



Universidad
Politécnica
de Cartagena

MIEMBRO DE



EUROPEAN
UNIVERSITY OF
TECHNOLOGY



UNIVERSIDAD POLITÉCNICA DE CARTAGENA

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN



Trabajo Fin de Estudios

Grado en Ingeniería en Sistemas de Telecomunicaciones

MODELOS TEMPORALES PARA SEÑALES PROCEDENTES DEL MAR MENOR

Mercedes Navarro Martínez

Trabajo realizado en el marco del programa "ENCUENTRO TF"

Director: Dr. Mathieu Kessler

Codirectora: Dra. María del Carmen Bueso Sánchez

Codirectora: Rosa Martínez Álvarez - Castellanos

2023

Agradecimientos

En primer lugar, me gustaría agradecer a mi director Mathieu Kessler y a mi codirectora María del Carmen Bueso por su ayuda y apoyo.

Agradecer a mi otra codirectora Rosa Martínez, y a Ivan Felis por acompañarme y aconsejarme y compartirme su conocimiento en todo el proceso.

A mi novio, Dani, la persona que me ha aguantado día a día durante este proceso, gracias por tu paciencia y cariño.

Gracias a la vida y a la carrera por permitirme conocer a mi mejor amiga, Miriam, y gracias a ella por estos 5 años de amistad, consejos y apoyo incondicional.

A mi mejor amigo, Jose, que no le vale solo con ser mi amigo también tiene que trabajar a mi lado.

Por último, gracias a mis padres, que luchan día a día porque yo pueda hacer lo que me gusta. Nunca podré agradecerlos lo suficiente.

Índice

1. Introducción	5
2. Objetivos	6
3. Estado del arte	7
3.1. Componentes de las series temporales.....	8
3.1.1. Estacionariedad.....	8
3.1.2. Estacionalidad	8
3.1.3. Tendencia.....	9
3.1.4. Residuo.....	9
3.2. Tipos de descomposición de una serie temporal.....	10
3.3. Modelos de series temporales.....	11
3.3.1. Modelo Autorregresivo (AR)	12
3.3.2. Modelo de Medias Móviles (MA)	13
3.3.3. Modelo ARMA	14
3.3.4. Modelo ARIMA	15
3.3.5. Modelo SARIMA	15
3.3.6. Modelo Prophet	16
3.4. Resumen del comportamiento de las funciones de autocorrelación de los modelos	17
4. Metodología	18
4.1. Fase 0: Visualización y limpieza de datos	19
4.1.1. Visualización de datos	19
4.1.2. Limpieza de datos	20
4.2. Fase 1: Identificación de patrones	21
4.2.1. Descomposición de la serie	21
4.2.2. Media móvil y desviación típica	21
4.2.3. Prueba de Dickey - Fuller	22
4.2.4. Función de autocorrelación (ACF).....	22
4.2.5. Función de autocorrelación parcial (PACF)	22
4.3. Fase 2: Intento de mejora de ajuste: transformación de datos.....	23
4.3.1. Diferenciación.....	23
4.3.2. Desestacionalización	23
4.4. Fase 3: Selección del modelo y explicación de patrones.....	24
4.5. Fase 4: Aplicación del modelo predictivo.....	25
4.5.1. División del conjunto de datos en prueba y entrenamiento	25
4.5.2. Predicción.....	25
4.5.3. Evaluación bondad del modelo predictivo	26
4.6. Software utilizado.....	27
5. Aplicación a datos sintéticos	28
5.1. AR	28
5.2. MA	30
5.3. ARMA.....	31
5.4. Prophet.....	32
6. Aplicación a datos del Mar Menor. Caso real	33
6.1. Fase 0: Visualización y limpieza de datos	34
6.2. Fase 1: Identificación de patrones	35
6.3. Fase 2: Intento de mejora de ajuste: diferenciación y desestacionalización	41
6.4. Fase 3: Selección del modelo y explicación de patrones.....	42
6.5. Fase 4: Aplicación de modelos predictivos.....	43
7. Conclusiones	45
8. Líneas futuras	46
Anexos	47
A. Aplicación a datos reales libro Daniel Peña.....	47
Bibliografía	53
Lista de figuras	54
Lista de tablas	55

RESUMEN

La problemática respecto a la variabilidad de las series temporales, junto con la creciente preocupación sobre el estado del Mar Menor, ha impulsado la necesidad de desarrollar una metodología efectiva para el análisis y pronóstico de datos temporales. En este Trabajo de Fin de Estudios (TFE), se ha seguido una metodología sistemática que abarca desde la exploración y clasificación de la serie temporal hasta la selección del modelo más adecuado para su predicción.

Mediante la aplicación de esta metodología a datos reales del Mar Menor, se ha validado su eficacia y precisión en el análisis de series temporales complejas. Se destacó que los modelos estadísticos son idóneos para predicciones a corto plazo, mientras que los modelos de inteligencia artificial son más apropiados para proyecciones futuras. Los resultados obtenidos ofrecen un enfoque sólido para la comprensión y gestión de datos temporales, contribuyendo al avance en la toma de decisiones en el contexto del Mar Menor y en otros ámbitos que enfrentan desafíos similares.

ABSTRACT

The growing concern about the state of the Mar Menor and the problem of the variability of the time series, has driven the need to develop an effective methodology for the analysis and forecasting of time series data. In this TFE, a systematic methodology has been proposed, ranging from the exploration and classification of the time series to the selection of the most appropriate model for its prediction.

By applying this methodology to real data from the Mar Menor, its effectiveness and accuracy in the analysis of complex time series has been validated. It was highlighted that statistical models are suitable for short-term predictions, while artificial intelligence models are more appropriate for future projections. The results obtained provide a robust approach to understanding and managing temporal data, contributing to the advancement of decision making in Mar Menor context and other areas facing similar challenges.

1. Introducción

En el ámbito de la ciencia de datos y el aprendizaje automático, las series temporales juegan un papel fundamental en el análisis y la predicción de datos que evolucionan a lo largo del tiempo. Estas series representan observaciones secuenciales registradas en intervalos regulares, como datos económicos o datos climáticos, entre otros. El objetivo principal al trabajar con series temporales es poder comprender su comportamiento y predecir su evolución futura. En este contexto, el desarrollo de modelos predictivos precisos y efectivos es esencial para lograr resultados confiables. Sin embargo, seleccionar el modelo más adecuado para una serie temporal en particular puede ser un desafío, ya que existen una amplia variedad de enfoques y algoritmos disponibles. Cada modelo tiene sus propias suposiciones y capacidades, y su rendimiento puede variar significativamente según las características de la serie temporal en cuestión.

En este Trabajo Fin de Estudios (TFE), se aborda la necesidad de establecer un proceso estandarizado para la selección del modelo predictivo adecuado a una serie temporal en función de sus características específicas. El **objetivo principal es llevar a cabo una metodología** que permita identificar el enfoque más idóneo para cada situación, maximizando así el rendimiento de los modelos y minimizando los errores de predicción. Para lograr este objetivo, se llevará a cabo un análisis exhaustivo de las características de las series temporales que pueden influir en la elección del modelo predictivo. Estas características podrían incluir la presencia de tendencias, estacionalidad o dependencia temporal. Además, se investigarán y evaluarán diferentes enfoques y algoritmos populares utilizados en el análisis de series temporales, como los modelos autorregresivos (AR) y los modelos de medias móviles (MA) o el modelo Facebook Prophet.

El proceso estandarizado propuesto se basará en un enfoque sistemático y objetivo, utilizando criterios claros y métricas de evaluación relevantes. Así, se aplicará una metodología que guíe a los usuarios a través de diferentes etapas, desde la exploración inicial y el preprocesamiento de los datos hasta la selección y ajuste de los modelos predictivos más adecuados.

En resumen, este TFE se centra en el desarrollo de un proceso estandarizado para la selección del modelo predictivo más adecuado a una serie temporal, considerando sus características particulares. Al establecer una metodología sólida y objetiva, se busca proporcionar una guía práctica y así mejorar la calidad de las predicciones en diversos dominios y optimizar la toma de decisiones basada en datos temporales.

Para ello, este trabajo se ha estructurado de la siguiente manera: en primer lugar, en el Capítulo 3, se revisarán y se recopilará un estado del arte sobre las series temporales y sus principales modelos clásicos. Esto irá seguido de una metodología para el análisis de las series, que se realizará en el Capítulo 4 mediante las diferentes fases entre las que se encuentra la limpieza de los datos o la descomposición de la serie. Por último, en los Capítulos 5 y 6, esta metodología será aplicada tanto a datos sintéticos como reales del Mar Menor comprobando así su correcto funcionamiento.

2. Objetivos

El **objetivo principal** de esta investigación consiste en describir una metodología que permita seleccionar el modelo más apropiado de los que se evalúan en el trabajo en función de las características específicas de cada serie temporal y con ayuda de técnicas de evaluación y comparación como los criterios de información de Akaike (AIC) y Bayesiano (BIC).

Para ello, se tendrán que llevar a cabo una serie de objetivos adicionales que se detallan a continuación:

- Realizar una revisión actualizada del estado del arte en técnicas de análisis y modelos de series temporales. Esta revisión irá desde los fundamentos teóricos hasta las aplicaciones prácticas, permitiendo obtener una visión completa de las metodologías existentes y su importancia en el análisis de series temporales.
- Analizar las características, patrones y componentes de cada serie temporal. Esto implicará una visualización de la serie, descomposición de la serie en sus componentes y análisis de métricas de clasificación de series. Esto permitirá clasificar y comprender mejor las particularidades de cada serie temporal.
- Obtener la predicción de las series temporales utilizando el modelo seleccionado. Se llevará a cabo la predicción tanto en datos sintéticos como en series temporales de datos observados en el Mar Menor, lo que permitirá evaluar el desempeño y la robustez de la metodología desarrollada. Se compararán los resultados obtenidos con las observaciones reales y se evaluará la precisión de los pronósticos mediante métricas de error, como la raíz del error cuadrático medio (RMSE).

3. Estado del arte

Las series temporales son un conjunto de datos tomados a intervalos de tiempo regulares que se ordenan secuencialmente dependiendo del tiempo y de su ocurrencia y representan matemáticamente un sistema físico permitiendo y siendo de interés su análisis, modelización, caracterización y predicción a medio y largo plazo. Además, una serie temporal puede ser considerada como la superposición de varios componentes elementales: tendencia, estacionalidad y residuo. El análisis y modelización permiten encontrar una descripción del sistema y determinar sus propiedades y componentes fundamentales [1], que se verán con más detalle en la sección 3.1, estas características particulares distinguen las series temporales de otros tipos de datos [2]. A la hora de analizar las propiedades ya mencionadas de las series se debe tener en cuenta la descomposición que se aplicará en cada caso, ya que existe una descomposición aditiva y otra de tipo multiplicativa, que se seleccionarán, en gran parte, dependiendo del análisis gráfico de la serie, como se explica en la sección 3.2.

Cabe destacar que uno de los aspectos más importantes en el análisis y predicción de una serie de tiempo es la representación gráfica y el instrumento fundamental para analizar las propiedades de una serie temporal es el coeficiente de autocorrelación, que mide la correlación y el grado de asociación entre observaciones, lo que permite construir el modelo apropiado para los datos [3]. Algunos de los modelos estadísticos clásicos más usados para el análisis y pronóstico de una serie son los modelos autorregresivos (AR), de medias móviles (MA), autorregresivos de medias móviles (ARMA), autorregresivos integrados de medias móviles (ARIMA) y ARIMA estacional (SARIMA) y se van a describir con detalle en la sección 3.3.

Para el análisis de las series temporales se suelen seguir una serie de pasos, en primer lugar, es importante realizar un preprocesado de los datos asegurando la calidad y evitando errores en los datos o datos faltantes y desnormalizados [4]. Otro paso es clasificar la serie temporal para así poder tratarla de la manera correcta y elegir el modelo adecuado para el tipo de serie temporal. El último paso sería, según el modelo seleccionado, predecir la serie temporal, las buenas predicciones capturan los cambios y patrones en datos históricos [5].

3.1. Componentes de las series temporales

3.1.1. Estacionariedad

El primer paso para analizar las series temporales es estudiar si la serie es estacionaria, requerimiento para algunos de los modelos de series temporales que vamos a estudiar más adelante. Una serie es estacionaria cuando es estable a lo largo del tiempo, es decir, su media y su varianza son constantes y, además, no presenta tendencia o estacionalidad. En general, una serie procedente de un proceso estacionario no cumple patrones a largo plazo, aunque sí se permiten algunos comportamientos cíclicos que no tiene duración fija [6]. Gráficamente una serie estacionaria tiene forma estrictamente horizontal como vemos en la Figura 1, ejemplo (a), en la que la serie varía de forma aleatoria, pero se mantiene en torno a una línea horizontal y con media constante.

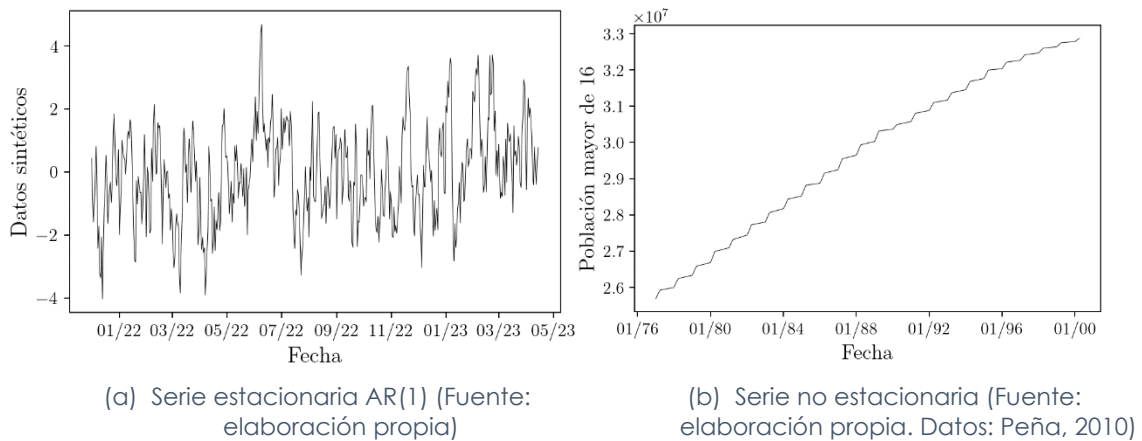


Figura 1. Ejemplo de estacionariedad

3.1.2. Estacionalidad

La estacionalidad es la propiedad de las series temporales que varían periódicamente y de forma predecible en un periodo inferior o igual a un año. La estacionalidad tiene siempre un periodo fijo y conocido [5]. Por ejemplo, una serie estacional sería una que represente la temperatura ya que aumentaría en verano y en invierno descendería.

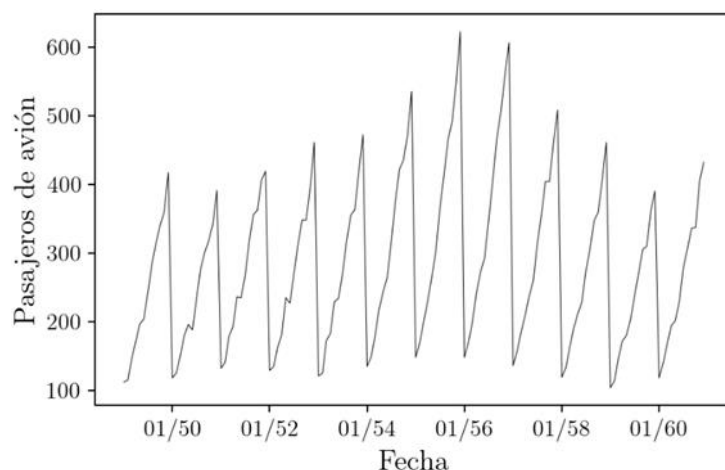


Figura 2. Serie estacional sin tendencia (Fuente: elaboración propia. Datos: Peña, 2010)

3.1.3. Tendencia

La tendencia es el movimiento a largo plazo de una serie de tiempo. Puede ser de varios tipos como creciente, decreciente, exponencial, lineal, estocástica, etc. Al tener una tendencia no podemos tener estacionariedad, ya que la serie no evoluciona en torno a un nivel constante [3].

