

Predicción de producción de energía fotovoltaica a corto plazo

by

JUAN CARLOS SOSA GARCÍA

A thesis submitted in conformity with the requirements
for the MSc in Economics, Finance and Computer Science

University of Huelva & International University of Andalusia

uhu.es

un
i Universidad
Internacional
de Andalucía
A

November 2017

Predicción de producción de energía fotovoltaica a corto plazo

Juan Carlos Sosa García

Máster en Economía, Finanzas y Computación

Dr. Juan Manuel Bravo Caro
Universidad de Huelva y Universidad Internacional de Andalucía

2017

Abstract

« Currently, with the increase of the pollution due to the excessive use of fossil fuels and the increase of environmental awareness, renewables energies growth has been favored. Among these energies, photovoltaic energy stands out because of its natural wealth in resources, safety and reliability. In this context, making accurate predictions of the energy that will be produced becomes a key factor. In this regard, from an annual data set of predicted solar radiation and real production, six different prediction models were created, in order to reach the best possible estimation. Among the prediction methods used, local linear regression is found, which is the one that offers the best results, reaching R^2 coefficients above 90% for every month. »

JEL classification: C16, C14, C51, Q42.

Key words: renewables energies, solar energy production forecasting, photovoltaic energy, solar radiation, non-parametric predictive models.

Predicción de producción de energía fotovoltaica a corto plazo

Juan Carlos Sosa García

Máster en Economía, Finanzas y Computación

Dr. Juan Manuel Bravo Caro
Universidad de Huelva y Universidad Internacional de Andalucía

2017

Resumen

« En la actualidad, el aumento de la contaminación debido al uso excesivo de combustibles fósiles y la mayor concienciación acerca del cuidado del medioambiente, han favorecido el auge de las llamadas energías renovables. Entre estas energías destaca la energía fotovoltaica, por su riqueza en recursos, seguridad y fiabilidad. En este contexto, realizar predicciones exactas de la energía que se producirá, se convierte en un factor de vital importancia. En este sentido, a partir de un conjunto de datos anual de radiación solar prevista y producción real, se crearon seis distintos modelos de predicción, con el fin de alcanzar la mejor estimación posible. Entre los métodos de predicción utilizados, se encuentra la regresión lineal local, siendo éste el que mejores resultados ofrece, alcanzando coeficientes R^2 superiores al 90% para todos los meses. »

Clasificación JEL: C16, C14, C51, Q42.

Palabras clave: energías renovables, predicción de producción solar, energía fotovoltaica, radiación solar, modelos predictivos no paramétricos.

Tabla de contenidos

1	Introducción	1
1.1	Crisis energética y contaminación	1
1.2	Energías renovables	3
2	Objetivos	4
3	Producción de energía fotovoltaica	5
4	Metodología	8
4.1	Datos	8
4.2	Variables	9
4.3	Modelos de predicción.....	9
4.3.1	Método basado en la persistente	10
4.3.2	Regresión lineal.....	10
4.3.3	Regresión polinómica	11
4.3.4	Regresión lineal local.....	11
4.3.5	Regresión basada en los k-vecinos más cercanos (KNN).....	12
4.3.6	Regresión basada en función Kernel.....	13
4.4	Índices de errores	14
4.4.1	Suma de los errores al cuadrado	14
4.4.2	Error estándar de la regresión	15
4.4.3	Suma total de los cuadrados.....	15
4.4.4	Coeficiente de determinación R^2	15
4.4.5	Error cuadrático medio.....	16
4.4.6	Error absoluto medio.....	16
5	Resultados	17
5.1	Método basado en la persistente	17

5.2 Regresión lineal.....	19
5.3 Regresión polinómica	21
5.4 Regresión lineal local.....	23
5.5 Regresión basada en los k-vecinos más cercanos (KNN).....	26
5.6 Regresión basada en función Kernel.....	27
5.7 Resultados globales.....	29
5.7.1 Residual Sum of Square (RSS)	29
5.7.2 Residual Standard Error (RSE).....	30
5.7.3 R square (R^2).....	31
5.7.4 Mean Square Error (MSE)	32
5.7.5 Mean Absolute Error (MAE)	33
6 Conclusiones	34
Referencias bibliográficas.....	36
Anexo I. Índices de errores mensuales.....	40
Anexo II. Evolución de la producción por el método de la persistente	42
Anexo III. Producción real vs Producción estimada. Método de la persistente	45
Anexo IV. Evolución de la producción por el método de la regresión lineal.....	47
Anexo V. Producción real vs Producción estimada. Método de la regresión lineal	50
Anexo VI. Evolución de la producción por el método de la regresión polinómica	52
Anexo VII. Producción real vs Producción estimada. Método de la regresión polinómica	55
Anexo VIII. Evolución de la producción por el método de la regresión lineal local	57
Anexo IX. Producción real vs Producción estimada. Método de la regresión lineal local.....	60
Anexo X. Evolución de la producción por el método de los k-vecinos más cercanos	62
Anexo XI. Producción real vs Producción estimada. Método de los k-vecinos más cercanos.....	65
Anexo XII. Evolución de la producción por el método basado en función Kernel	67
Anexo XIII. Producción real vs Producción estimada. Método basado en función Kernel	70

Lista de Tablas

Tabla 1. Índices de errores según método basado en la persistente	17
Tabla 2. Índices de errores según regresión lineal	19
Tabla 3. Índices de errores según regresión polinómica	21
Tabla 4. Índices de errores según regresión lineal local	23
Tabla 5. Índices de errores según regresión KNN	26
Tabla 6. Índices de errores según regresión basada en función Kernel	27
Tabla 7. Resumen de los índices de errores según los diferentes modelos empleados	29

Lista de Figuras

Figura 1. Emisiones de CO ₂ en los diez países más contaminantes. Fuente: IAE.	2
Figura 2. Emisiones de CO ₂ a nivel mundial, derivadas del uso de combustibles fósiles por sector en 2014. Fuente: IAE.	3
Figura 3. Crecimiento acumulativo global de la capacidad fotovoltaica. Fuente: IEA.	6
Figura 4. Producción fotovoltaica por países. Fuente: SPV Market Research	6
Figura 5. Producción solar por horas prevista vs real en Alemania, 2013. Fuente: Burger.	7
Figura 6. Producción estimada vs producción real. Método de la persistente (marzo y julio). Fuente: Elaboración propia	18
Figura 7. Evolución de la producción por el método de la persistente (marzo y julio). Fuente: Elaboración propia	18
Figura 8. Evolución del R ² (Persistente y Lineal). Fuente: Elaboración propia	20
Figura 9. Evolución de la producción real y prevista en marzo. Métodos de la persistente y lineal. Fuente: Elaboración propia	21
Figura 10. Evolución del RSS. Regresiones lineal y polinómica. Fuente: Elaboración propia	22
Figura 11. Evolución del R ² . Regresión lineal local. Fuente: Elaboración propia	23
Figura 12. Evolución del RSE. Regresiones lineal y lineal local. Fuente: Elaboración propia	24
Figura 13. Y real vs Y estimada. Regresiones lineal y lineal local. Fuente: Elaboración propia	25
Figura 14. Evolución de la producción por el método de la regresión lineal local (julio). Fuente: Elaboración propia.	25
Figura 15. Evolución del R ² . Regresión lineal local y KNN. Fuente: Elaboración propia.	26
Figura 16. Evolución del RSS. Regresión lineal local y Kernel. Fuente: Elaboración propia.	28

Figura 17. Evolución de la producción (marzo). Regresión basada en función Kernel y regresión lineal local. Fuente: Elaboración propia.	28
Figura 18. Evolución de la suma de los errores al cuadrado por método. Fuente: Elaboración propia.	30
Figura 19. Evolución del error estándar de la regresión por método. Fuente: Elaboración propia.	31
Figura 20. Evolución del R^2 por método. Fuente: Elaboración propia.	32
Figura 21. Evolución del error cuadrático medio por método. Fuente: Elaboración propia.	33
Figura 22. Evolución del error absoluto medio por método. Fuente: Elaboración propia.	34

1 Introducción

En los últimos años, problemas políticos y sociales importantes relacionados con la localización de las fuentes energéticas tradicionales, han despertado un interés cada vez más extendido por encontrar nuevas fuentes de energía, caracterizados por reponerse a un ritmo igual o superior al que son consumidas y a las que se denomina energías renovables (Velasco, 2009).

Por otro lado, a raíz de la floreciente crisis energética y el calentamiento global, el estudio y la promoción de las fuentes de energía renovables se ha convertido en una necesidad candente, en el panorama internacional. Por lo tanto, optimizar el rendimiento del sistema de energías renovables es una obligación para que sea económicamente viable en comparación con otras fuentes de energía convencionales (Jha, 2016).

1.1 Crisis energética y contaminación

El uso de combustibles fósiles tiene un papel dominante en las emisiones de gases de efecto invernadero en la atmósfera (Wu & Xu, 2017). Se estima que el sector energético representa aproximadamente dos tercios de todas las emisiones antropogénicas de gases de efecto invernadero (IEA, 2015). Aproximadamente el 80% del uso total de energía en el mundo se basa en combustibles fósiles, que se utilizan para el transporte y los sectores de energía estacionaria, incluida la generación de energía eléctrica (Franzitta et al., 2013).

El incremento de consumo de energía a nivel mundial ha llevado al mundo a dos crisis agudas (Memar et al., 2017). La primera es la contaminación ambiental debido al consumo de combustibles fósiles, y la segunda es la aceleración de crecimiento drástico en la explotación de estos recursos energéticos (Hosseini et al., 2017; Shahinzadeh et al., 2016).

La Agencia Internacional de Energía (IAE) ha publicado una lista de los países más contaminantes, en relación con la emisión total de dióxido de carbono (CO₂) y el total de combustibles fósiles consumidos (IAE, 2015). Los diez países más contaminantes de acuerdo a la IAE son: China, Estados Unidos, India, Rusia, Japón, Alemania, Corea, Canadá, Irán y Arabia Saudí. La evolución de las emisiones de CO₂ de dichos países se encuentra recogida en la Figura 1, que mostramos a continuación:

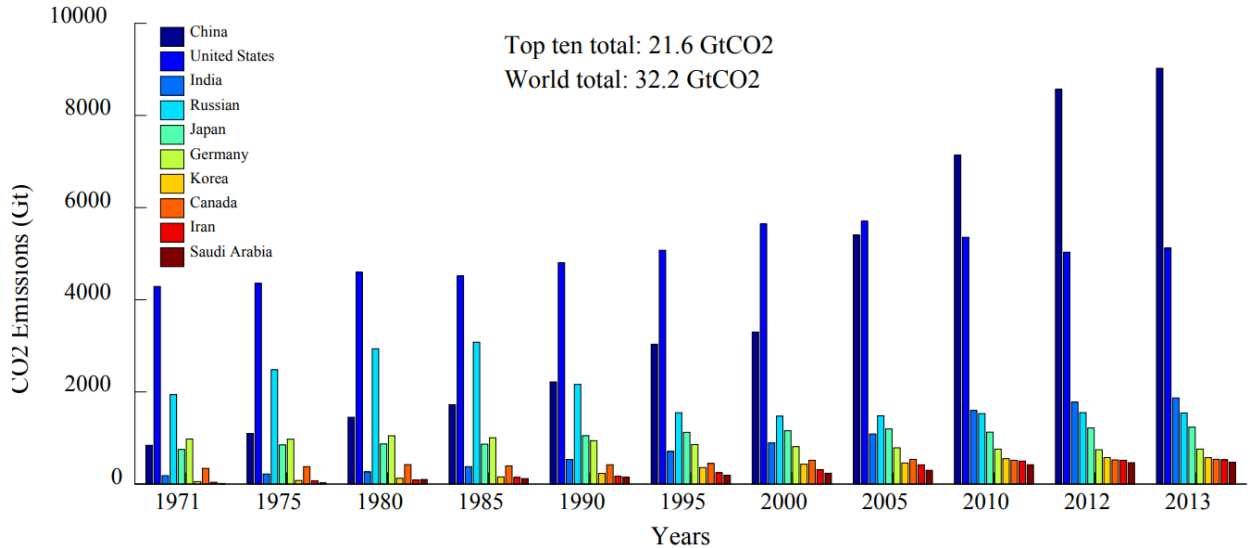


Figura 1. Emisiones de CO₂ en los diez países más contaminantes. Fuente: IAE.

Muchos de los países mencionados en la Figura 1 tienen industrias pesadas, alta capacidad de generación de electricidad y una gran población. Por lo tanto, para preservar el medioambiente y luchar ante las consecuencias devastadoras de la contaminación, es necesario que estos países adopten estrategias adecuadas para la reducción del consumo de energía, la utilización de energías alternativas para los recursos actuales, la mejora de la flota de transporte público y la reducción de las emisiones de CO₂. (Memar et al., 2017).

Dos sectores produjeron en 2014 dos tercios de las emisiones globales de CO₂ derivados del uso de combustibles fósiles (IAE, 2016). La electricidad y generación de calor fue la mayor con diferencia con un 42% del total de emisiones, seguido por el transporte con un 23%, tal y como se puede apreciar en la Figura 2.

Con el aumento de la demanda de energía y la concienciación acerca del cuidado del medioambiente, se ha convertido en una tendencia inexorable el desarrollo y utilización de energías renovables y el aumento de la energía limpia en relación a la energía de uso final. El desarrollo de la red eléctrica en el futuro, no solo debería considerar la demanda de energía, sino también la energía renovable a gran escala (Jin et al., 2017).

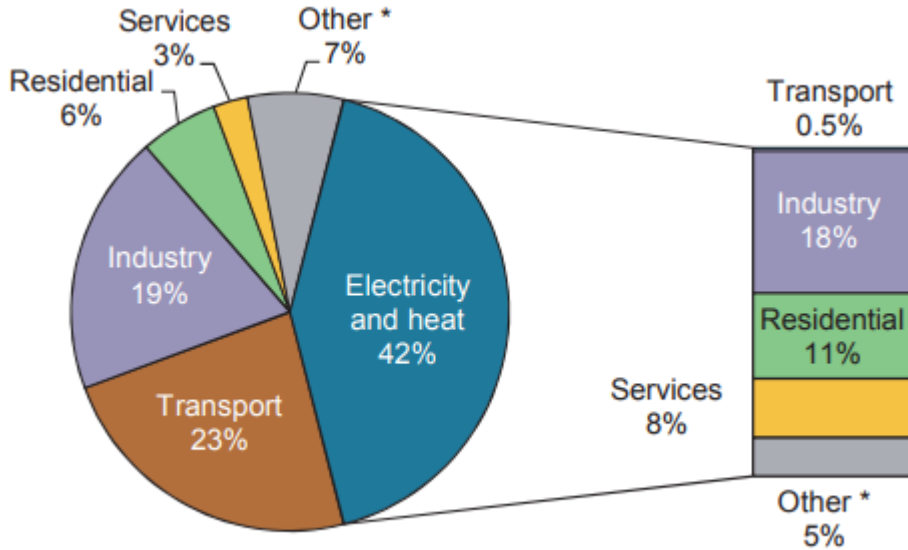


Figura 2. Emisiones de CO₂ a nivel mundial, derivadas del uso de combustibles fósiles por sector en 2014. Fuente: IAE.

1.2 Energías renovables

Se denomina energía renovable a la energía que se obtiene de fuentes naturales virtualmente inagotables, ya sea por la inmensa cantidad de energía que contienen, o porque son capaces de regenerarse por medios naturales (Casas et al., 2008). Entre las energías renovables se encuentran la eólica, geotérmica, hidroeléctrica, mareomotriz, solar, undimotriz, la biomasa y los biocarburantes.

Recientemente, debido a esta preocupación cada vez mayor por la escasez de energía y la contaminación del medioambiente, una gran cantidad de fuentes de energía sostenibles y renovables están emergiendo gradualmente, como la generación de energía fotovoltaica, la generación de energía eólica, la generación de biomasa, etc (Yang et al., 2017). El uso de estas energías renovables pueden contribuir a resolver, al menos parcialmente, las dificultades de abastecimiento que al parecer se presentarán inevitablemente en un futuro no tan lejano (Velasco, 2009), y a reducir la contaminación.

El uso de fuentes de energía renovables, producirá una reducción de la dependencia energética de los combustibles fósiles, mejorando la seguridad del suministro energético, ya que la mayor parte de las reservas de combustibles fósiles a nivel mundial pertenecen a países caracterizados por

la inestabilidad política. Además, el uso de energías renovables también contribuye a la reducción de las emisiones de gases de efecto invernadero, la contaminación ambiental y el aumento de la eficiencia de la red eléctrica (Franzitta et al., 2013).

Por estos motivos, en los últimos años los países europeos centraron su atención en la producción de energía eléctrica a partir de fuentes renovables (Asdrubali et al., 2013).

Las energía renovables como la energía fotovoltaica y la energía eólica se han propuesto como una sustitución de los combustibles fósiles con el fin de resolver los problemas medioambientales urgentes en todo el mundo, como la reducción de las emisiones de gases de efecto invernadero (Jiang et al., 2016).

La energía fotovoltaica está siendo reconocida cada vez más como una fuente de energía rentable (Kollimalla et al., 2014), utilizándose ampliamente para satisfacer la creciente demanda de energía (Cabrane et al., 2017), por lo que resulta interesante realizar un estudio sobre ella.

2 Objetivos

El objetivo de este trabajo se centra en predecir la producción de energía fotovoltaica. Esto resulta interesante puesto que las empresas que se dedican la producción de energía mediante plantas fotovoltaicas necesitan facilitar una estimación de su producción a Red Eléctrica de España, grupo empresarial español que actúa como operador del sistema eléctrico en el mercado eléctrico nacional.

Según la legislación vigente, Red Eléctrica de España calcula la desviación hora a hora y día a día entre las previsiones y el rendimiento real. El productor deberá pagar una penalización por todos esos desvíos acumulados, ya sean por exceso o por defecto. Por tanto, uno de los objetivos que persiguen las eléctricas actualmente es la obtención de un sistema que minimice las diferencias en sus predicciones y poder conseguir una mayor rentabilidad en su producción, sin costes añadidos (Jiménez-Pérez & Mora-López, 2016).

Para lograr el objetivo general del trabajo realizaran distintos modelos de predicción:

- Método basado en la persistente
- Regresión lineal

- Regresión polinómica
- Regresión lineal local
- Regresión basada en los k-vecinos más cercanos (KNN)
- Regresión basada en función Kernel

Este estudio tratará de buscar modelos de predicción que mejoren al modelo de la persistente, que como veremos en el apartado 4.3, consiste en asignar a nuestra estimación, el valor de las veinticuatro horas previas.

Todas las regresiones serán medidas a través de diferentes índices de errores, los cuáles serán explicados en el apartado 4.4 Índices de errores.

3 Producción de energía fotovoltaica

La tecnología de generación de energía fotovoltaica es rica en recursos, segura, fiable y libre de contaminación. Es una de las tecnologías de generación de energía más prometedoras de la actualidad. En algunas áreas remotas, montañas e islas, la generación de energía fotovoltaica se ha convertido en la fuente de energía más efectiva y práctica (Yang et al., 2017).

En la actualidad se consigue hacer un uso directo de la radiación solar aprovechando de forma activa la radiación térmica por medio de colectores solares, que son dispositivos que convierten energía radiante en calor de baja temperatura (<100 °C), para suministro de calefacción y agua caliente en viviendas y edificios institucionales y para calentar el agua de piscinas (Velasco, 2009).

Los paneles fotovoltaicos exhiben una fuerte fluctuación que cambia con factores climáticos como la radiación solar y la temperatura (Ongaro et al., 2012). Además, la integración de un sistema de almacenamiento de energía en un sistema basado en la energía fotovoltaica brinda la oportunidad de obtener mejores respuestas de voltaje y corriente para garantizar sus estabilidad y calidad de energía en diferentes modos de funcionamiento, especialmente durante las fluctuaciones de la radiación solar y las variaciones de carga (Jiang et al., 2016). Los dispositivos de almacenamiento de energía también mitigan el problema de variación de voltaje bajo una alta penetración fotovoltaica (Liu et al., 2012; Sugihara et al., 2013).

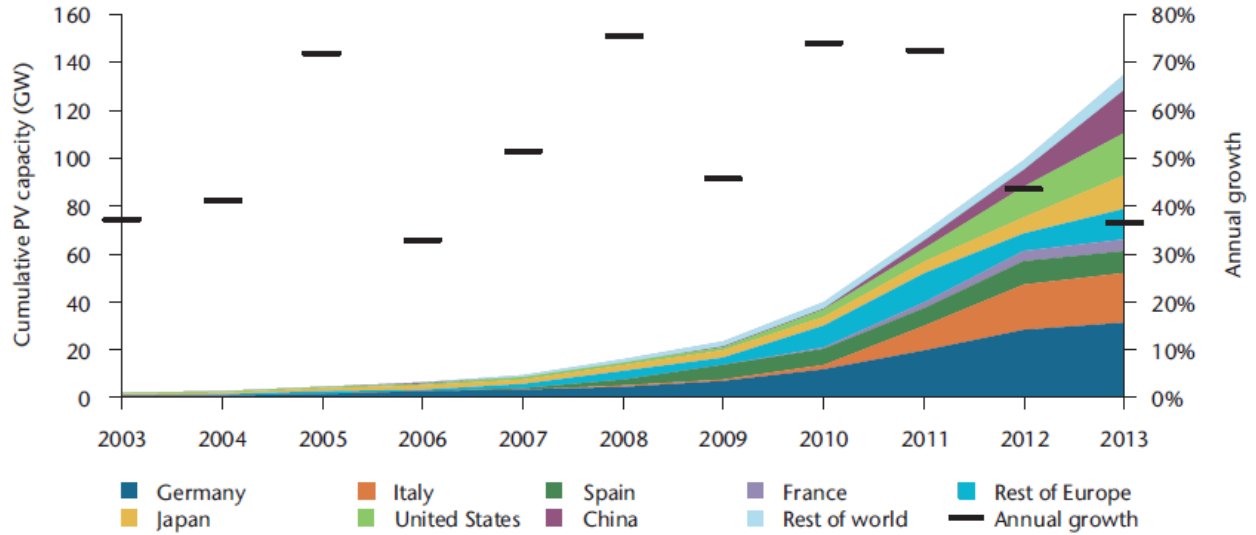


Figura 3. Crecimiento acumulativo global de la capacidad fotovoltaica. Fuente: IEA.

Desde 2003 a 2013, la capacidad instalada acumulada de energía fotovoltaica creció a una tasa promedio del 49% por año (Figura 3). En 2013, se instalaron aproximadamente 37 gigavatios (GW) de nueva capacidad fotovoltaica en aproximadamente 30 países, unos 100 megavatios (MW) por día (IEA, 2014).

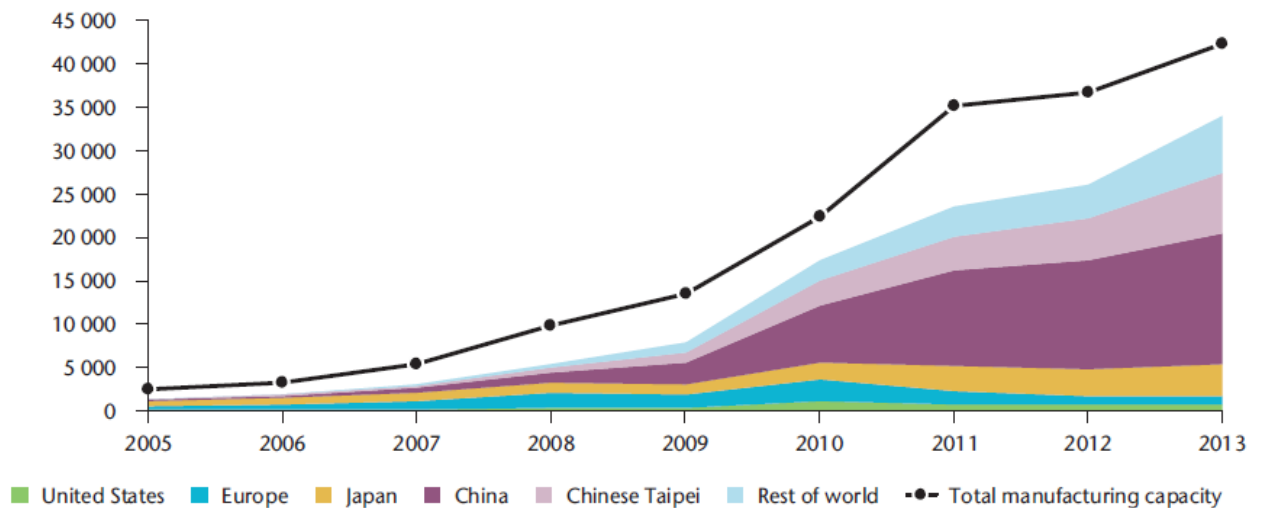


Figura 4. Producción fotovoltaica por países. Fuente: SPV Market Research

En Alemania, más de 1,3 millones de plantas fotovoltaicas generaron casi 30 teravatios hora (TWh) en 2013, el equivalente al 5,3% del consumo de energía del país (Burger, 2014) y la capacidad total alcanzó los 36 GW a finales de 2013. En Italia, los sistemas fotovoltaicos

generaron 22 TWh en 2013, equivalente al 7% del consumo de electricidad, con una capacidad total de 17 GW a finales de 2013. La generación de energía fotovoltaica superó el 3% de la demanda de electricidad en otros cinco países en 2013: Bélgica, Bulgaria, República Checa, Grecia y España (PVPS IA, 2014; RED, 2014). La Figura 4 muestra la evolución de la producción fotovoltaica por regiones geográficas.

La producción de energía solar fotovoltaica depende de los patrones de luz diurna y del clima, especialmente la capa de nubes y la turbidez atmosférica (IEA, 2014). La nubosidad es solo parcialmente predecible en áreas pequeñas, pero la incertidumbre con respecto a la cobertura agregada de nubes, que debe distinguirse de la variabilidad, se reduce a escalas geográficas más grandes. De hecho, la generación de energía solar ahora es más fácil de planificar gracias a pronósticos cada vez más fiables. Los cambios en la cobertura de nubes generalmente no son capaces de generar cambios repentinos imprevistos en la generación en algún nivel de agregación (Figura 5).

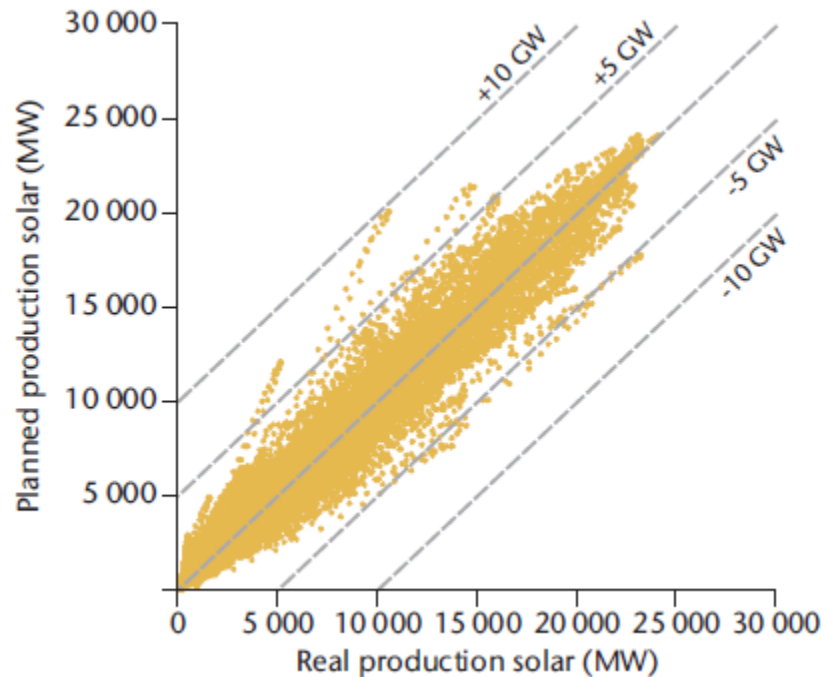


Figura 5. Producción solar por horas prevista vs real en Alemania, 2013. Fuente: Burger.

4 Metodología

La metodología realizada durante este estudio fue la siguiente: en primer lugar, se eligió un conjunto de datos, el cual contenía 3 variables, una temporal, otra con datos de radiación solar estimada, y por último, otra variable con la producción fotovoltaica real de cada hora.

Posteriormente, se eligieron una serie de modelos a través de los cuales realizar las predicciones de producción solar. Estos modelos fueron los siguientes: método basado en la persistente, regresión lineal, regresión polinómica, regresión lineal local, regresión basada en los k-vecinos más cercanos, y por último, regresión basada en función Kernel.

Por último, se establecieron una serie de índices a través de los cuales se medirían cada uno de los métodos de predicción aplicados, con el fin de buscar el método que obtuviese las predicciones con una mayor precisión.

4.1 Datos

Los datos analizados en este trabajo corresponden a predicciones de radiación solar de la Agencia Estatal de Meteorología (AEMET) realizadas por horas y durante un año completo (2013); junto con la producción que realmente se registró en cada hora. Los datos de la producción real fueron aportados por el coordinador, el Dr. José Manuel Bravo Caro, no pudiéndose revelar la localización de las plantas fotovoltaicas de donde fueron extraídos por motivos contractuales. Solo es posible indicar que fueron recogidos en España, en un año natural completo y de forma horaria.

Los datos están formados por una tabla de 8.760 registros (los 365 días del año multiplicados por las 24 horas del día) y tres columnas:

1. Fecha y hora en formato MatLab/Octave.
2. Datos de radiación prevista por AEMET.
3. Datos de producción real recogidos por las plantas fotovoltaicas.

4.2 Variables

Las variables usadas para este modelo fueron 2: radiación prevista por AEMET y producción real a través de datos recogidos en plantas fotovoltaicas. La radiación solar fue definida como la variable independiente o explicativa (x) y la producción solar como la variable dependiente o explicada.

La radiación solar es la radiación electromagnética procedente del Sol. La radiación solar incidente en el límite de la atmósfera terrestre se denomina radiación solar extraterrestre; el 97% de la misma está contenida dentro del intervalo espectral comprendido entre 290 y 3000 nm y se denomina radiación de onda corta (AEMET). La radiación solar se mide en W/m^2 según las unidades del sistema internacional.

La energía solar fotovoltaica es una fuente de energía que produce electricidad de origen renovable (Pearce, 2002), obtenida directamente a partir de la radiación solar mediante un dispositivo semiconductor denominado célula fotovoltaica (Chemistry Encyclopedia, 2017), o bien mediante una deposición de metales sobre un sustrato denominada célula solar de película fina. La producción solar fotovoltaica se mide en vatios.

