



MÁSTER EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MÁSTER

GENERACIÓN DE ESCENARIOS DE ENERGÍA SOLAR EN MERCADOS ELÉCTRICOS CONSIDERANDO DEPENDENCIAS ESPACIO TEMPORALES

Autor: Diego Benito Adrados

Director: Antonio Bello Morales

Director: Geovanny Alberto Marulanda García

Madrid

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
Generación de escenarios de energía solar en mercados eléctricos considerando
dependencias espacio temporales

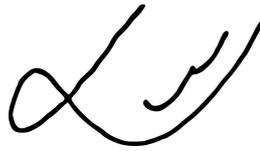
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2020/21 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

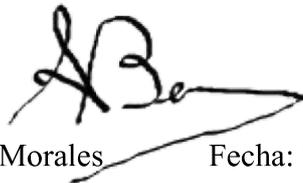


Fdo.: Diego Benito Adrados

Fecha: 21/ 07/ 2021

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Antonio Bello Morales

Fecha: 21/ 07/ 2021

EL DIRECTOR DEL PROYECTO



Fdo.: Geovanny Alberto Marulanda García

Fecha: 21/ 07/ 2021

AUTORIZACIÓN PARA LA DIGITALIZACIÓN, DEPÓSITO Y DIVULGACIÓN EN RED DE PROYECTOS FIN DE GRADO, FIN DE MÁSTER, TESIS O MEMORIAS DE BACHILLERATO

1º. Declaración de la autoría y acreditación de la misma.

El autor D. Diego Benito Adrados DECLARA ser el titular de los derechos de propiedad intelectual de la obra: Generación de escenarios de energía solar en mercados eléctricos considerando dependencias espacio-temporales, que ésta es una obra original, y que ostenta la condición de autor en el sentido que otorga la Ley de Propiedad Intelectual.

2º. Objeto y fines de la cesión.

Con el fin de dar la máxima difusión a la obra citada a través del Repositorio institucional de la Universidad, el autor **CEDE** a la Universidad Pontificia Comillas, de forma gratuita y no exclusiva, por el máximo plazo legal y con ámbito universal, los derechos de digitalización, de archivo, de reproducción, de distribución y de comunicación pública, incluido el derecho de puesta a disposición electrónica, tal y como se describen en la Ley de Propiedad Intelectual. El derecho de transformación se cede a los únicos efectos de lo dispuesto en la letra a) del apartado siguiente.

3º. Condiciones de la cesión y acceso

Sin perjuicio de la titularidad de la obra, que sigue correspondiendo a su autor, la cesión de derechos contemplada en esta licencia habilita para:

- a) Transformarla con el fin de adaptarla a cualquier tecnología que permita incorporarla a internet y hacerla accesible; incorporar metadatos para realizar el registro de la obra e incorporar “marcas de agua” o cualquier otro sistema de seguridad o de protección.
- b) Reproducirla en un soporte digital para su incorporación a una base de datos electrónica, incluyendo el derecho de reproducir y almacenar la obra en servidores, a los efectos de garantizar su seguridad, conservación y preservar el formato.
- c) Comunicarla, por defecto, a través de un archivo institucional abierto, accesible de modo libre y gratuito a través de internet.
- d) Cualquier otra forma de acceso (restringido, embargado, cerrado) deberá solicitarse expresamente y obedecer a causas justificadas.
- e) Asignar por defecto a estos trabajos una licencia Creative Commons.
- f) Asignar por defecto a estos trabajos un HANDLE (URL *persistente*).

4º. Derechos del autor.

El autor, en tanto que titular de una obra tiene derecho a:

- a) Que la Universidad identifique claramente su nombre como autor de la misma
- b) Comunicar y dar publicidad a la obra en la versión que ceda y en otras posteriores a través de cualquier medio.
- c) Solicitar la retirada de la obra del repositorio por causa justificada.
- d) Recibir notificación fehaciente de cualquier reclamación que puedan formular terceras personas en relación con la obra y, en particular, de reclamaciones relativas a los derechos de propiedad intelectual sobre ella.

5º. Deberes del autor.

El autor se compromete a:

- a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.
- b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.
- c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e intereses a causa de la cesión.
- d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción

de derechos derivada de las obras objeto de la cesión.

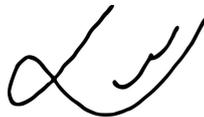
6º. Fines y funcionamiento del Repositorio Institucional.

La obra se pondrá a disposición de los usuarios para que hagan de ella un uso justo y respetuoso con los derechos del autor, según lo permitido por la legislación aplicable, y con fines de estudio, investigación, o cualquier otro fin lícito. Con dicha finalidad, la Universidad asume los siguientes deberes y se reserva las siguientes facultades:

- La Universidad informará a los usuarios del archivo sobre los usos permitidos, y no garantiza ni asume responsabilidad alguna por otras formas en que los usuarios hagan un uso posterior de las obras no conforme con la legislación vigente. El uso posterior, más allá de la copia privada, requerirá que se cite la fuente y se reconozca la autoría, que no se obtenga beneficio comercial, y que no se realicen obras derivadas.
- La Universidad no revisará el contenido de las obras, que en todo caso permanecerá bajo la responsabilidad exclusiva del autor y no estará obligada a ejercitar acciones legales en nombre del autor en el supuesto de infracciones a derechos de propiedad intelectual derivados del depósito y archivo de las obras. El autor renuncia a cualquier reclamación frente a la Universidad por las formas no ajustadas a la legislación vigente en que los usuarios hagan uso de las obras.
- La Universidad adoptará las medidas necesarias para la preservación de la obra en un futuro.
- La Universidad se reserva la facultad de retirar la obra, previa notificación al autor, en supuestos suficientemente justificados, o en caso de reclamaciones de terceros.

Madrid, a 21 de julio de 2021

ACEPTA



Fdo. Diego Benito Adrados

Motivos para solicitar el acceso restringido, cerrado o embargado del trabajo en el Repositorio Institucional:



MÁSTER EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MÁSTER

GENERACIÓN DE ESCENARIOS DE ENERGÍA SOLAR EN MERCADOS ELÉCTRICOS CONSIDERANDO DEPENDENCIAS ESPACIO TEMPORALES

Autor: Diego Benito Adrados

Director: Antonio Bello Morales

Director: Geovanny Alberto Marulanda García

Madrid

GENERACIÓN DE ESCENARIOS DE ENERGÍA SOLAR EN MERCADOS ELÉCTRICOS CONSIDERANDO DEPENDENCIAS ESPACIO TEMPORALES

Autor: Benito Adrados, Diego.

Director: Bello Morales, Antonio.

Director: Marulanda García, Geovanny Alberto

Entidad Colaboradora: Instituto de Investigación Tecnológica

RESUMEN DEL PROYECTO

En este proyecto se propone una metodología para realizar una generación de escenarios de energía solar fotovoltaica a largo plazo en sistemas multiárea, teniendo en cuenta las dependencias espacio-temporales existentes entre las áreas.

Palabras clave: Predicción, fotovoltaica, modelos ARIMA, simulaciones Monte Carlo.

1. Introducción

La energía solar fotovoltaica se está expandiendo muy rápidamente por todo el mundo, debido sobre todo a la necesidad de los países de descarbonizar sus economías y a la reducción de los precios de la tecnología durante los últimos años, que ha hecho que esta tecnología sea muy competitiva.

Sin embargo, también tienen ciertos inconvenientes, siendo el principal que esta tecnología no es gestionable. Generan en función de la radiación solar que llegue a los paneles fotovoltaicos, por lo que la generación por la noche es nula y puede cambiar mucho de un día a otro en función de las condiciones climatológicas. Además, no existe una infraestructura de almacenamiento lo suficientemente grande en ningún país del mundo que pueda ayudar a paliar este problema.

Debido a este inconveniente, las predicciones de energía solar son fundamentales para gestionar los sistemas eléctricos. En la literatura académica cada vez se pueden encontrar más artículos sobre predicciones de solar, pero la mayoría se enfocan en instalaciones concretas y en horizontes temporales muy cortos. Las predicciones a largo plazo y en grandes áreas son también muy necesarias, ya que ayudan enormemente tanto a inversores como a los operadores de la red.

Existen muchos métodos para elaborar predicciones para energía solar fotovoltaica. Normalmente se clasifican en 3 categorías:

- Métodos físicos: estos métodos elaboran sus predicciones a partir de variables atmosféricas. Dentro de esta categoría se pueden encontrar los Numerical Weather Predictor (NWP), las imágenes satélites, etc.
- Métodos estadísticos: estos métodos elaboran sus predicciones a partir de series temporales de generación eléctrica o variables atmosféricas. Las redes neuronales, los modelos de regresión o el Support Vector Machine (SVM) entrarían dentro de esta categoría.

- Métodos combinados: como su propio nombre indican, combinan los dos anteriores para elaborar la predicción. Los hay cooperativos o competitivos.

En este trabajo se ha optado por utilizar una metodología estadística basada en la agregación de los modelos de descomposición aditivos y los modelos ARIMA. Esta metodología busca capturar las principales dinámicas de las series temporales y encontrar las correlaciones existentes, y permiten generar escenarios reproduciendo dichas relaciones.

2. Definición del Proyecto

En este trabajo se ha propuesto una metodología para realizar una generación de escenarios de energía fotovoltaica en sistemas multiárea, teniendo en cuenta las dependencias espacio-temporales entre ellos. Las áreas que se han estudiado han sido España, Portugal y Francia y se evalúa, de forma específica, la calidad de los escenarios en un horizonte temporal de dos años. Las predicciones se han hecho a partir de la unión de modelos de descomposición aditivos con modelos ARIMA sin parte estacional, y para la generación de escenarios se han hecho simulaciones de Monte Carlo.

Se comparan dos métodos distintos.

- Estático: se ajusta un modelo ARIMA para la serie temporal de cada área, por lo que habría un total de 3 modelos.
- Dinámico: se divide la serie temporal de cada área en 12 partes, una por cada mes, y se ajusta un modelo ARIMA para cada uno de los meses. Habría un total de 12 modelos ARIMA por área.

3. Descripción del modelo

Se ha partido de series temporales históricas con desagregación horaria de generación fotovoltaica. Se han extraído las componentes principales de estas series, que son la tendencia, la estacionalidad anual y la estacionalidad diaria. Las transformaciones que se han realizado sobre los datos originales, explicadas de manera resumida, son:

1. Eliminación de la dependencia de la capacidad instalada, dividiendo los datos de generación entre la potencia fotovoltaica.
2. Corrección de *outliers* y N/As, interpolando los datos adyacentes
3. Corrección de los cambios de hora de marzo y octubre.
4. Eliminación de las horas nocturnas, en las que no hay generación o es insignificante.
5. Igualación de la frecuencia diaria a lo largo de las series temporales mediante interpolación por *splines*.
6. Normalización de la serie mediante transformada *logit*.
7. Captura de la estacionalidad anual mediante distintos métodos, escogiendo al final el que mejor funcione.
8. Captura de la estacionalidad diaria mediante medias.

Una vez se han hecho todas estas transformaciones, se ha llegado a tres series temporales (una por área) estacionarias. Para el caso estático se ajustará un modelo ARIMA para cada área, y para el dinámico, uno para cada área y cada mes. Asimismo, se emplea la

metodología Box-Jenkins para el ajuste de los modelos ARIMA. La siguiente figura muestra un resumen de la metodología:

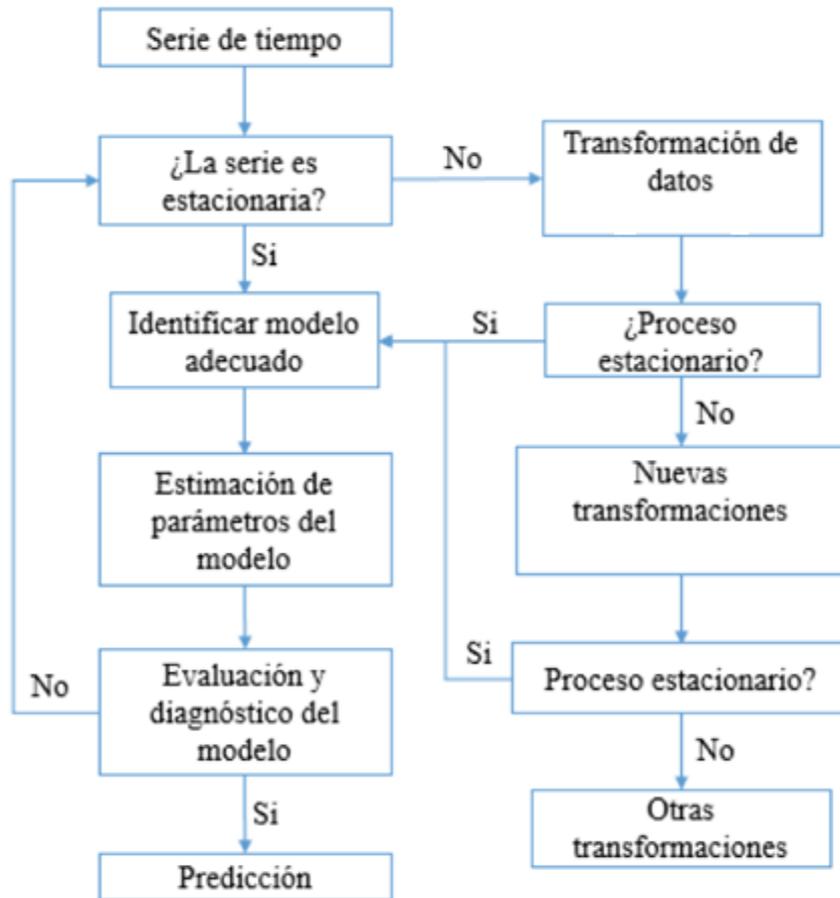


Ilustración 1: Resumen de la metodología Box-Jenkins. Tomado de [1].

Para elegir el orden del modelo, se ha minimizado lo máximo posible el AICc. Los coeficientes del modelo se han calculado mediante el criterio de máxima verosimilitud.

Una vez calculado el modelo, se han generado los escenarios mediante simulaciones de Monte Carlo. Estas simulaciones permiten generar distribuciones completas para cada hora que incluyan las dependencias temporales identificadas por los modelos ARIMA.

Generados los escenarios, se desharán todas las transformaciones que se han realizado y se evaluará el rendimiento de las simulaciones empleando dos indicadores, el Pinball Loss Function (PLF) y el Winkler Score (WS).

4. Resultados

A continuación, se van a mostrar los datos de generación históricos de cada área y los escenarios simulados por los dos métodos que se han nombrado antes.

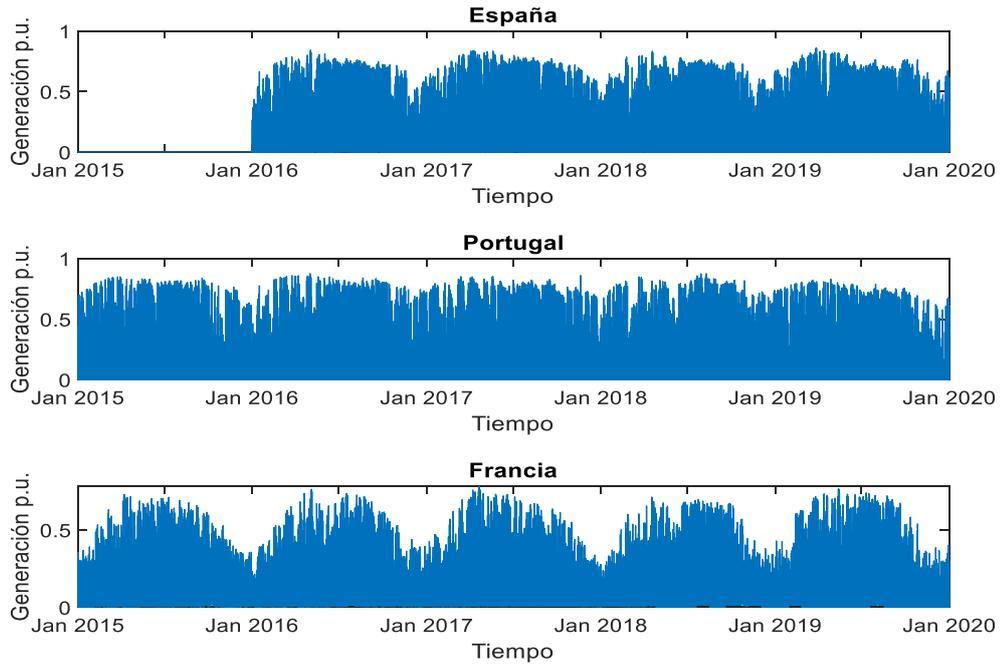


Figura 1: Series históricas de generación fotovoltaica en p.u.

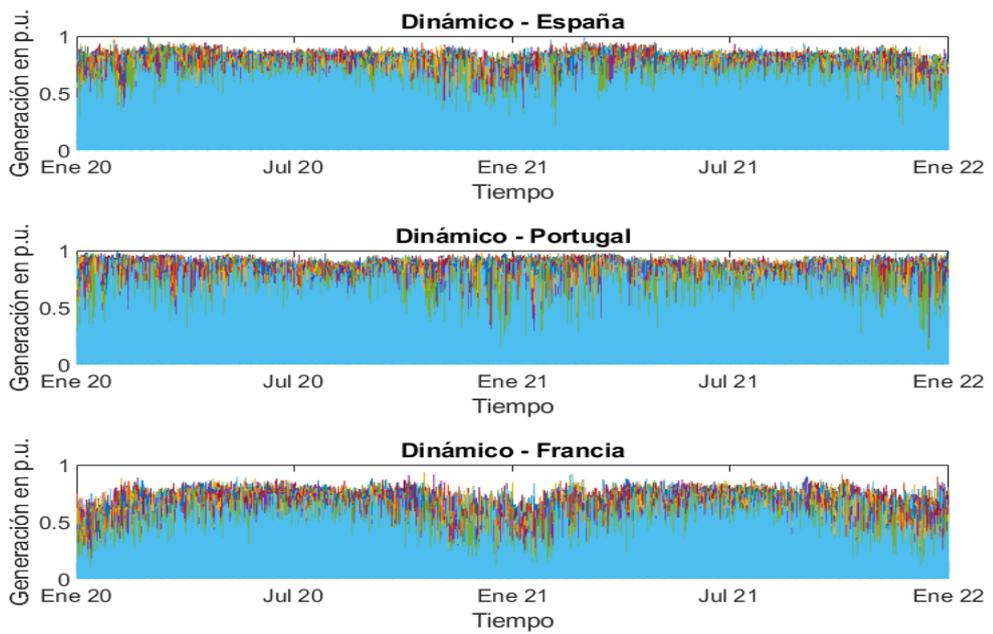


Figura 2: Escenarios generados con el método dinámico.

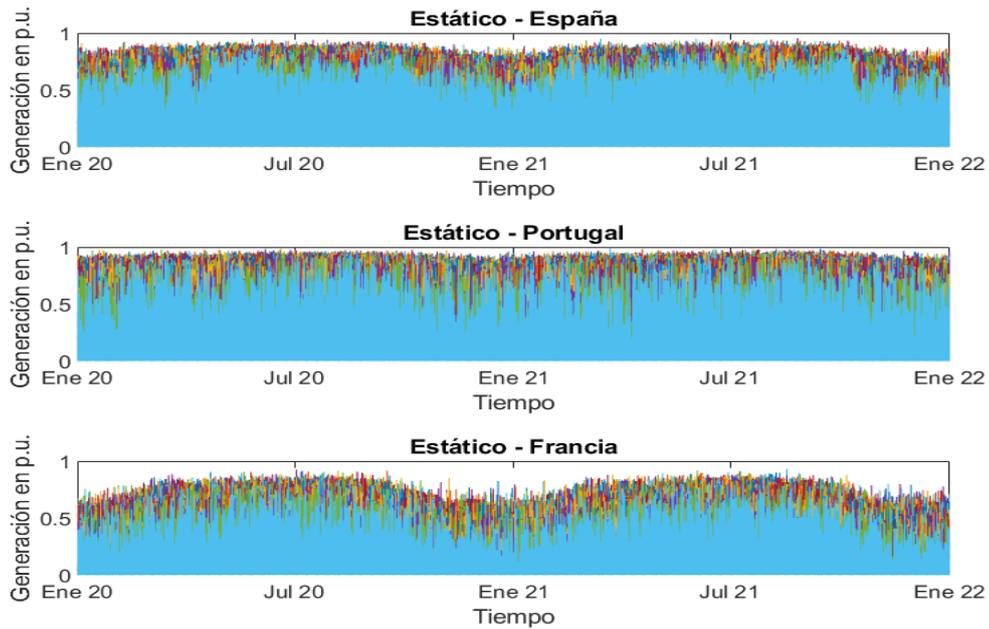


Figura 3: Escenarios generados con el método estático.

Las sendas arrojan resultados muy satisfactorios para capturar la dinámica de las series y su distribución asociada, teniendo en cuenta sus idiosincrasias. Se logra captar la estacionalidad anual y la diaria. Para evaluar el rendimiento de la predicción se han empleado los indicadores PLF y WS. El primero es una medida del error para las previsiones cuantílicas y el segundo permite evaluar a la vez la cobertura y la amplitud. Cuanto más bajos sean los dos, más precisa será la predicción. Para el WS se han escogido los intervalos 50%, 80%, 90% y 98% y para el PLF los percentiles P1, P5, P30, P50, P70, P95 y P99. Se muestra a continuación la media de estos indicadores para dichos intervalos y percentiles.

Tabla 1: Resultados de las métricas probabilísticas.

ÁREA	Modelo	WS-F1	WS-F2	WS-Spl	PLF-F1	PLF-F2	PLF-Spl
España	Dinámico	0.1918	0.1928	0.1868	0.0090	0.0089	0.0085
	Estático	0.2055	0.2026	0.2039	0.0092	0.0089	0.0089
Portugal	Dinámico	0.2681	0.2607	0.2559	0.0130	0.0126	0.0124
	Estático	0.3356	0.3323	0.3390	0.0156	0.0153	0.0153
Francia	Dinámico	0.2781	0.2766	0.2482	0.0115	0.0112	0.0094
	Estático	0.2690	0.2642	0.2487	0.0098	0.0094	0.0092

5. Conclusiones

Se puede concluir que:

- Los modelos ARIMA se pueden emplear de forma satisfactoria para generar escenarios a largo plazo.
- De forma general, el modelo dinámico da mejores resultados que el estático, excepto para el PLF de Francia.

- El método de interpolación por splines es el que mejor captura la estacionalidad diaria.
- No es necesario incluir una parte estacional en el modelo ARIMA.

GENERATION OF SOLAR ENERGY SCENARIOS IN ELECTRICITY MARKETS CONSIDERING SPATIO-TEMPORAL DEPENDENCIES

Author: Benito Adrados, Diego.

Supervisor: Bello Morales, Antonio.

Collaborating Entity: Instituto de Investigación Tecnológica

ABSTRACT

This project proposes a methodology for the generation of long-term solar PV scenarios in multi-area systems, taking into account the spatio-temporal dependencies between areas. Keywords: prediction, photovoltaics, ARIMA models, Monte Carlo simulations.

1. Introduction

Solar photovoltaic energy is expanding rapidly around the world, mainly due to the need for countries to decarbonise their economies and the reduction in technology prices over the last few years, which has made this technology very competitive.

However, they also have certain drawbacks, the main one being that this technology is not manageable. They generate according to the solar radiation that reaches the photovoltaic panels, so generation at night is nil and can change a lot from one day to the next depending on weather conditions. Moreover, there is no storage infrastructure large enough in any country in the world to help alleviate this problem.

Due to this drawback, solar energy forecasts are essential for managing electricity systems. More and more articles on solar forecasting can be found in the academic literature, but most of them focus on specific installations and very short time horizons. Long-term and large-area forecasts are also much needed, as they greatly help both investors and grid operators.

There are many methods to produce forecasts for solar PV. They usually fall into 3 categories:

- Physical methods: these methods elaborate their predictions from atmospheric variables. Within this category we can find the Numerical Weather Predictor (NWP), satellite images, etc.
- Statistical methods: these methods make their predictions based on time series of electricity generation or atmospheric variables. Neural networks, regression models or the Support Vector Machine (SVM) fall into this category.
- Combined methods: as their name suggests, they combine the two previous methods to produce the prediction. There are cooperative or competitive methods.

In this paper we have chosen to use a statistical methodology based on the aggregation of additive decomposition models and ARIMA models. This methodology seeks to capture the main dynamics of the time series and find the existing correlations, and allows us to generate scenarios reproducing these relationships.

2. Project definition

In this work, a methodology has been proposed to generate photovoltaic energy scenarios in multi-area systems, taking into account the spatio-temporal dependencies between them. The areas studied are Spain, Portugal and France, and the quality of the scenarios is specifically evaluated over a two-year time horizon. The predictions have been made by combining additive decomposition models with ARIMA models without a seasonal part, and Monte Carlo simulations have been used to generate the scenarios. Two different methods will be compared.

- Static: an ARIMA model is fitted for the time series of each area, so there would be a total of 3 models.
- Dynamic: the time series for each area is divided into 12 parts, one for each month, and an ARIMA model is fitted for each month.

3. Model description

The starting point was hourly time series of photovoltaic generation. The main components of these series have been extracted, which are the trend, the annual seasonality and the daily seasonality. The transformations carried out on the original data, explained in summary form, are as follows:

1. Elimination of the dependence on installed capacity by dividing the generation data by the PV power.
2. Correction of outliers and N/As by interpolating adjacent data.
3. Correction of March and October time shifts.
4. Elimination of night hours, where there is no or negligible generation.
5. Equalisation of the daily frequency along the time series by means of spline interpolation.
6. Normalisation of the series by logit transform.
7. Capture of the annual seasonality using different methods, choosing the one that works best in the end.
8. Capture of daily seasonality using averages.

Once all these transformations have been made, three stationary time series (one per area) have been obtained. For the static case, an ARIMA model will be fitted for each area, and for the dynamic case, one for each area and each month. The Box-Jenkins methodology is also used to fit the ARIMA models. The following figure shows a summary of the methodology:

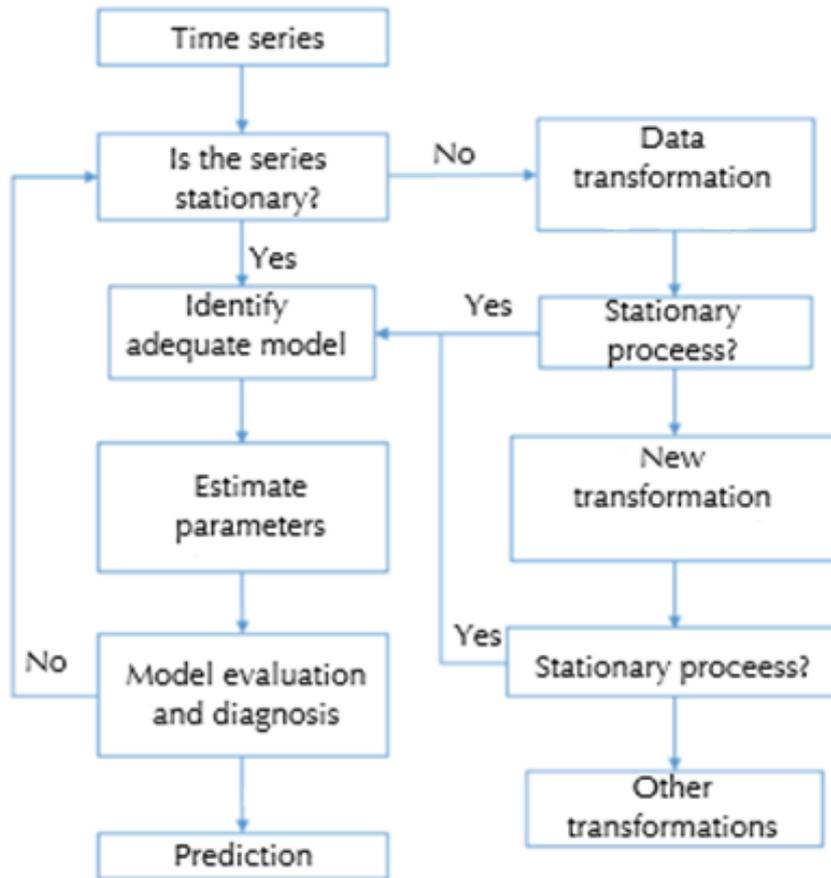


Figure 1: Diagram of Box-Cox methodology.

