta con un mercado de energía, son un insumo importante para los operadores de la red y los generadores, en el momento de hacer transacciones de bloques de generación en el mercado [5].

Pronóstico de persistencia. Es el modelo de pronóstico más simple. En términos prácticos, se basa en el supuesto de que la energía generada en el futuro cercano será similar a la generada actualmente y/o en el pasado reciente. El pronóstico de persistencia no necesita más insumos que la energía generada en el último tiempo. Por la metodología, las desviaciones aumentan con el horizonte del pronóstico, y tiene cierta validez para un PMCP. Este modelo es usado actualmente, tanto por operadores de la red como de las centrales fotovoltaicas cuando no existe otro tipo de pronóstico ver Figura 10 [5].



Figura 10 Diagrama a bloques del modelo de persistencia

Pronóstico análogo. Este modelo de pronósticos analiza variables meteorológicas recientes, y buscan en un historial un evento igual o muy similar. Así como en el método de persistencia, sólo se necesita un conjunto de datos recientes de la generación. Se utiliza un coeficiente de correlación para poder encontrar aquellos fenómenos históricos que se le asemejen. Con esto se pueden hacer diferentes tipos de regresiones o métodos estadísticos para hacer el pronóstico, ver Figura 11 [5].



Figura 11 Diagrama a bloques del pronóstico análogo.

Pronóstico estadístico. Un pronóstico estadístico analiza relaciones entre variables, a partir de información histórica. Uno de los métodos estadísticos que ha presentado mayor éxito en la generación de pronósticos es el generado a partir de las Redes Neuronales Artificiales (por sus siglas en inglés ANN). A menudo, las ANN se describen como algoritmos biológicos que trabajan imitando el funcionamiento del cerebro humano. A partir de un proceso de entrenamiento basado en información histórica, son capaces de encontrar las relaciones entre la variable de interés y otras variables, lo que permite hacer pronósticos, ver Figura 12 [5].



Figura 12 Diagrama a bloques sobre el proceso de pronóstico estadístico.

Pronóstico físico. Transforma el pronóstico meteorológico generado con un NWP (Numerical Weather Prediction) en un pronóstico de energía, a partir de la dependencia de la variable de potencia de las tecnologías –aerogeneradores o paneles fotovoltaicos - de las variables meteorológicas, declaradas por el fabricante, ver Figura 13 [5].

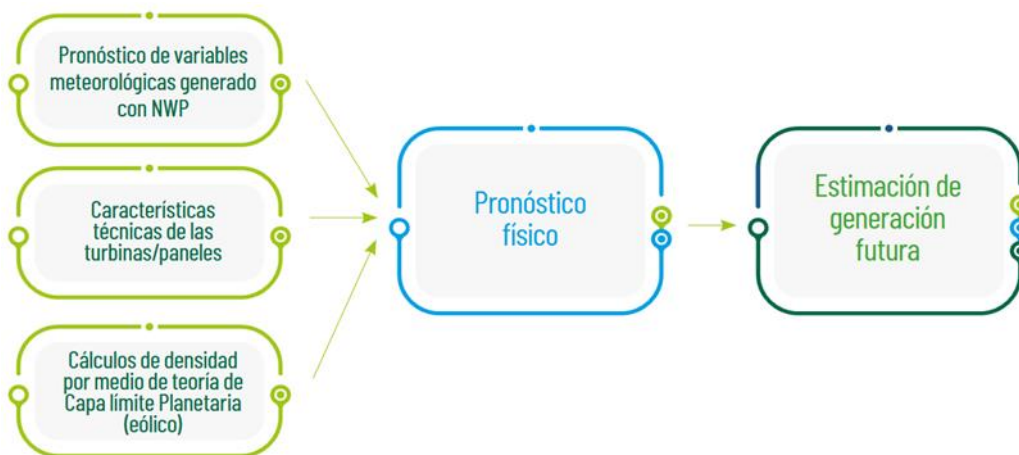


Figura 13 Diagrama a bloques sobre el proceso de pronóstico físico.

2.5 Supuestos de la regresión lineal

Estos supuestos proporcionan las bases para interpretar correctamente los resultados y las conclusiones del modelo.

La importancia de cumplir con estos supuestos radica en garantizar la validez de las inferencias estadísticas y las conclusiones derivadas del modelo de regresión. Además, el incumplimiento de estos supuestos puede afectar la capacidad predictiva del modelo y conducir a estimaciones sesgadas y poco fiables de los coeficientes de regresión. Es esencial realizar pruebas de diagnóstico para evaluar la violación de estos supuestos y, si es necesario, aplicar técnicas de corrección o considerar modelos alternativos.

2.5.1 Linealidad

Este supuesto implica que la relación entre las variables independientes y la variable dependiente debe ser lineal. Si la relación es no lineal, los resultados de la regresión (lineal) pueden ser poco confiables y conducir a interpretaciones erróneas sobre la relación entre las variables [6].

2.5.2 Media de los residuos

Los residuos, son las diferencias entre el valor verdadero y el valor predicho. Una de las suposiciones de la regresión lineal es que la media de los residuos debería ser cero [6].

2.5.3 Homocedasticidad

Este supuesto implica que la varianza de los errores debe ser constante en todos los niveles de las variables predictoras. Cuando se viola este supuesto, se produce heterocedasticidad, lo que significa que la dispersión de los errores varía en diferentes rangos de las variables predictoras. La presencia de heterocedasticidad puede

distorsionar los intervalos de confianza y los valores p-value, lo que puede afectar la precisión de las pruebas de hipótesis [6].

2.5.4 Normalidad

El supuesto de normalidad establece que los errores de la regresión deben seguir una distribución normal. Cuando este supuesto se cumple, las pruebas de hipótesis y los intervalos de confianza pueden interpretarse con mayor precisión. Si la normalidad no se cumple, los intervalos de confianza y las pruebas de hipótesis pueden verse afectados, lo que puede conducir a conclusiones erróneas [6].

2.5.5 Autocorrelación de residuos

El supuesto de independencia indica que los errores de la regresión no deben estar correlacionados entre sí. Si hay autocorrelación presente, puede afectar la precisión de los coeficientes y las pruebas de hipótesis, lo que lleva a conclusiones erróneas sobre la importancia de las variables predictoras. La presencia de autocorrelación en los residuos indica que los errores del modelo muestran cierto patrón sistemático en su distribución a lo largo del tiempo. Recordad que los errores o residuos de un modelo de regresión deberían distribuirse de manera aleatoria y seguir una distribución normal con media cero y varianza constante [6].

2.5.6 Multicolinealidad

En regresión, la multicolinealidad se refiere al grado en que las variables independientes están correlacionadas. La multicolinealidad afecta los coeficientes y los valores p, pero no influye en las predicciones, la precisión de las predicciones y las estadísticas [6].

2.6 Correlación de variables

La correlación es una medida de la relación (covariación) lineal entre dos variables cuantitativas continuas (x , y). La manera más sencilla de saber si dos variables están correlacionadas es determinar si covarían (varían conjuntamente). Es importante hacer notar que esta covariación no implica necesariamente causalidad, la correlación puede ser fortuita, como en el caso clásico de la correlación entre el número de venta de helados e incendios, debido al efecto de una tercera variable, la temperatura ambiental [7].

La correlación es en esencia una medida normalizada de asociación o covariación lineal entre dos variables. Esta medida o índice de correlación “ r ” puede variar entre -1 y +1, ambos extremos indicando correlaciones perfectas, negativa y positiva respectivamente. Un valor de $r = 0$ indica que no existe relación lineal entre las dos variables. Una correlación positiva indica que ambas variables varían en el mismo sentido. Una correlación negativa significa que ambas variables varían en sentidos opuestos. Lo interesante del índice de correlación es que “ r ” es en sí mismo una medida del tamaño del efecto, que suele interpretarse de la siguiente manera [7]:

- correlación **despreciable**: $r < |0.1|$
- correlación **baja**: $|0.1| < r \leq |0.3|$
- correlación **mediana**: $|0.3| < r \leq |0.5|$
- correlación **fuerte o alta**: $r > |0.5|$

2.7 Modelo de regresión lineal simple

El análisis de regresión es una técnica estadística que se emplea para investigar y modelar la relación entre variables, comúnmente a través de gráficos o diagramas de dispersión [8].

La ecuación (1) se llama modelo de regresión lineal. Por común acuerdo se dice que la “ x ” es la variable independiente y “ y ” la variable dependiente. Sin embargo, eso causa

confusión con el concepto de la dependencia estadística, así que llamaremos a la “ x ” como la variable predictora o regresora y “ y ” la variable de respuesta, como la ecuación (1) solo tiene una variable regresora, se llama modelo de regresión lineal simple [8].

$$y = \beta_0 + \beta_1 x \quad (1)$$

Donde

β_0 Ordenada al origen

β_1 Pendiente

Un objetivo importante del análisis de regresión es estimar los parámetros desconocidos en el modelo de regresión. También se le llama a este proceso ajuste del modelo a los datos [8].

Un aspecto esencial del análisis de regresión es la recolección, recopilación o adquisición de datos. Todo análisis de regresión es tan bueno como lo son los datos sobre los que se basa. Hay tres métodos básicos de recolectar datos:

- Un estudio retrospectivo basado en datos históricos.
- Un estudio observacional.
- Un experimento diseñado.

Un buen esquema de recolección de datos puede asegurar un análisis simplificado y un modelo de aplicación más general. Un mal esquema de recolección de datos induce a graves problemas en el análisis y su interpretación [8].

Los modelos de regresión se usan con varios fines, que incluyen los siguientes:

1. Descripción de datos.
2. Estimación de parámetros.
3. Predicción y estimación.
4. Control.

2.8 Estimación de los parámetros por Mínimos Cuadrados

Los parámetros β_0 y β_1 son desconocidos, y se deben estimar con los datos de la muestra. Supongamos que hay n pares de datos: $(y_1, x_1), (y_2, x_2), \dots (y_n, x_n)$, esos datos pueden obtenerse en un experimento controlado, diseñado en forma específica para recolectarlos, o en un estudio observacional, o a partir de registros históricos existentes (lo que se llama un estudio retrospectivo) [8].

Para estimar β_0 y β_1 se usa el método de mínimos cuadrados. Esto es, se estiman β_0 y β_1 tales que la suma de los cuadrados de las diferencias entre las observaciones y_i y la línea recta sea mínima. Según la ecuación (1) se puede escribir como:

$$y_i = \beta_0 + \beta_1 x_i \quad i = 1, 2, \dots, n \quad (2)$$

Se puede considerar que la ecuación (1) es un modelo poblacional de regresión, mientras que la ecuación (2) es un modelo muestral de regresión, escritos en términos de los n pares de datos (y_i, x_i) ($i = 1, 2, \dots, n$). Así el criterio de mínimos cuadrados es:

$$S(\beta_0 + \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (3)$$

Los estimadores, por mínimos cuadrados de β_0 y β_1 que se designarán por $\hat{\beta}_0$ y $\hat{\beta}_1$ deben satisfacer:

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\beta_0, \beta_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (4)$$

y

$$\left. \frac{\partial S}{\partial \beta_1} \right|_{\beta_0, \beta_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \quad (5)$$

Se simplifican estas dos ecuaciones y se obtiene:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (6)$$

y

$$\hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i \quad (7)$$

Las ecuaciones (6) y (7) son llamadas ecuaciones normales de mínimos cuadrados. Su solución es la siguiente:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (8)$$

y

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \quad (9)$$

En donde

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (10)$$

y

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (11)$$

son los promedios de y_i y x_i respectivamente. Por consiguiente, $\hat{\beta}_0$ y $\hat{\beta}_1$ en las ecuaciones (10) y (11) son los estimadores por mínimos cuadrados de la ordenada al origen y la

pendiente, respectivamente. El modelo ajustado de regresión lineal simple es, entonces,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (12)$$

La ecuación (12) produce un estimado puntual, de la media de y para una determinada x [8].

2.8 Error cuadrático medio MSE

Es una métrica utilizada para evaluar el rendimiento de un modelo de regresión. Se calcula como el promedio de los cuadrados de los errores, es decir, la diferencia entre los valores predichos y los valores reales [8].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

Donde:

- y_i es el valor real o verdadero.
- $\hat{y}_i$ es el valor predicho por el modelo.
- n es el número total de observaciones.

Un MSE más bajo indica que el modelo tiene mejor precisión, ya que los errores de predicción son menores [8].

En MSE alto significa que el modelo está cometiendo errores grandes en sus predicciones [8].

2.9 Coeficiente de determinación R^2

Es una métrica que mide qué tan bien un modelo de regresión explica la variabilidad de los datos. Se usa principalmente en regresión lineal para evaluar la calidad del ajuste del

modelo a través de la ecuación (14).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (14)$$

Donde:

$$SS_{res} = \sum (y_i - \hat{y}_i)^2 \quad \text{Suma de los cuadrados (residuales).}$$

$$SS_{tot} = \sum (y_i - \bar{y})^2 \quad \text{Suma total de los cuadrados (variabilidad total de los datos).}$$

y_i Son los valores reales.

$\bar{y}$ Son los valores predichos.

$\hat{y}_i$ Es el promedio de los valores reales.

Los valores de R^2 son:

$R^2 = 1$ El modelo predice perfectamente los datos.

$R^2 = 0$ El modelo no explica ninguna variabilidad en los datos.

R^2 negativo El modelo es peor que simplemente usar el promedio de los valores reales.

2.10 Tratamiento de datos

El análisis de datos es el proceso de exploración, transformación y examinación de datos para identificar tendencias y patrones que revelen ideas importantes y aumenten la eficiencia para respaldar la toma de decisiones [9].

El proceso de análisis de datos se basa en varios pasos y fases. Es posible que las conclusiones de fases posteriores requieran volver a trabajar en una fase anterior, lo que implica un proceso más cíclico que lineal [9].

El proceso de análisis se divide de mejor manera en los siguientes pasos y fases:

-
- Entrada de datos: determina los requisitos y recopila los datos. Esto implica un poco de trabajo de investigación, como hablar con las partes interesadas, averiguar quiénes son los responsables de los datos y obtener acceso a estos últimos.
 - Preparación de datos: se trata de la estrategia y la táctica de preparación de los datos para su objetivo principal de producir ideas de análisis. Esto incluye limpiar y consolidar los datos sin procesar para transformarlos en datos estructurados y listos para el análisis. También incluye probar los resultados en cada parte del proceso de preparación para garantizar que el análisis esté generando los resultados deseados.
 - Exploración de datos: La exploración de datos, o análisis exploratorio de datos, es el proceso de estudio e investigación de un gran conjunto de datos mediante el muestreo, el análisis estadístico, la identificación de patrones y la elaboración de perfiles visuales, entre otros. Los métodos no son necesariamente científicos ni concluyentes, sino que sirven para comprender mejor la transformación de los datos.
 - Enriquecimiento de datos: se enriquecen y aumentan los datos con entradas y conjuntos de datos adicionales para mejorar el análisis. Este paso en el proceso de análisis de datos es fundamental para revelar nuevas ideas mediante la observación de datos desde una nueva perspectiva.
 - Ciencia de datos: se trata de aplicar métodos más avanzados de extracción de datos para obtener significados e ideas más profundos y difíciles de extraer, que son en gran medida imposibles de obtener mediante modalidades más rudimentarias de procesamiento de datos. Esto incluye algoritmos, entrenamiento de modelos, aprendizaje automático (ML) e inteligencia artificial (IA), por mencionar algunos.
 - Inteligencia de negocios: se pueden obtener resultados de negocios a través de la combinación de datos, software, infraestructura, procesos comerciales e intuición humana de una organización. Los resultados ofrecen ideas accionables a través de informes, paneles de control y visualizaciones para ayudar a tomar decisiones comerciales, se pueden obtener resultados de negocios a través de la

combinación de datos, software, infraestructura, procesos comerciales e intuición humana de una organización.

- Generación de informes: los resultados del análisis de datos deben compartirse de una manera eficaz que preserve los conocimientos adquiridos. La generación de informes se trata de organizar ese conocimiento y sus resultados en un formato fácil de comprender.
- Optimización: dado que las variables cambian con el tiempo, es necesario optimizar y mejorar los modelos para que sigan cumpliendo su propósito inicial o para que evolucionen a partir de este propósito basándose en nuevas entradas o características cambiantes.

2.11 Tipos de análisis de datos

Existen varios tipos diferentes de análisis de datos. Estos son los siguientes:

- Análisis descriptivo: responde la pregunta “¿Qué sucedió?” (¿Cuáles fueron nuestras ventas de la semana pasada?)
- Análisis de diagnóstico: responde la pregunta “¿Por qué sucedió esto?” (¿Por qué aumentaron nuestras ventas con respecto a la semana anterior?)
- Análisis predictivo: responde la pregunta “¿Qué sucederá?” (¿Cómo creemos que serán nuestras ventas en esas mismas tiendas durante la temporada de fiestas?)
- Análisis prescriptivo: responde la pregunta “¿Qué debo hacer?” (Según nuestras predicciones, recomendamos enviar más cantidad de un producto determinado a fin de evitar que se agote) [9].

2.12 Python

Python es un lenguaje de programación potente y fácil de aprender. Tiene estructuras de datos de alto nivel eficientes y un simple pero efectivo sistema de programación orientado a objetos. La elegante sintaxis de Python y su tipado dinámico, junto a su naturaleza interpretada lo convierten en un lenguaje ideal para scripting y desarrollo rápido de aplicaciones en muchas áreas, para la mayoría de las plataformas [10].

El intérprete de Python y la extensa librería estándar se encuentran disponibles libremente en código fuente y de forma binaria para la mayoría de las plataformas desde la Web de Python, se pueden distribuir libremente. El mismo sitio también contiene distribuciones y referencias a muchos módulos libres de Python de terceros, programas, herramientas y documentación adicional [10].

El intérprete de Python es fácilmente extensible con funciones y tipos de datos implementados en C o C++ (u otros lenguajes que permitan ser llamados desde C). Python también es apropiado como un lenguaje para extender aplicaciones modificables [10].

2.12.1 Colab

También conocido como "Colaboratory", te permite programar y ejecutar Python en tu navegador con las siguientes ventajas:

- No requiere configuración
- Acceso a GPUs sin coste adicional
- Permite compartir contenido fácilmente
- Colab puede facilitar tu trabajo, ya seas estudiante, científico de datos o investigador de IA.

Los cuadernos de Colab te permiten combinar código ejecutable y texto enriquecido en un mismo documento, además de imágenes, HTML, LaTeX y mucho más. Los cuadernos que creas en Colab se almacenan en tu cuenta de Google Drive. Puedes compartir tus cuadernos de Colab fácilmente con compañeros de trabajo o amigos, lo que les permite comentarlos o incluso editarlos [11].

2.12.2 Librería PANDAS

Es una librería de Python especializada en el manejo y análisis de estructuras de datos [12].

Las principales características de esta librería son:

- Define nuevas estructuras de datos basadas en los arrays de la librería NumPy, pero con nuevas funcionalidades.
- Permite leer y escribir fácilmente ficheros en formato CSV, Excel y bases de datos SQL.
- Permite acceder a los datos mediante índices o nombres para filas y columnas.
- Ofrece métodos para reordenar, dividir y combinar conjuntos de datos.
- Permite trabajar con series temporales.
- Realiza todas estas operaciones de manera muy eficiente.

2.12.3 Librería Matplotlib

Es una librería Python open source, desarrollada inicialmente por el neurobiólogo John Hunter en 2002. El objetivo era visualizar las señales eléctricas del cerebro de personas epilépticas. Para conseguirlo, quería replicar las funcionalidades de creación gráfica de MATLAB con Python [13].

Es posible crear trazados, histogramas, diagramas de barra y cualquier tipo de gráfica con ayuda de algunas líneas de código. Se trata de una herramienta muy completa, que permite generar visualizaciones de datos muy detalladas.

Esta librería es particularmente útil para las personas que trabajan con Python o NumPy. Se emplea particularmente en servidores de aplicación web, shells y scripts de Python. Con las API de matplotlib, también es posible que los desarrolladores integren gráficas con aplicaciones de interfaz gráfica [13].

2.12.4 Librería NUMPY

Numpy es un paquete Python dedicado al cálculo científico. Se trata de una dependencia indispensable para matplotlib, puesto que este último usa las funciones Numpy para los datos digitales y los arrays multidimensionales [13]

2.12.5 Librería Seaborn

Seaborn es una biblioteca para crear gráficos estadísticos en Python. Está basada en Matplotlib, y se integra con las estructuras de Pandas [14].

Esta biblioteca es tan potente como Matplotlib, pero aporta simplicidad y funciones inéditas. Permite explorar y comprender rápidamente los datos [14].

Se pueden capturar marcos de datos enteros, y las funciones internas de mapeo semántico y agregación estadística permiten convertir los datos en visualizaciones gráficas [14].

Seaborn absorbe toda la complejidad de Matplotlib. Sin embargo, es posible crear gráficos que satisfagan todas tus necesidades y requisitos [14].

2.13 Variables meteorológicas de influencia

A menudo es deseable establecer que existe una correlación entre alguna variable dependiente o de respuesta y algunas variables independientes o predictoras [15].

Además, puede ser de interés poder pronosticar los valores de alguna variable de respuesta en función de alguna relación supuesta. Es necesario recopilar datos de la población y utilizarlos para establecer la relación empírica entre las variables. Cuando la relación entre la variable de respuesta y las variables predictoras es lineal, los modelos de regresión son apropiados.

De acuerdo con los resultados de una evaluación de publicaciones, se clasifican en dos categorías: las que pronostican generación de energía solar y las que pronostican irradiancia solar, la Tabla 1 muestra los resultados de esta publicación [15].

Tabla 1. Resultados de la publicación.

Item No.	Ref. No.	Area	Year	Authors	Response variables	Predictor variables
1	[9]	Solar output power	2018	El Mouatasim & Darmane	Output power	module & ambient temperatures, & solar irradiance
2	[10]		2017	Yadav & Chandel	Maximum power	voltage open circuit, speed of wind, air & module temperatures, & radiation solar
3	[11]		2013	Mellit <i>et al.</i>	Power profile	solar irradiance & temperature air
4	[12]		1997	Hiyama & Kitabayashi	Daily maximum power	temperature, irradiation, & wind velocity
5	[13]		2012	Oudjana <i>et al.</i>	Week ahead PV power	global irradiance & temperature
6	[14]		2015	Abuella & Chowdhury	Hourly steps solar power forecasts	liquid & ice cloud water, pressure, humidity, cloud cover, wind, temperature, thermal radiation, solar radiation & precipitation
7	[15]		2019	Xu <i>et al.</i>	Day-ahead power PV	hour, month, weather type & solar radiation
8	[16]		2016	Verma <i>et al.</i>	Power solar generation	wind speed, temperature, azimuthal angle, cloud cover, elevation angle, rainfall, solar panels degradation factor, & humidity
9	[17]		2019	Shapsough <i>et al.</i>	PV power output	ambient temperature & irradiation solar
10	[18]		2020	Al-Dahidi <i>et al.</i>	Solar PV production	solar irradiation, wind speed & ambient temperature
11	[19]		2013	Juban & Quach	Daily expected PV power output	short-wave radiative flux (down/up wards), air pressure, precipitation & maximum temperature
12	[20]		2018	Fernandez & Prajapati	Maximum power	solar radiation, short circuit current, temperature & circuit voltage
13	[21]		2017	Dias <i>et al.</i>	PV energy estimation	irradiance, temperature and wind speed
14	[22]		2021	Lehloka <i>et al.</i>	Output voltage	time of day

Como se puede observar las variables de predicción en la mayoría de los casos está presente la irradiancia solar.

De acuerdo con el resultado de un análisis de la influencia de los factores meteorológicos en la generación solar fotovoltaica, utilizando la correlación de rangos de Spearman's, se concluye que dichas variables son: irradiancia solar, humedad relativa y la temperatura.

La Figura 14Figura 14. Correlación de rangos de Sperman's. muestra el resultado de esta correlación, así como la Tabla 2 muestra el resultado numérico de dichos rangos [16].

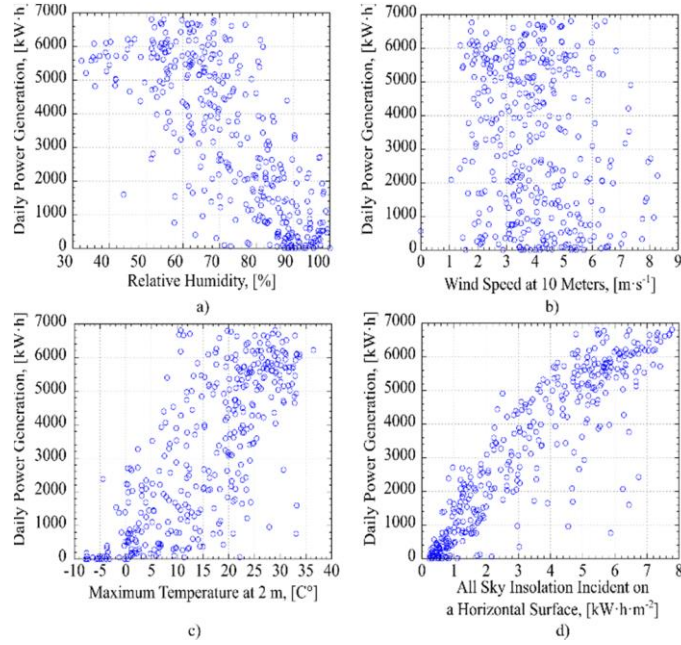


Figura 14. Correlación de rangos de Spearman's.

Tabla 2. Coeficiente de correlación de los rangos de Spearman's.

Name of influential factors	Description	Spearman's rank correlation coefficient
PRECTOT	Precipitation	-0,3705321
QV2M	Specific Humidity at 2 Meters	0,5582885
RH2M	Relative Humidity at 2 Meters	-0,7637474
PS	Surface Pressure	0,0986246
T2M_RANGE	Temperature Range at 2 Meters	0,7054887
TS	Earth Skin Temperature	0,7157898
T2MDEW	Dew/Frost Point at 2 Meters	0,561229
T2MWET	Wet Bulb Temperature at 2 Meters	0,5598176
T2M_MAX	Maximum Temperature at 2 Meters	0,7463827
T2M_MIN	Minimum Temperature at 2 Meters	0,6297871
T2M	Temperature at 2 Meters	0,7129819
WS50M_RANGE	Wind Speed Range at 50 Meters	0,0426838
WS10M_RANGE	Wind Speed Range at 10 Meters	-0,02647898
WS50M_MIN	Minimum Wind Speed at 50 Meters	-0,1895334
WS10M_MIN	Minimum Wind Speed at 10 Meters	-0,1990136
WS50M_MAX	Maximum Wind Speed at 50 Meters	-0,1583861
WS10M_MAX	Maximum Wind Speed at 10 Meters	-0,1732506
WS50M	Wind Speed at 50 Meters	-0,1893845
WS10M	Wind Speed at 10 Meters	-0,2264369
KT	Insolation Clearness Index (dimensionless)	0,7818045
CLRSKY_SFC_SW_DWN	Clear Sky Insolation Incident on a Horizontal Surface	0,4708504
ALLSKY_SFC_SW_DWN	All Sky Insolation Incident on a Horizontal Surface	0,9009986

Otro estudio de análisis de variables utiliza la correlación de Pearson para el pronóstico de generación solar, los resultados indican que las variables meteorológicas que presentan mayor impacto en la generación son: la irradiancia y la temperatura del aire. La Tabla 3 muestra estos resultados, así como la comparativa con la correlación de Spearman's [17].