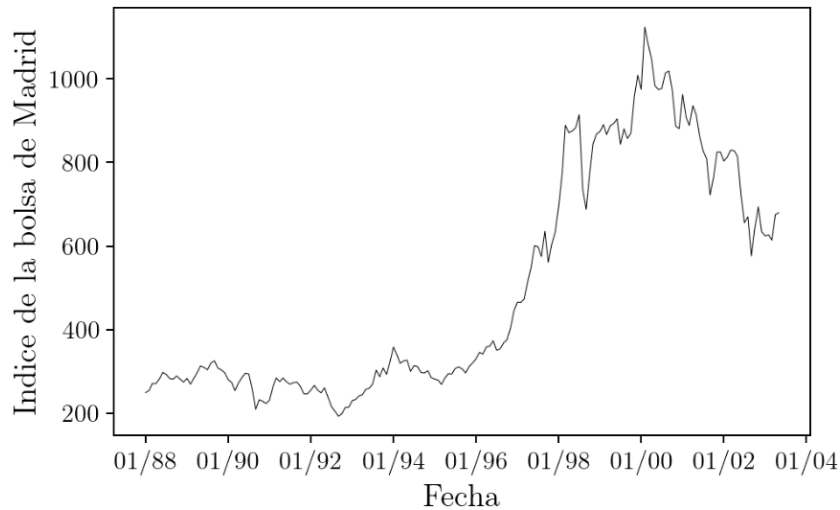


Figura 3. Serie con tendencia creciente (Fuente: elaboración propia. Datos: Peña, 2010)

3.1.4. Residuo

Después de eliminar las componentes de tendencia y estacionalidad de la serie temporal, la parte restante se conoce como componente irregular o residuo. Los residuos no se pueden explicar por ninguna tendencia o comportamiento estacional pero aun así pueden causar un impacto en la dinámica de la serie temporal. Mediante el análisis de residuos podemos contrastar cuál de los modelos elegidos es el más adecuado o si el modelo elegido es adecuado ya que si la serie está bien identificada, y el modelo ajustado es el mejor, estos residuos no deben presentar ninguna estructura (lo que denominamos ruido blanco) [7].

3.2. Tipos de descomposición de una serie temporal

El tipo de descomposición de la serie define el procedimiento a seguir a la hora de analizarla. La descomposición clásica considera que una serie temporal dada está compuesta por su tendencia, componente estacional y residuo. Hay dos posibles modelos y se asume uno u otro en función del comportamiento de la serie. Si denotamos por S_t a la componente estacional, T_t a la tendencia y R_t al residuo, cuando las fluctuaciones no varíen con el nivel de la serie temporal, es decir, no haya fluctuaciones elevadas, tendremos una descomposición aditiva que se expresa [8]:

$$y_t = S_t + T_t + R_t \quad (1)$$

Si la tendencia o estacionalidad por el contrario presentan grandes variaciones en el tiempo es más recomendable una descomposición multiplicativa [6]:

$$y_t = S_t * T_t * R_t \quad (2)$$

En el análisis de una serie temporal, el primer paso es elegir el tipo de descomposición que vamos a usar, con ello se podrá realizar el gráfico con las componentes de la serie como se muestra en la siguiente figura:

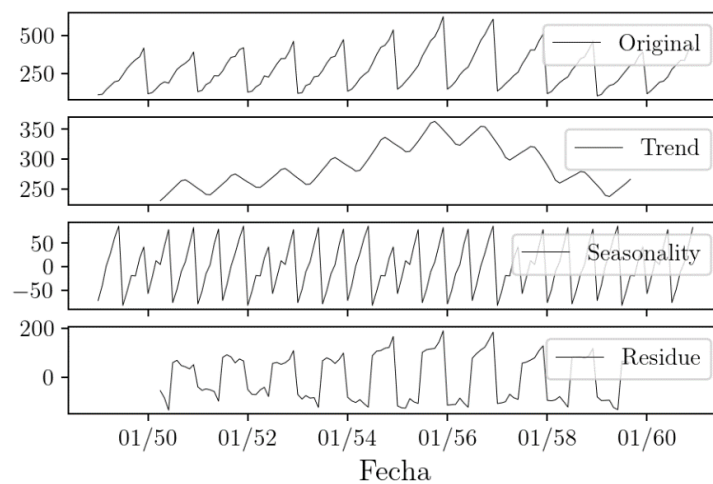


Figura 4. Descomposición de la serie del número de pasajeros mensuales (en miles) de las aerolíneas norteamericanas desde 1949 hasta 1960 (Fuente: elaboración propia. Datos: Peña, 2010)

En esta gráfica se usa una descomposición multiplicativa ya que la serie temporal presenta una estacionalidad clara y no tiene valores que sean 0.

3.3. Modelos de series temporales

Los modelos de series temporales desempeñan un papel fundamental en el análisis, predicción de datos e identificación de patrones que evolucionan a lo largo del tiempo. Estos modelos son herramientas estadísticas que permiten modelar y comprender el comportamiento y patrones de una serie temporal, así como predecir su evolución utilizando técnicas de análisis de series. Estos modelos serán utilizados una vez analizadas las características de la serie temporal, usando el que más se ajuste a estas asegurando una cierta calidad en la modelización [9]. En este contexto, existen varios enfoques ampliamente utilizados, entre ellos, el modelo de medias móviles (MA), el modelo autorregresivo (AR), el modelo autorregresivo de medias móviles (ARMA), el modelo autorregresivo integrado de medias móviles (ARIMA), el modelo autorregresivo integrado de medias móviles estacional (SARIMA) y el modelo Facebook Prophet.

Estos modelos que vamos a estudiar son para series univariantes y estocásticas, ya que una serie no estocástica es una en la cual no existe incertidumbre y no sería de interés. Un proceso estocástico es un conjunto de variables aleatorias, que se observa de manera secuencial en diferentes momentos [10]. Los modelos MA y AR son ampliamente utilizados en el análisis de series temporales debido a su simplicidad y efectividad en ciertos contextos. Sin embargo, en muchos casos, los datos de series temporales presentan tanto dependencia temporal como la necesidad de promedios móviles. Es ahí donde surge el modelo ARMA, que combina las características de los modelos MA y AR para capturar tanto la dependencia temporal como los promedios móviles. En situaciones en las que la serie temporal exhibe tendencias y/o estacionalidad, los modelos ARIMA y SARIMA se convierten en una opción poderosa. Estos modelos permiten diferenciar la serie temporal para eliminar tendencias o patrones estacionales, lo que proporciona un enfoque más robusto para el análisis y la predicción de la serie [11]. En estos modelos es fundamental tener en cuenta la forma de las gráficas de las funciones de autocorrelación simple y parcial (ACF, PACF), que facilita en algunos casos la selección tanto del modelo más adecuado como su orden, como se verá más adelante [12].

Además de estos modelos clásicos, recientemente ha surgido Prophet, un modelo desarrollado por Facebook que ha ganado popularidad debido a su capacidad para manejar múltiples fuentes de incertidumbre en las series temporales, como cambios estacionales y efectos de días festivos. Prophet utiliza un enfoque aditivo para descomponer la serie en sus componentes principales, lo que permite una modelización más flexible y precisa de los datos [13].

En este contexto, en este trabajo se explorará en detalle cada uno de estos modelos de series temporales: MA, AR, ARMA, ARIMA, SARIMA y Prophet. Se analizarán sus fundamentos teóricos y características, así como su aplicabilidad en diferentes escenarios. A través de este estudio, se busca proporcionar una comprensión sólida y práctica de estos modelos, brindando una base sólida para su implementación y selección adecuada en el análisis y pronóstico de series temporales.

3.3.1. Modelo Autorregresivo (AR)

Los modelos autorregresivos son modelos estadísticos que generalizan la idea de regresión para representar la dependencia entre dos variables aleatorias. Estos modelos son los primeros que se introdujeron y se basan en la suposición de que los valores futuros de una serie temporal están relacionados linealmente con sus valores pasados [10]. La predicción con un modelo AR se realiza multiplicando los valores pasados correspondientes por sus respectivos coeficientes y sumando los resultados. El modelo se denota como $AR(p)$, su expresión adopta la siguiente forma:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (3)$$

En los modelos autorregresivos habrá que prestar especial atención a la gráfica de la función de autocorrelación parcial (PACF), ya que sus retardos significativos indicarán el orden del modelo, es decir, los retardos significativos de esta función proporcionarán un valor aproximado del parámetro p . Los modelos AR pueden seguir un patrón en la función de autocorrelación como vemos en la Figura 5 y esto nos permitirá identificar más fácilmente el modelo más adecuado para los datos de una serie de tiempo.

Si p tiene valor 1, tendremos un proceso autorregresivo de primer orden $AR(1)$, que será estacionario si $|\phi_1| < 1$ y $\epsilon_t \sim N(0, \sigma_A^2)$, es decir, si es un proceso de ruido blanco univariante con media 0 y varianza σ_A^2 . La expresión de un modelo $AR(1)$ es:

$$Y_t = \delta + \phi_1 Y_{t-1} + \epsilon_t \quad (4)$$

Para este proceso $AR(1)$ estacionario, se comprueba:

Media	Varianza
$\mu_Y = \frac{\delta}{1 - \phi_1}$	$\sigma_Y^2 = \frac{\sigma_A^2}{1 - \phi_1^2}$
ACF	PACF
$\rho_k = \phi_1 \rho_{k-1}$	$\phi_{kk} = \phi_1$ si $k = 1$ $\phi_{kk} = 0$ si $k > 1$

Tabla 1. Características de un modelo $AR(1)$

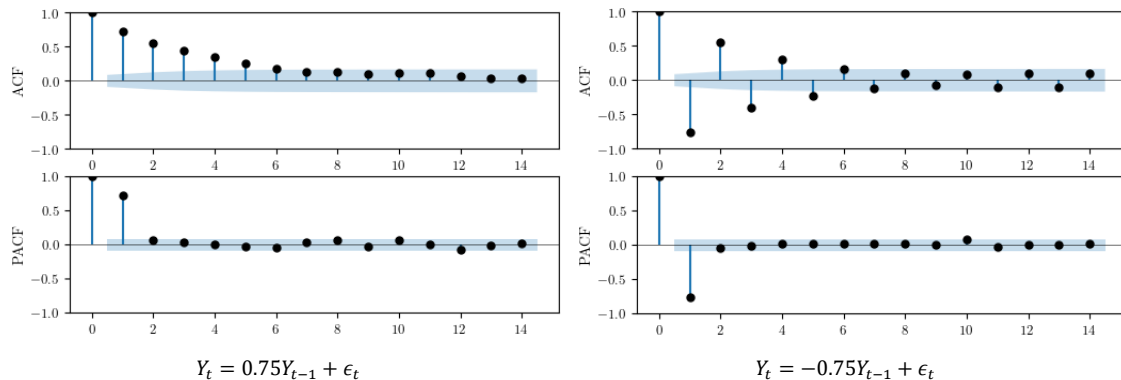


Figura 5. Autocorrelación simple y parcial de un modelo sintético $AR(1)$ (Fuente: elaboración propia)

En esta gráfica podemos apreciar que la autocorrelación parcial para los modelos AR con $p = 1$ solo tendrán un retardo significativo, y dependerá del valor que se le haya dado a ϕ_1 . Por otro lado, la correlación simple decaerá lentamente para valores de ϕ_1 positivos y decaerá de la misma manera para ϕ_1 negativos, pero alternando el signo, haciendo que podamos identificar el modelo AR siempre que las correlaciones tengan esta forma.

3.3.2. Modelo de Medias Móviles (MA)

Los procesos MA tienen la característica de tener memoria muy corta, donde el valor actual de la serie solo está correlacionado con un número pequeño de valores anteriores. Este modelo calcula su valor actual tomando un promedio de errores pasados y realizando la suma una vez que se multiplican por sus respectivos coeficientes. El orden de este modelo indica el retardo en los errores que se usan para obtener el valor actual. Establece una relación lineal entre la variable de estudio y los errores de predicciones pasadas. Este modelo se denota como $MA(q)$ y su expresión es:

$$Y_t = \delta + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} \dots + \theta_p\epsilon_{t-p} \quad (5)$$

A diferencia del modelo autorregresivo en este modelo para estimar el orden de la componente de medias móviles q , tenemos que fijarnos en la forma de la gráfica de autocorrelación simple y en sus retardos significativos. De igual manera, esta función puede seguir un patrón y así indicarnos que la serie se comporta como un modelo MA, como en la Figura 6, donde se aprecian diferentes autocorrelaciones para un modelo MA(1). La expresión de un MA(1) es:

$$Y_t = \delta + \epsilon_t - \theta\epsilon_{t-1} \quad (6)$$

Para este orden y cumpliendo la condición de invertibilidad $|\theta_1| < 1$ y $\epsilon_t \sim N(0, \sigma_A^2)$:

Media	Varianza
$\mu_Y = \delta$	$\sigma_Y^2 = (1 + \theta_1^2) \sigma_A^2$
ACF	PACF
$\rho_k = \frac{\theta_1}{1 + \theta_1^2}$; para $k = 1$	$\phi_{kk} = - \left[\frac{1 - \theta_1^2}{1 - \theta_1^{2(k+1)}} \right] \theta_1^k$
$\rho_k = 0$; para $k > 1$	

Tabla 2. Media, varianza y autocorrelaciones de un modelo MA(1)

Además, para ver la forma que tendrían las autocorrelaciones con distintos valores de p, q se muestra en la Tabla 4 un resumen de la forma general que siguen las autocorrelaciones del modelo MA para valores mayores de q [14].

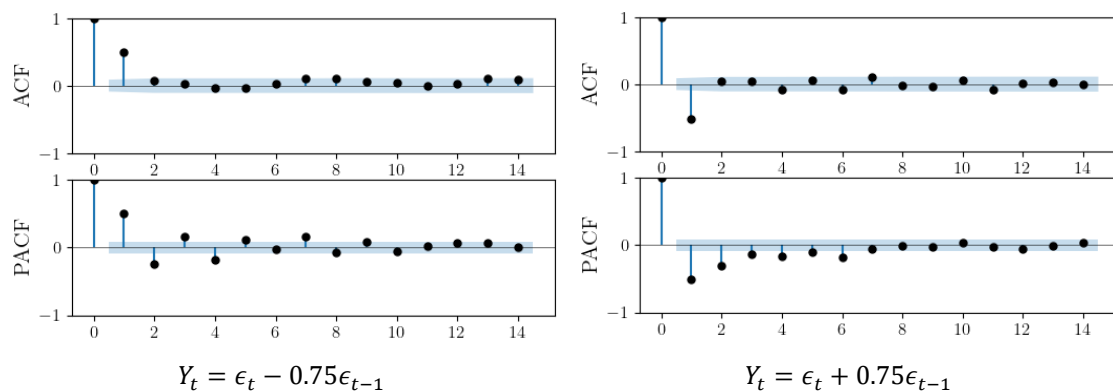


Figura 6. Autocorrelación simple y parcial modelo MA (1) (Fuente: elaboración propia)

Como podemos ver en estas gráficas, los retardos significativos de la gráfica de ACF nos indican el orden del modelo, en este caso 1, $q = 1$ y será positivo o negativo según θ . Además, para saber que es un modelo MA nos fijamos en que la PACF va decreciendo, alternando su signo o con signo negativo, siguiendo un patrón, como ya se había mencionado anteriormente.