4.3 Modelos de predicción

El estudio se llevó a cabo realizando diferentes modelos de predicción programados en MatLab. En primer lugar, se tomó como predicción de la producción, el valor de las 24 horas previas. Posteriormente, se realizaron regresiones lineales y polinómicas; para concluir con regresiones más sofisticadas, como la regresión lineal local, regresión basada en los k-vecinos más cercanos (KNN) y regresión basada en función Kernel.

Para todos los modelos, se tomó en cuenta unos datos de entrenamiento correspondientes a los 15 días previos.

En el caso de que la predicción tomase valores inferiores a cero, se incluyó una condición en el código para que dicha predicción fuese cero, ya que no existen producciones negativas. Igualmente, para las predicciones que resultaran superiores al máximo de los 15 días de entrenamiento, se programó que ésta fuese igual al máximo.

```

if y_p < 0
    y_p = 0;
elseif y_p > max(y_training)
    y_p = max(y_training);
end

```

4.3.1 Método basado en la persistente

En el modelo basado en la persistente, se otorgó al valor de la predicción de la producción solar, el valor que ésta tuviese veinticuatro horas previas. Es decir:

$$\hat{y} = y_{24}$$

Siendo y_{24} , la producción real 24 horas antes

4.3.2 Regresión lineal

El modelo de regresión lineal se define como (Bravo, 2015):

$$Y \approx X\theta$$

$$X^T Y = X^T X \theta$$

$$(X^T X)^{-1} X^T Y = \theta^*$$

En nuestro caso, usaremos la variable X de entrenamiento a la que añadiremos un término independiente:

$$X_{training} = \begin{bmatrix} 1 & x_{15} \\ 1 & x_{14} \\ \vdots & \vdots \\ 1 & x_1 \end{bmatrix} \quad Y_{training} = \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix} \quad X_{test} = [1 \quad x_0]$$

Donde x_k corresponderá al valor de x los k días previos.

Theta será calculado como:

$$\theta^* = (X^T X)^{-1} X^T Y = \left(\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{15} & x_{14} & \dots & x_1 \end{bmatrix}^T \begin{bmatrix} 1 & x_{15} \\ 1 & x_{14} \\ \vdots & \vdots \\ 1 & x_1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{15} & x_{14} & \dots & x_1 \end{bmatrix}^T \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix}$$

Obteniéndose el valor de la predicción tras multiplicar X_{test} por θ^* .

$$\hat{y} = X_{test} \theta^* = [1 \quad x_0] \begin{bmatrix} \theta^1 \\ \theta^2 \end{bmatrix}$$

4.3.3 Regresión polinómica

Para calcular la regresión polinómica, añadiremos a nuestro modelo anterior una columna más a la matriz X (training y test), en la que se incluirá el valor de x_k^2 .

$$X_{training} = \begin{bmatrix} 1 & x_{15} & x_{15}^2 \\ 1 & x_{14} & x_{14}^2 \\ \vdots & \vdots & \vdots \\ 1 & x_1 & x_1^2 \end{bmatrix} \quad Y_{training} = \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix} \quad X_{test} = [1 \quad x_0 \quad x_0^2]$$

El valor de θ^* será calculado como:

$$\theta^* = (X^T X)^{-1} X^T Y = \left(\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{15} & x_{14} & \dots & x_1 \\ x_{15}^2 & x_{14}^2 & \dots & x_1^2 \end{bmatrix}^T \begin{bmatrix} 1 & x_{15} & x_{15}^2 \\ 1 & x_{14} & x_{14}^2 \\ \vdots & \vdots & \vdots \\ 1 & x_1 & x_1^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{15} & x_{14} & \dots & x_1 \\ x_{15}^2 & x_{14}^2 & \dots & x_1^2 \end{bmatrix}^T \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix}$$

Y nuestra estimación se obtendrá así:

$$\hat{y} = X_{test} \theta^* = [1 \quad x_0 \quad x_0^2] \begin{bmatrix} \theta^1 \\ \theta^2 \\ \theta^3 \end{bmatrix}$$

4.3.4 Regresión lineal local

La regresión lineal local se define como (Bravo, 2015):

$$Y \approx X\theta$$

$$WY \approx WX\theta$$

$$X^T WY = X^T WX\theta$$

$$(X^T WX)^{-1} X^T WY = \theta^*$$

Donde W corresponde a los pesos asignados en función de la distancia a x_k .

Las variables en nuestro caso, serán las siguientes:

$$X_{training} = \begin{bmatrix} 1 & x_{15} \\ 1 & x_{14} \\ \vdots & \vdots \\ 1 & x_1 \end{bmatrix} \quad Y_{training} = \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix} \quad X_{test} = [1 \quad x_0] \quad W = \begin{bmatrix} w_0 & 0 & \dots & 0 \\ 0 & w_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{15} \end{bmatrix}$$

De donde despejaremos Theta como:

$$\begin{aligned} \theta^* &= (X^T W X)^{-1} X^T W Y = \\ &= \left(\begin{bmatrix} 1 & \dots & 1 \\ x_{15} & \dots & x_1 \end{bmatrix}^T \begin{bmatrix} w_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_{15} \end{bmatrix} \begin{bmatrix} 1 & x_{15} \\ 1 & x_{14} \\ \vdots & \vdots \\ 1 & x_1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \dots & 1 \\ x_{15} & \dots & x_1 \end{bmatrix}^T \begin{bmatrix} w_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_{15} \end{bmatrix} \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix} \end{aligned}$$

Obteniéndose el valor de nuestra estimación como:

$$\hat{y} = X_{test} \theta^* = [1 \quad x_0] \begin{bmatrix} \theta^1 \\ \theta^2 \end{bmatrix}$$

4.3.5 Regresión basada en los k-vecinos más cercanos (KNN)

El método de los k vecinos más cercanos (KNN) es un método de clasificación supervisada (Aprendizaje, estimación basada en un conjunto de entrenamiento y prototipos) que sirve para estimar la función de densidad $F(x/C_j)$ de los predictores x por cada clase C_j (Silverman & Jones, 1989; Fix & Hodges, 1951).

Este es un método de clasificación no paramétrico, que estima el valor de la función de densidad de probabilidad o directamente la probabilidad a posteriori de que un elemento x pertenezca a la clase C_j a partir de la información proporcionada por el conjunto de prototipos. En el proceso de aprendizaje no se hace ninguna suposición acerca de la distribución de las variables predictoras.

En nuestro modelo:

$$X_{training} = \begin{bmatrix} x_{15} \\ x_{14} \\ \vdots \\ x_1 \end{bmatrix} \quad Y_{training} = \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix}$$

Creamos una matriz que contendrá la distancia entre el valor de x_k y nuestra x de test (x_0):

$$distancia = \begin{bmatrix} |x_{15} - x_0| \\ |x_{14} - x_0| \\ \vdots \\ |x_1 - x_0| \end{bmatrix}$$

Creamos una nueva matriz que contenga las Y de training y distancia:

$$vecinos = [Y_{training} \quad distancia] = \begin{bmatrix} y_{15} & |x_{15} - x_0| \\ y_{14} & |x_{14} - x_0| \\ \vdots & \vdots \\ y_1 & |x_1 - x_0| \end{bmatrix}$$

Ordenamos la matriz anterior de forma ascendente según la columna de las diferencias, y calculamos nuestra predicción, dependiendo del número de vecinos (knn) que elijamos.

$$\hat{y} = \frac{1}{knn} (vecinos_{1,1} + \dots + vecinos_{knn,1})$$

4.3.6 Regresión basada en función Kernel

En estadística no paramétrica, una función Kernel es una función de ponderación utilizada en técnicas de estimación no paramétrica (Epanechnikov, 1969). Fue propuesta por primera vez por Rosenblatt (1956), y poseen ciertas propiedades interesantes: son funciones simétricas centradas en cero y su integral vale la unidad; puesto que el Kernel es una función de densidad, su estimador también lo es; y en la mayoría de los casos (dependiendo del Kernel) se trata también de funciones continuas y diferenciables.

En el estudio, se empleó como función Kernel la función Gaussiana.

En nuestro caso, las variables son:

$$X_{training} = \begin{bmatrix} x_{15} \\ x_{14} \\ \vdots \\ x_1 \end{bmatrix} \quad Y_{training} = \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix}$$

Creamos una matriz que contendrá la distancia entre el valor de x_k y nuestra x de test (x_0):

$$distancia = \begin{bmatrix} |x_{15} - x_0| \\ |x_{14} - x_0| \\ \vdots \\ |x_1 - x_0| \end{bmatrix}$$

Los pesos serán calculados como:

$$Pesos = \frac{1}{\gamma \text{ distancia}} = \frac{1}{\gamma} \begin{bmatrix} |x_{15} - x_0| \\ |x_{14} - x_0| \\ \vdots \\ |x_1 - x_0| \end{bmatrix} = \begin{bmatrix} \frac{|x_{15} - x_0|}{\gamma} \\ \frac{|x_{14} - x_0|}{\gamma} \\ \vdots \\ \frac{|x_1 - x_0|}{\gamma} \end{bmatrix}$$

El valor de Gamma estará comprendido entre ± 5 , creándose un bucle con tamaño de paso 0,05 con el que buscaremos el valor del Gamma óptimo, es decir, el que logre la predicción con minimizando el error.

Finalmente, el valor de nuestra predicción vendrá determinado por:

$$\hat{y} = pesos' Y_{training} = \begin{bmatrix} \frac{|x_{15} - x_0|}{\gamma} & \frac{|x_{14} - x_0|}{\gamma} & \dots & \frac{|x_1 - x_0|}{\gamma} \end{bmatrix} \begin{bmatrix} y_{15} \\ y_{14} \\ \vdots \\ y_1 \end{bmatrix}$$

4.4 Índices de errores

4.4.1 Suma de los errores al cuadrado

La suma de los errores al cuadrado o residual sum of squares (RSS), es la suma de los cuadrados de las diferencias entre los valores reales y estimados de una regresión. Es una medida de discrepancia entre los datos y un modelo estimado (Draper & Smith, 2014).

En un modelo como el nuestro con una sola variable explicatoria, RSS viene dado por:

$$RSS = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Siendo:

- y los valores medidos
- \hat{y} los valores estimados
- N el número de observaciones

4.4.2 Error estándar de la regresión

El error estándar de la regresión o residual standard error (RSE), es el valor que muestra la diferencia entre los valores reales y los estimados de una regresión. Es utilizado para valorar si existe una correlación entre la regresión y los valores medidos. Muchos autores prefieren este dato a otros como el coeficiente de correlación lineal, ya que el error estándar se mide en las mismas unidades que los valores que se estudian. La fórmula sería (CITA):

$$RSE = \sqrt{\frac{1}{N-2} \sum_{i=1}^N (y_i - \hat{y}_i)^2} = \sqrt{\frac{1}{N-2} RSS}$$

4.4.3 Suma total de los cuadrados

En el análisis estadístico de datos, la suma total de los cuadrados o total sum of squares (TSS or SST) está definida como la suma de las diferencias de los cuadrados de todas las observaciones respecto a la media global (Everitt & Skrondal, 2002).

En modelos de regresión estándar, TSS es la suma de los cuadrados de las diferencias de la variable dependiente y su media:

$$TSS = \sum_{i=1}^N (y_i - \bar{y})^2$$

Siendo \bar{y} la media.

4.4.4 Coeficiente de determinación R^2

El coeficiente de determinación, denominado R^2 , es un estadístico usado en el contexto de un modelo estadístico cuyo principal propósito es predecir futuros resultados o probar una hipótesis.

El coeficiente determina la calidad del modelo para replicar los resultados, y la proporción de variación de los resultados que puede explicarse por el modelo (Steel & Torrie, 1960).

El coeficiente de determinación R^2 se define como una proporción de varianza explicada de la varianza "total" de la variable dependiente y (Hayashi, 2000):

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} = 1 - \frac{RSS}{TSS}$$

4.4.5 Error cuadrático medio

En estadística, el error cuadrático medio o mean square error (MSE) de un estimador, mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. El MSE es una función de riesgo, correspondiente al valor esperado de la pérdida del error al cuadrado o pérdida cuadrática. La diferencia se produce debido a la aleatoriedad o porque el estimador no tiene en cuenta la información que podría producir una estimación más precisa (Lehmann & Casella, 2006).

El MSE viene dado por:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \frac{RSS}{N}$$

4.4.6 Error absoluto medio

En estadística, el error absoluto medio o mean absolute error (MAE) es una medida de la diferencia entre dos variables continuas. MAE es la distancia vertical media entre cada punto de dos variables continuas (Willmott & Matsuura, 2005).

En nuestro caso, el MAE viene dado por:

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N}$$

5 Resultados

A continuación, se mostrarán los resultados obtenidos de las predicciones realizadas para cada uno de los modelos explicados en el apartado 4.3.

5.1 Método basado en la persistente

Los resultados obtenidos bajo el modelo de predicción basado en la persistente vienen recogidos en la Tabla 1. En dicha tabla se puede apreciar como este método obtiene resultados muy buenos en los meses de verano, con R^2 superiores al 95% en junio, julio y agosto. Por otro lado, en los meses primaverales y otoñales, los R^2 se sitúan alrededor del 80%. Son los meses invernales, en los que este método se muestra más deficiente, especialmente el mes de marzo en el que los errores mediante este método se disparan, y el coeficiente de determinación alcanza el 31%.

Tabla 1. Índices de errores según método basado en la persistente

Mes	RSS	RSE	TSS	R^2	MSE	MAE
Febrero	188.007.060,84	529,72	751.164.818,05	0,75	279.772,41	214,85
Marzo	456.872.982,07	784,69	662.854.519,22	0,31	614.076,59	393,67
Abril	174.592.906,28	493,12	896.502.126,62	0,81	242.490,15	200,97
Mayo	157.224.773,26	460,32	905.724.951,67	0,83	211.323,62	203,61
Junio	31.293.912,51	208,77	851.945.769,48	0,96	43.463,77	77,13
Julio	22.217.780,22	173,04	832.819.151,76	0,97	29.862,61	62,69
Agosto	36.015.329,62	220,31	823.948.667,44	0,96	48.407,70	80,61
Septiembre	85.459.448,46	345,00	665.173.150,88	0,87	118.693,68	147,11
Octubre	144.125.313,61	440,73	690.568.703,87	0,79	193.716,82	205,66
Noviembre	149.149.177,94	455,77	700.738.419,75	0,79	207.151,64	167,48
Diciembre	165.103.976,95	471,71	622.946.790,20	0,73	221.913,95	173,75

Gráficamente, la Figura 6 muestra la producción estimada en relación a la producción real. Los puntos cercanos a la diagonal entre el extremo inferior izquierdo y el extremo superior derecho significan que son buenas estimaciones, es decir, con poco error. Por el contrario, los puntos lejanos a esta diagonal se consideran malas predicciones. En el mes de julio, los puntos se aproximan de manera casi uniforme a esta diagonal, mientras que en el mes de marzo no se puede apreciar ningún patrón concreto, es más una nube de puntos. De ahí que el R^2 obtenido sea muy bajo.

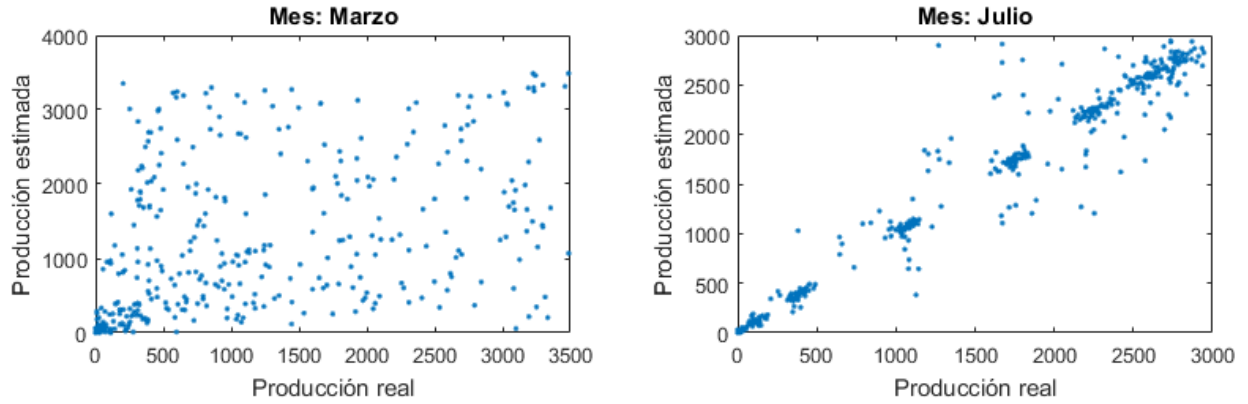


Figura 6. Producción estimada vs producción real. Método de la persistente (marzo y julio).
Fuente: Elaboración propia

Otra forma de ver gráficamente la producción real junto con la producción estimada, la podemos encontrar en la Figura 8.

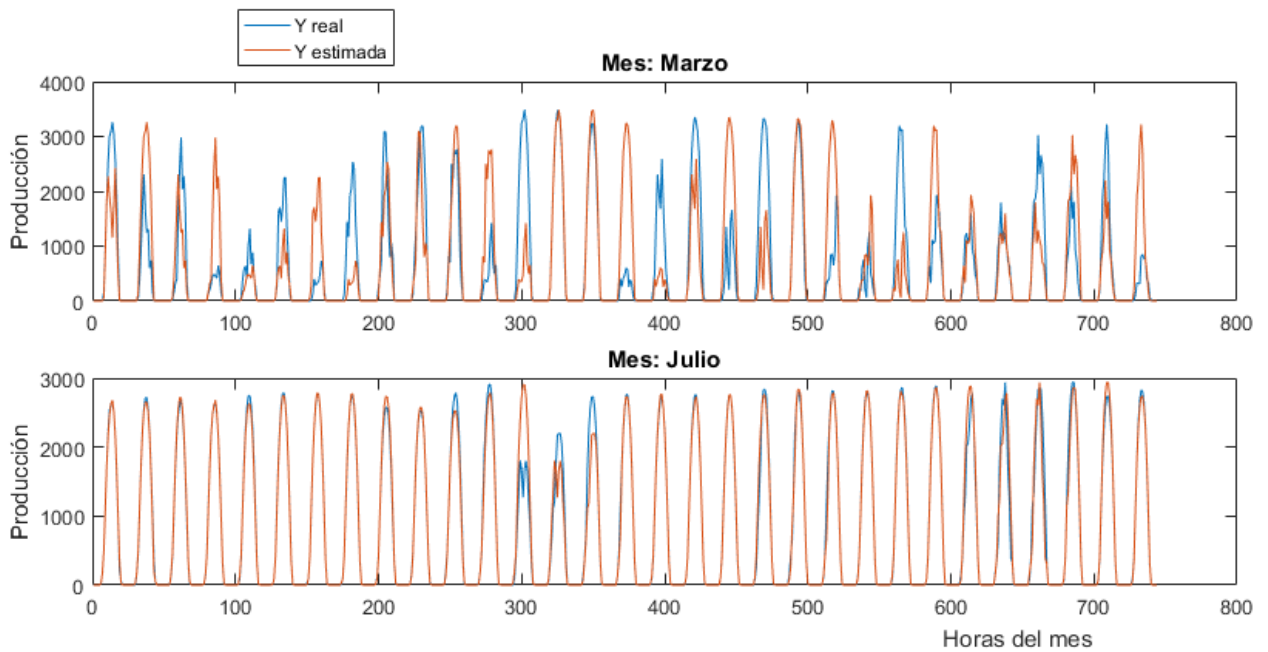


Figura 7. Evolución de la producción por el método de la persistente (marzo y julio). Fuente: Elaboración propia

En este gráfico se representan en el mismo eje tanto la producción real como la estimada, y su evolución a lo largo del mes. Los picos más altos muestran las horas centrales del día, mientras que los puntos en los que la producción es igual a cero, representan las noches, en las que no

existe producción de energía obviamente. En julio se aprecia como las predicciones son prácticamente idénticas a la producción real en la mayoría de días, salvo en algunos cercanos a la mitad del mes. Por el contrario, en el mes de marzo las diferencias se pueden apreciar fácilmente, ya que solo en algunos días se pueden considerar las predicciones como buenas.

El resto de gráficos de todos los meses relativos al modelo basado en la persistente los podemos encontrar en los anexos II y III.

5.2 Regresión lineal

El segundo modelo de predicción calculado fue la regresión lineal. La Tabla 2 muestra los índices de errores de este método. En ella se puede apreciar como este método mejora al de la persistente en todos los meses salvo en dos, junio y agosto, aunque la diferencia en el R^2 es tan solo de un 1 por ciento en cada mes.

Gráficamente, podemos ver la evolución del R^2 en la Figura 8.

En este caso se obtienen R^2 superiores al 80% en todos los casos, a excepción de marzo, aunque se obtiene un valor más de dos veces superior al obtenido por el método de la persistente (69% frente al 31%). Esta diferencia entre ambos métodos en marzo, se puede observar con mayor claridad en la Figura 9. La producción real se asemeja en mayor medida en el caso de la regresión lineal, no existiendo diferencias extremadamente grandes como sí ocurre en el método basado en la persistente.

Tabla 2. Índices de errores según regresión lineal

Mes	RSS	RSE	TSS	R^2	MSE	MAE
Febrero	147.204.714,93	468,73	751.164.818,05	0,80	219.054,64	212,49
Marzo	203.995.452,94	524,33	662.854.519,22	0,69	274.187,44	251,77
Abril	126.850.005,14	420,32	896.502.126,62	0,86	176.180,56	197,81
Mayo	79.772.186,78	327,89	905.724.951,67	0,91	107.220,68	166,60
Junio	43.077.684,56	244,94	851.945.769,48	0,95	59.830,12	94,09
Julio	17.000.049,50	151,36	832.819.151,76	0,98	22.849,53	65,64
Agosto	43.368.871,27	241,76	823.948.667,44	0,95	58.291,49	82,31
Septiembre	54.065.006,79	274,41	665.173.150,88	0,92	75.090,29	132,64
Octubre	77.087.915,37	322,32	690.568.703,87	0,89	103.612,79	164,74
Noviembre	120.052.296,42	408,91	700.738.419,75	0,83	166.739,30	178,42
Diciembre	56.856.942,42	276,82	622.946.790,20	0,91	76.420,62	118,65

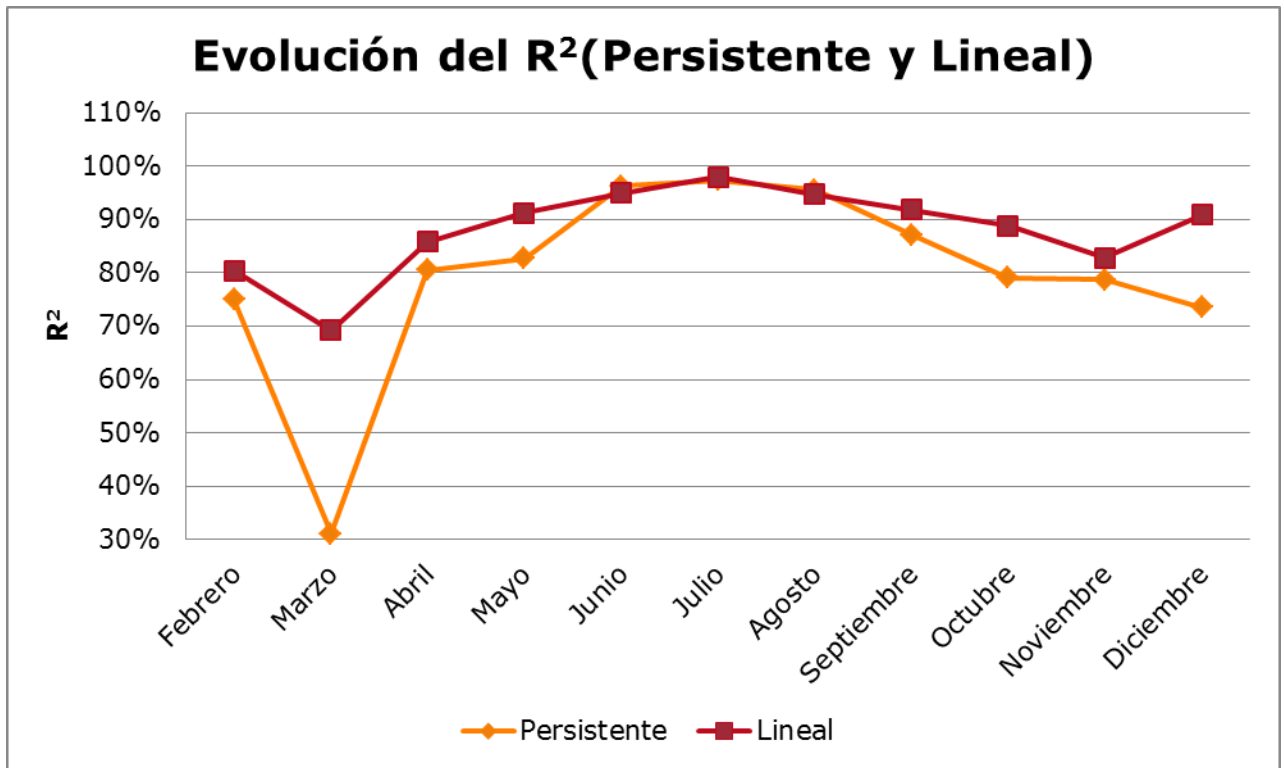


Figura 8. Evolución del R^2 (Persistente y Lineal). Fuente: Elaboración propia

En este caso se obtienen R^2 superiores al 80% en todos los casos, a excepción de marzo, aunque se obtiene un valor más de dos veces superior al obtenido por el método de la persistente (69% frente al 31%). Esta diferencia entre ambos métodos en marzo, se puede observar con mayor claridad en la Figura 9. La producción real se asemeja en mayor medida en el caso de la regresión lineal, no existiendo diferencias extremadamente grandes como sí ocurre en el método basado en la persistente.

En definitiva, podemos asegurar que este modelo mejora al de la persistente, cumpliendo así uno de los objetivos del estudio.

Los gráficos correspondientes al resto de meses, los podemos encontrar en los anexos IV y V.

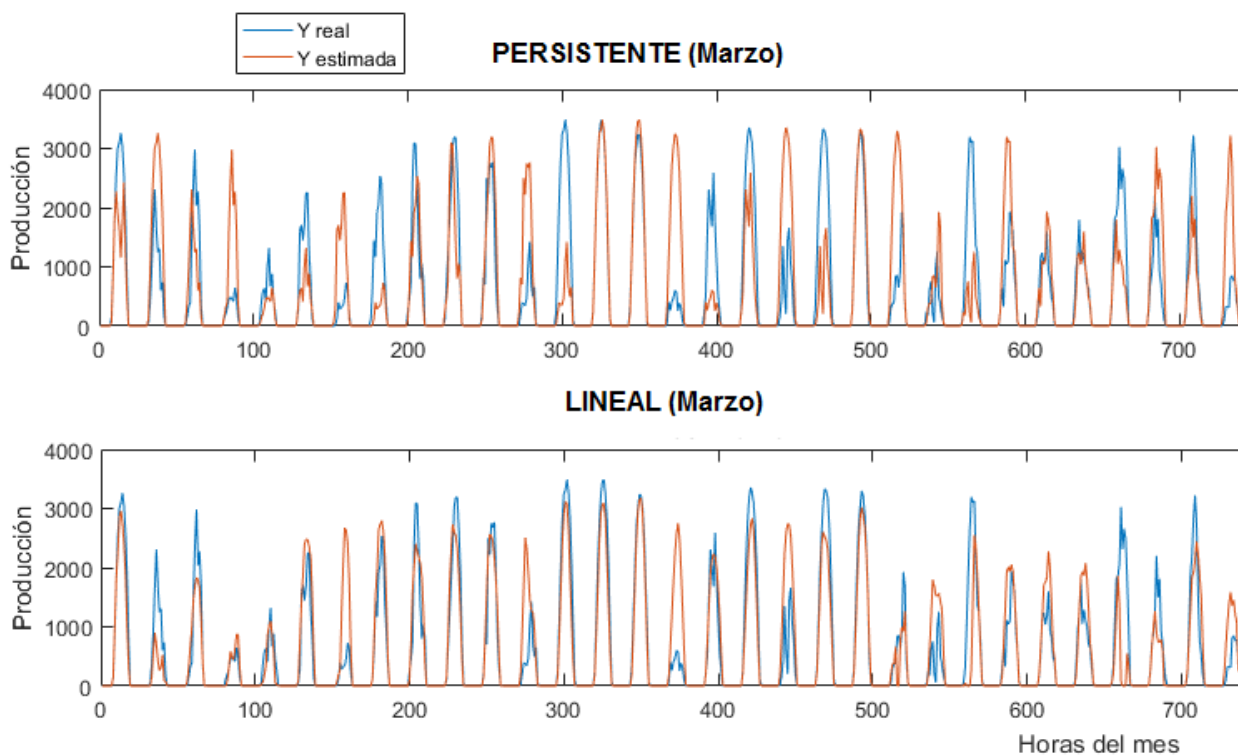


Figura 9. Evolución de la producción real y prevista en marzo. Métodos de la persistente y lineal. Fuente: Elaboración propia

5.3 Regresión polinómica

Al añadir un mayor número de regresores polinómicos, tratamos de mejorar al modelo lineal. La Tabla 3 recoge los índices de errores del modelo polinómico.

Tabla 3. Índices de errores según regresión polinómica

Mes	RSS	RSE	TSS	R^2	MSE	MAE
Febrero	143.224.129,88	462,35	751.164.818,05	0,81	213.131,15	189,96
Marzo	201.747.243,86	521,44	662.854.519,22	0,70	271.165,65	257,78
Abril	157.230.810,20	467,96	896.502.126,62	0,82	218.376,13	206,09
Mayo	100.106.595,20	367,31	905.724.951,67	0,89	134.551,88	174,95
Junio	39.713.733,17	235,18	851.945.769,48	0,95	55.157,96	95,35
Julio	33.676.822,02	213,04	832.819.151,76	0,96	45.264,55	70,56
Agosto	57.468.697,83	278,30	823.948.667,44	0,93	77.242,87	95,63
Septiembre	79.059.377,21	331,83	665.173.150,88	0,88	109.804,69	149,89
Octubre	94.762.200,21	357,37	690.568.703,87	0,86	127.368,55	179,21
Noviembre	139.288.253,79	440,45	700.738.419,75	0,80	193.455,91	190,63
Diciembre	75.890.100,28	319,81	622.946.790,20	0,88	102.002,82	129,15

En este caso vamos a comparar los Residual Sum of Squares (RSS) de la regresión lineal y polinómica. Vemos como los RSS de la regresión polinómica solo son inferiores a los mismos de la regresión lineal en tres meses: febrero, marzo y junio. La reducción de los errores de estos meses es inferior en los tres casos a cuatro millones. En cambio, la reducción de los errores en el resto de meses de la regresión lineal, respecto a la polinómica, es de media superior a 20 millones. Gráficamente, lo podemos ver en la Figura 10.