Tabla 3. Resultados de la correlación de Pearson y Spearman's.

Data source	Pearson Correlation		
	irradiance	Air temperature	Wind Speed
INMET	0.81	0.23	0.19
NASA Power	0.88	0.44	0.033
StormGlass and OpenWeather	0.74	-0.059	-0.11
Data source	Spearman Correlation		
	irradiance	Air temperature	Wind Speed
INMET	0.79	0.22	0.21
NASA Power	0.86	0.45	0.063
StormGlass and OpenWeather	0.70	-0.062	-0.13

Otro estudio aborda los temas de regresión lineal, regresión de Poisson y la regresión de Lasso, en función de variables meteorológicas como la temperatura, humedad relativa y la intensidad luminosa. La Figura 15 muestra la comparación de los resultados de las regresiones, como se puede observar, la regresión lineal presenta el mejor comportamiento respecto a los otros dos, con datos de entrenamiento [18].

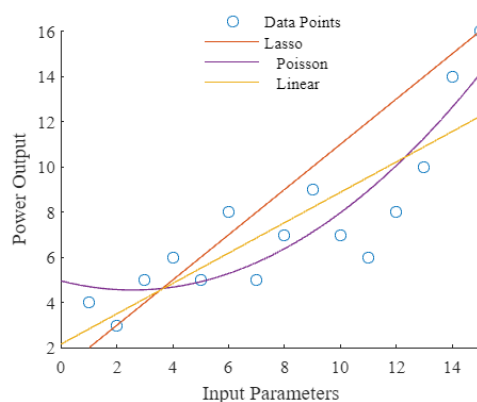


Figura 15 Comparación de modelos de regresión para pronóstico.

Capítulo 3 Desarrollo

La Figura 16 muestra el diagrama de flujo del desarrollo de la investigación, el cual corresponde desde el análisis de datos y tratamiento hasta la obtención de los modelos de predicción de generación solar fotovoltaica de corto plazo.

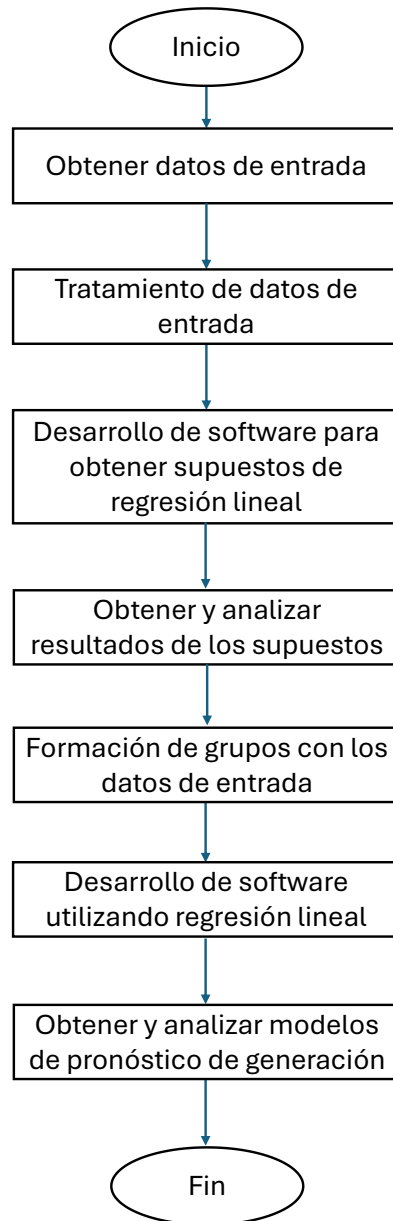


Figura 16 Diagrama de flujo del desarrollo del proyecto.

3.1 Obtener datos de entrada

La Tabla 4 muestra los datos de variables meteorológicas de un parque solar fotovoltaico para el año 2023 de CENACE-GCROR. Los datos están organizados en columnas que incluyen la estación, fecha, hora y/o mes para los 365 días del año.

Las variables meteorológicas indican:

- Temperatura ambiente
- Sensación térmica
- Índice de calor
- Precipitación
- Humedad
- Velocidad del viento
- Nubosidad
- Radiancia solar

Tabla 4. Variables meteorológicas de la ubicación de un parque de generación solar fotovoltaico.

ESTACIÓN	FECHA	HORA	TEMPERATURA	TEMPERATURA A.	INDICE CALOR	PRECIPITACIÓN	HUMEDAD	V. VIENTO	NUBOSIDAD	R. SOLAR
P1	01/01/2023	1	6.9	5.7	6.9	0	86	1.9	27.7	0
P1	01/01/2023	2	7	5.7	7	0	85.6	1.9	7.6	0
P1	01/01/2023	3	6.8	5.5	6.8	0	85.8	1.9	0	0
P1	01/01/2023	4	6.6	5.3	6.6	0	86.3	1.9	0.4	0
P1	01/01/2023	5	6.3	5.1	6.3	0	86.9	1.9	1.6	0
P1	01/01/2023	6	6	4.7	6	0	88.1	1.8	2.2	0
P1	01/01/2023	7	5.6	4.3	5.6	0	88.8	1.8	1.9	0
P1	01/01/2023	8	6.9	6.2	6.9	0	85.5	1.8	1.7	50.1
P1	01/01/2023	9	10.1	12.9	10.1	0	74.5	1.6	0.8	277.4
P1	01/01/2023	10	10.8	17.9	10.8	0	70.8	2.7	0.1	494.9
P1	01/01/2023	11	14.1	21.2	14.1	0	62	2	1.7	668.2
P1	01/01/2023	12	17.5	23.2	17.5	0.07	54.5	1.2	11.9	763.2
P1	01/01/2023	13	18.4	22.3	18.4	0.27	56.6	0.3	42.2	635.1
P1	01/01/2023	14	17.6	22	17.6	2.23	63.3	0	73.2	629.9
P1	01/01/2023	15	17.8	19.6	17.8	0	62.7	0	78.9	455.3
P1	01/01/2023	16	18.3	19.1	18.3	0	59.9	0	63.8	406.9
P1	01/01/2023	17	18.2	18	18.2	0	59.3	0	44.2	263.1
P1	01/01/2023	18	16.8	14.7	16.8	0	66.5	0	39.2	60.5
P1	01/01/2023	19	12.9	12.9	12.9	0	66.1	0.9	41.7	0
P1	01/01/2023	20	12.2	12.2	12.2	0	67.5	0.8	56.4	0
P1	01/01/2023	21	10.5	10.5	10.5	0	72.4	1.5	85.3	0
P1	01/01/2023	22	9.7	8.9	9.7	0	73.8	1.9	99.5	0
P1	01/01/2023	23	9.4	8.5	9.4	0	74.6	1.9	100	0
P1	01/01/2023	24	8.8	7.9	8.8	0	78.1	2	100	0

La Tabla 5 muestra los datos de generación en MW/h, clasificados en año, mes, día y hora. Estos datos son proporcionados por CENACE-GCROR.

Tabla 5. Datos de generación eléctrica de parque solar fotovoltaico.

A	SIN_HORA	FECHA	SIN_HORA - Month	FECHA SIN_HORA - Day	Hora	Generación Real
	2023	January		1	0	0.00
	2023	January		1	1	0.00
	2023	January		1	2	0.00
	2023	January		1	3	0.00
	2023	January		1	4	0.00
	2023	January		1	5	0.00
	2023	January		1	6	0.00
	2023	January		1	7	8.98
	2023	January		1	8	124.37
	2023	January		1	9	176.28
	2023	January		1	10	180.13
	2023	January		1	11	178.74
	2023	January		1	12	179.61
	2023	January		1	13	182.81
	2023	January		1	14	145.92
	2023	January		1	15	96.59
	2023	January		1	16	50.99
	2023	January		1	17	43.87
	2023	January		1	18	0.00
	2023	January		1	19	0.00
	2023	January		1	20	0.00
	2023	January		1	21	0.00
	2023	January		1	22	0.00
	2023	January		1	23	0.00

3.2 Tratamiento de datos de entrada

Una vez recolectada la información se procede a reorganizar las variables meteorológicas junto con la variable de generación dentro de la misma hoja de procesamiento de información, lo cual se puede apreciar en la Tabla 6

Tabla 6. Variables meteorológicas y generación eléctrica.

ESTACIÓN	FECHA	HORA	TEMPERATURA	TEMPERATURA A.	INDICE CALOR	PRECIPITACIÓN	HUMEDAD	V. VIENTO	NUBOSIDAD	R. SOLAR	GENERACIÓN
P1	01/01/2023	1	6.9	5.7	6.9	0	86	1.9	27.7	0	0
P1	01/01/2023	2	7	5.7	7	0	85.6	1.9	7.6	0	0
P1	01/01/2023	3	6.8	5.5	6.8	0	85.8	1.9	0	0	0
P1	01/01/2023	4	6.6	5.3	6.6	0	86.3	1.9	0.4	0	0
P1	01/01/2023	5	6.3	5.1	6.3	0	86.9	1.9	1.6	0	0
P1	01/01/2023	6	6	4.7	6	0	88.1	1.8	2.2	0	0
P1	01/01/2023	7	5.6	4.3	5.6	0	88.8	1.8	1.9	0	0
P1	01/01/2023	8	6.9	6.2	6.9	0	85.5	1.8	1.7	50.1	8.981
P1	01/01/2023	9	10.1	12.9	10.1	0	74.5	1.6	0.8	277.4	124.371
P1	01/01/2023	10	10.8	17.9	10.8	0	70.8	2.7	0.1	494.9	176.278
P1	01/01/2023	11	14.1	21.2	14.1	0	62	2	1.7	668.2	180.125
P1	01/01/2023	12	17.5	23.2	17.5	0.07	54.5	1.2	11.9	763.2	178.735
P1	01/01/2023	13	18.4	22.3	18.4	0.27	56.6	0.3	42.2	635.1	179.611
P1	01/01/2023	14	17.6	22	17.6	2.23	63.3	0	73.2	629.9	182.809
P1	01/01/2023	15	17.8	19.6	17.8	0	62.7	0	78.9	455.3	145.92
P1	01/01/2023	16	18.3	19.1	18.3	0	59.9	0	63.8	406.9	96.589
P1	01/01/2023	17	18.2	18	18.2	0	59.3	0	44.2	263.1	50.991
P1	01/01/2023	18	16.8	14.7	16.8	0	66.5	0	39.2	60.5	43.865
P1	01/01/2023	19	12.9	12.9	12.9	0	66.1	0.9	41.7	0	0
P1	01/01/2023	20	12.2	12.2	12.2	0	67.5	0.8	56.4	0	0
P1	01/01/2023	21	10.5	10.5	10.5	0	72.4	1.5	85.3	0	0
P1	01/01/2023	22	9.7	8.9	9.7	0	73.8	1.9	99.5	0	0
P1	01/01/2023	23	9.4	8.5	9.4	0	74.6	1.9	100	0	0
P1	01/01/2023	24	8.8	7.9	8.8	0	78.1	2	100	0	0

Se tiene un total de 78,840 datos que corresponden a los históricos del año 2023 de un parque solar fotovoltaico, los cuales se deben distribuir de distintas formas para llevar a cabo el estudio estadístico y la obtención de modelos de pronóstico.

3.3 Desarrollo de software para obtener supuestos de regresión lineal

Para realizar la aplicación de supuestos de la regresión lineal se implementa en el lenguaje de programación Python, a través de la plataforma Colab, utilizando las librerías necesarias para leer y manejar los archivos previamente clasificados en un formato .xlsx.

La Figura 17 muestra la importación de uno de los archivos para su tratamiento en el cuaderno u entorno de Colab.

```
# Importar biblioteca
import pandas as pd

# Crear archivo de datos desde documento .xlsx
data_file_name = '/content/Datos_gen_hora.xlsx'
```

Figura 17 Código fuente para importar datos en Colab.

Fuente: Elaboración propia.

3.4 Obtener y analizar resultados de los supuestos

Para el análisis de datos general se realiza la aplicación de los supuestos de la regresión lineal, la Figura 18 muestra parte del código fuente para llevar a cabo el supuesto de la linealidad.

```
[ ] # ajuste del modelo lineal
x = df11.drop(["norm_GENERACIÓN"], axis=1)
y = df11.norm_GENERACIÓN

[ ] from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X = sc.fit_transform(x)

▶ from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state = 0, test_size=0.3)

[ ] from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from sklearn import linear_model

regr = linear_model.LinearRegression()
regr.fit(X_train, y_train)
y_pred = regr.predict(X_train)
y_pred
print(regr.coef_)

↗ [-3.21644482  0.0558693  3.16780648 -0.01468054 -0.01076417  0.00362417
   -0.02372261  0.30069868]

[ ] print(regr.intercept_)

↗ 0.2740436380256111

[ ] print("R squared: {}".format(r2_score(y_true=y_train, y_pred=y_pred)))

↗ R squared: 0.807868468321945
```

Figura 18 Código fuente para identificar si los datos presentan linealidad.

Fuente: Elaboración propia.

La Figura 19 muestra el supuesto para la media de los residuos.

```
[ ] # Media de los residuos
residuals = y_train.values-y_pred
mean_residuals = np.mean(residuals)
print("Mean of Residuals {}".format(mean_residuals))
```

Figura 19 Código fuente para determinar la media de los residuos.

La Figura 20 muestra el código fuente para determinar la homocedasticidad de los datos.

```
✓ 1 s # 3er Comprobación de homocedasticidad cuando la varianza de los errores estocásticos
p = sns.scatterplot(x=y_pred, y=residuals)
plt.xlabel('y_pred/predicted values')
plt.ylabel('Residuals')
plt.ylim(-75,75)
plt.xlim(.26)
p = sns.lineplot(x=[0, 230], y=[0,0],color='blue')
p = plt.title('Residuals vs fitted values plot for homoscedasticity check')
```

Figura 20 Código fuente para determinar la homocedasticidad de los datos.

La Figura 21 muestra el código fuente para determinar si el residuo presenta distribución normal.

```
✓ 0 s # 4to Supuesto Verificar la normalidad de los términos de error/residuos
p = sns.distplot(residuals,kde=True)
p = plt.title('Normality of error terms/residuals')
```

Figura 21Código fuente para determinar si los residuos presentan una distribución normal.

La Figura 22 muestra el código fuente para determinar si el residuo presenta autocorrelación.

```
✓ 1 s # 5to supuesto autocorrelación
plt.figure(figsize=(10,6))
p = sns.lineplot(x=y_pred,y=residuals,marker='o',color='blue')
plt.xlabel('y_pred/predicted values')
plt.ylabel('Residuals')
plt.ylim(-200,100)
plt.xlim(0,200)
p = sns.lineplot(x=[0,200],y=[0,0],color='red')
p = plt.title('Residuals vs fitted values plot for autocorrelation check')
```

Figura 22 Código fuente para determinar si los residuos presentan autocorrelación.

La Figura 23 muestra el código fuente para determinar la correlación de las variables independientes contra la variable dependiente.

```
✓ 5 # 6. multicolinealidad
plt.figure(figsize=(10,10)) # on this line I just set the size of figure to 12 by 10.
p=sns.heatmap(dfs['dfh12'].corr(), annot=True,cmap='RdYlGn',square=True)
```

Figura 23 Código fuente para determinar la correlación de las variables.

3.5 Formación de grupos con los datos de entrada

Para la obtención de los modelos de estudio se conformaron distintos grupos de datos, los cuales se pueden ver en la Figura 24.

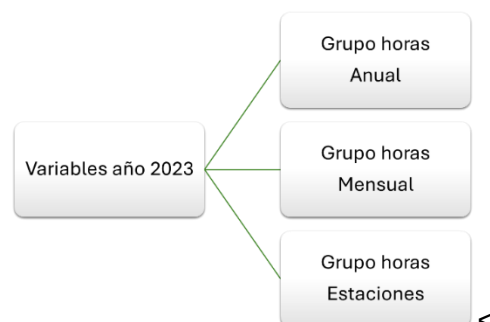


Figura 24 Formación de grupos para su análisis.

La Tabla 7 muestra la clasificación del grupo horas-Anual, la cual considera un agrupamiento de 24 horas para los 365 días del año 2023, este agrupador es el que concentra el mayor número de datos respecto para el análisis a los otros dos y su objetivo principal es realizar un análisis estadístico de regresión lineal para cada hora, con la finalidad de obtener un modelo general con base a todas las horas y los días del año 2023.

Tabla 7. Grupo horas-Anual para la hora 8 de cada día del mes de enero 2023.

ÓN_8	FECHA_8	HORA_8	TEMPERATURA_8	TEMPERATURA_A_8	INDICE CALOR_8	PRECIPITACIÓN_8	HUMEDAD_8	V. VIENTO_8	NUBOSIDAD_8	R. SOLAR_8	GENERACIÓN_8
.	01/01/2023	8	6.9	6.2	6.9	0	85.5	1.8	1.7	50.1	8.981
.	02/01/2023	8	6.2	6.7	6.2	0	78.2	1.4	98.4	43.4	14.815
.	03/01/2023	8	7.2	6	7.2	0	81.8	2.5	99.9	43.2	21.701
.	04/01/2023	8	7.7	9.1	7.7	0	74.9	0	0	48.8	17.572
.	05/01/2023	8	7.3	6.7	7.3	0	88.2	0	65.2	39.5	21.211
.	06/01/2023	8	5	5.4	5	0	95.7	0.2	43.5	41.2	23.458
.	07/01/2023	8	5.2	5.1	5.2	0	80.3	0.4	0	47.9	4.662
.	08/01/2023	8	6.4	6.9	6.4	0	62.5	0.7	0.5	47.3	5.894
.	09/01/2023	8	5.6	6.2	5.6	0	97.5	0	97	32.4	4.477
.	10/01/2023	8	7.1	7.5	7.1	0	83.6	0.5	60.4	35.6	7.507
.	11/01/2023	8	6.1	5.7	6.1	0	83	1.4	6.1	44.2	5.835
.	12/01/2023	8	7.6	8.5	7.6	0	72.1	0	0	44.2	14.532
.	13/01/2023	8	4.5	-0.3	4.5	0	76.3	1.6	71.2	31.3	5.222
.	14/01/2023	8	1.3	0.4	1.3	0	78.8	0	3.5	46.3	24.367
.	15/01/2023	8	1.2	3.1	1.2	0	69.6	1.3	0	46.2	16.193
.	16/01/2023	8	4.5	3.3	4.5	0	71	3	0	45.8	6.617
.	17/01/2023	8	5.1	4.3	5.1	0	75.3	1	64.4	44.6	14.874
.	18/01/2023	8	6.3	5.2	6.3	0	86.1	3.3	50.3	43.6	13.323
.	19/01/2023	8	7.9	9.1	7.9	0	77.9	0.3	66	43.4	16.014
.	20/01/2023	8	6	5.7	6	0	73.3	1.5	95.6	43	3.659
.	21/01/2023	8	5.6	5.4	5.6	0	78.7	2	47.8	45.6	0
.	22/01/2023	8	6	5.6	6	0	81.2	2	0	45.3	4.584
.	23/01/2023	8	7.6	6.8	7.6	0	92.4	2.4	100	28.3	3.12
.	24/01/2023	8	5.6	5.3	5.6	0	83.6	1.2	31.7	47.1	2.149
.	25/01/2023	8	8.9	7.5	8.9	0	86.1	0.6	90.7	24.5	9.758
.	26/01/2023	8	6.6	7.5	6.6	0	98.4	1.1	85.7	26.5	2.263
.	27/01/2023	8	3.1	2.7	3.1	0	100	1.4	71.7	39.4	2.723
.	28/01/2023	8	5.3	4.8	5.3	0	100	1.5	93.8	36.3	7.795
.	29/01/2023	8	9.3	9.8	9.3	0	74.2	0.5	0	49.1	21.705
.	30/01/2023	8	7.7	9.5	7.7	0	59	0.3	0	51.2	16.437
.	31/01/2023	8	5.9	7.1	5.9	0	66.9	0.5	28.7	52.5	11.954

La Tabla 8 muestra la clasificación del grupo horas-Mensual, la cual considera un agrupamiento de 24 horas para cada mes del año 2023, este agrupador reduce significativamente la muestra de datos, su objetivo es obtener un modelo de regresión lineal más ajustado debido a las condiciones de operación y climatológicas de cada mes.

Tabla 8. Grupo horas-Mensual para la hora 12 de cada día del mes de junio 2023.

HORA	12								
FECHA	Temperatura	Temperatura	Indice calor	Precipitacion	Humedad	V. Viento	Nubosidad	R. Solar	Generacion
01/06/2023	23.2	29.9	23.2	0	44.3	1.2	29.2	884	183.813
02/06/2023	24.6	31.7	24.6	0	54.6	2.8	4.4	1016.3	181.576
03/06/2023	25.9	29.9	25.9	0	37.2	4.7	63.1	884.1	170.241
04/06/2023	24.8	31.3	24.8	0	41.7	2.2	44.2	991.1	156.016
05/06/2023	20.6	28.1	20.6	0.69	51.7	1.1	48	922.7	135.095
06/06/2023	21.9	28.2	21.9	0	49.7	2.2	21.3	917.3	120.358
07/06/2023	22	27	22	0.09	45.8	0.3	60.2	745.2	144.647
08/06/2023	23.4	28.8	23.4	0.08	42.7	1	72.4	855.5	166.646
09/06/2023	25.5	30.1	25.5	0	47.8	2.5	29.9	900.5	132.284
10/06/2023	25.3	32.7	25.3	0	41.7	1.5	42.1	1031.7	196.581
11/06/2023	25.7	30.8	25.7	0	36.4	1.5	34.7	825.7	183.993
12/06/2023	25.6	33.5	25.6	0	41.1	1.3	12.7	1051	192.295
13/06/2023	24.7	33.1	24.7	0	54.9	0.5	0.2	1073.2	197.282
14/06/2023	25.2	33.8	25.2	0	26.6	1.3	0.4	1060.9	194.409
15/06/2023	25.5	34	25.5	0	34.3	0.6	0.2	1053.6	197.647
16/06/2023	25.5	33.1	25.5	0	26.2	3.1	47	1061.6	199.267
17/06/2023	26	32.9	26	0	17.6	5.1	54.7	1067.2	199.402
18/06/2023	25.2	32.6	25.2	0	30.1	4.4	0	1070.6	199.355
19/06/2023	25.1	32.6	25.1	0	29.2	2.1	0	1057.6	197.679
20/06/2023	26.1	33.3	26.1	0	30.3	3.1	0.2	1050.2	192.87
21/06/2023	24.9	32.6	24.9	0	35.3	1.2	20	1021	174.342
22/06/2023	24.5	31	24.5	0	37.3	0.9	4.3	1048.9	189.26
23/06/2023	22.8	29.7	22.8	0	43.7	1.7	2	1038.8	194.228
24/06/2023	22.3	28.6	22.3	0.06	55.5	2.7	57.7	893.2	132.097
25/06/2023	18.4	27.2	18.4	0.15	74.7	1.8	78.5	882.8	128.64
26/06/2023	22.4	23.5	22.4	0	50.8	6.6	76	564.2	179.076
27/06/2023	21.4	29.1	21.4	0.12	57.8	3.1	72.6	953.6	85.699
28/06/2023	20.4	28.9	20.4	0.23	75.7	3	98.4	900.2	134.795
29/06/2023	20.6	24	20.6	0.26	56.1	2.3	97.3	642.8	142.67
30/06/2023	21.9	28.1	21.9	0.23	56.3	1.3	83.6	832.4	161.157

La Tabla 9 muestra la clasificación del grupo horas-Estación, la cual considera un agrupamiento de 24 horas para cada estación del año 2023, para el caso de la primavera considera los días desde el 20 de marzo del 2023 hasta el 21 de junio del mismo año, en el caso del verano, considera los días 22 de junio 2023 hasta 23 de septiembre 2023, para el otoño, los días 24 de septiembre hasta el 22 de diciembre del 2023, y para el invierno se toma el resto del año y el inicio, es decir los días 23 al 31 de diciembre del 2023, así como los días primero hasta el 19 de marzo del 2023. Al realizar este agrupamiento obtendremos un mejor conjunto de variables climatológicas.

Tabla 9. Grupo horas-Estación para primavera 2023 (extracto) hora uno.

Hora	1								
Fecha	Temperatura	Temperatura	Indice calor	Precipitacion	Humedad	V. Viento	Nubosidad	R. Solar	Generacion
20/03/2023	5	5	5	0	88.5	1.3	98.5	0	0
21/03/2023	7.1	6.2	7.1	0	82.5	1.7	0.2	0	0
22/03/2023	10.4	10.3	10.4	0	58.9	2.1	6	0	0
23/03/2023	11.3	10.8	11.3	0	33	1.5	91.2	0	0
24/03/2023	11.1	10.9	11.1	0	68.5	2.5	18.5	0	0
25/03/2023	12.4	11.7	12.4	0	82.1	2.6	3.7	0	0
26/03/2023	10.7	10.9	10.7	0	98.2	1.7	11.6	0	0
27/03/2023	11.1	11.2	11.1	0	81.7	1.1	18.2	0	0
28/03/2023	14.1	12.9	14.1	0	97.8	0.7	37.5	0	0
29/03/2023	12	12.3	12	0	94.9	2	69.9	0	0
30/03/2023	11.5	11.1	11.5	0	72.3	0.4	8.1	0	0
31/03/2023	9.7	10.6	9.7	0	76.7	0	99.9	0	0
01/04/2023	10.7	10.8	10.7	0	85.1	1.1	29	0	0
02/04/2023	11.8	12.2	11.8	0	99.8	1.9	90.1	0	0
03/04/2023	11.9	12.1	11.9	0	69.1	0.6	0	0	0
04/04/2023	14.7	13.3	14.7	0	50.6	3.3	12	0	0
05/04/2023	14.1	12.9	14.1	0	50.3	2.7	41.7	0	0
06/04/2023	11.8	12.1	11.8	0	80.9	0.6	0.8	0	0
07/04/2023	13.1	14.3	13.1	0	97.8	0.4	90.5	0	0
08/04/2023	12.4	12.9	12.4	0	94.5	1.3	92.2	0	0
09/04/2023	12.2	12.8	12.2	0	92.2	0.7	83.7	0	0
10/04/2023	12.1	12.5	12.1	0	91.1	1.3	98.9	0	0
11/04/2023	11.2	11.5	11.2	0	91.3	0.7	92.9	0	0
12/04/2023	9.7	9.7	9.7	0	95.7	1.2	71.7	0	0
13/04/2023	8	7.1	8	0	96	1.7	26.1	0	0
14/04/2023	10	10	10	0	84.6	2.3	1	0	0

3.6 Desarrollo de software utilizando regresión lineal

Para aplicar la regresión lineal a los datos previamente clasificados en archivos .xlsx, nuevamente empleamos la plataforma de Google Colab, se generan 3 programas semejantes para poder analizar cada uno de ellos de forma individual.

Antes de realizar la regresión lineal, se debe dividir el conjunto de datos en entrenamiento y prueba 80 % y 20 % respectivamente.

La Figura 25 muestra el código fuente para dividir los datos en entrenamiento y prueba, así como el entrenamiento del modelo de regresión lineal considerando los datos anteriores.

```

# Dividir en X (predictores) e y (variable objetivo)
X = df_hora[[variable_x]]
y = df_hora[variable_y]

# Dividir en conjuntos de entrenamiento y prueba (80% - 20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
train_data[hora] = (X_train, y_train)
test_data[hora] = (X_test, y_test)

# Entrenar el modelo de regresión lineal en el conjunto de entrenamiento
modelo = LinearRegression()
modelo.fit(X_train, y_train)

```

Figura 25 Código fuente para la división de datos e inicio de la regresión lineal.