3.3.3. Modelo ARMA

Si combinamos las propiedades de los procesos AR y MA surge un proceso llamado ARMA, donde se asume la estacionariedad de la serie temporal y da lugar a una familia muy amplia y flexible de procesos estocásticos estacionarios útiles para representar procesos, utilizando pocos parámetros, cuyos primeros q coeficientes son cualesquiera, mientras que los siguientes decrecen según leyes simples que se describen en la Tabla 4. Un proceso $ARMA(p, q)$ es estacionario si puede ser expresado en términos de su error actual y errores pasados, su expresión es [1]:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \dots + \theta_q \epsilon_{t-q} \quad (7)$$

Cuando un proceso estacionario sigue un modelo $ARMA(p, q)$ depende únicamente de los $p + q + 2$ parámetros del modelo. Si los valores de p y q son igual a 1, tendremos un modelo $ARMA(1,1)$ donde las autocorrelaciones tienen una estructura similar como apreciamos en la Figura 7, su expresión es:

$$Y_t = \delta + \phi_1 Y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} \quad (8)$$

Y con $|\phi_1| < 1, |\theta_1| < 1$ (condiciones necesarias e indispensables para la estacionariedad) y $\epsilon_t \sim N(0, \sigma_A^2)$, comprobamos:

Media	Varianza
$\mu_Y = \frac{\mu}{1 - \phi_1}$	$\sigma_Y^2 = \left[1 + \frac{(\phi_1 - \theta_1)^2}{1 - \phi_1^2} \right] \sigma_A^2$
ACF	PACF
$\rho_k = \frac{[(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)]}{(1 - 2\phi_1 \theta_1 + \theta_1^2)} \text{ para } k = 1$ $\rho_k = \rho_1 \phi_1^{k-1}$	Valor inicial $\phi_{11} = \rho_1$ seguido de: $\phi_{kk} = - \left[\frac{1 - \theta_1^2}{1 - \theta_1^{2(k+1)}} \right] \theta_1^k$

Tabla 3. Media, varianza y autocorrelación de un modelo $ARMA(1,1)$

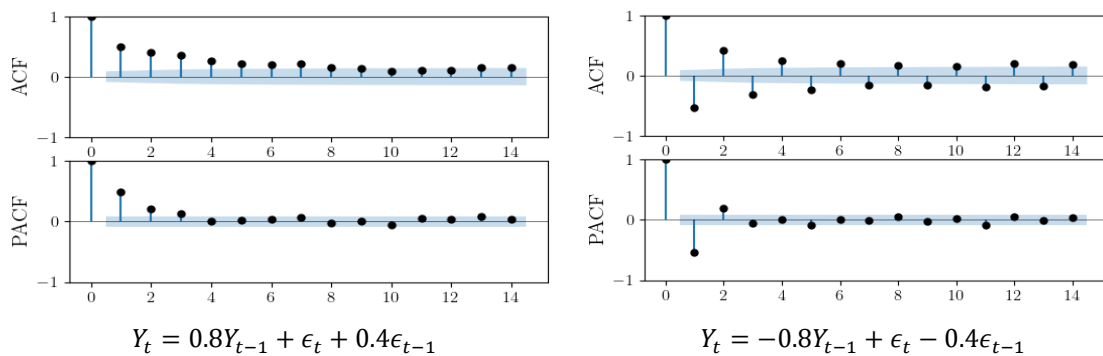


Figura 7. Autocorrelaciones simple y parcial de un modelo $ARMA(1,1)$ (Fuente: elaboración propia)

3.3.4. Modelo ARIMA

Los modelos ARIMA (Autorregresivo Integrado de Medias Móviles) son una generalización de los modelos ARMA, que se utilizan cuando la serie temporal no es estacionaria y requiere diferenciación, es decir, son una versión "integrada" de una serie estacionaria. Estos modelos asumen la estacionariedad en la serie después de aplicar la diferenciación. Un modelo ARIMA ideal se caracteriza por tener residuos que se comportan como un ruido blanco, y el modelo estimado es estacionario e invertible. Los modelos ARIMA son considerados los modelos clásicos más generales en la predicción de series de tiempo, se denotan como *ARIMA* (p, d, q) y su expresión es:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \dots + \theta_q \epsilon_{t-q} \quad ; \quad \text{con } \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (9)$$

El enfoque principal de los modelos ARIMA es transformar la serie temporal en una forma estacionaria mediante la diferenciación. Una serie es estacionaria si sus propiedades estadísticas son constantes en el tiempo. La ecuación ARIMA representa una relación lineal entre la variable dependiente y los retardos de la variable, junto con los retardos de los errores de predicción. El parámetro "d" denota el grado de diferenciación necesario para lograr la estacionariedad. La identificación de los órdenes p, d y q en los modelos ARIMA se realiza de acuerdo con los componentes autorregresivos, diferenciales y de media móvil, respectivamente. La selección adecuada de estos órdenes es fundamental para obtener un modelo ARIMA eficaz y ajustado a los datos específicos de la serie temporal [15].

3.3.5. Modelo SARIMA

La estacionalidad es una componente de las series temporales y se repite regularmente cada s periodos de tiempo. Como se ha visto anteriormente, para eliminar una tendencia se necesita diferenciar y para la estacionalidad es necesario realizar una diferenciación estacional. Surgen así los modelos ARIMA que consideran los comportamientos estacionales y mejoran la precisión de las predicciones: modelos SARIMA o ARIMA estacionales [6].

El modelo SARIMA es un modelo muy poderoso que presenta un conjunto adicional de componentes autorregresivas y de medias móviles, además, permite diferenciar los datos por la frecuencia estacional. Este modelo se denota como *SARIMA*(p, d, q)(P, D, Q, s) y comparte una parte con ARIMA, pero incluye otra parte para especificar el efecto de la estacionalidad, a esta parte la llamamos orden estacional. El parámetro P hace referencia al orden autorregresivo estacional, D representa el orden de integración estacional y Q denota el orden de la media móvil estacional. Por último, el cuarto termino "s" representa la duración del ciclo, es decir, el número de periodos necesarios para que la tendencia se repita [16].

3.3.6. Modelo Prophet

Facebook Prophet es un modelo basado en la técnica de ajuste de curvas del modelo bayesiano. Sus parámetros son fáciles de entender y no requiere una serie de tiempo con muchos datos para realizar la predicción. La técnica es más adecuada cuando tenemos una gran estacionalidad y es robusta ante datos perdidos o variaciones en la tendencia [17]. Su fórmula se escribe como:

$$Y_t = g_t + h_t + s_t + e_t \quad (10)$$

donde g_t es la tendencia, h_t la componente de vacaciones, s_t la estacionalidad y e_t el término de error. Las características de Prophet son las siguientes:

- Se trata de un modelo de autorregresión aditivo con componentes no lineales.
- Los datos de la serie de tiempo se observan cada hora, cada día y cada mes durante un año o más. Esto se representará en el gráfico de componentes del modelo, como en la Figura 8.
- Tiene en cuenta las vacaciones y días festivos.
- Tiene en cuenta tendencia, estacionalidad, valores atípicos y datos faltantes, haciendo que sea un modelo robusto.

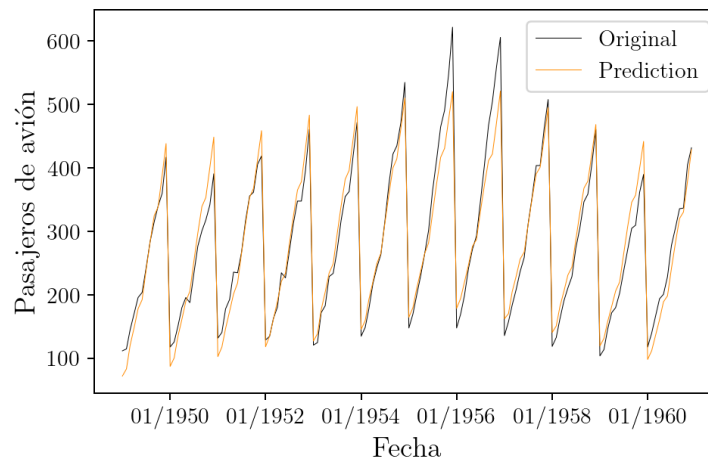


Figura 8. Predicción de serie de pasajeros en vuelos internacionales con Prophet (1949-1960)
(Fuente: elaboración propia. Datos: Peña, 2010)

Para las predicciones con este modelo existe una librería llamada *prophet*, es una librería de código abierto para la predicción de series univariantes desarrollada por Facebook disponible para usar en lenguajes como Python o R. Se desarrolla para satisfacer la necesidad de previsión desde el punto de vista empresarial [16]. Además, esta librería ofrece una gráfica donde se pueden ver las componentes del modelo como tendencia y estacionalidad anual, semanal o diaria.

3.4. Resumen del comportamiento de las funciones de autocorrelación de los modelos

La tabla 4 muestra un resumen de la forma que tienen las funciones de autocorrelación para cada modelo dependiendo del valor de sus hiperparámetros y del valor de los coeficientes ϕ y θ .

Proceso	ACF	PACF
AR(1)	Decaimiento exponencial: con signo positivo si $\phi_1 > 0$ y alternado en signo empezando negativo si $\phi_1 < 0$	Pico en el retardo 1, decrece hasta 0: positivo si $\phi_1 > 0$, negativo si $\phi_1 < 0$
AR(p)	Decaimiento rápido de tipo geométrico con alternancia de signos, sinusoidal o mezcla de varios tipos	Picos en retardos de 1 hasta p, va decreciendo/se anula para retardos superiores a p
MA(1)	Pico en el retardo 1 y decae hasta 0: con pico positivo si $\theta_1 < 0$ y negativo si $\theta_1 > 0$	Decaimiento exponencial: negativo si $\theta_1 > 0$ y alternado en signo empezando positivo si $\theta_1 < 0$
MA(q)	Picos en los retardos 1 hasta q, decrece hasta cero/ se anula para retardos superiores a q	Decaimiento exponencial o en forma sinusoidal
ARMA(p, q)	Los primeros valores sin patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas.	Los primeros valores sin patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas.

Tabla 4. Resumen de la forma de las funciones de autocorrelación para los modelos MA, AR y ARMA

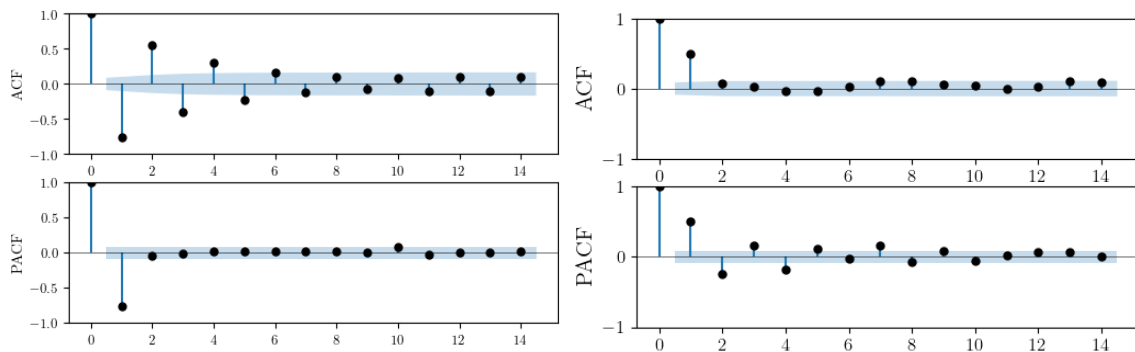


Figura 9. Ejemplo de funciones de autocorrelación para un modelo AR(1) (izquierda) y de un modelo MA(1) (derecha) (Fuente: elaboración propia)

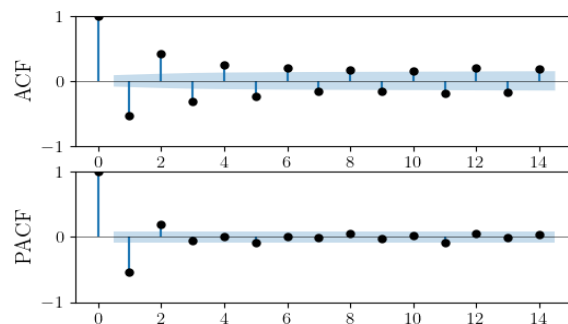


Figura 10. Ejemplo de las funciones de autocorrelación para un modelo ARMA(1,1) (Fuente: elaboración propia)

4. Metodología

La metodología utilizada en este TFE representa una parte fundamental para alcanzar el objetivo de desarrollar un proceso de estandarización para la selección del modelo predictivo adecuado a una serie temporal en base a sus características. A continuación, se introducen las fases de la metodología seguida, desde la visualización y preprocesamiento de los datos hasta la evaluación de la bondad del modelo.

- Fase 0: En primer lugar, se realizará una exploración inicial de los datos de la serie temporal, centrándose en su visualización y comprensión. La visualización permite detectar patrones visuales, tendencias, estacionalidad y posibles anomalías en los datos. Además, se llevará a cabo un preprocesamiento de los datos, que puede incluir la limpieza de valores atípicos, la interpolación de datos faltantes y la normalización de la serie.
- Fase 1: A continuación, se procederá a descomponer la serie temporal en sus componentes: tendencia, estacionalidad y residuo. Esta descomposición permite comprender mejor la estructura de la serie y cómo se relacionan sus componentes. Posteriormente, se analizará la media móvil y la desviación estándar de la serie para obtener información sobre su distribución y variabilidad. Esto proporciona una base para comprender el comportamiento general de los datos. A continuación, se llevará a cabo un test de estacionariedad utilizando la prueba de Dickey-Fuller aumentada. Esta prueba estadística nos ayudará a determinar si la serie temporal es estacionaria de igual forma que para identificar la estacionalidad se analiza la función de autocorrelación.
- Fase 2: Una vez identificadas las características de la serie se le aplicarán las transformaciones adecuadas si las requiere como diferenciación o desestacionalización. Estas transformaciones ayudarán a estabilizar la serie y mejorar la calidad y capacidad de pronóstico del modelo.
- Fase 3: Una vez que se ha realizado una exploración exhaustiva de la serie temporal y se han identificado las características relevantes, se procederá a estimar el modelo adecuado. Esto se realizará mediante el análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), que proporcionan información valiosa sobre la estructura de los datos. Estos gráficos ayudarán a seleccionar el orden de los componentes autorregresivos y de medias móviles. La selección final del modelo se realizará utilizando los criterios de información de Akaike (AIC) y bayesiano (BIC). Estos criterios permiten evaluar y comparar la calidad relativa de diferentes modelos, considerando tanto su capacidad predictiva como su complejidad.
- Fase 4: Una vez seleccionado y ajustado el modelo final, se procederá a la etapa de predicción y evaluación, donde se realizará la predicción sobre el 20% de los datos de la serie y se evaluará la bondad del modelo mediante métricas de error, como el error cuadrático medio (RMSE), que proporciona una medida cuantitativa de la precisión de los pronósticos realizados, comprobando así que el modelo seleccionado es adecuado para modelizar el comportamiento de la serie temporal, verificando el proceso estandarizado de selección del modelo.