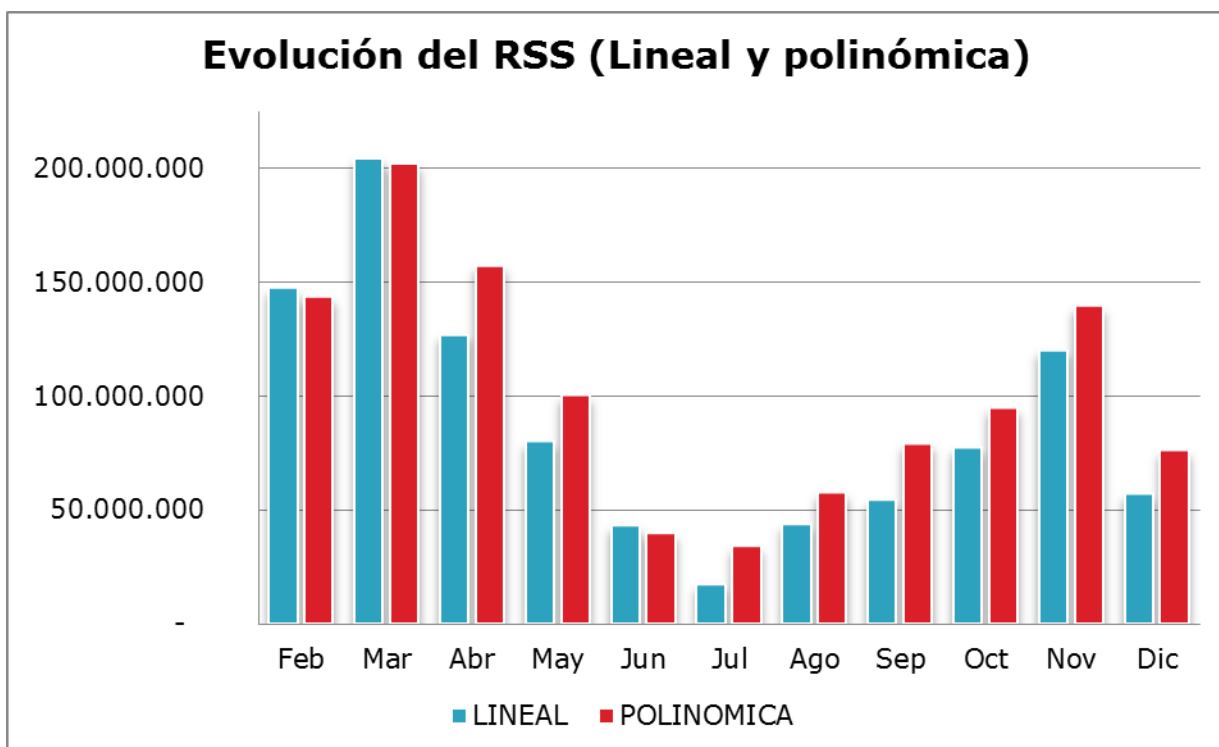


Figura 10. Evolución del RSS. Regresiones lineal y polinómica. Fuente: Elaboración propia

Por tanto, no podemos asegurar que en nuestro caso la regresión polinómica sea mejor a la regresión lineal, por lo que ante la mayor complejidad del modelo, sería más aconsejable utilizar el lineal, ya que incluso en los meses en los que las predicciones son algo peores, las diferencias no son relativamente grandes.

Los gráficos de evolución mensual de la producción y comparativos entre la producción real y producción estimada, los encontraremos en el anexo VI y VII.

5.4 Regresión lineal local

Una vez realizados los modelos basados en parámetros, comenzaremos el estudio de modelos no paramétricos. En primer lugar, analizaremos la regresión lineal local. Los resultados de los índices de errores de este método aparecen recogidos en la Tabla 4.

Tabla 4. Índices de errores según regresión lineal local

Mes	RSS	RSE	TSS	R ²	MSE	MAE
Febrero	39.306.280,10	242,21	751.164.818,05	0,95	58.491,49	77,78
Marzo	51.249.438,60	262,81	662.854.519,22	0,92	68.883,65	100,25
Abril	35.189.202,79	221,38	896.502.126,62	0,96	48.873,89	77,01
Mayo	24.517.968,23	181,78	905.724.951,67	0,97	32.954,26	65,12
Junio	9.422.229,07	114,56	851.945.769,48	0,99	13.086,43	29,76
Julio	3.225.682,91	65,93	832.819.151,76	1,00	4.335,60	17,29
Agosto	13.540.793,81	135,09	823.948.667,44	0,98	18.199,99	24,75
Septiembre	15.826.598,40	148,47	665.173.150,88	0,98	21.981,39	52,00
Octubre	25.719.871,42	186,18	690.568.703,87	0,96	34.569,72	66,54
Noviembre	42.653.309,44	243,73	700.738.419,75	0,94	59.240,71	69,03
Diciembre	12.342.798,90	128,97	622.946.790,20	0,98	16.589,78	38,16

Los resultados de este método son excelentes, obteniéndose para todos los meses coeficientes de determinación superiores a 0,90. Incluso uno de los meses (julio) se acerca a 1 (0,9961). La Figura 11 muestra la evolución del R², a lo largo de todo el año.

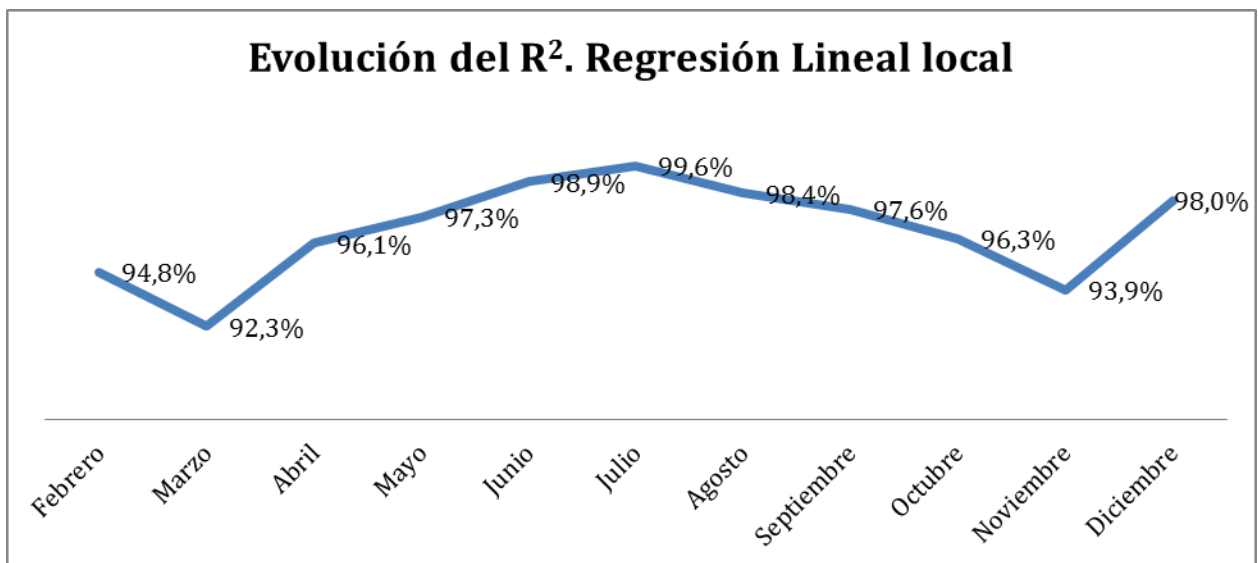


Figura 11. Evolución del R². Regresión lineal local. Fuente: Elaboración propia

Para comparar la regresión lineal local y la regresión lineal utilizaremos el error estándar de la regresión (RSE). La Figura 12 muestra la evolución de este índice de error a través de los dos modelos.

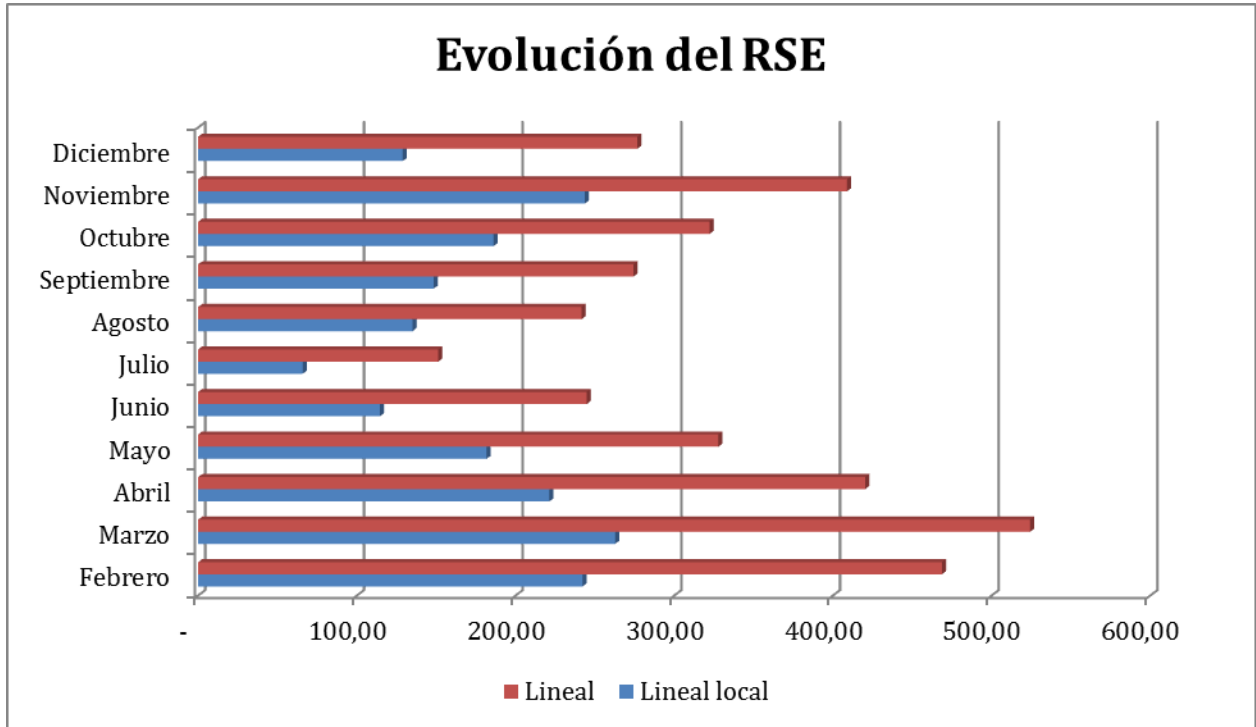


Figura 12. Evolución del RSE. Regresiones lineal y lineal local. Fuente: Elaboración propia

Vemos como la reducción de los errores de la regresión lineal local respecto de la lineal es de aproximadamente la mitad, una reducción de un 47,80% de media, siendo noviembre el mes en el que este índice se reduce en menor medida en términos porcentuales (40,39%), y julio el que más (56,44%).

En la Figura 13 vemos la comparativa entre la producción real y la producción estimada, según los modelos de regresión lineal y lineal local. En el caso de la regresión lineal, los puntos están mucho más dispersos, mientras que en la regresión lineal local los puntos, en su mayoría, se sitúan alrededor de la diagonal entre el extremo superior derecho y el inferior izquierdo, lográndose por tanto, mejores predicciones.

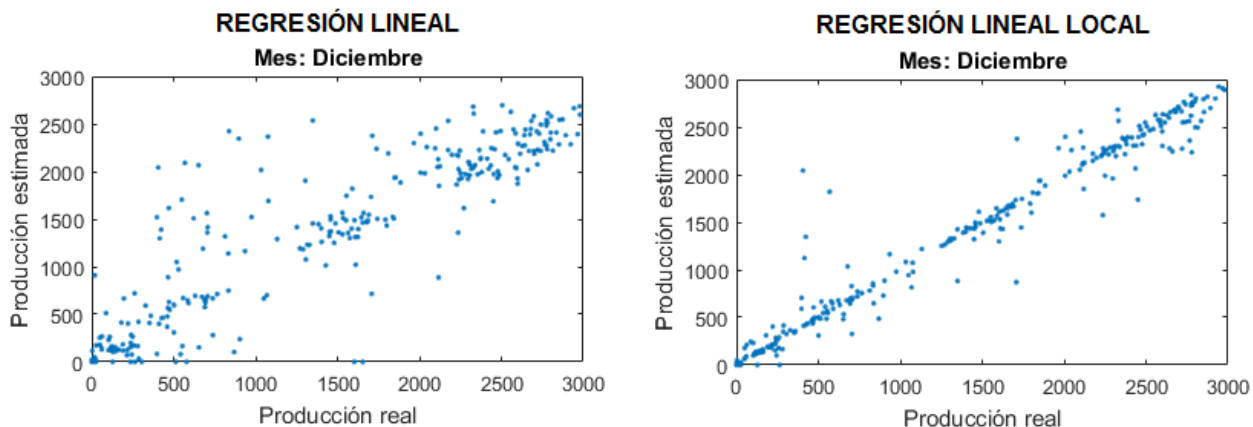


Figura 13. Y real vs Y estimada. Regresiones lineal y lineal local. Fuente: Elaboración propia

El resto de gráficos Y real vs Y estimada de la regresión lineal local se encuentran recogidos en el Anexo IX.

Otro gráfico que muestra la precisión de este modelo lo vemos en la Figura 14. La línea naranja que representa la producción estimada, sigue casi a la perfección las curvas trazadas por la producción real (línea azul).

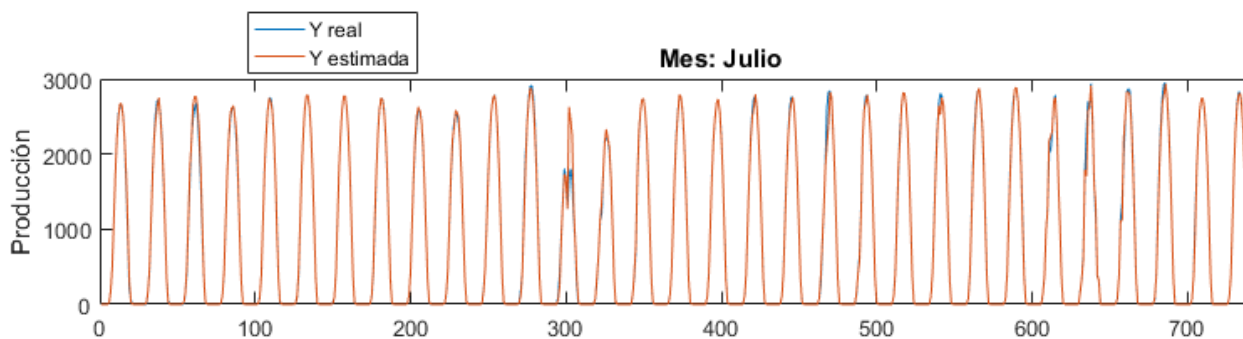


Figura 14. Evolución de la producción por el método de la regresión lineal local (julio). Fuente: Elaboración propia.

La evolución del método de la regresión lineal local, muestra como a partir de marzo los errores comienzan a reducirse hasta llegar a su pico más bajo en verano (julio). Posteriormente, vuelve a crecer de forma progresiva hasta noviembre, volviendo a decrecer en diciembre.

Los gráficos de evolución de la producción del resto de meses se encuentran en el Anexo VIII.

5.5 Regresión basada en los k-vecinos más cercanos (KNN)

El segundo método no paramétrico utilizado fue la regresión basada en los k-vecinos más cercanos o regresión KNN. Los resultados de los índices de errores de este método aparecen recogidos en la Tabla 5.

Tabla 5. Índices de errores según regresión KNN

Mes	RSS	RSE	TSS	R ²	MSE	MAE
Febrero	70.389.961,88	324,13	751.164.818,05	0,91	104.746,97	110,01
Marzo	64.117.414,03	293,96	662.854.519,22	0,90	86.179,32	123,70
Abril	60.447.313,93	290,15	896.502.126,62	0,93	83.954,60	102,60
Mayo	38.169.976,76	226,81	905.724.951,67	0,96	51.303,73	85,79
Junio	11.524.870,17	126,69	851.945.769,48	0,99	16.006,76	36,57
Julio	8.341.487,41	106,03	832.819.151,76	0,99	11.211,68	31,84
Agosto	14.305.531,53	138,85	823.948.667,44	0,98	19.227,86	36,42
Septiembre	35.498.683,86	222,35	665.173.150,88	0,95	49.303,73	81,62
Octubre	34.623.890,45	216,02	690.568.703,87	0,95	46.537,49	85,38
Noviembre	68.669.665,94	309,26	700.738.419,75	0,90	95.374,54	99,14
Diciembre	25.414.856,40	185,07	622.946.790,20	0,96	34.159,75	54,05

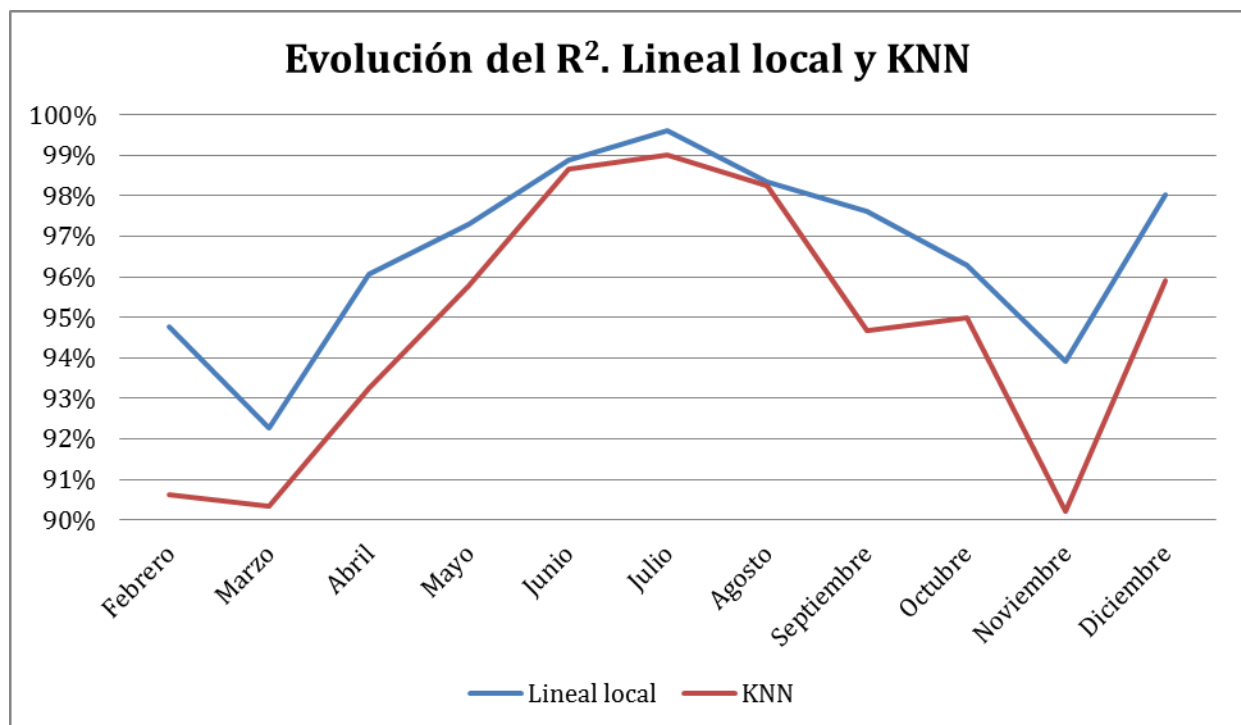


Figura 15. Evolución del R². Regresión lineal local y KNN. Fuente: Elaboración propia.

Los resultados de este método siguen siendo excelentes, obteniéndose para todos los meses coeficientes de determinación superiores a 0,90. No obstante, dichos resultados no mejoran a la regresión lineal local. La Figura 15 muestra la evolución del R^2 , a lo largo de todo el año, según ambos modelos. En ella se ve como el R^2 de la regresión KNN, no supera en ningún mes a la regresión lineal local, aunque si se acerca mucho, especialmente en los meses veraniegos.

Los gráficos tanto de la evolución de la producción, como de la producción real vs producción estimada según el modelo KNN son similares a los de la regresión lineal local, por lo que solo serán añadidos a los Anexos X y XI.

5.6 Regresión basada en función Kernel

El último de los modelos no paramétricos utilizados fue la regresión basada en función Kernel. Los resultados de este método se encuentran reflejados en la Tabla 6.

Tabla 6. Índices de errores según regresión basada en función Kernel

Mes	RSS	RSE	TSS	R^2	MSE	MAE
Febrero	57.732.057,28	293,54	751.164.818,05	0,92	85.910,80	94,51
Marzo	49.311.288,49	257,79	662.854.519,22	0,93	66.278,61	110,01
Abril	60.600.699,74	290,52	896.502.126,62	0,93	84.167,64	103,94
Mayo	23.219.569,83	176,90	905.724.951,67	0,97	31.209,10	66,21
Junio	13.580.889,05	137,53	851.945.769,48	0,98	18.862,35	36,25
Julio	7.878.524,35	103,04	832.819.151,76	0,99	10.589,41	26,95
Agosto	16.058.777,42	147,11	823.948.667,44	0,98	21.584,38	39,53
Septiembre	42.667.244,39	243,77	665.173.150,88	0,94	59.260,06	87,07
Octubre	38.699.319,17	228,38	690.568.703,87	0,94	52.015,21	88,10
Noviembre	47.855.862,45	258,17	700.738.419,75	0,93	66.466,48	85,22
Diciembre	30.687.242,19	203,37	622.946.790,20	0,95	41.246,29	64,00

Los resultados del modelo basado en función Kernel fueron similares a los obtenidos según el modelo KNN. No obstante, este método sí consigue mejorar a la regresión lineal local en dos meses (marzo y mayo), aunque eso sí, la mejora es mínima. La Figura 16 muestra la comparación de la evolución del índice RSS, según la regresión lineal local y la regresión basada en función Kernel.

La Figura 17 muestra la evolución de la producción real y estimada, en el mes de marzo, uno de los meses en los que la regresión basada en función Kernel presenta mejores resultados que la

regresión lineal local. A simple vista, ambos gráficos son muy similares, puesto que como dijimos con anterioridad la diferencia entre ambos modelos es mínima.

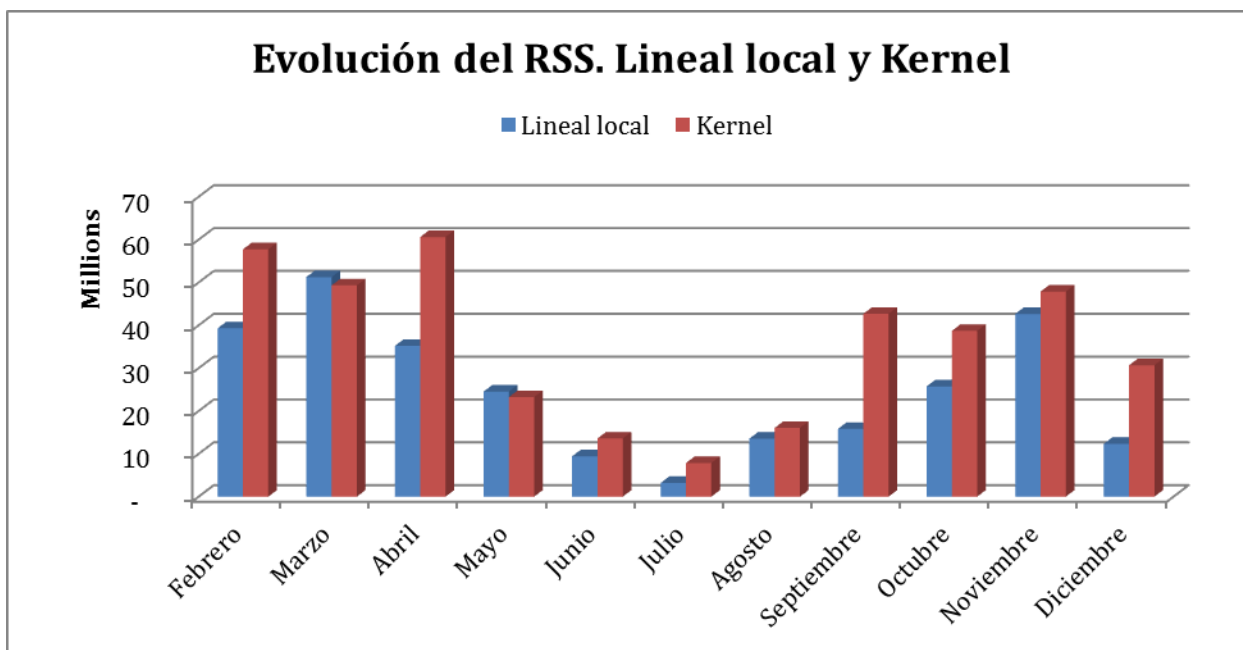


Figura 16. Evolución del RSS. Regresión lineal local y Kernel. Fuente: Elaboración propia.

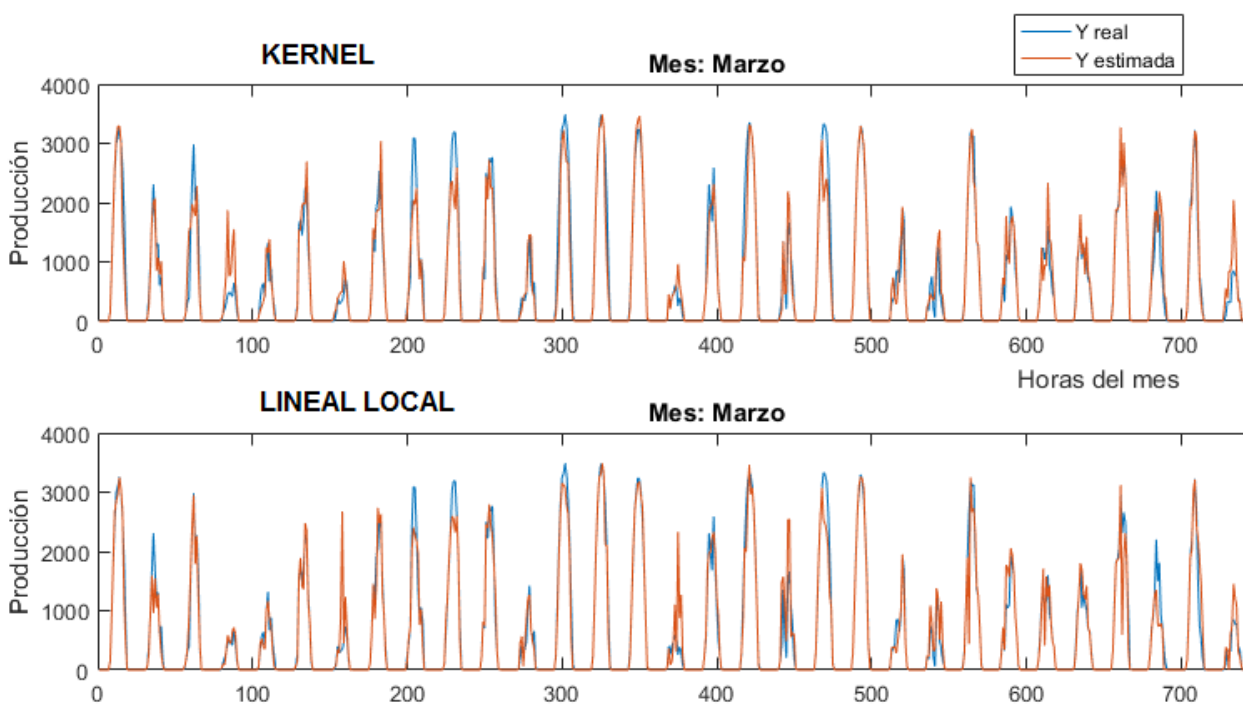


Figura 17. Evolución de la producción (marzo). Regresión basada en función Kernel y regresión lineal local. Fuente: Elaboración propia.

El resto de gráficos de evolución de la producción por horas y meses se encuentran en el Anexo XII; y los gráficos de producción real vs producción estimada forman parte del Anexo XIII.

5.7 Resultados globales

Para comparar de forma global que método fue mejor, se aplicaron los índices de error de forma anual, obteniéndose los resultados que la Tabla 7 recoge.

Tabla 7. Resumen de los índices de errores según los diferentes modelos empleados

Método	RSS	RSE	R ²	MSE	MAE
Lineal local	272.994.173,68	184,57	96,80%	34.056,16	55,95
Kernel	388.291.474,35	220,12	95,45%	48.439,56	72,63
KNN	431.503.652,35	232,04	94,95%	53.830,30	76,68
Lineal	969.331.126,12	347,79	88,65%	120.924,54	150,84
Polinómica	1.122.167.963,63	374,20	86,86%	139.991,01	157,79
Persistente	1.610.062.661,78	448,23	81,14%	200.856,12	175,20

Los métodos no paramétricos son los que mejores resultados nos ofrecen, siendo el mejor método la regresión lineal local, seguido de cerca por los métodos de regresión basados en función Kernel y en los k-vecinos más cercanos. A continuación, los métodos de regresión lineal y polinómica muestran resultados similares, algo mejores en el supuesto lineal. Por último, el método basado en la persistente, por lo que todos los métodos habrían cumplido con uno de los objetivos del estudio, mejorar a este modelo.

Los resultados de todos los índices de error por meses y modelo se encuentran recogidos en el Anexo I, ordenados en forma descendente según el coeficiente de determinación R².

5.7.1 Residual Sum of Square (RSS)

La Figura 18 muestra la evolución de la suma de los residuos al cuadrado según los distintos medios utilizados. Se puede apreciar como el mejor método según este indicador sería la regresión lineal local, seguido por los modelos Kernel y KNN. Un escalón por debajo se situaría la regresión lineal, seguido por la regresión polinómica. Por último, la regresión basada en la persistente, que solo ofrece resultados similares a los métodos no paramétricos en verano.