3.7 Obtener y analizar modelos de pronóstico de generación

La imagen 43 muestra el código fuente para calcular las predicciones y métricas del conjunto de entrenamiento, prueba, así como el cálculo de la pendiente y del intercepto.

```

# Calcular predicciones y métricas en el conjunto de entrenamiento
y_train_pred = modelo.predict(X_train)
r2_train = round(r2_score(y_train, y_train_pred), 3)
mse_train = round(mean_squared_error(y_train, y_train_pred), 3)

# Calcular predicciones y métricas en el conjunto de prueba
y_test_pred = modelo.predict(X_test)
r2_test = round(r2_score(y_test, y_test_pred), 3)
mse_test = round(mean_squared_error(y_test, y_test_pred), 3)

# Calcular la pendiente y el intercepto
pendiente = round(modelo.coef_[0], 3)
intercepto = round(modelo.intercept_, 3)

# Predicción para el valor del predictor dado (irradiancia)
dato_predictor = predictores[i]
generacion_prediccion = modelo.predict([[dato_predictor]])
prediccion_generacion = round(generacion_prediccion[0], 3)

```

Figura 26 Código fuente para el cálculo de predicciones y métricas del conjunto de datos.

La figura 43 muestra el código fuente para crear una tabla de resultados de la regresión lineal, así como el código para realizar los gráficos con los resultados de las iteraciones.

```

# Crear un DataFrame con los resultados de esta hora
resultado = pd.DataFrame({
    'Hora': [hora],
    'Pendiente': [pendiente],
    'Intercepto': [intercepto],
    'Ecuación': [f'y = {pendiente}x + {intercepto}'],
    'Dato Predictor': [dato_predictor],
    'Predicción Generación': [prediccion_generacion],
    'R2 Entrenamiento': [r2_train],
    'MSE Entrenamiento': [mse_train],
    'R2 Prueba': [r2_test],
    'MSE Prueba': [mse_test]
})

# Graficar regresión con intervalo de confianza para esta hora
plt.figure()
sns.regplot(x=X_train[variable_x], y=y_train, ci=95, line_kws={"color": "red"})
plt.title(f'Regresión lineal para la hora {hora} (entrenamiento)')
plt.xlabel(variable_x)
plt.ylabel(variable_y)
plt.show()

# Concatenar el resultado al DataFrame final
tabla_final = pd.concat([tabla_final, resultado], ignore_index=True)

```

Figura 27 Código fuente para la creación de una tabla de resultados y sus gráficos.

La Figura 28 muestra el código fuente para la generación de gráficos de predicción por hora por día del conjunto de datos.

```

# Graficar la Predicción de Generación por Hora
plt.figure(figsize=(8, 6))
plt.plot(tabla_final['Hora'], tabla_final['Predicción Generación'], marker='o', color='b', linestyle='-')
plt.title('Predicción de Generación por Hora')
plt.xlabel('Hora')
plt.ylabel('Predicción de Generación')
plt.grid(True)
plt.show()

```

Figura 28 Código fuente graficar el resultado del pronóstico por hora por día o de corto plazo.

Capítulo 4 Resultados

Este capítulo integra los resultados de la investigación conforme al diagrama de flujo de la Figura 16, principalmente la aplicación de los supuestos para poder aplicar una regresión lineal a los datos proporcionados por CENACE-GCROR hasta la obtención de los modelos de pronóstico para cada grupo de datos de la Figura 24, los cuales son datos anuales, mensuales y estacionales.

4.1 Resultados de linealidad

La Figura 29 muestra los resultados de la aplicación de la prueba de linealidad de los datos, como se puede apreciar el resultado es positivo, ya que presenta un valor de $R^2 = 0.809$.



Figura 29 Resultado de la prueba de linealidad.

4.2 Resultado de la media de los residuos

La Figura 30 muestra los resultados de la prueba de la media de los residuos, la cual presenta un resultado satisfactorio, ya que el valor es muy cercano a cero.

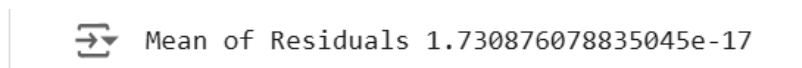


Figura 30 Resultado de la prueba de media de los residuos.

4.3 Resultado de homocedasticidad

La Figura 31 muestra el resultado de la homocedasticidad de los datos, la figura muestra que no se representa ningún patrón dentro del conjunto de datos.

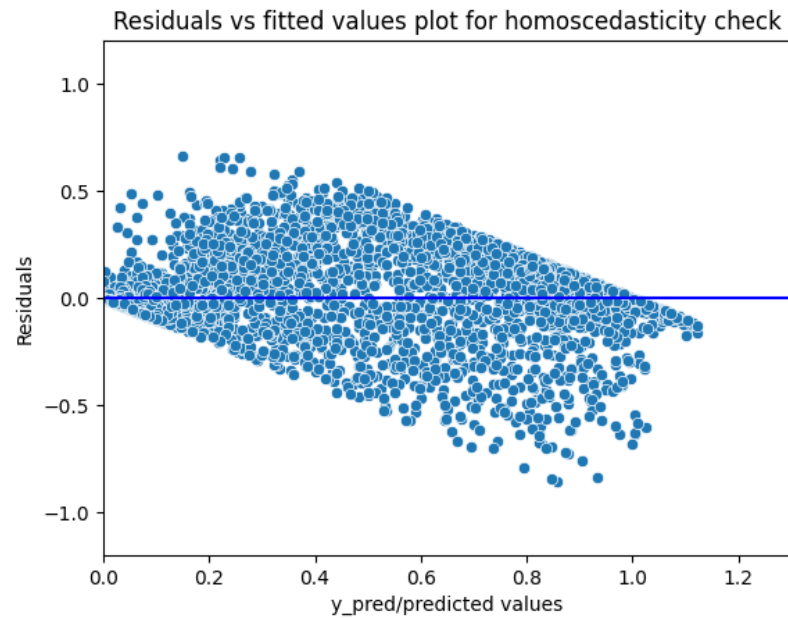


Figura 31 Resultado de la prueba de homocedasticidad.

4.4 Distribución de los datos

La Figura 32 muestra la distribución del conjunto de datos, la cual presenta una distribución normal.

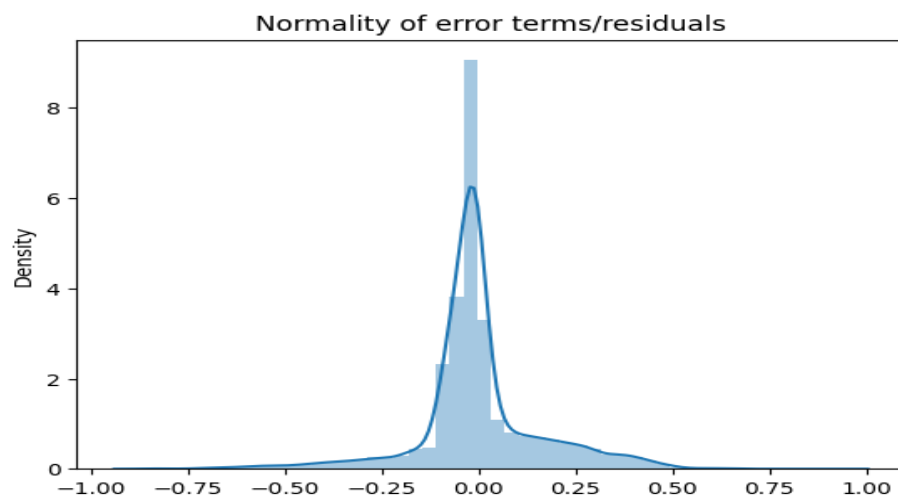


Figura 32 Resultado de la distribución de los residuos.

4.5 Autocorrelación de los residuos

La Figura 33 muestra el resultado de autocorrelación para los residuos,

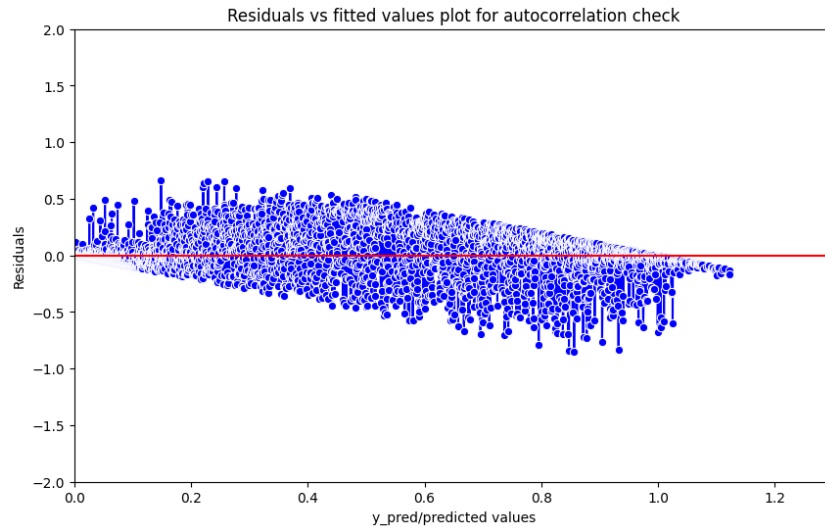


Figura 33 Resultado de autocorrelación.

4.6 Correlación de las variables

La Figura 34 muestra el resultado para correlación entre todas las variables, esta última nos indica que variable presenta la mayor correlación para considerarla en el modelo de regresión lineal, esta variable es la irradiancia solar la cual presenta una correlación de 0.92 respecto a la generación, la segunda variable con mayor correlación es la de sensación térmica con un valor de 0.81, para una correlación negativa se tiene el caso de humedad con -0.47. Los resultados cuentan con datos normalizados con escalado simple.

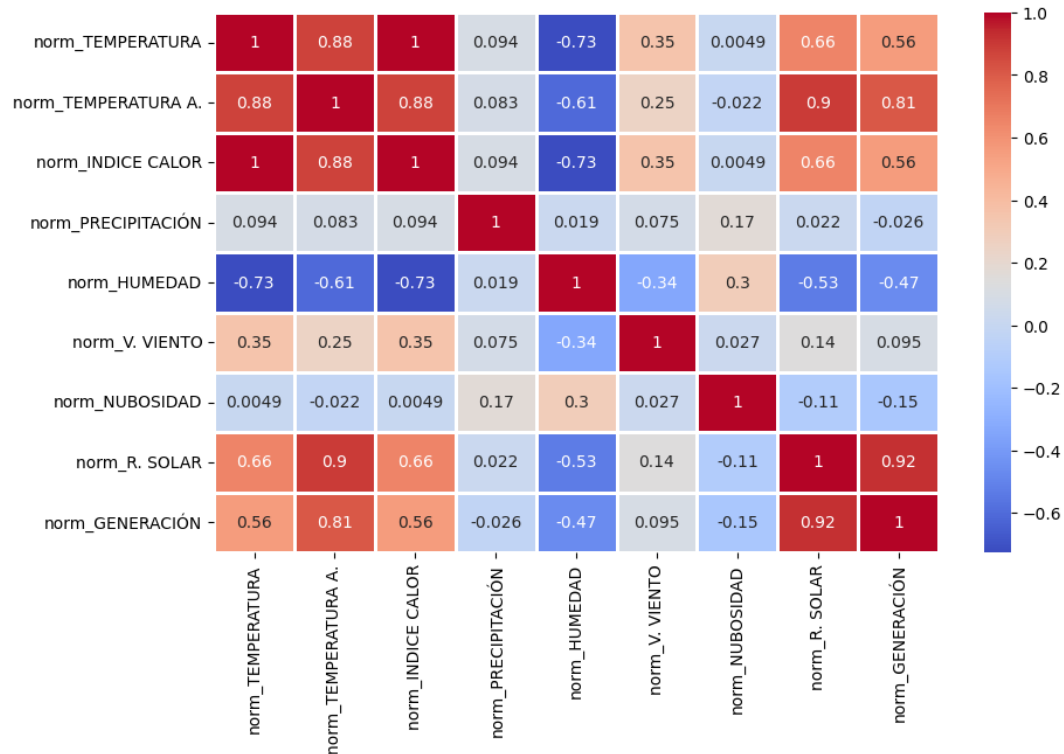


Figura 34 Resultado de correlación.

4.7 Regresión lineal para el conjunto de datos hora-anual

Una vez identificada la variable de irradiancia a través del mapa de calor o de la correlación de variables, se procede a ejecutar la regresión lineal, primero se considera el primer grupo relacionado con horas anual.

La Figura 35 muestra el comportamiento para las horas 7 hasta la hora 19, en ellas se puede apreciar la dispersión de los datos y su comportamiento, en color rojo se muestra la recta que representa dicho comportamiento.

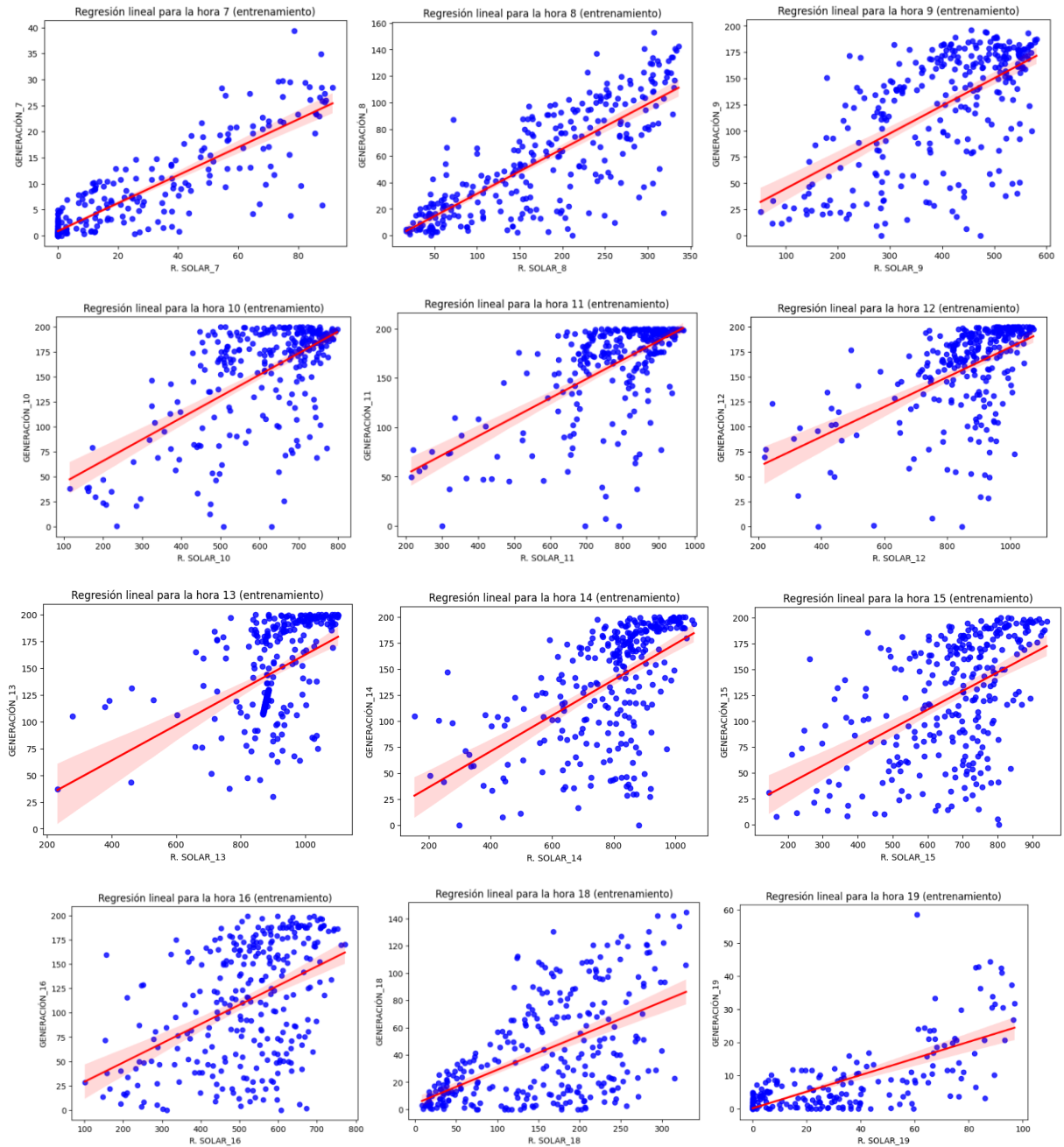


Figura 35 Resultado de la regresión de la hora 7 a la 19.

En algunos casos como los son las horas 15 y 16, se parecía mayor dispersión, en comparación con el resto de las horas.

4.8 Resultados de la regresión lineal grupo horas-anual

Ahora bien, se procede a analizar el resultado de las ecuaciones de la regresión lineal.

La muestra la tabla de resultados con las ecuaciones para cada una de las horas de la clasificación de datos horas-anual, la cual permite efectuar el pronóstico a corto plazo con datos de irradiancia.

Tabla 10. Tabla de resultados con las ecuaciones para cada hora en el grupo de datos hora-anual.

Tabla de ecuaciones y predicciones por hora:

	Hora	Pendiente	Intercepto	Ecuación	Dato Predictor
0	1	0.000	0.000	$y = 0.0x + 0.0$	0.000
1	2	0.000	0.000	$y = 0.0x + 0.0$	0.000
2	3	0.000	0.000	$y = 0.0x + 0.0$	0.000
3	4	0.000	0.000	$y = 0.0x + 0.0$	0.000
4	5	0.000	0.000	$y = 0.0x + 0.0$	0.000
5	6	0.000	0.003	$y = 0.0x + 0.003$	0.000
6	7	0.269	0.880	$y = 0.269x + 0.88$	21.290
7	8	0.338	-2.323	$y = 0.338x + -2.323$	167.790
8	9	0.263	18.569	$y = 0.263x + 18.569$	395.405
9	10	0.216	22.423	$y = 0.216x + 22.423$	608.000
10	11	0.193	13.727	$y = 0.193x + 13.727$	767.000
11	12	0.149	30.088	$y = 0.149x + 30.088$	855.000
12	13	0.164	-2.002	$y = 0.164x + -2.002$	898.000
13	14	0.172	1.958	$y = 0.172x + 1.958$	784.000
14	15	0.180	2.880	$y = 0.18x + 2.88$	663.000
15	16	0.198	9.301	$y = 0.198x + 9.301$	509.000
16	17	0.206	16.928	$y = 0.206x + 16.928$	326.000
17	18	0.249	4.103	$y = 0.249x + 4.103$	147.000
18	19	0.251	0.024	$y = 0.251x + 0.024$	21.530
19	20	0.000	0.019	$y = 0.0x + 0.019$	0.000
20	21	0.000	0.000	$y = 0.0x + 0.0$	0.000
21	22	0.000	0.000	$y = 0.0x + 0.0$	0.000
22	23	0.000	0.000	$y = 0.0x + 0.0$	0.000
23	24	0.000	0.000	$y = 0.0x + 0.0$	0.000

La Figura 36 muestra el gráfico de predicción, el cual se obtiene a través de un vector con 24 datos de irradiancias y las ecuaciones de la Tabla 10.

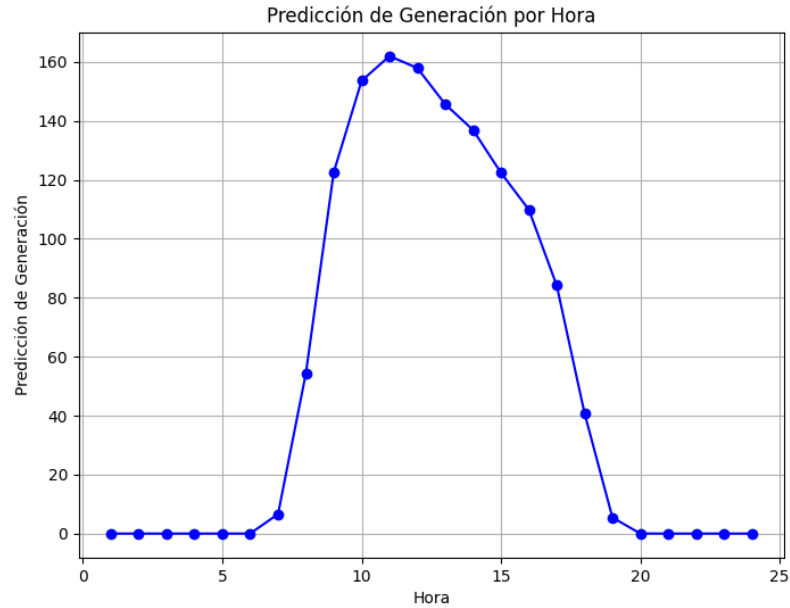


Figura 36 Gráfico de pronóstico para el grupo de datos horas-anual con datos de prueba.

La Figura 37 muestra la comparativa entre los valores de R^2 de entrenamiento y de prueba.

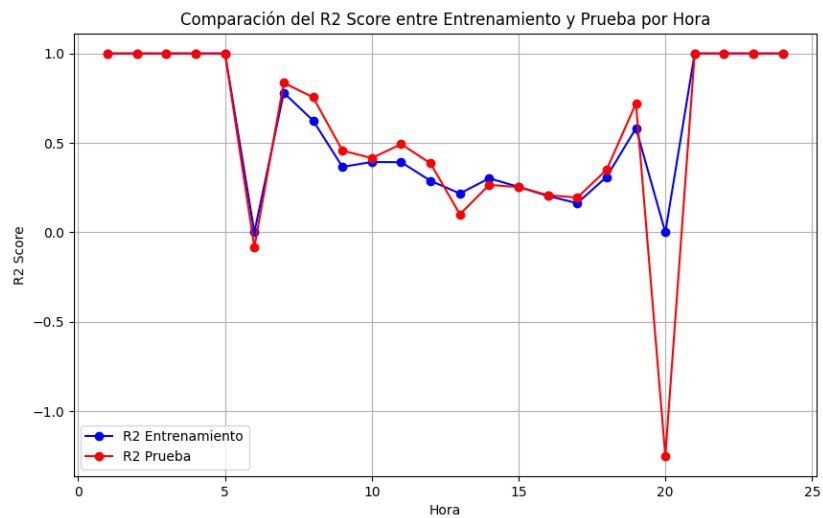


Figura 37 Gráfico de R^2 para datos de entrenamiento y prueba del conjunto horas-anual.

La Figura 38 muestra la comparativa entre los valores de MSE de entrenamiento y de prueba.

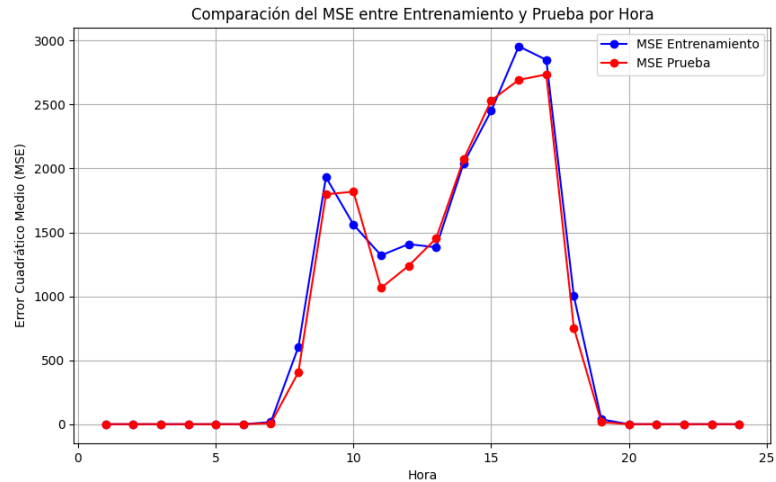
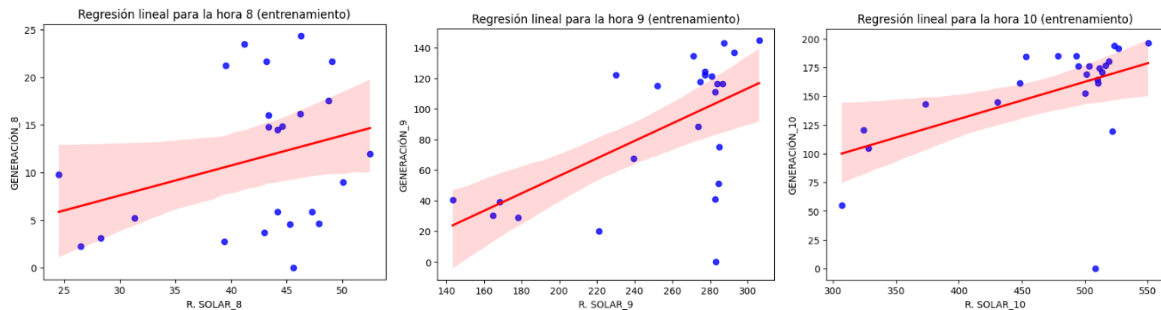


Figura 38 Gráfico de MSE para datos de entrenamiento y prueba del conjunto horas-anual.

4.9 Regresión lineal para el conjunto de datos hora-mensual

Los siguientes resultados corresponden al conjunto de datos agrupados mensualmente, en este apartado se obtienen 12 conjuntos de 24 ecuaciones (una ecuación para cada hora), así como el cálculo de la R^2 y MSE. Figura 39 muestra el conjunto de gráficos de la regresión lineal desde la hora 8 hasta la hora 18 para el mes de enero, para el resto de las horas el resultado es cero o muy cercano a cero.