A través de diferentes pruebas, se pretende ajustar el procedimiento de estandarización hasta establecer cuál es el modelo adecuado para cada serie. Las pruebas se aplicarán a datos de diferentes tipos para así considerar el mayor tipo de series posibles. En este caso serán series de datos sintéticos, series reales de datos como consumo de gasolina o población e incluso datos reales provenientes del CTN, como es la temperatura, o clorofila del Mar Menor. Todas las pruebas y análisis realizados en el presente trabajo se llevaron a cabo utilizando el framework de Python.

4.1. Fase 0: Visualización y limpieza de datos

4.1.1. Visualización de datos

La visualización de datos desempeña un papel fundamental al proporcionar una comprensión inicial y una perspectiva visual de las características y patrones presentes en la serie, permitiendo comprender su comportamiento y realizar un análisis preciso. En esta etapa, se llevará a cabo una exploración detallada de la serie de tiempo, con el objetivo de identificar tendencias, posibles datos faltantes y otros aspectos relevantes.

En primer lugar, se procederá a la visualización de la serie de tiempo en su totalidad. Esto permitirá tener una visión de cómo se desarrollan los datos a lo largo del tiempo y si se observan patrones o tendencias evidentes. Además, se buscarán posibles datos faltantes. Estos datos faltantes pueden tener un impacto significativo, por lo que es crucial identificarlos y abordarlos correctamente como se describirá más adelante. La visualización de datos también nos brinda información valiosa sobre la presencia de estacionalidad en la serie temporal mediante la descomposición estacional donde es posible detectar patrones repetitivos en la serie. Esto es fundamental para determinar si se deben aplicar técnicas de transformación. Además, se visualizarán la media y la varianza de la serie para evaluar la consistencia y la estabilidad de la serie a lo largo del tiempo. Otra herramienta importante en la visualización de datos de series temporales son las gráficas de las funciones de autocorrelación. Estas gráficas muestran las correlaciones entre los valores de la serie en diferentes retardos temporales. La visualización de las autocorrelaciones puede revelar patrones de dependencia temporal, ayudando así a identificar el orden de los componentes autorregresivos y de promedios móviles en los modelos posteriores.

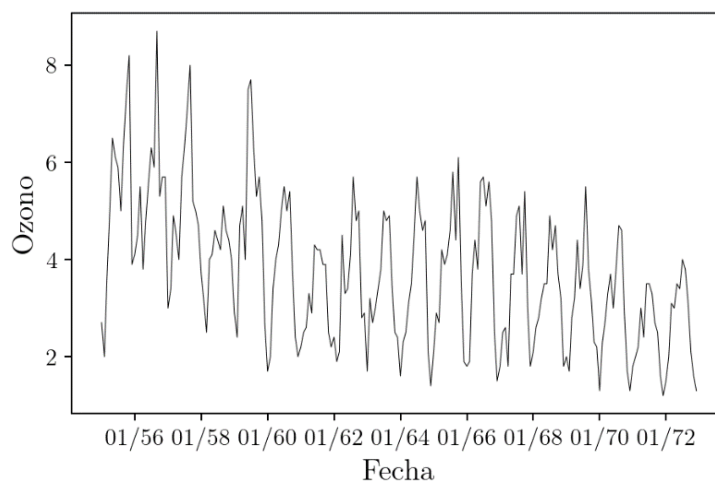


Figura 11. Serie de tiempo de Ozono de 1955 a 1972. (Fuente: elaboración propia. Datos: Peña, 2010)

4.1.2. Limpieza de datos

La limpieza de datos es una técnica de preprocesado que nos ayuda a mejorar la calidad de los datos ya que eliminamos datos faltantes o atípicos, asegurando así la precisión del análisis posterior de la serie temporal. En el caso de tener **valores faltantes** hay varias estrategias que se pueden adoptar, por ello es importante evaluar cada caso específico y considerar el impacto que tendrá cada estrategia en los resultados del análisis. Además, es importante documentar y justificar cualquier método de los siguientes utilizado para tratar los datos faltantes.

- Interpolación: estimar los valores faltantes en función de los valores cercanos.
- Media móvil: reemplazar los valores faltantes por la media de los valores adyacentes.
- Forward/Backward fill: completar los valores faltantes con el valor anterior o posterior.
- Media condicional: se calcula la media en unas condiciones (por ejemplo, día de la semana) y se utiliza esta media para completar los valores faltantes en esas mismas condiciones.
- Eliminación de datos: si los datos faltantes son pocos en comparación con el tamaño total de la serie, se pueden eliminar esas observaciones siempre y cuando no suponga una pérdida de información.

Por otro lado, en el caso de **datos atípicos** u outliers, es importante detectarlos y tratarlos de forma correcta para que no se distorsionen los resultados. Una vez detectados se pueden eliminar, corregir o dejarlos como en la serie original, pues se puede tratar de datos correctos y, en ocasiones, los más interesantes [18]. Por todo ello, es importante investigar la naturaleza del valor atípico antes de decidir su tratamiento. Los outliers se eliminan cuando

- Son erróneos.
- No modifican los resultados, pero afectan a las suposiciones.
- Crean una asociación fuerte que no existiría en caso de no existir el valor

Por último, los valores atípicos no serán tratados y por tanto se analizarán junto con la serie cuando

- No se trate de un error y no afecten a las suposiciones.
- Aporten información y modifiquen nuestro análisis posterior [19].

Una vez realizada la limpieza de datos vamos a pasar a clasificar la serie de tiempo para ver qué modelo es el más adecuado según las características de dicha serie.

4.2. Fase 1: Identificación de patrones

4.2.1. Descomposición de la serie

La descomposición de la serie es el primer paso para visualizar cómo se comporta una serie, ya que podemos detectar claramente si esta serie tiene tendencia y/o estacionalidad, así confirmamos las estimaciones realizadas en la Fase 0 de visualización de datos. La presencia de estas componentes se va a corroborar en esta fase mediante otras técnicas de clasificación que se complementan con la descomposición de la serie.

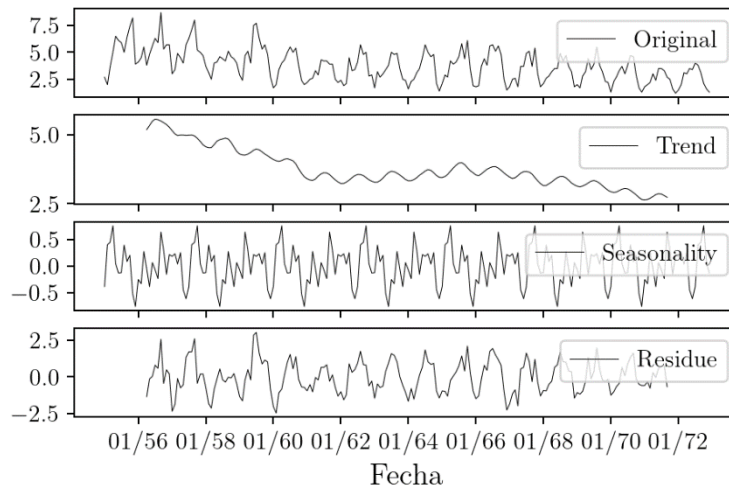


Figura 12. Descomposición de la serie de tiempo de Ozono (Fuente: elaboración propia. Datos: Peña, 2010)

4.2.2. Media móvil y desviación típica

El estudio de la media móvil y la desviación típica en una serie de tiempo es una técnica ampliamente utilizada para analizar y suavizar los datos a lo largo del tiempo, lo que ayuda a identificar patrones y tendencias de los datos ya que elimina el ruido y las fluctuaciones aleatorias de los datos originales. La media móvil se aplica a una serie de tiempo para obtener un valor promedio en un período determinado, se calcula tomando un conjunto de valores consecutivos y realizando su media.

La desviación típica también se puede calcular en función de la media móvil y se conoce como desviación típica móvil y mide la dispersión o volatilidad de los valores alrededor de la media móvil. La desviación típica móvil puede ayudar a identificar cambios significativos en la variabilidad de los datos.

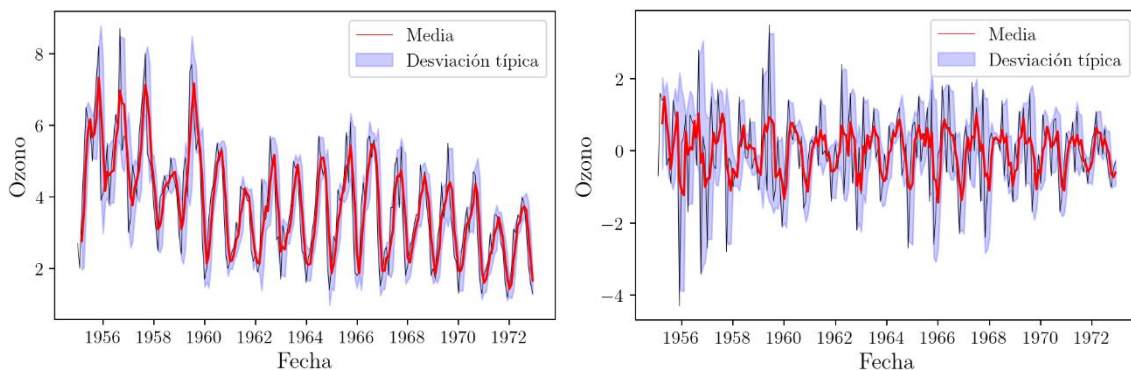


Figura 13. A la izquierda la media y desviación de una serie no estacionaria y a la derecha media y desviación de la misma serie transformada (Fuente: elaboración propia. Datos: Peña, 2010)

4.2.3. Prueba de Dickey - Fuller

La prueba de Dickey – Fuller permite detectar si hay presencia significativa de tendencia en una serie y descartar así su estacionariedad. Determina la existencia o no de raíces unitarias en una serie de tiempo. Se considera como hipótesis nula que la serie temporal no es estacionaria y como hipótesis alternativa que es estacionaria [16]. La prueba de Dickey – Fuller analiza la estacionariedad de un AR(1) pero puede generalizarse para otros modelos, en cuyo caso recibe el nombre de prueba de Dickey – Fuller aumentada, esta es una de las herramientas más fiables para determinar si una serie es estacionaria.

4.2.4. Función de autocorrelación (ACF)

La función de autocorrelación (ACF) permite identificar cómo están de correlacionados entre sí los valores en una serie de tiempo. El coeficiente de correlación varía desde -1 a 1, donde un valor de 0 indica que no existe relación entre las variables. Se puede usar para detectar patrones sistemáticos en conjuntos de datos. Además, la representación gráfica de la ACF permite determinar si hay una estacionalidad en el conjunto de los datos mediante el análisis del retardo de esta función en el valor del periodo de la serie. En caso de ser significativo tendremos una serie con estacionalidad. Consideramos un valor significativo cuando el coeficiente de correlación es mayor a 0.5. Por otro lado, esta gráfica de la ACF, como ya se ha comentado anteriormente, también es útil para para estimar el valor de q , es decir, el orden de la componente de Media Móvil (MA) como en la Figura 14, donde en la gráfica superior de autocorrelación observamos que tenemos un retardo significativo, por lo que tendremos unos datos que se comportan como un MA(1).

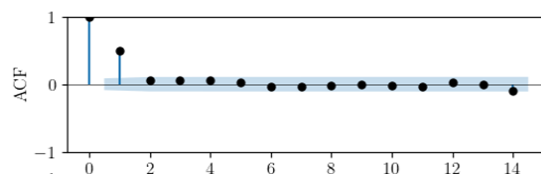


Figura 14. Autocorrelación simple para una serie estacionaria MA(1) Fuente: elaboración propia)

4.2.5. Función de autocorrelación parcial (PACF)

La función de autocorrelación parcial muestra la correlación entre dos variables teniendo en cuenta el efecto de otras variables. Puede ser útil para estimar el orden de la componente autorregresiva de un modelo AR, es decir, el parámetro p [20] con el valor del último desfase o coeficiente de correlación significativo, como se ha comentado en la sección 3.3.1. La estimación de los parámetros p y q con la autocorrelación tanto simple como parcial nos ayuda a obtener cuál es el modelo predictivo más apropiado para la serie, modelo que se comprobará más adelante mediante el criterio de información adecuado.

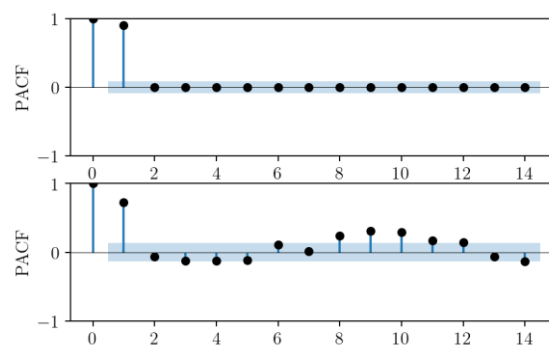


Figura 15. PACF de un modelo AR(2) y de serie de Ozono (Fuente: elaboración propia)

4.3. Fase 2: Intento de mejora de ajuste: transformación de datos

4.3.1. Diferenciación

La diferenciación es una técnica de transformación de datos que será necesaria en el caso de tener una serie temporal no estacionaria, se utiliza para eliminar o reducir la tendencia y así obtener residuos estacionarios. Esta técnica funciona calculando la diferencia entre los valores consecutivos de una serie temporal con el objetivo de estabilizar la media de la serie temporal y que parezca estacionaria, aunque en ocasiones la serie diferenciada no muestre estacionariedad y sea necesario hacer una diferenciación de segundo orden. Esto nos ayudará a definir mejor cuál es el modelo más adecuado para nuestra serie temporal según sus características. La diferenciación se refleja en los modelos con el parámetro d . La diferenciación se denota como (∇Y_t) y su expresión es:

$$\nabla Y_t = Y_t - Y_{t-1} \quad (11)$$

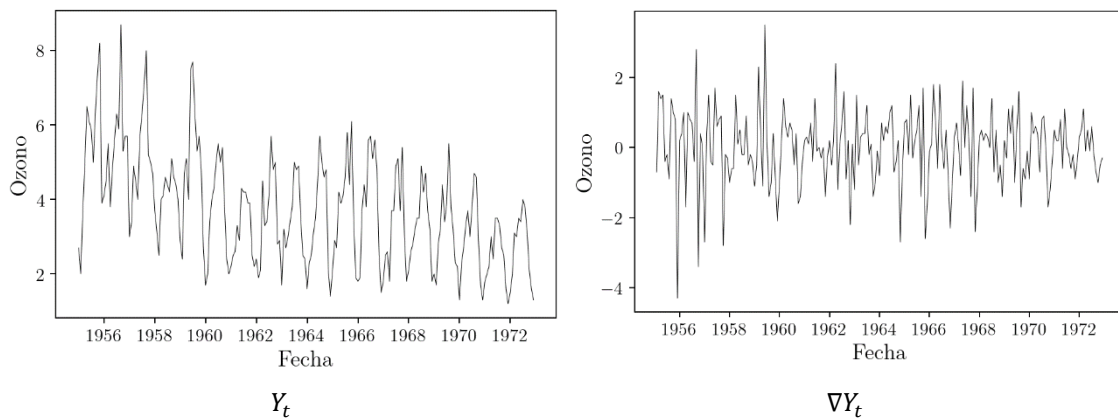


Figura 16. Serie de Ozono original (izquierda) y serie diferenciada (derecha) (Fuente: elaboración propia. Datos: Peña, 2010)

Un ejemplo de cómo varía una serie al diferenciarla se puede ver en la figura anterior en la que originalmente se tiene una serie con una tendencia decreciente la cual queda estable y con valores en torno a cero una vez diferenciada como se observa a la derecha de la Figura 16.