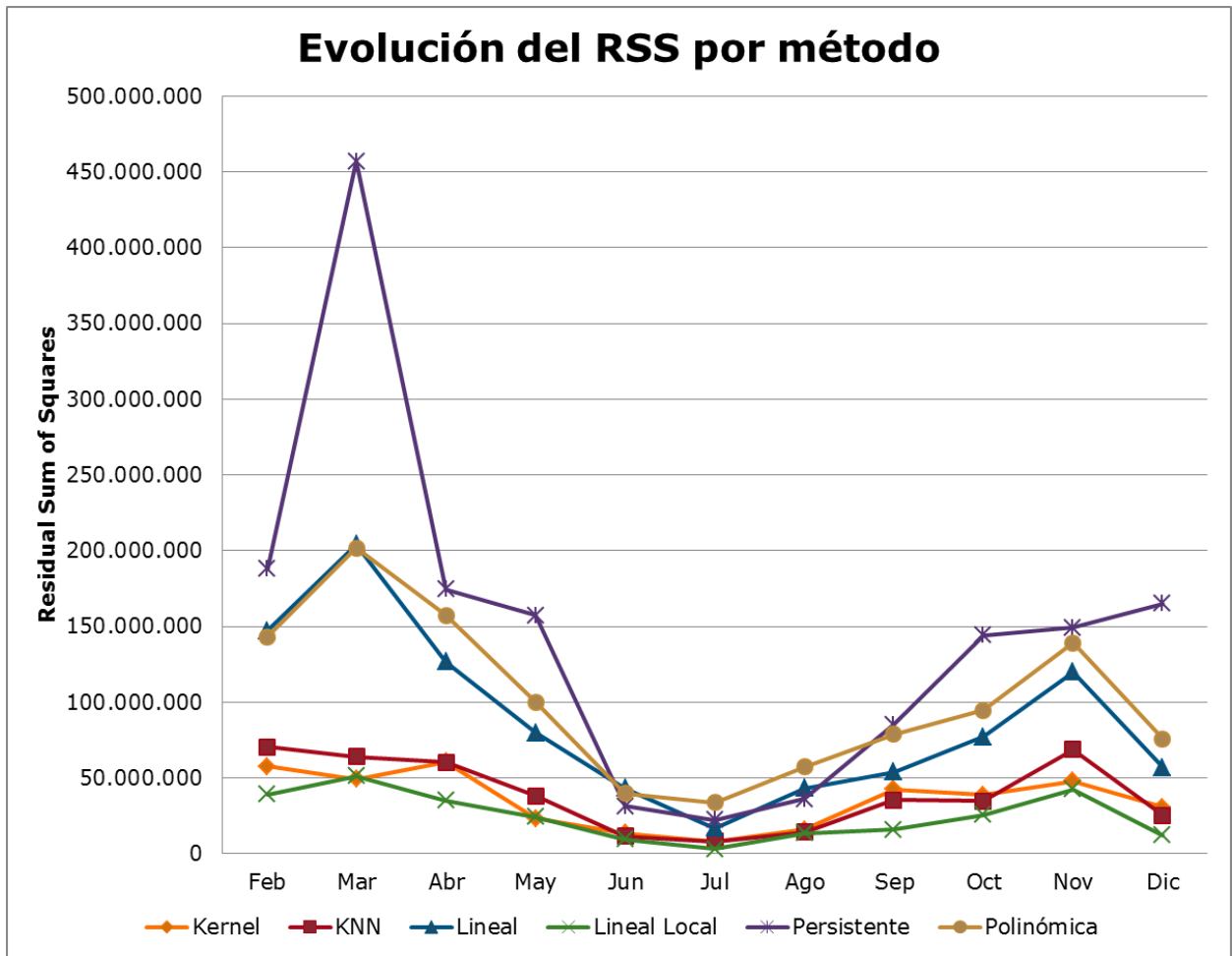


Figura 18. Evolución de la suma de los errores al cuadrado por método. Fuente: Elaboración propia.

5.7.2 Residual Standard Error (RSE)

La evolución del error estándar de la regresión según los diferentes modelos utilizados viene reflejada en la Figura 19.

De nuevo los métodos no paramétricos son los que obtienen mejores resultados, como ya viéramos en el gráfico de la evolución del RSS, siendo el mes de julio el que mejores resultados obtiene para cada uno de los métodos de predicción utilizados.

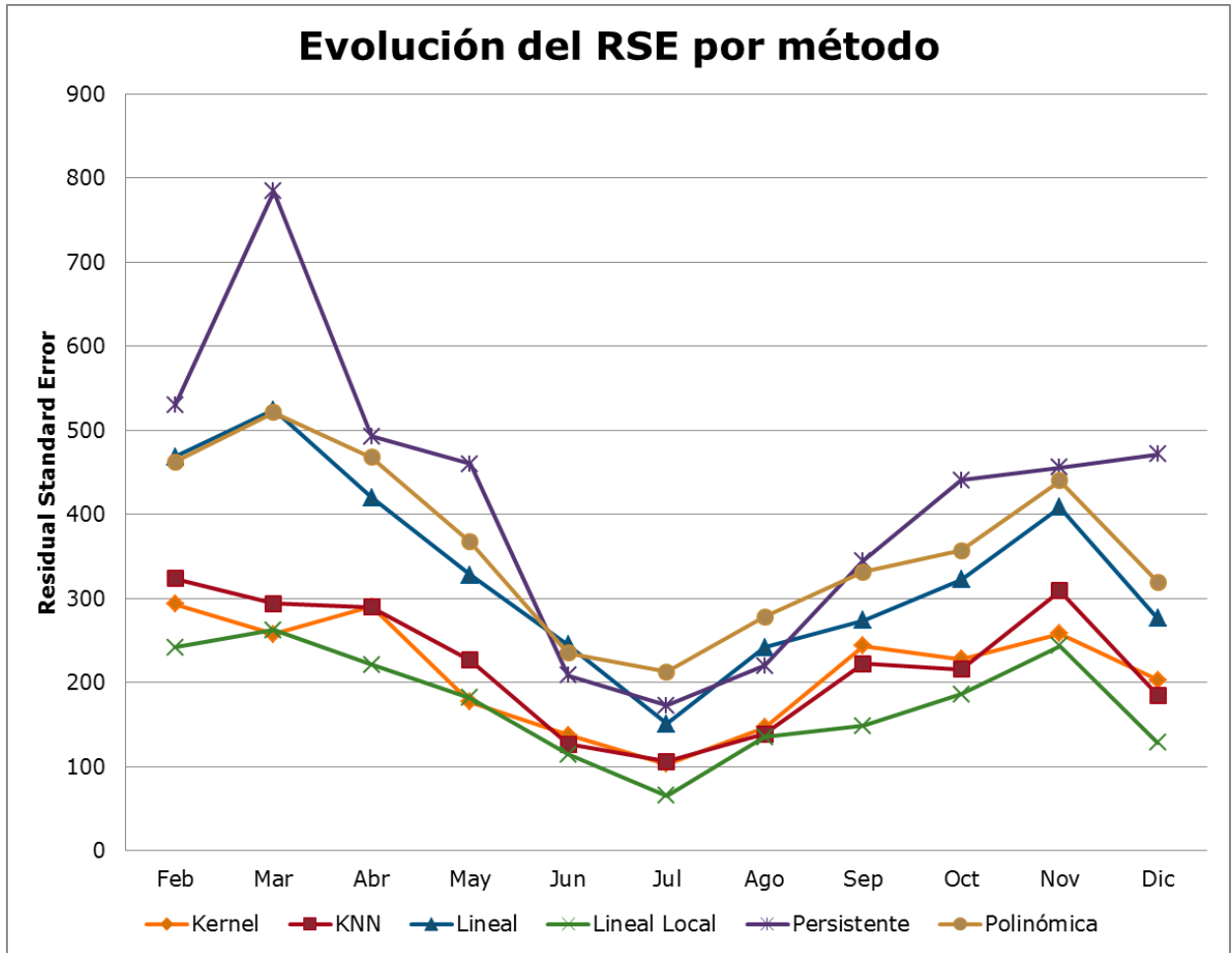


Figura 19. Evolución del error estándar de la regresión por método. Fuente: Elaboración propia.

5.7.3 R square (R^2)

La Figura 20 muestra la evolución del coeficiente de determinación a lo largo de los 11 meses del estudio. Vemos como los método no paramétricos obtienen valores superiores a 0,9 para todos los meses, mientras que el resto de métodos solo obtienen estas cifras en los meses de verano.

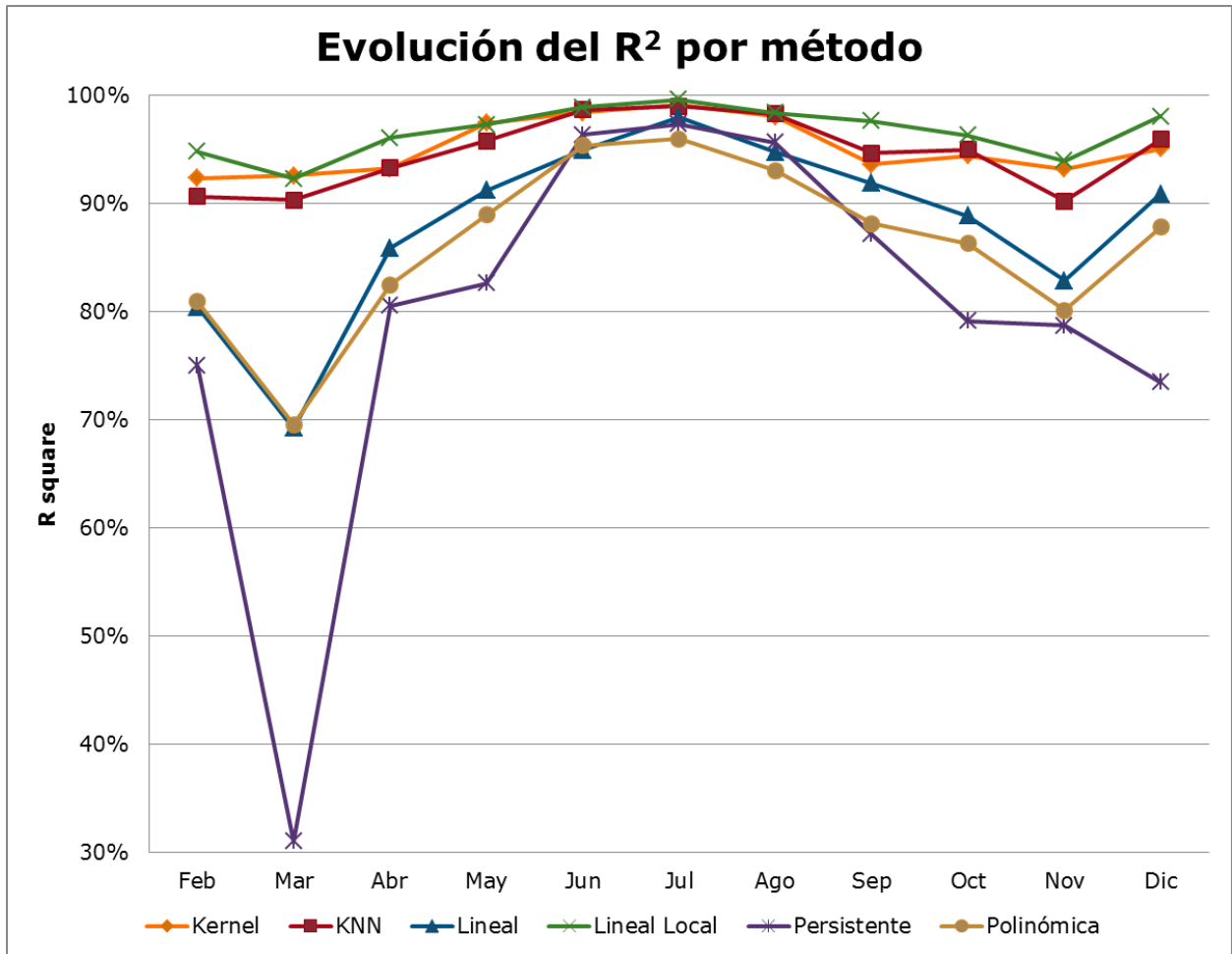


Figura 20. Evolución del R^2 por método. Fuente: Elaboración propia.

5.7.4 Mean Square Error (MSE)

En la Figura 21 se observa la evolución del error cuadrático medio a lo largo del tiempo, según los diferentes índices utilizados. La regresión lineal local vuelve a ser la que mejores resultados ofrece según este índice, mientras que la regresión basada en función Kernel y la regresión basada en los k-vecinos más cercanos, reflejan resultados similares, pero algo peores a la regresión lineal local. En el caso de los métodos no paramétricos, vemos como el MSE por mes no supera los 100.000 puntos, salvo en el caso de la regresión KNN en febrero que los supera ligeramente (104.747). Del mismo modo que para los otros índices, vemos que las regresiones lineal, polinómica y basada en la persistente, muestran mejores resultados conforme los meses se van aproximando al verano, en los que las predicciones son más precisas.

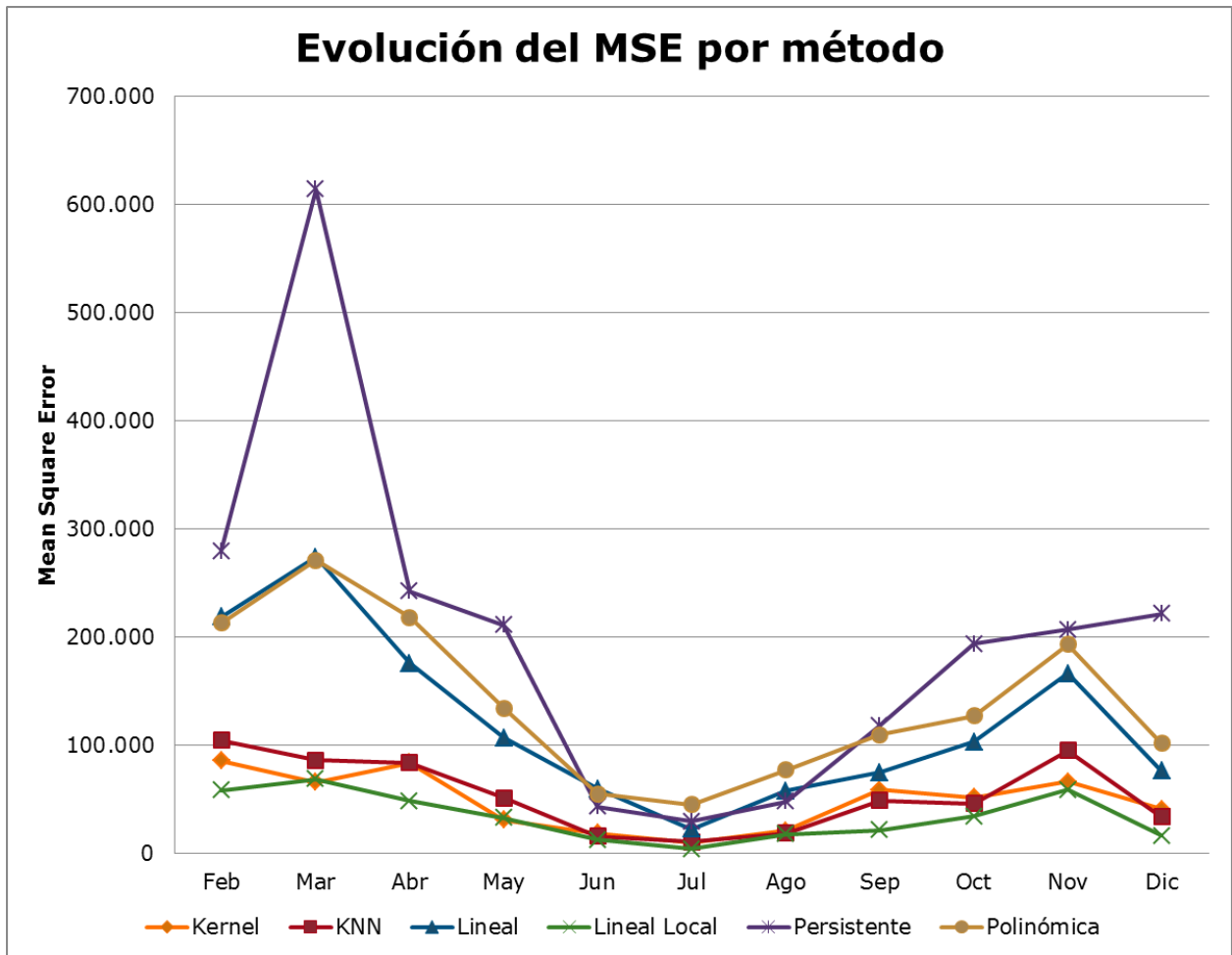


Figura 21. Evolución del error cuadrático medio por método. Fuente: Elaboración propia.

5.7.5 Mean Absolute Error (MAE)

El último índice de error utilizado fue el error absoluto medio. La Figura 22 nos enseña la evolución mensual de dicho índice según los seis modelos de predicción empleados. El orden de selección de método de predicción según este índice sería el siguiente:

1. Regresión lineal local
2. Regresión basada en función Kernel/Regresión basada en los k-vecinos más cercanos.
3. Regresión lineal
4. Regresión polinómica

5. Regresión basada en la persistente.

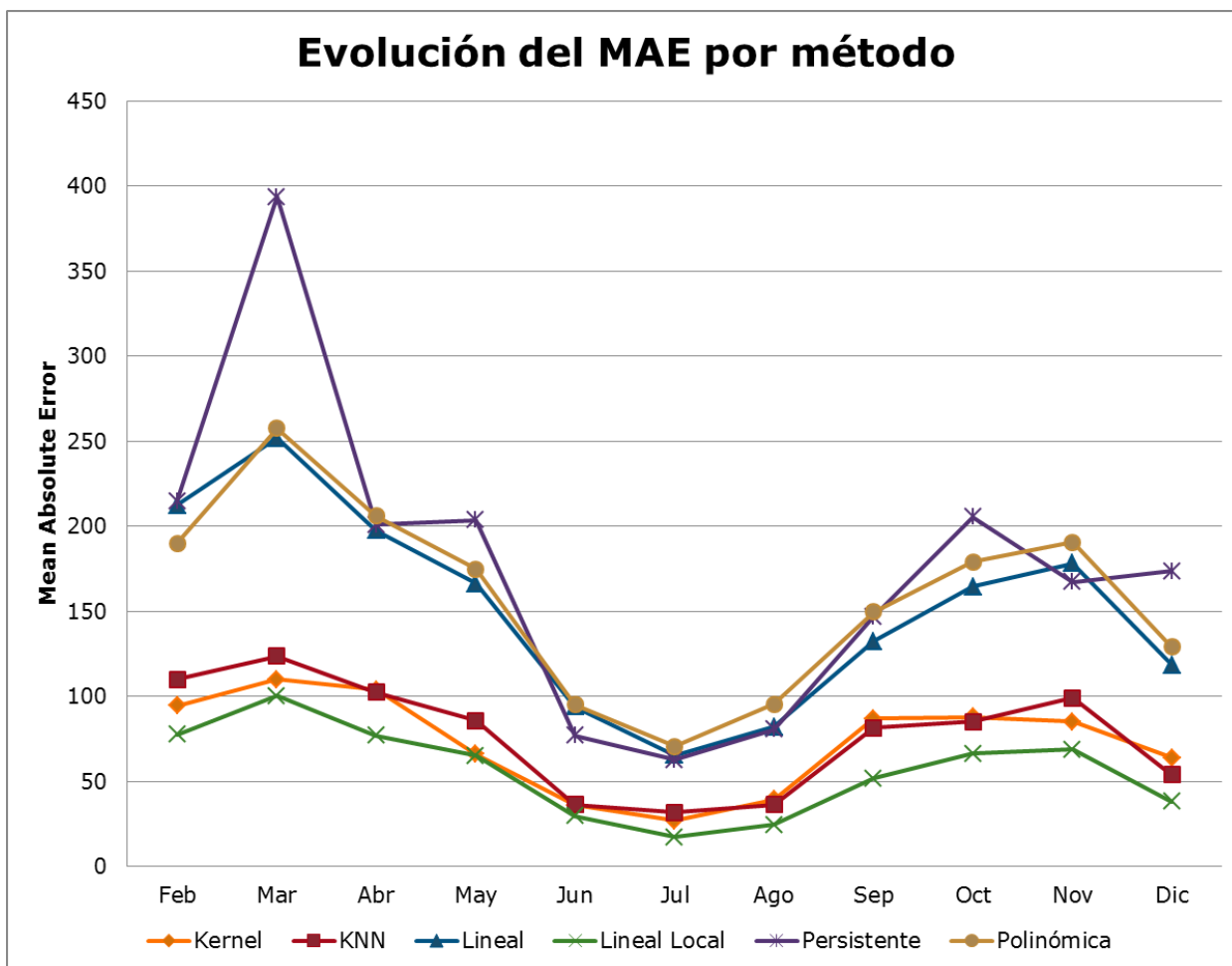


Figura 22. Evolución del error absoluto medio por método. Fuente: Elaboración propia.

6 Conclusiones

En los últimos años, con el aumento de la contaminación y la mayor concienciación por parte de la sociedad respecto al cuidado del medioambiente, han propiciado un auge en las denominadas energías renovables, entre las que destacan: la energía fotovoltaica, la energía eólica, energía mareomotriz y la biomasa. El uso de estas energías limpias podría facilitar la reducción de emisiones de gases de efecto invernadero y mejorar la eficiencia de la red eléctrica.

Una de las fuentes de energías renovables propuesta como alternativa al uso de los combustibles fósiles es la energía fotovoltaica, por ser rica en recursos, segura, confiable y, por supuesto, libre

de contaminación. Por todas estas razones, la energía fotovoltaica es una de las tecnologías generadoras de energía más prometedoras en la actualidad.

El objetivo de este trabajo se centraba en la predicción de producción de energía fotovoltaica a través de un conjunto de datos que contenía información horaria de la radiación solar prevista por AEMET y la producción real que se produjo. En este sentido, para lograr dicho objetivo se realizaron seis modelos diferentes de predicción. El primero de ellos fue simplemente establecer la producción prevista como la producción del día anterior a la misma hora, denominando este método como persistente. Posteriormente, se realizó un modelo de regresión lineal y otro de regresión polinómica, con la intención de mejorar los resultados del anterior. Por último, se ejecutaron modelos de predicción no paramétricos con mayor grado de complejidad. Estos últimos fueron: regresión lineal local, regresión basada en los k-vecinos más cercanos y regresión basada en función Kernel.

Como objetivo secundario se planteó lograr modelos de predicción que mejorasen al modelo inicial de la persistente, puesto que si el mejor modelo de predicción es asignar al valor de ésta, el mismo valor que se obtuvo el día anterior, el análisis predictivo carecería de sentido. Este objetivo fue cumplido por todos los métodos de predicción realizados. No obstante, el modelo de la persistente llega a ser mejor que el lineal y polinómico en los meses de verano como ya ha sido comentado en el apartado dedicado a los resultados.

Para medir los resultados de cada uno de los métodos, se tuvieron en cuenta diferentes índices para la evaluación de los errores. Estos índices fueron: la suma de los errores al cuadrado (RSS), el error estándar de la regresión (RSE), el coeficiente de determinación (R^2), el error cuadrático medio (MSE), y por último, el error absoluto medio (MAE).

Los resultados obtenidos reflejan que en los meses de verano (junio, julio y agosto) los métodos predictivos utilizados no mejoran en gran medida al método inicial de la persistente. En el resto de meses sí se produce una amplia mejoría por parte de los métodos no paramétricos, y una leve mejoría por parte de los modelos lineal y polinómico. Cabe reseñar, que el método que mejores resultados obtiene es el de la regresión lineal local, seguidos de cerca por el método basado en función Kernel y KNN. Por otro lado, el método lineal ofrece mejores resultados que el polinómico, por lo que sería más interesante utilizar este método ya que siendo más sencillo obtiene unas mejores cifras en términos de error.

Referencias bibliográficas

- AEMET. (2017). Radiación y ozono. *Radiación solar*. Recuperado de http://www.aemet.es/es/serviciosclimaticos/vigilancia_clima/radiacion_ozono
- Asdrubali, F., Presciutti, A., & Scrucca, F. (2013). Development of a greenhouse gas accounting GIS-based tool to support local policy making - Application to an Italian municipality. *Energy Policy*, *61*, 587-594.
- Bravo-Caro, J.M. (2015). Técnicas de Aprendizaje Automático.
- Burger, B. (2014). Electricity production from solar and wind in Germany in 2014. *Fraunhofer Institute for Solar Energy Systems ISE*, *29*.
- Cabrane, Z., Ouassaid, M., & Maaroufi, M. (2017, March). Management and control of the integration of supercapacitor in photovoltaic energy storage. In *Green Energy Conversion Systems (GECS), 2017 International Conference on* (pp. 1-6). IEEE.
- Casas, J. M., Gea, F., Javaloyes, E., Martín, A., Pérez, J. A., Triguero, I., & Vives, F. (2008). Educación medioambiental. *Espanya.: Ed. Club Universitario*.
- Draper, N. R., & Smith, H. (2014). *Applied regression analysis*. John Wiley & Sons.
- Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, *14*(1), 153-158.
- Everitt, B., & Skrondal, A. (2002). *The Cambridge dictionary of statistics* (Vol. 106). Cambridge: Cambridge University Press.
- Franzitta, V., Milone, A., Milone, D., Pitruzzella, S., Trapanese, M., & Viola, A. (2014). A Case Study to evaluate the indoor global quality. In *Advanced Materials Research* (Vol. 864, pp. 1054-1058). Trans Tech Publications.
- Franzitta, V., Milone, A., Milone, D., Pitruzzella, S., Trapanese, M., & Viola, A. (2014). Experimental evidence on the thermal performance of opaque surfaces in Mediterranean climate. In *Advanced Materials Research* (Vol. 860, pp. 1227-1231). Trans Tech Publications.
- Franzitta, V., Milone, D., Trapanese, M., Viola, A., Di Dio, V., & Pitruzzella, S. (2013). Energy and economic comparison of different conditioning system among traditional and eco-sustainable building. In *Applied Mechanics and Materials* (Vol. 394, pp. 289-295). Trans Tech Publications.
- Hayashi, F. (2000). *Econometrics*.

- Hosseini, S.J., Moazzami, M., Shahinzadeh, H. & Gharehpetian, G.B. (2017). Iran's long-term energy planning based on ground source heat pumps (GSHP) Scenario. In *Renewable Energy and Distributed Generation (ICREDG), 2016 Fourth Iranian Conference, 2016 Conference*. IEEE.
- International Energy Agency. (2014). *Technology Roadmap: Solar Photovoltaic Energy*. Paris, France: International Energy Agency.
- International Energy Agency. (2015). *CO2 emissions from fuel combustion Highlights*. Paris, France: International Energy Agency
- International Energy Agency. (2015). *Energy and Climate Change: World Energy Outlook Special Report*. Paris, France: International Energy Agency
- International Energy Agency. (2016). *CO2 emissions from fuel combustion Highlights*. Paris, France: International Energy Agency
- Jha, S. K. (2016). Implication of optimal performance of renewable energy system. *IEEE 7th Power India International Conference (PIICON)*, Bikaner, 2016, pp. 1-5
- Jiang, W., Zhang, L., Zhao, H., Huang, H., & Hu, R. (2016). Research on power sharing strategy of hybrid energy storage system in photovoltaic power station based on multi-objective optimisation. *IET Renewable Power Generation*, 10(5), 575-583.
- Jiménez-Pérez, P. F., & Mora-López, L. (2016). Modeling and forecasting hourly global solar radiation using clustering and classification techniques. *Solar Energy*, 135, 682-691.
- Jin, X., Wen, Z., Weidong, L., Yan, L., Xinshou, T., Chao, L., & Linjun, W. (2017, August). Study on the driving force and challenges of developing power grid with high penetration of renewable energy. In *Transportation Electrification Asia-Pacific (ITEC Asia-Pacific), 2017 IEEE Conference and Exp* (pp. 1-5). IEEE.
- Kollimalla, S. K., Mishra, M. K., & Narasamma, N. L. (2014). Design and analysis of novel control strategy for battery and supercapacitor storage system. *IEEE Transactions on Sustainable Energy*, 5(4), 1137-1144.
- Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.
- Liu, X., Aichhorn, A., Liu, L., & Li, H. (2012). Coordinated control of distributed energy storage system with tap changer transformers for voltage rise mitigation under high photovoltaic penetration. *IEEE Transactions on Smart Grid*, 3(2), 897-906

- Memar, M. R., Moazzami, M., Shahinzadeh, H., & Fadaei, D. (2017, April). Techno-economic and environmental analysis of a grid-connected photovoltaic energy system. In *Electrical Power Distribution Networks Conference (EPDC), 2017 Conference on* (pp. 124-130). IEEE.
- Ongaro, F., Saggini, S., & Mattavelli, P. (2012). Li-ion battery-supercapacitor hybrid storage system for a long lifetime, photovoltaic-based wireless sensor network. *IEEE Transactions on Power Electronics*, 27(9), 3944-3952.
- Pearce, J. M. (2002). Photovoltaics—a path to sustainable futures. *Futures*, 34(7), 663-674.
- PVPS IA (Implementing Agreement for a Co-operative Programme on Photovoltaic Power Systems) (2014), *PVPS Report Snapshot of Global PV 1992-2013*, www.iea-pvps.org.
- PVPS IA (Implementing Agreement for a Co-operative Programme on Photovoltaic Power Systems) (2014), *PV Cost Data for the IEA*, personal communication, January.
- RED (RED eléctrica de España) (2014), *The Spanish Electricity System – Preliminary Report 2013*, RED, Madrid, Spain, www.ree.es/sites/default/files/downloadable/preliminary_report_2013.pdf.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3), 832-837.
- Shahinzadeh, H., Fathi, S. H., and Hasanalizadeh-Khosroshahi, A. (2016). Long-Term Energy Planning in IRAN using LEAP Scenario: Using Combined Heat and Power (CHP). In *Renewable Energy and Distributed Generation (ICREDG), 2017 Fifth Iranian Conference, 2017 Conference*. IEEE.
- Silverman, B., & Jones, M. (1989). E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). *International Statistical Review / Revue Internationale De Statistique*, 57(3), 233-238.
- Solar Cells – Chemistry Encyclopedia – structure, metal, equation, The pn Junction. Chemistryexplained.com. Consultado el 30 de octubre de 2017.
- SPV Market Research (2014), *Photovoltaic manufacturer Shipments: Capacity, Price & Revenues 2013/2013*, Report SPV-Supply2, April.
- Steel, R. G., & Torrie, J. (1960). *Principles And Procedures Of Statistics*. Mcgraw-Hill Book Company, Inc.,; New York; Toronto; London.
- Sugihara, H., Yokoyama, K., Saeki, O., Tsuji, K., & Funaki, T. (2013). Economic and efficient voltage management using customer-owned energy storage systems in a distribution network

with high penetration of photovoltaic systems. *IEEE Transactions on Power Systems*, 28(1), 102-111.