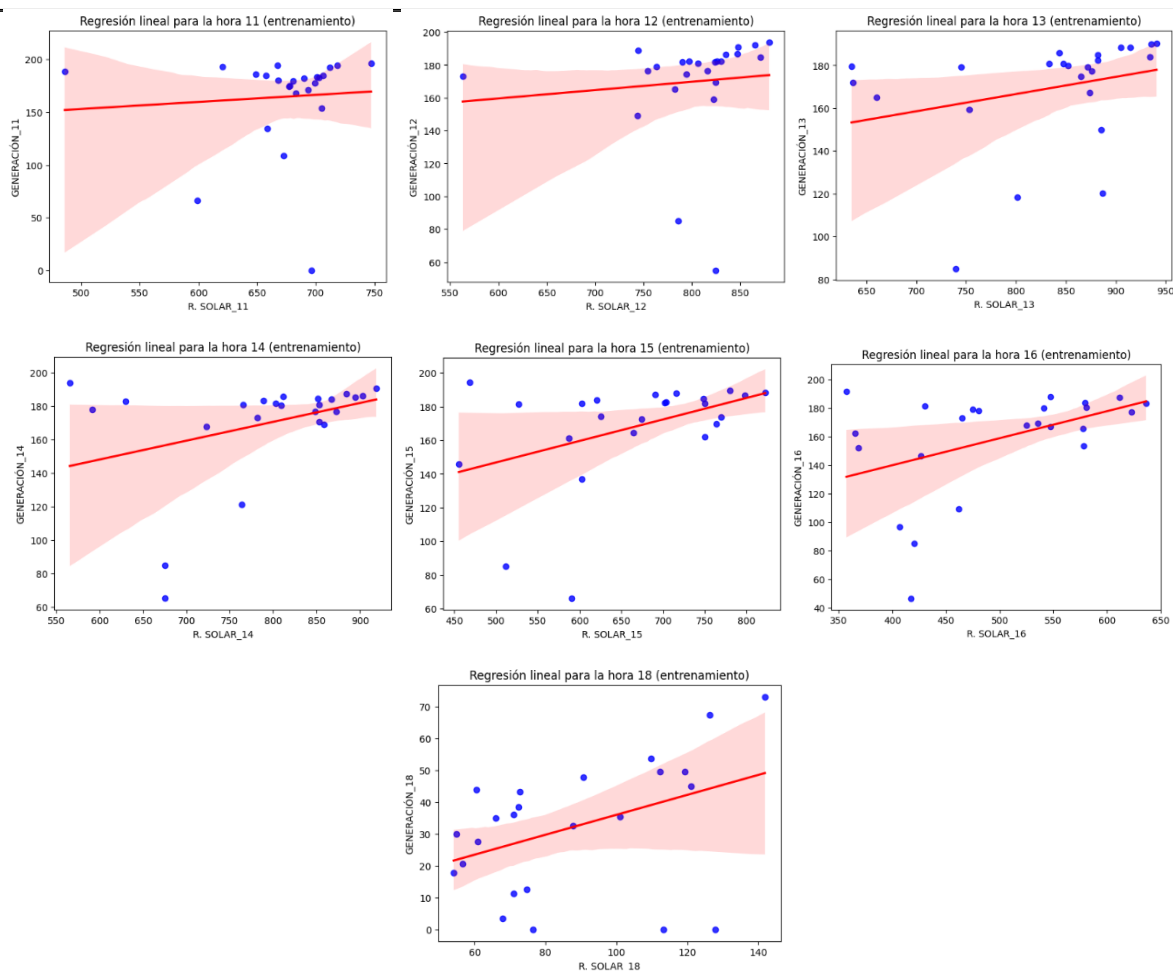
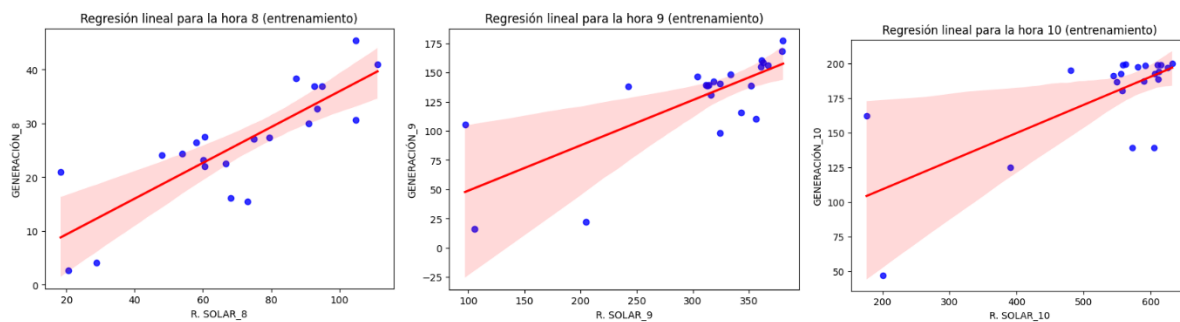


Figura 39 Resultado de la regresión de la hora 8 a la 18 para el mes de enero.

La Figura 40 muestra el conjunto de gráficos de la regresión lineal desde la hora 8 hasta la hora 19 para el mes de febrero, para el resto de las horas el resultado es cero o muy cercano a cero.



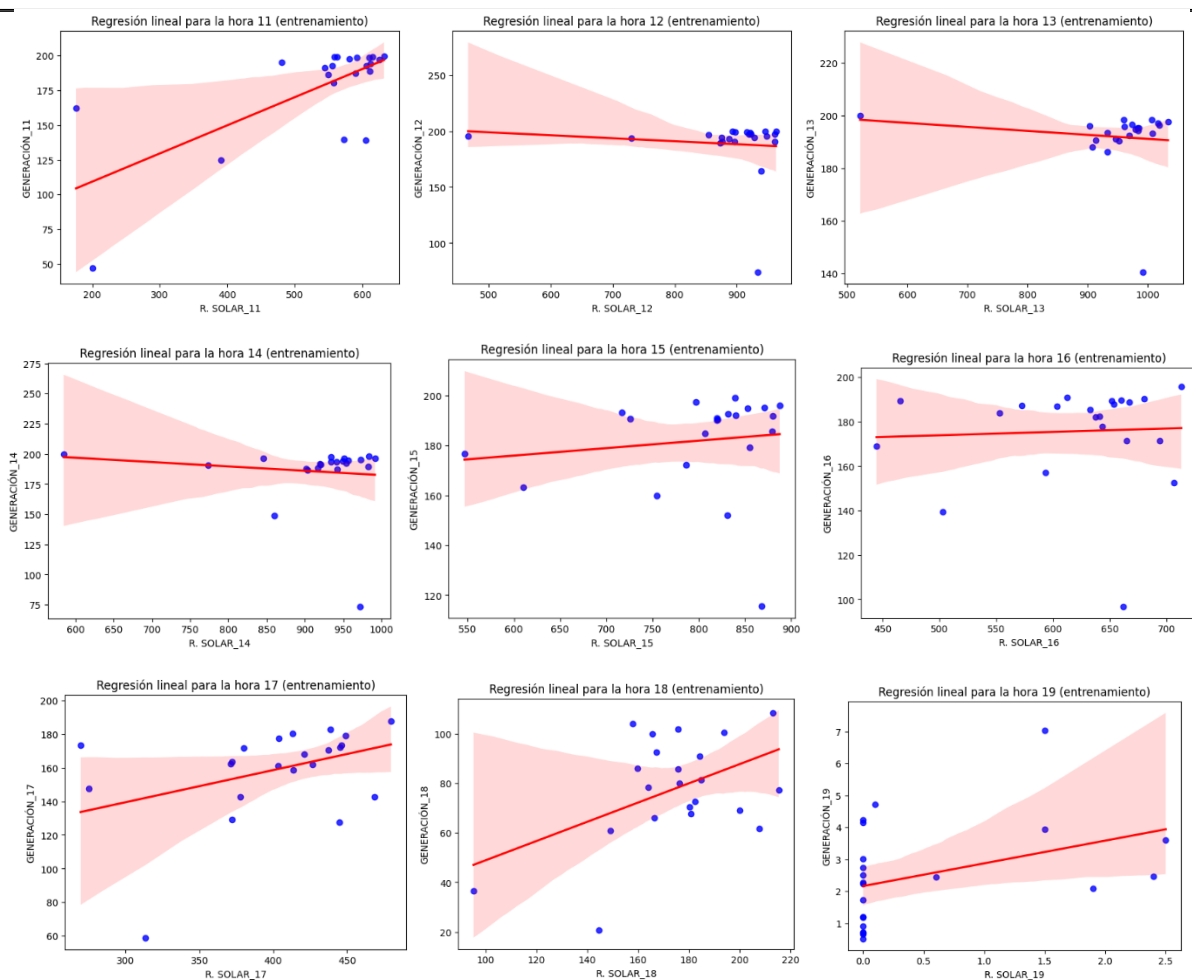
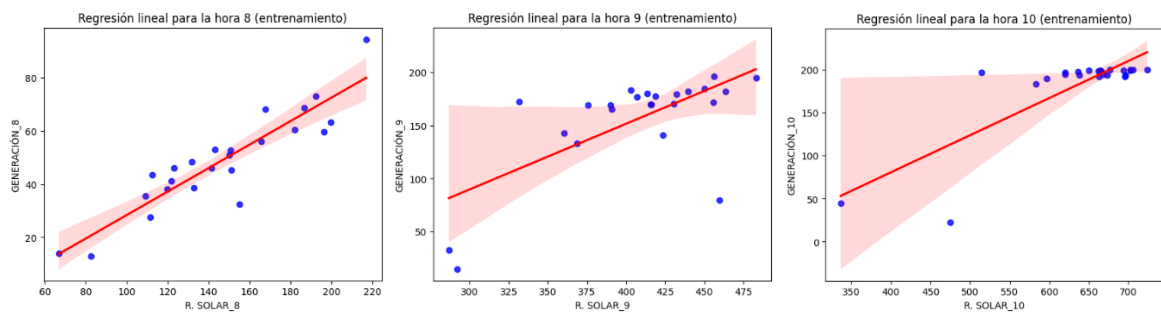


Figura 40 Resultado de la regresión de la hora 8 a la 19 para el mes de febrero.

La Figura 41 muestra el conjunto de gráficos de la regresión lineal desde la hora 8 hasta la hora 19 para el mes de marzo, para el resto de las horas el resultado es cero o muy cercano a cero.



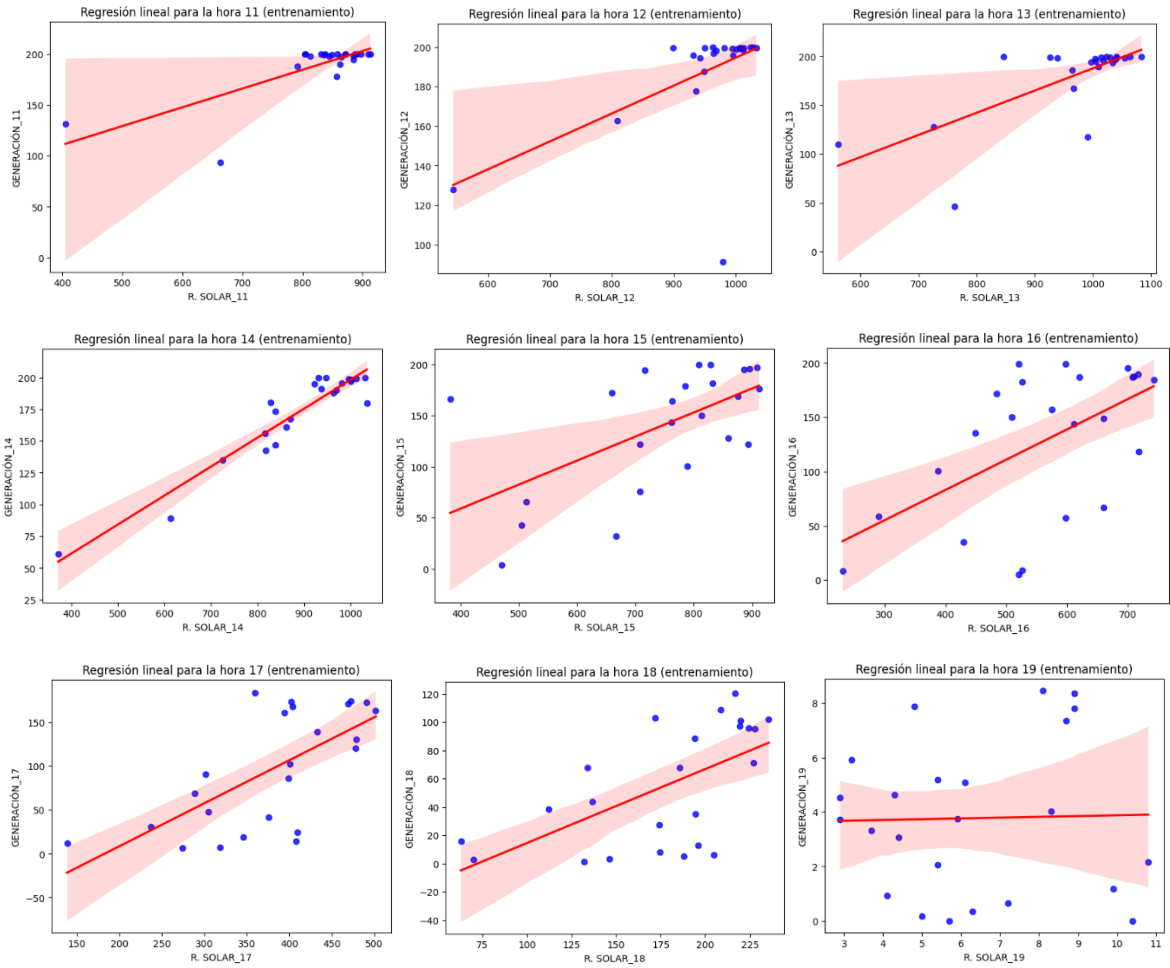
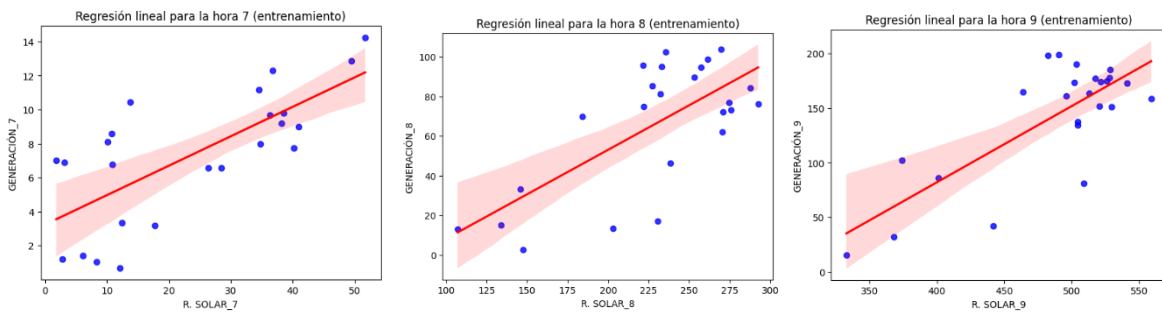


Figura 41 Resultado de la regresión de la hora 8 a la 19 para el mes de marzo.

La Figura 42 muestra el conjunto de gráficos de la regresión lineal desde la hora 7 hasta la hora 19 para el mes de abril, para el resto de las horas el resultado es cero o muy cercano a cero.



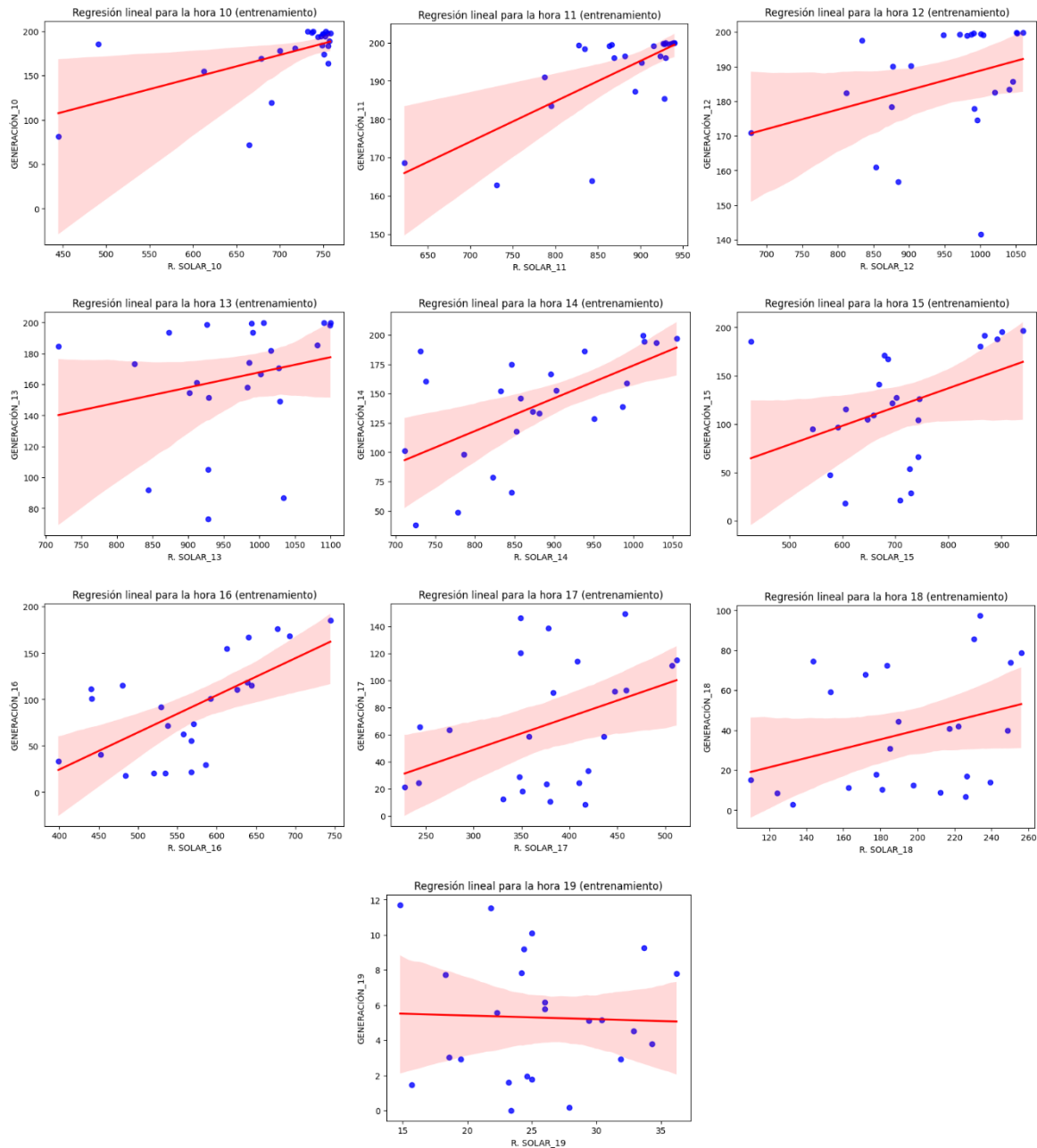
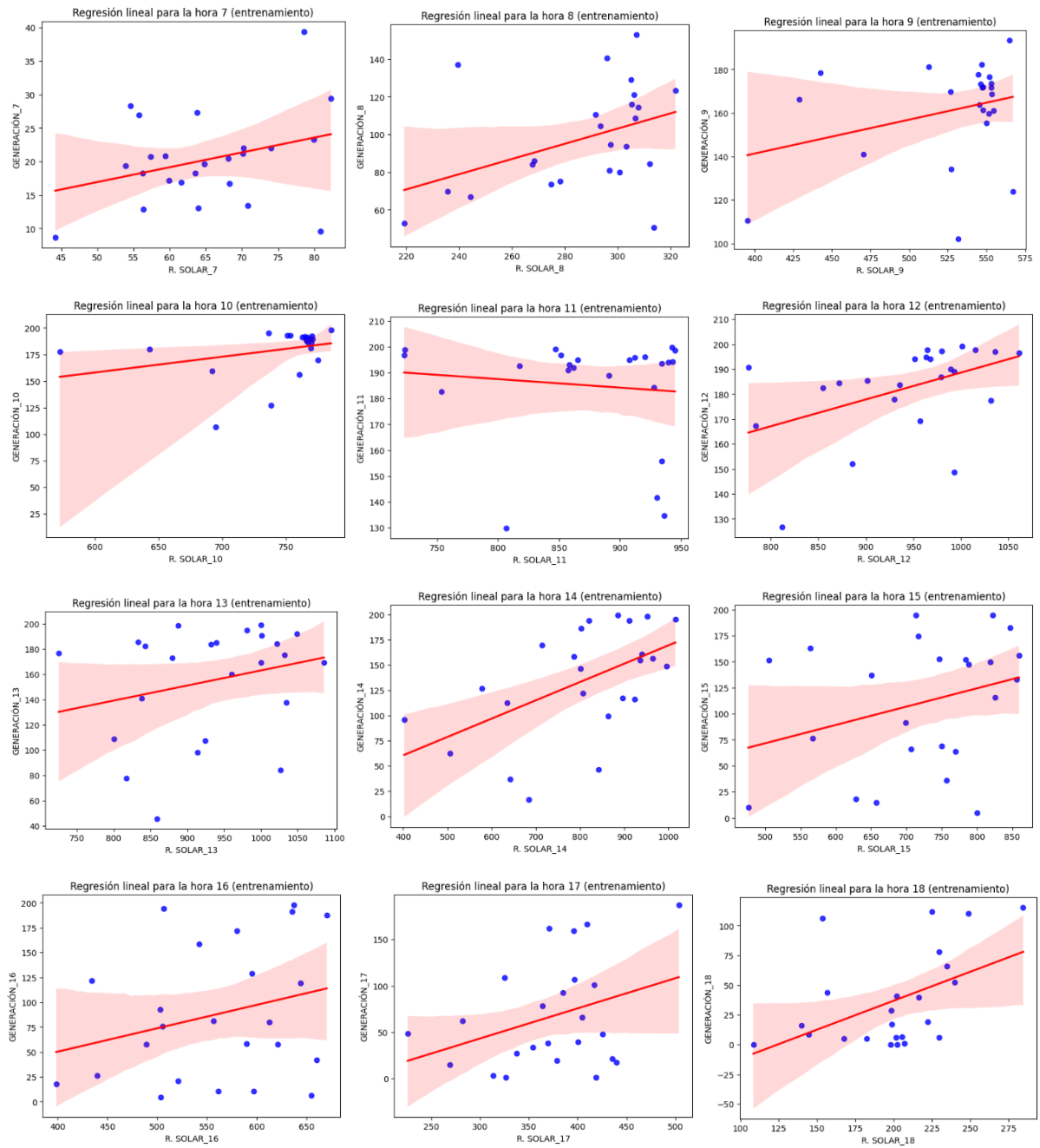


Figura 42 Resultado de la regresión de la hora 7 a la 19 para el mes de abril.

La Figura 43 muestra el conjunto de gráficos de la regresión lineal desde la hora 7 hasta la hora 19 para el mes de mayo, para el resto de las horas el resultado es cero o muy cercano a cero.



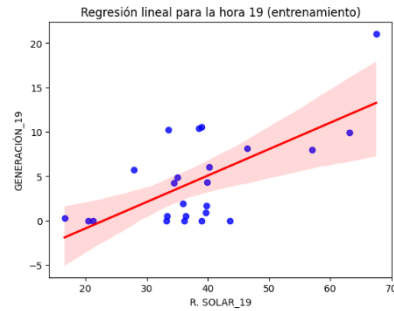
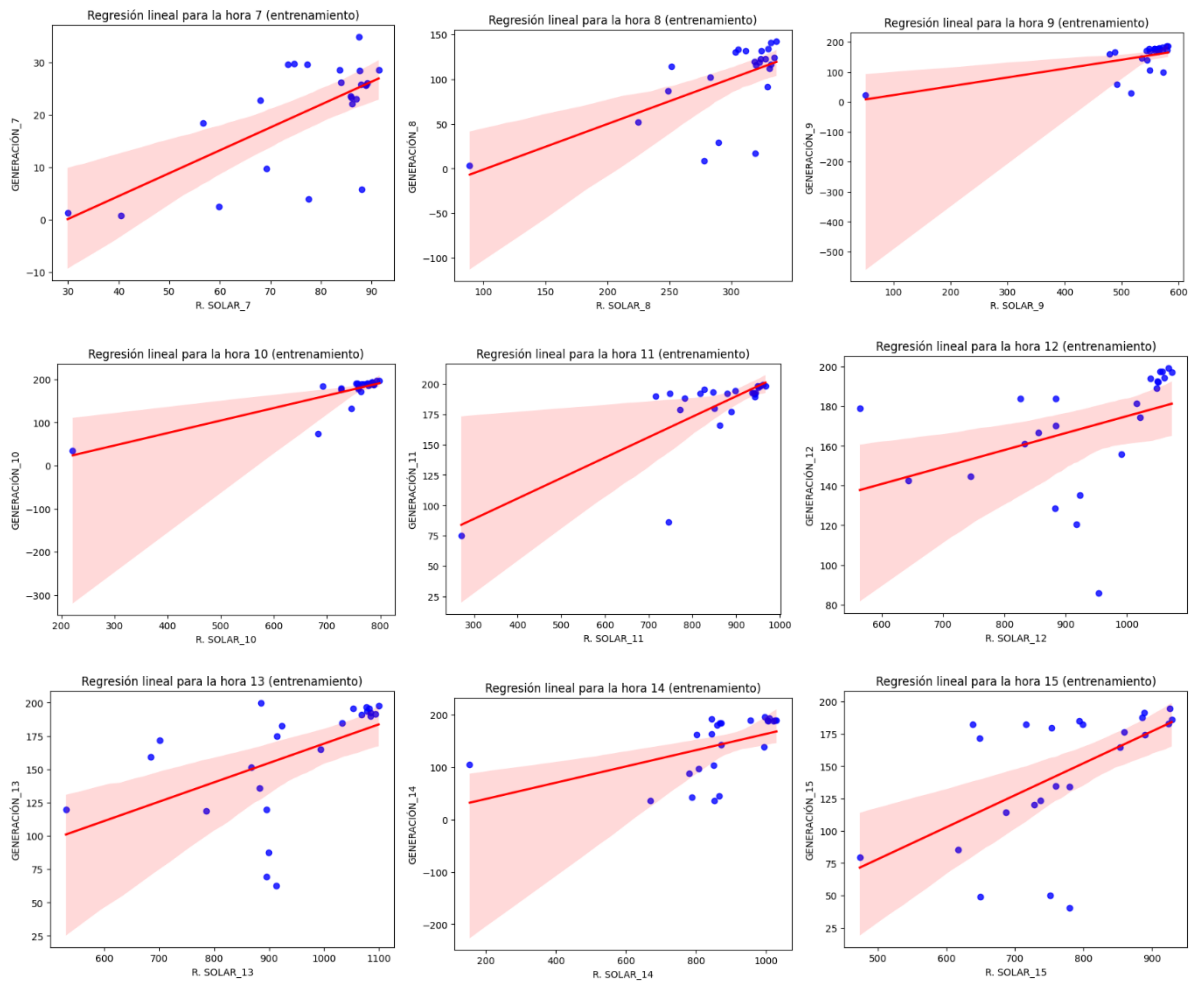


Figura 43 Resultado de la regresión de la hora 7 a la 19 para el mes de mayo.

La Figura 44 muestra el conjunto de gráficos de la regresión lineal desde la hora 7 hasta la hora 19 para el mes de junio, para el resto de las horas el resultado es cero o muy cercano a cero.