4.3.2. Desestacionalización

La desestacionalización de una serie de tiempo es un proceso similar a la diferenciación, pero en este caso se busca eliminar o mitigar el componente estacional presente en los datos y es importante en el análisis de series temporales, ya que permite separar el efecto de la estacionalidad de otros componentes, facilitando la comprensión e interpretación de los datos, así como la precisión de los modelos utilizados para el pronóstico. Existen diferentes técnicas para desestacionalizar una serie. Un enfoque popular es la descomposición de la serie en sus componentes principales y, una vez descompuesta, se toma el residuo de la serie como serie sin estacionalidad, pudiendo así estimar el modelo adecuado. La desestacionalización transforma el proceso Y_t de la forma:

$$\nabla_s Y_t = Y_t - Y_{t-s} \quad (12)$$

Se debe tener en cuenta que esto lo realizamos para estimar el modelo que vamos a usar y sus parámetros, para la predicción siempre usaremos la serie original sin transformaciones para que los datos estén en la escala original.

4.4. Fase 3: Selección del modelo y explicación de patrones

La selección del modelo más adecuado para cada serie de tiempo atendiendo a sus características se va a llevar a cabo aplicando dos criterios de información como son el Akaike (AIC) y el bayesiano de Schwarz (BIC). Estos dos criterios de información comparan modelos, seleccionando el modelo con valor menor de AIC y BIC, con un buen ajuste y con un número de parámetros reducido como modelo más adecuado.

- Criterio de información de Akaike (AIC)

Evalúa cómo de bien se ajusta el modelo a los datos. Este criterio mide la bondad del ajuste y penaliza que el modelo considere una gran cantidad de datos históricos sin ser necesarios, así, cuanto menor sea este valor mejor balance entre ajuste y complejidad tiene.

- Criterio de información bayesiano de Schwarz (BIC)

Este criterio está relacionado con el AIC, pero con la diferencia de que interviene el número de observaciones que posee la serie. De igual manera, cuanto menor sea indicará que la selección del orden es más apropiada. Este criterio es más adecuado para muestras grandes, a diferencia del AIC, que se usa para muestras pequeñas.

Criterio de información de Akaike	Criterio de información bayesiano
$AIC \equiv -\frac{2L}{n} + \frac{2m}{n}$	$BIC \equiv -\frac{2L}{n} + m \frac{\log(n)}{n}$

Tabla 5. Fórmulas de los criterios de selección del modelo predictivo

En la tabla 5 presentamos las expresiones de ambos criterios de información, donde m es el número de parámetros que se estiman en el modelo y n el número de observaciones.

Modelo	AIC	BIC
ARIMA(3, 1, 0)	691.597	705.061
ARIMA(1, 1, 0)	729.350	739.448
ARIMA(0, 1, 0)	784.41	788.180

Tabla 6. Valor de los criterios de información para modelos estimados para los datos de Ozono

Por otro lado, en esta otra tabla se presentan los valores de los criterios AIC y BIC para tres modelos ARIMA de diferentes parámetros ajustados a los datos de Ozono.

4.5. Fase 4: Aplicación del modelo predictivo

4.5.1. División del conjunto de datos en prueba y entrenamiento

Una vez elegido el modelo, se debe de entrenar el mismo. En este caso, el entrenamiento no se hará con el conjunto de datos completo, ya que, aunque un modelo actúe bien sobre un conjunto de datos no asegura que las predicciones posteriores vayan a ser precisas. La calidad de una predicción solo puede determinarse midiendo cómo un modelo actúa sobre datos que no se usan para entrenar e implementar el modelo. Por ello, en el análisis de una serie de tiempo se divide el conjunto de datos en dos partes, el conjunto de entrenamiento (*training set*) que se usa para entrenar y ajustar el modelo, que suele ser del 80%; y otra el porcentaje o conjunto de prueba (*test set*), sobre el que se realizan las predicciones y se miden los errores. El conjunto de prueba suele ser del 20%, aunque variara según las características de la serie temporal [6].

Cabe destacar que, con la cantidad de datos de entrenamiento inicial se predice una cantidad de valores siguientes que viene dado por el orden de MA, y en cada instante temporal se actualiza la secuencia de entrenamiento con el valor de test que ya se ha predicho, retroalimentando así el modelo con datos originales que se usarán para las predicciones posteriores. En ningún caso introduce el valor de test antes de ser predicho. Se realiza de esta manera ya que se han realizado diferentes pruebas y debido a la variabilidad de los datos y la disponibilidad casi en tiempo real de los datos es el método que arroja mejores resultados.

4.5.2. Predicción

Para validar que se ha elegido el modelo adecuado, se va a realizar la predicción sobre el conjunto de datos test, es decir, el 20% del conjunto de datos total. Esta predicción se va a realizar actualizando el conjunto de entrenamiento en cada iteración como se ha indicado anteriormente.

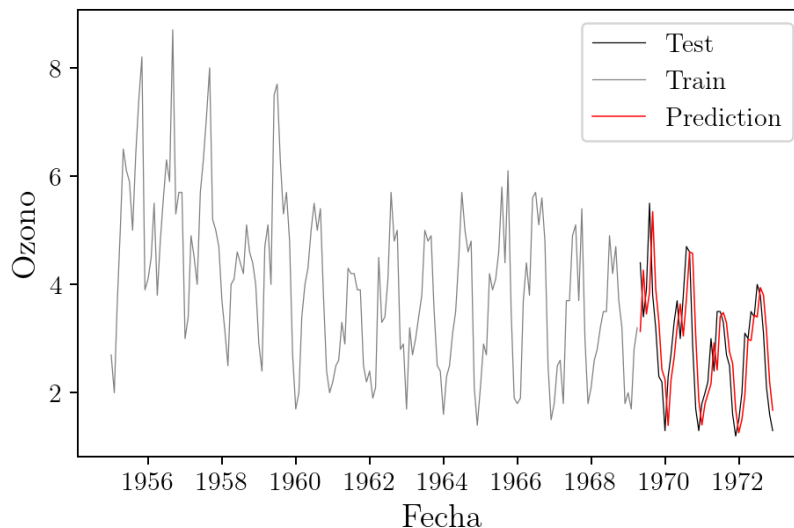


Figura 17. Predicción con un modelo ARIMA de la serie de datos de Ozono (Fuente: elaboración propia. Datos: Peña, 2010)

En esta gráfica podemos ver la predicción del conjunto de datos de Ozono, una predicción que a simple vista parece buena, esto nos confirma que la selección del modelo predictivo es la adecuada. Se evaluará lo bien que se ajusta el modelo a estos datos con más precisión gracias a las métricas de error que se presentarán en la siguiente sección.

4.5.3. Evaluación de la bondad del modelo predictivo

Posteriormente a la predicción de una serie, tenemos que medir cómo de bien se ajusta el modelo con el que realizamos dicha predicción a los datos. Para ello, usaremos una serie de métricas que definiremos a continuación. Estas métricas no estudian la calidad de la predicción, solo el ajuste del modelo a los datos de la serie, cuanto menor sea el valor de estas métricas mejor ajuste tendrá el modelo. Estas métricas son las siguientes:

- Error Cuadrático Medio (MSE¹) y Raíz del Error Cuadrático Medio (RMSE²)

Se calcula restando los valores reales menos los estimados, se elevan al cuadrado y así obtenemos el MSE. Si a esto le hacemos la raíz cuadrada tenemos el RMSE que comparte las ventajas del MSE, pero es más representativo y fácil de interpretar directamente al mostrarnos el error en las mismas unidades que la variable predicha [21]. RMSE es la métrica más simple y común para la evaluación de regresión. Es útil si tenemos valores inesperados que nos deberían interesar. Consigue penalizar con más severidad errores grandes en comparación a errores más pequeños. La desventaja es que es especialmente problemático si tenemos datos muy ruidosos [16].

- Error Absoluto Medio (MAE³)

Se calcula haciendo el promedio de diferencias absolutas entre valores observados y predicciones. Es una puntuación lineal, todas las diferencias individuales se ponderan por igual. Nos da también el error en las mismas unidades que la variable predicha. No es tan sensible como el MSE a los valores atípicos. En comparación a MSE, calcula el error absoluto y no penaliza con la misma severidad errores grandes, es menos restrictivo. Es la solución más sencilla para la inexactitud del error medio [16].

- Error Porcentual Absoluto Medio (MAPE⁴)

Es la media de la diferencia absoluta entre los valores predichos y los verdaderos, dividido por el valor verdadero, como se muestra en la Tabla 7. Se suele usar ya que es fácil de interpretar y explicar. Por ejemplo, un MAPE de 11,5% significa que la diferencia entre el valor real y previsto es en promedio 0,115 veces el valor real. Cuanto menor sea el MAPE mejor podrá pronosticar valores un modelo. Un modelo de MAPE 2% tiene mayor precisión que uno con el MAPE al 10%. Funciona mejor si no hay extremos en los datos y no hay ceros, ya que no es válido dividir entre cero.

MSE	RMSE	MAE	MAPE
$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$	$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$	$MAE = \frac{1}{n} \sum_{t=1}^n y_t - \hat{y}_t $	$MAPE = \frac{1}{n} \sum_{t=1}^n 100 \left \frac{y_t - \hat{y}_t}{y_t} \right $

Tabla 7. Expresiones matemáticas de las métricas de evaluación

RMSE	MAE	MAPE
0.781	0.643	0.251

Tabla 8. Métricas de bondad del modelo de Ozono

¹ Mean square error.

² Root mean square error.

³ Mean absolute error.

⁴ Mean absolute percentage error

4.6. Software utilizado

El framework utilizado en este TFE es el lenguaje de programación Python. Python es un lenguaje de programación de alto nivel, conocido por su simplicidad, legibilidad y su amplia variedad de bibliotecas y herramientas. En particular, se han utilizado diversas bibliotecas y paquetes de Python especializados en la manipulación de datos y en el análisis de series temporales, por ejemplo, *pandas* y *numpy* para el tratamiento y procesado de datos; *matplotlib* para las representaciones gráficas; *statsmodels* para modelos estadísticos y descomposición estacional de la serie [22]; y, *sklearn* [23] o *Prophet* [24] para el análisis predictivo. Estas bibliotecas proporcionan funcionalidades avanzadas para la manipulación, visualización, preprocesamiento, modelado y evaluación de series temporales. La elección de Python como framework para este TFE se debe a varias razones. En primer lugar, Python es ampliamente utilizado en el ámbito de la ciencia y el análisis de datos. Además, Python cuenta con una sintaxis clara y sencilla, lo que facilita el desarrollo y comprensión del código. Otra ventaja de utilizar Python es la disponibilidad de numerosas bibliotecas especializadas en el análisis gráfico de series temporales. Estas bibliotecas ofrecen implementaciones eficientes de algoritmos y modelos, así como herramientas para la visualización y evaluación de resultados.

5. Aplicación a datos sintéticos

Los datos sintéticos se crean a medida mediante programación para satisfacer una serie de características, simular a los datos reales y ayudar a crear y entrenar modelos. Por ello se han creado una serie de conjuntos sintéticos con las características de modelos determinados que son interesantes para estudiar diferentes características en las series e ilustrar cómo se aplica la metodología anterior a estos modelos.

5.1. AR

Para crear un conjunto AR teórico tenemos que seleccionar un valor del parámetro ϕ , este puede variar entre -1 y 1. Estos datos se han generado mediante la función ArmaProcess en Python, en la que introducimos el orden del parámetro AR y el valor de ϕ y genera un conjunto de datos con las características de una serie de tiempo AR. La amplitud del primer retardo de la gráfica de la función de autocorrelación parcial dependerá de este valor. En este caso el parámetro tendrá un valor de 0.75. Según la metodología del Capítulo 4, se realiza un preprocesado de datos, eliminamos los valores atípicos si fuese necesario y realizamos su descomposición estacional. Esto se ve en la siguiente figura:

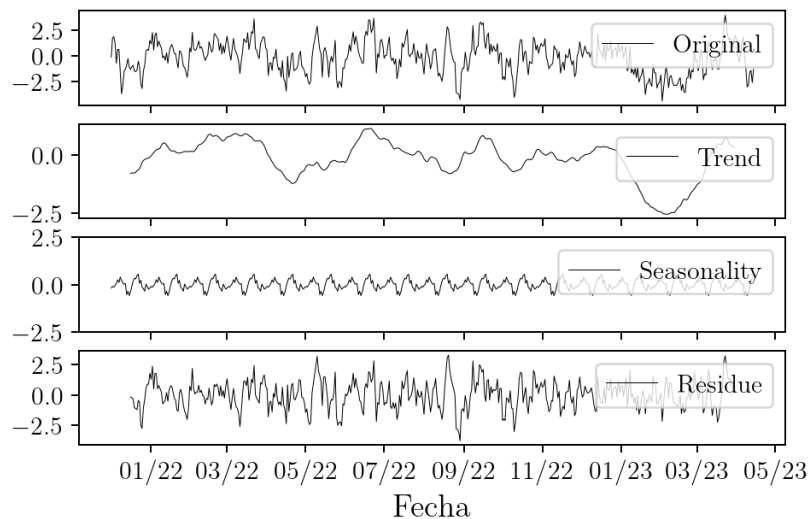


Figura 18. Representación gráfica del modelo AR(1) (Fuente: elaboración propia)

En esta gráfica está representada la serie original procesada, la tendencia, la estacionalidad y el residuo, se trata de una serie que, como se sabe por ser un conjunto autorregresivo, no parece presentar tendencia y que cuenta con una componente de estacionalidad poco elevada, y, aunque teóricamente no debe presentar estacionalidad, se va a comprobar la presencia de estacionalidad con la técnica de clasificación del análisis de correlación. Se va a clasificar este conjunto de datos con las técnicas que se han visto anteriormente como la prueba de Dickey Fuller y la ya mencionada del estudio de la autocorrelación.

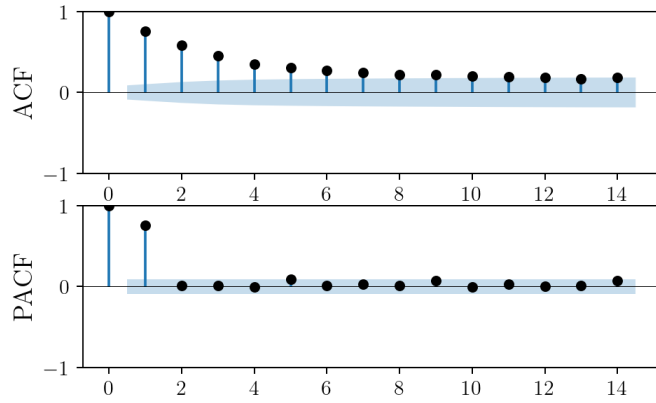


Figura 19. Autocorrelación simple y parcial AR(1) (Fuente: elaboración propia)

Como se ve en la anterior figura se trata de una serie sin estacionalidad, ya que el retardo en el valor del periodo de la autocorrelación (12) no es significativo, y, además, se trata de una serie sin tendencia al ser el p -valor = $4,01e^{-17}$. Esto nos indica que se puede obviar la parte de transformaciones y podemos aplicar directamente el modelo predictivo que se ha identificado como adecuado. Con el gráfico de autocorrelación parcial podemos estimar el parámetro autorregresivo p para esta serie. Como ya se sabía, la serie de datos sigue un patrón AR(1) ya que dicha autocorrelación tiene un único retardo significativo, los criterios de información AIC y BIC confirman esta suposición. Para aplicar este modelo se define el conjunto de entrenamiento que incluye los datos de la serie hasta una fecha, con el objetivo de predecir a partir de esa fecha. Aplicamos entonces el modelo predictivo AR(1), a la serie de tiempo, obteniendo la predicción sobre el 20% de los datos como se observa en la Figura 20.