Velasco, J. G. (2009). *Energías renovables*. Reverte.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.

Wu, Z., & Xu, C. (2017, June). Integrated sea wave and off-shore photovoltaic energy assessment along the Sardinian coasts. In *OCEANS 2017-Aberdeen* (pp. 1-6). IEEE.

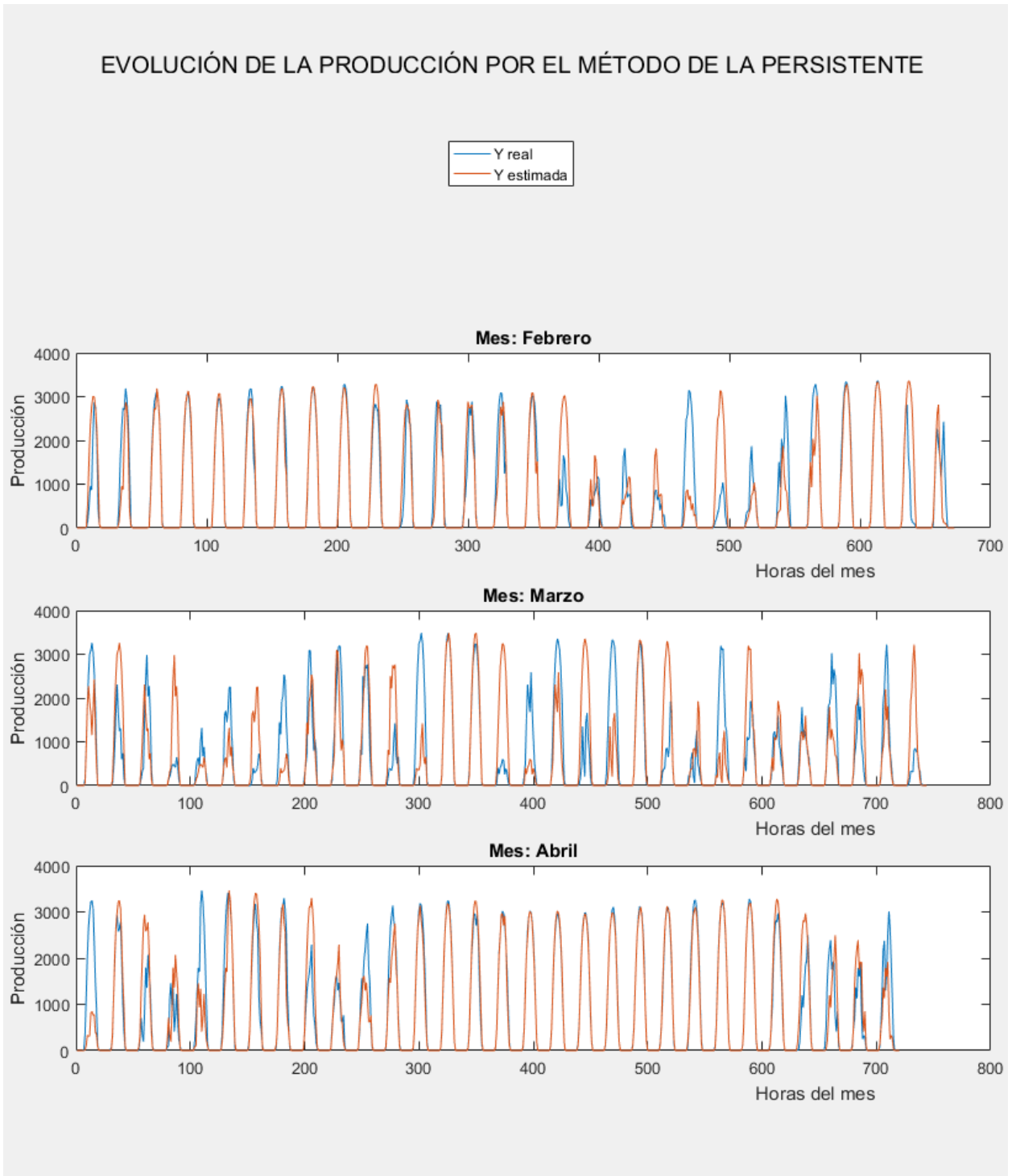
Yang, H., Luo, S., Sun, X., Zhang, Q., & Zhou, C. (2017, August). Research on photovoltaic energy storage system based on LADRC. In *Transportation Electrification Asia-Pacific (ITEC Asia-Pacific), 2017 IEEE Conference and Exp* (pp. 1-6). IEEE.

Anexo I. Índices de errores mensuales

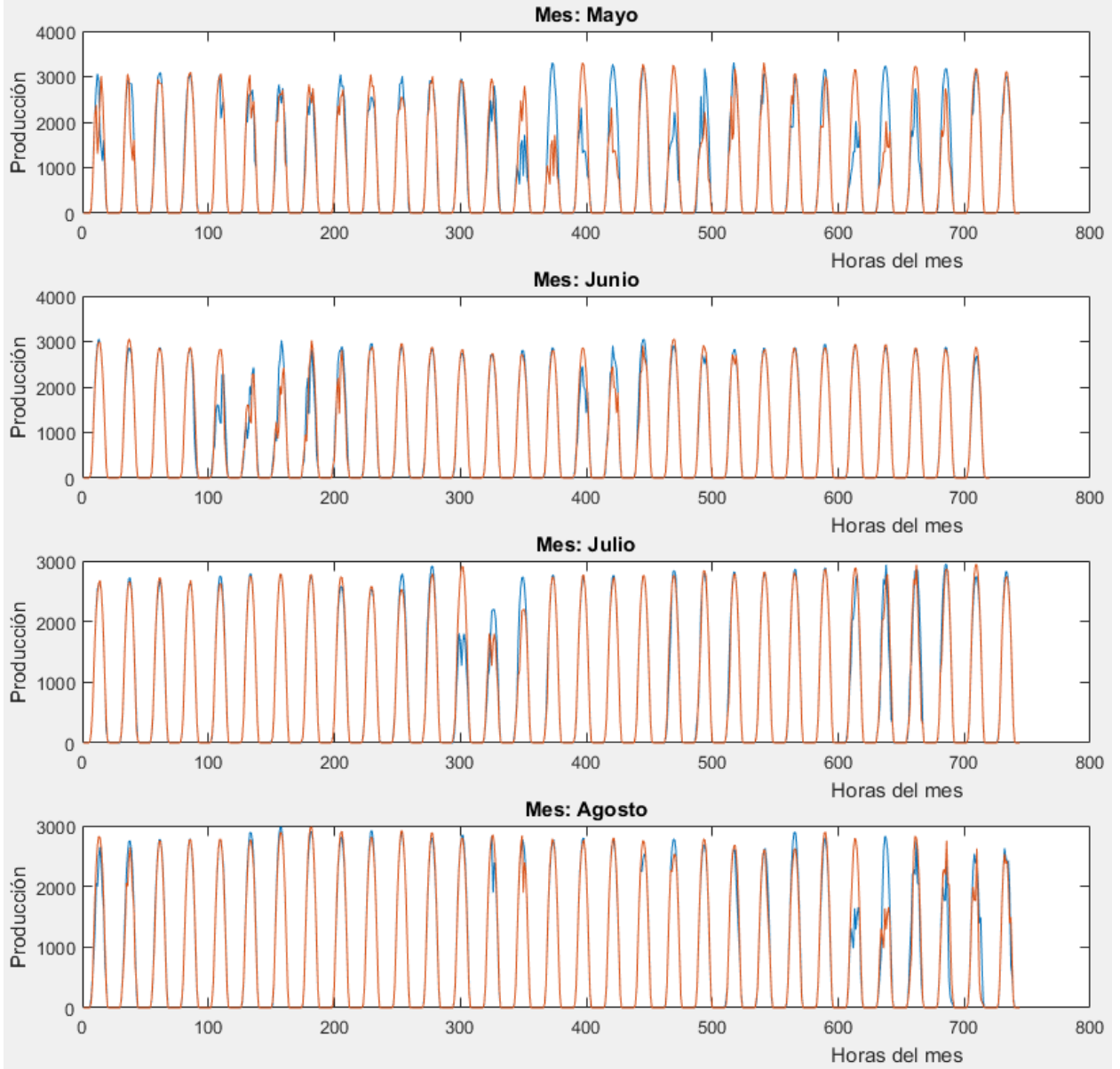
	Mín. de RSS	Mín. de RSE	Máx. de R ²	Mín. de MSE	Mín. de MAE
Febrero	39.306.280,10	242,21	94,77%	58.491,49	77,78
Lineal Local	39.306.280,10	242,21	94,77%	58.491,49	77,78
Kernel	57.732.057,28	293,54	92,31%	85.910,80	94,51
KNN	70.389.961,88	324,13	90,63%	104.746,97	110,01
Polinómica	143.224.129,88	462,35	80,93%	213.131,15	189,96
Lineal	147.204.714,93	468,73	80,40%	219.054,64	212,49
Persistente	188.007.060,84	529,72	74,97%	279.772,41	214,85
Marzo	49.311.288,49	257,79	92,56%	66.278,61	100,25
Kernel	49.311.288,49	257,79	92,56%	66.278,61	110,01
Lineal Local	51.249.438,60	262,81	92,27%	68.883,65	100,25
KNN	64.117.414,03	293,96	90,33%	86.179,32	123,70
Polinómica	201.747.243,86	521,44	69,56%	271.165,65	257,78
Lineal	203.995.452,94	524,33	69,22%	274.187,44	251,77
Persistente	456.872.982,07	784,69	31,07%	614.076,59	393,67
Abril	35.189.202,79	221,38	96,07%	48.873,89	77,01
Lineal Local	35.189.202,79	221,38	96,07%	48.873,89	77,01
KNN	60.447.313,93	290,15	93,26%	83.954,60	102,60
Kernel	60.600.699,74	290,52	93,24%	84.167,64	103,94
Lineal	126.850.005,14	420,32	85,85%	176.180,56	197,81
Polinómica	157.230.810,20	467,96	82,46%	218.376,13	206,09
Persistente	174.592.906,28	493,12	80,53%	242.490,15	200,97
Mayo	23.219.569,83	176,90	97,44%	31.209,10	65,12
Kernel	23.219.569,83	176,90	97,44%	31.209,10	66,21
Lineal Local	24.517.968,23	181,78	97,29%	32.954,26	65,12
KNN	38.169.976,76	226,81	95,79%	51.303,73	85,79
Lineal	79.772.186,78	327,89	91,19%	107.220,68	166,60
Polinómica	100.106.595,20	367,31	88,95%	134.551,88	174,95
Persistente	157.224.773,26	460,32	82,64%	211.323,62	203,61
Junio	9.422.229,07	114,56	98,89%	13.086,43	29,76
Lineal Local	9.422.229,07	114,56	98,89%	13.086,43	29,76
KNN	11.524.870,17	126,69	98,65%	16.006,76	36,57
Kernel	13.580.889,05	137,53	98,41%	18.862,35	36,25
Persistente	31.293.912,51	208,77	96,33%	43.463,77	77,13
Polinómica	39.713.733,17	235,18	95,34%	55.157,96	95,35
Lineal	43.077.684,56	244,94	94,94%	59.830,12	94,09

Julio	3.225.682,91	65,93	99,61%	4.335,60	17,29
Lineal Local	3.225.682,91	65,93	99,61%	4.335,60	17,29
Kernel	7.878.524,35	103,04	99,05%	10.589,41	26,95
KNN	8.341.487,41	106,03	99,00%	11.211,68	31,84
Lineal	17.000.049,50	151,36	97,96%	22.849,53	65,64
Persistente	22.217.780,22	173,04	97,33%	29.862,61	62,69
Polinómica	33.676.822,02	213,04	95,96%	45.264,55	70,56
Agosto	13.540.793,81	135,09	98,36%	18.199,99	24,75
Lineal Local	13.540.793,81	135,09	98,36%	18.199,99	24,75
KNN	14.305.531,53	138,85	98,26%	19.227,86	36,42
Kernel	16.058.777,42	147,11	98,05%	21.584,38	39,53
Persistente	36.015.329,62	220,31	95,63%	48.407,70	80,61
Lineal	43.368.871,27	241,76	94,74%	58.291,49	82,31
Polinómica	57.468.697,83	278,30	93,03%	77.242,87	95,63
Septiembre	15.826.598,40	148,47	97,62%	21.981,39	52,00
Lineal Local	15.826.598,40	148,47	97,62%	21.981,39	52,00
KNN	35.498.683,86	222,35	94,66%	49.303,73	81,62
Kernel	42.667.244,39	243,77	93,59%	59.260,06	87,07
Lineal	54.065.006,79	274,41	91,87%	75.090,29	132,64
Polinómica	79.059.377,21	331,83	88,11%	109.804,69	149,89
Persistente	85.459.448,46	345,00	87,15%	118.693,68	147,11
Octubre	25.719.871,42	186,18	96,28%	34.569,72	66,54
Lineal Local	25.719.871,42	186,18	96,28%	34.569,72	66,54
KNN	34.623.890,45	216,02	94,99%	46.537,49	85,38
Kernel	38.699.319,17	228,38	94,40%	52.015,21	88,10
Lineal	77.087.915,37	322,32	88,84%	103.612,79	164,74
Polinómica	94.762.200,21	357,37	86,28%	127.368,55	179,21
Persistente	144.125.313,61	440,73	79,13%	193.716,82	205,66
Noviembre	42.653.309,44	243,73	93,91%	59.240,71	69,03
Lineal Local	42.653.309,44	243,73	93,91%	59.240,71	69,03
Kernel	47.855.862,45	258,17	93,17%	66.466,48	85,22
KNN	68.669.665,94	309,26	90,20%	95.374,54	99,14
Lineal	120.052.296,42	408,91	82,87%	166.739,30	178,42
Polinómica	139.288.253,79	440,45	80,12%	193.455,91	190,63
Persistente	149.149.177,94	455,77	78,72%	207.151,64	167,48
Diciembre	12.342.798,90	128,97	98,02%	16.589,78	38,16
Lineal Local	12.342.798,90	128,97	98,02%	16.589,78	38,16
KNN	25.414.856,40	185,07	95,92%	34.159,75	54,05
Kernel	30.687.242,19	203,37	95,07%	41.246,29	64,00
Lineal	56.856.942,42	276,82	90,87%	76.420,62	118,65
Polinómica	75.890.100,28	319,81	87,82%	102.002,82	129,15
Persistente	165.103.976,95	471,71	73,50%	221.913,95	173,75

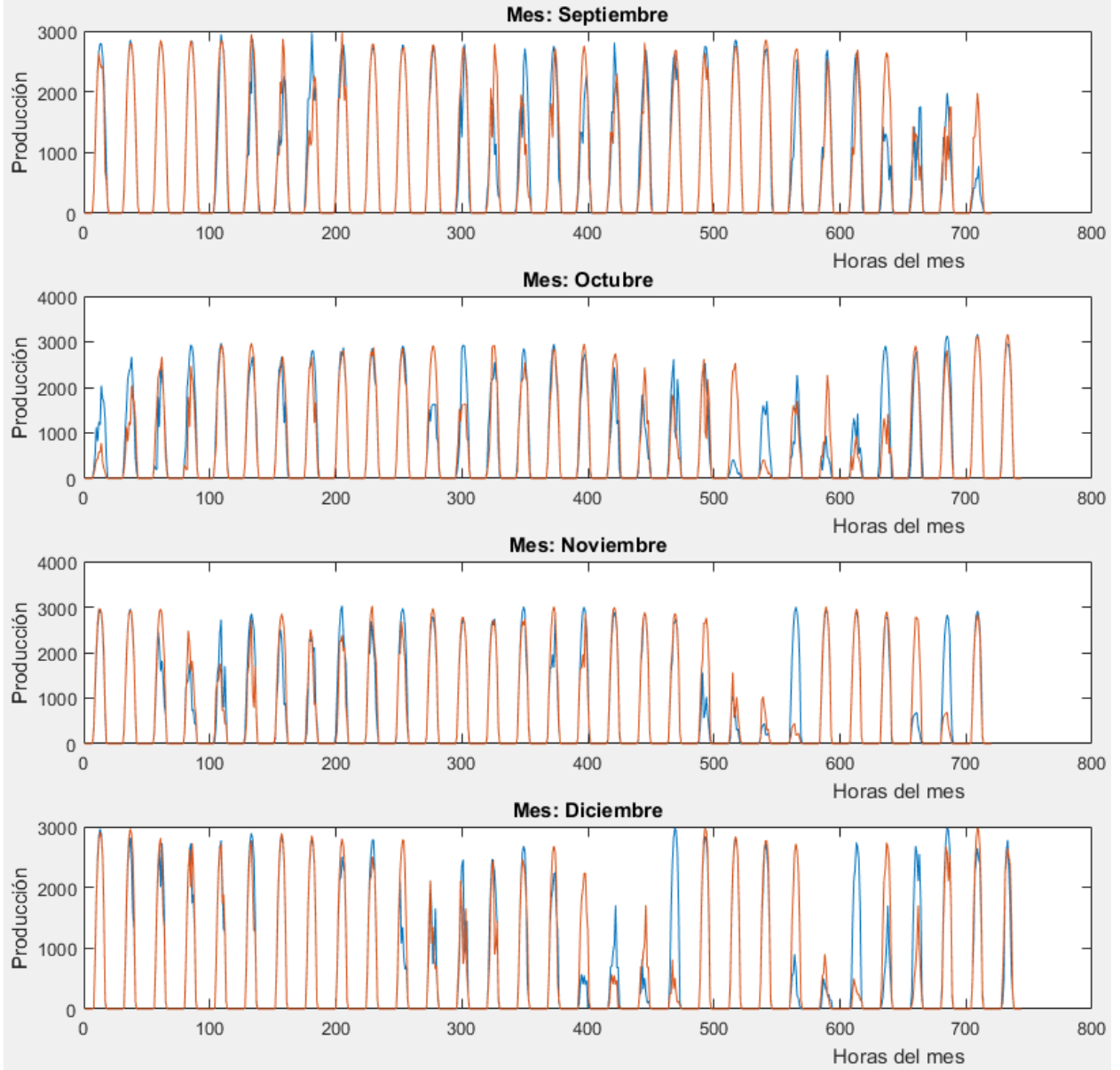
Anexo II. Evolución de la producción por el método de la persistente



EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LA PERSISTENTE

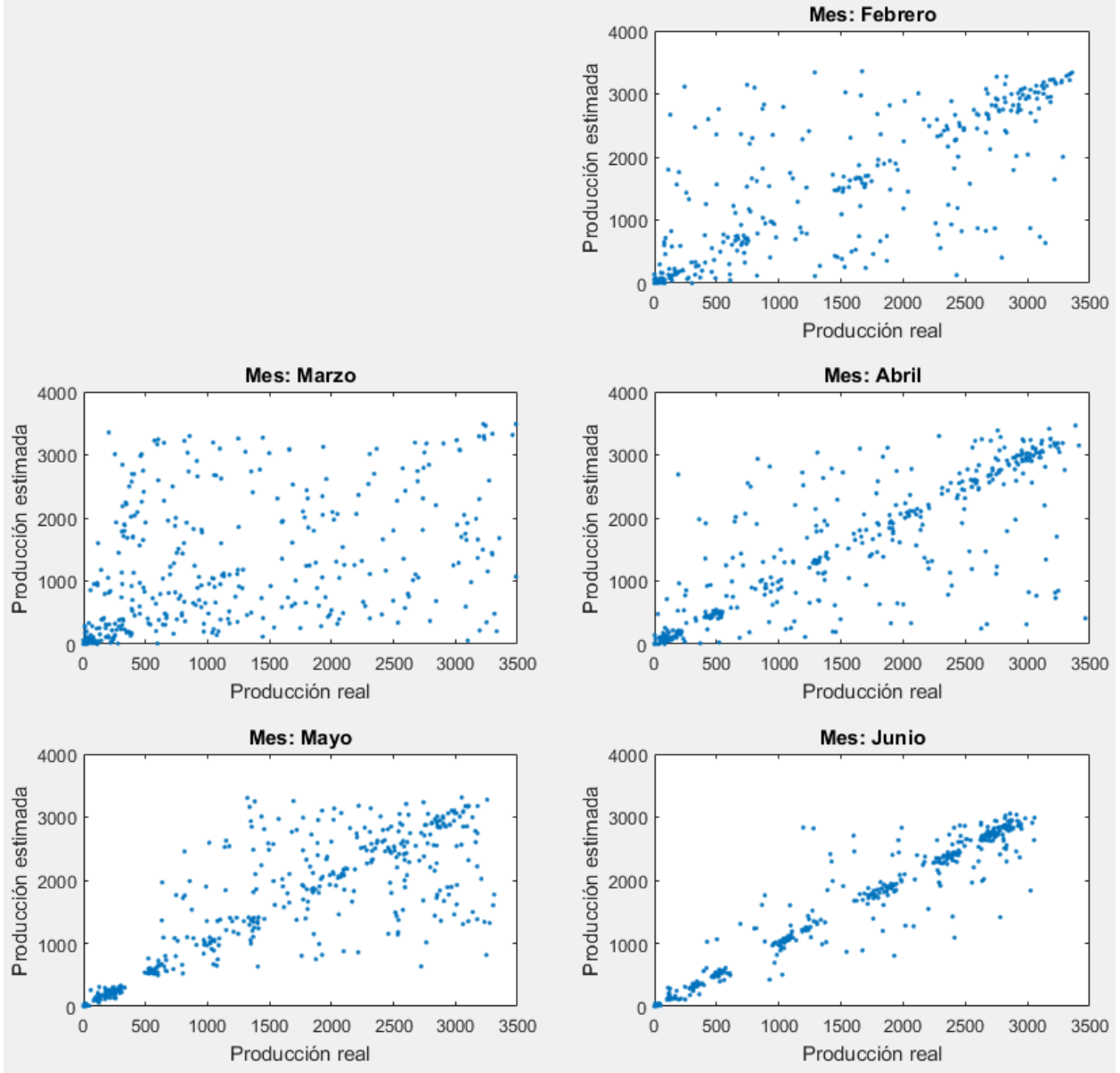


EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LA PERSISTENTE

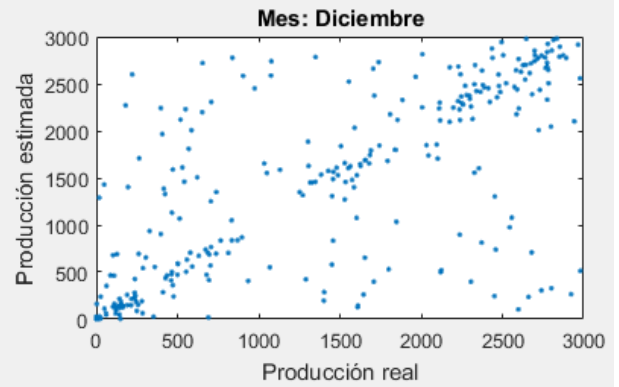
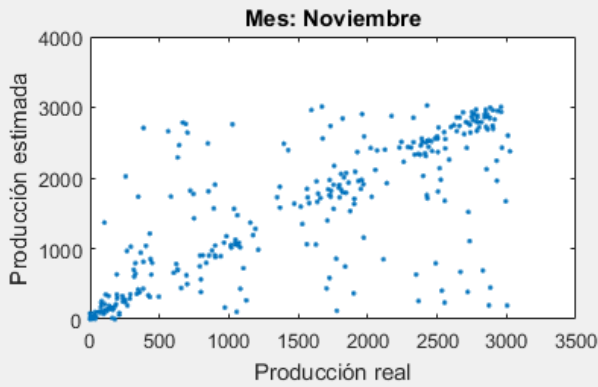
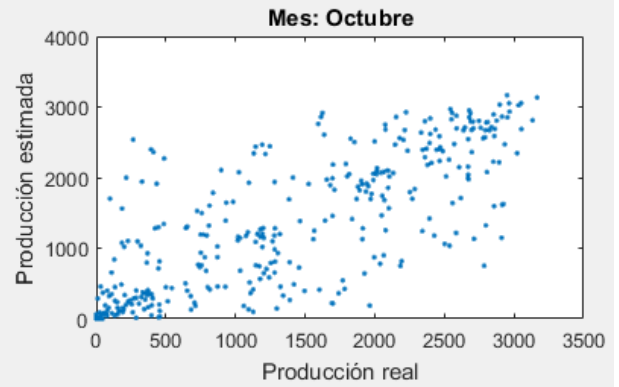
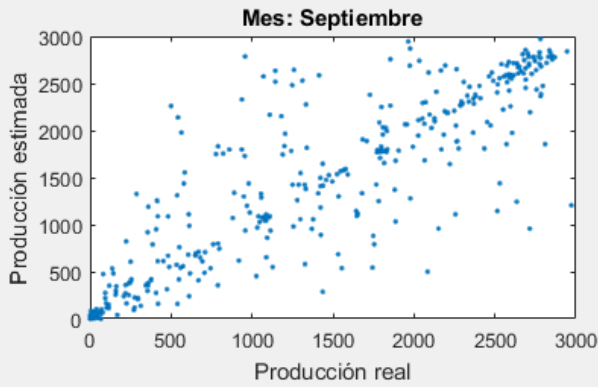
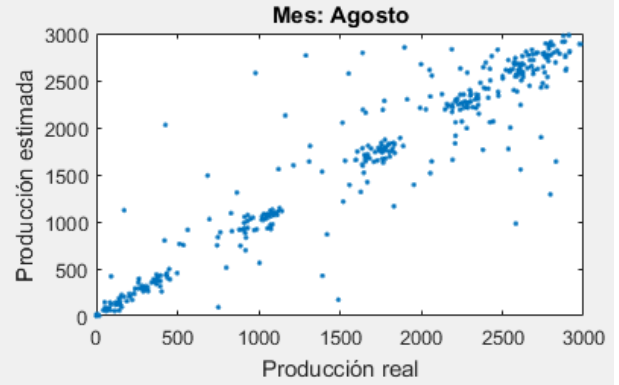
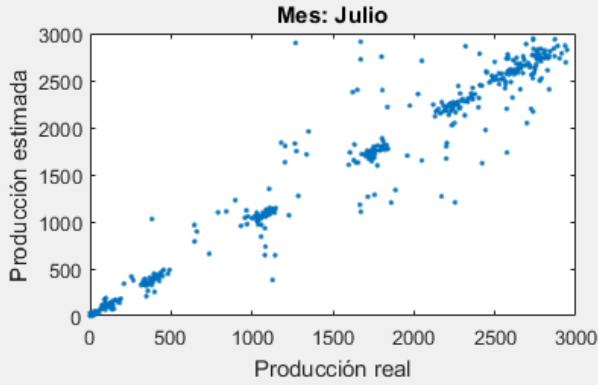


Anexo III. Producción real vs Producción estimada. Método de la persistente

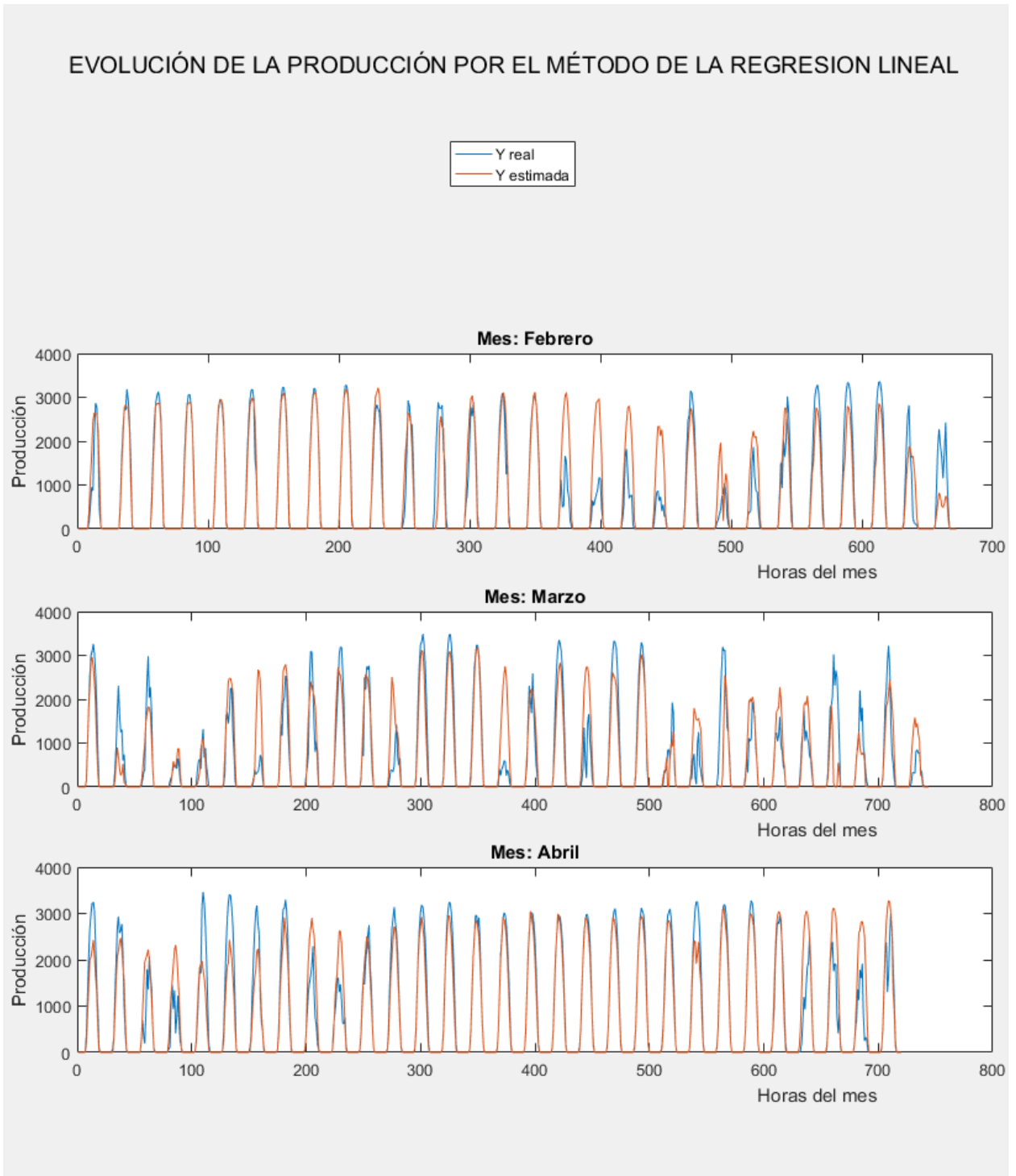
Y REAL vs Y ESTIMADA. MÉTODO DE LA PERSISTENTE