In order to choose the order of the model, the AICc has been minimised as much as possible. The model coefficients were calculated using the maximum likelihood criterion.

Once the model has been calculated, the scenarios have been generated using Monte Carlo simulations. These simulations allow the generation of complete distributions for each hour that include the time dependencies identified by the ARIMA models.

Once the scenarios have been generated, all the transformations that have been performed will be undone and the performance of the simulations will be evaluated using two indicators, the Pinball Loss Function (PLF) and the Winkler Score (WS).

4. Results

The historical generation data for each area and the scenarios simulated by the two methods mentioned above are shown below.

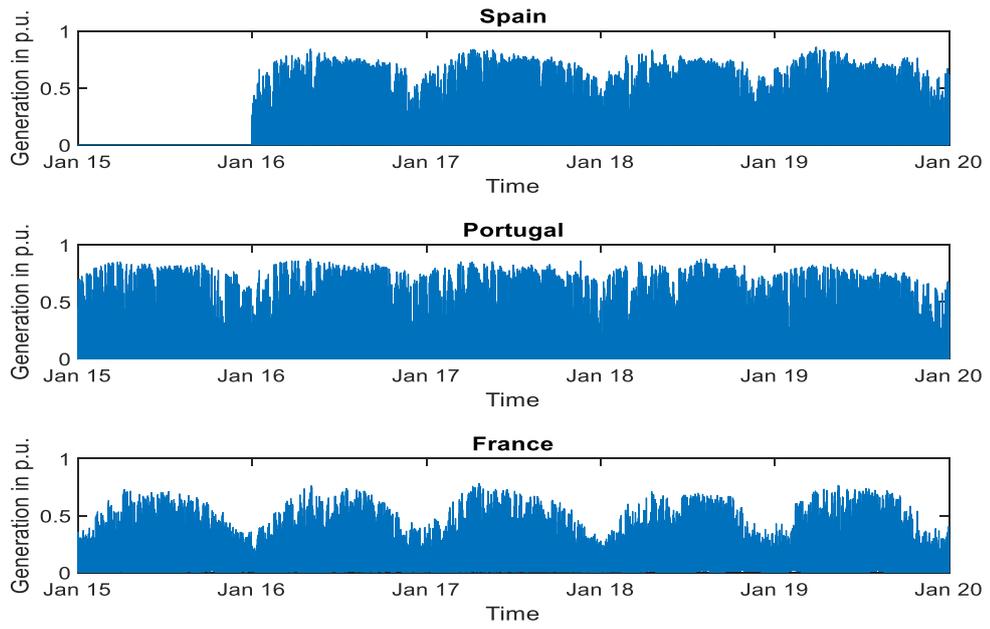


Figure 2: Historical time series of generation in p.u.

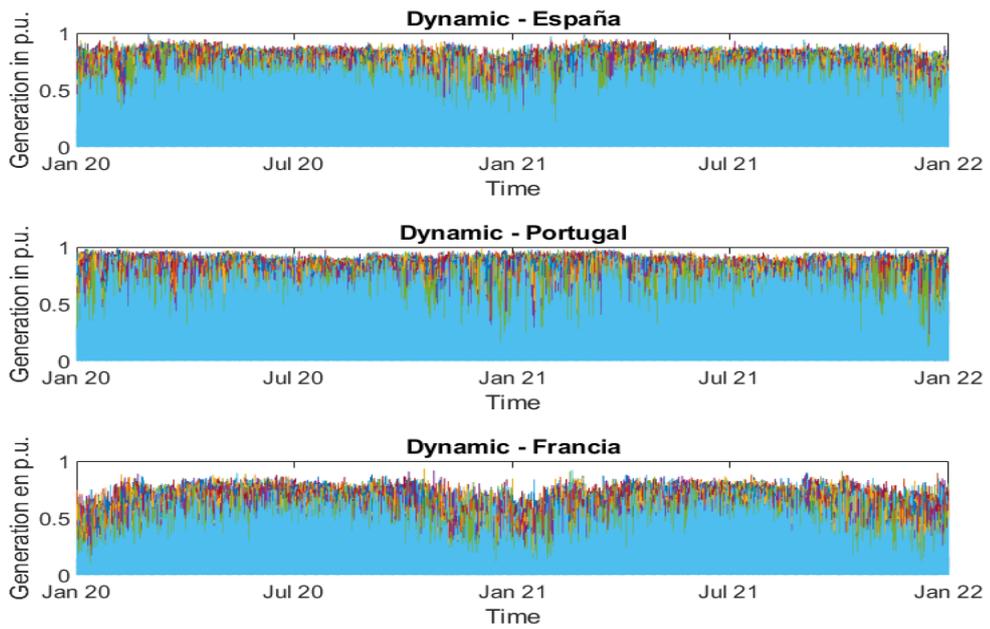


Figure 3: Generated scenarios with the dynamic model.

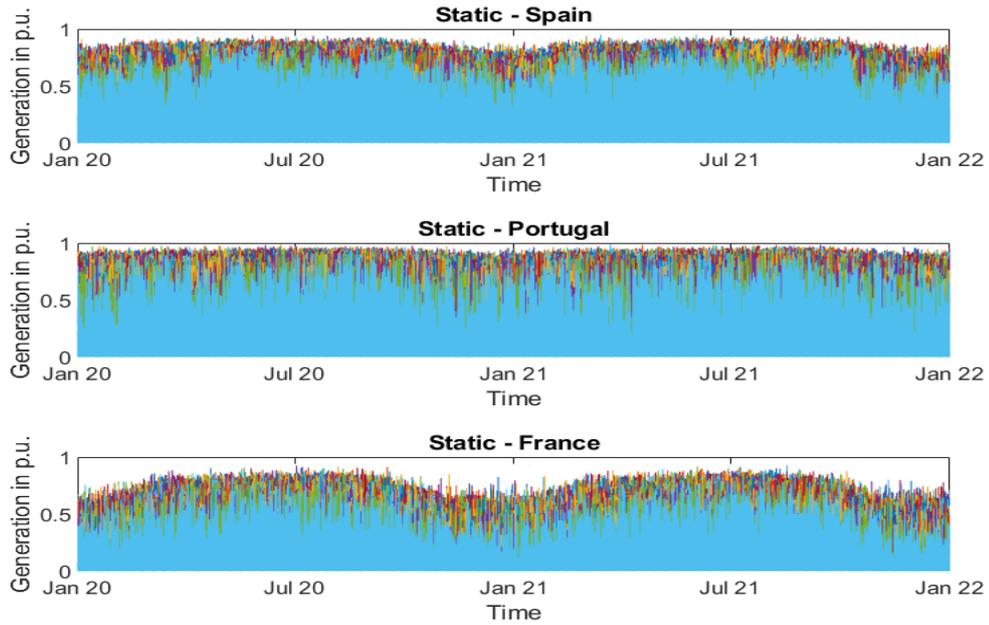


Figure 4: Generated scenarios with the static model.

The paths yield very satisfactory results in capturing the dynamics of the series and their associated distribution, taking into account their idiosyncrasies. Both annual and daily seasonality are captured. The PLF and WS indicators have been used to assess the forecasting performance. The former is a measure of the error for quantile forecasts and the latter allows both coverage and amplitude to be assessed. The lower the PLF and WS, the more accurate the prediction. For the WS the intervals 50%, 80%, 90% and 98% have been chosen and for the PLF the percentiles P1, P5, P30, P50, P70, P95 and P99 have been chosen. The mean of these indicators for these intervals and percentiles is shown below.

Table 1: Results of the probabilistic metrics.

Country	Model	WS-F1	WS-F2	WS-Spl	PLF-F1	PLF-F2	PLF-Spl
Spain	Dynamic	0.1918	0.1928	0.1868	0.0090	0.0089	0.0085
	Static	0.2055	0.2026	0.2039	0.0092	0.0089	0.0089
Portugal	Dynamic	0.2681	0.2607	0.2559	0.0130	0.0126	0.0124
	Static	0.3356	0.3323	0.3390	0.0156	0.0153	0.0153
France	Dynamic	0.2781	0.2766	0.2482	0.0115	0.0112	0.0094
	Static	0.2690	0.2642	0.2487	0.0098	0.0094	0.0092

5. Conclusions

It can be concluded that:

- ARIMA models can be used satisfactorily to generate long-term scenarios.
- In general, the dynamic model gives better results than the static model, except for the PLF of France.
- The spline interpolation method best captures daily seasonality.

- It is not necessary to include a seasonal part in the ARIMA model.

Índice de la memoria

1. Introducción	9
1.1 Cambio climático y medio ambiente	9
1.2 Estado actual de la generación solar	12
1.3 Generalidades de la generación solar	16
1.4 Motivación	19
1.5 Objetivos	20
2. Estado de la cuestión	22
2.1 Aspectos básicos de las técnicas de predicción	22
2.2 Métodos de predicción para energía fotovoltaica	24
2.3 Estado del arte	28
2.4 Comparativa entre modelos	32
2.5 Estado de la energía solar en las áreas de estudio	33
3. Metodología	46
3.1 Caracterización de los datos	47
3.2 Preprocesado	47
3.3 Descomposición	50
3.4 Detección de dependencias espacio-temporales	52
3.5 Generación de escenarios	54
3.6 Evaluación del rendimiento	55
4. Resultados	56
4.1 Estudio de los datos	56
4.2 preprocesado	62
4.3 Descomposición de la serie	68
4.4 Detección de dependencias espacio-temporales	72
4.5 Generación de escenarios	73
4.6 Evaluación del rendimiento	75
5. Conclusiones y trabajos futuros	80
6. Bibliografía	82

Anexo I **88**

Anexo II **89**

Anexo III **91**

Índice de figuras

Figura 1: Series históricas de generación fotovoltaica en p.u.	12
Figura 2: Escenarios generados con el método dinámico.....	12
Figura 3: Escenarios generados con el método estático.	13
Figura 4: Temperaturas globales anuales en el periodo 1850-2017. Diferencia de 1,35°C entre las barras más extremas. Tomado de [4]	9
Figura 5: Generación eléctrica global 2019 por fuente [7].....	11
Figura 6: Mapa del potencial de la energía solar a nivel mundial. Fuente: EOM [10].	13
Figura 7: Evolución de la potencia fotovoltaica instalada a nivel mundial. Fuente: IRENA [8].	14
Figura 8: Evolución de los precios de las subastas de fotovoltaica y capacidad instalada. Tomado de [15].	17
Figura 9: Clear Sky Index. Tomado de [27]......	31
Figura 10: Capacidad fotovoltaica acumulada 2000-2020. Tomado de [29].	33
Figura 11: Capacidad instalada en España 2020. Fuente: REE.....	35
Figura 12: Generación eléctrica en España en 2020. Fuente: REE.....	36
Figura 13: Evolución de la capacidad fotovoltaica instalada en España 2003-2019. Tomado de REE [31].	37
Figura 14: Capacidad fotovoltaica en España por comunidades autónomas a 31/12/2019. Tomado de REE [31]......	37
Figura 15: Capacidad instalada en Portugal por tecnología 2011-2020. Tomado de REN [32].	39
Figura 16: Generación eléctrica en Portugal por tecnología 2011-2020. Tomado de REN [32].	39
Figura 17: Capacidad fotovoltaica Portugal 2010-2020.....	40
Figura 18: Distribución de las instalaciones fotovoltaicas en Portugal en 2016. Fuente: Global Power Plant Database [33].....	41

Figura 19: Evolución de la capacidad fotovoltaica instalada en Francia 2007-2020. Tomado de RTE [34].	43
Figura 20: Distribución de la capacidad fotovoltaica instalada en Francia en 2020. Tomado de RTE [34].	44
Figura 21: Mapa de interconexiones entre las 3 áreas de estudio. Tomado de [31].	45
Figura 22: Resumen de la metodología Box-Jenkins. Tomado de [1].	52
Figura 23: Series temporales de generación fotovoltaica.	57
Figura 24: Histogramas, gráficos de dispersión y coeficiente de Pearson de las 3 áreas. ...	58
Figura 25: Generación y capacidad fotovoltaica en cada área 2015-2019.	59
Figura 26: Serie de generación fotovoltaica España 3D.	60
Figura 27: Serie de generación fotovoltaica Portugal 3D.	61
Figura 28: Serie de generación fotovoltaica Francia 3D.	61
Figura 29: Series temporales de generación fotovoltaica en p.u.	63
Figura 30: Generación fotovoltaica en p.u. a las 9 am España.	64
Figura 31: Generación en España en función de hora y día.	65
Figura 32: Generación en España en función de hora y día con frecuencia corregida.	66
Figura 33: Series temporales de generación fotovoltaica con frecuencia corregida.	67
Figura 34: Distribuciones de la serie original y las transformadas de España (sin horas nocturnas).	67
Figura 35: Series temporales con transformada logit aplicada.	68
Figura 36: Tendencia de las series temporales.	69
Figura 37: Métodos de captura de la estacionalidad anual.	70
Figura 38: Series temporales sin estacionalidad anual.	71
Figura 39: Series temporales sin estacionalidad anual ni diaria.	72
Figura 40: Sendas simuladas con modelo dinámico.	74
Figura 41: Sendas simuladas con modelo estático.	75
Figura 42: Métricas deterministas mensuales España.	76
Figura 43: Métricas deterministas mensuales Portugal.	77
Figura 44: Métricas deterministas mensuales de Francia.	77
Figura 45: Métricas deterministas horarias España.	78

Figura 46: Métricas deterministas horarias Portugal.....	78
Figura 47: Métricas deterministas horarias Francia.	79
Figura 48: Objetivos de Desarrollo Sostenible de la ONU.	88

Índice de tablas

Tabla 1: Resultados de las métricas probabilísticas.	13
Tabla 2: Países por potencia fotovoltaica instalada 2019. Fuente: UNEF.	14
Tabla 3: Objetivos de capacidad fotovoltaica instalada 2030. Nota: España escenario objetivo y Francia escenario B para 2028.	34
Tabla 4: Calendario de subastas fotovoltaicas Francia 2019-2023. Potencia en MW.	44
Tabla 5: Resumen de la metodología seguida.	46
Tabla 6: Transformada Ordered Quantile y Logit.	49
Tabla 7: Coeficientes de correlación de Pearson de capacidad instalada con respecto a generación.	60
Tabla 8: Resumen indicadores PLF y WS.	75
Tabla 9: Resultados Winkler Score.	89
Tabla 10: Resultados Pinball Loss Function.	90

Índice de abreviaturas

ACF	Autocorrelation Function (Función de Autocorrelación)
AEM	Asociación Española de Climatología
AIC	Akaike Information Criteria (Criterio de Información Akaike)
AICc	Corrected Akaike Information Criteria (Criterio de Información Akaike Corregido)
AIE	Agencia Internacional de la Energía
ANN	Artificial Neural Network (Red Neuronal Artificial)
AR	Autoregressive (Autorregresivo)
ARIMA	Autoregressive Integrated Moving Average (Autorregresivo Integrado de Media móvil)
ARMA	Autoregressive Moving Average (Autorregresivo de Media móvil)
BIC	Bayesian Information Criteria (Criterio de Información Bayesiano)
GEI	Gases de Efecto Invernadero
GHI	Global Horizontal Irradiance (Irradiación Global Horizontal)
IPCC	Intergovernmental Panel on Climate Change (Panel Intergubernamental del Cambio Climático)
IRENA	International Renewable Energy Agency (Agencia internacional de la Energía Renovable)
kNN	k Nearest Neighbour
MA	Moving Average (Media móvil)
MAE	Mean Absolute Error (Error Absoluto Medio)
MBE	Mean Bias Error (Error de Sesgo Medio)
MLP	Multilayer Perceptron (Perceptrón Multicapa)
NECP	National Energy and Climate Plan (Plan Nacional de Energía y Clima)
NWP	Numerical Weather Predictor (Predictor Numérico del Tiempo)
PACF	Partial Autocorrelation Function (Función de Autocorrelación Parcial)
PLF	Pinball Loss Function (Función de Pérdida Pinball)
PNIEC	Plan Nacional Integrado de Energía y Clima
REE	Red Eléctrica Española
RMSE	Root Mean Square Error (Raíz del Error Cuadrático Medio)
SARIMA	Seasonal Autoregressive Integrated Moving Average (Estacional Autorregresivo Integrado de Media móvil)
SSE	Sum of Squared Error (Suma de Errores Cuadrados)
SVR	Support Vector Regression (Regresor de Vectores de Soporte)
UE	Unión Europea
WEO	World Energy Outlook (Panorama Energético Mundial)

WS Winkler Score (Puntuación Winkler)

1. INTRODUCCIÓN

1.1 CAMBIO CLIMÁTICO Y MEDIO AMBIENTE

A lo largo de las últimas décadas el cambio climático se ha convertido en uno de los temas que más debate genera y se considera que es una de las mayores amenazas a las que se enfrenta el planeta. Una encuesta del año 2018 realizada por el Pew Research Center mostraba que el cambio climático es considerado por la opinión pública la mayor amenaza global que existe [2].

Según la Asociación Española de Meteorología “se trata de una variación en el estado del sistema climático que perdura durante periodos de tiempo suficientemente largos para suponer que se ha alcanzado un nuevo equilibrio” [3]. A continuación, se muestra la Figura 4, en la que cada barra vertical representa un año, correspondiendo las más azules con los años más fríos y las más rojas con los más cálidos. Se puede apreciar que los años más cálidos se concentran todos al final, lo que indica que existe un aumento sostenido de la temperatura media global.

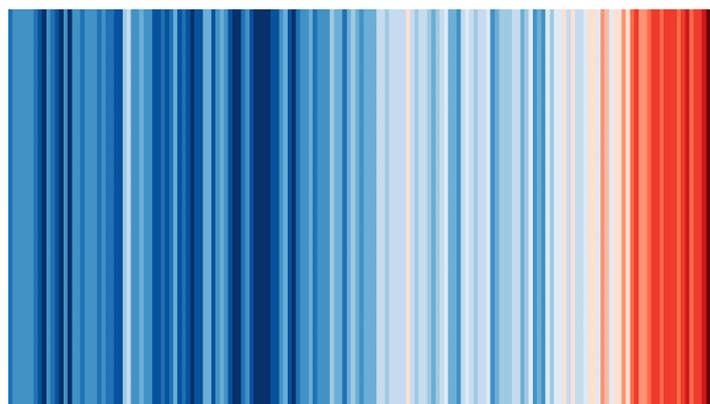


Figura 4: Temperaturas globales anuales en el periodo 1850-2017. Diferencia de 1,35°C entre las barras más extremas. Tomado de [4]

Ahora bien, los cambios climáticos se han dado a lo largo de toda la historia y se pueden producir por causas naturales tales como cambios en el eje de la Tierra, variabilidad solar,

actividad volcánica, etc. Sin embargo, dada la brusquedad y la rapidez del cambio actual, la práctica totalidad de la comunidad científica señala que este cambio climático está dado por causas antropogénicas. De hecho, el Panel Intergubernamental del Cambio climático (IPCC, por sus siglas en inglés) señala que “la influencia humana en el cambio climático es clara y va en aumento” [5].

Es por este motivo por el que desde hace unos años los gobiernos de distintos países y algunas entidades supranacionales se están poniendo de acuerdo en elaborar tratados en los que se comprometen a atajar las causas que producen el cambio climático. Algunos de los acuerdos o tratados más importantes son el Protocolo de Kyoto o los Acuerdos de París de 2015.

Para mitigar el cambio climático lo primero que habría que preguntarse es cuáles son las causas que lo producen. El IPCC señala la emisión de Gases de Efecto Invernadero (GEI) como una de las causas más importantes del calentamiento global. Entre estos gases están el vapor de agua, el metano, el dióxido de nitrógeno o el dióxido de carbono. Según la NASA, este último es el impulsor más importante del calentamiento global y los humanos hemos aumentado su concentración atmosférica en un 47% desde el inicio de la Revolución Industrial [6]. La reducción de las emisiones de estos gases es uno de los objetivos cruciales para frenar el cambio climático. Es por esto por lo que los grandes tratados sobre energía y clima hacen hincapié en este punto.

Para saber qué medidas hay que tomar para reducir las emisiones de GEI hay que conocer de dónde salen dichas emisiones. Realmente, hay muchas fuentes distintas de GEI, pero una gran parte de ellas sale de la quema de combustibles fósiles, tales como el carbón, el gas natural o los derivados del petróleo. Estos combustibles tienen una variedad de usos enorme. Sirven para climatizar edificios, para mover los motores de combustión de los vehículos y, por supuesto, para generar electricidad. Según la Agencia Internacional de la Energía (AIE), el sector eléctrico ha sido responsable de la emisión de 13699,1 Mt de CO₂ en 2019, lo que representa un alrededor de un 41% de las emisiones mundiales [7].

La Figura 5 presenta el mix energético global por tipo de fuente. Los combustibles fósiles, principalmente el carbón y el gas natural, suponen un 62% de la generación de energía eléctrica, muy por encima de lo que actualmente suponen las energías renovables [7].

Distribución eléctrica global por fuente 2019

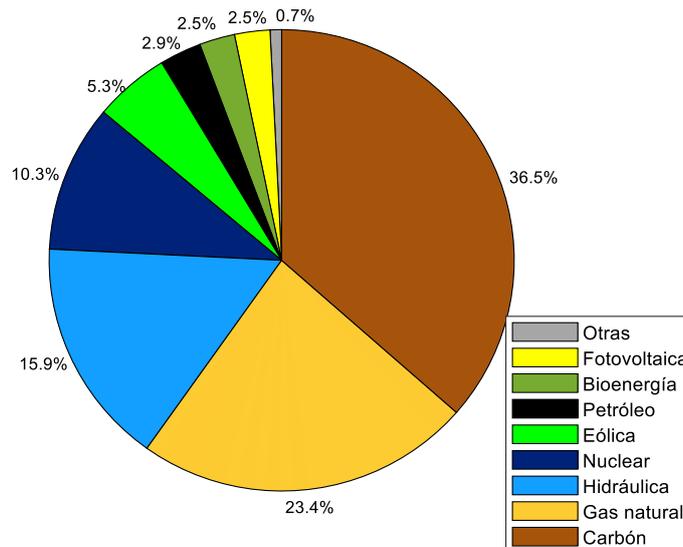


Figura 5: Generación eléctrica global 2019 por fuente [7].

Aunque todas las tecnologías de generación existentes producen algún tipo de impacto ambiental durante alguna de las etapas de su ciclo de vida, ya sea durante la construcción, el abastecimiento de materias primas u otras, existe un consenso en el que para frenar el cambio climático es necesario que se produzca un cambio tecnológico profundo en los mixes de generación. En particular, habrá que sustituir aquellas tecnologías que supongan importantes emisiones de GEI por otras “limpias” que apenas produzcan estos gases a lo largo de su ciclo de vida, y cuyos impactos medioambientales sean menos relevantes y más asumibles.

Entre las tecnologías que se pueden suponer bajas en emisiones encontramos tanto energías renovables, como la hidráulica, la eólica y la solar, como no renovables, como la nuclear. Dentro de este grupo, la solar se está erigiendo como una de las fuentes de energía más importantes a nivel mundial, y es que el crecimiento que ha experimentado a lo largo de la

última década supera el de cualquier otro tipo de tecnología de generación eléctrica renovable [8]. Se va a explicar a continuación cómo funciona y cuáles son los principios de la tecnología solar fotovoltaica, que es la que se va a tratar en este estudio.

1.2 ESTADO ACTUAL DE LA GENERACIÓN SOLAR

La generación solar fotovoltaica se está expandiendo a un ritmo sorprendente, pero no es una tecnología de última generación ni que se haya desarrollado en los últimos tiempos. De hecho, las primeras células solares se construyeron a finales del siglo XIX y la que se podría considerar la predecesora de las actuales fue patentada por Russell Ohl en 1946 [9]. La evolución actual se debe especialmente a que la tecnología se ha hecho mucho más competitiva a nivel económico y a la necesidad de descarbonizar los mixes de generación eléctrica.

1.2.1 POTENCIAL SOLAR

La generación fotovoltaica depende de la irradiación solar recibida por las células fotovoltaicas. Esta irradiación podrá variar mucho en función de la ubicación de la planta. En el siguiente mapa se puede apreciar cómo se distribuye el potencial de energía solar alrededor del mundo.

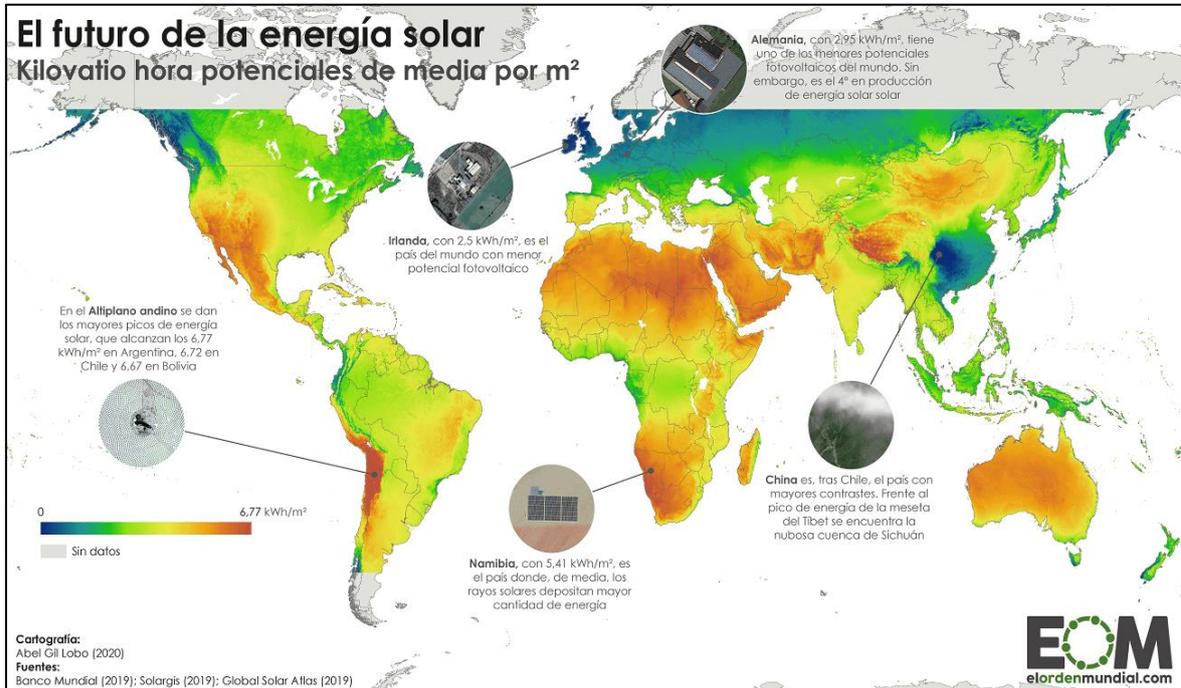


Figura 6: Mapa del potencial de la energía solar a nivel mundial. Fuente: EOM [10].

Como se puede observar en la Figura 6, se puede apreciar claramente como África, Oriente Medio y algunas regiones de Sudamérica son zonas donde la energía fotovoltaica podría tener unos altos rendimientos. Si no se han instalado grandes cantidades en estas zonas es por falta de recursos u otros factores como la accesibilidad, problemas en cadena de suministro, problemas sociales, etc. Se espera, debido a la reducción de precios y al creciente interés por estas tecnologías, que estos lugares van a experimentar un desarrollo notable en los próximos años o décadas.

1.2.2 CAPACIDAD INSTALADA

La energía fotovoltaica ha experimentado un notable crecimiento en las últimas décadas. La amenaza del calentamiento global y la reducción de los precios de los paneles han sido los dos factores principales que han promovido este crecimiento. A continuación, se muestra como la potencia fotovoltaica mundial instalada se ha multiplicado por 10 en la última década.