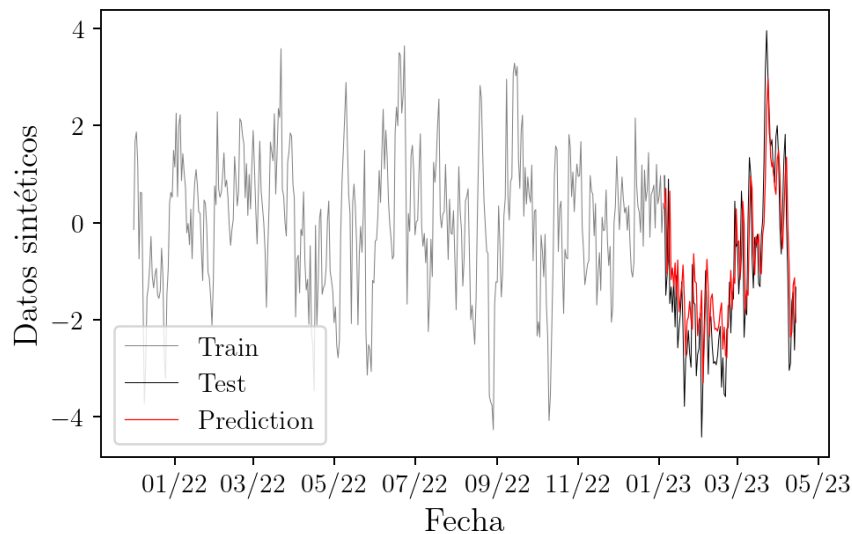


Figura 20. Predicción modelo teórico AR(1) (Fuente: elaboración propia)

Como estos conjuntos de datos son sintéticos y aleatorios, y cada vez se genera un conjunto diferente, aunque con la misma distribución de los datos, vemos la tabla de errores para esta realización en particular (valor individual de la Tabla 9) junto con los errores medios de haber generado diez realizaciones diferentes AR(1), pudiendo así observar también la desviación típica.

RMSE	\overline{RMSE}	MAE	\overline{MAE}	MAPE	\overline{MAPE}
1,055	$0,997 \pm 0,057$	0,830	$0,797 \pm 0,055$	1,454	$2,032 \pm 0,990$

Tabla 9. Errores predicción modelo AR(1), para una realización y para la media de 10 realizaciones

5.2. MA

Creamos un modelo sintético MA(2), que no tendrá tampoco tendencia ni estacionalidad por lo que no es necesario realizar una transformación de datos. En este caso como se ve en la Ecuación (5) la variable a la cual tenemos que asignar un valor es θ . Para este caso le damos los valores $\theta_1 = 0.75$ y $\theta_2 = 0.4$. Tras el preprocesado y la descomposición clasificamos la serie como estacionaria, sin tendencia ni estacionalidad, según el test de Dickey Fuller y el estudio de la autocorrelación simple. Apoyándonos en la representación de la autocorrelación de la Figura 21 podemos estimar cual será el modelo predictivo utilizado.

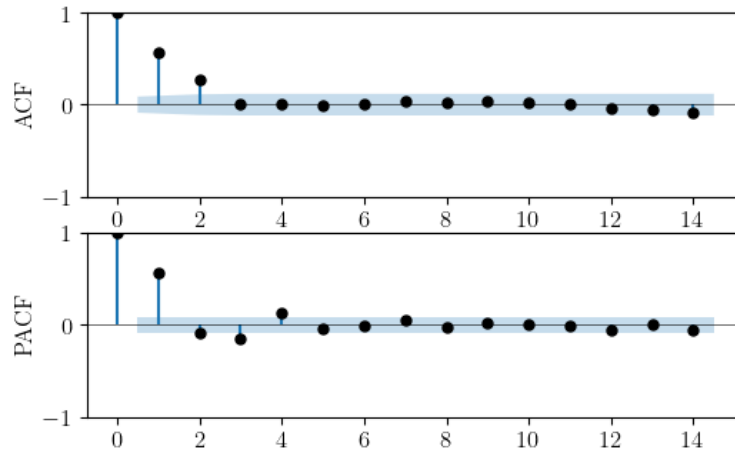


Figura 21. Gráfica autocorrelaciones modelo MA(2) (Fuente: elaboración propia)

Podemos observar que, en este caso de una serie MA, la gráfica que contiene información sobre el modelo predictivo a utilizar es la de autocorrelación simple, que cuenta con dos retardos significativos que representan el orden del modelo de medias móviles, por lo que el modelo predictivo utilizado será el MA(2), obteniendo la siguiente predicción:

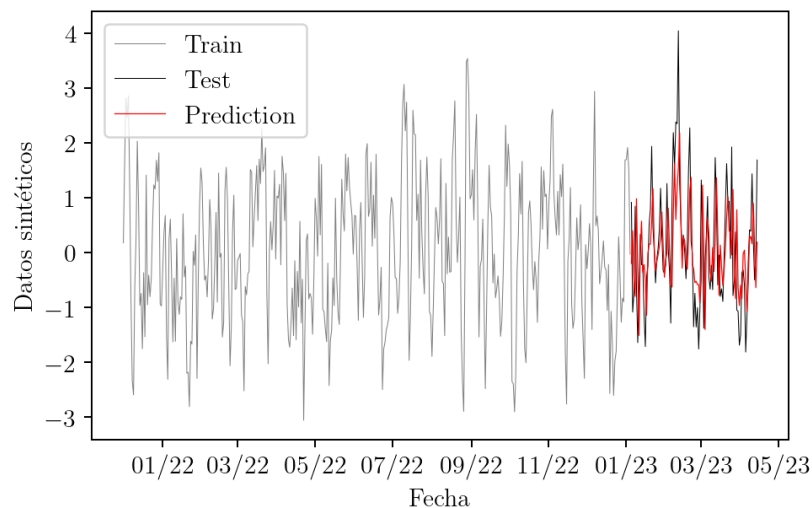


Figura 22. Predicción modelo MA(2) teórico (Fuente: elaboración propia)

RMSE	$\overline{\text{RMSE}}$	MAE	$\overline{\text{MAE}}$	MAPE	$\overline{\text{MAPE}}$
0,928	$1,014 \pm 0,059$	0,749	$0,815 \pm 0,045$	3,322	$2,581 \pm 1,499$

Tabla 10. Errores de predicción del modelo MA(2) junto con la media de errores y desviación típica de 10 realizaciones

Como en el caso anterior, en esta tabla se pueden observar los errores del conjunto visto en este capítulo en particular junto con los de diez series generadas de la misma forma.

5.3. ARMA

Generamos ahora una realización del modelo ARMA con $\phi=0.8$ y $\theta=-0.4$. Una serie ARMA, también es una serie estacionaria sin estacionalidad ni tendencia y su autocorrelación tendrá los primeros valores sin patrón fijo seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas como se ha visto en la Tabla 4 y como se puede ver en la siguiente gráfica.

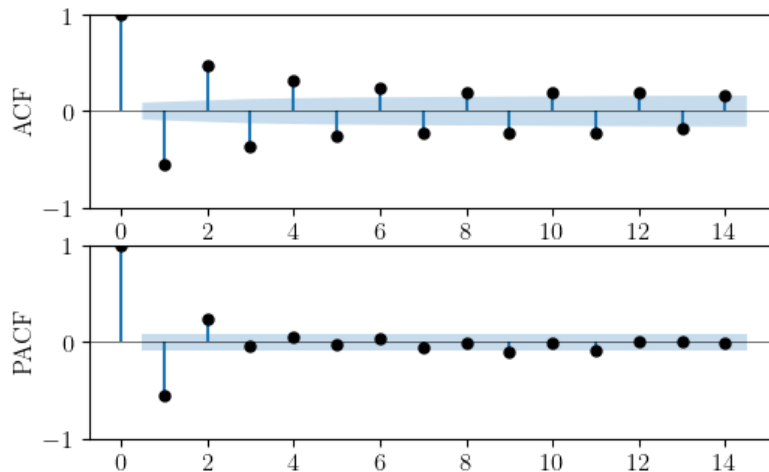


Figura 23. Gráfica de autocorrelación de un modelo ARMA(1,1) (Fuente: elaboración propia)

Después del preprocesado y clasificación de la serie, gracias a los criterios de información conocidos (AIC, BIC) se selecciona el modelo predictivo correspondiente, es decir, el modelo ARMA(1,1), con este modelo entrenamos para obtener la siguiente predicción:

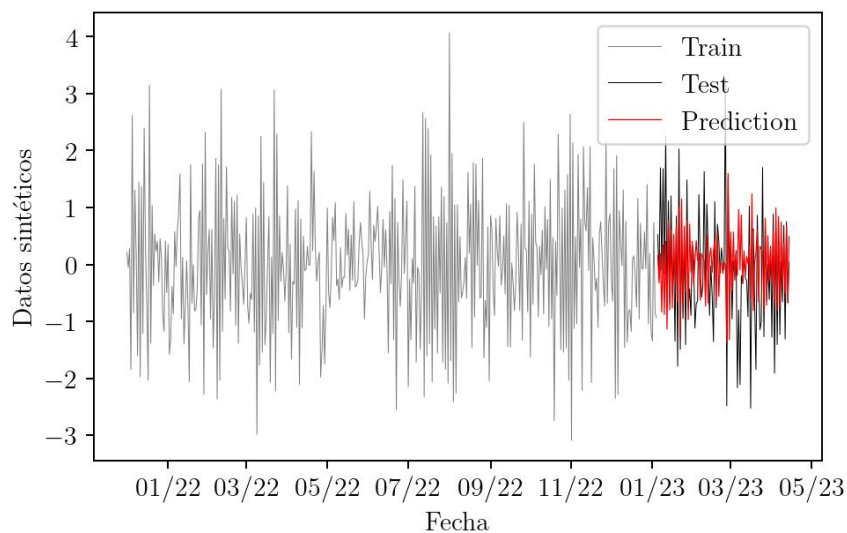


Figura 24. Predicción serie temporal ARMA(1,1) (Fuente: elaboración propia)

Como se observa en esta serie de pruebas de datos sintéticos las predicciones se obtienen con alta precisión ya que creamos la serie datos sintéticos con las características conocidas y deseadas, así los errores del modelo tanto para esta serie como para los errores medios de 10 realizaciones son:

RMSE	$\overline{\text{RMSE}}$	MAE	$\overline{\text{MAE}}$	MAPE	$\overline{\text{MAPE}}$
0.986	$0,990 \pm 0,074$	0.805	$0,788 \pm 0,049$	1.649	$1,545 \pm 0,571$

Tabla 11. Errores de predicción medios y desviación típica media de 10 realizaciones ARMA(1, 1)

5.4. Prophet

Para este modelo vamos a generar un conjunto de datos sintéticos pero esta vez sin seguir un patrón AR, MA o ARMA, si no que establecemos las características de los datos para que se ajusten al modelo Prophet. Los datos serán de un año, es decir, tendremos un conjunto de datos con 365 valores, y los parámetros que se han introducido son una componente estacional, una tendencia lineal y un ruido aleatorio, parámetros que se suman a los datos para crear la serie que se va a entrenar con Prophet, obteniendo la siguiente predicción:

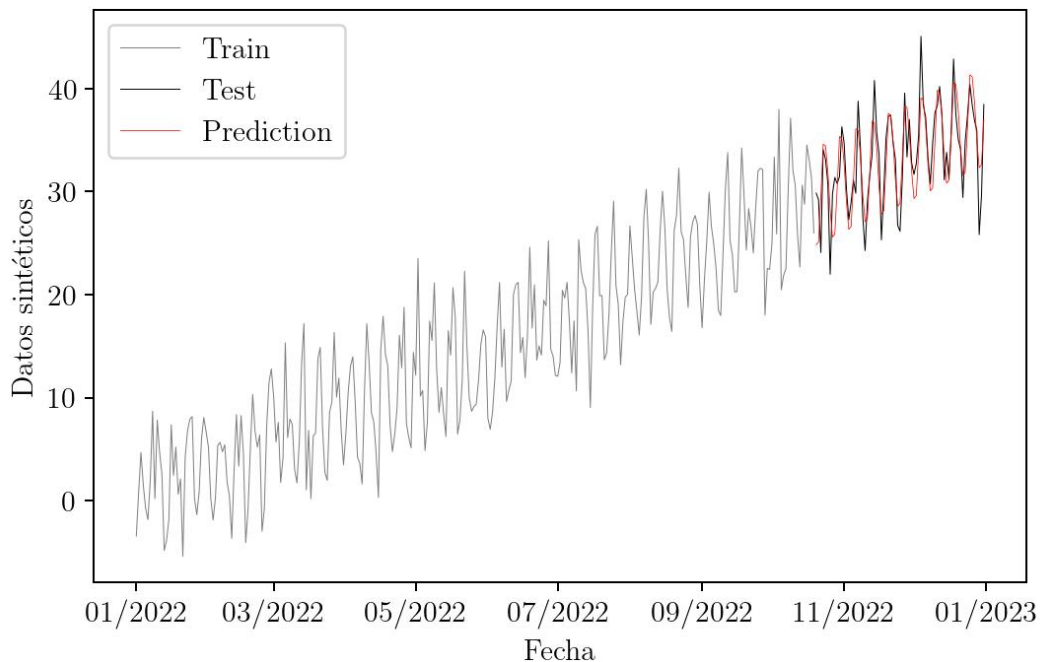


Figura 25. Serie sintética original junto con la predicción con el modelo Prophet (Fuente: elaboración propia)

Además de la predicción del modelo Prophet nos ofrece una gráfica con sus componentes de tendencia y estacionalidad ya sea diaria, semanal o anual. Se puede ver en la Figura 25 que la predicción del modelo sintético es bastante precisa, y podemos ver a continuación los errores predictivos de este modelo

RMSE	MAE	MAPE
3.245	2.549	0.315

Tabla 12. Errores de predicción del modelo Prophet

Si la media de los datos de la serie es 17.9 estos errores nos indican que el ajuste de la predicción a los datos es muy bueno.

6. Aplicación a datos del Mar Menor. Caso real

El Mar Menor es una laguna costera situada en la Región de Murcia (España) que se enfrenta a problemas ambientales y ecológicos de una gran magnitud, lo que ha generado la necesidad de analizar y comprender su evolución, así como la tendencia de sus indicadores ambientales a lo largo del tiempo. El análisis de series temporales en el contexto del Mar Menor permite obtener información valiosa sobre la dinámica y los cambios que experimenta esta laguna costera. Estos datos pueden proporcionar información clave para la toma de decisiones en la gestión y conservación de este ecosistema, así como para la implementación de medidas adecuadas de protección y restauración. No obstante, el análisis de series temporales presenta desafíos particulares. Las series temporales pueden ser complejas y estar influenciadas por múltiples factores, como los ciclos estacionales, los fenómenos climáticos y las actividades humanas. Además, pueden presentar irregularidades, datos faltantes y ruido, lo que dificulta su interpretación y modelado. En este contexto, la aplicación de la metodología expuesta en este trabajo a los datos reales del Mar Menor busca abordar estas problemáticas y proporcionar una comprensión más profunda de la dinámica y los patrones presentes en las series temporales.