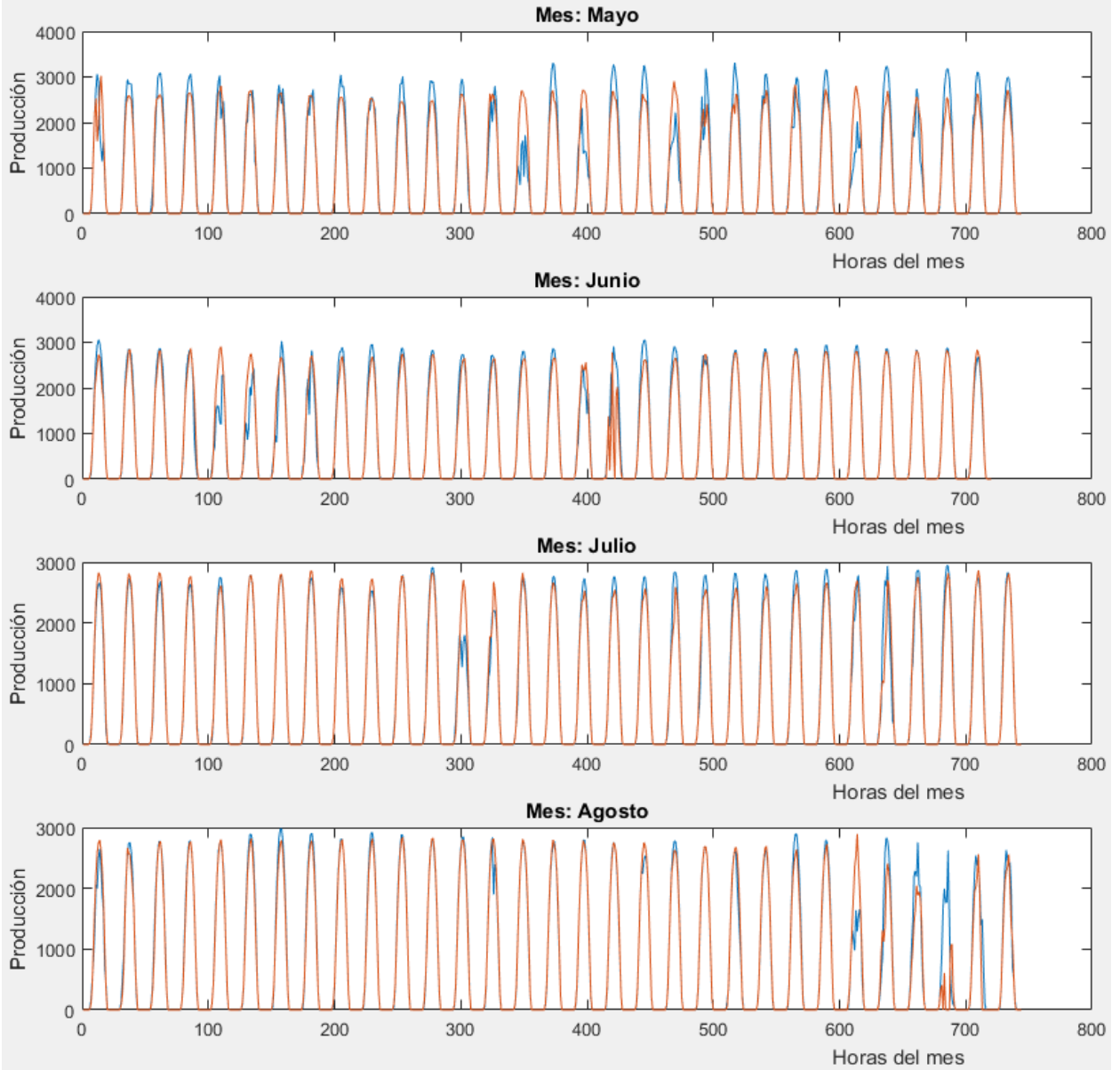
Y REAL vs Y ESTIMADA. MÉTODO DE LA PERSISTENTE



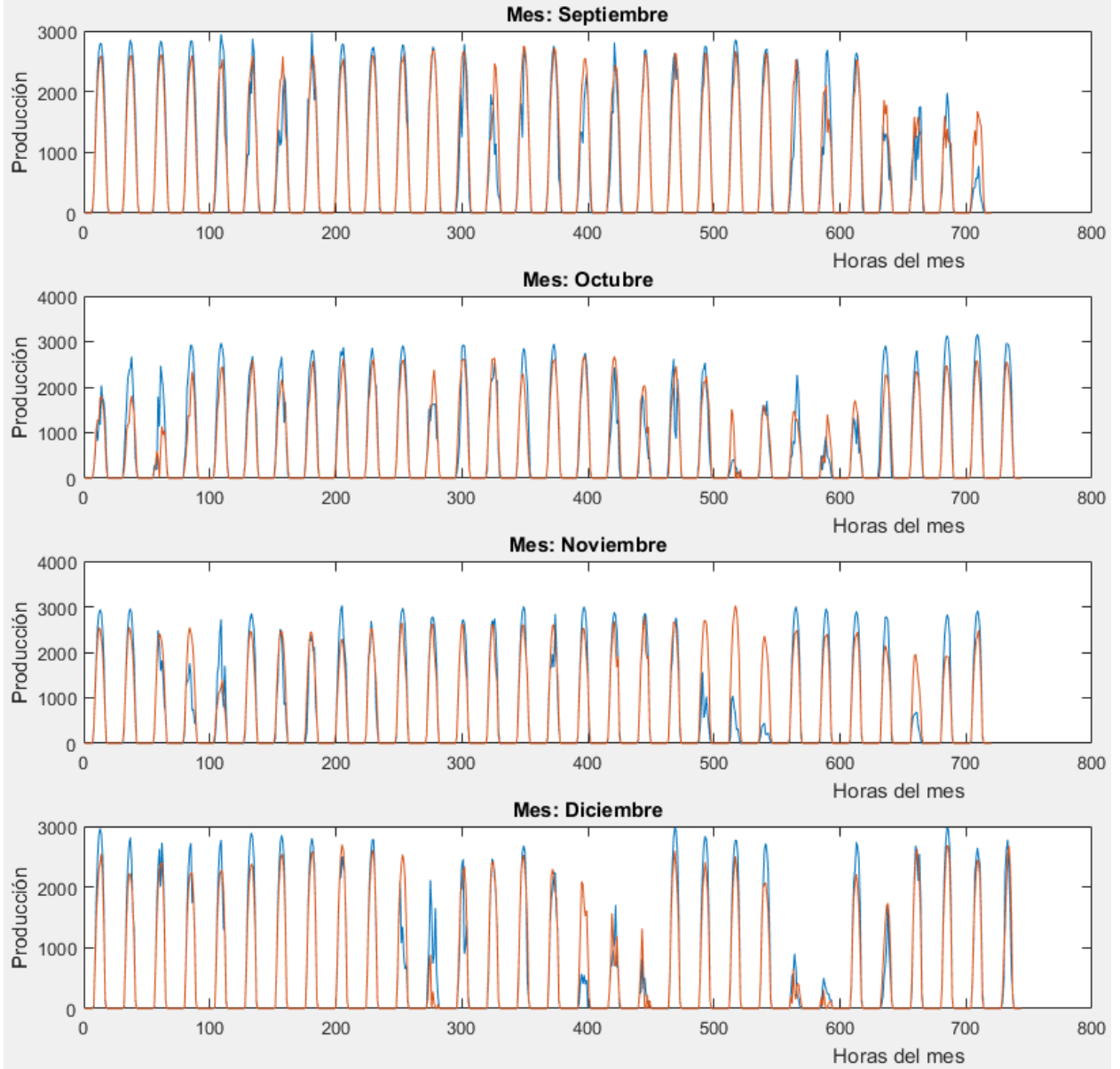
Anexo IV. Evolución de la producción por el método de la regresión lineal



EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LA REGRESION LINEAL

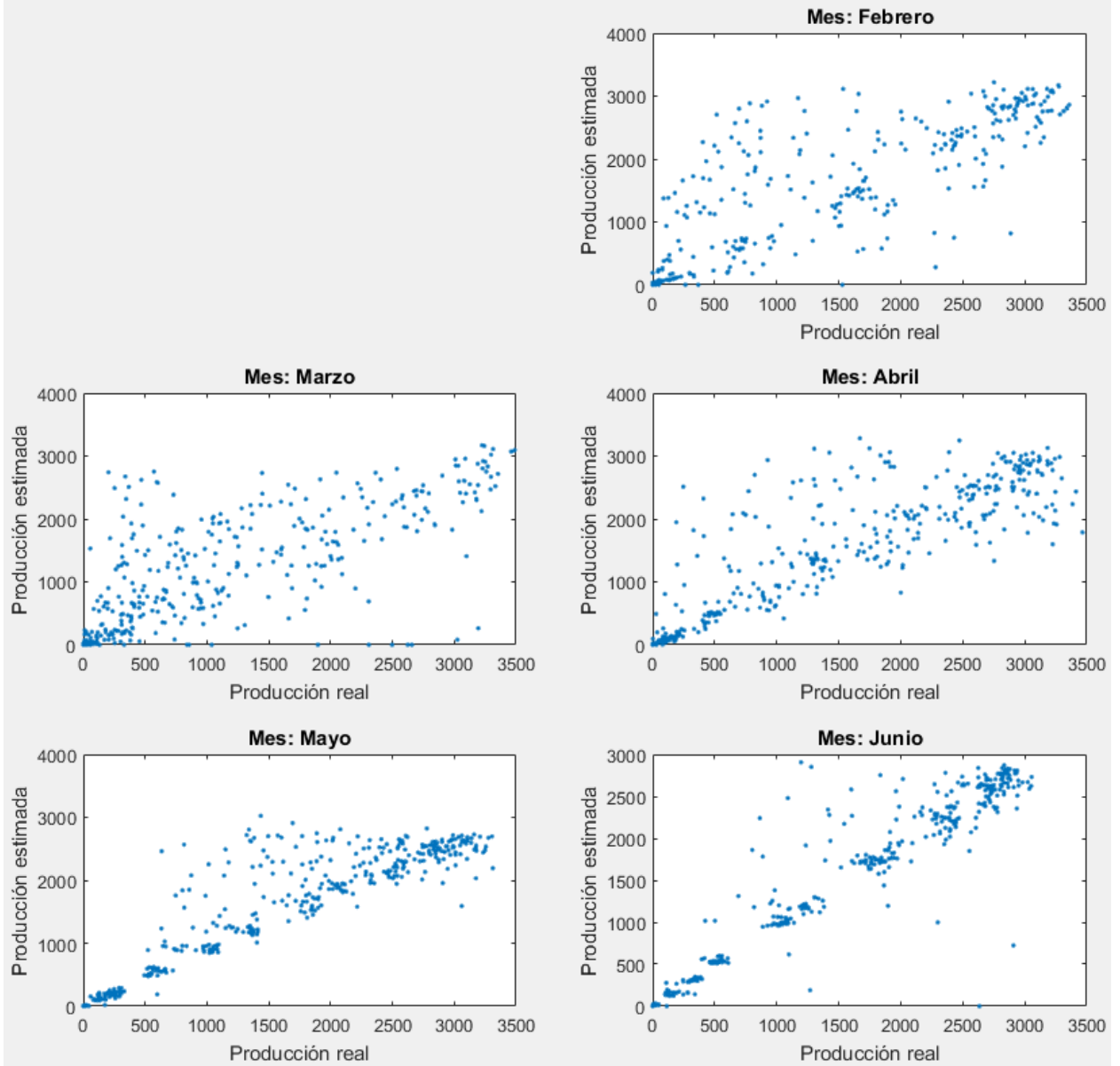


EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LA REGRESION LINEAL

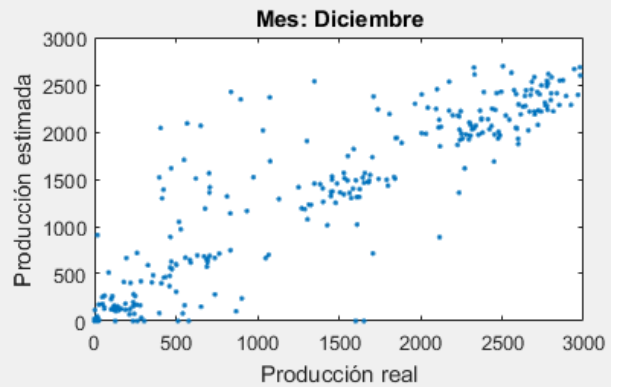
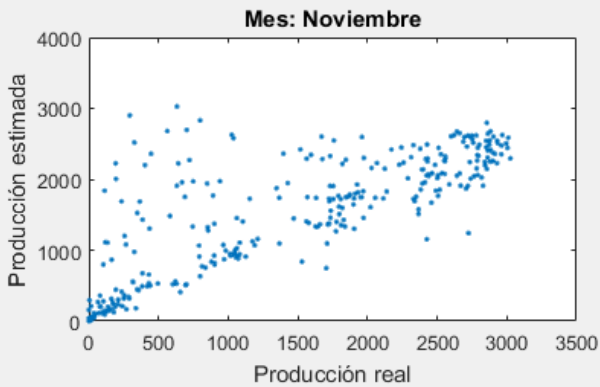
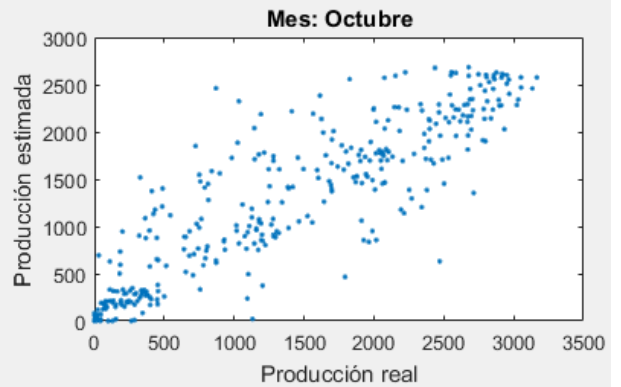
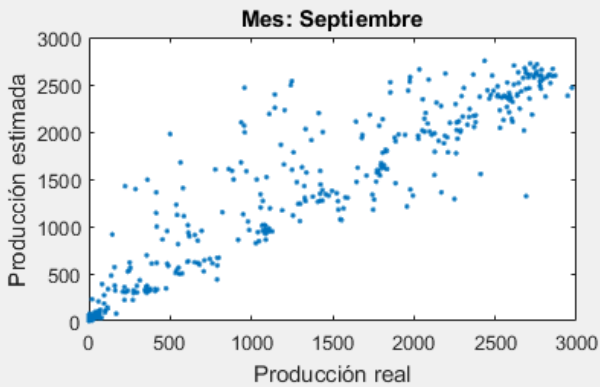
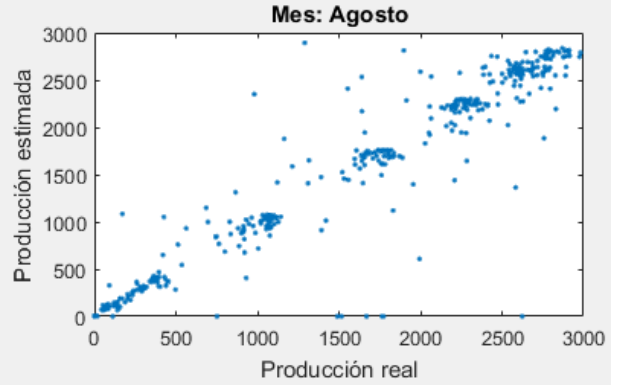
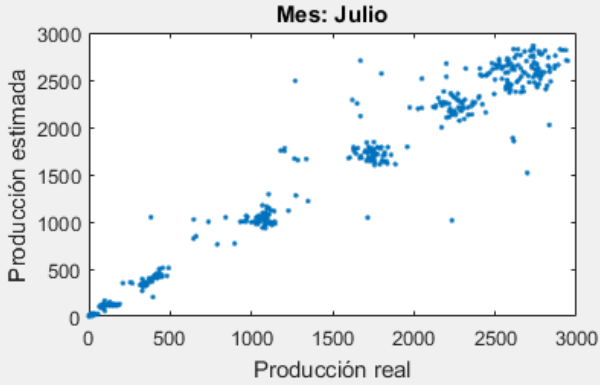


Anexo V. Producción real vs Producción estimada. Método de la regresión lineal

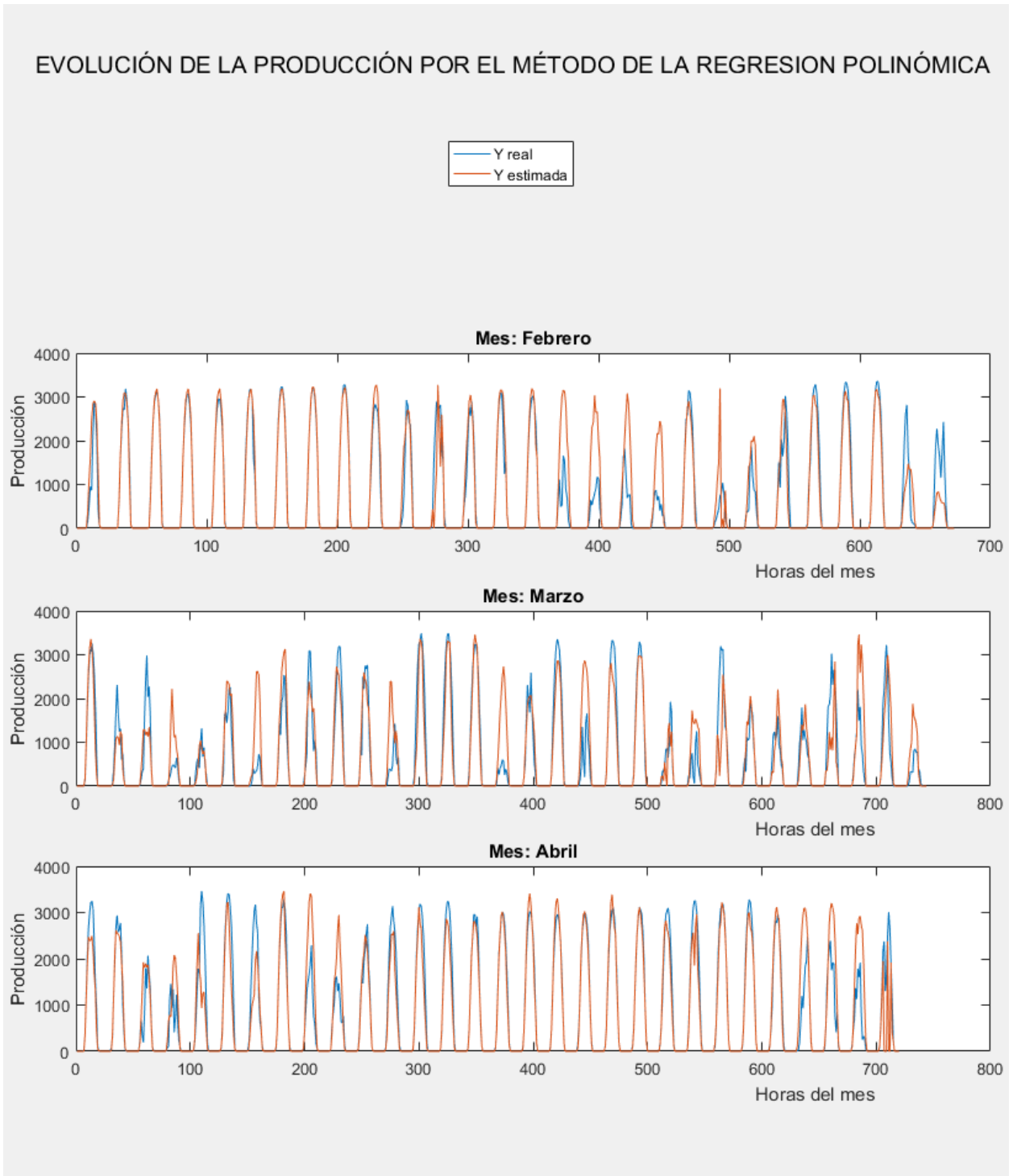
Y REAL vs Y ESTIMADA. MÉTODO DE LA REGRESIÓN LINEAL



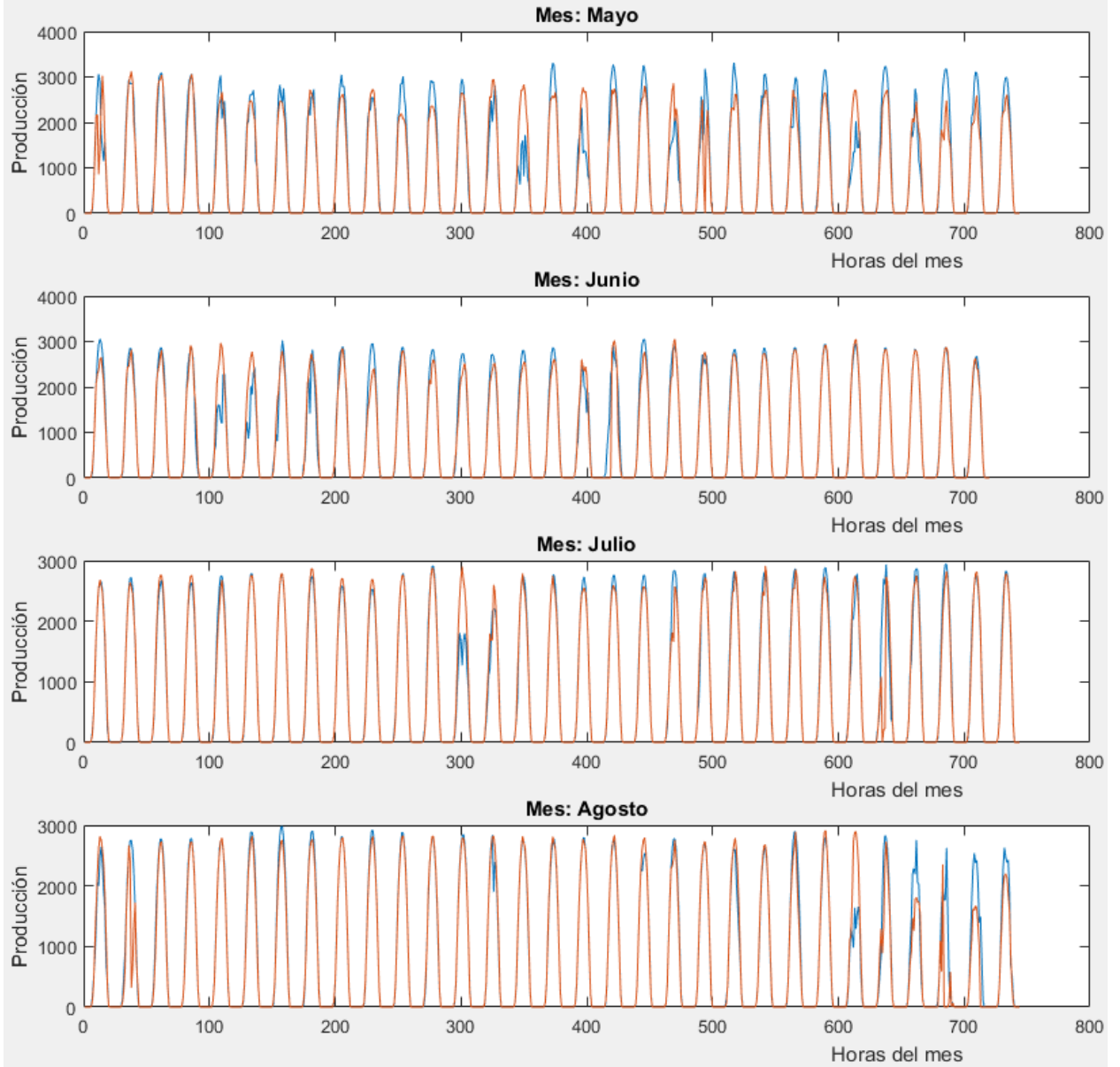
Y REAL vs Y ESTIMADA. MÉTODO DE LA REGRESIÓN LINEAL



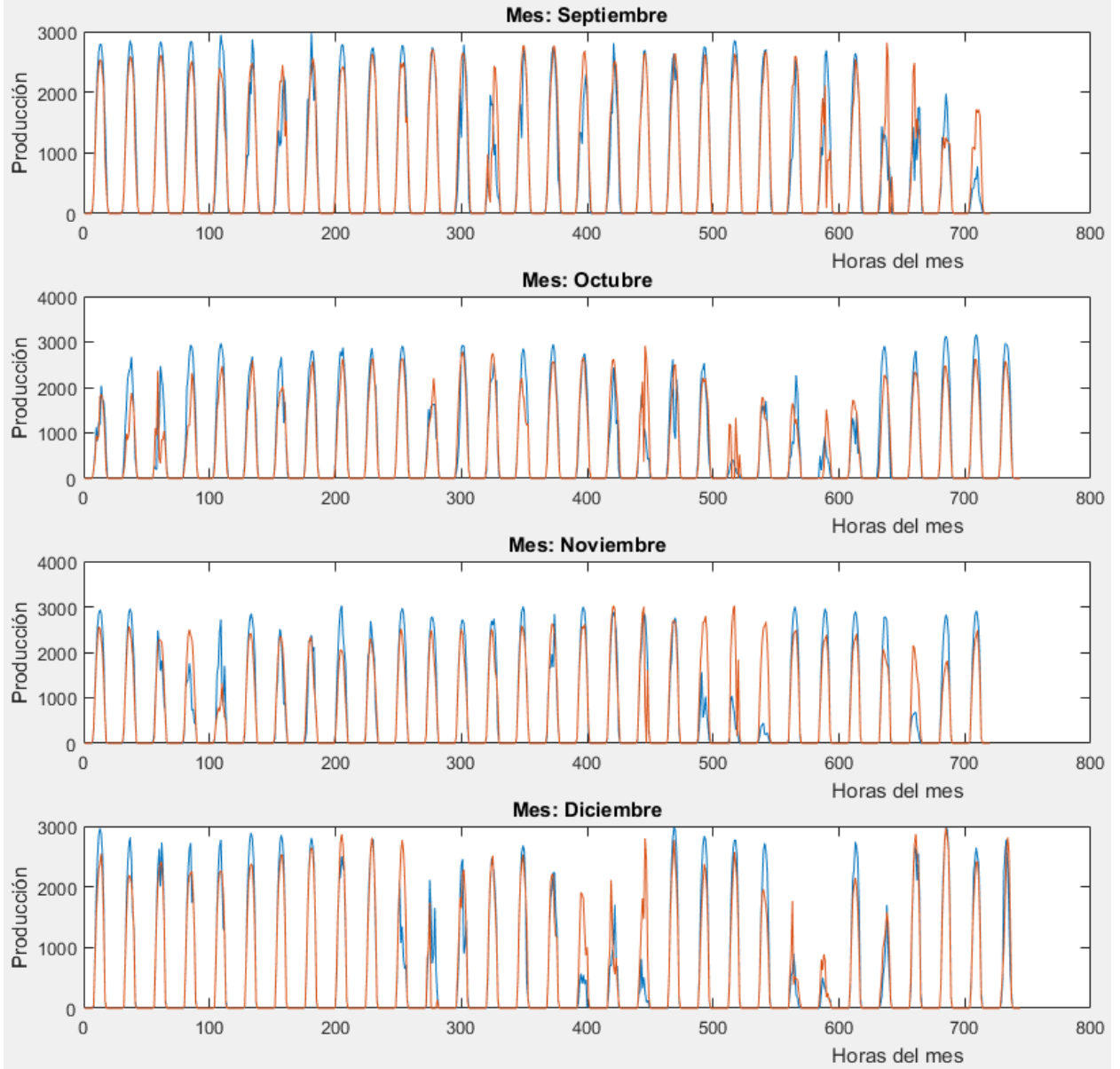
Anexo VI. Evolución de la producción por el método de la regresión polinómica



EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LA REGRESION POLINÓMICA

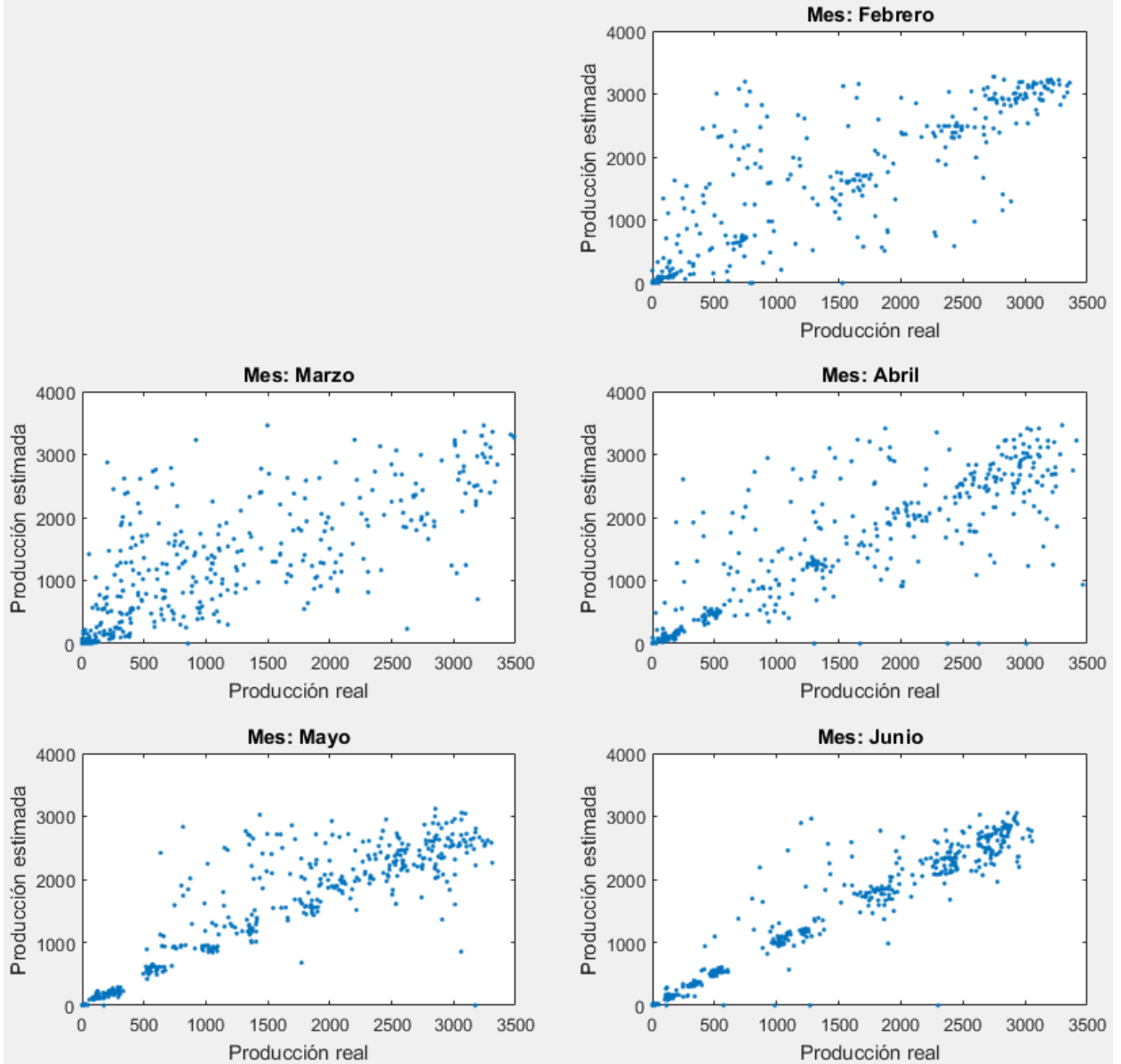


EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LA REGRESION POLINÓMICA

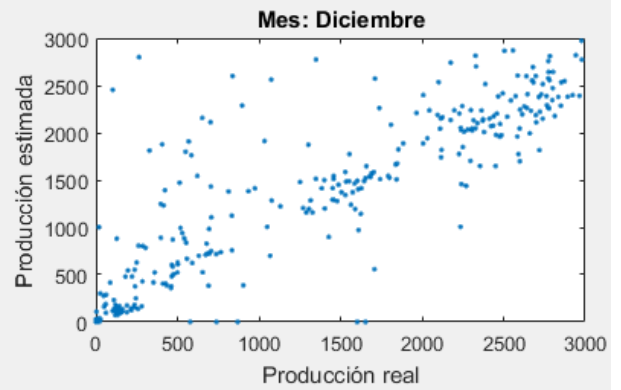
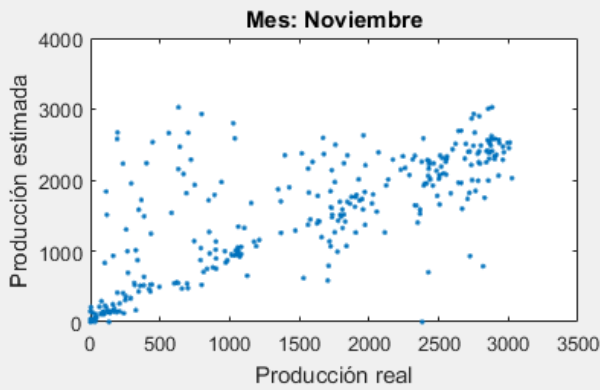
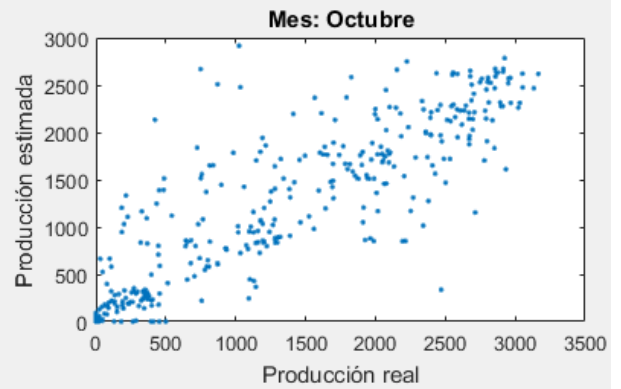
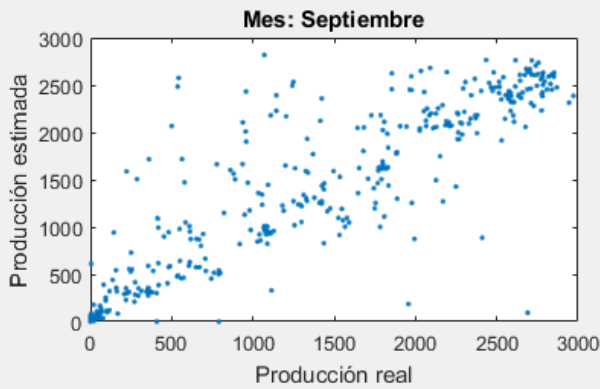
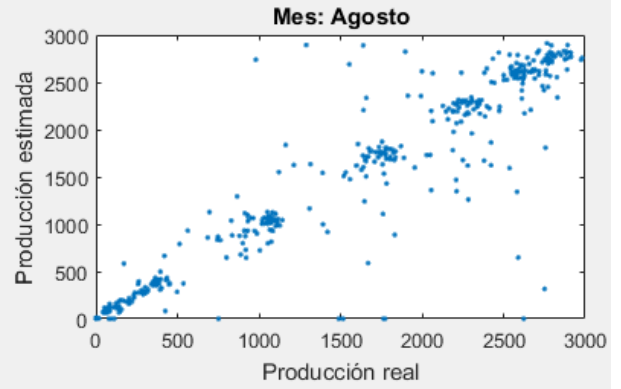
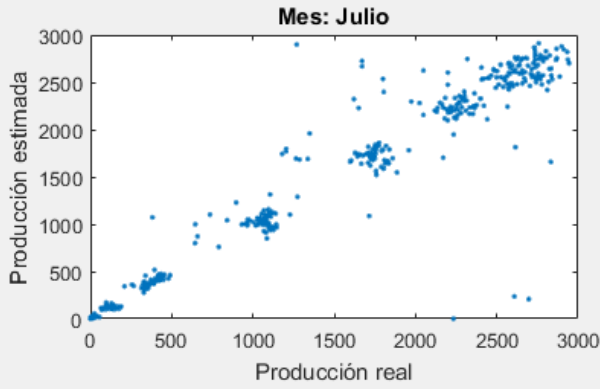


Anexo VII. Producción real vs Producción estimada. Método de la regresión polinómica

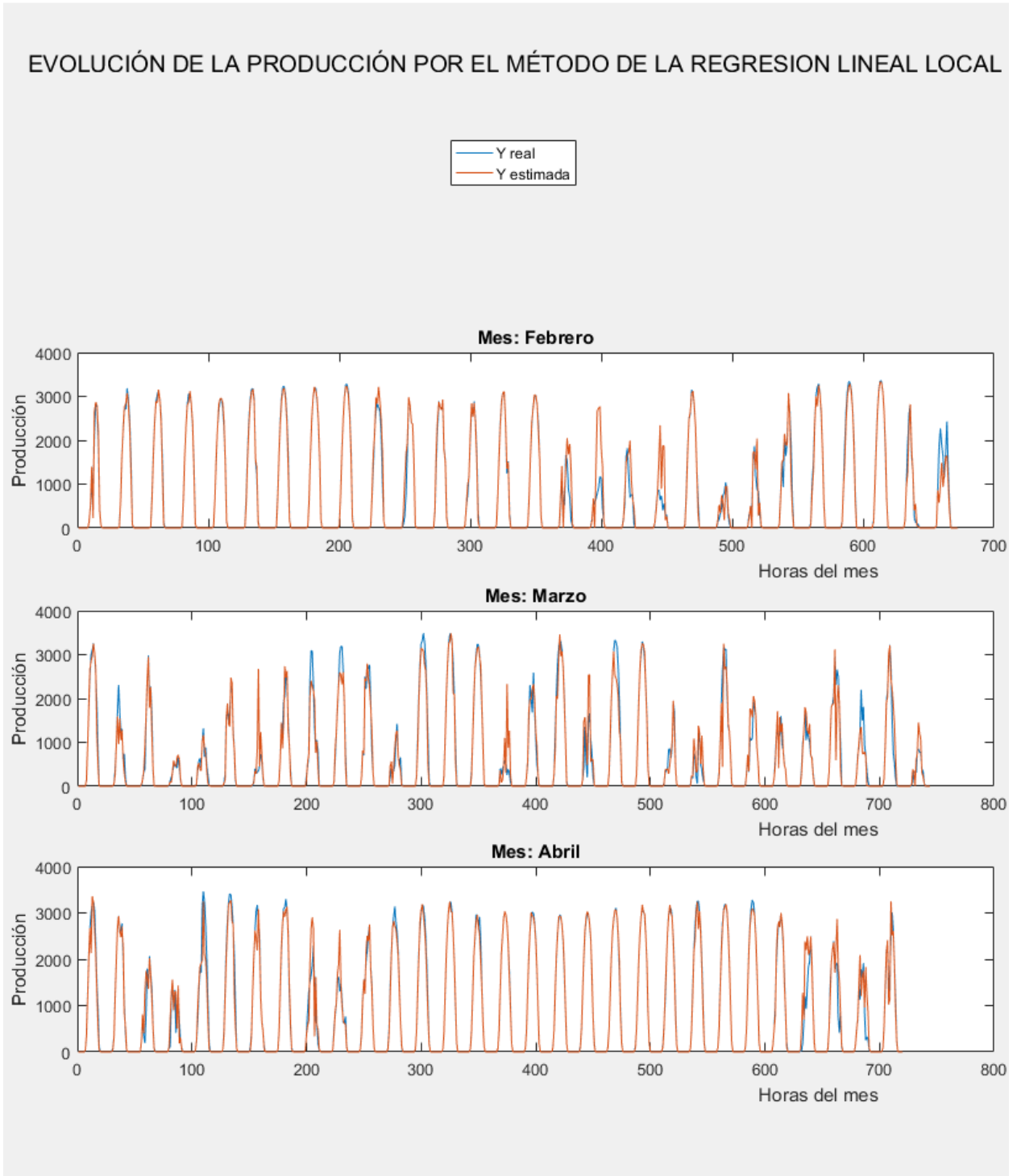
Y REAL vs Y ESTIMADA. MÉTODO DE LA REGRESIÓN POLINÓMICA



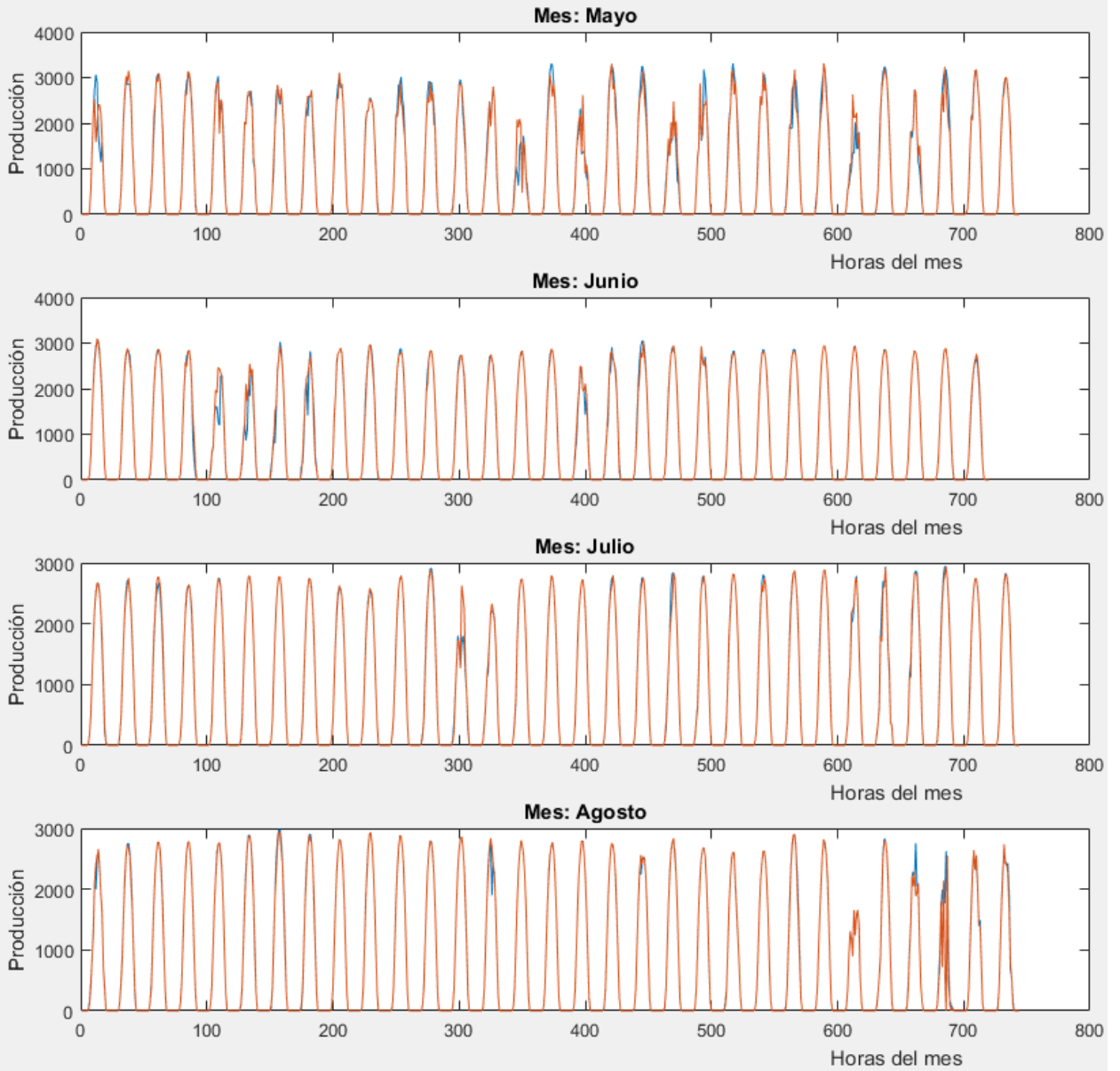
Y REAL vs Y ESTIMADA. MÉTODO DE LA REGRESIÓN POLINÓMICA



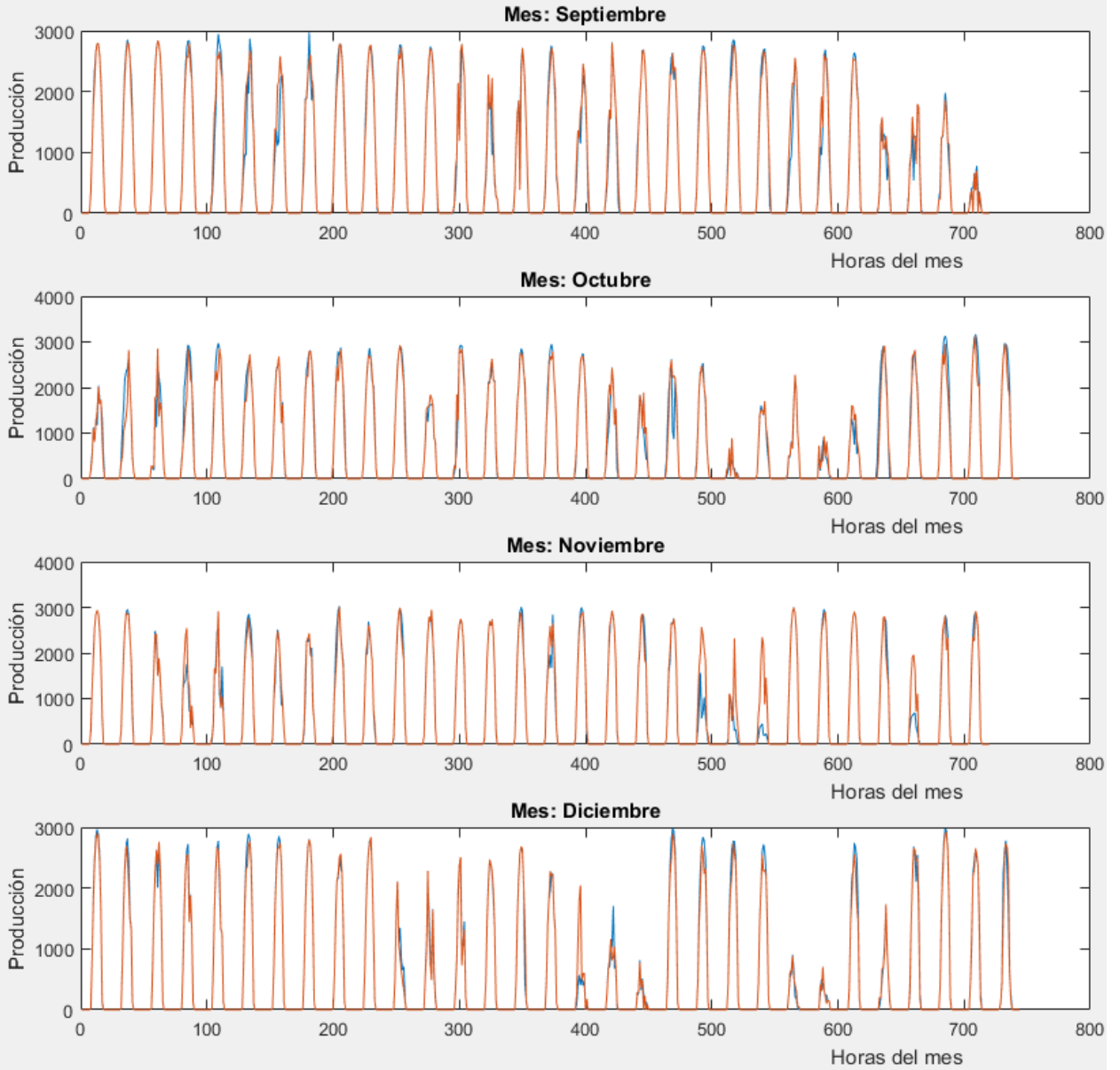
Anexo VIII. Evolución de la producción por el método de la regresión lineal local



EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LA REGRESION LINEAL LOCAL

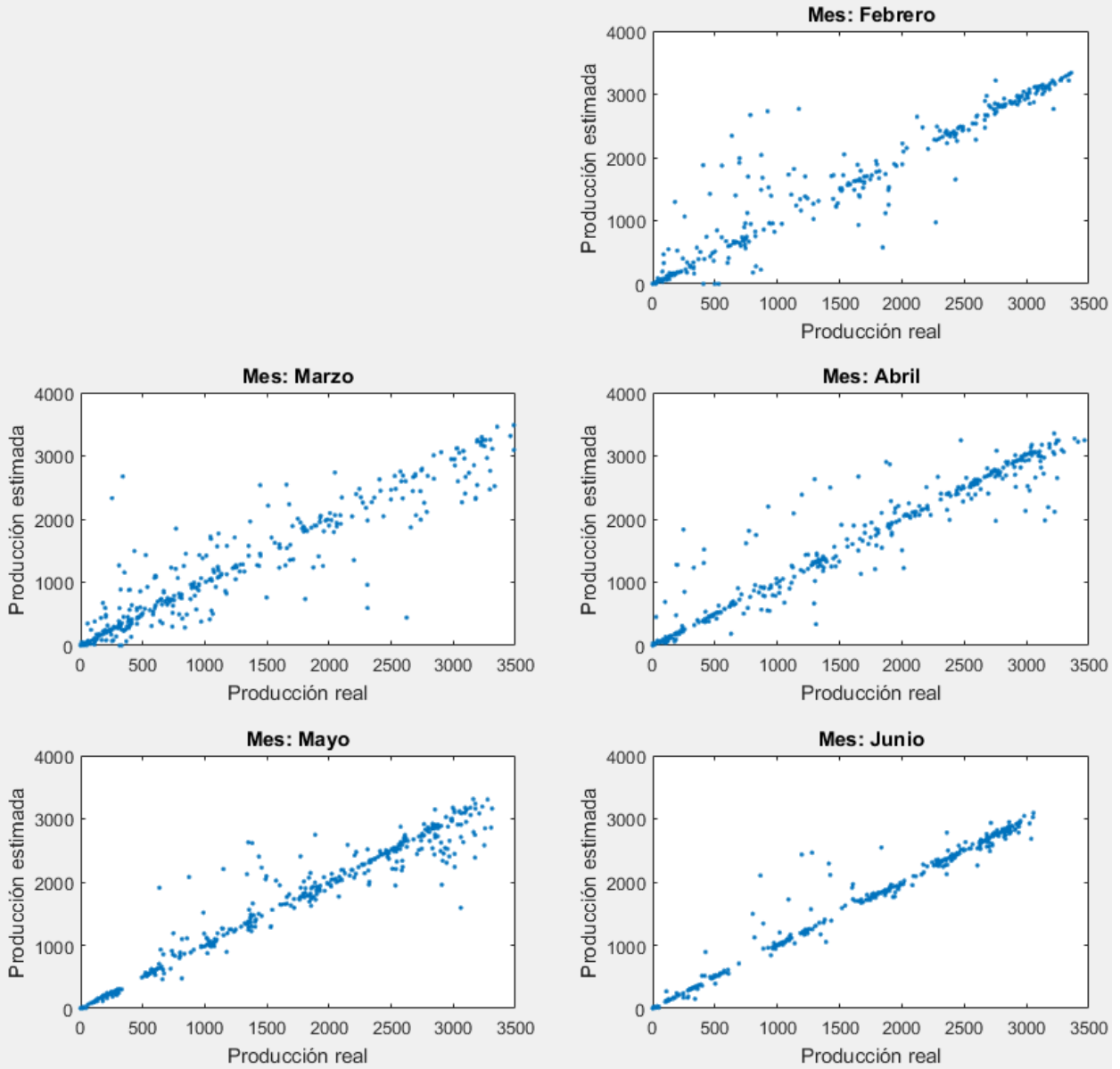


EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LA REGRESION LINEAL LOCAL

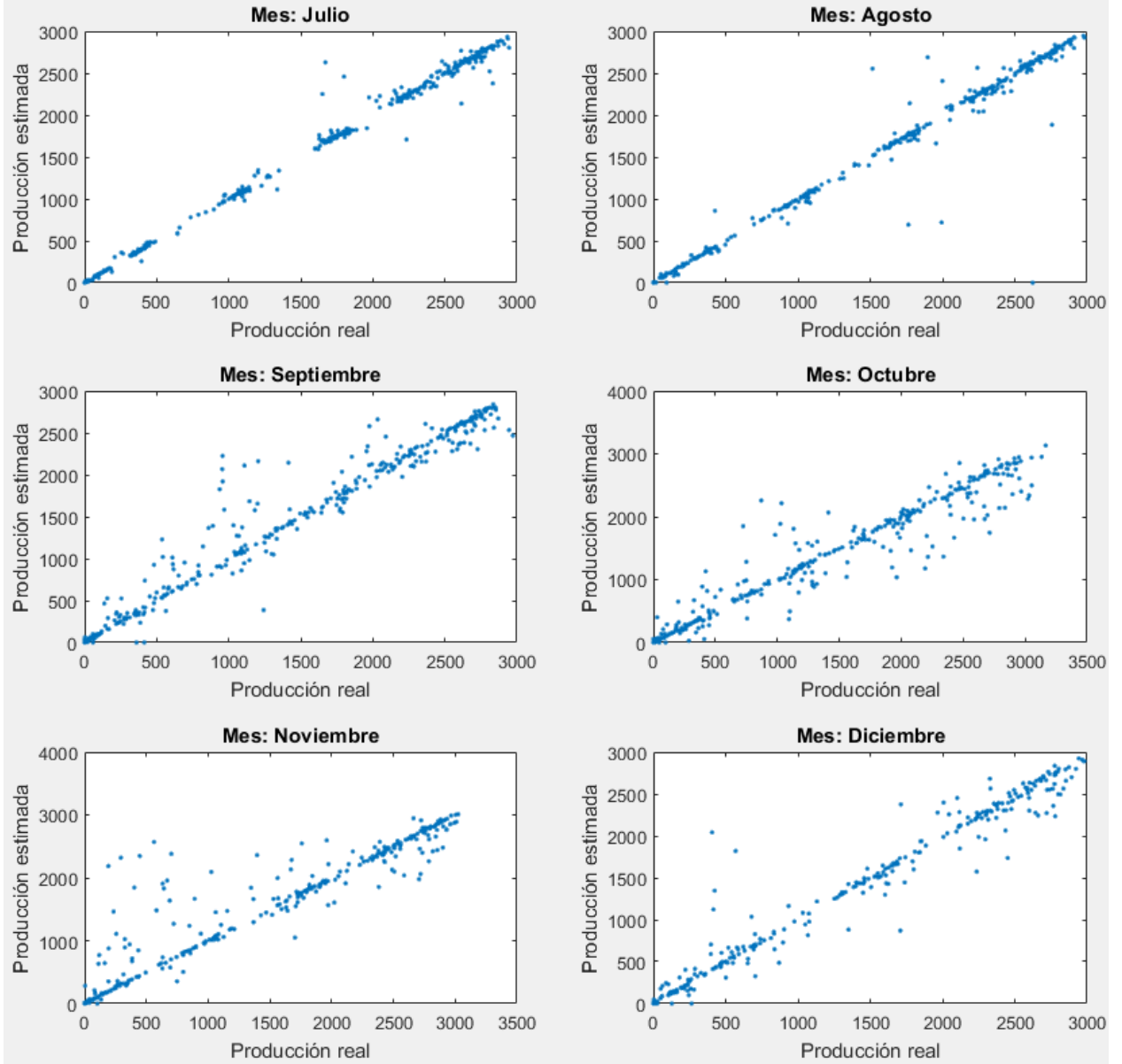


Anexo IX. Producción real vs Producción estimada. Método de la regresión lineal local