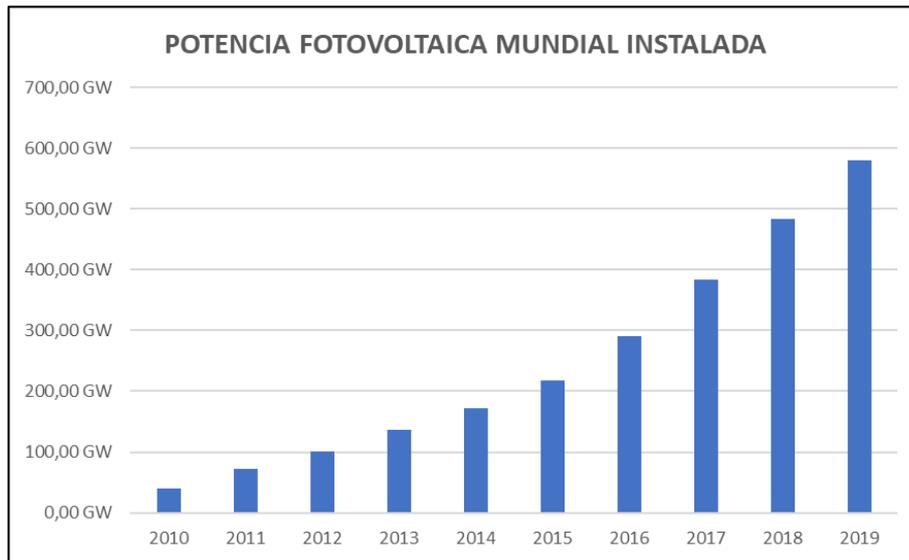


Figura 7: Evolución de la potencia fotovoltaica instalada a nivel mundial. Fuente: IRENA [8].

Sin embargo, este crecimiento no ha sido homogéneo geográficamente. Por regiones, Asia concentra más de la mitad de la potencia, seguida de Europa con un 22% y de América con un 15%. China es, con mucha diferencia, el país con más potencia instalada, y es que acapara un tercio de toda la potencia mundial. Le siguen Estados Unidos, Japón, algunos países de la Unión Europea, India, Australia y Corea del Sur [11].

Tabla 2: Países por potencia fotovoltaica instalada 2019. Fuente: UNEF.

	País	Potencia
	China	204,7 GW
	EE. UU.	75,9 GW
	Japón	63,0 GW
	Alemania	49,2 GW
	India	42,2 GW
	Italia	20,8 GW
	Australia	14,6 GW
	Reino Unido	13,3 GW
	Corea del Sur	11,2 GW
	Francia	9,9 GW

En la Tabla 2: Países por potencia fotovoltaica instalada 2019. Fuente: UNEF. Tabla 2 se puede apreciar que los grandes productores de energía solar son países desarrollados (EE.

UU., Japón, Alemania, etc.) o que tienen un poderío económico muy destacable (China e India).

1.2.3 POLÍTICAS DE INCENTIVOS A LA SOLAR

Las necesidades de descarbonización han hecho que muchos países adopten medidas para promover las energías bajas en emisiones y con pocos impactos ambientales. Estas medidas son cada vez menos relevantes en el caso de la solar fotovoltaica, ya que la reducción de precios de la misma ha hecho que ahora sea una tecnología competitiva en gran parte del mundo y que no necesite ayudas para funcionar.

Algunas de las políticas más aplicadas globalmente para impulsar la tecnología solar son:

Feed-in tariffs

Las feed-in tariffs son políticas destinadas a promover la inversión en renovables. Su objetivo es ofrecer una compensación de costes a los generadores de energía renovable, dándoles una fijación de precios y contratos a largo plazo. Los precios varían en función de la energía que se quiera promover y suelen ir reduciéndose gradualmente. Si no se planifican bien puede que el precio sea demasiado bajo para los inversores o demasiado alto para el estado.

Cuotas y obligaciones

Las autoridades pueden establecer objetivos medioambientales y establecer obligaciones medioambientales que empresas o individuos deben cumplir. Se diferencian objetivos por tecnología y sector. Su efectividad depende del contexto en el que se aplican y de las penalizaciones si no se cumplen los objetivos. En India se introdujeron este tipo de medidas en 26 estados, pero como no iban acompañadas de certificados o validaciones no fueron plenamente efectivas en un principio. Algunos países como Italia o Polonia han pasado de este tipo de medidas a otras más efectivas como las feed-in tariffs.

Subastas

En las subastas de energías renovables, al contrario que en la feed-in tariffs, “la industria determina el precio del proyecto a través de licitaciones competitivas entre los licitadores” [12]. Las subastas de energías renovables se están extendiendo rápidamente y están sustituyendo a las feed-in tariffs, sobre todo en proyectos a gran escala. Pueden asegurar transparencias y compromiso. Son muy útiles para averiguar el precio real de una tecnología.

1.3 GENERALIDADES DE LA GENERACIÓN SOLAR

El desarrollo de la tecnología solar se encuentra estrechamente relacionado con el tipo de generación que se desee implementar y las ventajas y desventajas de incluir esta tecnología en las redes eléctricas. Las siguientes secciones abordan este tema de forma general.

1.3.1 TIPOS DE GENERACIÓN SOLAR

Existen dos formas principales de aprovechar la radiación solar para generar energía eléctrica, que son:

- Energía termosolar: se aprovecha la radiación solar para calentar un fluido que se usa en un ciclo termodinámico de Rankine para generar electricidad.
- Energía fotovoltaica: se convierte directamente la energía solar en energía eléctrica mediante el efecto fotovoltaico.

Esta última está teniendo un desarrollo muy superior a la termosolar y generó en el año 2019 44 veces más energía eléctrica [7].

1.3.2 VENTAJAS E INCONVENIENTES DE LA GENERACIÓN SOLAR

La principal ventaja que presenta la tecnología fotovoltaica es, por supuesto, que no genera emisiones de GEI durante su operación, lo que hace de ella una herramienta fundamental en la lucha contra el cambio climático.

Por otro lado, es una tecnología que puede implementarse a distintas escalas, lo que hace que sea factible tanto construir grandes plantas con miles de paneles fotovoltaicos como instalaciones de menor envergadura que se puedan utilizar a nivel residencial.

Los elevados costes han sido el principal hándicap que ha afrontado esta energía desde sus inicios. Sin embargo, los precios de esta tecnología ahora son mucho más competitivos, y la agencia Lazard estima que en el periodo 2009-2020 los precios han caído un 90% [13] [14]. De hecho, la AIE ha llegado a decir en el World Energy Outlook de 2020 que “para los proyectos con financiación de bajo costo que aprovechan recursos de alta calidad, la energía solar fotovoltaica es ahora la fuente de electricidad más barata de la historia” [7]. En la Figura 8, que se muestra a continuación, se puede ver de manera clara como los precios de las subastas de fotovoltaica han ido cayendo de manera drástica los últimos años, llegando incluso a casi alcanzar los de la eólica terrestre, que eran tres veces menores hace una década.

Figure 1.2 Global weighted average prices resulting from auctions, 2010-2018, and capacity awarded each year

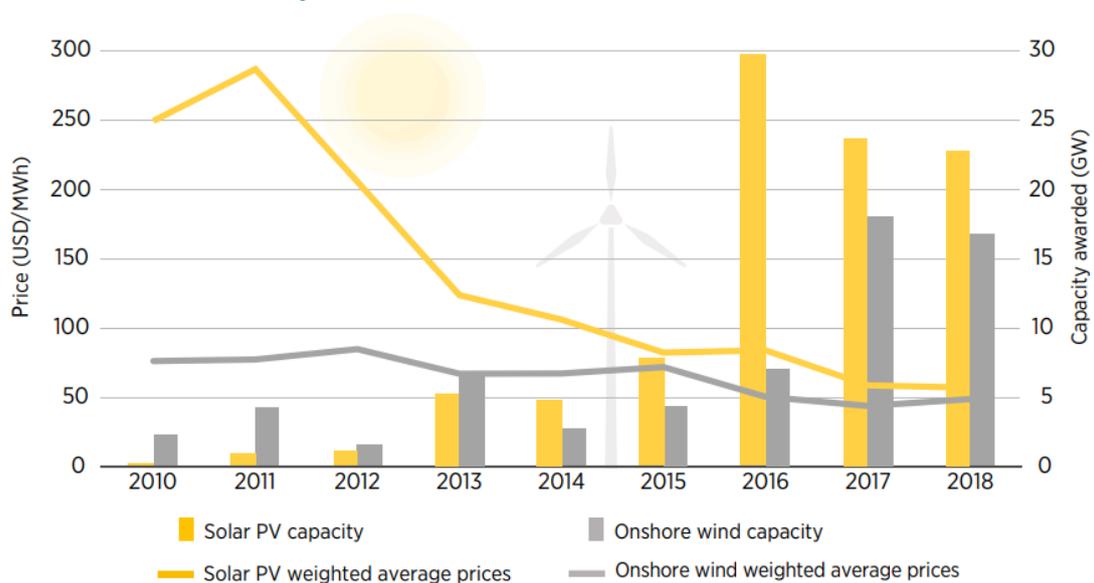


Figura 8: Evolución de los precios de las subastas de fotovoltaica y capacidad instalada. Tomado de [15].

El precio medio se situó en 2018 un poco por encima de los 50\$, pero en ciertos lugares del mundo se han alcanzado mínimos espectaculares. En Portugal y España se han llegado a dar

precios de 11 y 15 €/MWh respectivamente, mientras que en otros países como Chile o Arabia Saudí también se ha conseguido bajar de los 20 €/MWh. Esto, unido a que, en muchos lugares del mundo, y concretamente de Europa, se están eliminando las subvenciones a la generación con combustibles fósiles y se están elevando los precios de derechos de emisiones, hace la energía fotovoltaica más atractiva todavía.

En cuanto a los inconvenientes, es necesario recordar que la fotovoltaica también tiene impactos ambientales. Necesita muchos recursos minerales y las grandes plantas pueden provocar un elevado impacto paisajístico. Además, esta tecnología necesita mucha más superficie para generar la misma cantidad de energía que, por ejemplo, una central de carbón o de ciclo combinado.

El otro gran problema es la intermitencia en la generación. La fotovoltaica no es gestionable, y sólo produce electricidad durante el día, por lo que no puede adecuarse a la demanda en cada momento. Se necesitaría una gran capacidad de almacenamiento para sostener un mix eléctrico basado mayoritariamente en energía fotovoltaica, y estas soluciones, al menos hoy en día, están limitadas o son demasiado costosas. Si no se tiene esta capacidad de almacenamiento pueden llegar a suceder situaciones como las que han ocurrido en California en 2018-2020, que se tengan que desconectar muchas plantas fotovoltaicas durante el día por exceso de generación o que se tenga que deslastrar demanda por insuficiencia [16].

Para paliar el problema de la intermitencia de la generación, es necesario elaborar predicciones fiables de generación fotovoltaica a diferentes escalas temporales. Las predicciones tienen utilidad para muchos agentes del mercado. Por ejemplo, ayudan a planear la infraestructura de la generación a medio y largo plazo, consiguiendo así un mix de generación balanceado, dan soporte para la toma de decisiones que realizan los agentes en lo relativo a la operación del sistema, y cubren la necesidad que tienen los inversores de tener una visión de futuro del mercado [17].

1.4 MOTIVACIÓN

Una vez visto cómo funciona la tecnología fotovoltaica, cómo está siendo su expansión y cuáles son los principales métodos de predicción que existen, se va a hablar de la motivación de este proyecto, así como de su alcance y objetivos principales.

Las predicciones fiables de energía solar fotovoltaica a largo plazo son importantes por dos motivos principales, a saber:

- Que la fotovoltaica es una tecnología en expansión en todo el mundo y que cada vez tiene más peso en los mixes energéticos.
- Que su generación es intermitente y que no existe ahora mismo una capacidad importante de almacenamiento.

Por estos motivos, el número de artículos en la literatura académica sobre este tema no ha dejado de crecer en los últimos años.

Al hacer una revisión extensa del estado del arte, se ha encontrado que hay una carencia de estudios sobre los métodos de predicción a largo plazo. La mayor parte de los *papers* se centran en el muy corto, corto y medio plazo, que son útiles sobre todo para la operación de los parques fotovoltaicos. Sin embargo, las predicciones a largo plazo también son tremendamente útiles. Ayudan a los agentes de mercado en su operación y búsqueda de maximización de sus beneficios y permiten gestionar mejor el riesgo, ya que las predicciones de fotovoltaica, combinadas con otros factores, pueden llevar a una predicción de los precios de la electricidad. También son muy valiosas para los operadores del sistema y de las redes de distribución, debido a que pueden conocer mejor qué infraestructuras nuevas necesitan para que la red siga siendo estable.

Otra de las grandes carencias que se ha encontrado en el estado del arte es que existen muy pocos trabajos que hagan predicciones para grandes áreas geográficas, como países. Lo más común es que los trabajos se centren en una instalación fotovoltaica concreta. Esto tiene cierto sentido, ya que cuando se trabaja con grandes áreas, la distribución de las variables

medioambientales no suele ser homogénea. Si se quiere elaborar una predicción de energía solar para Estados Unidos, la radiación solar o la presencia de nubes no va a ser igual en California que en Minnesota.

Tampoco se han encontrado suficientes estudios que elaboren predicciones en contextos multiárea, es decir, entre áreas que están cercanas geográficamente y que tienen relaciones entre sí. Estos casos de estudio son interesantes, ya que las predicciones de una de las áreas pueden afectar en gran medida a las otras.

En este proyecto se desarrollará una metodología robusta que permita generar escenarios de energía solar fotovoltaica realistas para contextos multiárea de mercados eléctricos interconectados, teniendo en cuenta las dependencias espacio temporales existentes entre diferentes zonas. Es fundamental incluir estas dependencias en el proceso, debido a la cada vez mayor interconectividad que existe entre distintos países. Se ha elegido Europa como zona de estudio por ser la segunda región del mundo por capacidad fotovoltaica instalada [11], sólo por detrás de China, y porque desde la propia Comisión Europea se están estableciendo objetivos de mínima generación renovable que son un impulso para este tipo de tecnología. Según el escenario de desarrollo sostenible del WEO 2020 la potencia total instalada podría triplicarse en esta región hasta alcanzar un 25% del total [7].

Dentro de Europa, debido a la dificultad de trabajar con todos los mercados eléctricos a la vez, se han escogido como áreas de trabajo España, Portugal y Francia. Estas tres áreas juntas suponen más de un 15% de la potencia instalada en la UE, están interconectadas entre sí y tienen distintos grados de penetración de la tecnología solar, por lo que el caso de estudio se considera suficientemente diverso y relevante.

1.5 OBJETIVOS

Los objetivos principales de este proyecto son:

- Hacer una revisión del estado del arte de métodos de generación de escenarios de energía solar.

- Hacer una revisión de los Planes Nacionales de Energía y Clima (NCEP, por sus siglas en inglés) de las áreas de estudio.
- Caracterizar el comportamiento de la generación de energía solar a diferentes escalas temporales.
- Desarrollar una metodología robusta que permita generar escenarios de energía solar realistas para diferentes mercados con nivel de detalle horario, teniendo en cuenta las dependencias espacio temporales existentes entre diferentes países.
- Aplicar la metodología desarrollada en un caso de estudio de gran alcance centrado en diferentes mercados europeos, de forma que se valide su robustez y capacidad para capturar la dinámica de la generación solar.

2. ESTADO DE LA CUESTIÓN

2.1 ASPECTOS BÁSICOS DE LAS TÉCNICAS DE PREDICCIÓN

La predicción es una herramienta que está muy presente en el día a día y que se puede aplicar en muchos campos de estudio. Quizás las predicciones que la gente tiene más presentes son las meteorológicas, pero también se utilizan en otras áreas como la economía, la energía o incluso el ocio.

En este apartado, se pretende dar respuesta de manera concisa a preguntas como qué fenómenos se pueden predecir, qué métodos de predicción existen y cuáles son las particularidades que tiene la predicción para el caso de este estudio, la generación fotovoltaica.

Para explicar la predicción en general se ha tomado como referencia el libro “*Forecasting: Principles and Practice*” de Rob J. Hyndman y George Athanasopoulos, uno de los documentos de referencia en cuanto a predicción se refiere.

Para empezar, la predictibilidad de un evento, es decir, la capacidad de averiguar con antelación lo que va a ocurrir, va a depender sobre todo de 3 factores:

- Lo bien que se comprendan los factores que contribuyen a ese evento.
- La cantidad de datos disponibles referentes a ese evento.
- Si la predicción puede afectar a la variable que intentamos predecir.

Por ejemplo, la demanda eléctrica de un territorio se puede predecir con relativa facilidad, ya que se satisfacen esas tres condiciones. Factores como el día de la semana o la temperatura tiene una gran correlación con la demanda, suele haber exceso de datos y la predicción no va a afectar de ninguna manera a los hábitos de consumo eléctrico de la sociedad. En cambio, sería muy difícil predecir con precisión fenómenos como, por ejemplo, el valor en bolsa de una empresa, ya que, aunque puede haber una gran cantidad de datos disponible, no se

comprende bien qué factores afectan a este valor y la predicción sí que puede alterar el mercado bursátil.

Suele asumirse de manera errónea que no se pueden elaborar predicciones en entornos cambiantes, pero esto no es así. Un buen modelo de predicción es capaz de averiguar cómo está cambiando el entorno y extender este cambio a futuro. Eso sí, en entornos muy cambiantes y con grandes fluctuaciones, las predicciones van a ser también muy cambiantes y con grandes fluctuaciones.

En cuanto a los horizontes temporales, las predicciones se pueden dividir según el tiempo que abarcan en:

- Predicciones a corto plazo: principalmente se necesitan para la programación de la producción, el personal y el transporte.
- Predicciones a medio plazo: son necesarias para determinar los recursos necesarios, como materiales, maquinaria, personal, etc.
- Predicciones a largo plazo: fundamentales para las decisiones estratégicas.

El tiempo que abarque cada una de estas categorías dependerá principalmente del problema a estudiar. De hecho, en algunas predicciones relativas a la energía solar se distingue entre cuatro horizontes de tiempo y no entre tres, que son el muy corto, el corto, el medio y el largo plazo [18]. La elección del método de predicción dependerá del horizonte de tiempo de esta.

De manera general, se pueden distinguir dos grandes métodos de predicción, que dependerán en gran medida de la cantidad de información que haya disponible. Estos métodos son:

- Métodos cualitativos: se utilizan cuando no hay datos disponibles o estos son irrelevantes. Se basan en conocimiento avanzado para elaborar las predicciones (método Delphi).
- Métodos cuantitativos: para poder utilizar estos métodos se han de cumplir dos condiciones.
 - Que haya datos suficientes.

- Que se pueda asumir que los patrones del pasado se van a repetir en el futuro.

La mayoría usan series temporales en las que los datos se toman a intervalos de tiempo regulares.

Existen varios métodos que se pueden agrupar dentro de esta última categoría y que van a ser explicados posteriormente.

En cuanto a la forma de elaborar una predicción, normalmente se siguen 5 pasos:

1. Definición del problema: requiere entender la manera en la que se va a utilizar la predicción, quién necesita la predicción y cómo puede ayudar a la organización que la necesita.
2. Adquisición de información: engloba tanto la adquisición de los datos como el conocimiento necesario para llevar a cabo la predicción.
3. Análisis exploratorio preliminar: estudiar la tendencia, la estacionalidad, la presencia de ciclos, correlación entre variables, etc.
4. Elección y ajuste de los modelos: como se ha mencionado anteriormente, la elección del modelo dependerá de numerosos factores, como la disponibilidad de los datos, la relación entre la variable a predecir y los predictores o la manera en la que se va a emplear la predicción.
5. Uso y evaluación del modelo de predicción: una vez se ha escogido el modelo y se han ajustado los parámetros del mismo, será necesario evaluar su calidad. Existen numerosos métodos para determinar si un modelo es adecuado o no.

2.2 MÉTODOS DE PREDICCIÓN PARA ENERGÍA FOTOVOLTAICA

En este apartado sólo se van a considerar los métodos de predicción que se suelen utilizar para el caso de la generación solar.

Existen varios métodos de predicción de escenarios de energía fotovoltaica, y se deberá escoger el método en función del problema que se quiera estudiar. Cada uno tiene sus

fortalezas y debilidades que lo hacen adecuado o no según el tipo de circunstancias. Se pueden dividir los métodos de predicción en 3 categorías fundamentales [19]:

- Modelos estadísticos.
- Modelos físicos.
- Modelos combinados.

Modelos estadísticos

Estos modelos elaboran sus predicciones a partir de series temporales que recogen datos meteorológicos o de generación eléctrica. Lo que se busca es extrapolar los datos pasados para predecir, con mayor o menor acierto, los futuros. Para elaborar predicciones sobre generación fotovoltaica, los datos de entrada de las series temporales pueden ser variables atmosféricas (temperatura, irradiación solar, ángulo de incidencia del sol, etc.) y/o directamente datos generación solar [19].

Existen varios modelos estadísticos que se han utilizado con éxito en las predicciones de generación fotovoltaica, como, por ejemplo:

Redes neuronales

Según el IBM *Knowledge Center*, “una red neuronal es un modelo simplificado que emula el modo en que el cerebro humano procesa la información. Funciona simultaneando un número elevado de unidades de procesamiento interconectadas que parecen versiones abstractas de neuronas” [20]. Se parece al cerebro humano en dos aspectos [21]:

- La red adquiere conocimiento desde su entorno a través de un proceso de aprendizaje.
- Los conocimientos adquiridos se almacenan en los puntos fuertes de la conexión interneuronal, conocidos como pesos sinápticos.

El modelado de las redes neuronales se puede dividir en tres etapas:

1. Etapa de diseño: elegir el tipo de red neuronal, el número de “neuronas” o nodos por capa, el número de capas, los parámetros de entrada y salida, el entrenamiento y la validación de los datos.
2. Etapa de entrenamiento: se ajustan los valores de los parámetros.
3. Etapa de validación: se hace a partir de datos distintos a los usados en la fase de entrenamiento. Se analiza el grado de precisión conseguido.

Support Vector Machine

Se trata de una técnica de Machine Learning creada originalmente para problemas de clasificación y cuya aplicación a la regresión se conoce como Support Vector Regression (SVR). Es un algoritmo de regresión no lineal en el que la entrada se sitúa en un espacio de mayor dimensión mediante un mapeo no lineal para realizar posteriormente una regresión lineal en ese espacio [22]. De manera más sencilla, lo que se consigue es transformar una regresión no lineal en una lineal. Han mostrado un gran rendimiento comparados con otros métodos como las redes neuronales u otros modelos estadísticos tradicionales. Evitan el overfitting, descartan ajustes iterativos de los parámetros, requieren pocos núcleos, son computacionalmente más rápidos y tienen una buena generalización y convergencia.

Cadenas de Márkov

Las cadenas de Márkov son aquellas en las que la probabilidad de que ocurra un estado sólo depende del estado anterior. El proceso se define mediante el estado inicial, el número de estados y la matriz de transición que describe las posibilidades de cambio. Se han encontrado trabajos en la literatura en los que este método supera a otros como el SVR para predicciones de muy corto plazo [19].

Modelos ARIMA

Estos modelos tratan de buscar autocorrelaciones en los datos y replicarlas para obtener la predicción. Se podrían englobar dentro de los modelos regresivos. Están formados por:

- Un modelo autorregresivo (AR), en el que el regresor es la variable a predecir, pero desfasada en el tiempo.
- Un modelo de media móvil (MA), que es una autorregresión en el que el regresor son los errores pasados de predicción.
- Una parte integrada (I), que permite trabajar con series no estacionarias.

También es posible incluir una parte estacional si la serie temporal tiene alguna estacionalidad muy marcada. En ese caso hablaríamos de modelos SARIMA.

Este tipo de modelo estadístico va a ser el que se emplee en este trabajo. Se puede encontrar más información en el Anexo III

Modelos físicos

Estos métodos elaboran sus predicciones a partir de variables atmosféricas como la cantidad de aerosoles en la atmósfera, la presencia de nubes, el ángulo con el que la luz del sol alcanza la superficie, etc. Se pueden subdividir en [19]:

Predicción numérica del tiempo

El objetivo es proporcionar información sobre las condiciones atmosféricas en una cierta escala temporal. Este método puede ser subdividido en dos categorías: global y local. Los modelos globales simulan las condiciones atmosféricas a escala global mientras que los locales lo hacen para áreas más limitadas. Suelen tener un límite de tiempo de 15 días. Estos modelos de NWP están basados en la integración numérica de ecuaciones diferenciales acopladas que explican el mecanismo de transporte de la radiación y la dinámica de la atmósfera.

Imágenes celestes

En este método una cámara digital va tomando imágenes del cielo constantemente. De esta forma, se puede detectar la aparición de nubes, medir su altura sobre el suelo y determinar

sus movimientos. A partir de la localización de las nubes y su movimiento se hacen predicciones a corto plazo de radiación solar y potencia fotovoltaica.

Imágenes de satélite

Muy parecido al método anterior. Los patrones de nubes se determinan mediante imágenes tomadas por satélites. Las imágenes se mezclan para generar campos vectoriales de movimiento de nubes que puede ser utilizado para pronosticar las futuras ubicaciones de las nubes. Estos métodos se utilizaron eficazmente en las predicciones de irradiación en el rango de 1-min a 5-h adelante.

Métodos combinados

Estos métodos pueden resolver las debilidades de los métodos estadísticos y físicos, y mejorar sus aspectos positivos y precisión. Pueden ser clasificados en dos categorías [23]:

Previsión cooperativa

La predicción se divide en subtarear para las que se seleccionan determinados predictores. La predicción final será la suma de las salidas de las distintas subtarear.

Previsión competitiva

En este caso, las variables predictoras se entrenan individualmente con distintos conjuntos de datos o con el mismo conjunto y distintos parámetros. Después se escogen aquellos que hayan dado mejores resultados y la predicción se obtendrá de la media de los resultados de cada variable.

2.3 ESTADO DEL ARTE

En este apartado se va a hacer un resumen de los principales artículos académicos que existen en los que se intenta hacer predicciones de energía fotovoltaica mediante el empleo de modelos ARIMA o similares (ARMA, SARMA, AR, etc.).

Para empezar, se van a definir los distintos horizontes temporales que existen para la predicción de energía fotovoltaica. Se ha tomado como referencia la clasificación hecha en [19], que distingue 4 tipos de horizontes:

- Muy corto plazo: desde unos segundos hasta 30 minutos.
- Corto plazo: desde 30 minutos hasta 6 horas.
- Medio plazo: desde 6 horas hasta un día, aunque también se puede llegar hasta una semana o incluso un mes.
- Largo plazo: de un día en adelante, aunque también se puede considerar que va desde un mes hasta un año o más [22].

Empezando por los estudios a corto plazo, que son mayoría, en el capítulo 15 del libro “Solar Energy Forecasting and Resource Assessment” de Jan Kleissl se menciona el estudio de Pedro y Coimbra [24] en el que se hace una comparativa entre varios métodos de predicción para un problema concreto. Se trataba predecir la generación fotovoltaica en una planta de 1 MW en Merced con dos horizontes temporales distintos, de una y de dos horas. Para hacer la predicción sólo se tenían disponibles los datos horarios de generación de dicha instalación de noviembre de 2009 a agosto de 2011, sin ninguna variable exógena. Los métodos de predicción que se compararon fueron:

- *Clear Sky index persistence*
- Modelo ARIMA
- k-Nearest Neighbour (kNN)
- Red neuronal artificial (ANN)
- Algoritmos genéticos combinados con ANN (GA/ANN)

Las métricas de error con las que se compararon los resultados fueron el Error Medio Absoluto, el Error de Sesgo Medio, la Raíz del Error Cuadrático Medio (MAE, MBE, RMSE, por sus siglas en inglés) y R^2 . En general los dos métodos que empleaban redes neuronales tuvieron mejor resultados en casi todas las métricas, excepto en el MBE, donde el modelo ARIMA se impuso en el horizonte temporal de una hora. Ahora bien, las redes neuronales son significativamente más difíciles y menos interpretables que los modelos

ARIMA y, además, están especialmente indicadas para el corto plazo. El modelo ARIMA consiguió mejorar sus resultados para el horizonte de tiempo de dos horas.