Desde el portal de datos del Mar Menor, se han descargado los datos históricos estandarizados y tratados (nivel L4) de todos los parámetros ambientales allí provistos, de los que se han seleccionado los siguientes seis parámetros y se analizarán en las consiguientes secciones con el orden dispuesto en la lista:

- Clorofila A [mg/L].
- Ficoeritrina (PE).
- Oxígeno disuelto [mg/L].
- Salinidad [PSU].
- Temperatura del agua [°C].
- Transparencia [m].

El intervalo temporal abarcado por estas series históricas es de unos 5 años aproximadamente, desde 2017 hasta 2022. Los datos provienen de las diferentes estaciones de monitorización repartidas por todo el Mar Menor (nivel L1), y han sido posteriormente interpolados y estandarizados por el suministrador a un mallado común, tal y como se muestra en la figura, donde se representa, en azul las estaciones de la OISMA (Oficina de Impulso Socioeconómico del Medio Ambiente) y en rojo las estaciones del Servicio de Pesca.



Figura 26. Área de selección empleada para el presente proyecto. (Fuente: SDC – UPCT)

6.1. Fase 0: Visualización y limpieza de datos

Para cada indicador se va a hacer una primera representación gráfica que servirá para observar los posibles valores atípicos que puedan mostrar los datos (Tabla 13). En este caso, tal y como se ha mencionado, estos datos ya vienen tratados por el suministrador, por lo que los posibles valores atípicos que se encuentren aportarán un valor añadido a las series temporales y, por ello, no se eliminarán.

Datos	Nº datos total	Rango	Nº datos atípicos
Clorofila	975	06/2017-01/2020	47
Ficoeritrina (PE)	487	07/2021-10/2022	22
Oxígeno	1737	04/2017-12/2021	0
Salinidad	2041	04/2017-10/2022	0
Temperatura	974	04/2017-11/2019	0
Transparencia	2252	09/2016-10/2022	0

Tabla 13. Dimensión de datos reales junto con el número de datos atípicos

Así, las únicas dos series que cuentan con datos atípicos son las de Clorofila y Ficoeritrina.

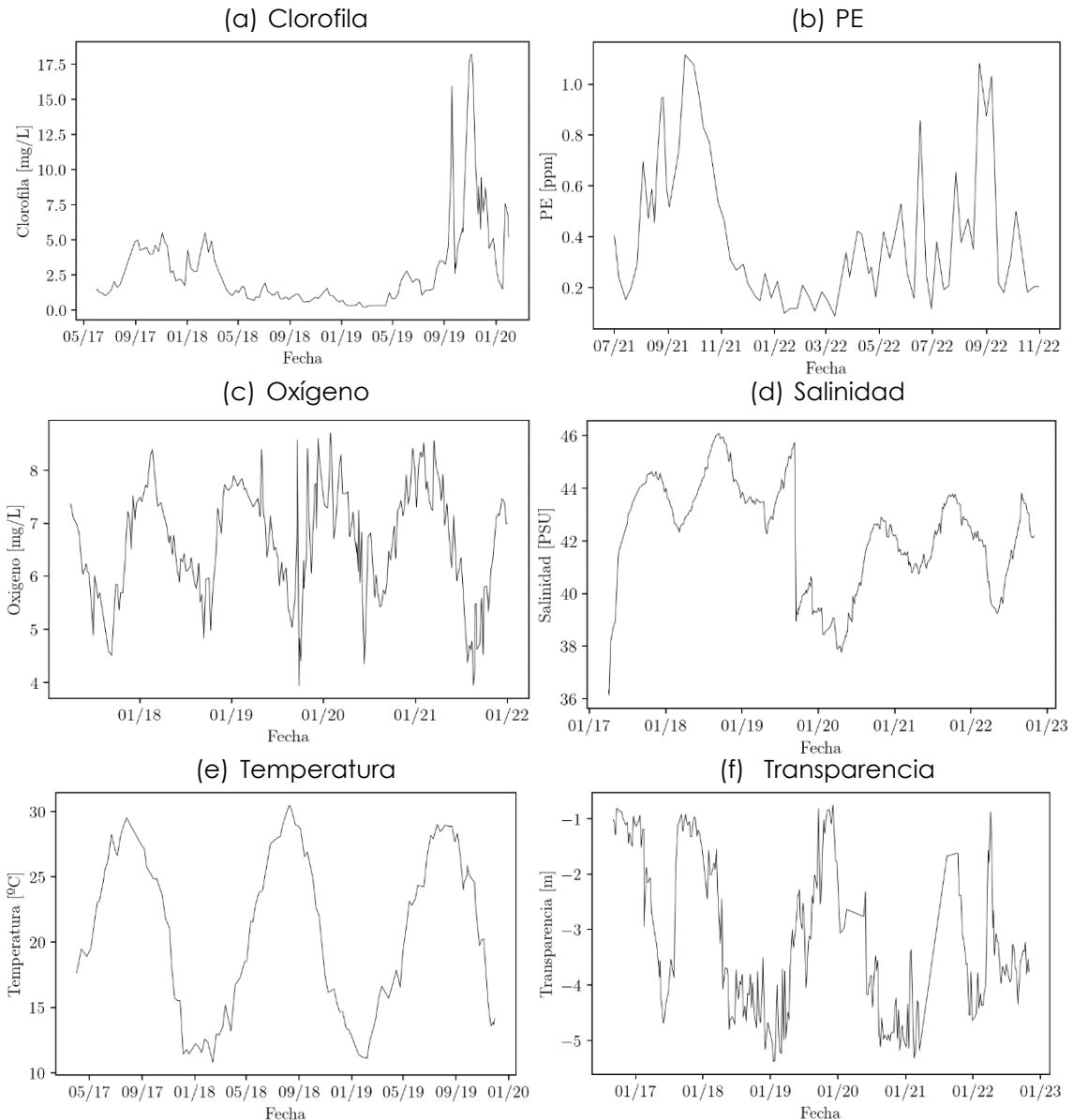


Figura 27. Series de tiempo de indicadores del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)

6.2. Fase 1: Identificación de patrones

- **Clorofila**

En esta fase se van a identificar los patrones que tiene el conjunto de datos y esto se hará con la ayuda de la descomposición estacional. La Figura 28 indica que la serie parece ser **estacional** con pequeños tendencias iniciales y finales. Debido a esta tendencia, la serie **no es estacionaria**. Finalmente, la prueba de Dickey-Fuller confirma la ausencia de estacionariedad como se observa en la Tabla 14, donde se muestran los p-valores asociados a los test realizados para los conjuntos de datos analizados. El p-valor supera el nivel de referencia de 0.05, aunque cabe destacar que el valor, aunque lo supera, es muy cercano a 0.05, y si se hubiera considerado un umbral de confianza del 90% esta serie podría considerarse estacionaria.

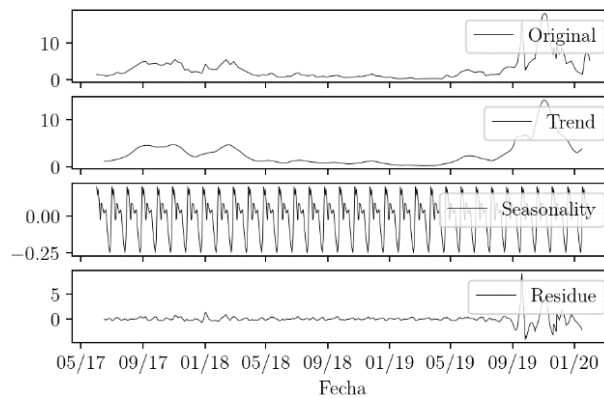


Figura 28. Descomposición estacional de la serie Clorofila (Fuente: elaboración propia. Datos: SDC - UPCT)

Se hará uso de las gráficas de autocorrelación para detectar si el conjunto de datos es estacional, y además para estimar cuál podría ser el modelo más adecuado para tratar el conjunto de datos, modelo que confirmaremos en la Fase 3, con los criterios y métricas correspondientes.

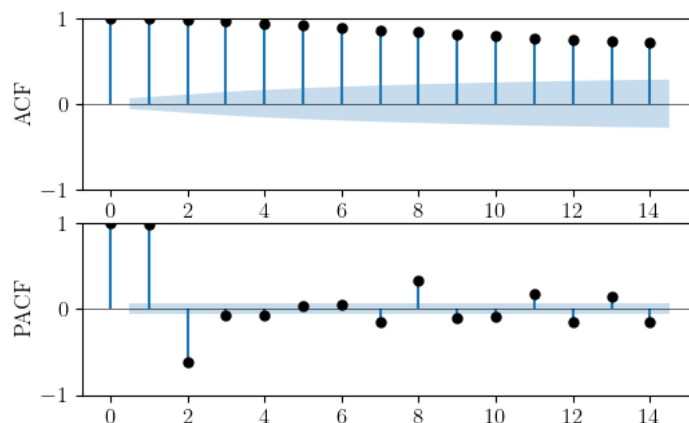


Figura 29. Funciones de autocorrelación simple y parcial para la serie Clorofila (Fuente: elaboración propia. Datos: SDC - UPCT)

Se aprecia claramente que esta serie es una serie estacional, en la gráfica de ACF el retardo en el periodo de esta serie (12) es significativo. Según todo lo visto en esta Fase, podríamos estimar que el mejor modelo de esta serie es un modelo SARIMA, por no ser estacionaria y a su vez es estacional, y el parámetro autorregresivo para la serie podría ser $p=3$, por los retardos significativos que se aprecian en la gráfica de PACF. El modelo completo y los demás parámetros lo confirmamos en la Fase 3.

- PE

Vamos a realizar la fase de identificación de patrones para la serie de tiempo de PE. De la misma manera que la serie de Clorofila tiene una tendencia al inicio y al final. En el caso de la estacionalidad parece tener, pero también podemos ver que las unidades son bastante pequeñas. Por ello habrá que comprobar estas suposiciones.

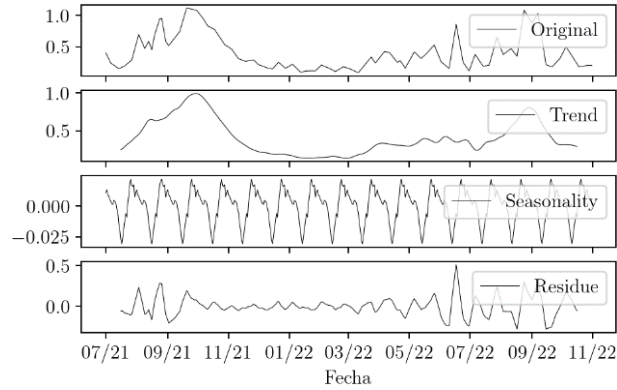


Figura 30. Descomposición estacional de la serie PE del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)

Para la estacionariedad, según la Tabla 14, donde el p-valor superan el umbral de 0.05, concluimos que se trata de una serie **no estacionaria** y tendremos que aplicar una diferenciación. Por otro lado, los resultados de la estacionalidad indican que tenemos una serie sin esta componente, ya que el retardo de la gráfica de autocorrelación en el valor del periodo (30) no es significativo, es menor a 0.5, que es el umbral que se ha establecido para saber si una serie presenta estacionalidad. Por tanto, concluimos la serie **no es estacional** y no será necesaria la diferenciación estacional.

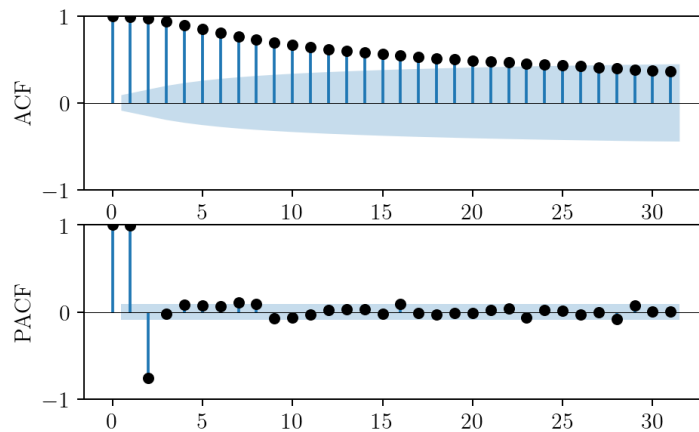


Figura 31. Autocorrelación simple y parcial para la serie PE (Fuente: elaboración propia. Datos: SDC - UPCT)

En este caso no podemos estimar claramente un valor del parámetro autorregresivo p ya que la gráfica de autocorrelación parcial no es muy estable, aunque por los dos retardos más significativos podría ser $p=2$, el modelo se estimará más adelante siguiendo los criterios Akaike y Bayesiano.

- **Oxígeno**

En el conjunto de concentración de oxígeno en el agua, si observamos su descomposición estacional observamos que se trata de una serie claramente estacional, ya que la serie se repite de forma periódica. Además, en la Figura 33 se observa una autocorrelación simple que tiene el retardo correspondiente al periodo de este conjunto con un valor significativo, confirmando la **estacionalidad**.

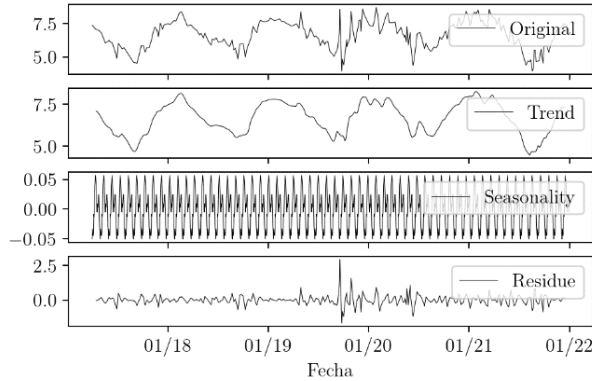


Figura 32. Descomposición estacional de la serie de Oxígeno del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)

Para la estacionariedad se puede observar en primer lugar la tendencia de la gráfica anterior viéndose cómo tiene tendencia creciente y decreciente a lo largo del tiempo, por lo que tendremos una serie **sin estacionariedad**. Esto nos lo confirma la prueba de Dickey-Fuller, con un p-valor superior a 0.05, indicador de una serie no estacionaria. Aunque, como en la clorofila este valor de p es muy cercano a 0.05 y podría considerarse, dependiendo del nivel de significación, que la serie es estacionaria.

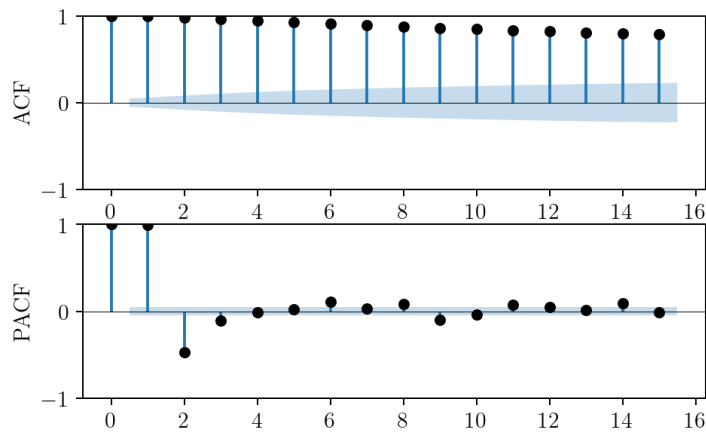


Figura 33. Autocorrelación simple y parcial de la serie de Oxígeno del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)

Por último, podemos fijarnos en la gráfica anterior la gráfica de autocorrelación parcial y así estimar el parámetro autorregresivo, que parece ser, debido a sus dos retardos significativos, igual a 2, $p = 2$.