Y REAL vs Y ESTIMADA. MÉTODO DE LA REGRESIÓN LINEAL LOCAL

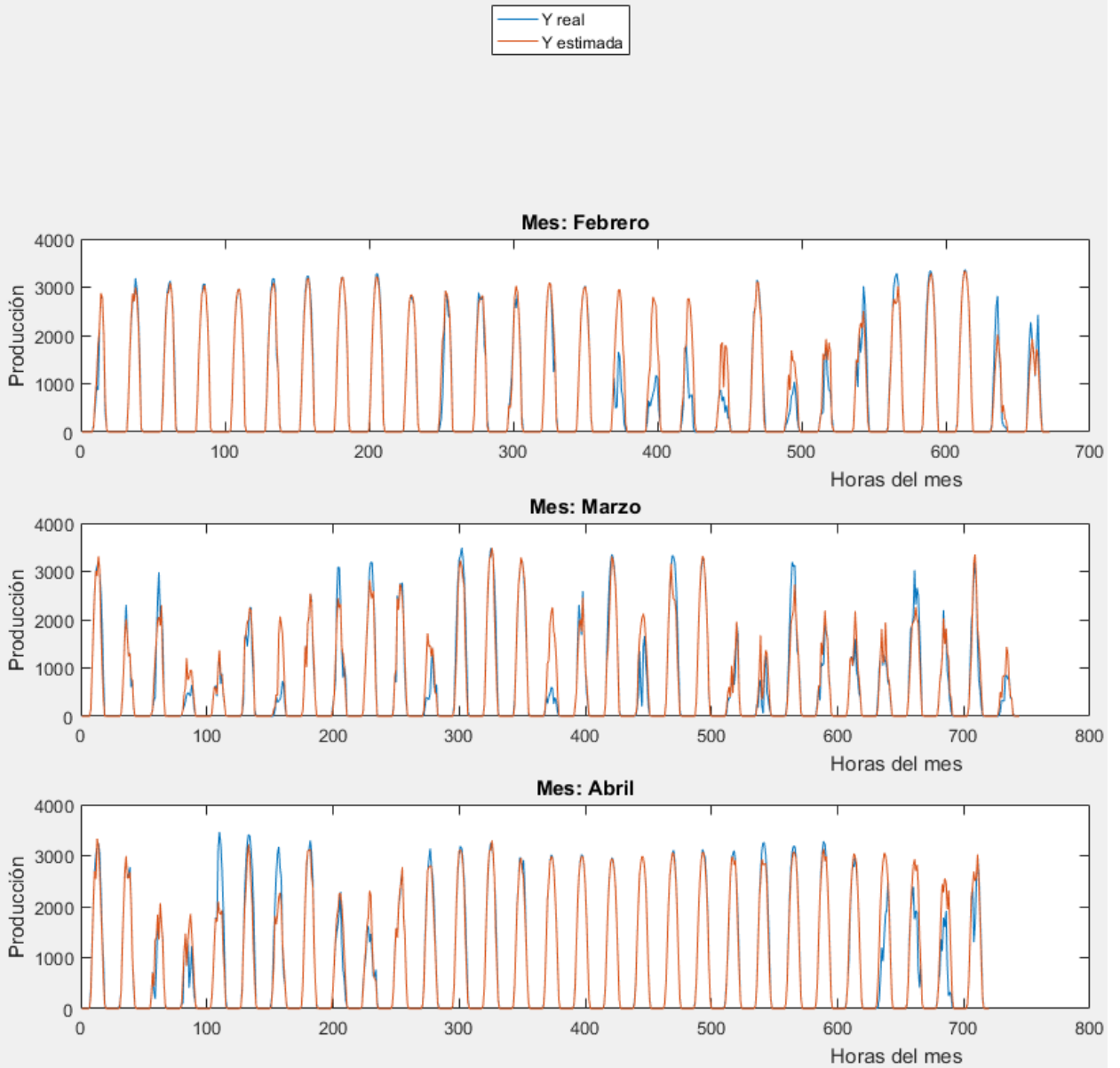


Y REAL vs Y ESTIMADA. MÉTODO DE LA REGRESIÓN LINEAL LOCAL

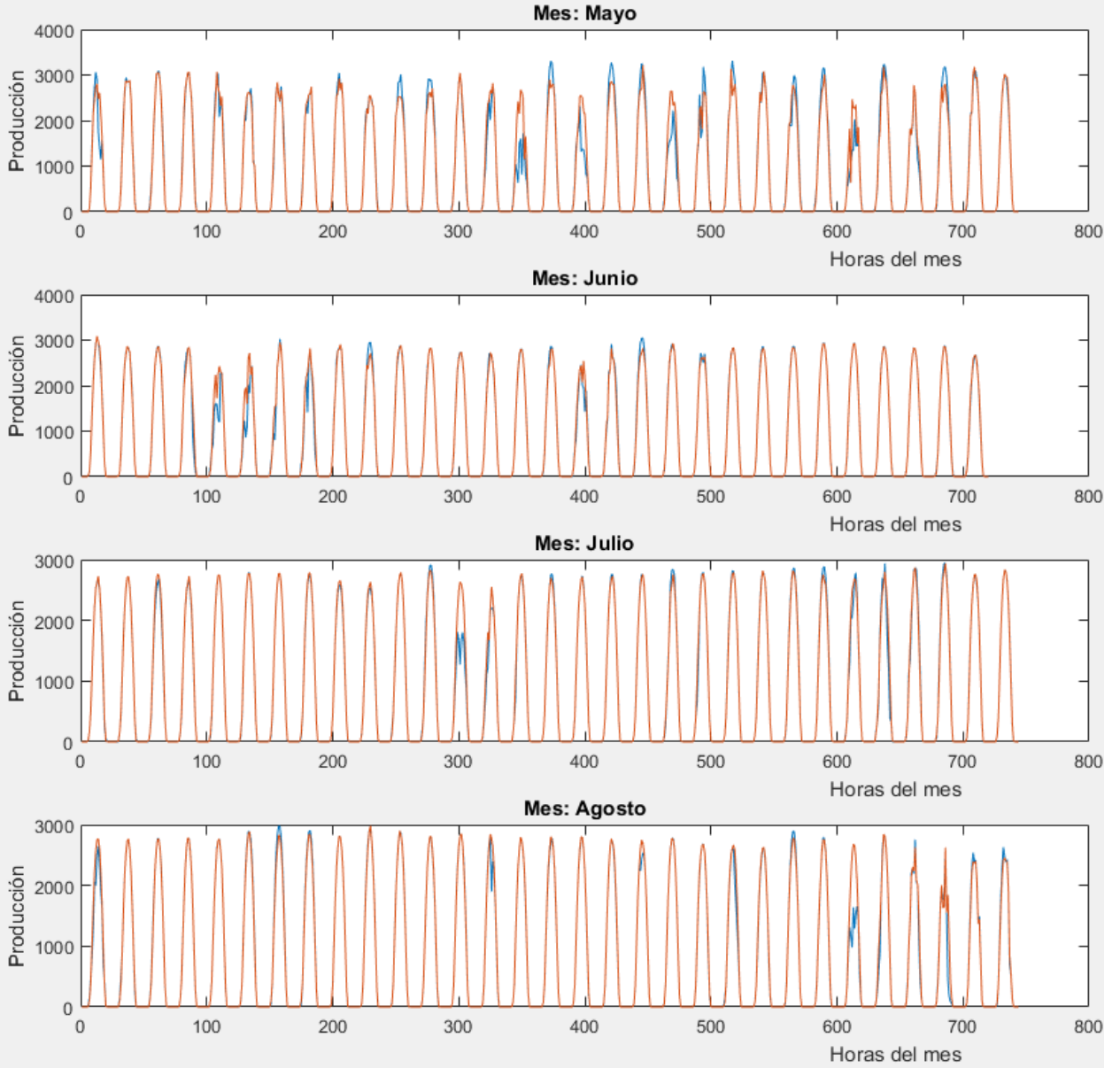


Anexo X. Evolución de la producción por el método de los k-vecinos más cercanos

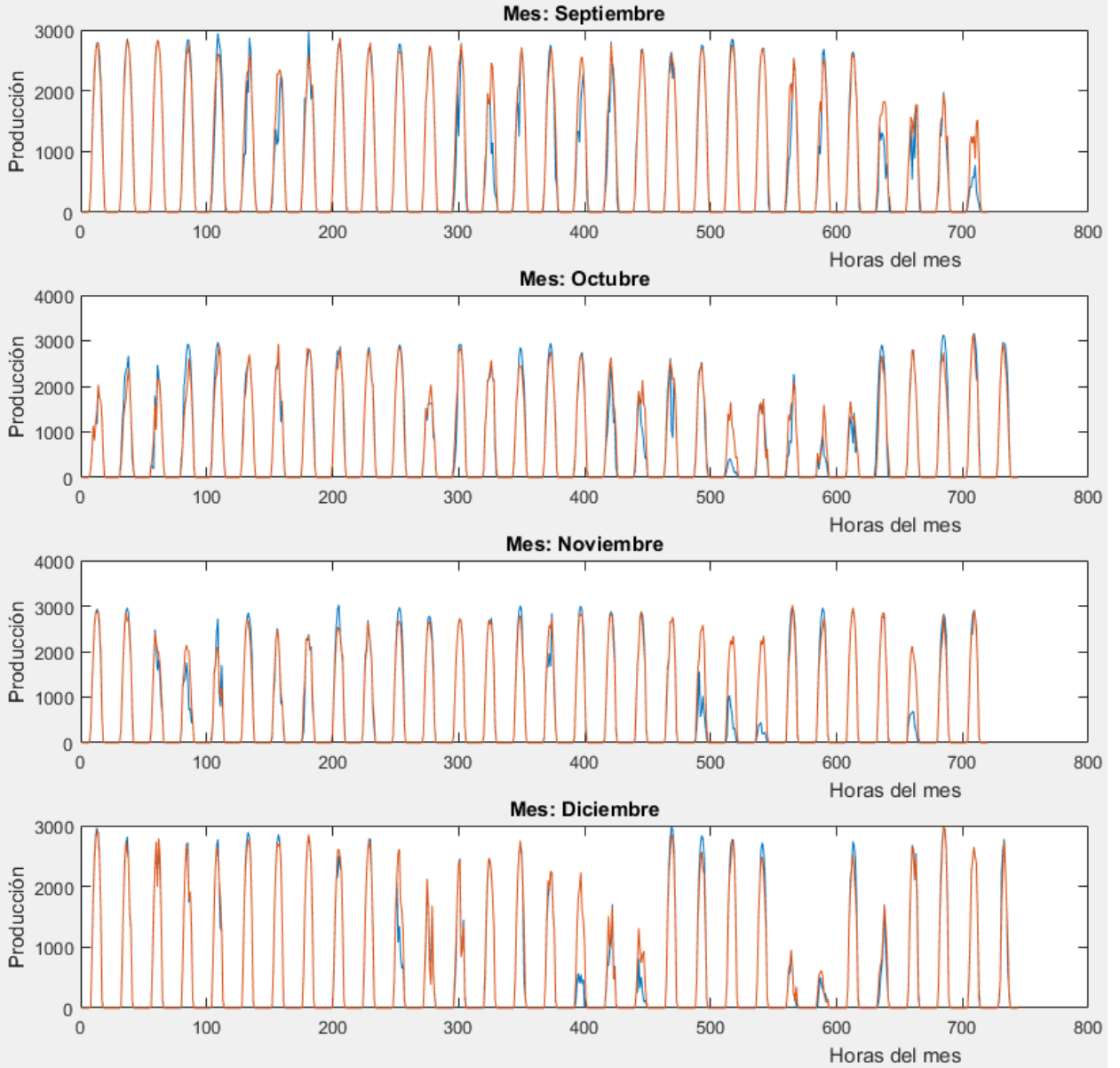
EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LOS K-VECINOS MÁS CERCANOS



EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LOS K-VECINOS MÁS CERCANOS

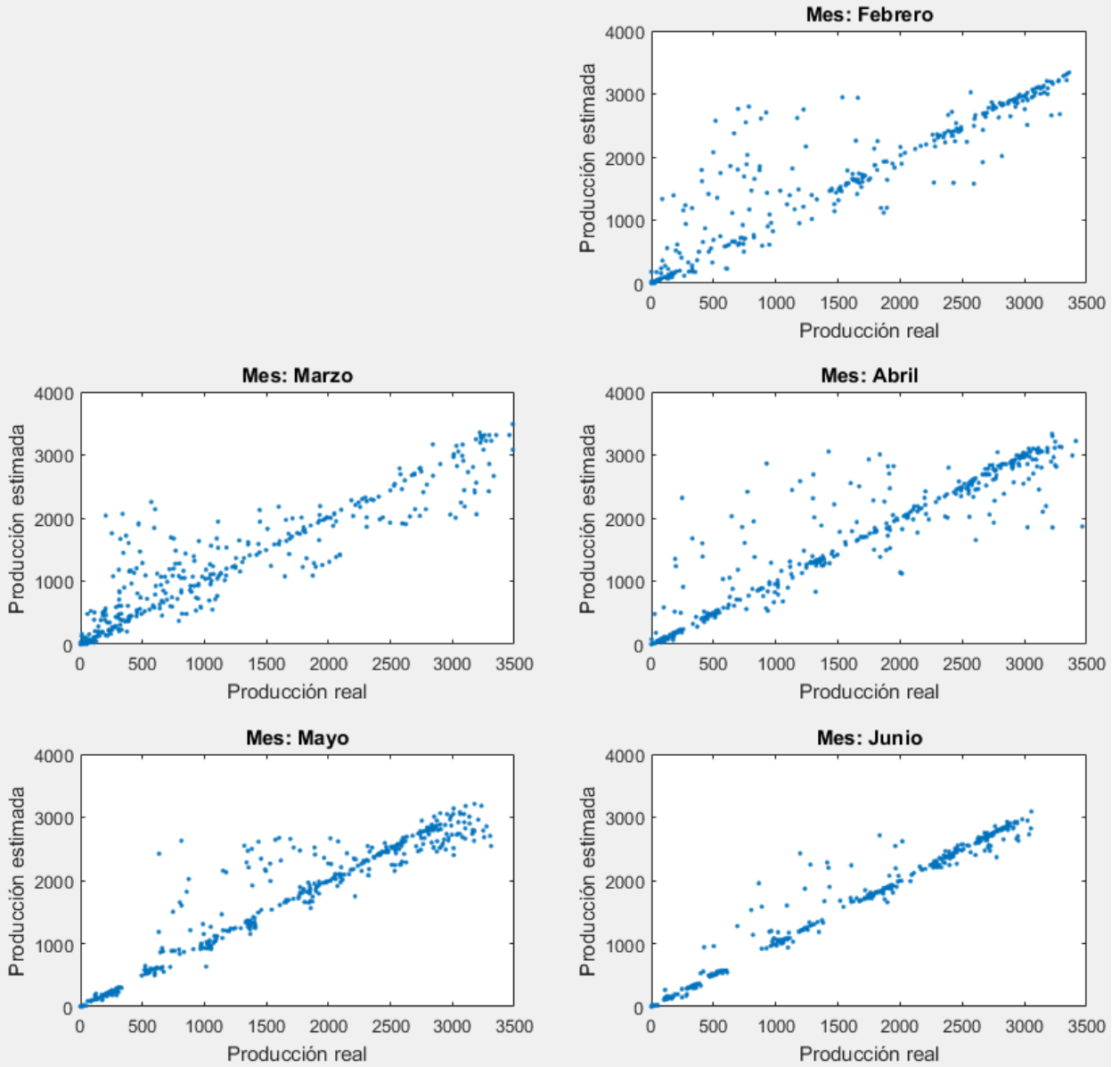


EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO DE LOS K-VECINOS MÁS CERCANOS

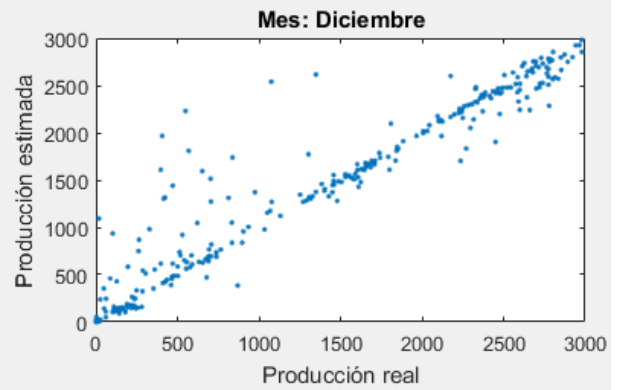
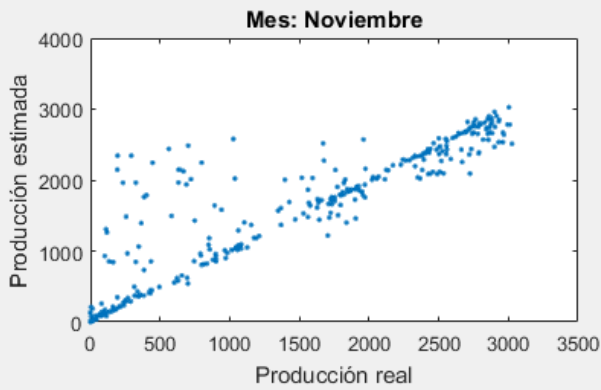
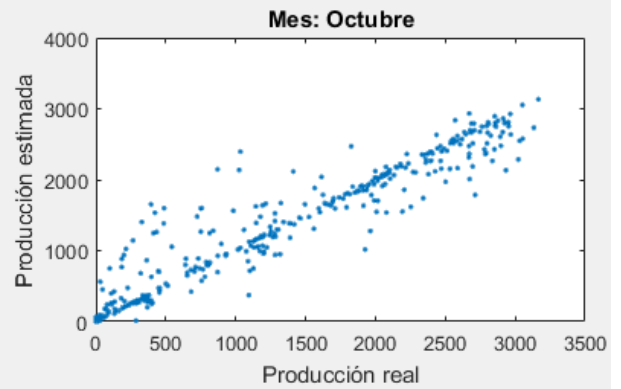
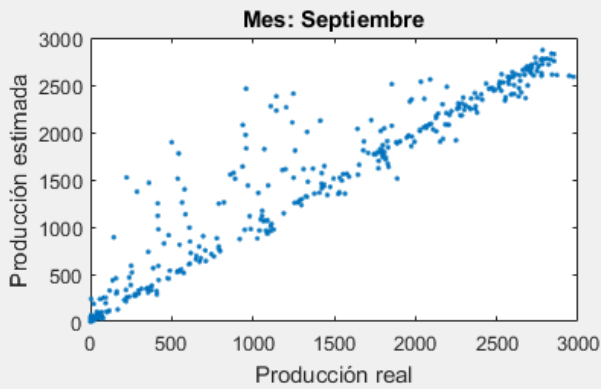
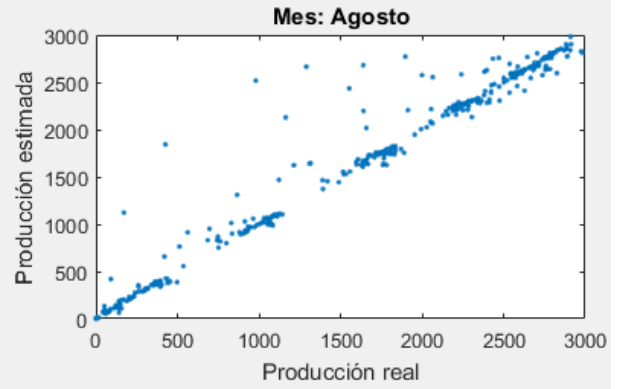
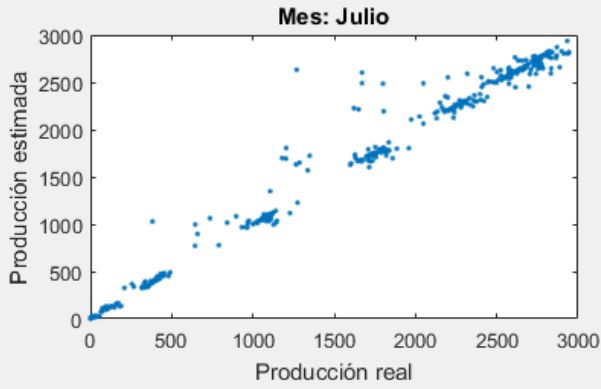


Anexo XI. Producción real vs Producción estimada. Método de los k-vecinos más cercanos

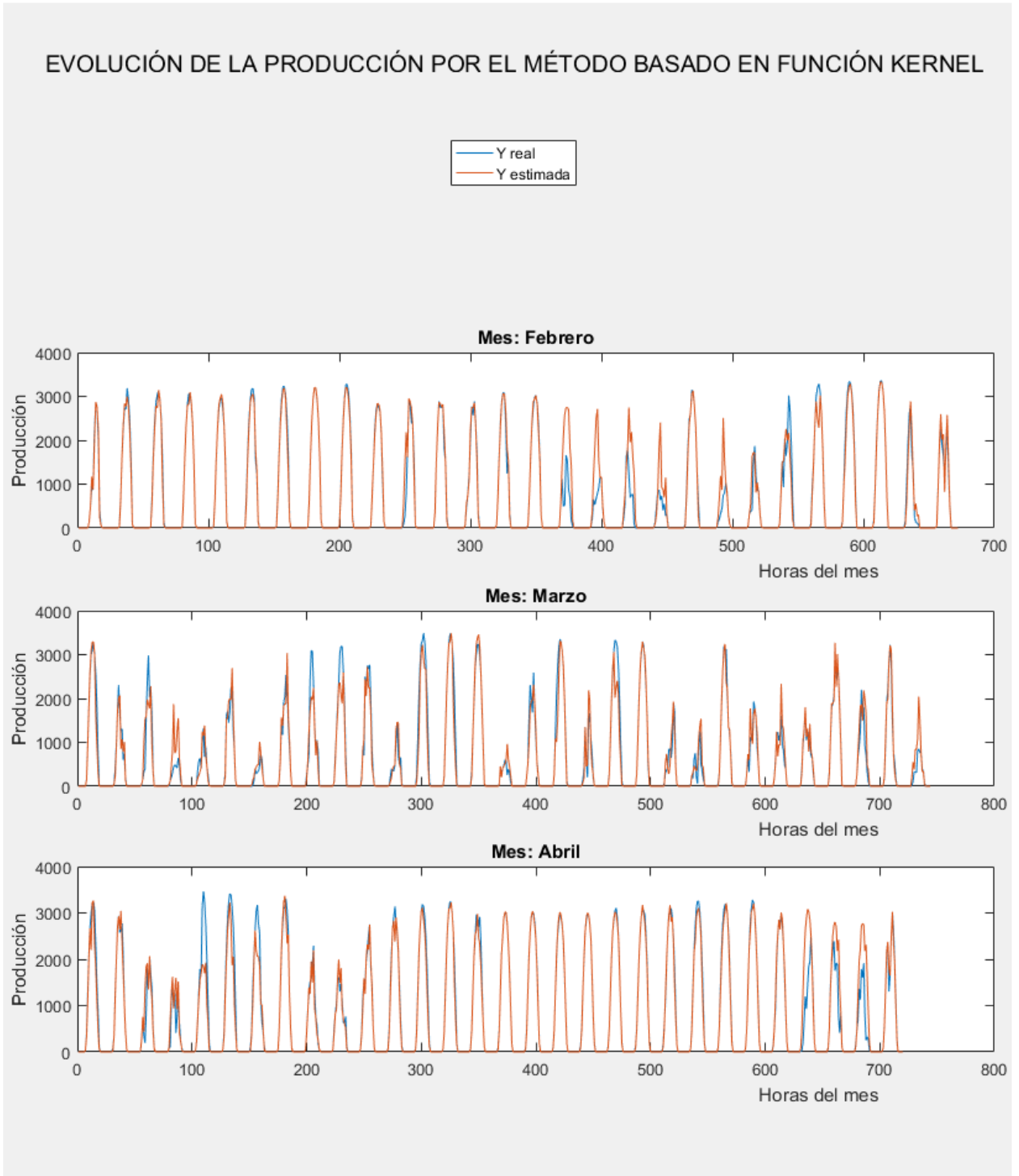
Y REAL vs Y ESTIMADA. MÉTODO DE LOS K-VECINOS MÁS CERCANOS



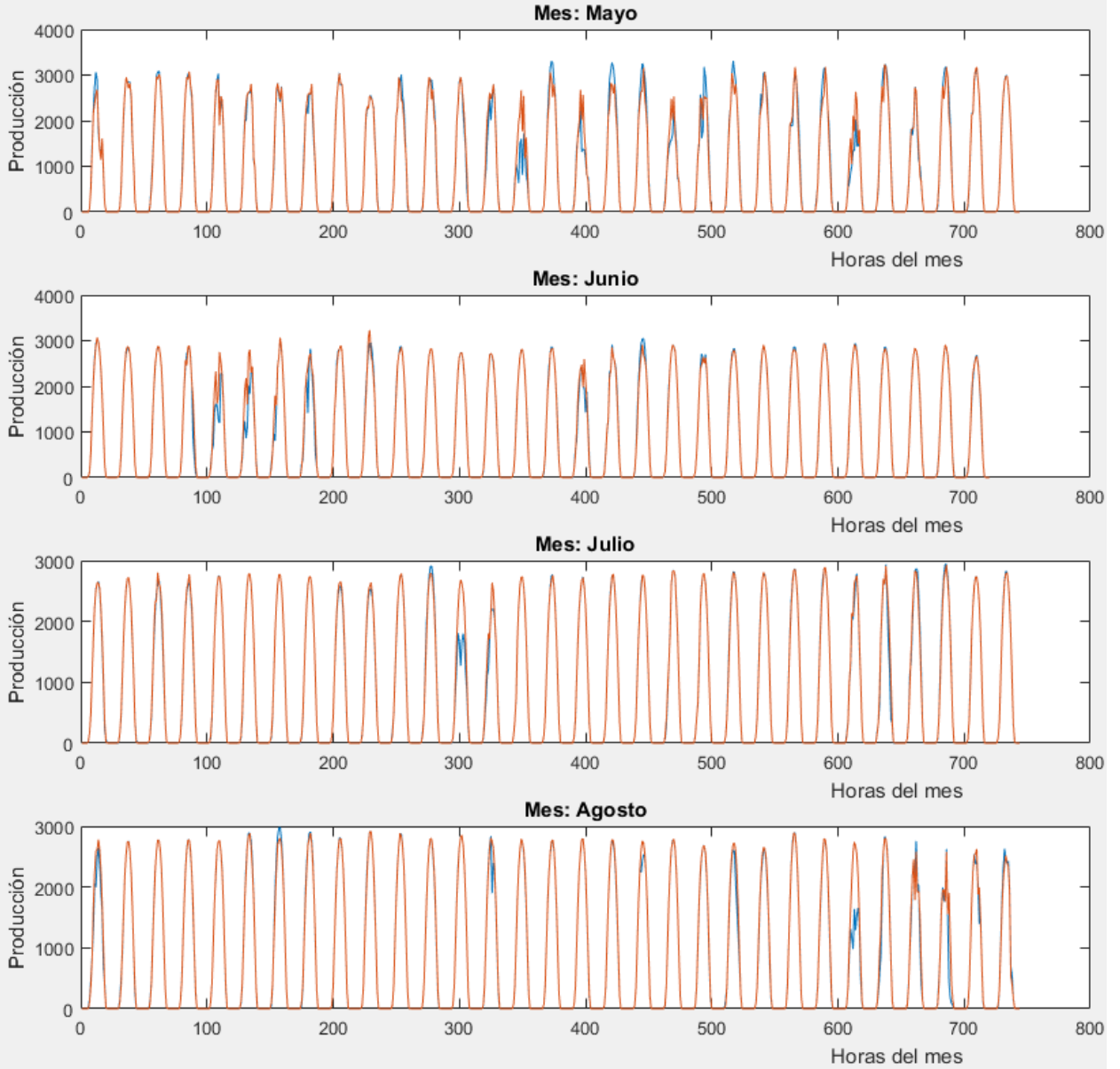
Y REAL vs Y ESTIMADA. MÉTODO DE LOS K-VECINOS MÁS CERCANOS



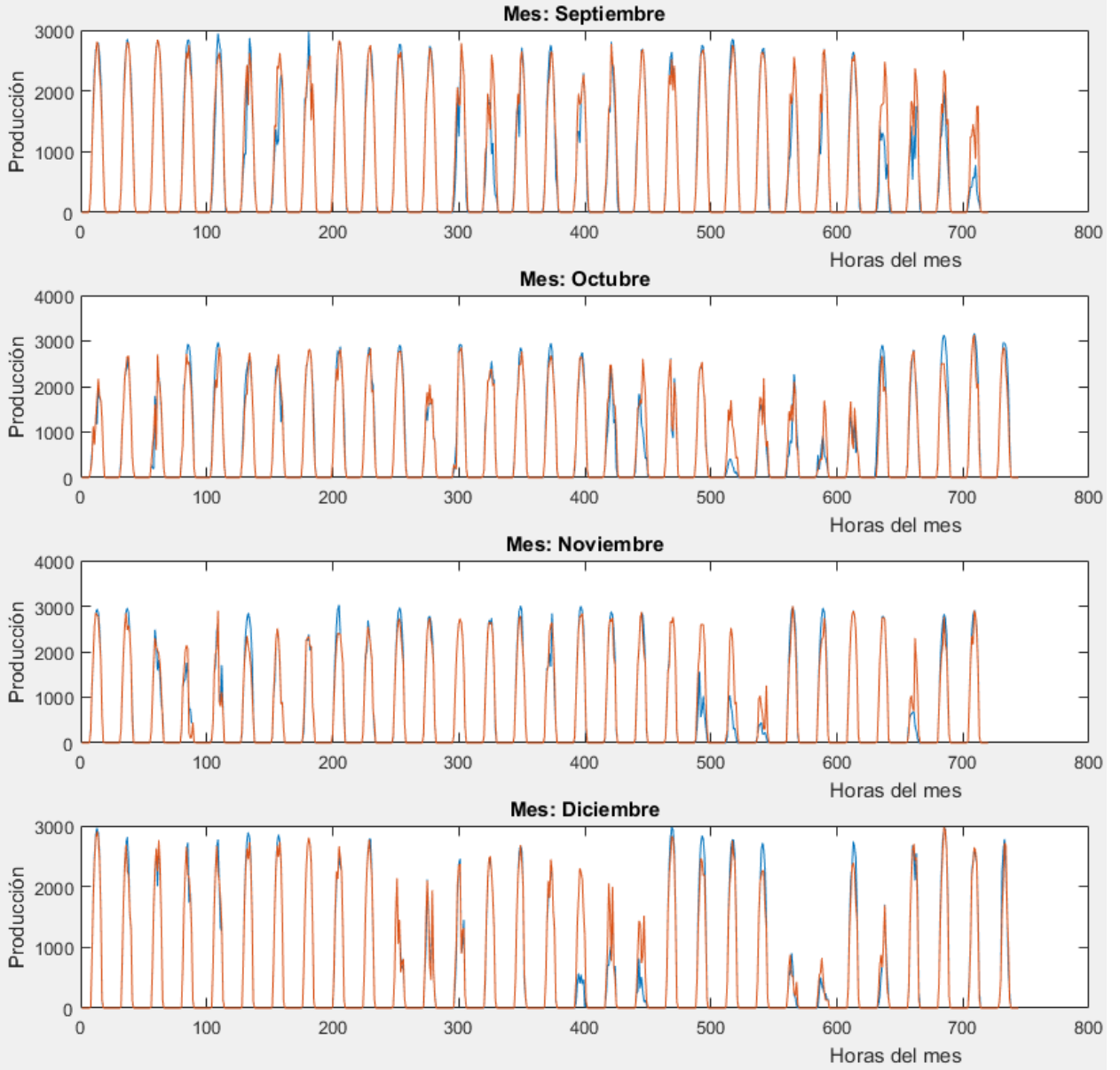
Anexo XII. Evolución de la producción por el método basado en función Kernel



EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO BASADO EN FUNCIÓN KERNEL

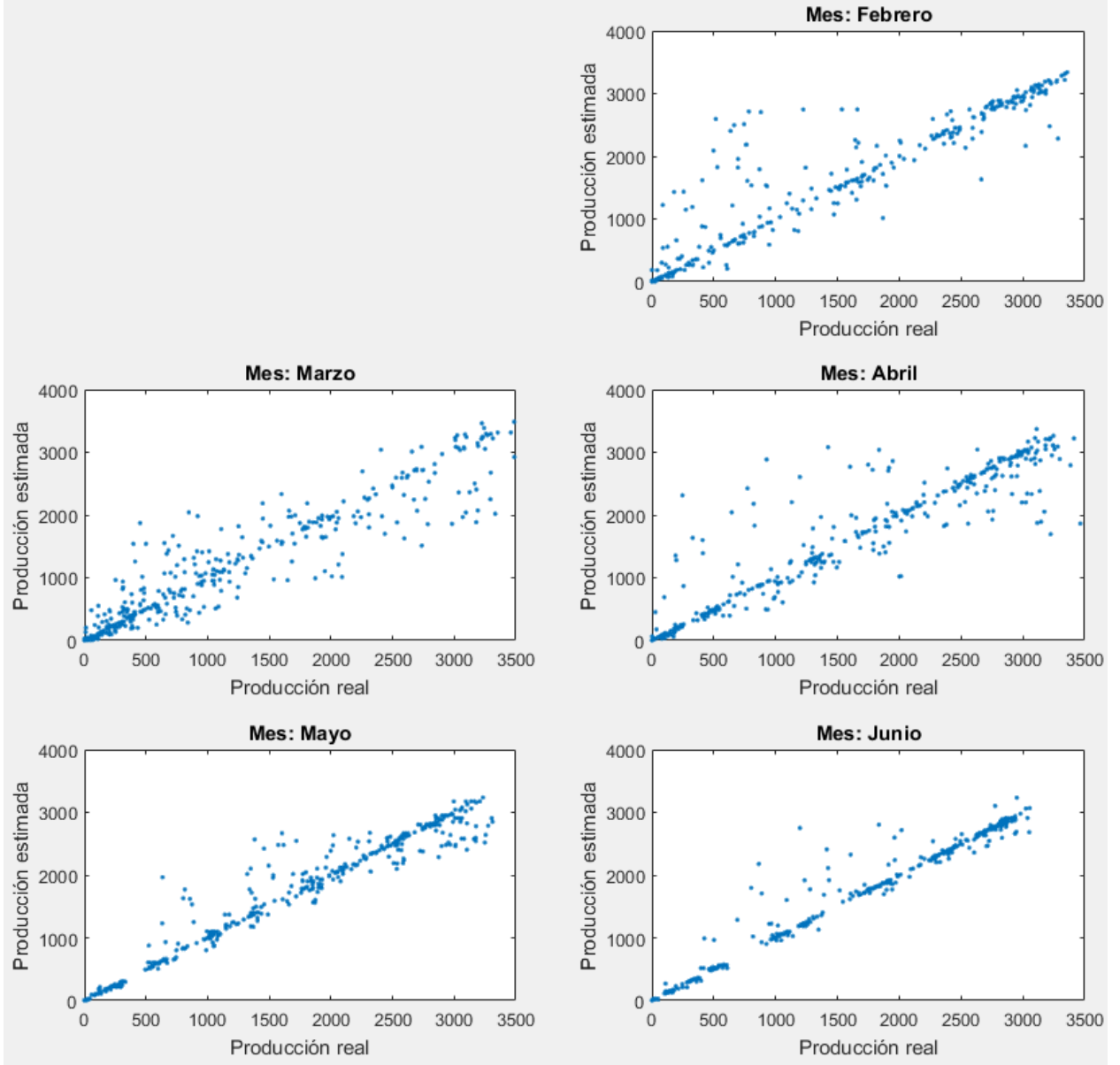


EVOLUCIÓN DE LA PRODUCCIÓN POR EL MÉTODO BASADO EN FUNCIÓN KERNEL



Anexo XIII. Producción real vs Producción estimada. Método basado en función Kernel

Y REAL vs Y ESTIMADA. MÉTODO BASADO EN FUNCIÓN KERNEL



Y REAL vs Y ESTIMADA. MÉTODO BASADO EN FUNCIÓN KERNEL