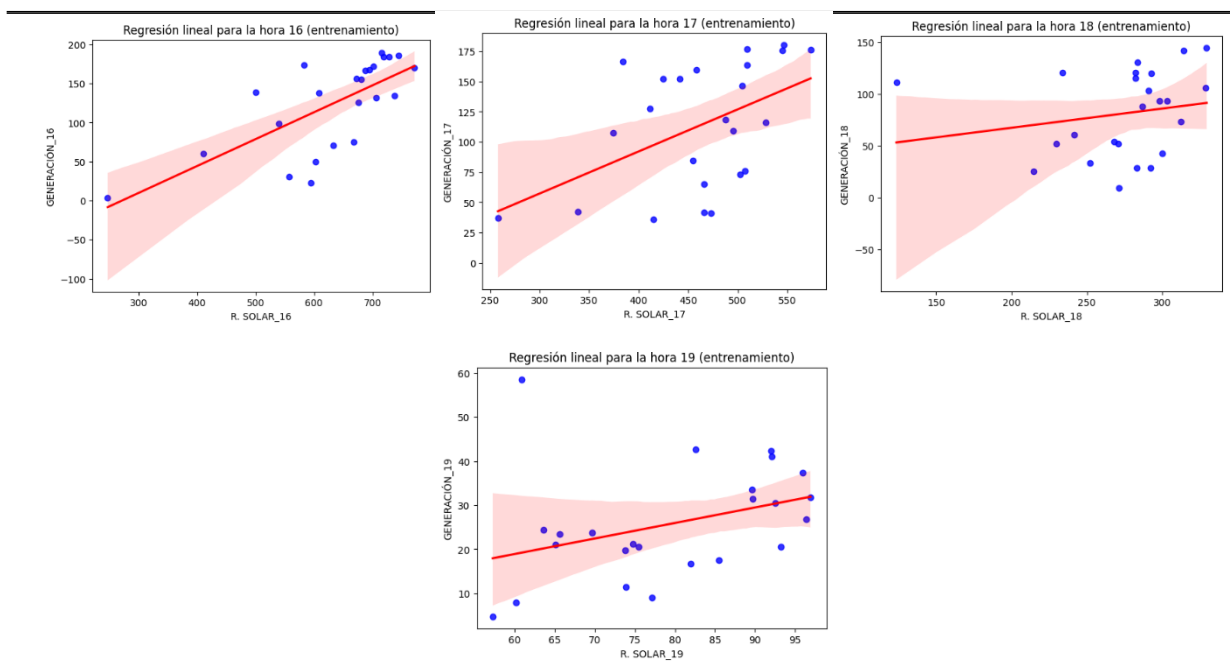
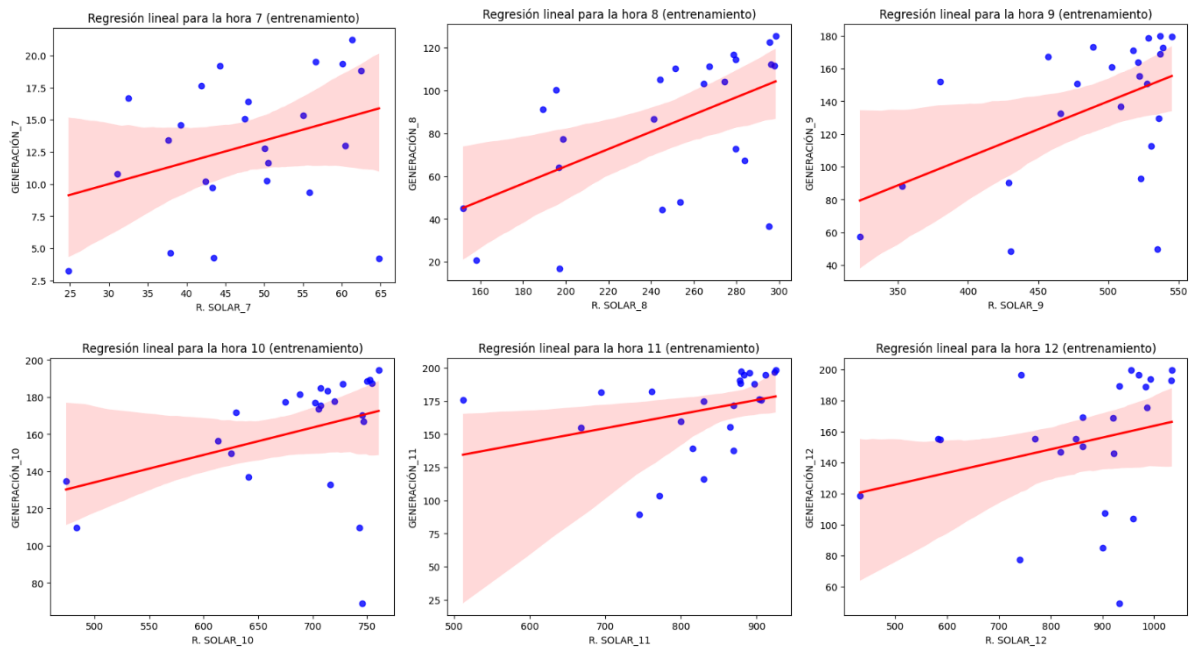


Figura 44 Resultado de la regresión de la hora 7 a la 19 para el mes de junio.

La Figura 45 muestra el conjunto de gráficos de la regresión lineal desde la hora 7 hasta la hora 19 para el mes de julio, para el resto de las horas el resultado es cero o muy cercano a cero.



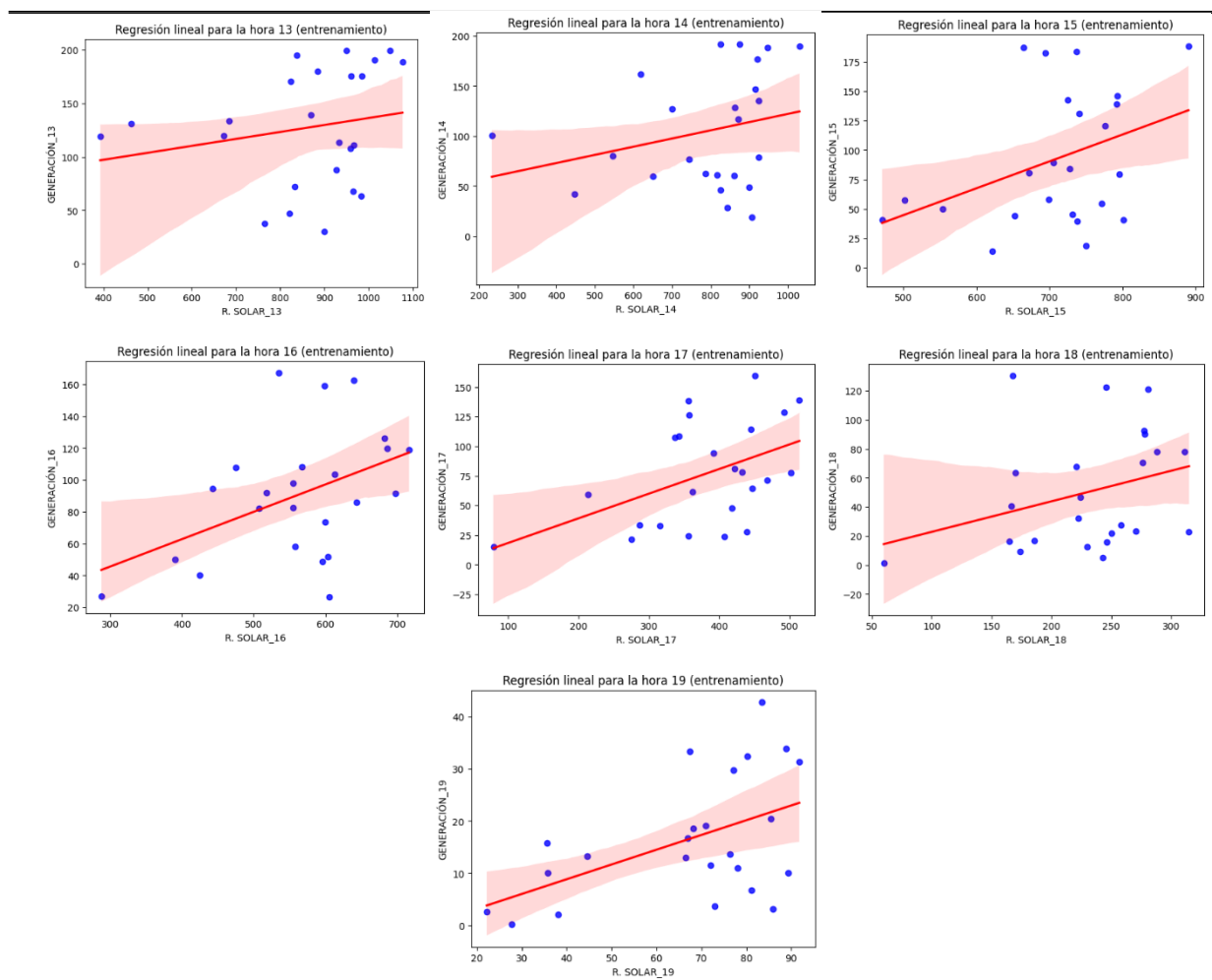
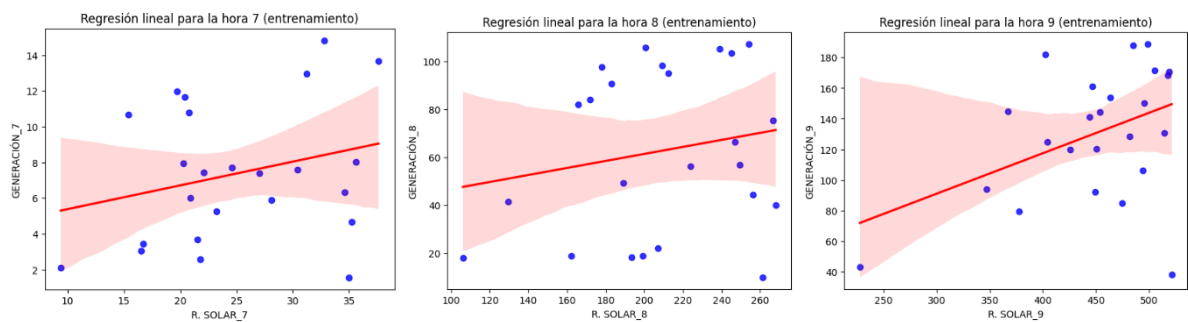


Figura 45 Resultado de la regresión de la hora 7 a la 19 para el mes de julio.

La Figura 46 muestra el conjunto de gráficos de la regresión lineal desde la hora 7 hasta la hora 19 para el mes de agosto, para el resto de las horas el resultado es cero o muy cercano a cero.



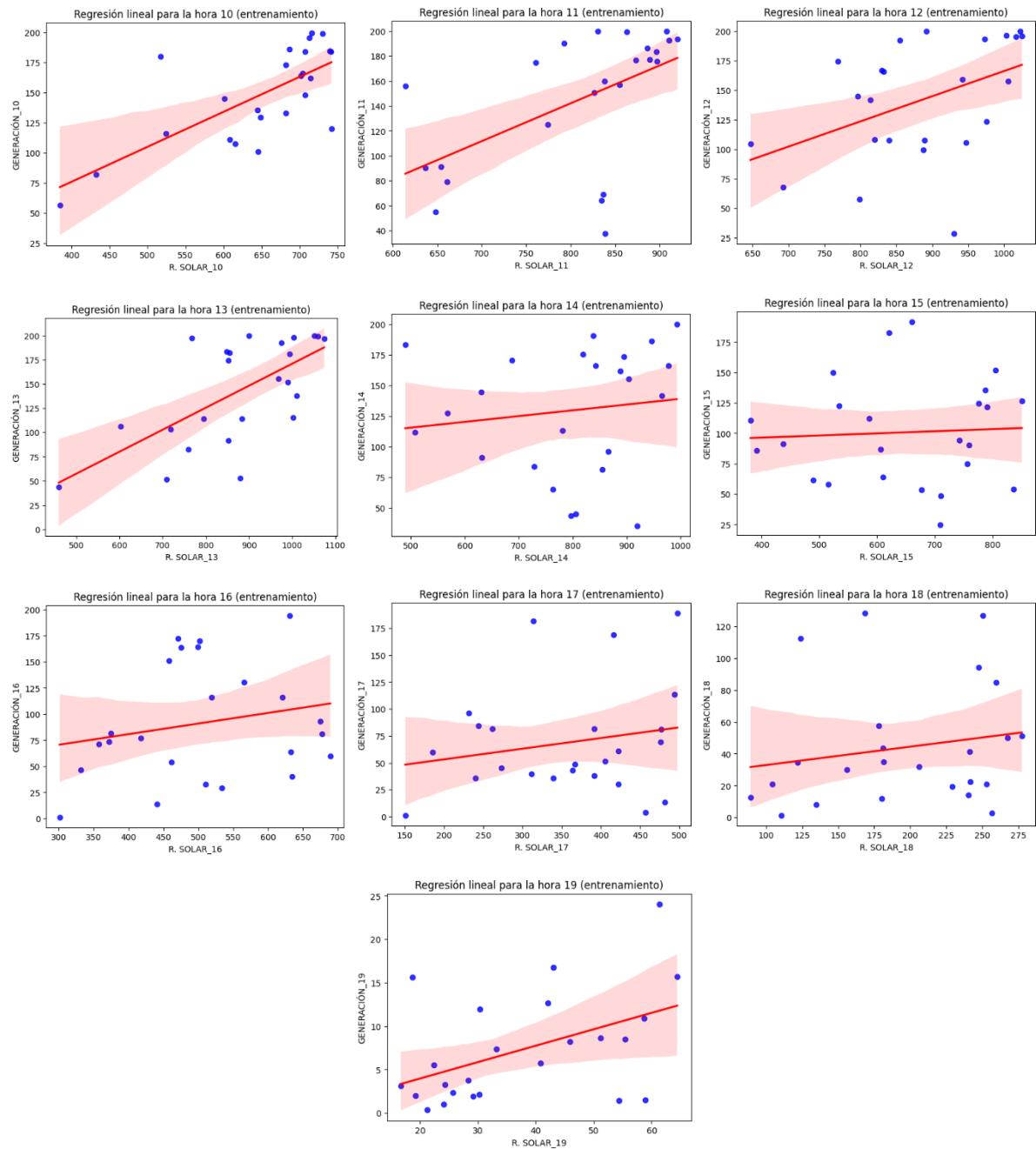
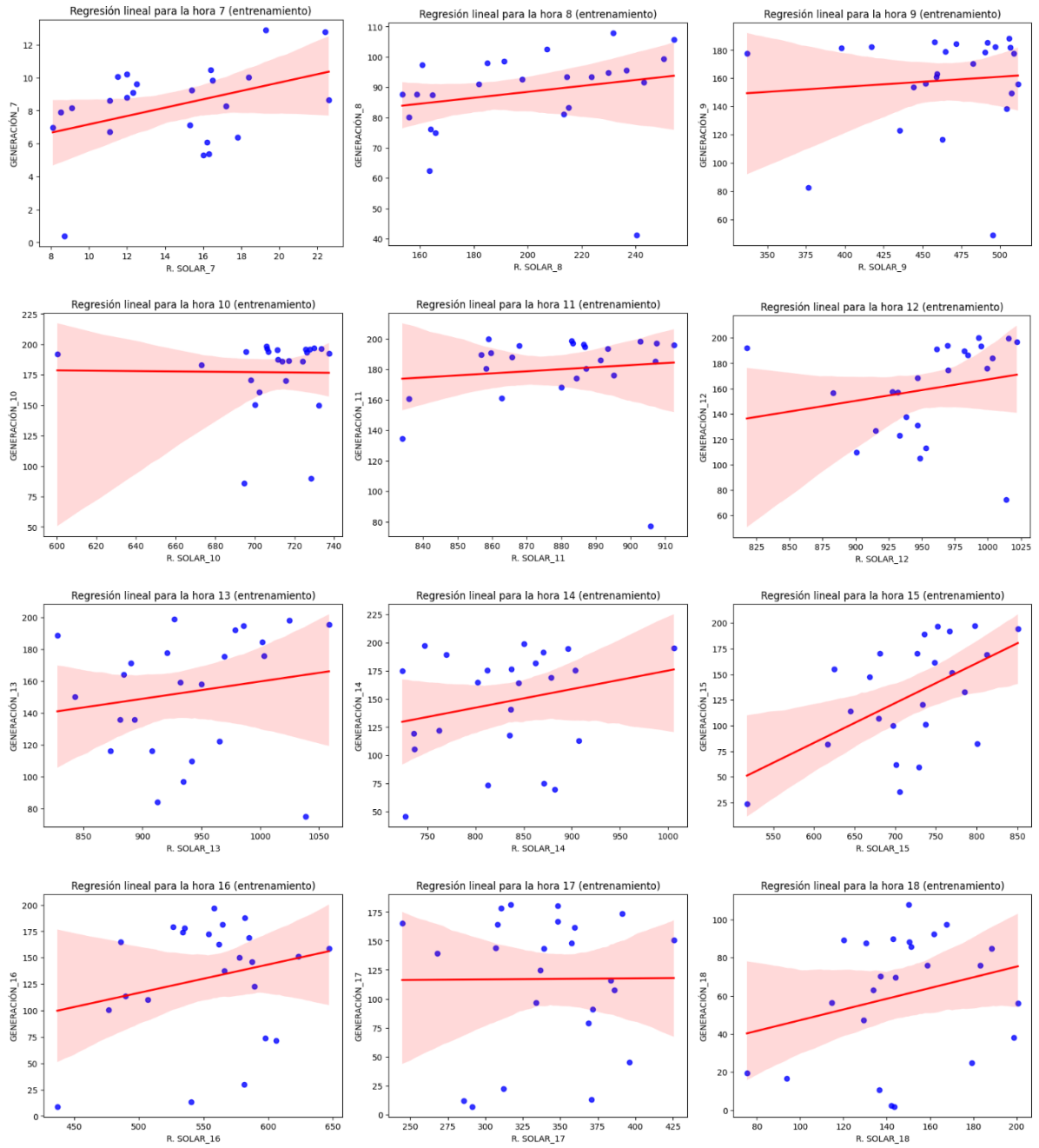


Figura 46 Resultado de la regresión de la hora 7 a la 19 para el mes de agosto

La Figura 47 muestra el conjunto de gráficos de la regresión lineal desde la hora 7 hasta la hora 19 para el mes de septiembre, para el resto de las horas el resultado es cero o muy cercano a cero.



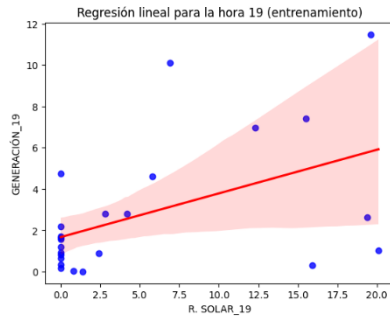
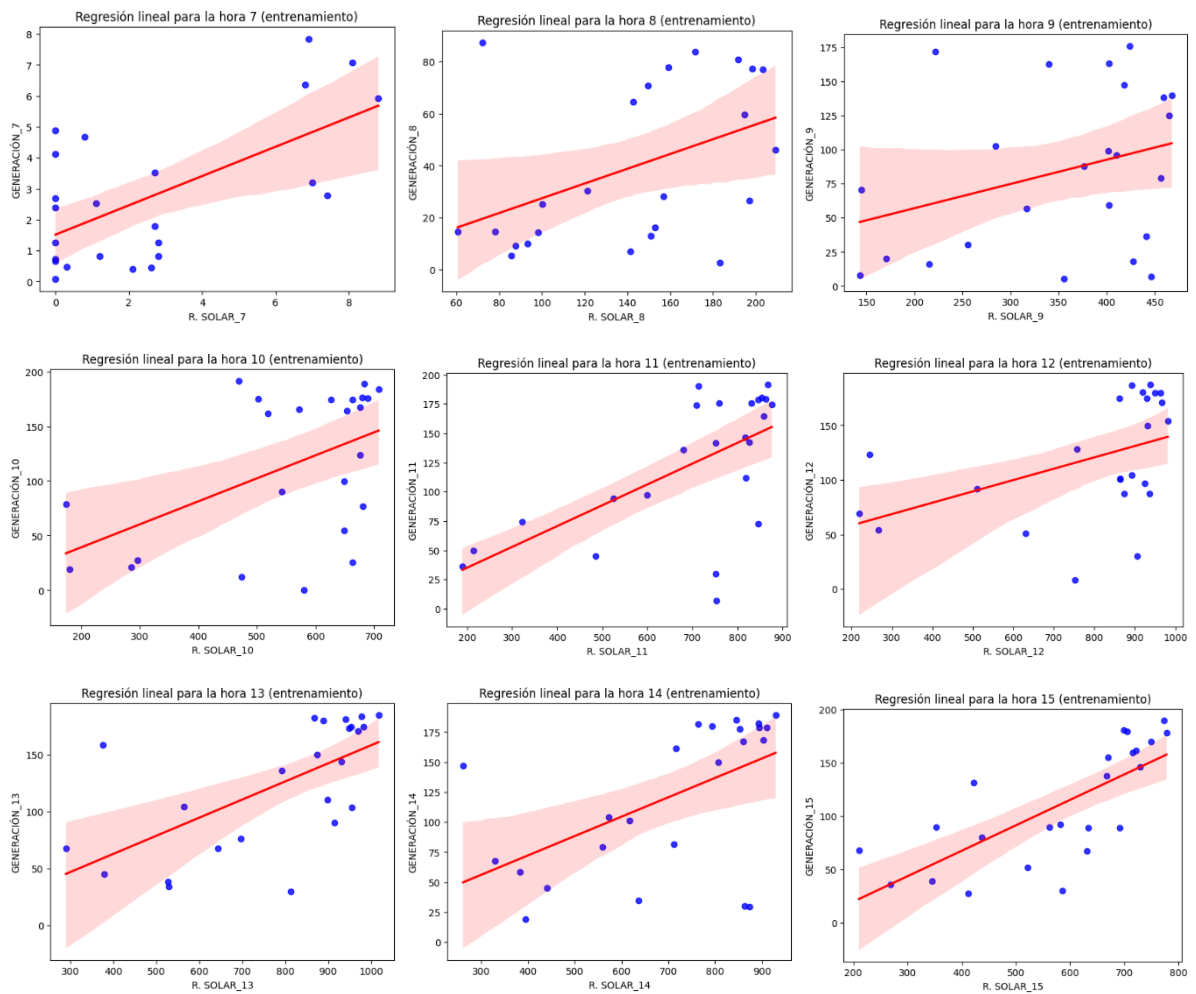


Figura 47 Resultado de la regresión de la hora 7 a la 19 para el mes de septiembre.

La Figura 48 muestra el conjunto de gráficos de la regresión lineal desde la hora 7 hasta la hora 18 para el mes de octubre, para el resto de las horas el resultado es cero o muy cercano a cero.



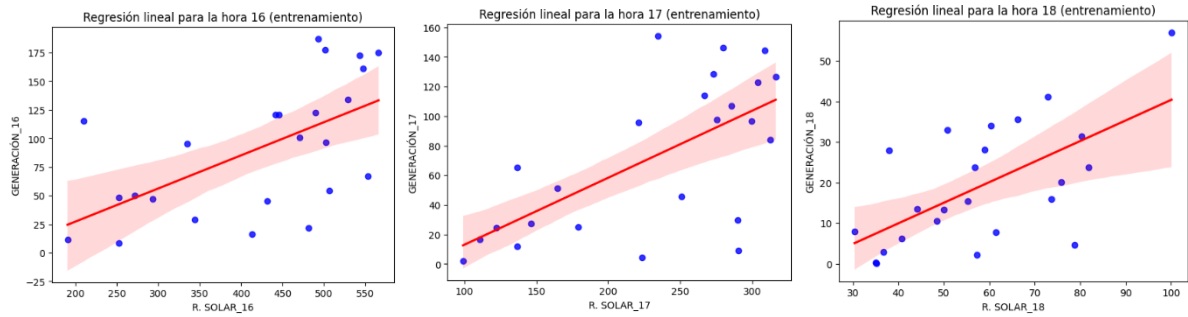
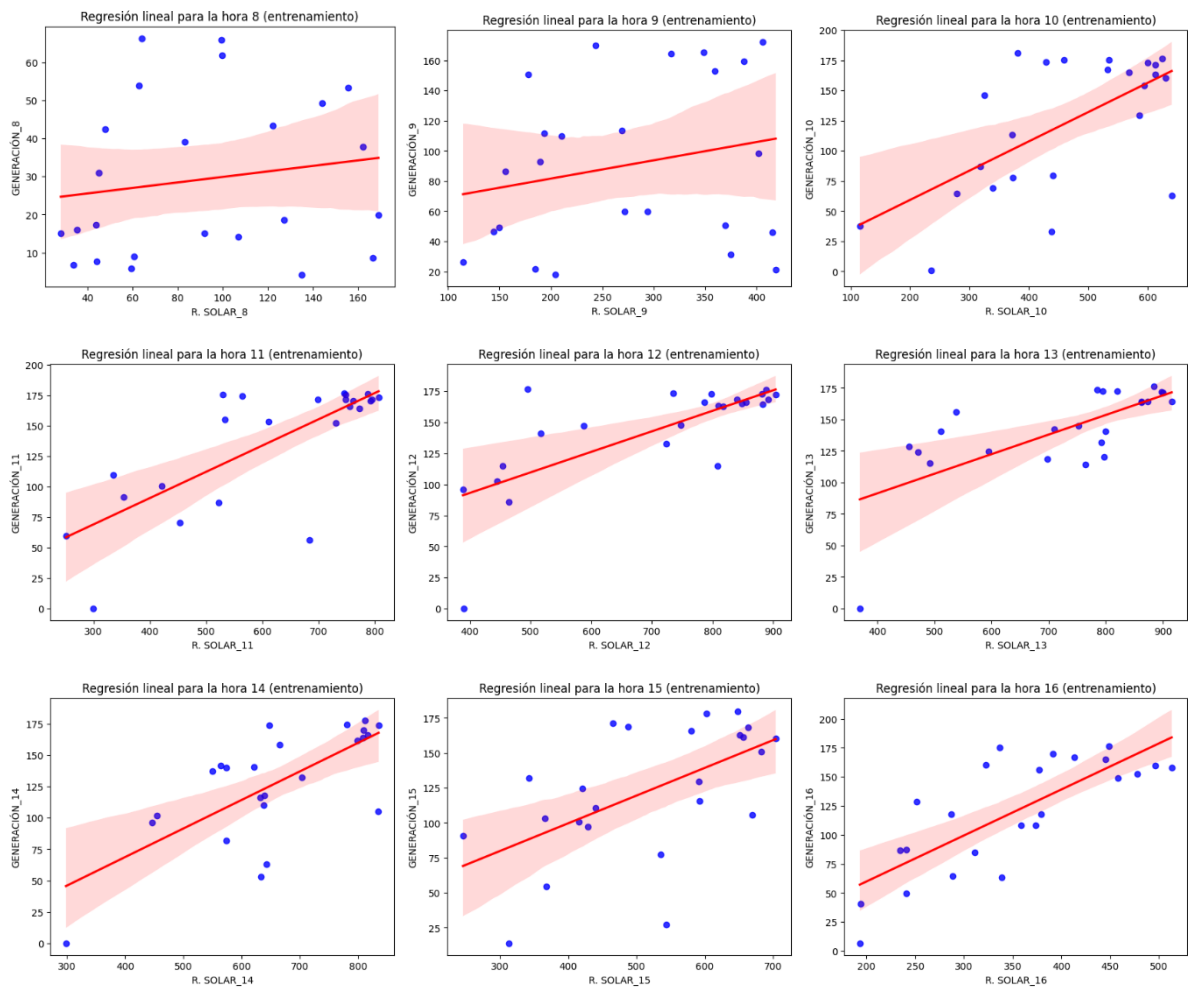


Figura 48 Resultado de la regresión de la hora 7 a la 18 para el mes de octubre.

La Figura 49 muestra el conjunto de gráficos de la regresión lineal desde la hora 8 hasta la hora 18 para el mes de noviembre, para el resto de las horas el resultado es cero o muy cercano a cero.