Además de este, existen numerosos artículos en los que se compara el rendimiento entre distintos métodos de predicción. Investigadores de la Universidad De Texas A&M quisieron hacer una previsión de la generación fotovoltaica de una planta de 20 MW situada en China. Para ello emplearon dos métodos distintos, una red neuronal artificial y un modelo ARIMA. El modelo ARIMA tenía parte estacional, por lo que sería más apropiado hablar de un modelo SARIMA. Para este dato se tenían datos de generación de 9 meses de la instalación cada 15 minutos, que se pasaron a horarios haciendo una media. Sólo se tuvo en cuenta la generación entre las 6:00 y las 19:00, desechando las horas nocturnas en las que la generación era nula o prácticamente. Se probaron varios modelos distintos, concretamente un ARMA, un ARIMA y un SARIMA, a diferentes escalas temporales. Para el orden de los modelos se inspeccionó visualmente el ACF y el PACF y finalmente se escogió el modelo que más minimizase el AIC. Los resultados fueron que el modelo más preciso fue el SARIMA, aunque el tiempo de computación de este también fue muy grande, y que la previsión era menos precisa si se aumentaba el horizonte temporal. En cuanto a la comparación con las redes neuronales, estas últimas dieron mejores resultados, pero en el artículo se especifica que ambos modelos son fiables.

En uno de los trabajos que se han encontrado se trataba de realizar predicciones de energía solar de una planta concreta ajustando un modelo ARMA para cada hora del día [25]. Se tenían datos de generación horarios de dicha planta que se habían recogido durante 9 meses. Sólo se consideraba la generación entre las 6:00 y las 19:00, el resto de los datos se descartaban. Una vez ajustados los 14 modelos ARMA, se realizaban las simulaciones de Monte Carlo para generar los distintos escenarios. Los resultados que se obtuvieron fueron relativamente buenos si se tiene en cuenta la simplicidad del modelo.

En cuanto al largo plazo, en otro trabajo llevado a cabo por investigadores norteamericanos se usó un modelo ARIMA para elaborar una predicción a 30 días de la generación de una instalación fotovoltaica de 10 kW situada en Texas [26]. En este caso, los datos de

generación no eran horarios, sino diarios, por lo que no se tuvo que modelar la estacionalidad diaria. Los datos históricos abarcaban un año entero. Se hizo una diferenciación para estacionarizar la serie, se hizo un test aumentado de Dickey-Fuller para comprobar que la serie temporal era estacionaria y se escogió el orden del modelo a partir del ACF y el PACF. Los parámetros del modelo se estimaron usando el AIC y la suma de errores cuadrados (SSE, por sus siglas en inglés). Los errores en los resultados fueron un poco elevados y los autores afirmaron que existen muchas posibles áreas de mejora.

En [27] se hace una predicción de energía solar a largo plazo. No se usan modelos ARIMA, pero la forma en la que se convierte la serie original en estacionaria es muy interesante. En este artículo se tiene como dato de entrada la radiación solar horizontal (GHI, por sus siglas en inglés) en un lugar determinado. Se usa un índice Clear Sky para transformar la serie. Este índice vendría a representar la radiación solar si las condiciones atmosféricas fuesen las mejores posibles, por lo que lograría capturar toda la información determinista de la serie temporal. Si se quita de la serie original, se tendría una serie estacionaria, tal y como se muestra en la Figura 9.

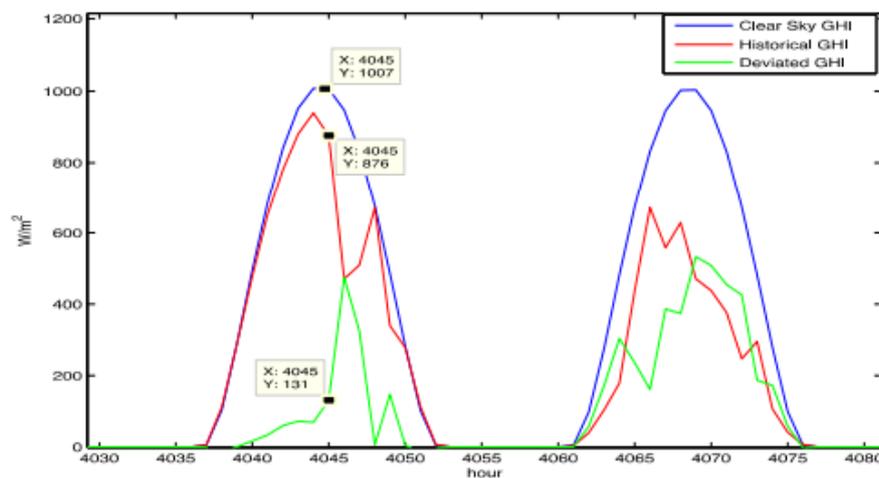


Fig. 4 GHI deviation between clear sky GHI and historical GHI

Figura 9: Clear Sky Index. Tomado de [27].

Al hacer una revisión extensa del estado del arte, se ha encontrado que hay una carencia de estudios sobre los métodos de predicción a largo plazo. La mayor parte de los papers se

centran en el corto y medio plazo, que son útiles sobre todo para la operación de los parques fotovoltaicos. Sin embargo, las predicciones a largo plazo también son tremendamente útiles para inversores, operadores de sistema, empresas que tienen activos en generación y otros agentes de mercado, como se ha mencionado en la sección 1.4.

Otra de las grandes carencias que se ha encontrado en el estado del arte, que ha sido mencionada previamente, es que existen muy pocos trabajos que hagan predicciones para grandes áreas geográficas, como países. Lo más común es que los trabajos se centren en una instalación fotovoltaica concreta. Esto tiene cierto sentido, ya que cuando se trabaja con grandes áreas, la distribución de las variables medioambientales no suele ser homogénea. Si se quiere elaborar una predicción de energía solar para Estados Unidos, la radiación solar o la presencia de nubes no va a ser igual en California que en Minnesota. Si se cogen valores medios para las variables de entrada se puede estar incurriendo en errores muy graves.

En este proyecto, se va a realizar una generación de posibles escenarios de energía solar con un horizonte temporal superior a un año para tres grandes áreas: España, Portugal y Francia. Además, como los sistemas eléctricos de estos tres países están interconectados entre sí, habrá dependencias entre ellos que se deberán tener en cuenta. Como única variable de entrada se va a utilizar los datos horarios de generación fotovoltaica en cada uno de estos tres países en por unidad (p.u.).

2.4 COMPARATIVA ENTRE MODELOS

No hay ningún modelo que sea incuestionablemente mejor que los demás. Todos los métodos de predicción pueden ser útiles en ciertas circunstancias y cada uno tiene sus puntos fuertes y débiles. Por ejemplo, los métodos físicos que captan imágenes para elaborar las predicciones suelen ser buenos para un horizonte temporal corto (menos de 1 hora) y para áreas pequeñas, mientras que los NWP pueden funcionar bien a largo plazo y para áreas grandes, e incluso a nivel global [23]. Los métodos estadísticos se suelen utilizar para áreas pequeñas y predicciones intradiarias. Sin embargo, también han mostrado buenos resultados cuando han sido utilizadas en grandes extensiones y a largo plazo [28]. Además, estos

modelos estadísticos son capaces de capturar las dependencias espacio-temporales que existen cuando se trabaja en contextos multiárea.

2.5 ESTADO DE LA ENERGÍA SOLAR EN LAS ÁREAS DE ESTUDIO

2.5.1 PANORAMA EUROPEO

Como se ha mencionado anteriormente, se ha escogido España, Portugal y Francia como zona de estudio por ser países importantes dentro de la segunda región del mundo con más capacidad fotovoltaica instalada, la UE. Asimismo, las instituciones europeas han establecido objetivos de descarbonización muy ambiciosos que pasen por promover las energías renovables, entre ellas la fotovoltaica. Se muestra a continuación cómo ha evolucionado la distribución de la fotovoltaica instalada en la UE por países.

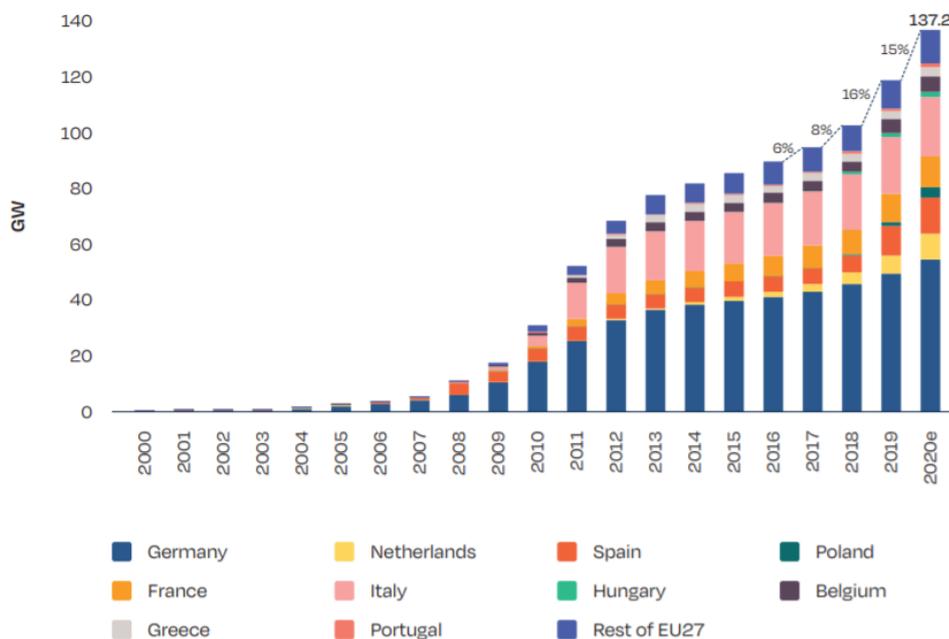


Figura 10: Capacidad fotovoltaica acumulada 2000-2020. Tomado de [29].

La Figura 10 muestra que Alemania es el país con mayor capacidad instalada de toda la UE, seguido de Italia, España, Francia y Países Bajos.

En cuanto a los planes a futuro, se ha realizado un estudio de los National Energy and Climate Plan (NECP o Planes Nacionales de Energía y Clima en español) de cinco de las mayores potencias de la UE: Alemania, Francia, Italia, España y Portugal. Todos estos países apuestan fuertemente en sus planes por la energía fotovoltaica y establecen mecanismos para aumentar su cuota dentro de los mixes. Las subastas aparecen en los NECP como la herramienta favorita para promover la fotovoltaica y la mayoría de los países tienen establecido un calendario para los próximos años.

Los objetivos de potencia fotovoltaica instalada en 2030 por país son:

Tabla 3: Objetivos de capacidad fotovoltaica instalada 2030. Nota: España escenario objetivo y Francia escenario B para 2028.

Tecnología	España	Alemania	Francia	Italia	Portugal
<i>Solar fotovoltaica</i>	39.18 GW	98.00 GW	44.00 GW	51.12 GW	9.00 GW

Debido a la complejidad de hacer este proyecto con todos los países de la Unión Europea, se ha decidido escoger una parte del territorio para hacer el estudio. Como se ha mencionado antes, se han escogido los mercados de España, Portugal y Francia. Estos países suponen alrededor de un 18% de la capacidad fotovoltaica total de la UE, tienen diferencias geográficas que hacen el estudio interesante, tienen sistemas eléctricos interconectados entre sí y el desarrollo de la tecnología ha sido distinto en cada zona. Por estos motivos, se considera que el estudio es diverso y que es una muestra representativa de lo que sucede en la UE.

A continuación, se va a hacer una breve descripción de cada una de las tres áreas de estudio. Se describirán los mixes energéticos, cuál es la situación actual de la tecnología fotovoltaica en el país y los planes de energía y clima que estos países han hecho para los próximos años.

2.5.2 ESPAÑA

2.5.2.1 Perspectiva general

El país está situado en la Península Ibérica, al suroeste de Europa. Tiene una extensión de aproximadamente 500.000 km² y una población de 47 millones de habitantes. En el año 2020 se consumieron en el país 249,82 TWh de energía eléctrica [30]. Esta cifra convierte al país en el cuarto de la UE que más energía consume, sólo por detrás de Alemania, Francia e Italia. Tiene un mix eléctrico muy diversificado, en el que no hay ninguna tecnología que predomine claramente sobre el resto. La potencia instalada a finales de 2020 era de 110,423 GW y se distribuía como se muestra en la Figura 11.

Desglose de potencia instalada en España a 31/12/2020

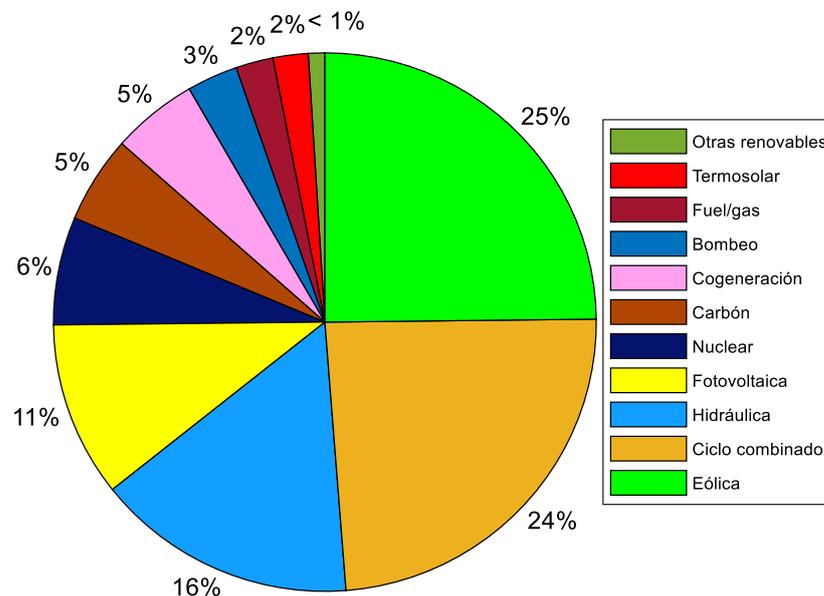


Figura 11: Capacidad instalada en España 2020. Fuente: REE.

La eólica y los ciclos combinados son las tecnologías más presentes en el mix, seguidas de la hidráulica, la solar fotovoltaica y las centrales térmicas de carbón.

Balace de la generación en España en 2020

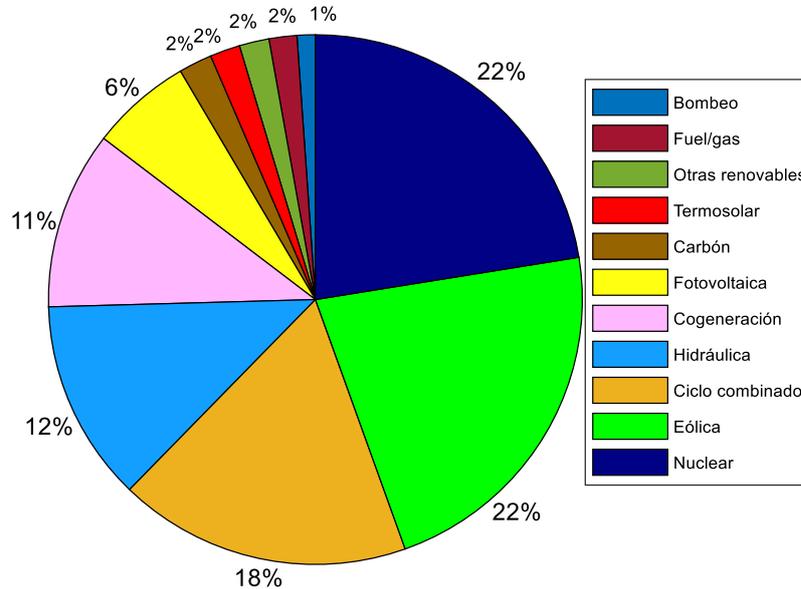


Figura 12: Generación eléctrica en España en 2020. Fuente: REE.

Si se comparan la Figura 11 y la Figura 12 se observará que las tecnologías con más capacidad instalada no coinciden siempre con las que más generan. La nuclear, que fue la quinta tecnología en cuanto a potencia, es la que más generó en 2020 dentro del mix, debido a su alto factor de carga. En el ranking de generación le seguirían la eólica, los ciclos combinados, la hidráulica, la cogeneración y la solar fotovoltaica.

2.5.2.2 Energía solar

El incremento y la distribución de la capacidad fotovoltaica en el país se muestran a continuación:

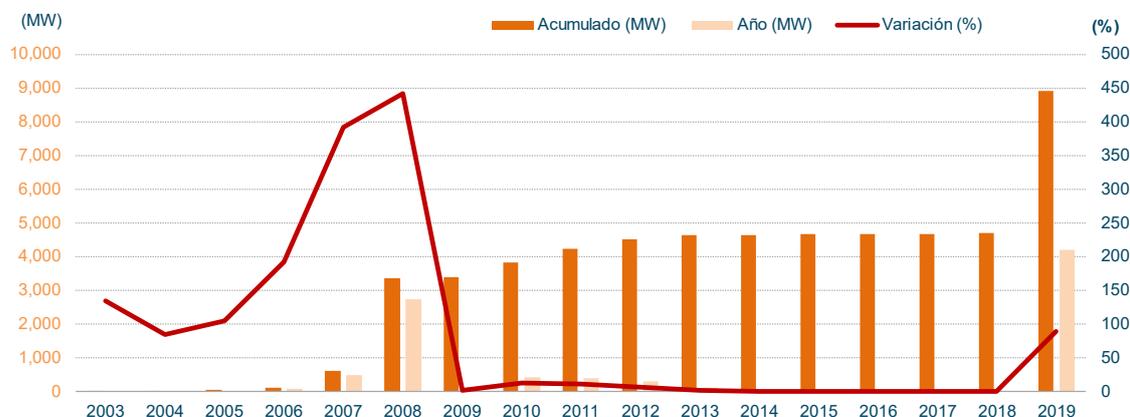


Figura 13: Evolución de la capacidad fotovoltaica instalada en España 2003-2019. Tomado de REE [31].

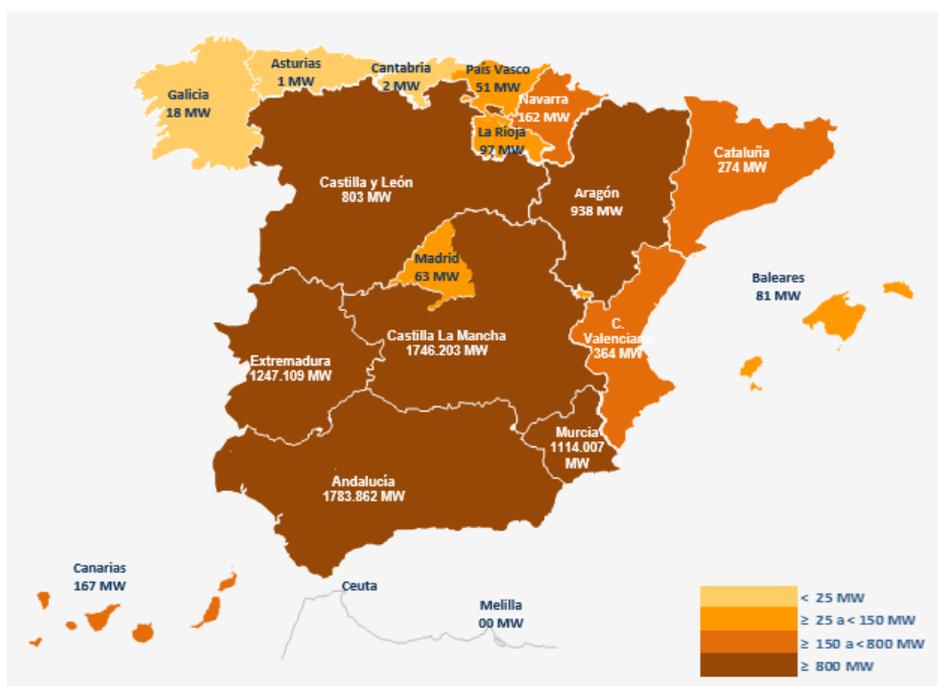


Figura 14: Capacidad fotovoltaica en España por comunidades autónomas a 31/12/2019. Tomado de REE [31].

En la Figura 13 se puede comprobar como la potencia fotovoltaica permaneció prácticamente constante de 2013 a 2018. Este parón se debió a los efectos de la crisis económica de 2008 y a que la tecnología todavía era más costosa en comparación con otras más tradicionales. La capacidad se incrementó notablemente en 2019. Esta subida se

concentró en los últimos meses del año. En cuanto a la distribución de las instalaciones, estas están más presentes en la parte sur de la península, donde el recurso solar es mayor.

2.5.2.3 Planes futuros

En su National Energy and Climate Plan (NECP) o Plan Nacional Integrado de Energía y Clima (PNIEC), los objetivos que se marca España para el año 2030 son:

- 39,5% de mejora de la eficiencia energética.
- 42% de renovables sobre el uso final de la energía.
- 74% de renovables en la generación eléctrica.
- 10% de grado de interconexión eléctrica.
- 23% de reducción de emisiones de GEI respecto a los niveles de 1990.
- 61% de dependencia energética.

El objetivo de reducción de emisiones quedaría por debajo del 40% establecido por la UE, aunque el documento señala que esta reducción estaría en línea con lo demandado por el IPCC.

En cuanto a la energía fotovoltaica, en el escenario tendencial, que es el más conservador, se espera que llegue a los 18.921 MW en 2030, mientras que el escenario objetivo sitúa esta cifra en 39.181 MW. Las subastas se mencionan como uno de los instrumentos más adecuados para aumentar la capacidad fotovoltaica.

2.5.3 PORTUGAL

2.5.3.1 Perspectiva general

Al igual que España, se encuentra situado en la Península Ibérica, al suroeste de Europa. Tiene una extensión de algo menos de 100.000 km² y una población aproximada de 10 millones de personas. El consumo eléctrico del país en el año 2020 fue de 48,8 TWh [32]. Su mix eléctrico también es bastante diversificado, y la única diferencia importante con respecto a las otras dos áreas de estudio es que no existe ningún reactor nuclear en el país. La capacidad instalada en el país y la producción eléctrica en los años 2011-2020 se muestran

en la **¡Error! No se encuentra el origen de la referencia.** y en la **¡Error! No se encuentra el origen de la referencia.**
origen de la referencia..



Figura 15: Capacidad instalada en Portugal por tecnología 2011-2020. Tomado de REN [32].

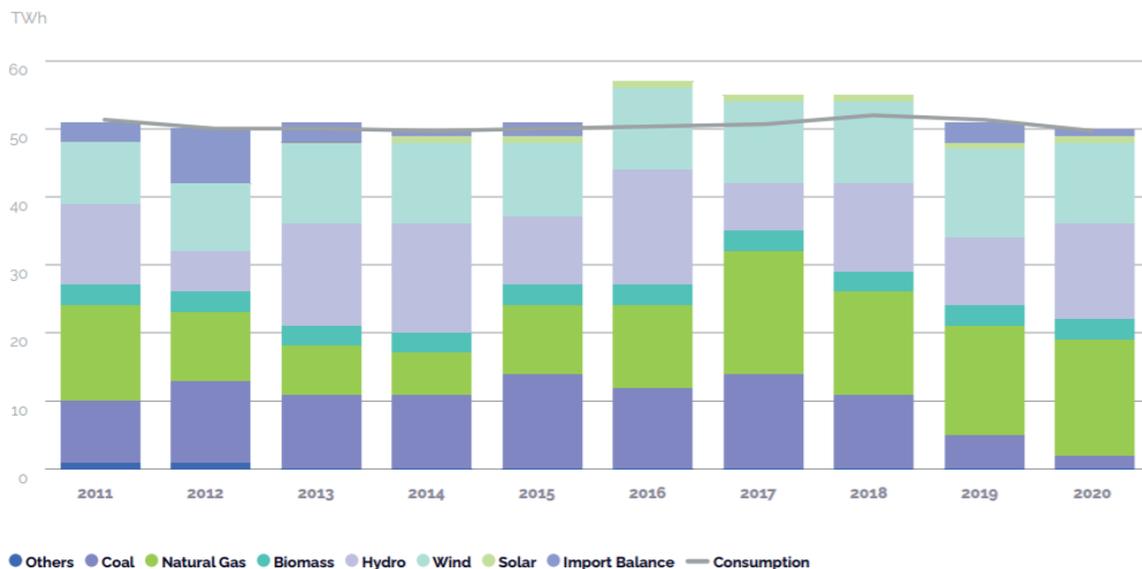


Figura 16: Generación eléctrica en Portugal por tecnología 2011-2020. Tomado de REN [32].

Como se puede comprobar en las imágenes, el carbón ha ido perdiendo mucha importancia con el paso de los años, y hay tres tecnologías que se reparten casi toda la generación: el gas

natural, la hidráulica y la eólica. Las energías renovables cubrieron en el año 2020 un 59% del consumo nacional.

2.5.3.2 Energía solar

Como se puede comprobar en la **¡Error! No se encuentra el origen de la referencia.** y en la **¡Error! No se encuentra el origen de la referencia.**, la energía solar no tiene todavía mucha importancia en el mix portugués. La capacidad instalada a finales de 2020 era de 879 MW y ese año cubrió un 2,6% del consumo total.

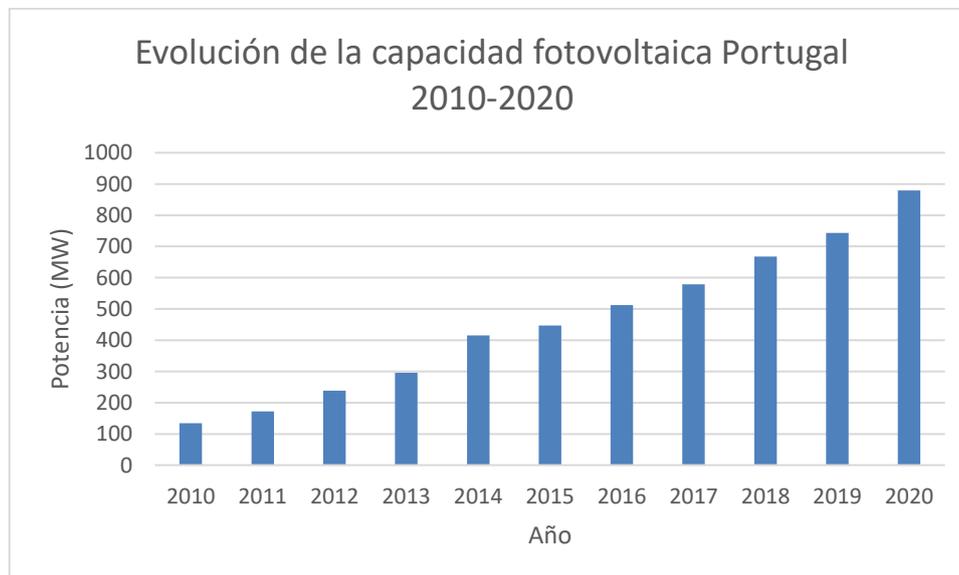


Figura 17: Capacidad fotovoltaica Portugal 2010-2020.

Como se puede comprobar en la Figura 17, la capacidad fotovoltaica no ha dejado de aumentar en Portugal durante los últimos 10 años. Además, este incremento se ha dado de manera muy constante y sin grandes variaciones.