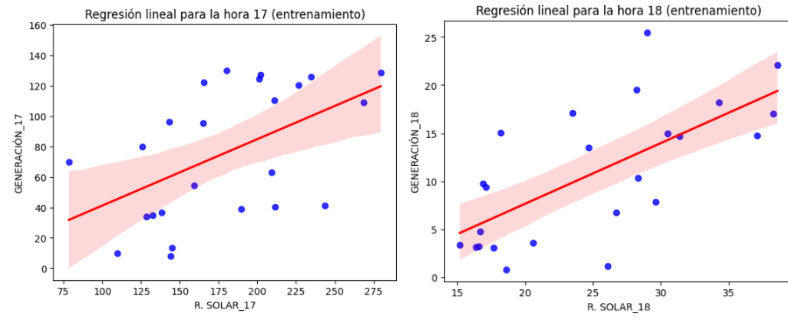
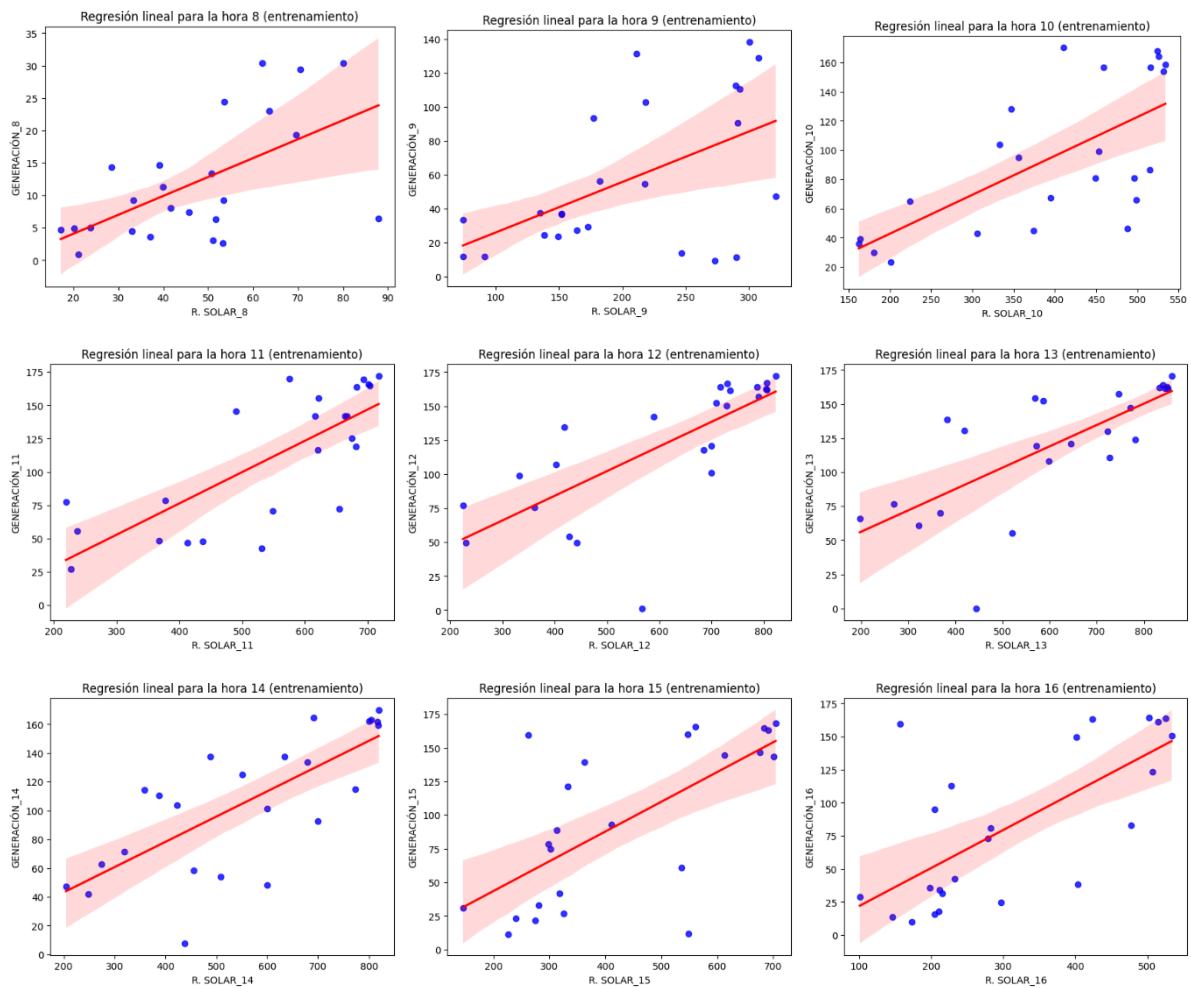


Figura 49 Resultado de la regresión de la hora 8 a la 18 la para el mes de noviembre.

La Figura 43 muestra el conjunto de gráficos de la regresión lineal desde la hora 8 hasta la hora 18 para el mes de diciembre, para el resto de las horas el resultado es cero o muy cercano a cero.



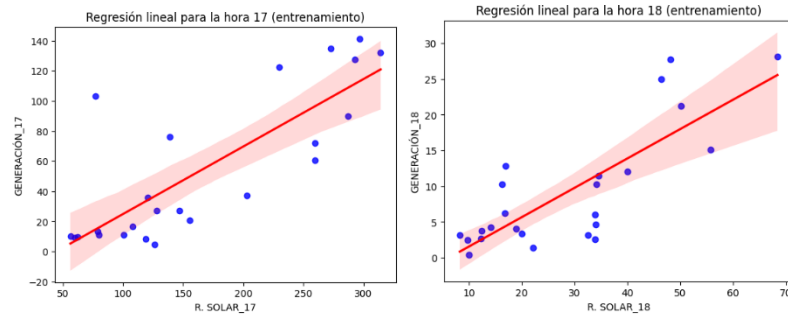


Figura 50 Resultado de la regresión de la hora 8 a la 18 para el mes de diciembre.

4.10 Resultados de la regresión lineal grupo horas-mensual

Ahora bien, se procede a analizar el resultado de las ecuaciones de la regresión lineal.

La Tabla 11 y la Tabla 12 muestran los resultados con las ecuaciones para cada una de las horas de la clasificación de datos horas-mensual, la cual permite efectuar el pronóstico a corto plazo con datos de irradiancia.

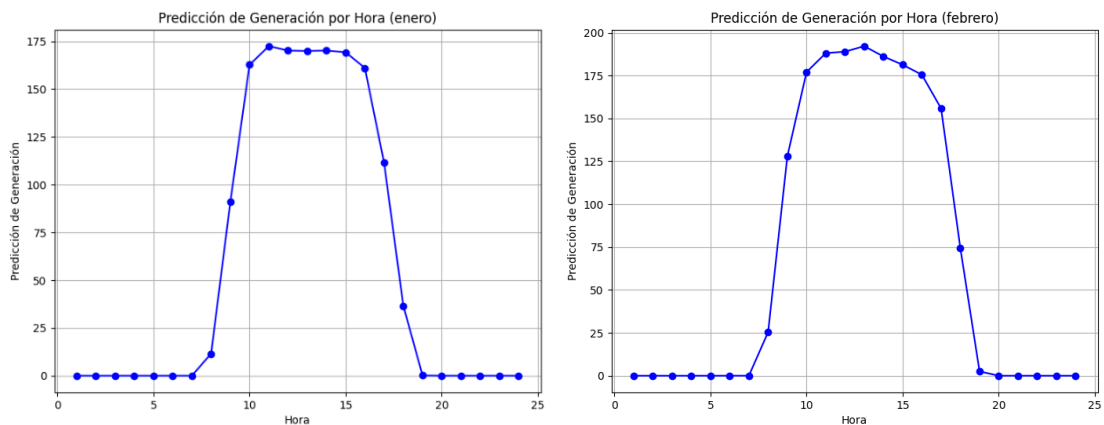
Tabla 11. Tabla de resultados con las ecuaciones para cada hora en el grupo de datos hora-mes.

Hora	Ecuación					
	enero	febrero	marzo	abril	mayo	junio
1	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
2	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
3	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
4	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
5	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
6	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.003$	$y = 0.0x + 0.03$
7	$y = 0.0x + 0.0$	$y = 0.0x + 0.014$	$y = 2.996x + 1.17$	$y = 0.173x + 3.236$	$y = 0.221x + 5.873$	$y = 0.436x + -12.917$
8	$y = 0.315x + -1.855$	$y = 0.332x + 2.739$	$y = 0.442x + -15.77$	$y = 0.449x + -36.865$	$y = 0.404x + -18.149$	$y = 0.511x + -52.372$
9	$y = 0.572x + -58.132$	$y = 0.387x + 10.14$	$y = 0.618x + -95.629$	$y = 0.697x + -196.675$	$y = 0.156x + 79.017$	$y = 0.295x + -6.734$
10	$y = 0.324x + 0.584$	$y = 0.203x + 68.692$	$y = 0.432x + -92.018$	$y = 0.257x + -6.86$	$y = 0.148x + 68.902$	$y = 0.289x + -40.457$
11	$y = 0.067x + 119.51$	$y = 0.203x + 68.692$	$y = 0.184x + 37.04$	$y = 0.105x + 100.295$	$y = -0.033x + 213.839$	$y = 0.168x + 38.442$
12	$y = 0.051x + 129.105$	$y = -0.027x + 212.176$	$y = 0.141x + 53.522$	$y = 0.056x + 132.597$	$y = 0.108x + 80.966$	$y = 0.086x + 89.547$
13	$y = 0.08x + 102.166$	$y = -0.015x + 206.228$	$y = 0.227x + -39.33$	$y = 0.098x + 69.967$	$y = 0.119x + 43.758$	$y = 0.145x + 24.393$
14	$y = 0.112x + 80.659$	$y = -0.036x + 218.494$	$y = 0.228x + -29.727$	$y = 0.279x + -104.962$	$y = 0.181x + -12.092$	$y = 0.155x + 8.462$
15	$y = 0.128x + 82.941$	$y = 0.03x + 158.009$	$y = 0.235x + -34.86$	$y = 0.194x + -18.242$	$y = 0.176x + -16.236$	$y = 0.247x + -45.279$
16	$y = 0.189x + 64.274$	$y = 0.015x + 166.274$	$y = 0.279x + -28.493$	$y = 0.401x + -135.772$	$y = 0.236x + -43.967$	$y = 0.345x + -93.64$
17	$y = 0.227x + 42.326$	$y = 0.191x + 82.135$	$y = 0.49x + -89.421$	$y = 0.243x + -23.857$	$y = 0.325x + -54.31$	$y = 0.348x + -46.964$
18	$y = 0.313x + 4.713$	$y = 0.388x + 10.17$	$y = 0.523x + -37.796$	$y = 0.232x + -6.442$	$y = 0.487x + -60.55$	$y = 0.186x + 30.226$
19	$y = 0.0x + 0.15$	$y = 0.711x + 2.164$	$y = 0.029x + 3.592$	$y = -0.021x + 5.828$	$y = 0.298x + -6.823$	$y = 0.353x + -2.303$
20	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.128$
21	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
22	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
23	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
24	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$

Tabla 12. Tabla de resultados con las ecuaciones para cada hora en el grupo de datos hora-mes continuación.

Hora	Ecuación					
	julio	agosto	septiembre	octubre	noviembre	diciembre
1	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
2	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
3	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
4	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
5	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
6	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
7	$y = 0.169x + 4.927$	$y = 0.133x + 4.055$	$y = 0.253x + 4.631$	$y = 0.474x + 1.514$	$y = 0.0x + 0.771$	$y = 0.0x + 0.03$
8	$y = 0.404x + -16.229$	$y = 0.146x + 32.081$	$y = 0.098x + 68.863$	$y = 0.284x + -0.963$	$y = 0.072x + 22.64$	$y = 0.291x + -1.745$
9	$y = 0.342x + -31.031$	$y = 0.264x + 11.916$	$y = 0.072x + 125.254$	$y = 0.178x + 21.415$	$y = 0.121x + 57.329$	$y = 0.297x + -3.709$
10	$y = 0.148x + 60.347$	$y = 0.291x + -40.473$	$y = -0.015x + 187.457$	$y = 0.21x + -2.735$	$y = 0.243x + 10.679$	$y = 0.266x + -10.496$
11	$y = 0.106x + 80.025$	$y = 0.303x + -100.724$	$y = 0.135x + 61.498$	$y = 0.178x + -0.56$	$y = 0.216x + 4.314$	$y = 0.235x + -17.608$
12	$y = 0.076x + 87.932$	$y = 0.214x + -47.376$	$y = 0.169x + -1.893$	$y = 0.104x + 37.498$	$y = 0.165x + 26.932$	$y = 0.181x + 11.721$
13	$y = 0.065x + 71.251$	$y = 0.227x + -56.174$	$y = 0.109x + 50.867$	$y = 0.159x + -0.838$	$y = 0.155x + 29.273$	$y = 0.157x + 24.979$
14	$y = 0.082x + 40.269$	$y = 0.047x + 91.914$	$y = 0.164x + 10.751$	$y = 0.161x + 7.749$	$y = 0.227x + -22.067$	$y = 0.176x + 7.964$
15	$y = 0.229x + -69.738$	$y = 0.017x + 89.392$	$y = 0.388x + -149.454$	$y = 0.239x + -28.187$	$y = 0.198x + 20.407$	$y = 0.222x + -0.889$
16	$y = 0.172x + -5.857$	$y = 0.101x + 40.133$	$y = 0.268x + -17.555$	$y = 0.29x + -30.72$	$y = 0.396x + -19.422$	$y = 0.287x + -6.77$
17	$y = 0.208x + -2.464$	$y = 0.099x + 33.413$	$y = 0.009x + 113.935$	$y = 0.453x + -32.266$	$y = 0.437x + -2.587$	$y = 0.448x + -20.02$
18	$y = 0.21x + 1.715$	$y = 0.116x + 21.324$	$y = 0.281x + 19.066$	$y = 0.507x + -10.352$	$y = 0.632x + -4.997$	$y = 0.411x + -2.565$
19	$y = 0.283x + -2.47$	$y = 0.189x + 0.173$	$y = 0.211x + 1.671$	$y = 0.0x + 0.02$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
20	$y = 0.0x + 0.083$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
21	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
22	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
23	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
24	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$

La Figura 51 y Figura 52 muestran los gráficos de predicción, los cuales se obtienen a través de un vector con 24 datos de irradiancias y las ecuaciones de la **¡Error! No se encuentra el origen de la referencia.** y la **¡Error! No se encuentra el origen de la referencia..** Para cada mes, de enero a junio y de julio a diciembre.



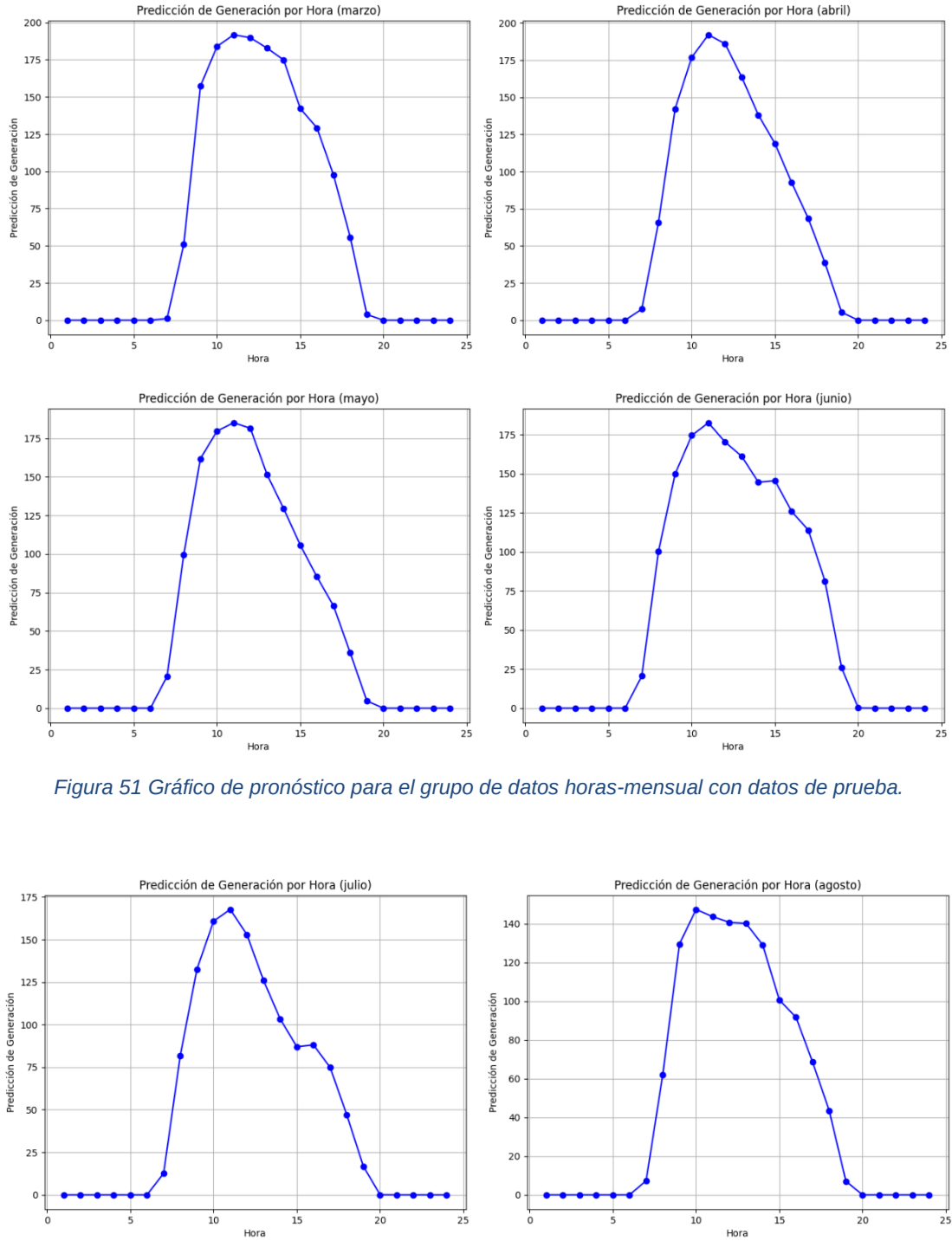


Figura 51 Gráfico de pronóstico para el grupo de datos horas-mensual con datos de prueba.

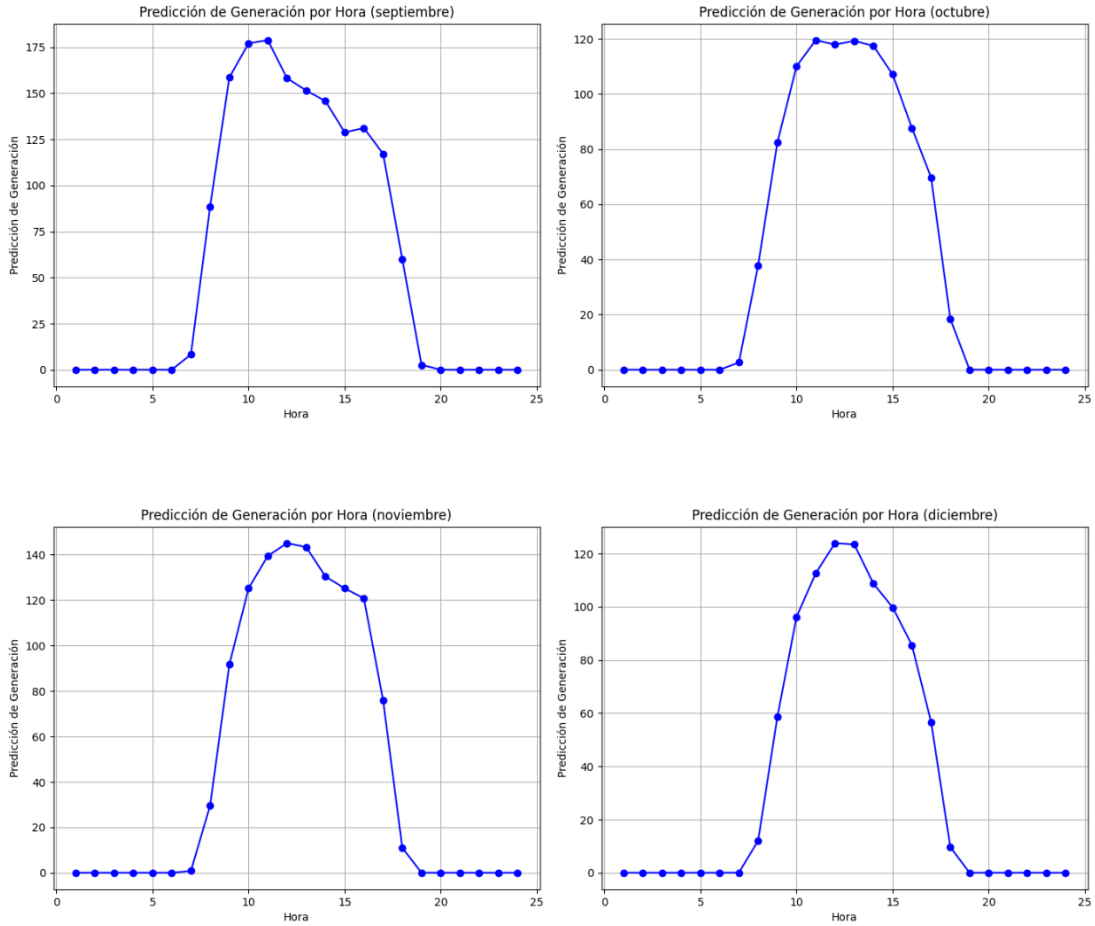
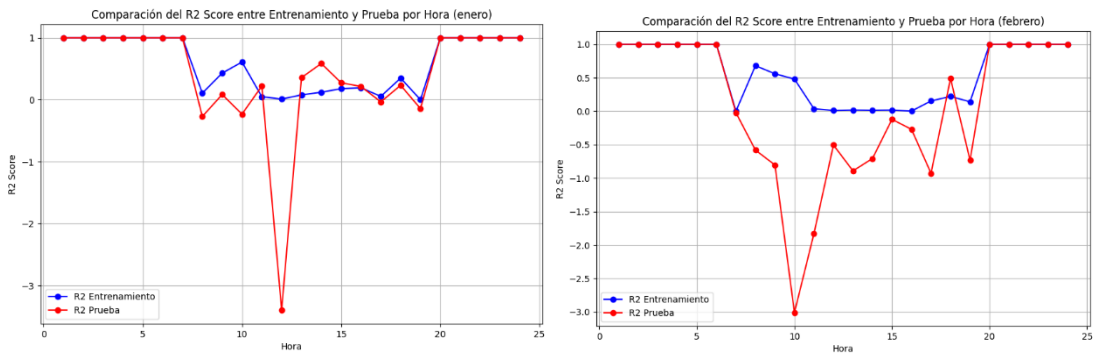


Figura 52 Gráfico de pronóstico para el grupo de datos horas-mensual con datos de prueba julio a diciembre.

La Figura 53 y la Figura 54 muestran la comparativa entre los valores de R^2 de entrenamiento y de prueba para el grupo horas-mensual.



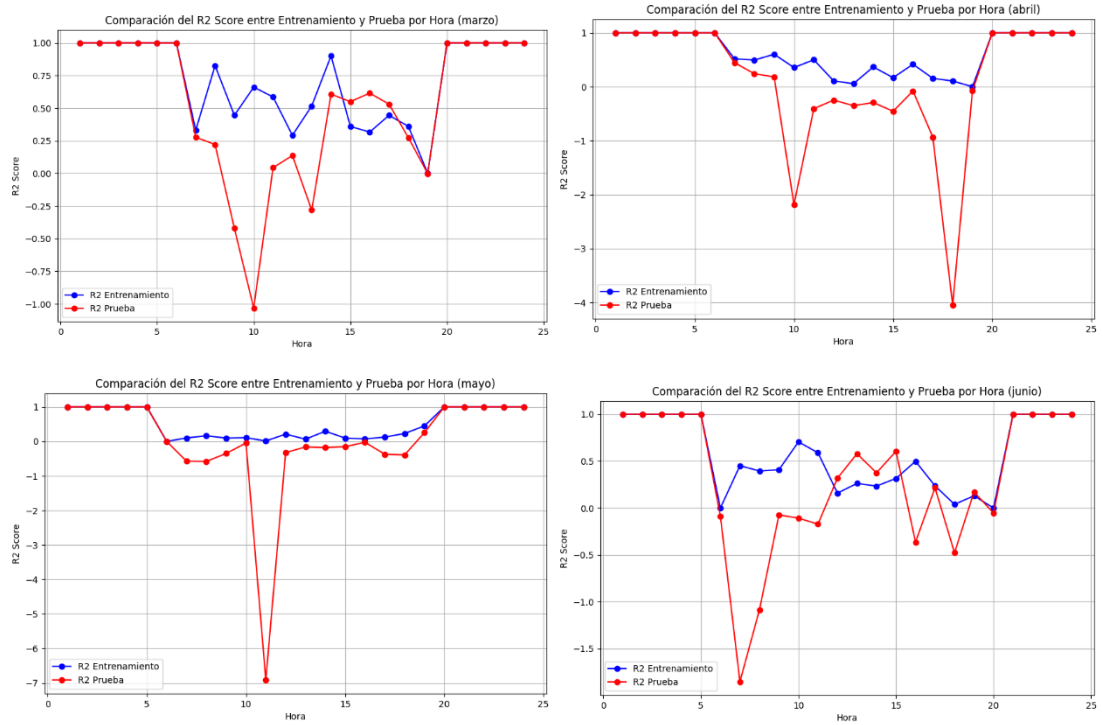
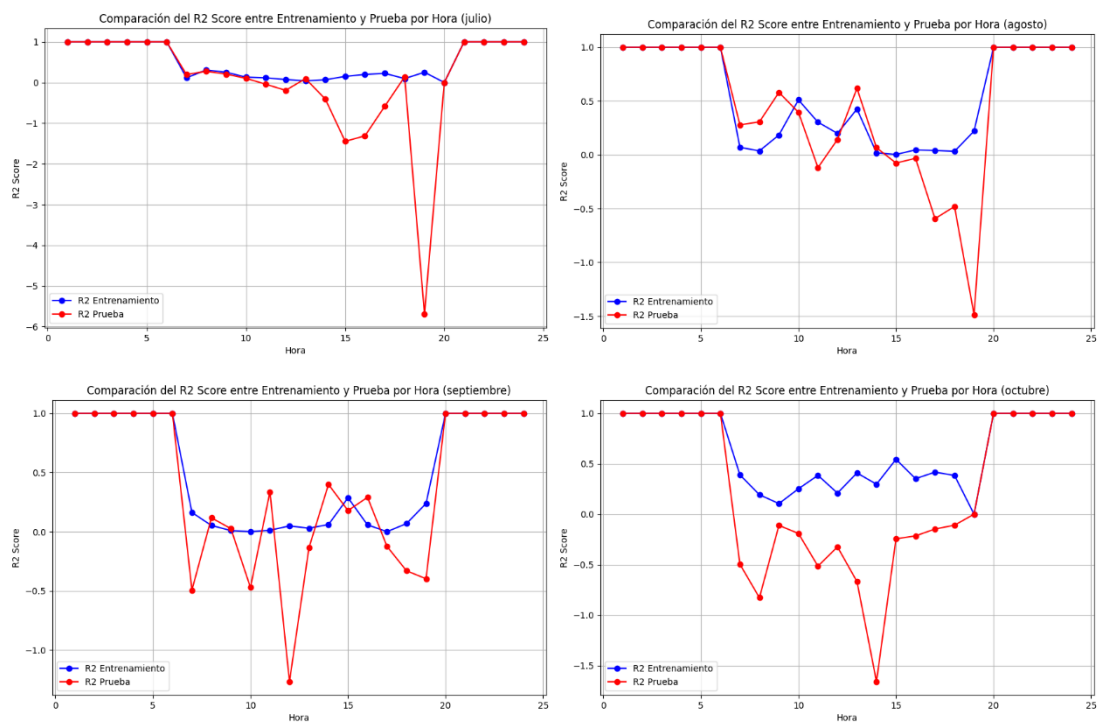


Figura 53 Gráfico de R^2 para datos de entrenamiento y prueba del conjunto horas-mensual.



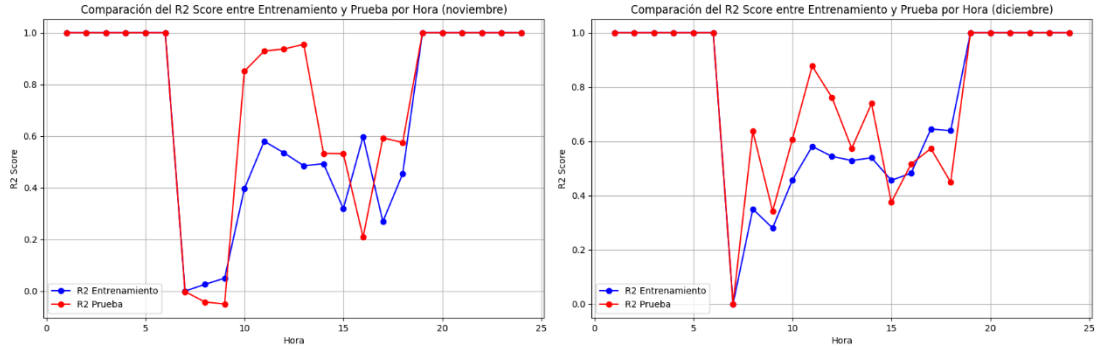
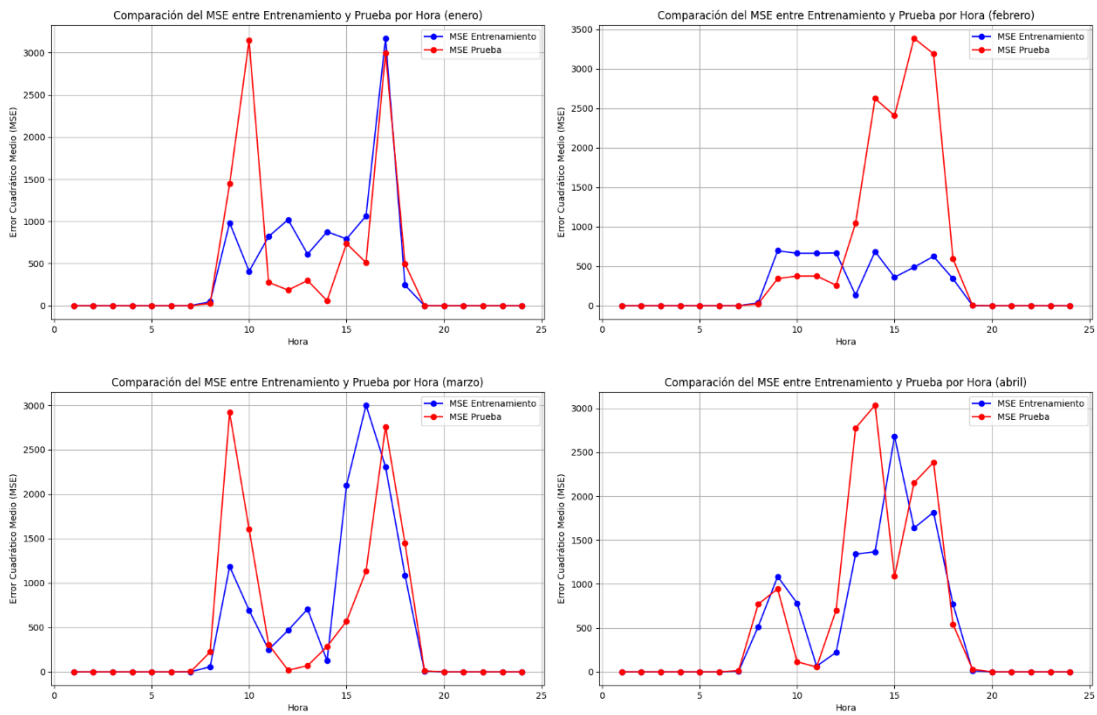


Figura 54 Gráfico de R^2 para datos de entrenamiento y prueba del conjunto horas-mensual (continuación).

La Figura 55 y la Figura 56 muestran la comparativa entre los valores de MSE de entrenamiento y de prueba para el conjunto de datos horas-mensual.



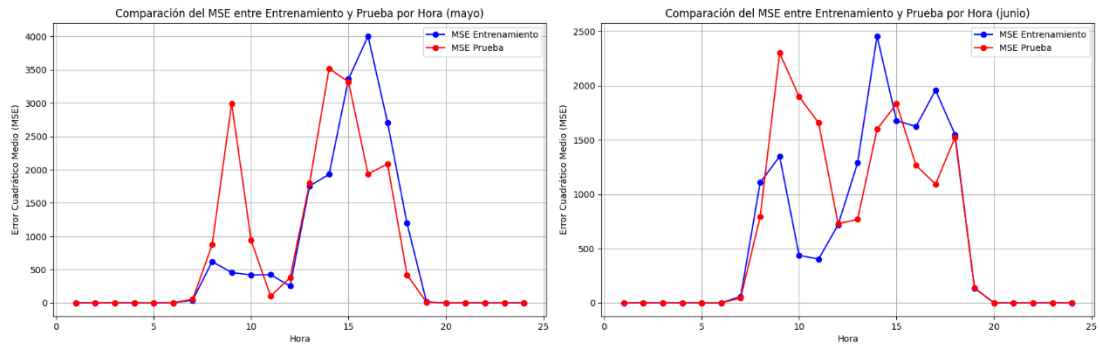


Figura 55 Gráfico de MSE para datos de entrenamiento y prueba del conjunto horas-anual.

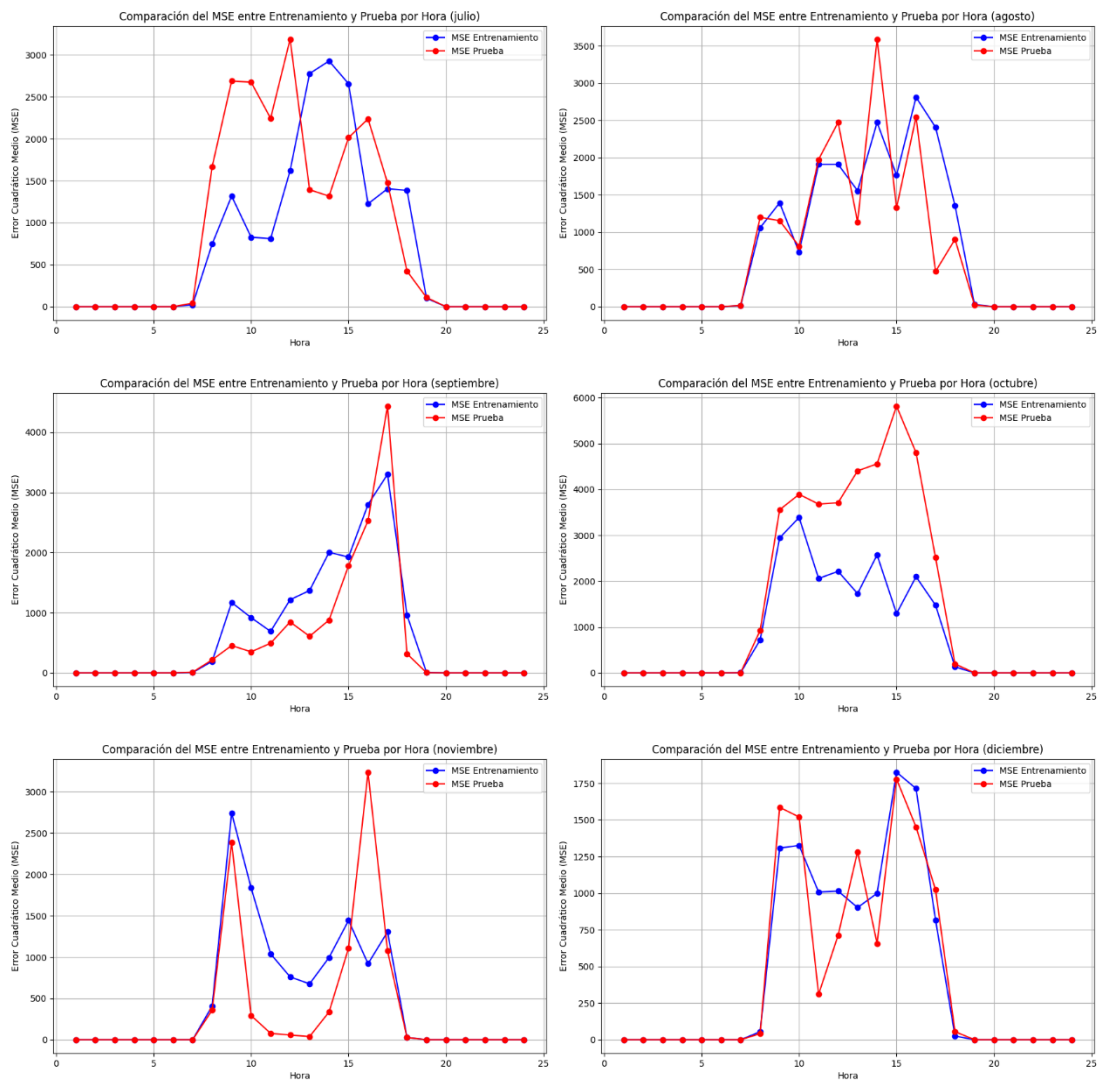
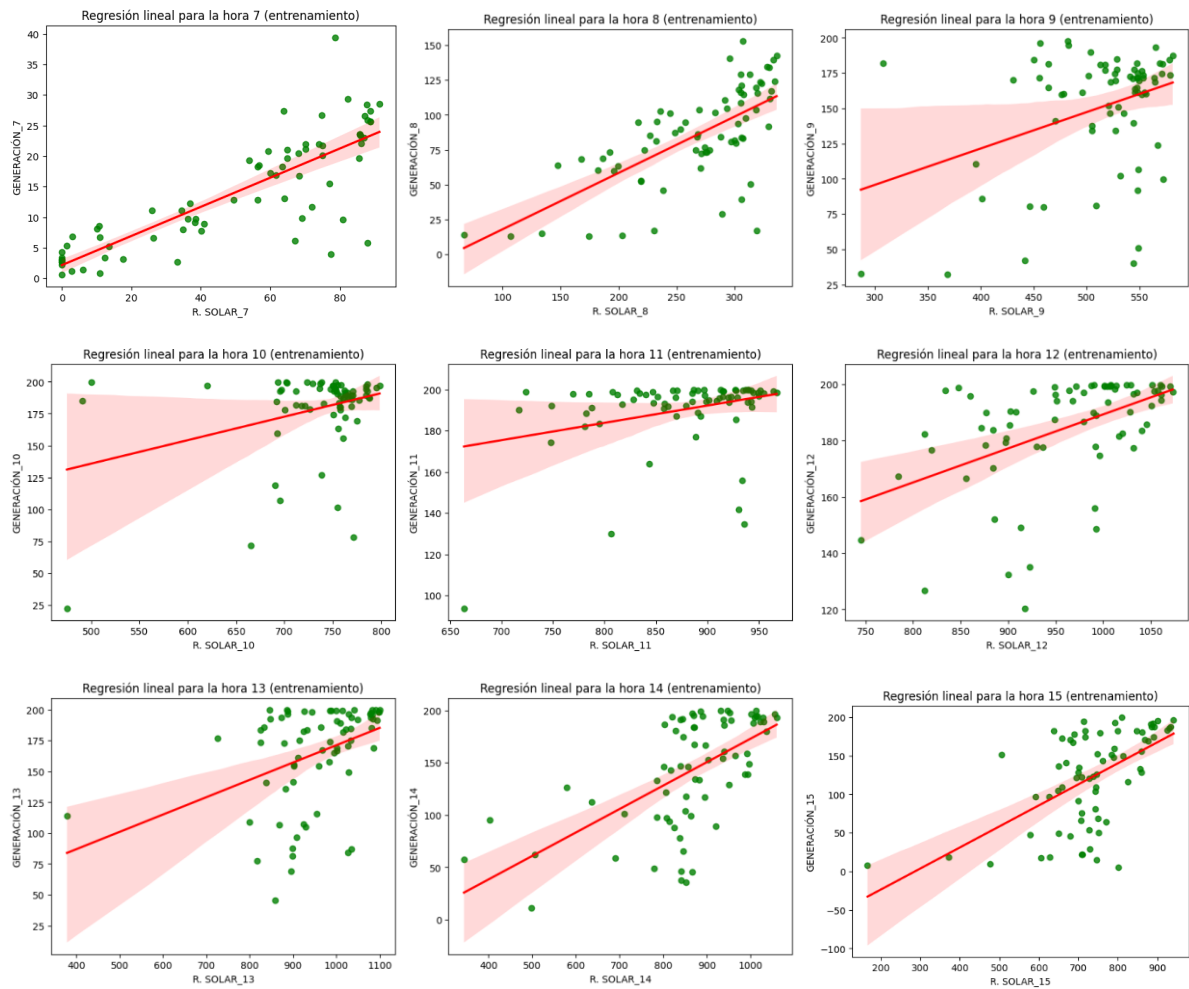


Figura 56 Gráfico de MSE para datos de entrenamiento y prueba del conjunto horas-anual (continuación)

4.11 Regresión lineal para el conjunto de datos hora-estación

Los siguientes resultados corresponden al conjunto de datos agrupados por cada estación del año, en este apartado se obtienen 4 conjuntos de 24 ecuaciones (una ecuación para cada hora), así como el cálculo de la R^2 y MSE.

La Figura 57 muestra el comportamiento de la regresión lineal de la estación primavera.



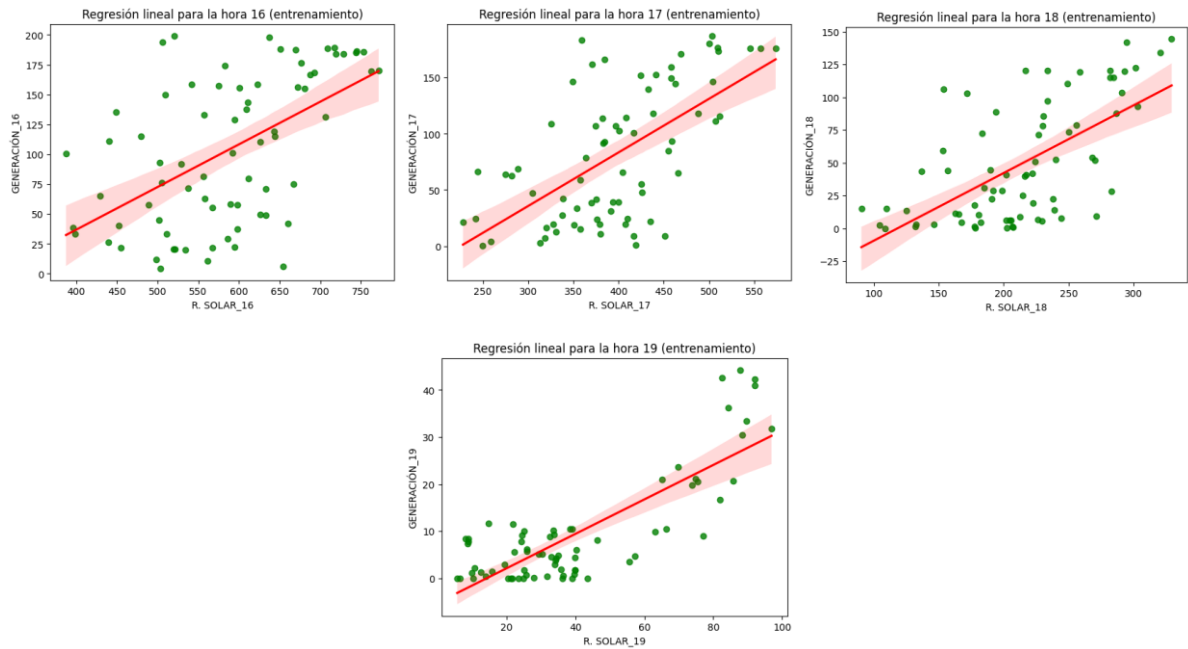
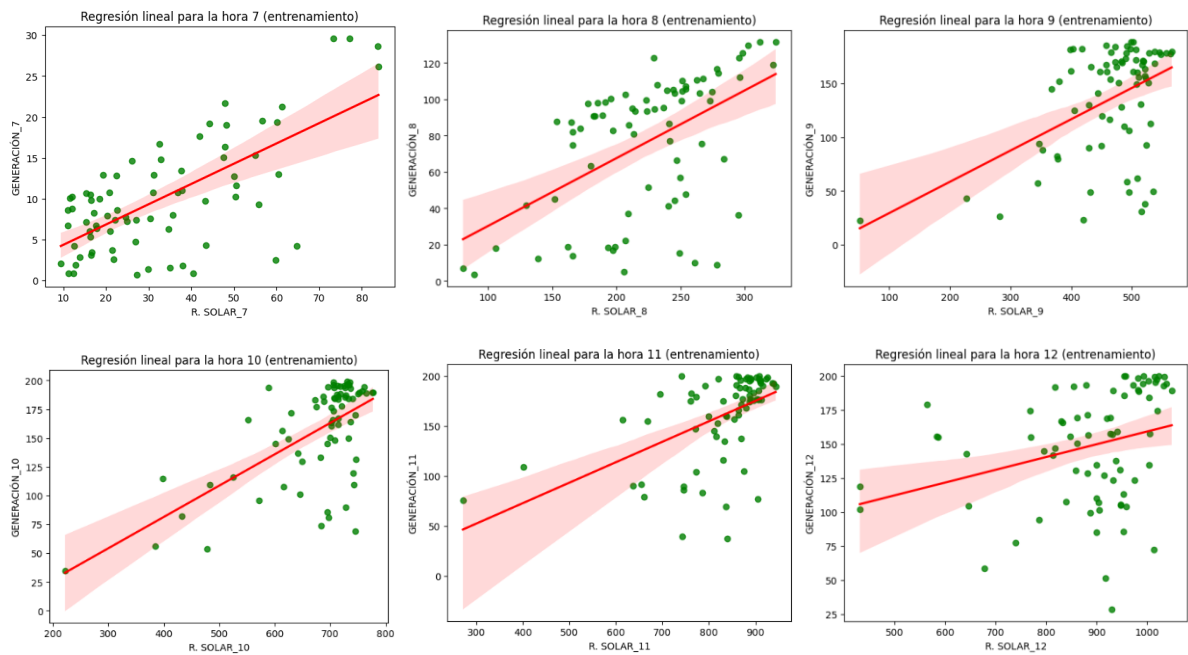


Figura 57 Resultado de la regresión de la hora 7 a la 19 para la estación primavera.

La Figura 58 muestra el comportamiento de la regresión lineal de la estación verano.



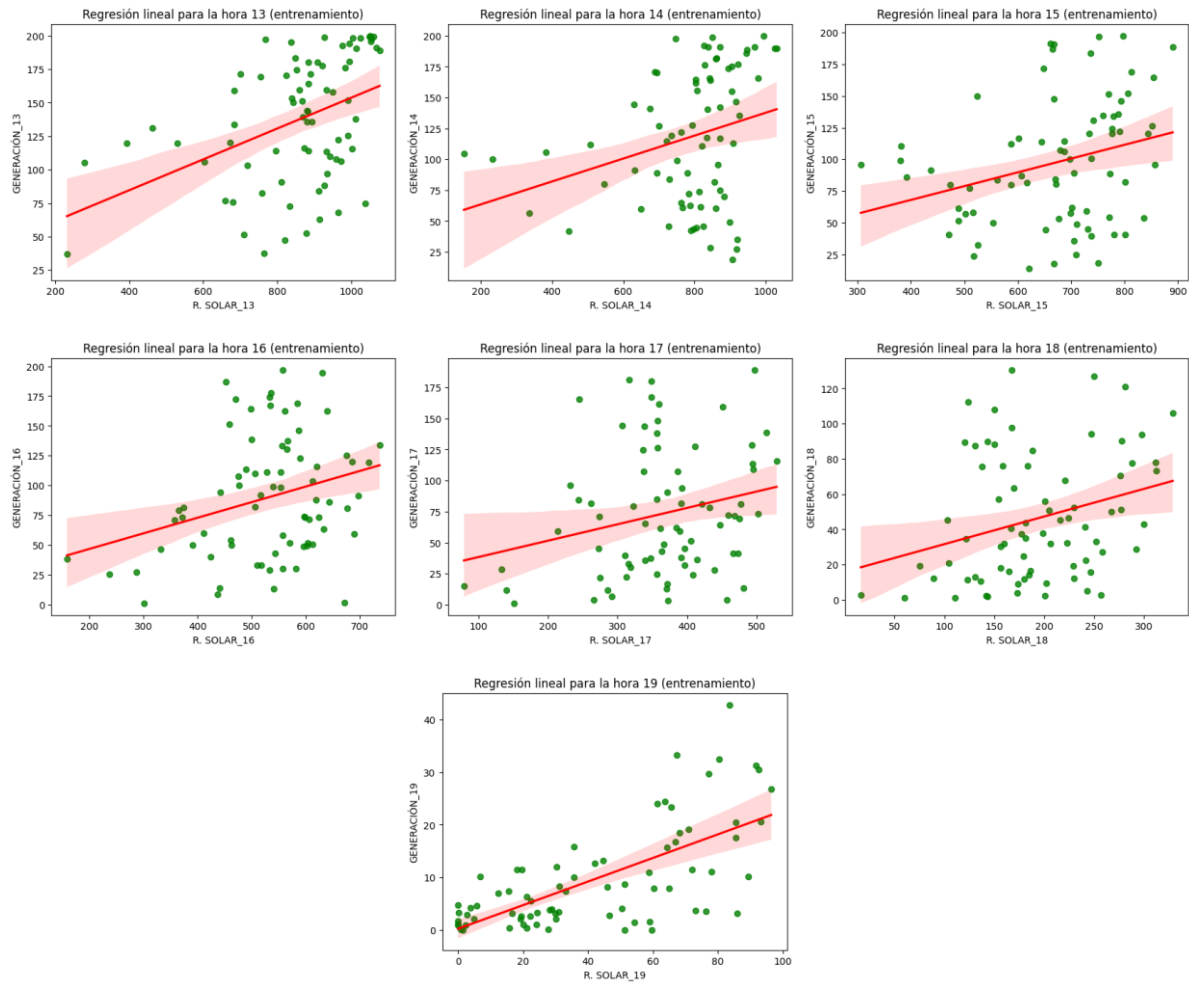


Figura 58 Resultado de la regresión de la hora 7 a la 19 para la estación verano.

La Figura 59 muestra el comportamiento de la regresión lineal de la estación otoño.



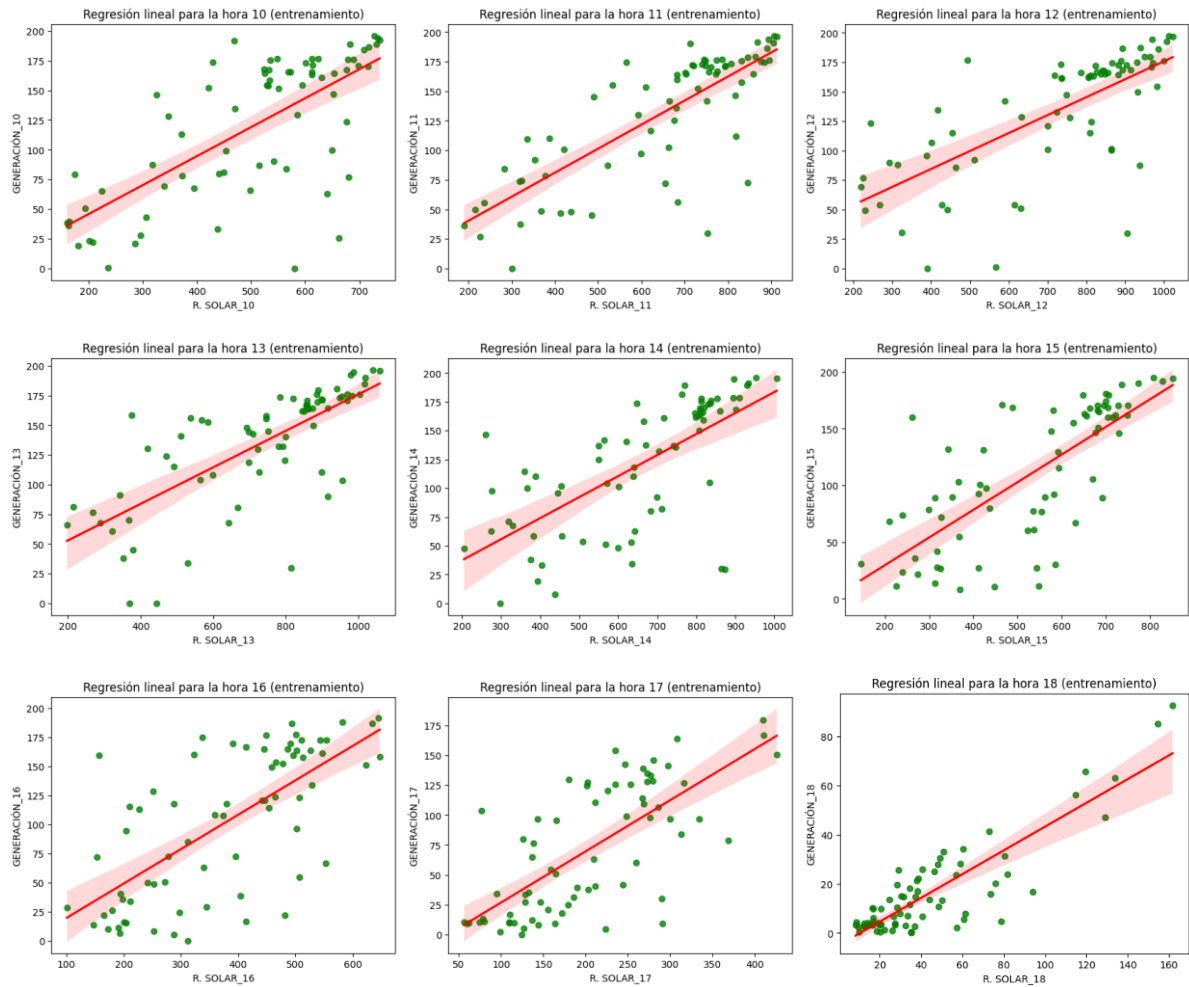
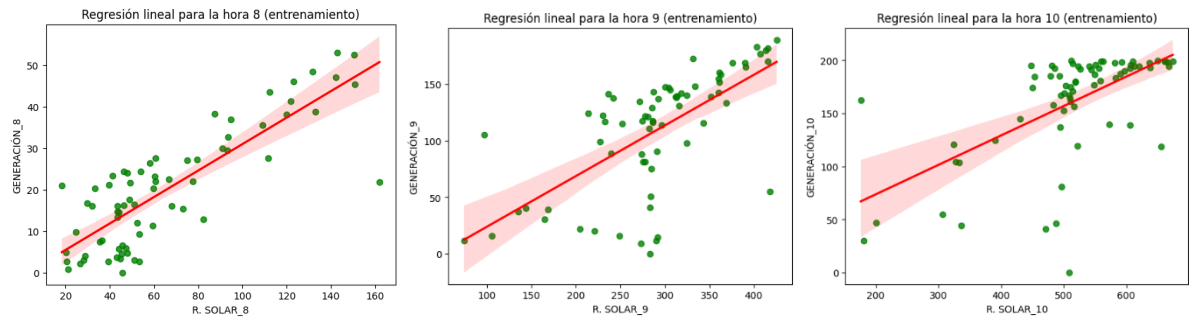


Figura 59 Resultado de la regresión de la hora 7 a la 18 para la estación otoño.