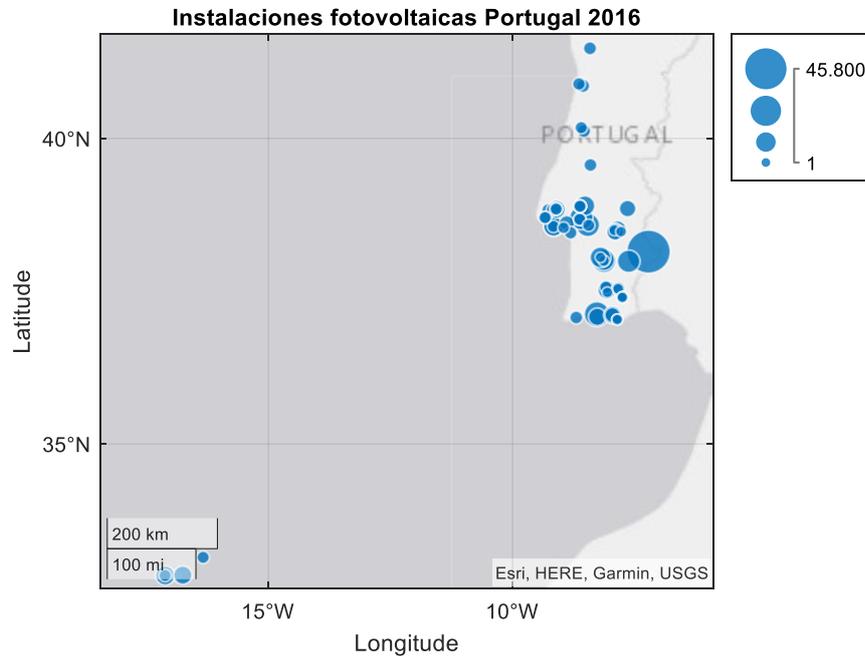


Figura 18: Distribución de las instalaciones fotovoltaicas en Portugal en 2016. Fuente: Global Power Plant Database [33].

Al igual que en el caso de España, las plantas tienden a situarse en el sur del país, donde el recurso solar es mayor a lo largo de todo el año.

2.5.3.3 Planes futuros

En su NECP, los objetivos que se marca Portugal para el año 2030 son:

- 35% de mejora de la eficiencia energética.
- 47% de renovables sobre el uso final de la energía.
- 80% de renovables en la generación eléctrica.
- 15% de grado de interconexión eléctrica.
- 45-55% de reducción de emisiones de GEI respecto a los niveles de 1990.

En el documento se especifica que se van a utilizar dos instrumentos para potenciar la energía solar fotovoltaica, que son las subastas y la posibilidad de que los promotores desarrollen expansiones de la red junto con el operador de la red en situaciones en las que no hay

capacidad de recepción. Para el año 2030 se pretende que haya 9 GW de potencia fotovoltaica, de los cuales 7 serían de instalaciones centralizadas y el resto descentralizadas.

2.5.4 FRANCIA

2.5.4.1 Perspectiva general

Se trata del país más grande de toda la UE con una extensión de 675.000 km² y sus casi 65 millones de habitantes lo convierten en el segundo más habitado. Es uno de los motores económico e industrial del continente y una de las mayores potencias a nivel global. En el año 2020 su consumo eléctrico fue de 460 TWh [34], casi el doble que el de España, y el segundo de toda la Unión Europea, sólo por detrás de Alemania.

Su mix energético es único en el mundo, y es que se trata del país en el que la energía de origen nuclear cuenta con mayor peso. En el año 2020 esta tecnología supuso el 67% de toda la producción eléctrica en el país, porcentaje que no se alcanza en ningún otro lugar del mundo.

2.5.4.2 Energía solar

La energía solar fotovoltaica, aunque sigue teniendo una aportación modesta dentro del mix francés (apenas un 2,2% en 2019), lleva unos años incrementándose de forma muy notable.



Figura 19: Evolución de la capacidad fotovoltaica instalada en Francia 2007-2020. Tomado de RTE [34].

En la Figura 19 se aprecia como desde el año 2012 se viene instalado aproximadamente 1 GW de nueva capacidad fotovoltaica al año. Esta cifra se tendrá que incrementar si se quiere cumplir con los objetivos marcados en el PPE para el año 2023. En cuanto a la distribución de las instalaciones, esta se muestra de manera gráfica en la **¡Error! No se encuentra el origen de la referencia..**

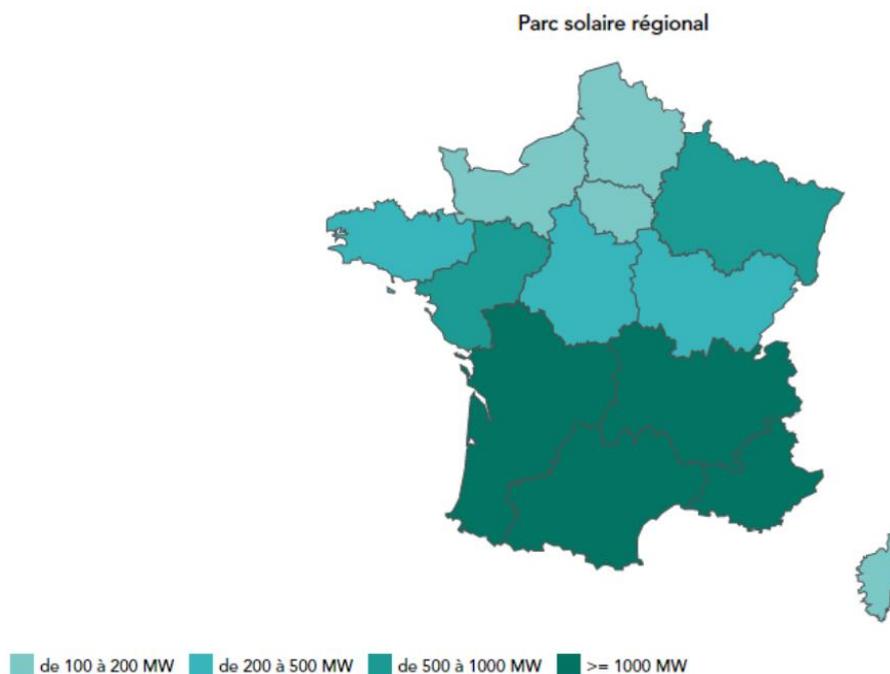


Figura 20: Distribución de la capacidad fotovoltaica instalada en Francia en 2020. Tomado de RTE [34].

Evidentemente, las instalaciones fotovoltaicas tienden a ubicarse en la zona sur del país, donde el recurso solar es mayor a lo largo de todo el año. En el norte abundan más las instalaciones eólicas.

2.5.4.3 Planes futuros

En su NECP, algunos de los objetivos que se marca Francia para el año 2030 son:

- 20% de mejora de la eficiencia energética.
- 33% de renovables sobre el uso final de la energía.
- 40% de renovables en la generación eléctrica.
- 40% de reducción de emisiones de GEI respecto a los niveles de 1990.

En cuanto a la energía solar fotovoltaica, en el plan se pone como objetivo para el año 2023 que la capacidad llegue a 20,1 GW, y para el año 2028 que se encuentre entre los 35,1 y 44 GW. Al igual que en el caso de España y Portugal, uno de los mecanismos que más se contempla para llegar a cumplir estos objetivos son las subastas. En el plan se especifica un calendario de subastas para la energía fotovoltaica que va de 2019 a 2024 y que se muestra a continuación.

Tabla 4: Calendario de subastas fotovoltaicas Francia 2019-2023. Potencia en MW.

Tecnología	2019	2020	2021	2022	2023	2024
Fotovoltaica en tierra	1700	2850	2000	2000	2000	2000
Fotovoltaica en tejado	900	900	900	900	900	900

2.5.5 INTERCONEXIONES

Las tres áreas de estudio se encuentran interconectadas, aunque Portugal y Francia de manera indirecta, a través de España. Estas interconexiones permiten el intercambio de energía eléctrica entre las distintas áreas, lo que genera dependencias espacio-temporales entre ellas. Es por este motivo por el que será necesario conocer sus características generales.

En la Figura 21 se pueden apreciar las líneas eléctricas que conectan estos países.

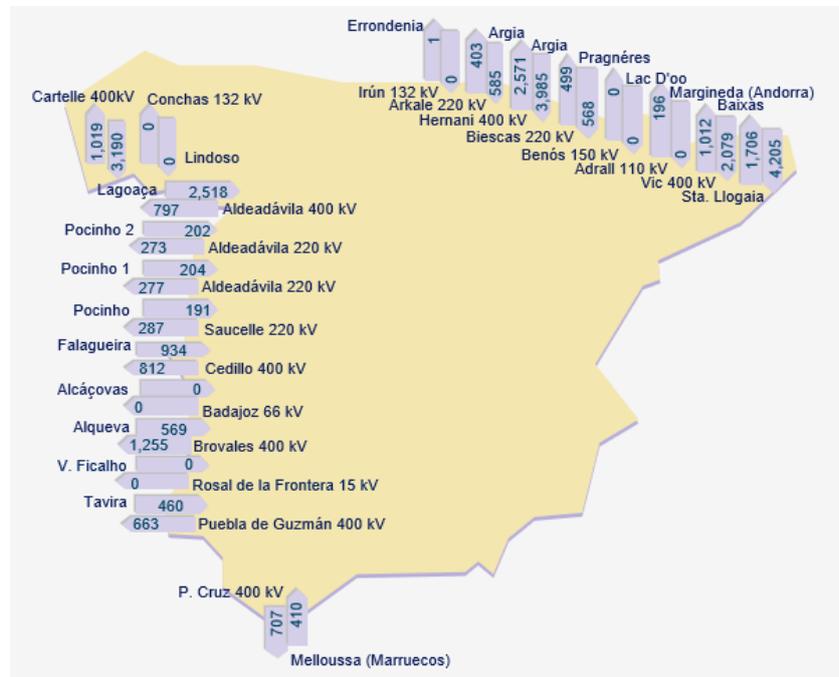


Figura 21: Mapa de interconexiones entre las 3 áreas de estudio. Tomado de [31].

La mayor parte de las líneas de interconexión son de corriente alterna y de alta tensión. Únicamente existe una, entre España y Francia, que es de corriente continua, que es la que une las subestaciones de Sta. Llogaia (España) y Baixas (Francia).

3. METODOLOGÍA

En esta sección se va a presentar la metodología propuesta para llevar a cabo la generación de escenarios de producción fotovoltaica, que consta de 6 pasos:

- 1) Caracterización de los datos
- 2) Preprocesado de los datos
- 3) Descomposición de la serie temporal
- 4) Detección de las dependencias espacio-temporales
- 5) Generación de escenarios
- 6) Evaluación de los resultados

Tabla 5: Resumen de la metodología seguida.

RESUMEN DE LA METODOLOGÍA	
1. CARACTERIZACIÓN DE LA SERIE	Análisis de distribución
	Detección de outliers y N/As
	Análisis de tendencia y estacionalidad
2. PREPROCESADO	Eliminar dependencia de la capacidad instalada
	Corregir cambio de hora
	Corregir outliers N/As mediante interpolación
	Eliminar horas nocturnas
	Igualar la frecuencia mediante interpolación
	Aplicar transformación logit
3. DESCOMPOSICIÓN DE LA SERIE	Capturar y sustraer la tendencia
	Capturar la estacionalidad anual mediante distintos métodos
	Capturar estacionalidad diaria mediante medias y extraerla
4. DETECCIÓN DE DEPENDENCIAS ESPACIO-TEMPORALES	Seguir la metodología Box-Jenkins
	Averiguar el orden mediante el AICc
	Estimar los coeficientes mediante máxima verosimilitud
5. GENERACIÓN DE ESCENARIOS	Realizar simulaciones de Monte Carlo
	Deshacer las transformaciones hechas anteriormente
6. EVALUACIÓN DEL RENDIMIENTO	Evaluar rendimiento mediante:
	• PLF
	• WS
	• Métricas deterministas

3.1 CARACTERIZACIÓN DE LOS DATOS

El objetivo principal de este apartado es analizar las series temporales de generación, estudiar sus idiosincrasias y averiguar qué las define, y estudiar las relaciones entre las distintas áreas.

Los dos datos más importantes con los que se ha trabajado son capacidad instalada y generación fotovoltaica. Los datos son horarios, aunque la capacidad instalada se actualiza mensualmente.

Lo primero que se hará será localizar posibles N/As que pueda haber en la serie. También se visualizará la distribución de los datos, mediante histogramas y boxplots, en busca de outliers.

Después se comprobará si las series temporales de generación tienen patrones que se visualicen fácilmente. Esto ayudará a reducir esfuerzos en etapas posteriores.

3.2 PREPROCESADO

En este apartado se va a hacer un tratamiento de las series temporales originales para corregir posibles errores que puedan tener y para facilitar el ajuste de los modelos ARIMA. Los pasos que se llevan a cabo son:

1. Quitar dependencia de la potencia instalada.
2. Corregir outliers y N/As.
3. Corregir cambio de hora.
4. Eliminar horas nocturnas.
5. Igualar la frecuencia de la serie.
6. Normalizar.

Se tiene una serie temporal con la generación solar horaria para cada una de las áreas. La generación va a depender enormemente de la capacidad total instalada, y esa dependencia puede dificultar la predicción, ya que el comportamiento de la serie temporal de la capacidad

instalada no se puede explicar a través de las dependencias espacio-temporales. El progresivo aumento de la capacidad fotovoltaica en los países del estudio hace que la serie temporal tenga una tendencia positiva, por lo que la serie temporal no será estacionaria. Por este motivo, se va a dividir la generación entre la potencia instalada, es decir, se va a pasar la generación a valores por unidad (*p.u.*). Este paso, además, va a permitir encontrar outliers de forma más sencilla, ya que se sabe que los valores tienen que estar necesariamente entre 0 y 1. Por otro lado, también eliminan el impacto de las magnitudes de generación entre las distintas áreas y hace que no se penalicen las correlaciones espacio-temporales entre las áreas más por las magnitudes de la generación que por los instantes en que varían la series.

La serie también puede presentar algunos datos que no estén rellenos, a los que se conoce como N/As. Estos datos, que si existen van a ser pocos, se podría eliminar directamente de la serie temporal. Sin embargo, en este caso se ha decidido que si existen se realizará una interpolación lineal entre los datos adyacentes al N/A. Otro posible error que puede haber es que algunos datos se salgan de los límites “naturales”, es decir, que sean menores que 0 o mayores que 1. En los casos en los que esto ocurra, se igualarán los valores negativos a 0 y en los que se supere la unidad se interpolará entre los valores adyacentes, al igual que con los N/As.

La siguiente corrección tiene que ver con los cambios de hora que se producen a finales de marzo y octubre. Estos cambios de hora dificultan el ajuste de un modelo ARIMA, ya que la generación en ciertas horas cambia radicalmente en estos dos momentos del año, y un modelo ARIMA no va a ser capaz de ajustarse a ese cambio tan brusco. Lo que se ha hecho para arreglar este inconveniente es que cada uno de los valores que estén entre las 3am del último domingo de marzo y las 2am del último domingo de octubre tomen el valor de la hora siguiente.

Como se ha mencionado repetidamente a lo largo de este trabajo, uno de los inconvenientes más importantes de la energía fotovoltaica es que no genera electricidad por la noche. Si no hay radiación, no hay generación. Este hecho provoca que a lo largo de la serie haya muchos valores donde la generación sea nula. La sobreabundancia de datos nulos distorsiona mucho

la serie y dificulta que se pueda encontrar un modelo fiable. Por este motivo, la mayoría de los proyectos en los que se intentan hacer predicciones de energía solar o de radiación, o bien hacen un tratamiento diferente de las horas nocturnas, o directamente las eliminan [27]. En este caso, se ha optado por esta última medida. Como criterio para decidir qué horas se eliminan se ha decidido eliminar para los tres países las horas que van entre las 22 y las 7 (no incluidas) y aquellas que, a lo largo de todo el mes, no lleguen a cubrir el 1% de la generación en dicho mes.

Al haber eliminado las horas nocturnas, la frecuencia diaria de la serie ya no es la misma. En los meses de invierno hay más horas nocturnas que en los de verano, por lo que en la nueva serie temporal se pueden encontrar días con bastantes más valores de generación que otros. Esto va a generar muchos inconvenientes a la hora de capturar la estacionalidad diaria. Antes la frecuencia diaria de los datos era de 24 horas durante todo el año, pero ahora se han eliminado muchas horas en las que no había generación, que son más en invierno que en verano. Para volver a igualar la frecuencia a lo largo de todo el año se va a realizar una interpolación de los valores diarios de generación. Esta interpolación va a devolver la misma serie, pero con una nueva frecuencia. La frecuencia diaria de la nueva serie se va a igualar a la frecuencia máxima una vez se han eliminado los valores nocturnos.

Una vez hecho esto, se pueden emplear varias transformaciones con el objetivo de simplificar el comportamiento de las series temporales. Existen numerosos métodos de normalización que se emplean habitualmente en la literatura. A continuación, se presentan tres de las transformadas más comunes, que son la Box-Cox, la Ordered Quantile y la Logit.

Tabla 6: Transformada Ordered Quantile y Logit.

Nombre	Transformada
Box-Cox	$w_t = \begin{cases} \log(y_t) & \text{si } \lambda = 0 \\ (y_t^\lambda - 1)/\lambda & \text{si } \lambda \neq 0 \end{cases}$

Ordered Quantile	$z = \Phi^{(-1)}\left(\frac{\text{rank}(y) - \frac{1}{2}}{\text{length}(y)}\right)$
Logit	$y^{(\lambda)} = \ln\left(\frac{y}{1-y}\right)$

El uso de una u otra de estas transformaciones dependerá de diversos factores, ya que cada una tiene sus ventajas y sus inconvenientes. La Logit es la más sencilla, pero no puede trabajar con valores mayores que uno o negativos. La Box-Cox probablemente sea la más empleada. Por otro lado, la Ordered Quantile suele funcionar muy bien, pero deshacer la transformación puede ser complicado.

3.3 *DESCOMPOSICIÓN*

Una vez se ha normalizado la serie, se procederá a descomponerla. Los métodos de descomposición buscar separar una serie en sus componentes o patrones principales y trabajar sobre cada uno de ellos por separada

Los métodos más habituales son los aditivos, multiplicativos y pseudoaditivos. La descomposición aditiva es la más adecuada para emplear en series de generación fotovoltaica en p.u., ya que el recurso solar no cambia con el paso del tiempo. Su forma sería:

$$y_t = T_t + S_t + R_t$$

Cuyos patrones son [35]:

- Tendencia (T_t): se dice que existe cuando una serie a largo plazo crece o decrece, ya sea de manera lineal o no. En el caso de estudio la tendencia se va a obtener por el método de los mínimos cuadrados. Sin embargo, al haber quitado la dependencia de la potencia instalada y al permanecer el recurso solar con el paso del tiempo, se verá

que las series temporales de energía fotovoltaica no tienen una tendencia que sea relevante.

- Estacionalidad (T_t): existe estacionalidad cuando los valores de la serie están influidos por el momento en el que se producen, como, por ejemplo, el día de la semana o la estación del año. Se espera que las series de generación solar tengan dos muy marcadas:

- Estacionalidad anual: la generación solar tendrá máximos en verano y mínimos en invierno. En este trabajo se van a capturar mediante dos métodos distintos y se escogerá el mejor de ellos. El primero será las series de Fourier que se muestra matemáticamente a continuación:

$$S_{a_t} = \sum_{j=1}^{\lfloor \frac{S}{2} \rfloor} \left\{ a_j * \cos\left(\frac{2\pi jt}{s}\right) + b_j * \sin\left(\frac{2\pi jt}{s}\right) \right\}$$

El segundo método será la interpolación por splines. Un spline no es más que una curva diferenciable definida a trozos mediante polinomios. Uno de los tipos más empleados es el spline cúbico, donde los polinomios empleados son de grado tres.

$$P(x) = ax^3 + bx^2 + cx + d$$

a, b, c y d serían las incógnitas y existirían ciertas condiciones de contorno que se deberían cumplir

- Estacionalidad diaria: la generación será muy baja en las primeras y las últimas horas del día y alcanzará los máximos en las horas intermedias del día (nótese que se han eliminado las horas nocturnas de la serie, donde la generación es nula o casi nula). Para este proyecto se ha decidido capturar hallando la media de cada hora y restándola a la serie original. Este método es muy adecuado para la estacionalidad diaria ya que esta mantiene la misma forma a lo largo de toda la serie y, al contrario del Fourier, no requiere escoger un orden ni optimizar constantes, por lo que es más sencillo.

- Término remanente (R_t): es aquella dinámica de la serie que no ha sido capturada por los términos deterministas y que será modelada con modelos ARIMA. Se busca que sea lo más parecido posible al ruido

3.4 DETECCIÓN DE DEPENDENCIAS ESPACIO-TEMPORALES

Una vez se ha quitado la tendencia y las diferentes estacionalidades de la serie se va a proceder a ajustar un modelo ARIMA para cada área que capture la dependencia de la generación fotovoltaica con respecto del tiempo.

Una de las formas más extendidas para ajustar un buen modelo ARIMA es emplear la metodología Box-Jenkins, que no es más que un proceso iterativo en el cual se ajustan modelos y se calculan residuos hasta que estos últimos sean los más parecidos posibles al ruido blanco. De forma esquematizado, la metodología sería como se muestra en la **¡Error! No se encuentra el origen de la referencia..**

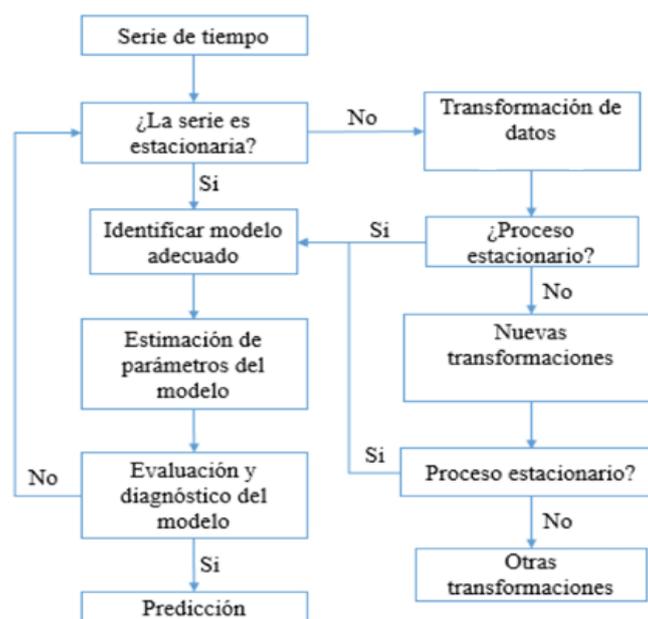


Figura 22: Resumen de la metodología Box-Jenkins. Tomado de [1].

Para la identificación del orden del modelo se inspeccionarán de forma visual las gráficas de la Función de Autocorrelación y la Función de Autocorrelación Parcial (ACF y PACF, por

sus siglas en inglés). El ACF mide la correlación existente entre un valor y_t e y_{t-k} . El problema es que si y_t e y_{t-1} están correlacionados, y_{t-1} e y_{t-2} también lo estarán, por tanto, y_t e y_{t-2} estarán correlacionados. El PACF viene a solucionar este problema eliminando los valores intermedios. Son muy útiles porque con un simple vistazo de las gráficas del ACF y del PACF se puede discernir de forma aproximada cuál va a ser el orden adecuado de los modelos.

Aparte de la revisión visual del ACF y el PACF, se suelen utilizar otros criterios para escoger el mejor orden de los modelos. Los que se utilizan normalmente para escoger el mejor orden del modelo ARIMA son el Criterio de Información Akaike, el Criterio de Información Bayesiano y el Criterio de Información Akaike Corregido (AIC, BIC y AICc respectivamente, por sus siglas en inglés). La estimación de los parámetros del modelo se hará de acuerdo con el criterio de máxima verosimilitud. Las ecuaciones de los criterios se muestran a continuación:

$$AIC = -2 \log(L) + 2(p + q + k + 1)$$

$$AICc = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2}$$

$$BIC = AIC + [\log(T) - 2](p + q + k + 1)$$

donde $k = 1$ si c es distinto de 0 y $k = 0$ si $c = 0$, y L es la verosimilitud, que sería:

$$L = \sum_{t=1}^T (y_t - \hat{y}_t)$$

Para obtener un buen modelo la idea sería escoger uno de estos tres indicadores y ver qué orden del modelo y qué parámetros lo minimizan lo máximo posible. Es decir, se tratan de escoger el número de coeficientes autorregresivos y de media móvil y estimar sus valores. Hay que tener en cuenta que estos tres indicadores tienen algunas diferencias entre sí. Por ejemplo, el BIC penaliza más por parámetros adicionales que el AIC. Por otro lado, el AICc funciona mejor cuando se tienen pocos datos que el AIC, y son prácticamente iguales cuando

hay muchos datos. En [35] se especifica que se prefiere comparar los AICc, y es lo que se va a hacer en este trabajo.

3.5 GENERACIÓN DE ESCENARIOS

Estimados los parámetros de los modelos ARIMA, será el momento de generar los escenarios futuros. Normalmente lo que se haría sería generar residuos descorrelacionados y con una distribución normal para generar los distintos escenarios, pero, en este caso, al estar las tres áreas de estudio muy correladas debido a factores geográficos y de la propia red eléctrica, será necesario incluir estas relaciones en los residuos, y también habrá que tener en cuenta que las relaciones entre las áreas pueden cambiar durante el año. Con el objetivo de capturar estas correlaciones, se han generado residuos horarios multivariados y correlacionados con funciones de distribución normal, por lo que incluyen correlaciones espaciales a través de matrices de covarianza obtenidas a partir de series de datos históricos.

$$f_x(x_1, \dots, x_k) = \frac{\exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)}{\sqrt{(2\pi)^k \Sigma^{1/2}}}$$

donde k es el número de residuos generados, x es un vector real de k dimensiones, Σ es la matriz de covarianzas simétrica, μ es la media de los residuos.

Una vez han sido realizados estos pasos, se hacen las simulaciones de Monte Carlo para generar los escenarios de producción fotovoltaica. Estas simulaciones permiten generar distribuciones completas para cada hora que incluyan las dependencias temporales identificadas por los modelos ARIMA.

Una vez están los escenarios generados, habrá que deshacer todas las transformaciones que se han hecho hasta el momento, pero en el orden contrario. Es decir, habrá que añadir las estacionalidades, deshacer la transformación la serie, añadir las horas nocturnas y corregir el cambio de hora.

3.6 EVALUACIÓN DEL RENDIMIENTO

Generados los posibles escenarios, es hora de evaluar la calidad de los mismos. Existen numerosos métodos para evaluar si una predicción ha sido más o menos precisa en la literatura. En este proyecto se han utilizado los indicadores probabilísticos Pinball Loss Function (PLF) y el Winkler Score (WS) para evaluar los escenarios. Además, también se utilizarán métricas deterministas, como comparaciones entre medias de los datos históricos y simulados, máximos, mínimos, etc. Las métricas probabilísticas permitirán comprobar cuánto se parece la distribución de los datos simulados a los históricos y las deterministas serán útiles para ver si se está capturando bien el comportamiento de los históricos.

En el PLF se asigna una puntuación en función de las diferencias entre los valores de la serie temporal real y los percentiles de los simulados. Es una medida del error para las previsiones cuantílicas [36] y cuanto más bajo sea el indicador, mejor será el modelo. Siendo y_a el percentil de los escenarios generado con $a = 1, 2, \dots, 99$ el PLF sería:

$$PLF(\hat{y}_a, y_t) = \begin{cases} \left(1 - \frac{a}{100}\right) (\hat{y}_a - y_t) & \text{si } y_t < \hat{y}_a \\ \frac{a}{100} (y_t - \hat{y}_a) & \text{si } y_t \geq \hat{y}_a \end{cases}$$

Cuando se trata de varios métodos con niveles de cobertura de precisión similar, lo mejor será escoger el modelo que tenga los intervalos más estrechos. El WS permite evaluar conjuntamente la cobertura y la amplitud del intervalo. Para un intervalo de predicción $(1-\alpha)100\%$, se define como:

$$W_t = \begin{cases} \delta_h & \text{si } L_h \leq y_h \leq U_h \\ \delta_h + \frac{2}{\alpha} (L_h - y_h) & \text{si } y_h < L_h \\ \delta_h + \frac{2}{\alpha} (y_h - U_h) & \text{si } y_h > U_h \end{cases}$$

4. RESULTADOS

Para llevar a cabo la simulación de escenarios de energía solar se han utilizado los softwares R y Matlab. En concreto, se ha utilizado Matlab como software principal y R como apoyo, porque tiene ciertas funciones sobre métodos de normalización que no incluye el primero y que han resultado útiles para el proyecto.

Se va a ir mostrando la aplicación de la metodología previamente explicada al caso de estudio real sobre España, Portugal y Francia, el tratamiento que se hace de las series temporales originales, el ajuste de los modelos ARIMA y finalmente los resultados obtenidos.

4.1 ESTUDIO DE LOS DATOS

En este apartado se van a mostrar las principales características de las series temporales que se van a estudiar. Para realizar este estudio se han empleado los datos de generación solar fotovoltaica horaria de los tres países que se contemplan en este estudio: España, Portugal y Francia. En el caso de España se tienen datos de los años 2016 a 2019 (ambos incluidos) y en el caso de Portugal y Francia de 2015 a 2019. En total, se tienen 35.064 valores para España y 43.824 para Francia y Portugal.

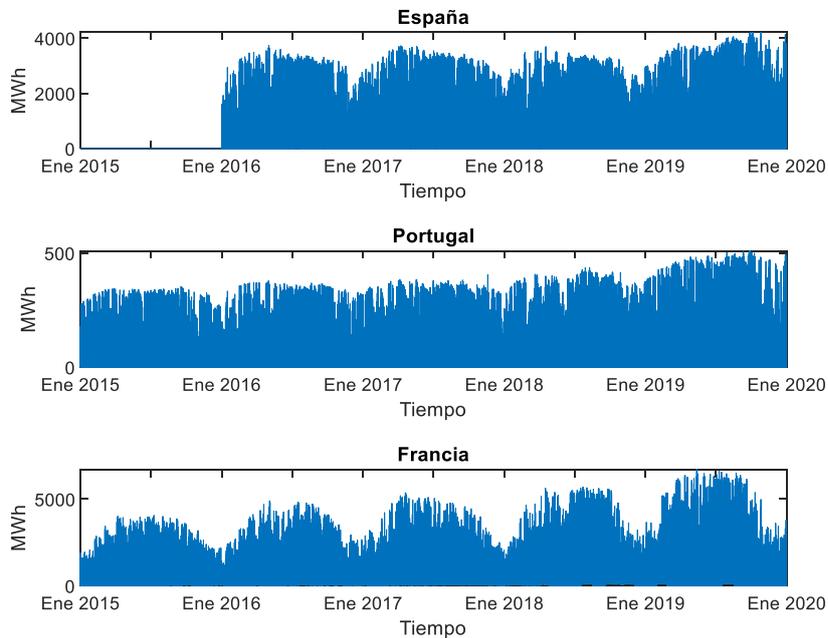


Figura 23: Series temporales de generación fotovoltaica.

Las gráficas de las series temporales contienen mucha información que merece la pena analizar. Se puede ver una tendencia positiva clara en los casos de Francia y Portugal. Además, hay una estacionalidad anual muy marcada que cambia un poco en función del país. Esto se tratará en detalle más adelante. En este momento se ha mirado cómo es la distribución de los datos de estas series temporales con la intención de encontrar posibles outliers y de tratar de averiguar qué métodos de normalización se podrían emplear. Lo único que se ha sacado en claro es que hay una sobreabundancia de datos nulos, debido a las horas nocturnas, que no permiten encontrar ninguna distribución lógica que explique la ocurrencia de esos datos. Además, en los casos de España y Francia se ha visto que hay unos pocos valores de generación que están por debajo del 0. Esto podría deberse a que algún parque fotovoltaico que tuviese baterías estuviera cogiendo energía de la red por la noche para recargarlas. Sin embargo, lo más probable es que se trate de error, por lo que habrá que corregirlo igualando estos valores a 0. Las horas nocturnas hacen que haya una sobreabundancia de datos nulos, lo que puede dificultar el ajuste del modelo.

Se va a mostrar ahora las correlaciones que existen entre las 3 áreas.

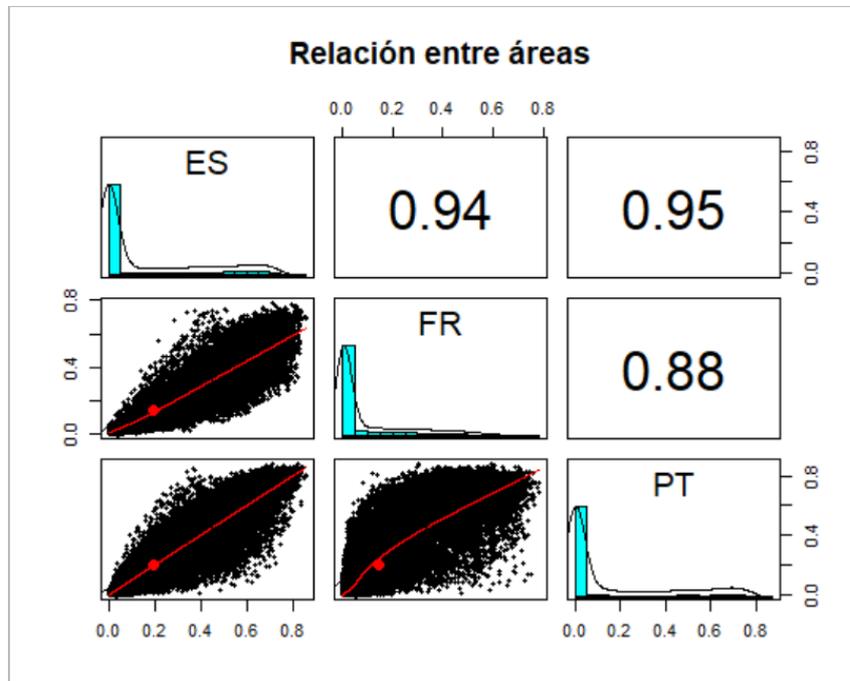


Figura 24: Histogramas, gráficos de dispersión y coeficiente de Pearson de las 3 áreas.

En la Figura 24 se muestran los histogramas de las tres series temporales en la diagonal, gráficos de dispersión por debajo de la diagonal y el coeficiente de correlación de Pearson por encima de esta. Como se puede observar, las tres áreas están fuertemente correlacionadas. La correlación entre Francia y Portugal es ligeramente menor, probablemente porque no abarcan ni la misma latitud ni longitud.

Se va a estudiar ahora la tendencia de las series temporales. Si se mira la Figura 23, se podrá comprobar que existe una tendencia clara en los casos de Francia y Portugal, mientras que en el de España permanece bastante constante. El recurso solar no cambia de un año a otro, y es difícil que las condiciones climatológicas cambien tanto de un año a otro, por lo que es lógico pensar que esta tendencia se debe a un aumento de la capacidad instalada. En la Figura 25, que se muestra a continuación, se aprecia la generación en cada área comparada con la evolución de la capacidad fotovoltaica instalada.

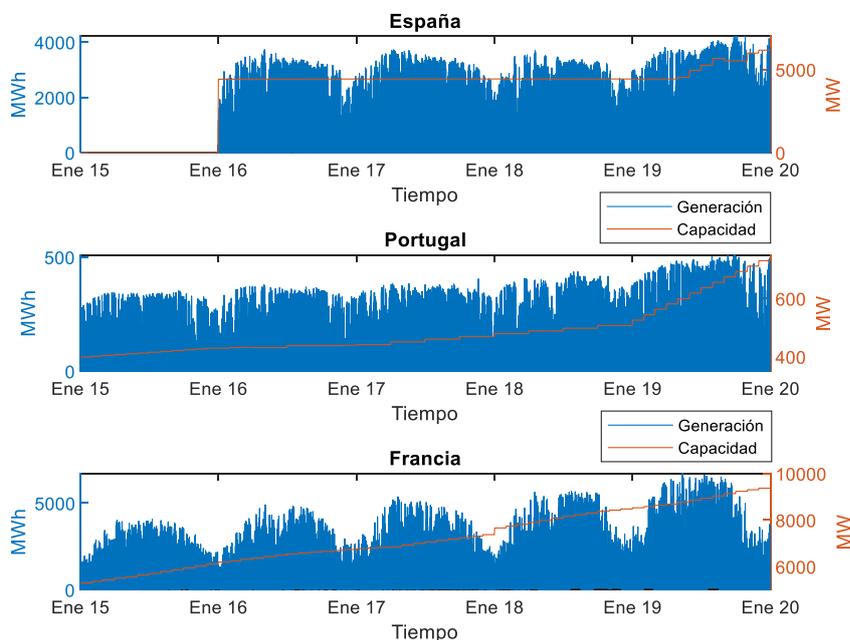


Figura 25: Generación y capacidad fotovoltaica en cada área 2015-2019.

Los ejes verticales de la parte izquierda de la Figura 25 representan la energía generada y los de la parte derecha la potencia instalada. Nótese que los rangos de los ejes varían en función del país. Esto se ha hecho así para evidenciar la relación entre la tendencia y la capacidad instalada.

En los casos de Portugal y Francia se observa claramente que la generación viene siendo mayor cada año, y que este crecimiento está fuertemente ligado al incremento de la capacidad instalada. En cuanto a España, en el periodo 2016-2018 la potencia instalada apenas cambió, y tampoco lo hizo la generación. Sólo a partir de 2019, cuando se pusieron en marcha nuevas instalaciones, se puede apreciar que la generación aumenta considerablemente. Ahora se va a estudiar la correlación que existe entre la capacidad fotovoltaica instalada y la generación. Está claro que el comportamiento de la generación no se puede explicar sólo a partir de la capacidad, ya que también entran en juego otros factores importantes, como la estacionalidad anual o la diaria. A continuación, se van a mostrar los coeficientes de correlación de Pearson de la capacidad instalada con respecto a los datos de

generación horarios, los datos de generación diarios, los datos de generación mensuales y con respecto a la tendencia de la serie sacada por el método de mínimos cuadrados.

Tabla 7: Coeficientes de correlación de Pearson de capacidad instalada con respecto a generación.

Área	Horarios	Diarios	Mensuales	Tendencia
España	0.0318	0.1253	0.1532	0.6038
Portugal	0.0838	0.2903	0.3853	0.8722
Francia	0.1088	0.3083	0.3448	0.9951

La tabla indica que, por ejemplo, en el caso de Francia, toda la tendencia puede explicarse únicamente con la capacidad instalada.

Se va a pasar ahora a tratar la estacionalidad diaria. Esta también es muy marcada y relevante en el caso de la generación fotovoltaica, pero no se puede apreciar bien en la Figura 23. Reordenando un poco la gráfica se obtienen las siguientes figuras.

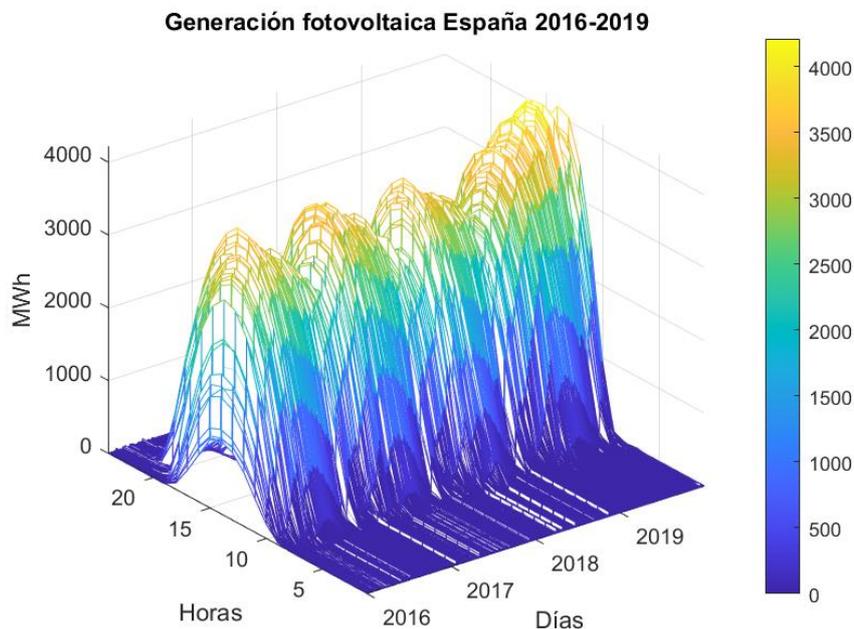


Figura 26: Serie de generación fotovoltaica España 3D.

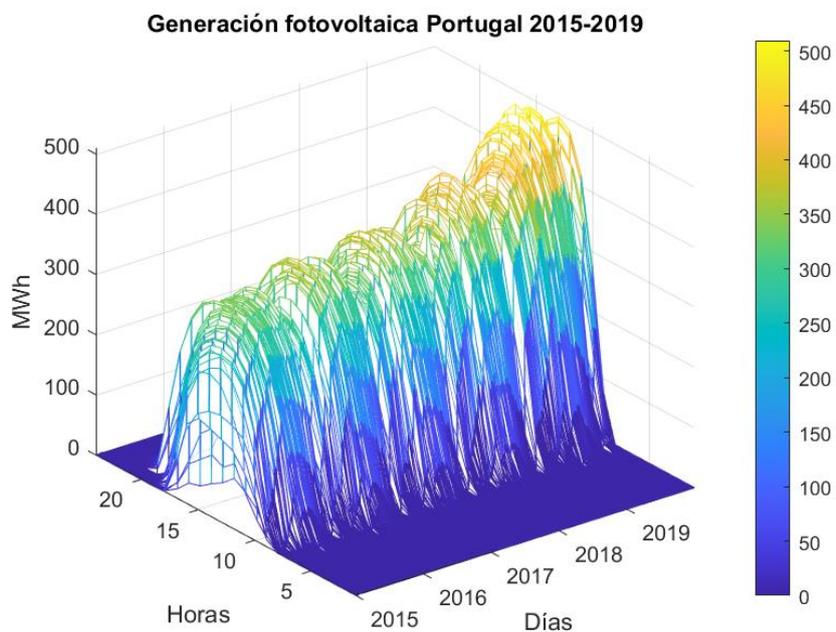


Figura 27: Serie de generación fotovoltaica Portugal 3D.

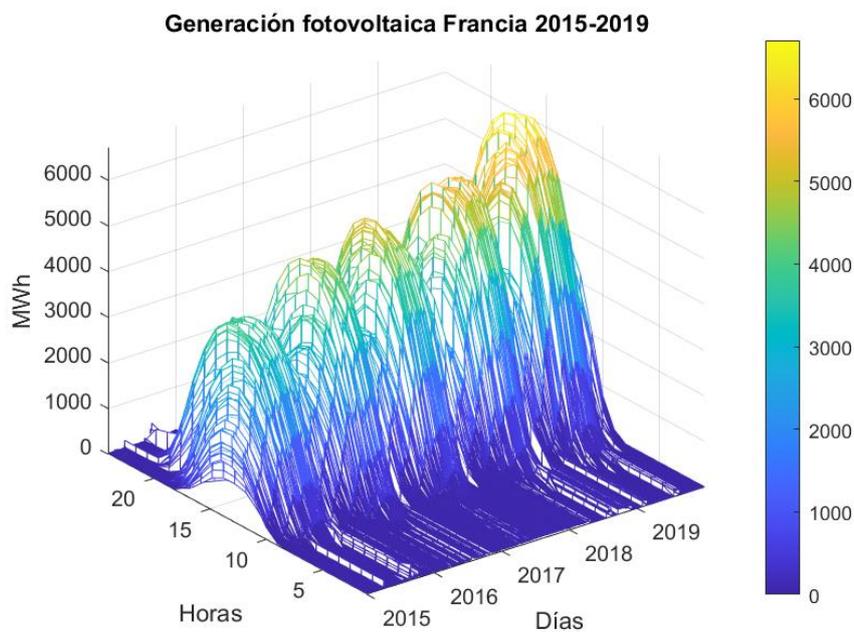


Figura 28: Serie de generación fotovoltaica Francia 3D.

En la Figura 26, la Figura 27 y la Figura 28 se puede observar claramente como la generación es nula durante las primeras horas del día, empieza a crecer después del amanecer, alcanza su pico en las primeras horas de la tarde, decrece y vuelve a ser cero por la noche. Esta estacionalidad diaria varía un poco en función de la época del año, habiendo más horas donde la generación es nula en invierno que en verano. Las diferencias a lo largo del año en la estacionalidad diaria serán más acusadas en el caso de Francia, ya que este país se encuentra en una latitud mayor que los otros dos.

4.2 PREPROCESADO

La tendencia provocada por el aumento de la capacidad apreciable en la Figura 25, supone un problema al trabajar con las series temporales de generación, ya que hace que la serie no sea estacionaria. Para quitar la dependencia de la generación de la potencia instalada se pasaron los datos a por unidad (p.u.). Un pequeño problema que surge al hacer esto es que los datos de potencia instalada se actualizan de mes a mes, por lo que la varianza intramensual puede aumentar ligeramente. La generación en p.u. se muestra a continuación.

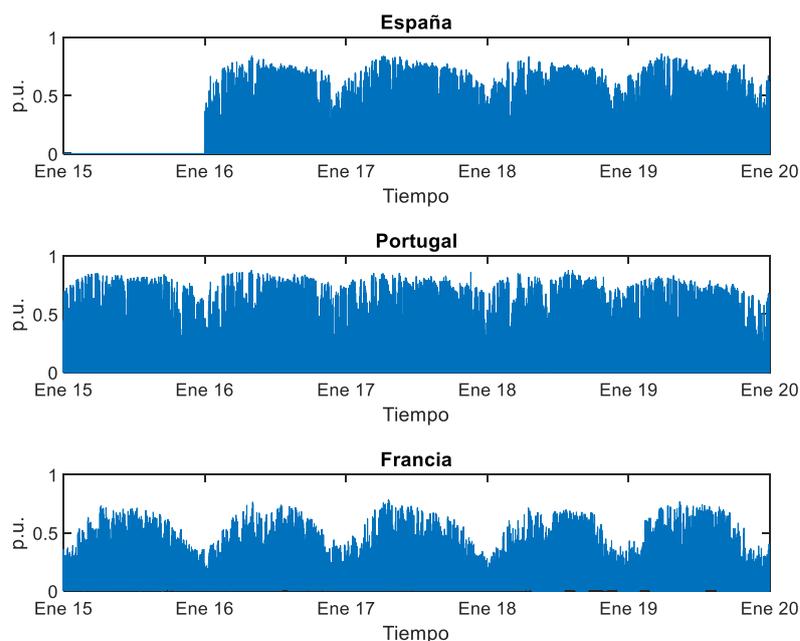


Figura 29: Series temporales de generación fotovoltaica en p.u.

Se puede ver en la Figura 29 como el factor de carga de la fotovoltaica en España y Portugal es considerablemente mayor que en Francia. Esto se puede debe a que el recurso solar es mayor en el caso de los primeros (ver Figura 6) o a otros motivos como la eficiencia de la tecnología o el mejor aprovechamiento en términos de ubicación de las instalaciones.

Una vez se ha eliminado la dependencia de la potencia instalada, se corrigieron los posibles errores que pueda haber en los datos. No se han encontrado N/A's, pero sí que hay valores negativos de generación en los datos de España y Francia. Estos datos aparecen en horas nocturnas y pueden deberse a que se use en ciertos momentos la energía de la red para cargar las baterías de algún parque fotovoltaico, pero lo más normal es que se deban a un error. Se han igualado estos valores negativos a 0, tal y como se mencionó en el apartado de metodología.

Ahora se ha solucionado el problema de los cambios de hora que se producen a finales de marzo y octubre. Como se explicó en 3.2, estos cambios de hora hacen que sea más complicado ajustar un modelo ARIMA, por lo que lo mejor será corregirlo. Lo que se ha hecho para arreglar este inconveniente es que cada uno de los valores que estén entre las 3 am del último domingo de marzo y las 2 am del último domingo de octubre tomen el valor de la hora siguiente.

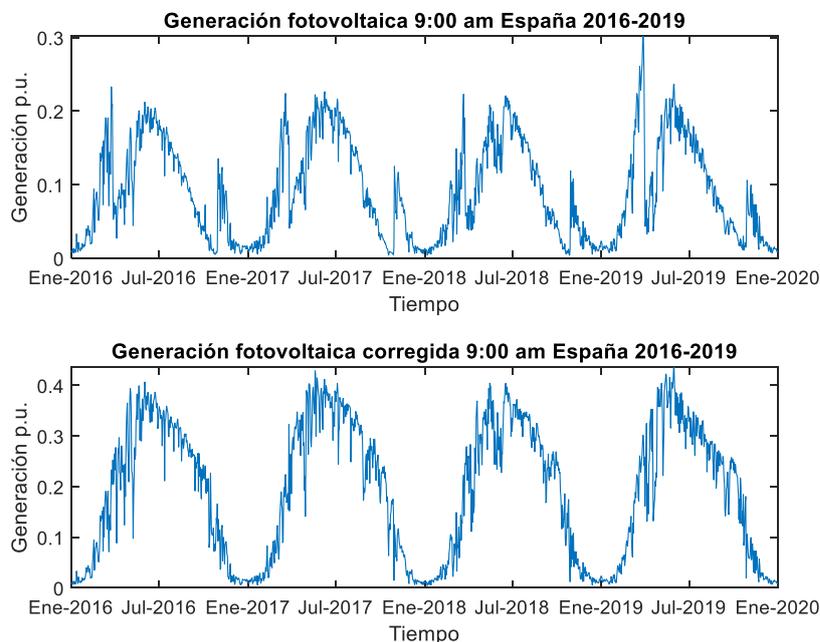


Figura 30: Generación fotovoltaica en p.u. a las 9 am España.

La Figura 30 muestra dos gráficas con la generación fotovoltaica en España en p.u. a las 9:00 antes (arriba) y después (abajo) de la corrección. En la gráfica de arriba se observa claramente como hay dos momentos del año en los que la generación cambia muy bruscamente, motivados por los cambios de hora de primavera y otoño. En la gráfica de abajo se aprecia como la corrección consigue eliminar estos cambios tan bruscos.

Las horas nocturnas deberán ser eliminadas en este momento. Para ver la influencia que tiene a lo largo de la serie, se va a volver a mostrar la Figura 26, pero vista desde otra perspectiva distinta que permita apreciar mejor el efecto de las horas nocturnas.

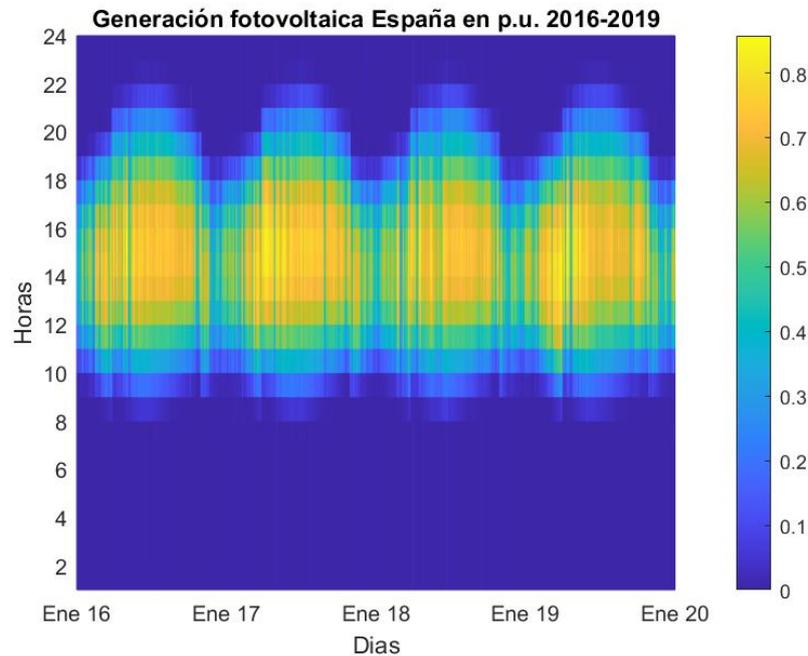


Figura 31: Generación en España en función de hora y día.

En la Figura 31 se puede apreciar claramente cómo más de la mitad de las horas la generación es nula o prácticamente. Además, la estacionalidad anual hace que en ciertos meses el número de horas productivas sea mucho menor que en otros. Como criterio para decidir qué horas se consideran diarias y qué horas no se ha establecido que sólo se considerarán diaria aquellos valores que estén entre las 7 y las 22 y cuya generación sea superior a 0,01 *p.u.*

Una vez se han eliminado estos datos surge un nuevo problema, y es que ahora la frecuencia de la serie no es constante. Para volver a igualar la frecuencia a lo largo de todo el año se va a realizar una interpolación de los valores diarios de generación. Esta interpolación va a devolver la misma serie, pero con una nueva frecuencia. La frecuencia diaria de la nueva serie va a ser 16, que es la frecuencia máxima que hay en las series una vez se han eliminado los valores nocturnos.

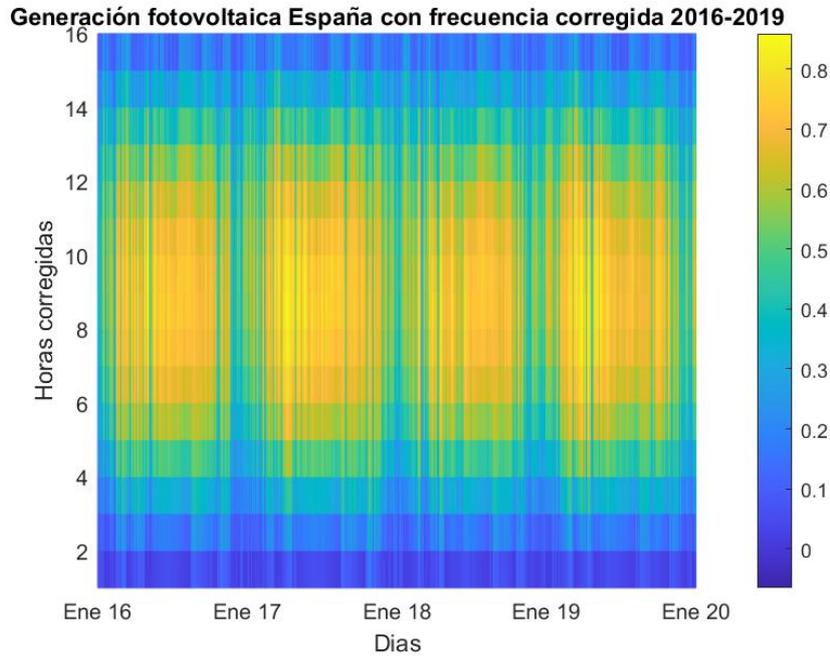


Figura 32: Generación en España en función de hora y día con frecuencia corregida.

Se observa en la Figura 32 que ahora la generación en cada hora no cambia tanto a lo largo del año. La serie vista desde la perspectiva original sería la siguiente.

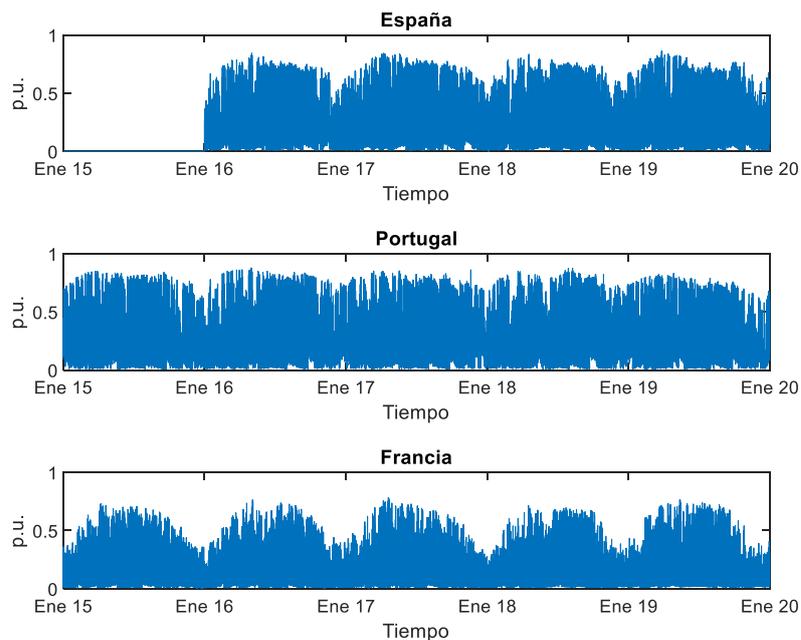


Figura 33: Series temporales de generación fotovoltaica con frecuencia corregida.