La Figura 60 muestra el comportamiento de la regresión lineal de la estación invierno.



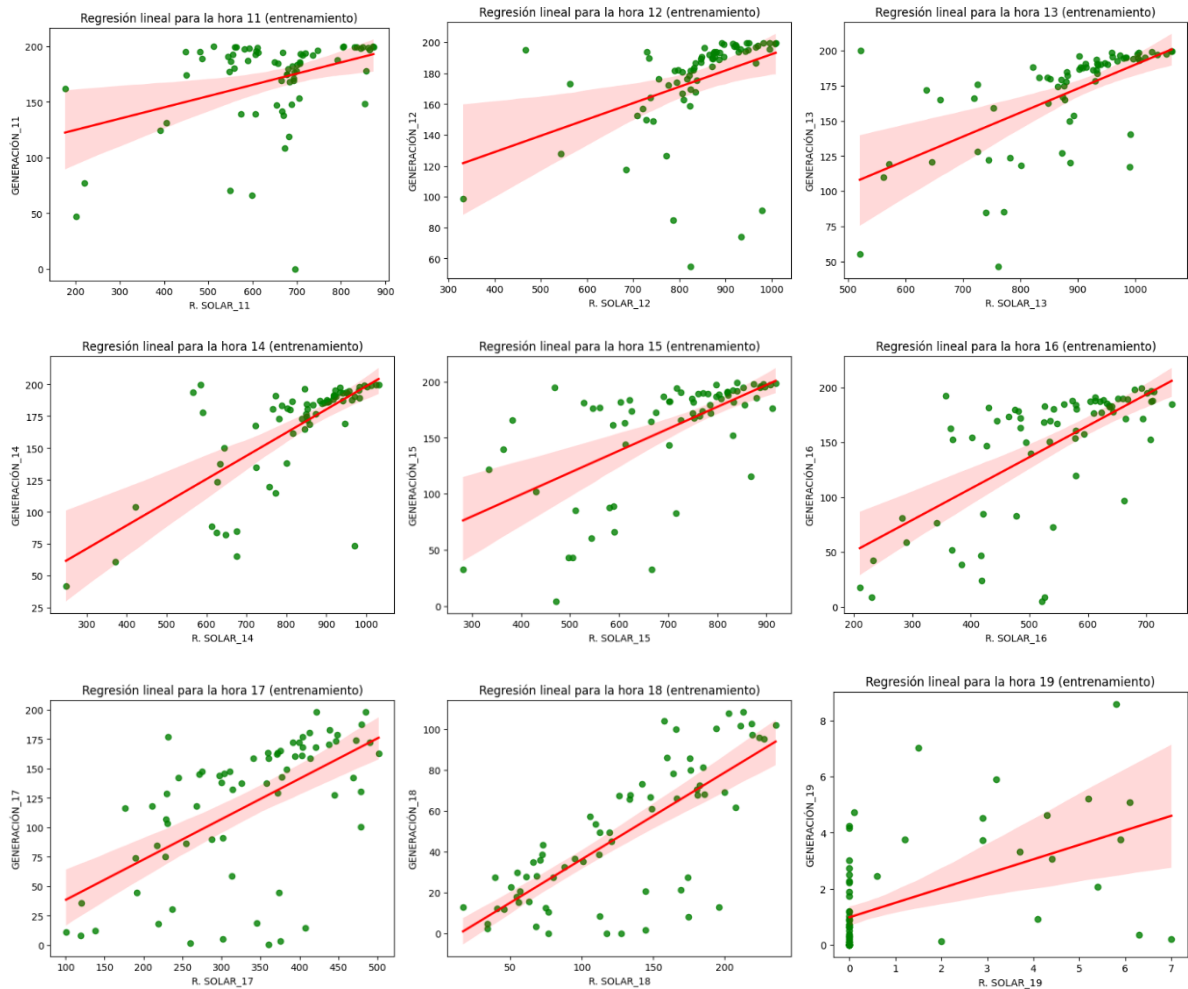


Figura 60 Resultado de la regresión de la hora 8 a la 19 para la estación invierno.

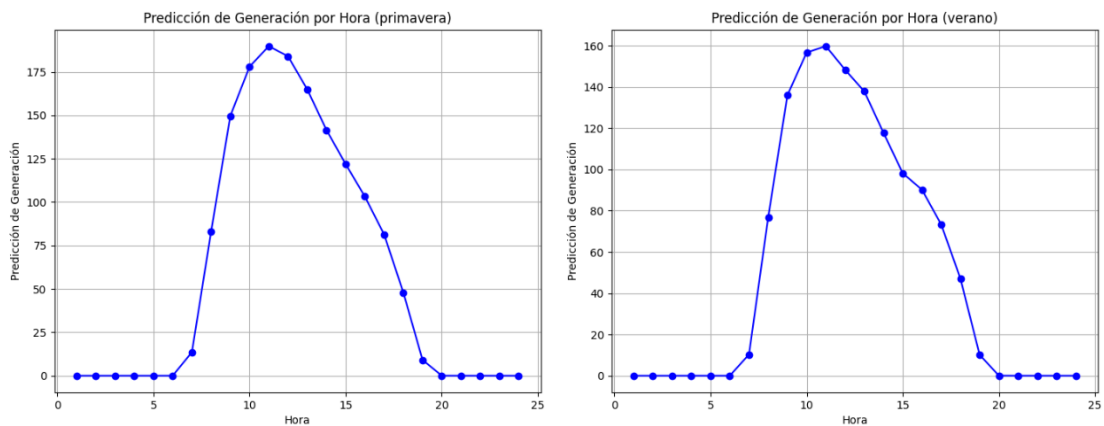
4.12 Resultados de la regresión lineal grupo horas-estación

Ahora bien, se procede a analizar el resultado de las ecuaciones de la regresión lineal. La Tabla 13 muestra la tabla de resultados con las ecuaciones para cada una de las horas de la clasificación de datos horas-anual, la cual permite efectuar el pronóstico a corto plazo con datos de irradiancia.

Tabla 13. Resultados con las ecuaciones para cada hora en el grupo de datos hora-estación.

Hora	Ecuación			
	primavera	verano	otoño	invierno
1	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
2	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
3	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
4	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
5	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
6	$y = 0.0x + 0.009$	$y = 0.0x + 0.001$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
7	$y = 0.238x + 2.158$	$y = 0.248x + 1.881$	$y = 0.616x + 0.762$	$y = 0.0x + 0.073$
8	$y = 0.405x + -22.583$	$y = 0.373x + -7.043$	$y = 0.29x + 2.999$	$y = 0.319x + -0.883$
9	$y = 0.258x + 18.317$	$y = 0.291x + 0.389$	$y = 0.249x + 15.177$	$y = 0.448x + -20.748$
10	$y = 0.184x + 43.896$	$y = 0.272x + -27.284$	$y = 0.244x + -2.885$	$y = 0.277x + 18.258$
11	$y = 0.084x + 116.654$	$y = 0.204x + -8.973$	$y = 0.203x + -0.247$	$y = 0.101x + 104.812$
12	$y = 0.121x + 68.417$	$y = 0.094x + 65.226$	$y = 0.153x + 23.404$	$y = 0.106x + 86.752$
13	$y = 0.14x + 30.768$	$y = 0.115x + 38.591$	$y = 0.154x + 22.212$	$y = 0.17x + 19.509$
14	$y = 0.224x + -51.084$	$y = 0.093x + 44.881$	$y = 0.183x + 0.944$	$y = 0.182x + 16.758$
15	$y = 0.273x + -78.335$	$y = 0.109x + 24.543$	$y = 0.244x + -19.203$	$y = 0.195x + 21.316$
16	$y = 0.357x + -106.051$	$y = 0.131x + 20.637$	$y = 0.296x + -9.887$	$y = 0.286x + -6.524$
17	$y = 0.477x + -107.463$	$y = 0.132x + 25.183$	$y = 0.429x + -16.22$	$y = 0.343x + 4.106$
18	$y = 0.517x + -61.06$	$y = 0.157x + 15.924$	$y = 0.483x + -5.035$	$y = 0.425x + -6.166$
19	$y = 0.366x + -5.185$	$y = 0.225x + 0.209$	$y = 0.0x + 0.113$	$y = 0.517x + 0.981$
20	$y = 0.0x + 0.022$	$y = 0.0x + 0.036$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
21	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
22	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
23	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$
24	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$	$y = 0.0x + 0.0$

La Figura 61 muestra el gráfico de predicción, el cual se obtiene a través de un vector con 24 datos de irradiancias y las ecuaciones de la figura 53.



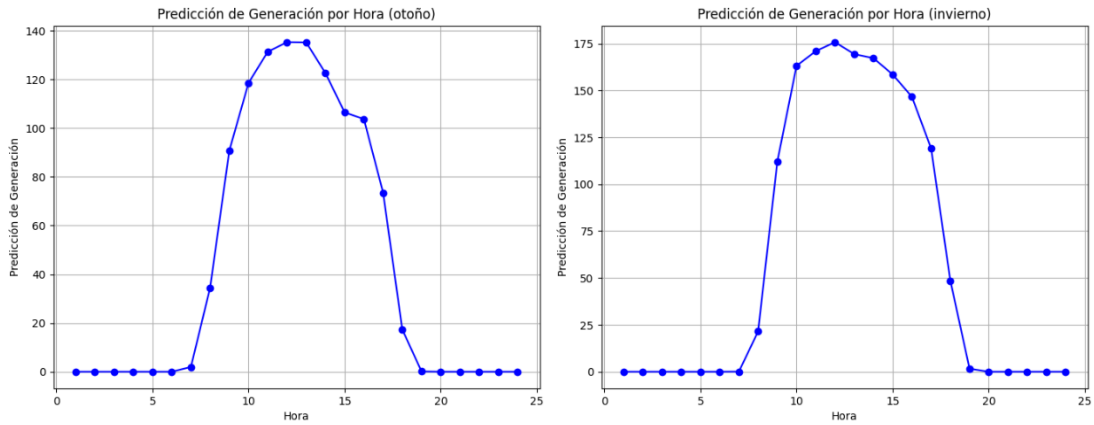


Figura 61 Gráfico de pronóstico para el grupo de datos horas-mensual con datos de prueba.

La Figura 62 muestra la comparativa entre los valores de R^2 de entrenamiento y de prueba.

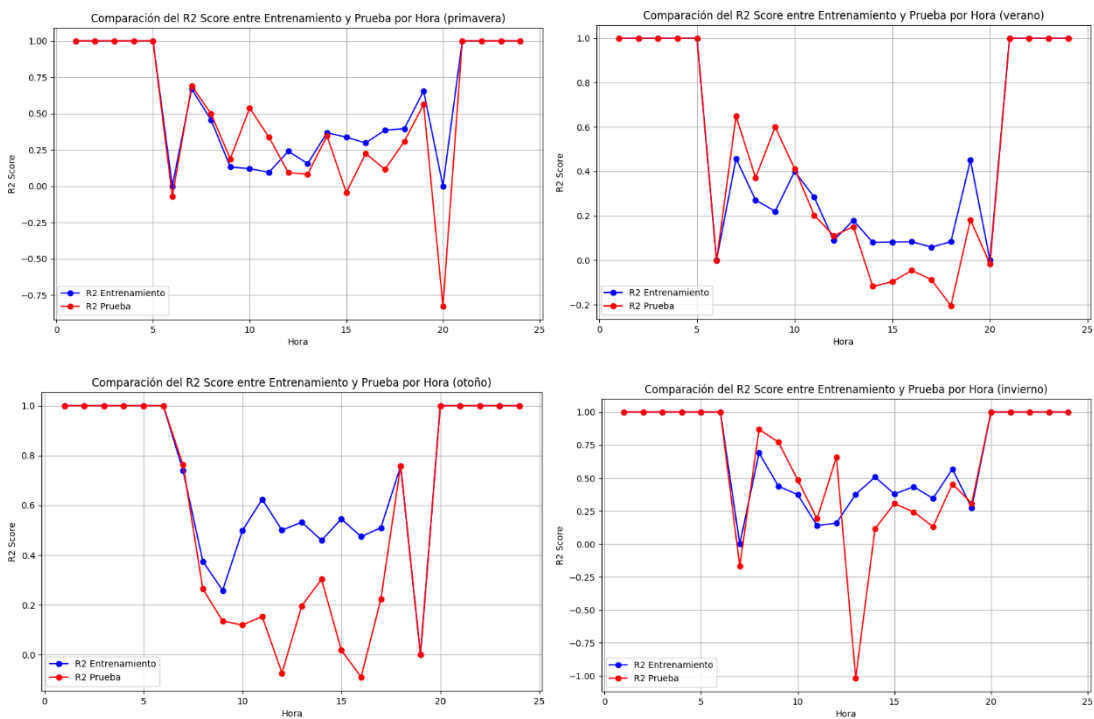


Figura 62 Gráfico de R^2 para datos de entrenamiento y prueba del conjunto horas-mensual.

La Figura 63 muestra la comparativa entre los valores de MSE de entrenamiento y de prueba.

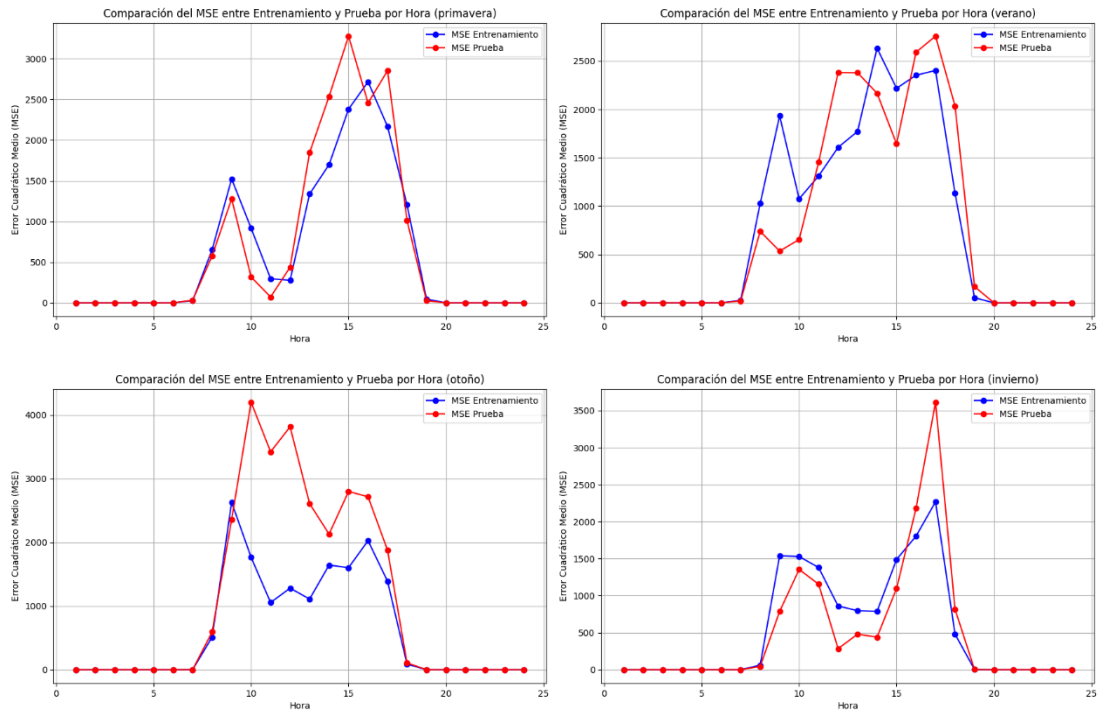


Figura 63 Gráfico de MSE para datos de entrenamiento y prueba del conjunto horas-anual.

4.13 Aplicación del modelo

Una vez que se obtienen los modelos de predicción para cada grupo y parque se realiza una comparativa con datos reales del año 2024, esta implementación se lleva a cabo en una hoja de cálculo y los resultados se muestran desde la Figura 64. Pronóstico de generación: 6 de febrero 2024 parque 1. Hasta la Figura 68.

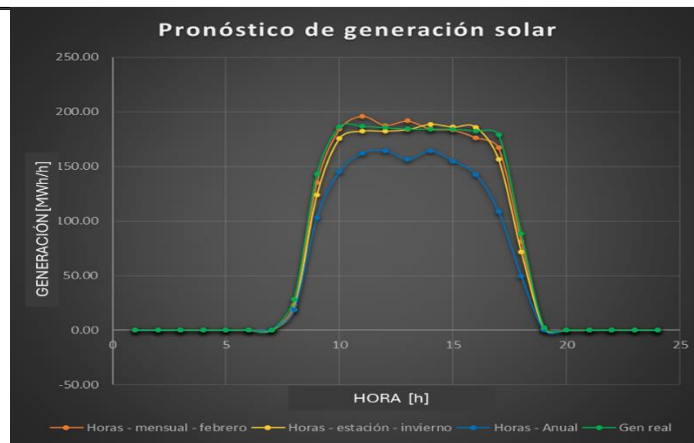


Figura 64. Pronóstico de generación: 6 de febrero 2024 parque 1.

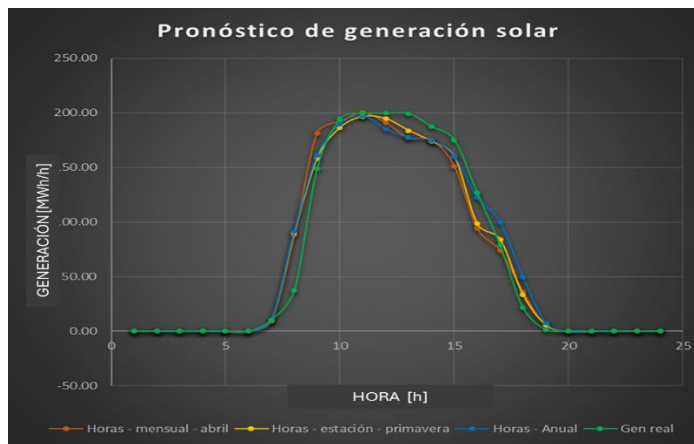


Figura 65 Pronóstico de generación: 17 de abril 2024 parque 1.

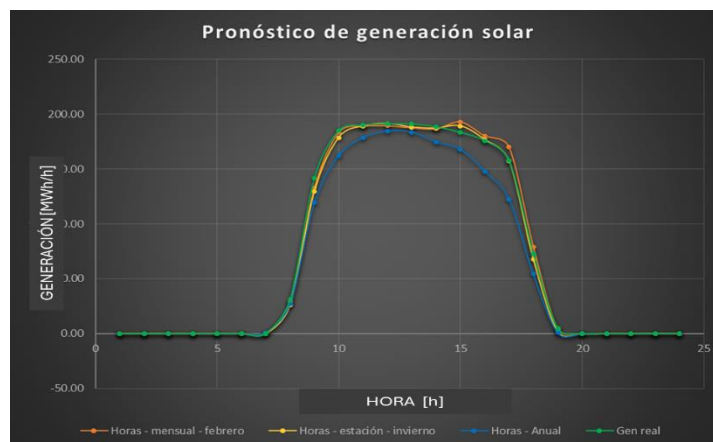


Figura 66 Pronóstico de generación: 24 de febrero 2024 parque 2.

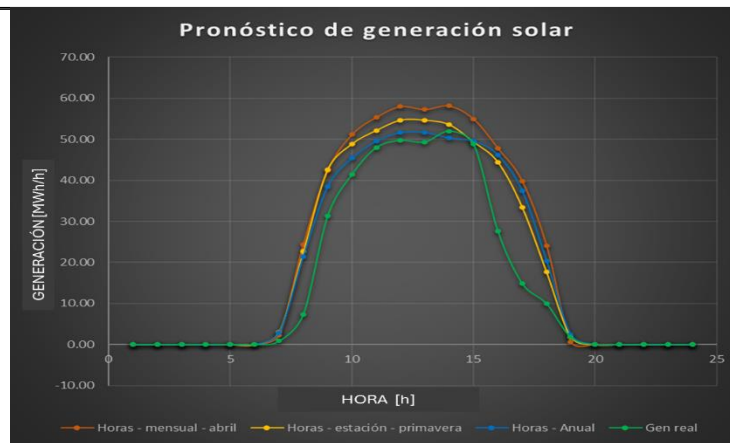


Figura 67 Pronóstico de generación: 26 de abril 2024 parque 4.

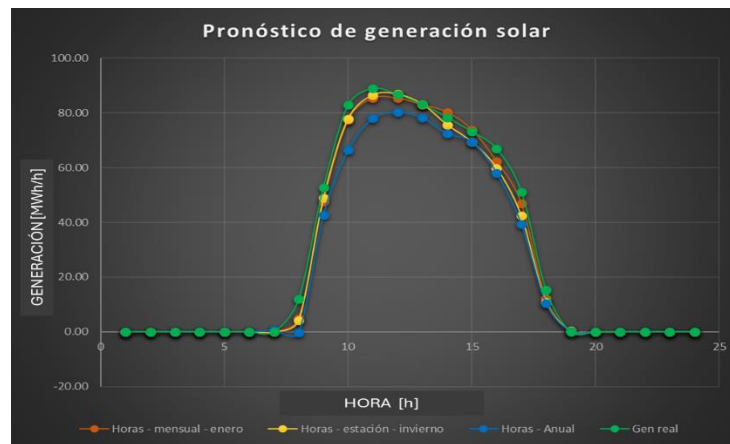


Figura 68 Pronóstico de generación: 6 de enero 2024 parque 5.

Capítulo 5 Conclusiones

Existen diferentes formas de realizar un pronóstico de generación solar fotovoltaica, en este trabajo se implementó la solución por el método estadístico de regresión lineal, una parte fundamental antes de la aplicación del método es la creación de grupos con los datos históricos del año 2023 de un parque solar ubicado en la zona oriente de México, estos grupos son anual, mensual y estacional.

Para el lograr el resultado de esta investigación es de suma importancia contar con el desarrollo de los software específicos para el análisis de datos y el uso del lenguaje de programación, en este caso Python.

Otro dato importante es resultado del análisis de correlación de variables, dónde se identificó plenamente que las variables Irradiancia solar y la temperatura ambiente presentan fuerte correlación positiva con la variable de generación eléctrica y con una correlación negativa para la variable de humedad relativa y la nubosidad.

El resultado del proyecto de investigación consta de tres modelos de predicción para cuatro parques que pertenecen a la GCROR los cuales abordaremos de la siguiente forma.

Para el grupo anual, se obtienen 24 ecuaciones, una para cada hora del día, teniendo como variable predictora a la irradiancia solar y como variable de respuesta a la generación solar, el modelo se entrena con el 80% de datos y el otro 20% para realizar la comparación a través del coeficiente de determinación y el MSE. Los resultados se muestran satisfactorios en la mayoría de las horas por lo que se concluye que es aceptable.

Para el grupo mensual, se obtienen 12 paquetes de 24 ecuaciones como el caso anual, cada paquete de ecuaciones corresponde a la predicción de la generación por cada mes

del año, al momento de realizar el análisis utilizando el coeficiente de determinación nos podemos percatar que sufre una mayor variación respecto al caso anual, los meses con mejor comportamiento son enero mayo y julio y utilizando el MSE el mejor comportamiento se espera en los meses de marzo, agosto y septiembre.

Para el grupo estacional, se obtienen 4 paquetes de 24 ecuaciones, una para cada estación del año, la estación que presenta mejor evaluación respecto al coeficiente de determinación y el MSE es primavera y posteriormente el verano.

Estos modelos de predicción corresponden a cuatro parques solares fotovoltaico en la zona oriente de México, es necesario realizar el análisis de datos para otros parques de otras gerencias de control y evaluar su cumplimiento.

Capítulo 6 Recomendaciones

El método estadístico de regresión lineal nos ofrece una predicción aceptable para la generación solar fotovoltaica, existen otros métodos estadísticos que pueden ofrecer mejores resultados tales como una regresión múltiple, exponencial o logarítmica y es su defecto utilizar redes neuronales con aprendizaje automático, este dependerá de la cantidad de datos y el objetivo de la aplicación.

Por otra parte, es posible mejorar el modelo actual realizando un mejor tratamiento de los datos, así como implementar los modelos en el pronóstico habitual de la gerencia de control.

Bibliografía

- [1] Secretaría de Energía (SENER) Programa de Desarrollo del Sistema Eléctrico Nacional 2024-2038. (2024). <https://www.gob.mx/sener/articulos/programa-de-desarrollo-del-sistema-electrico-nacional-2024-2038>
- [2] Harper. E. (2018). El ABC de las Energías Renovables en los sistemas eléctricos. Limusa Características Y Desventajas.
- [3] González Longatt F.M (2008). Introducción a los Sistemas de Potencia. Recuperado el día 16/09/2023 de https://fqlongatt.org/OLD/Archivos/Archivos/SP_I/PPT-IntroSP.pdf
- [4] Centro Nacional de Control de Energía (CENACE). (2023). Anexo 1 Infraestructura del Sistema Eléctrico Nacional.
- [5] Sonia Montecinos Geisse. (mayo 2021). Pronósticos de generación eólica y solar fotovoltaica en América Latina. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH Red sectorial GADeR-ALC. https://energypedia.info/images/a/af/Pronosticos_eolico_FV_LATM.pdf
- [6] Castro, A. 2023. Regresión Lineal supuestos. https://rpubs.com/acastro/regresion_lineal_supuestos
- [7] Pablo Vinuesa. (2016). Correlación: teoría y práctica. <http://www.ccg.unam.mx/~vinuesa/>
- [8] Montgomery y otros (2006). Introducción al análisis de regresión lineal. CECSA.
- [9] Alteryx. Análisis de datos. <https://www.alteryx.com/es/glossary/data-analytics>
- [10] Python Software Foundation. (2025). El tutorial de Python. <https://docs.python.org/es/3.13/tutorial/index.html>
- [11] Colaboratory. ¿Qué es Colaboratory? <https://colab.research.google.com/?hl=es#scrollTo=ISrWNR3MuFUS>

[12] Sánchez A. La librería Pandas. <https://aprendeconalf.es/docencia/python/manual/pandas/> pandas

[13] Datascientest. Matplotlib: todo lo que tienes que saber sobre la librería Python de Dataviz. <https://datascientest.com/es/todo-sobre-matplotlib> matplotlib

[14] Datascientest. Seaborn: todo sobre la herramienta de Data Visualization Python. <https://datascientest.com/es/seaborn-la-herramienta-de-data-visualization-python> seaborn

[15] Nhlanhla M, & Bongumsa M. (2024). A survey of applications of simple and multiple linear regression in solar power generation. International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE).

[16] Rubanenko O, & Yanovych V. (2020). Analysis of instability generation of Photovoltaic power station. IEEE.

[17] Danielsson G, et al. (2024). *Analysis of Meteorological Variables Used for Solar Power Generation Forecast*. Workshop on Communication Networks and Power Systems (WCNPS).

[18] Mathur A, et al. (2023). Comparative Study of Regression Models for PV Power Prediction. XIX International Scientific Technical Conference Alternating Current Electric Drives (ACED).