Ahora se va a tratar de normalizar la serie para que la varianza sea lo más constante posible a lo largo de la misma. Se han probado varios métodos de normalización, que son:

- Box-Cox.
- Logit.

A continuación, se muestran las distribuciones de las series transformadas.

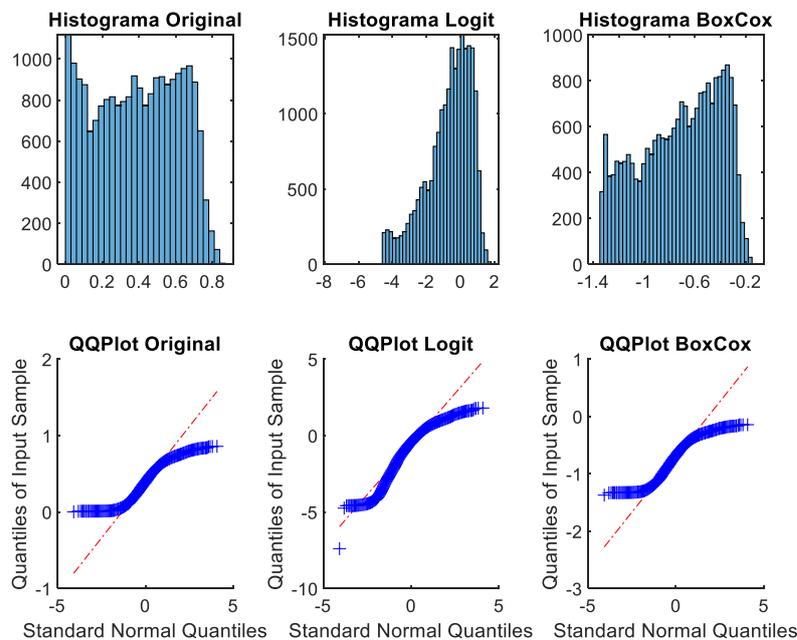


Figura 34: Distribuciones de la serie original y las transformadas de España (sin horas nocturnas).

Como se puede comprobar en la Figura 34, la serie sin transformar tiene una distribución relativamente constante. Entre 0 y 0,6 p.u. la distribución no cambia demasiado, y a partir de ahí empieza a caer muy rápido. De entre las dos transformaciones que se han evaluado, se ha escogido la normalización Logit debido a que la distribución de los datos es un poco mejor. La serie con la transformada Logit aplicada sería:

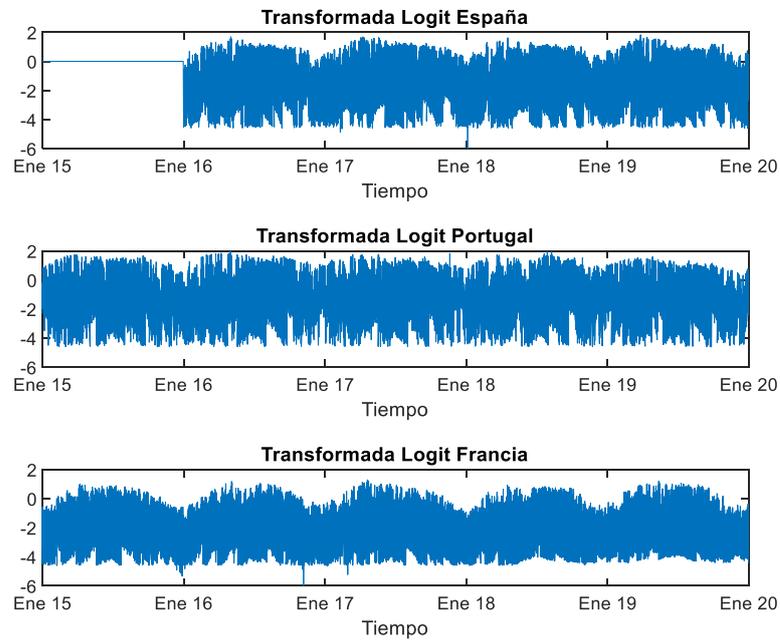


Figura 35: Series temporales con transformada logit aplicada.

4.3 DESCOMPOSICIÓN DE LA SERIE

La primera componente que se analizará de las series temporales de cada área será la tendencia. En la siguiente imagen se muestra la tendencia polinómica de cada serie.

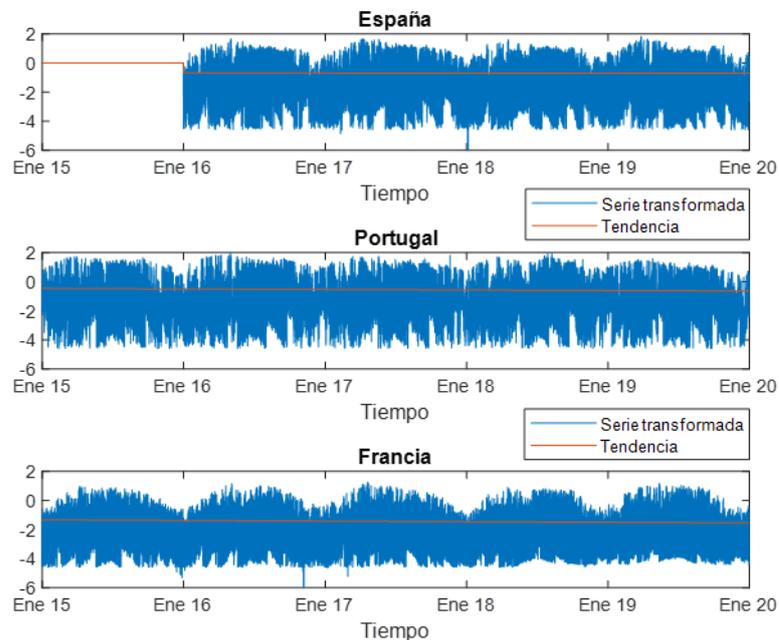


Figura 36: Tendencia de las series temporales.

La tendencia es insignificante en el caso de España, y poco relevante en el de Portugal y Francia. El motivo por el que la tendencia es poco importante es porque el recurso solar permanece constante a lo largo de los años y porque ya se ha quitado el efecto del incremento de la potencia instalada. Se ha decidido no quitar la tendencia porque mejoraría poco los resultados e incrementaría la complejidad del proceso. No obstante, sería generalizable la metodología al caso en el que fuese necesaria la tendencia como consecuencia de existir, por ejemplo, una mejora de la eficiencia en la generación solar que se vaya instalando en el futuro.

Una vez se ha hecho la normalización de la serie se ha capturado la estacionalidad anual de las series temporales. Se han empleado tres métodos para modelar la estacionalidad anual, que son:

- Serie de Fourier de orden 1.
- Serie de Fourier de orden 2.
- Estacionalidad mediante splines.

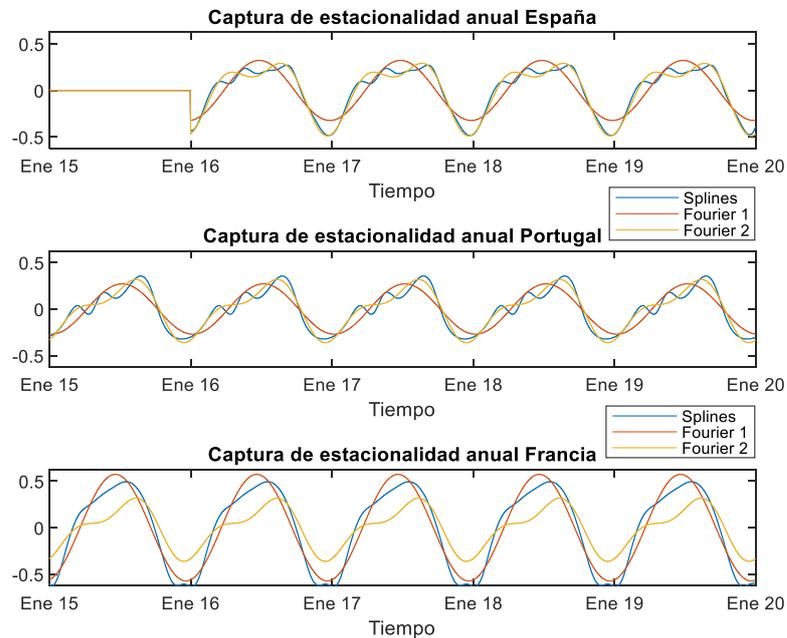


Figura 37: Métodos de captura de la estacionalidad anual.

Con la captura de la estacionalidad anual lo que se hace es quitar una componente determinista la serie. La radiación solar y el número de horas de luz aumenta paulatinamente desde el solsticio de invierno al de verano, y disminuye a partir de este. Como no está muy claro que una de las formas sea claramente mejor que las otras, se va a probar a hacer la simulación con las tres y se decidirá cuál es la mejor a partir de los indicadores de rendimiento (PLF y WS). De aquí en adelante, para no sobrecargar el documento de figuras, para las gráficas que se mostrarán se habrá escogido como método de captura el Fourier de orden 1. Las otras figuras se mostrarán en anexos.

La serie sin estacionalidad anual sería:

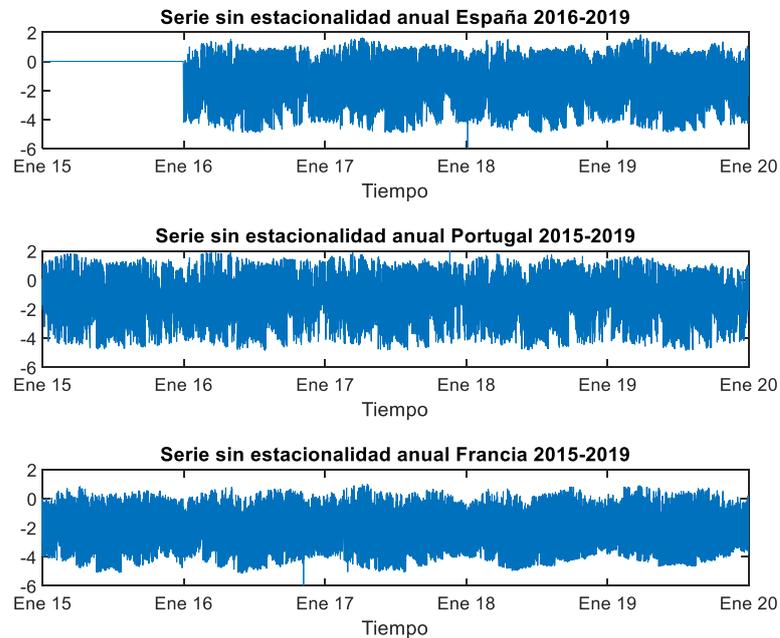
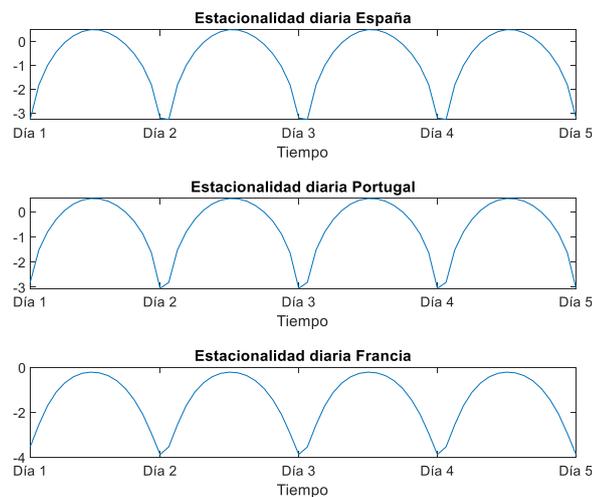


Figura 38: Series temporales sin estacionalidad anual.

La estacionalidad diaria, en cambio, se capturará sacando las medias de las distintas horas de la serie temporal. Este método es muy adecuado para la estacionalidad diaria ya que esta mantiene la misma forma a lo largo de toda la serie y, al contrario del Fourier, no requiere escoger un orden ni optimizar constantes, por lo que es más sencillo. La captura de la estacionalidad se muestra en la siguiente imagen.



Y la serie desestacionalizada sería:

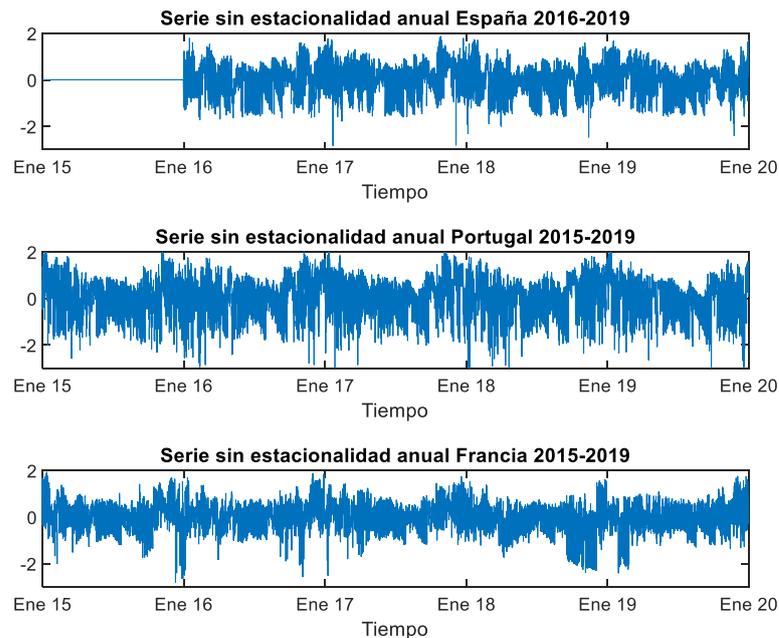


Figura 39: Series temporales sin estacionalidad anual ni diaria.

4.4 DETECCIÓN DE DEPENDENCIAS ESPACIO-TEMPORALES

Una vez normalizada y desestacionalizada la serie, es el momento de ajustar los mejores modelos ARIMA, tanto para el caso estático como para el dinámico. Para ello, se ha seguido la metodología Box-Jenkins. Se va a minimizar el indicador AICc a la vez que se lleva a cabo una inspección visual de los residuos con el ACF y el PACF. Esta inspección es especialmente importante, ya que, si sólo se tratase de minimizar el AICc, el modelo podría no capturar relaciones que fuesen esenciales. Una vez se tienen los órdenes de los modelos ARIMA, se van a estimar sus coeficientes mediante el criterio de máxima verosimilitud. Todo este proceso es más largo y computacionalmente más complicado para el caso dinámico, ya que hay que calcular treinta y seis modelos, por solo tres del estático.

4.5 GENERACIÓN DE ESCENARIOS

Obtenidos ya los modelos ARIMA para cada área, se simularán ahora las sendas de los escenarios que se obtienen con dichos modelos. Se han generado residuos multivariantes de media cero, considerando las covarianzas entre las tres áreas, mediante simulaciones Monte Carlo. En total se han generado 300 sendas distintas. Este número es totalmente parametrizable y se podría escoger de acuerdo a la aplicación. No obstante, 300 se considera un número estadísticamente óptimo para este caso particular. Es lo suficientemente grande como para asegurar gran variedad en los escenarios, pero no tan grande como para que la computación sea muy lenta. Generadas las sendas, se deberán deshacer los cambios que se han hecho previamente. Es decir, habrá que sumar la estacionalidad diaria, la anual, deshacer la transformada Logit, deshacer la interpolación a través de la cual se había igualado la frecuencia de la serie y añadir las horas nocturnas con valor 0.

A continuación, se van a mostrar las sendas simuladas para cada área, tanto de los modelos mensuales como de los anuales.

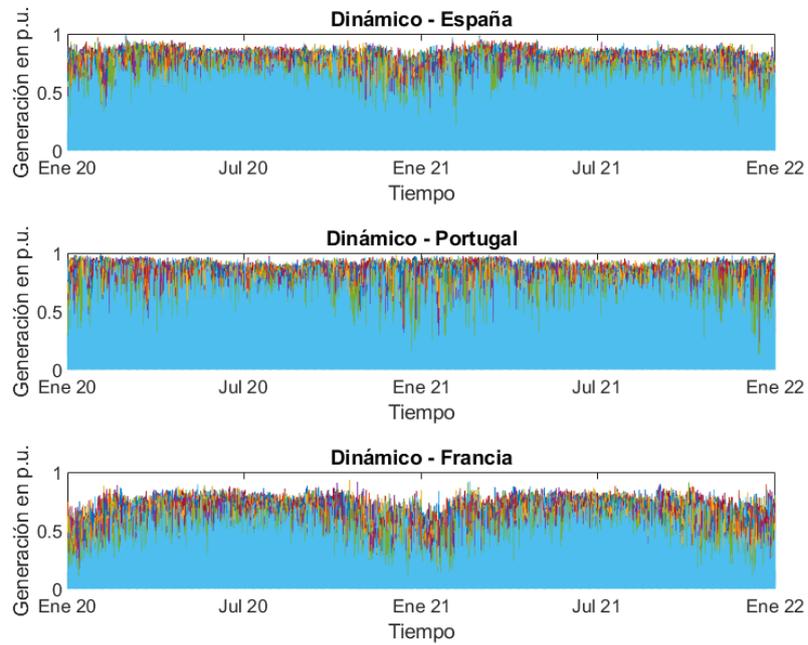


Figura 40: Sendas simuladas con modelo dinámico.

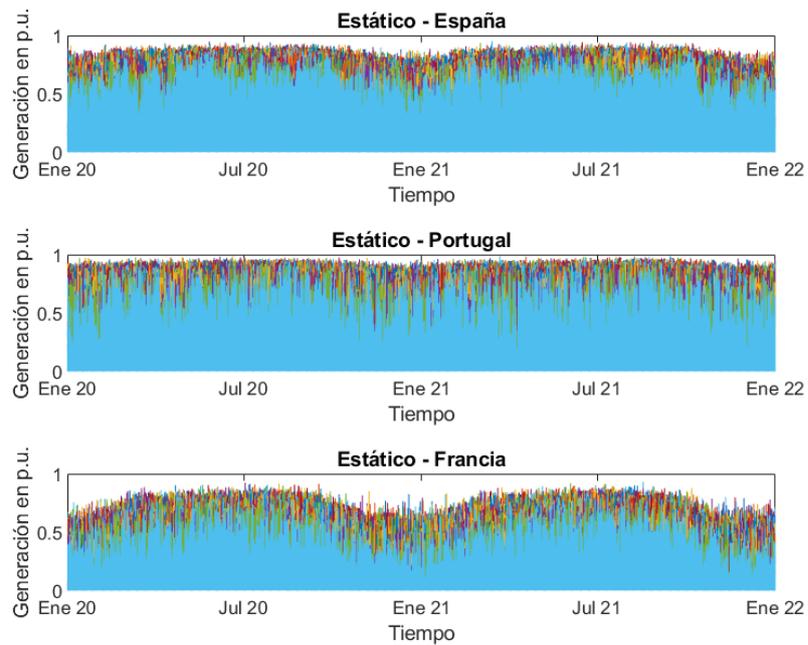


Figura 41: Sendas simuladas con modelo estático.

Las diferencias entre áreas son bastante notables, y también entre los dos métodos. En las sendas del método estático los cambios se producen de manera más suave, mientras que en las del dinámico suceden de forma mucho más abrupta.

4.6 EVALUACIÓN DEL RENDIMIENTO

Como se ha mencionado en la metodología, para comparar los escenarios generados se van a utilizar las métricas probabilísticas WS y PLF. Para el WS se han escogido intervalos del 50%, 80%, 90% y 98%. Para el PLF se han escogido los percentiles P1, P5, P30, P50, P70, P95 y P99. A continuación, se muestra un resumen de las medias de los resultados de estos indicadores para los intervalos y percentiles mencionados. Los resultados completos se pueden consultar en el Anexo II.

Tabla 8: Resumen indicadores PLF y WS.

ÁREA	Modelo	WS-F1	WS-F2	WS-Spl	PLF-F1	PLF-F2	PLF-Spl
España	Dinámico	0.1918	0.1928	0.1868	0.0090	0.0089	0.0085
	Estático	0.2055	0.2026	0.2039	0.0092	0.0089	0.0089
Portugal	Dinámico	0.2681	0.2607	0.2559	0.0130	0.0126	0.0124
	Estático	0.3356	0.3323	0.3390	0.0156	0.0153	0.0153
Francia	Dinámico	0.2781	0.2766	0.2482	0.0115	0.0112	0.0094
	Estático	0.2690	0.2642	0.2487	0.0098	0.0094	0.0092

Como se ha mencionado en 3.6, cuanto menor sea el valor de los indicadores WS y PLF, más parecida será la distribución de los datos históricos y de los simulados y mejor será el modelo ajustado. Como se puede ver en la Tabla 8, los modelos dinámicos funcionan mejor y el método que mejor captura la estacionalidad anual es la interpolación por splines.

Ahora se van a analizar distintas métricas deterministas de las series temporales. Se han sacado los valores medios, mínimos y máximos de las sendas simuladas para cada hora de cada mes. Después se ha realizado la media de generación de cada hora para todos los meses y de todos los meses para cada hora. De esta forma, se pueden comparar las simulaciones

con los datos históricos tanto a nivel horario como a nivel mensual. Las métricas mensuales serían:

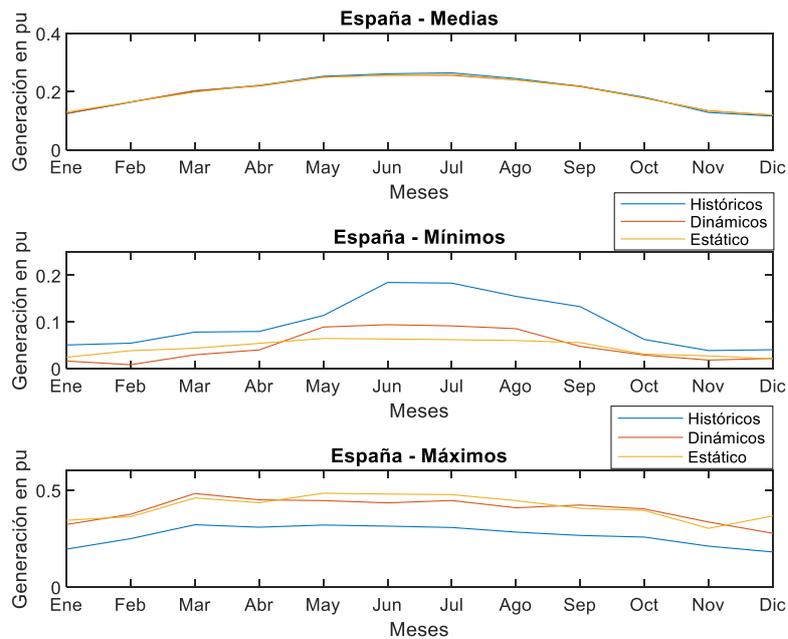


Figura 42: Métricas deterministas mensuales España.

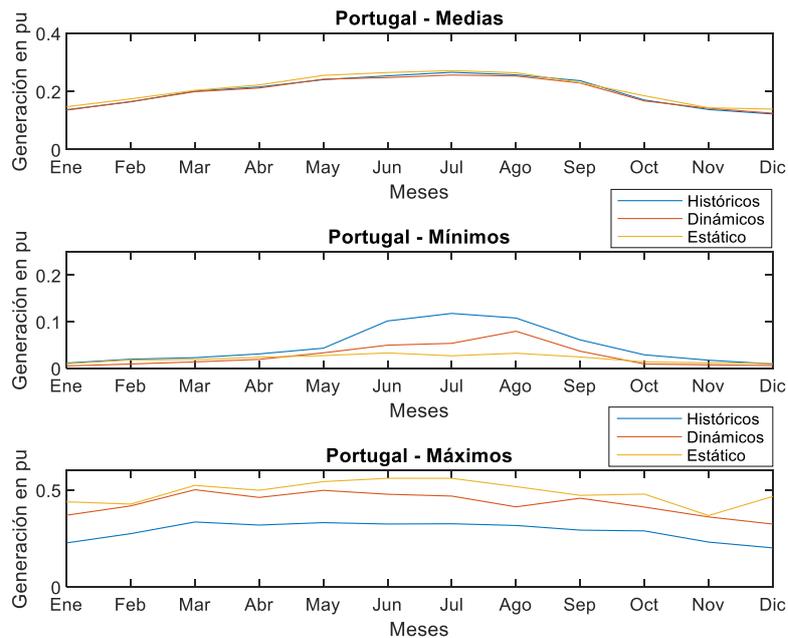


Figura 43: Métricas deterministas mensuales Portugal.

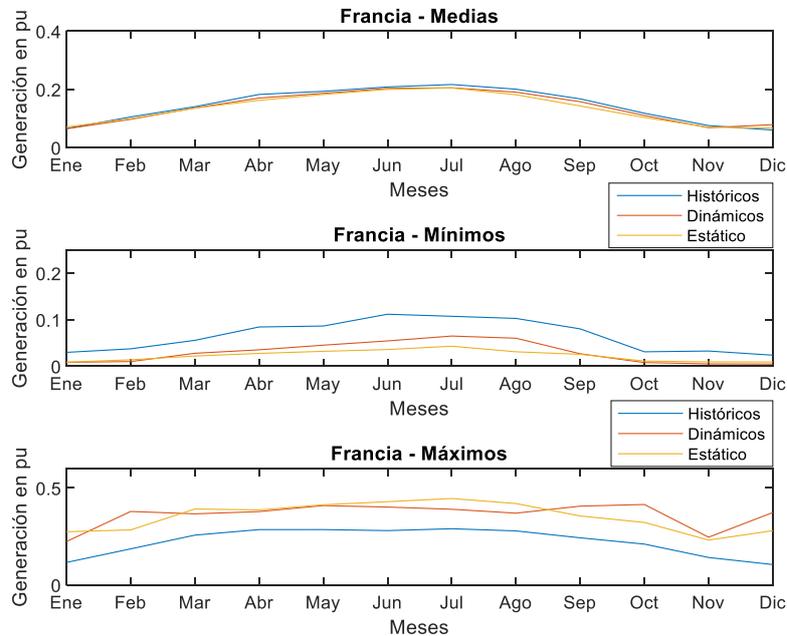


Figura 44: Métricas deterministas mensuales de Francia.

Como se puede apreciar en la Figura 42, la Figura 43 y la Figura 44, en las tres áreas tanto el modelo dinámico como el estático capturan muy bien la media de los históricos. En cuanto a los mínimos y los máximos, en los primeros las simulaciones son menores y en los segundos mayores. Esto se puede considerar como algo normal, ya que, dentro de las 300 sendas, por pura probabilidad, tienen que aparecer valores más extremos que los de los datos históricos, y son estos valores extremos los que se han cogido para las métricas. Se pueden observar algunos datos atípicos, por ejemplo, que en el caso de España el valor máximo del modelo dinámico del mes de marzo es muy superior al del histórico (Figura 42) o que en Francia el dato máximo del mes de diciembre del modelo dinámico es especialmente alto (Figura 44). Esto probablemente se deba a algunos picos de la estacionalidad anual sacada por splines, y es probable que se pudiese paliar si se tuviesen más datos históricos. De todas formas, lo importante de estas métricas es que los dos modelos consiguen capturar casi a la perfección la media de los históricos.

A nivel horario las métricas serían:

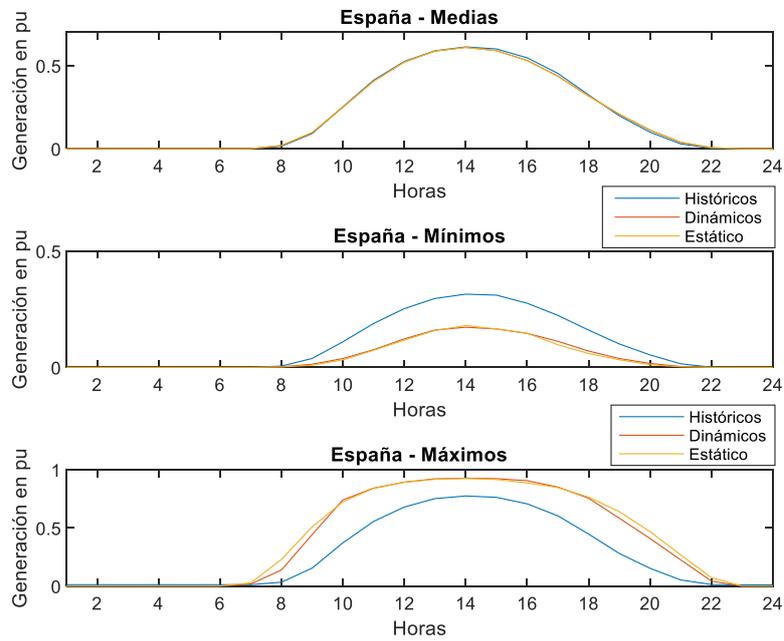


Figura 45: Métricas deterministas horarias España.

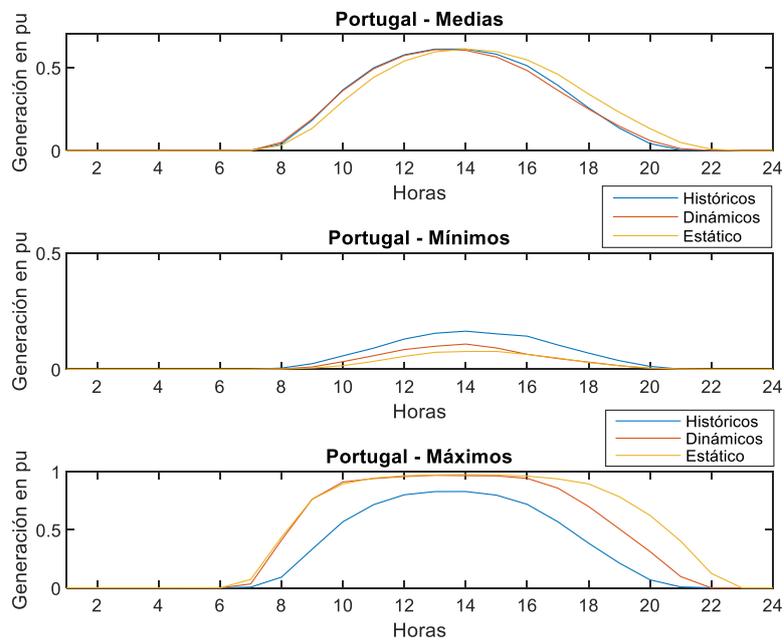


Figura 46: Métricas deterministas horarias Portugal.

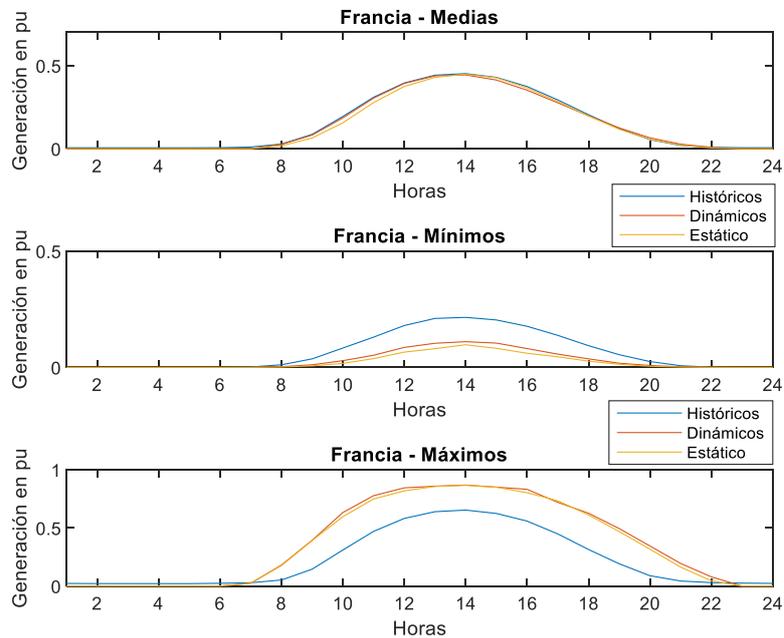


Figura 47: Métricas deterministas horarias Francia.

Nuevamente, las medias de los datos históricos y de las simulaciones coinciden muy bien, excepto en el caso del modelo estático de Portugal, donde parece que hay cierto desfase, que se debe a que este país se encuentra en una franja horaria distinta a la de los otros dos. En cuanto a los mínimos y máximos, las simulaciones quedan por debajo y por encima de los históricos, respectivamente. Como se ha mencionado antes, teniendo en cuenta la gran cantidad de escenarios que se generan, esta situación es bastante normal.

5. CONCLUSIONES Y TRABAJOS FUTUROS

Después de haber presentado los resultados obtenidos, en este apartado se van a presentar las conclusiones principales que se han sacado y se va a comentar cuáles son las posibles líneas de mejora que existen.

En primer lugar, a la vista de los resultados, se puede concluir que los escenarios generados a partir de la agregación de los modelos de descomposición aditivos y los ARIMA son realistas y arrojan resultados muy satisfactorios para capturar la dinámica de las series y su distribución asociada, teniendo en cuenta sus idiosincrasias. Se ha logrado capturar de forma precisa la dinámica propia de la generación solar y la distribución de los datos de las simulaciones coincide en gran medida con las de los datos históricos.

También se concluye que no es estrictamente necesario realizar una diferenciación de la serie para poder ajustar un modelo ARIMA. Se ha decidido optar por convertir la serie original en estacionaria mediante una serie de transformaciones (normalizaciones, capturas de estacionalidad, etc.) y los resultados han sido satisfactorios. Esto hace que el modelo gane mucho en interpretabilidad.

En cuanto a los métodos de captura de la estacionalidad anual, la interpolación por splines ha dado en este caso mejores resultados que las series de Fourier. Sin embargo, sería difícil tratar de generalizar esto, ya que pueden existir otro tipo de series temporales en los que esto no sea así.

El método de las medias ha resultado ser muy útil para capturar la estacionalidad diaria de la serie. De hecho, en un principio los modelos ARIMA que se habían ajustado tenían una parte estacional con un periodo de un día, pero, una vez se realizaron varias pruebas, quedo probado que no era necesario incluir esta parte y que con la captura de la estacionalidad por medias era suficiente.

En cuanto a los posibles trabajos futuros, aunque los resultados de este proyecto han sido bastante satisfactorios, sí que se han encontrado muchas líneas de mejora.

Uno de los primeros puntos de mejora, y también de los más sencillos, sería aumentar el número de datos históricos para las áreas. Para este caso concreto de la ratio entre el horizonte de la predicción con respecto de los históricos era demasiado alto. Con más datos sería de esperar que se pudiese mejorar el realismo de los escenarios generados a futuro.

Otro de los puntos que se podría estudiar sería el tratamiento de las horas nocturnas. En este y en muchos trabajos se ha optado por eliminar estas horas, pero existen otros en los que lo que se hace es un tratamiento especial de las mismas [37]. Con este tratamiento no sería necesario igualar la frecuencia a lo largo de la serie y se podría aprovechar mejor la información disponible en las series temporales.

Por otro lado, se podría mejorar la normalización de los datos empleando otro tipo de técnicas o incluso aplicando varias sucesivamente.

También sería interesante proponer alguna otra metodología basada en otras técnicas y comparar los distintos enfoques. Por ejemplo, técnicas de Machine Learning o técnicas/modelos matemáticos basados en distribuciones alejadas de la normalidad

Además, se podría emplear información adicional para mejorar la calidad de las predicciones, como la generación proveniente de otras renovables como, por ejemplo, la eólica. Estas fuentes de energía tienen correlaciones entre sí y la predicción ganaría tanto en precisión como en utilidad

Por último, y alejado quizás de lo que se ha hecho en este trabajo y similares, sería muy útil para inversores y agentes mejorar esta metodología para un horizonte de muy largo plazo en donde se incluyan factores como la mejora del rendimiento de los nuevos parques o nuevas ubicaciones que no sean del todo coherentes con los históricos considerados.

6. BIBLIOGRAFÍA

- [1] E. J. Fajardo Ortiz, «ResearchGate,» Agosto 2017. [En línea]. Available: https://www.researchgate.net/figure/Figura-1-Metodologia-Box-Jenkins-Fuente-Elaboracion-propia_fig1_318940178. [Último acceso: 11 Julio 2021].
- [2] J. Poushter y C. Huang, «Pew Research Center,» 2019 Febrero 2019. [En línea]. Available: <https://www.pewresearch.org/global/2019/02/10/climate-change-still-seen-as-the-top-global-threat-but-cyberattacks-a-rising-concern/>. [Último acceso: 11 Junio 2021].
- [3] AEC, ACOMET, «Vocabulario climático para comunicadores y divulgación general,» Madrid.
- [4] E. Hawkins, «Climate Lab Book,» 22 Mayo 2018. [En línea]. Available: <http://www.climate-lab-book.ac.uk/2018/warming-stripes/>. [Último acceso: 30 Octubre 2020].
- [5] Intergovernmental Panel on Climate Change, «Cambio climático 2013 Bases físicas,» 2013.
- [6] NASA, «Global Climate Change. Vital Signs on the Planet,» Nasa's Jet Propulsion Laboratory, 21 Octubre 2020. [En línea]. Available: <https://climate.nasa.gov/causes/>. [Último acceso: 26 Octubre 2020].
- [7] Agencia Internacional de la Energía, «World Energy Outlook 2020,» 2020.

- [8] International Renewable Energy Agency, «Renewable Capacity Statistics 2020,» IRENA, Masdar City, 2020.
- [9] Premium, «Premium Energia,» 22 Febrero 2017. [En línea]. Available: <http://premiumenergia.es/historia-de-la-energia-solar/>. [Último acceso: 30 Junio 2021].
- [10] El Orden Mundial, «El Orden Mundial,» 7 Enero 2021. [En línea]. Available: <https://elordenmundial.com/mapas/potencial-energia-solar-mundo/>. [Último acceso: 11 Enero 2021].
- [11] Unión Española Fotovoltaica UNEF, «El sector fotovoltaico: hacia una nueva era. Informe anual 2020,» 2020.
- [12] A. d. Roca, «Magnus Commodities,» 1 Octubre 2019. [En línea]. Available: <https://www.magnuscmd.com/es/subastas-de-renovables-que-son-y-que-hay-detras-de-la-diversidad-de-precios/>. [Último acceso: 1 Julio 2021].
- [13] Lazard, «Lazard's Levelized Cost of Energy Analysis - Version 14.0,» 2020.
- [14] Lazard, «Lazard,» 19 Octubre 2020. [En línea]. Available: <https://www.lazard.com/perspective/levelized-cost-of-energy-and-levelized-cost-of-storage-2020/>. [Último acceso: 28 Octubre 2020].
- [15] International Renewable Energy Agency (IRENA), «Renewable energy auctions: Status and trends beyond price.,» Abu Dhabi, 2019.
- [16] S. Roth, «California has too much solar power. That may be goo for ratepayers,» *Los Angeles Times*, 5 Junio 2019.

- [17] A. D. Rigal, Interviewee, *CEO de Aleasoft y doctor en Inteligencia Artificial*. [Entrevista]. 26 Noviembre 2020.
- [18] R. Ahmed, V. Sreeram, Y. Mishra y M. Arif, «A review and evaluation of the state-of-the-art in PV solar power,» *Renewable and Sustainable Energy Review*, nº 124, 2020.
- [19] S. Sobria, S. Koochi-Kamali y N. A. Rahim, «Solar photovoltaic generation forecasting methods: A review,» *Energy Conversion and Management*, vol. 156, pp. 459-497, 2018.
- [20] IBM Knowledge Center, «IBM,» [En línea]. Available: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/neuralnet/neuralnet_model.html. [Último acceso: 28 Octubre 2020].
- [21] S. Haykin, *Neural Networks and Learning Machines*, Hamilton: Pearson, 2009.
- [22] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. Idna Idris, W. Van Deventer, B. Horan y A. Stojcevski, «Forecasting of photovoltaic power generation and model optimization: A review,» *Renewable and Sustainable Energy Reviews*, nº 81, pp. 912-928, 2018.
- [23] Y. Ren, P. Suganthan y N. Srikanth, «Ensemble methods for wind and solar power forecasting — A state-of-the-art review.,» *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 82-91, 2015.
- [24] H. T. C. Pedro y C. F. M. Coimbra, «Assessment of forecasting techniques for solar power production with no exogenous inputs,» *Solar Energy*, nº 86, pp. 2017-2028, 2012.

- [25] B. Singh y D. Pozo, «A Guide to Solar Power Forecasting using ARMA Models,» 2018.
- [26] S. Atique, S. Noureen, V. Roy, V. Subburaj, S. Bayne y J. Macfie, «Forecasting of total daily solar energy generation using ARIMA: A case study,» *IEEE*, pp. 114-119, 2019.
- [27] M. Alanazi, A. Alanazi y A. Khodaei, «Long-Term Solar Generation Forecasting,» *IEEE*, 2016.
- [28] G. Marulanda, A. Bello, J. Cifuentes y J. Reneses, «Wind Power Long-Term Scenario Generation,» *Energies*, vol. 13, nº 3427, 2020.
- [29] SolarPower Europe, EU Market Outlook for Solar Power, Bruselas, 2021.
- [30] REE - Red Eléctrica de España, «El sistema eléctrico español. Avance 2020.,» Alcobendas, 2021.
- [31] REE - Red Eléctrica de España, «Las energías renovables en el sistema eléctrico español 2019,» Alcobendas, 2020.
- [32] REN – Redes Energéticas Nacionais, «Carrying the future: Reports & Accounts 2020.,» Lisboa, 2021.
- [33] Global Power Plant Database, «Global Power Plant Database,» 12 Junio 2019. [En línea]. Available: <https://datasets.wri.org/dataset/globalpowerplantdatabase>. [Último acceso: 1 Junio 2021].
- [34] RTE - Réseau de transport d'électricité, «Bilan électrique 2020,» París, 2021.
- [35] R. J. Hyndman y G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, 2018.

- [36] B. Liu, J. Nowotarski, T. Hong y R. Weron, «Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts,» *Transactions on smart grid*, pp. 730-737, 2017.
- [37] X. Ghislain Agoua, R. Girard y G. Karioniotakis, «Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production,» *IEEE Transactions on Sustainable Energy*, vol. 9, nº 2, pp. 538-546, 2018.
- [38] «Areatecnología,» [En línea]. Available: <https://www.areatecnologia.com/electricidad/paneles-solares.html>. [Último acceso: 11 Enero 2021].
- [39] Ingelibre, «Ingelibre,» 9 Noviembre 2014. [En línea]. Available: <https://ingelibreblog.wordpress.com/2014/11/09/influencia-de-la-irradiacion-y-temperatura-sobre-una-placa-fotovoltaica/>. [Último acceso: 11 Enero 2021].
- [40] R. G. Cordero, «SunFields Europe,» 2017. [En línea]. Available: <https://www.sfe-solar.com/paneles-solares/tipos/>. [Último acceso: 11 Enero 2021].
- [41] «Back 2 Back Distribution,» 16 Enero 2018. [En línea]. Available: <https://www.back2back.ae/solar-panels-are-they-worth-the-investment/>. [Último acceso: 11 Enero 2021].
- [42] «PREMIUM Energía,» PREMIUM Energía, 22 Febrero 2017. [En línea]. Available: <http://premiumenergia.es/historia-de-la-energia-solar/>. [Último acceso: 24 Diciembre 2020].
- [43] P. Rodó, «Economipedia,» Economipedia, [En línea]. Available: <https://economipedia.com/definiciones/modelo-autorregresivo-ar.html>. [Último acceso: 29 Octubre 2020].

- [44] Entsoe, «Entsoe,» 1 Enero 2019. [En línea]. Available: <https://www.entsoe.eu/data/map/>. [Último acceso: 31 Mayo 2021].
- [45] V. A. Golovko, «Deep Learning: An Overview and Main Paradigms,» *Optical Memory and Neural Networks*, vol. XXVI, nº 1, pp. 1-17, 2017.

ANEXO I

Alineación con los objetivos de desarrollo sostenible

Este proyecto está directamente alineado con varios de los ODS establecidos por la ONU e indirectamente relacionado con casi todos los demás.



Figura 48: Objetivos de Desarrollo Sostenible de la ONU.

Como se ha mencionado anteriormente, la elaboración de escenarios de generación solar fotovoltaica es una tarea fundamental para facilitar la integración de esta tecnología en la generación eléctrica, y para fomentar la sustitución de energías contaminantes por esta. Por tanto, el objetivo con el que más se relaciona es el 7. Además, como la energía solar tiene un impacto ambiental mucho menor que otras tecnologías actuales y ayuda a mitigar el cambio climático, también se puede relacionar con los objetivos 11, 13, 14 y 15. El objetivo 3 también estaría relacionado con este proyecto, ya que la sustitución de tecnologías contaminantes por solar supondría una reducción de la contaminación a nivel mundial y por tanto una reducción de las enfermedades asociadas a esta. Por último, la industria fotovoltaica requiere de trabajadores de varios sectores y con diversos grados de especialización, por lo que también se contribuiría a conseguir los objetivos 8 y 9.

ANEXO II

Se muestran en este anexo los resultados de las métricas probabilísticas PLF y WS para los dos métodos y los tres tipos de captura de la estacionalidad anual.

Tabla 9: Resultados Winkler Score.

Método	Área	Modelo	50%	80%	90%	98%	Media
Fourier 1	España	Dinámico	0.1161	0.1610	0.1948	0.2953	0.1918
		Estático	0.1196	0.1719	0.2095	0.3210	0.2055
	Portugal	Dinámico	0.1691	0.2355	0.2804	0.3875	0.2681
		Estático	0.2024	0.2842	0.3433	0.5125	0.3356
	Francia	Dinámico	0.1455	0.2050	0.2527	0.5093	0.2781
		Estático	0.1205	0.1833	0.2421	0.5302	0.2690
Fourier 2	España	Dinámico	0.1157	0.1623	0.1966	0.2966	0.1928
		Estático	0.1157	0.1687	0.2061	0.3199	0.2026
	Portugal	Dinámico	0.1637	0.2274	0.2712	0.3804	0.2607
		Estático	0.1978	0.2794	0.3391	0.5129	0.3323
	Francia	Dinámico	0.1410	0.2010	0.2525	0.5118	0.2766
		Estático	0.1151	0.1774	0.2375	0.5266	0.2642
Splines	España	Dinámico	0.1102	0.1565	0.1899	0.2904	0.1868
		Estático	0.1152	0.1687	0.2072	0.3244	0.2039
	Portugal	Dinámico	0.1609	0.2228	0.2666	0.3732	0.2559
		Estático	0.1976	0.2806	0.3428	0.5353	0.3390
	Francia	Dinámico	0.1162	0.1673	0.2181	0.4910	0.2482
		Estático	0.1129	0.1656	0.2175	0.4987	0.2487

Tabla 10: Resultados Pinball Loss Function.

Método	Área	Modelo	P1	P5	P30	P50	P70	P95	P99	Media
Fourier 1	España	Dinámico	0.0012	0.0049	0.0164	0.0183	0.0154	0.0049	0.0017	0.0090
		Estático	0.0013	0.0050	0.0163	0.0183	0.0162	0.0055	0.0019	0.0092
	Portugal	Dinámico	0.0020	0.0075	0.0248	0.0270	0.0216	0.0065	0.0019	0.0130
		Estático	0.0025	0.0089	0.0285	0.0317	0.0268	0.0083	0.0026	0.0156
	Francia	Dinámico	0.0012	0.0046	0.0183	0.0228	0.0214	0.0080	0.0039	0.0115
		Estático	0.0011	0.0044	0.0156	0.0186	0.0171	0.0077	0.0042	0.0098
Fourier 2	España	Dinámico	0.0013	0.0048	0.0161	0.0180	0.0154	0.0050	0.0017	0.0089
		Estático	0.0014	0.0051	0.0161	0.0176	0.0152	0.0053	0.0018	0.0089
	Portugal	Dinámico	0.0019	0.0072	0.0239	0.0261	0.0211	0.0063	0.0019	0.0126
		Estático	0.0025	0.0088	0.0279	0.0311	0.0263	0.0082	0.0026	0.0153
	Francia	Dinámico	0.0011	0.0046	0.0178	0.0221	0.0207	0.0081	0.0040	0.0112
		Estático	0.0011	0.0043	0.0151	0.0178	0.0161	0.0075	0.0042	0.0094
Splines	España	Dinámico	0.0012	0.0046	0.0155	0.0173	0.0146	0.0049	0.0017	0.0085
		Estático	0.0014	0.0051	0.0160	0.0176	0.0152	0.0053	0.0018	0.0089
	Portugal	Dinámico	0.0018	0.0071	0.0236	0.0256	0.0207	0.0062	0.0019	0.0124
		Estático	0.0027	0.0091	0.0279	0.0308	0.0260	0.0081	0.0026	0.0153
	Francia	Dinámico	0.0011	0.0040	0.0148	0.0181	0.0169	0.0069	0.0038	0.0094
		Estático	0.0009	0.0039	0.0146	0.0177	0.0162	0.0070	0.0040	0.0092

ANEXO III

En este anexo se proporciona información adicional sobre los modelos ARIMA.

MODELOS DE REGRESIÓN

El análisis de regresión se refiere al método que determina una relación funcional del modelo entre la respuesta (dependiente) y parámetros de predicción (independientes). Representan un proceso repetitivo, por lo que los resultados se utilizan para analizar, verificar, criticar y modificar los datos de entrada.

MODELOS AUTOREGRESIVOS (AR)

En un modelo de regresión múltiple, la variable que interesa se predice a partir de la combinación lineal de los predictores. En un modelo autorregresivo, como su propio nombre indica, la variable que interesa se predice a partir de la combinación lineal de los valores pasados de esa misma variable. La ecuación que describe un modelo autorregresivo es:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

donde ε_t es ruido blanco, ϕ son los parámetros que hay que estimar y p es un número entero no negativo que representa el orden del modelo autorregresivo. Es decir, si p es igual a cinco y los datos de la serie temporal son mensuales, para elaborar la predicción de cierto valor se utilizarán los cinco meses anteriores. A este modelo se le conoce como AR(p).

MODELOS DE MEDIA MÓVIL (MA)

En vez de usar valores pasados de la variable a predecir en una regresión, un modelo de media móvil usa los errores de predicción pasados en un modelo similar a una regresión:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

donde θ son los parámetros que hay que estimar y q es el orden del modelo de media móvil. No se observan los valores de ε_t , por lo que esto no es una regresión en sentido normal. Cada valor de y_t puede suponerse como una media móvil ponderada de los pasados errores de predicción.

MODELOS AUTOREGRESIVOS INTEGRADOS DE MEDIA MÓVIL (ARIMA)

Si se combinan los modelos autorregresivos con los de media móvil se obtiene un modelo ARMA. Una de las restricciones que tienen los modelos ARMA es que la serie temporal que se quiera predecir con estos modelos tiene que ser necesariamente estacionaria.

Una serie estacionaria es aquella cuyas propiedades no cambian en función del momento en el que sea observada. Es decir, la media y la varianza se mantienen constantes. Por este motivo, las series temporales que tengan tendencia o algún tipo de estacionalidad no pueden ser estacionarias. Por el contrario, una serie de ruido blanco es necesariamente estacionaria. Una serie que tenga ciclos esporádicos también es estacionaria, dado que los ciclos pueden aparecer en cualquier momento y ser de cualquier magnitud, por lo que no podemos predecirlos. En general, una serie estacionaria no tendrá patrones que nos ayuden a predecir su comportamiento a largo plazo.

En el caso que abarca el proyecto, la generación fotovoltaica, no se podrían emplear modelos ARMA directamente, ya que la serie temporal no es estacionaria, debido a que hay una componente estacional diaria muy clara (se genera durante el día y no por la noche) y también una anual (se genera más en verano que en invierno).

Una de las posibles formas de convertir una serie temporal no estacionaria en estacionaria es restar a cada dato de la serie temporal el valor anterior.

$$y'_t = y_t - y_{t-1}$$

De esta forma, se puede conseguir igualar la media y la varianza a lo largo de toda la serie, y por tanto, convertir la serie en estacionaria. La diferenciación se puede realizar, no sólo sobre el valor de la serie inmediatamente anterior, sino sobre valores en periodos anteriores

$$y'_t = (y_t - y_{t-m})$$

Ahora bien, no siempre se puede convertir una serie no estacionaria en estacionaria aplicando una sola diferenciación. Puede haber casos en los que sea necesario hacer más de una.

De esta forma, para poder elaborar predicciones sobre series no estacionarias se pueden emplear los modelos ARIMA, que no son más que modelos ARMA en los que se ha realizado alguna diferenciación sobre la variable a predecir.

La ecuación de los modelos ARIMA(p,d,q) quedaría de la siguiente manera.

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

donde d es el grado de la diferenciación e y'_t es el valor t de la serie diferenciada.

La notación que se suele emplear para mostrar las ecuaciones de los modelos ARIMA no es esta, sino la notación *Backshift*. B es un operador que nos permite saltar al dato anterior, tal y como muestra la ecuación:

$$By_t = y_{t-1}$$

También se puede saltar varias unidades de tiempo hacia atrás mediante el empleo de potencias:

$$B^2 y_t = y_{t-2} \text{ ó } B^m y_t = y_{t-m}$$

La ecuación final de un modelo ARIMA con esta notación quedaría de la siguiente manera:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(y'_t - \mu) = (1 + \theta_1 B + \dots + \theta_q B^q)\varepsilon_t$$

Las condiciones de invertibilidad y estacionaridad que se aplican a los modelos AR y MA también se aplican a ARIMA.

De lo mencionado anteriormente se puede concluir que si no existe ningún tipo de diferenciación se estaría hablando de un modelo ARMA. Sin embargo, en este trabajo y para estos casos, se hablará de modelos ARIMA con $d = 0$. Se mantiene el nombre de ARIMA aunque no haya diferenciación para no perder generalidad y para coincidir con el nombre que más se suele emplear en la literatura.

6.1.1 MODELOS AUTORREGRESIVOS INTEGRADOS DE MEDIA MÓVIL ESTACIONALES (SARIMA)

Muchas de las series que se quieren predecir, como es el caso de la generación fotovoltaica, tienen componentes estacionales, es decir, su comportamiento se repite cada cierto tiempo. A los modelos ARIMA se les puede añadir una componente estacional (Seasonal) para que sean capaces de capturar estos comportamientos.

La ecuación final de los modelos SARIMA sería la siguiente:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} +$$

$$\phi_1 y'_{t-1*T} + \dots + \phi_p y'_{t-p*T} + \theta_1 \varepsilon_{t-1*T} + \dots + \theta_Q \varepsilon_{t-Q*T} +$$

$$\varepsilon_t$$

Donde T es el periodo de la estacionalidad. Utilizando la notación Backshift el modelo SARIMA quedaría:

$$\theta_p(B)\Theta_P(B^S)(1-B)^d(1-B^S)^D y_t = w_q(B)W_Q(B^S)a_t$$

Donde $\theta_p(B)$, $w_q(B)$, $\Theta_P(B^S)$, y $W_Q(B^S)$, son polinomios de B de grados p, q, P y Q , respectivamente, y son:

$$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p$$

$$w_q(B) = 1 - w_1 B - w_2 B^2 - \dots - w_q B^q$$

$$\Theta_p(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_p B^{pS}$$

$$W_Q(B^S) = 1 - W_1 B^S - W_2 B^{2S} - \dots - W_Q B^{QS}$$