- **Salinidad**

Para el caso de la salinidad del Mar Menor tenemos una serie que a simple vista no parece estacionaria, ya que la tendencia no es clara y parece que, cuenta con estacionalidad, por lo que vamos a comprobar dicha estacionalidad con la gráfica de ACF de la Figura 35, que nos dice, como se había supuesto, que es un conjunto **con estacionalidad** al tener el retardo correspondiente al periodo valor significativo (> 0.5).

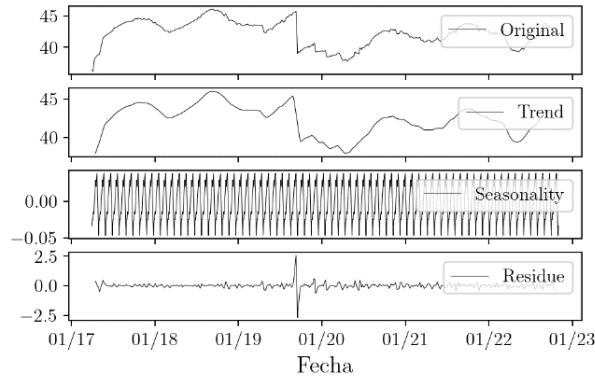


Figura 34. Descomposición estacional serie de salinidad (Fuente: elaboración propia. Datos: SDC - UPCT)

A diferencia de lo que habíamos supuesto viendo la descomposición estacional esta serie **no es estacionaria** ya que tenemos un p-valor que supera el umbral, aunque no sea mucha diferencia y como en los otros casos, con otro criterio podría considerarse estacionaria. Por lo tanto, esta serie va a necesitar tanto diferenciación como desestacionalización.

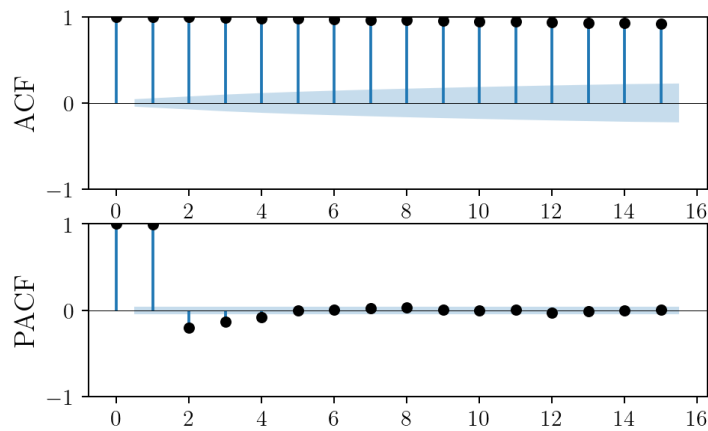


Figura 35. Autocorrelación simple y parcial de la serie de salinidad (Fuente: elaboración propia. Datos: SDC - UPCT)

Por último, con esta gráfica de PACF, viendo sus retardos significativos podemos estimar el posible valor que tendrá el parámetro autorregresivo, p parece ser 2. Lo comprobaremos más adelante mediante los criterios de información.

- **Temperatura**

Para la temperatura tenemos una serie **estacional** sin duda, ya que vemos como tenemos un patrón claro de repetición de la serie en el tiempo, y además, si se observa la gráfica de autocorrelación simple tenemos que el retardo en el valor del periodo (12) es mayor a 0.5, confirmando esto las suposiciones de estacionalidad y, por consiguiente, a necesidad de realizar una diferenciación estacional.

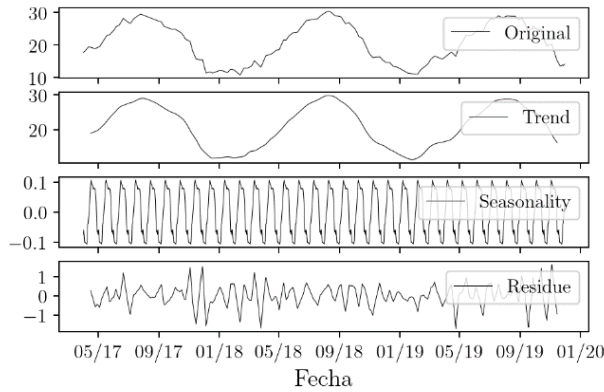


Figura 36. Descomposición estacional serie de temperatura del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)

Por otro lado, también contamos con una serie **sin estacionariedad** ya que a lo largo del tiempo va presentando diferentes tendencias. Esto se confirma con el p-valor, que es mayor al umbral como se ve en la Tabla 14

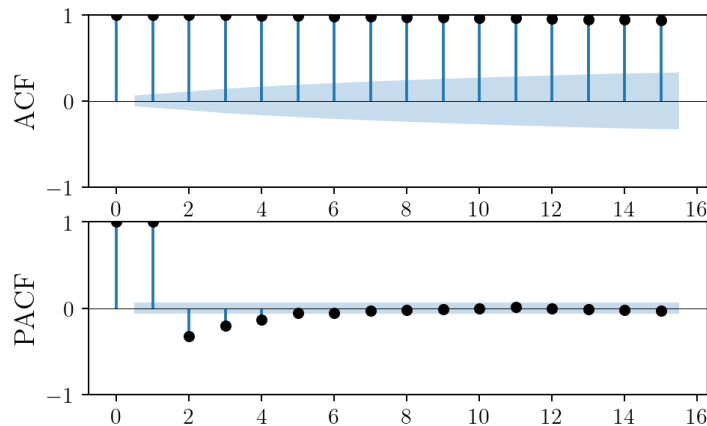


Figura 37. Autocorrelación parcial serie de temperatura (Fuente: elaboración propia. Datos: SDC - UPCT)

Viendo la gráfica de autocorrelación parcial de la gráfica anterior no queda claro cuál sería el valor de la componente autorregresiva de la serie ya que no hay un único retardo significativo, por lo que el modelo de la serie se estimará mediante los criterios de información conocidos (AIC, BIC)

- **Transparencia**

Para la transparencia tenemos un caso parecido a la temperatura, una gráfica con una forma que parece repetirse y que tiene tendencia creciente y decreciente durante todo su espacio temporal. Esta gráfica es **estacional** y, como en casos anteriores, se confirma con la Figura 38, de correlación simple y su retardo significativo en el valor del periodo de la serie.

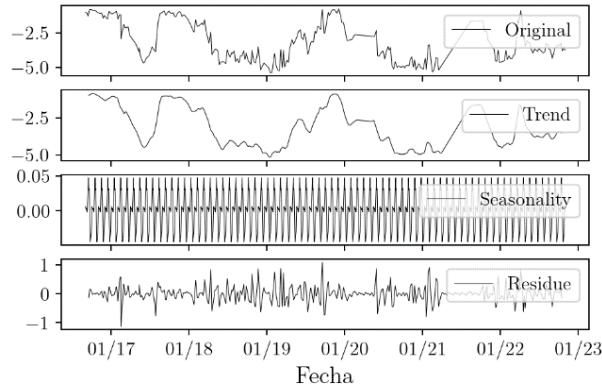


Figura 38. Descomposición estacional de la serie de transparencia (Fuente: elaboración propia. Datos: SDC - UPCT)

La ausencia de estacionariedad se confirma con la presencia de tendencia en la serie y con el p-valor, que es mayor a 0.05, aunque no es excesivamente alto, con el nivel de significación adoptado se considera una serie **no estacionaria**.

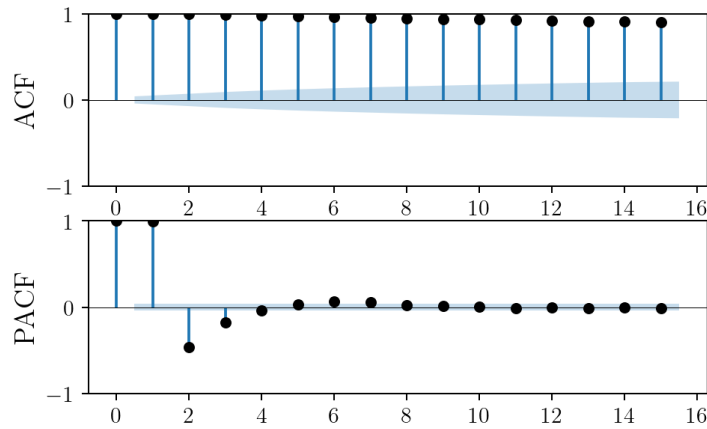


Figura 39. Autocorrelación simple serie de temperatura (Fuente: elaboración propia. Datos: SDC - UPCT)

En este caso, al igual que en casos anteriores, podemos estimar el parámetro autorregresivo a través de los retardos significativos de la autocorrelación parcial. Es posible que p tenga un valor de 3, aunque se verificará más adelante.

Conjunto	p-valor	Estacionariedad
Clorofila	0.065	No
Ficoeritrina (PE)	0.209	No
Oxígeno	0.055	No
Salinidad	0.067	No
Temperatura	0.072	No
Transparencia	0.061	No

Tabla 14. Valor de p de la prueba de Dickey-Fuller de los conjuntos de datos del Mar Menor

6.3. Fase 2: Intento de mejora de ajuste: diferenciación y desestacionalización

Con las conclusiones anteriores, en las que se clasifican las series según estacionalidad y estacionariedad, podemos ahora aplicar una diferenciación y/o desestacionalización y así intentar mejorar el ajuste de los modelos a las series de datos y poder estimar el modelo adecuado para estas series con más precisión, haciendo que las predicciones tengan menos error y aumenten su calidad. A continuación, se presentan los conjuntos de datos desestacionalizados o diferenciados según corresponda:

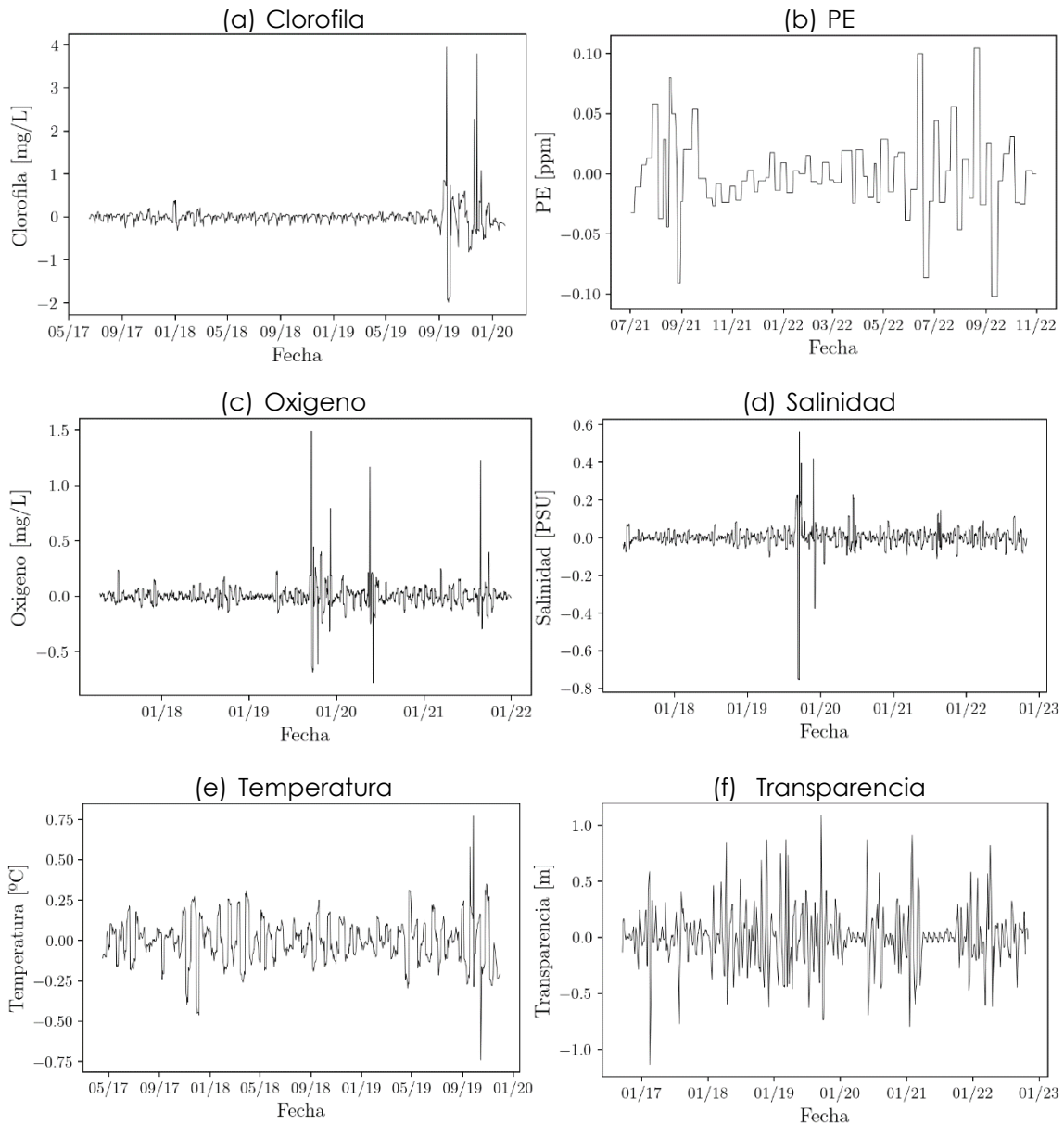


Figura 40. Series de tiempo del Mar Menor diferenciadas (Fuente: elaboración propia. Datos: SDC - UPCT)

6.4. Fase 3: Selección del modelo y explicación de patrones

En esta fase, utilizaremos los criterios Akaike y Bayesiano para realizar combinaciones de los parámetros de los modelos dentro del rango de 0 a 2 y observar la combinación que nos ofrezca un menor valor de AIC y BIC. Teniendo en cuenta el valor menor de los criterios junto con las características observadas en fases anteriores y los intentos de mejora mediante la diferenciación y desestacionalización, identificaremos los mejores modelos para cada serie de tiempo. Los resultados se presentarán en la siguiente tabla:

Conjunto	Modelo	AIC	BIC
Clorofila	SARIMA(2,1,1) (0,1,1)	393.866	417.483
Ficoeritrina (PE)	ARIMA(2,1,1)	-2418.080	-2401.080
Oxígeno	SARIMA(2,1,2) (0,1,1)	-2996.446	-2963.794
Salinidad	SARIMA(2,1,2) (0,1,1)	-6133.040	-6099.426
Temperatura	SARIMA(2,1,0) (0,1,1)	-1686.450	-1667.054
Transparencia	SARIMA(1,1,2) (0,1,1)	-4900.312	-4871.730

Tabla 15. Modelos predictivos para cada conjunto de datos según los criterios AIC y BIC

Se puede observar que el oxígeno y la salinidad comparten el modelo predictivo más adecuado para sus características. A modo de prueba, se va a aumentar el rango para realizar combinaciones de 0 a 3 y ver si el AIC y el BIC es menor para el oxígeno y la salinidad.

Conjunto	Modelo	AIC	BIC
Oxígeno	SARIMA(2, 1, 3) (0, 1, 1)	-3023.798	-2985.705
Salinidad	SARIMA(3, 1, 2) (0, 1, 1)	-6482.535	-6443.294

Tabla 16. Modelos predictivos para los conjuntos de oxígeno y salinidad según AIC y BIC

Como podemos ver si aumentamos el rango disminuyen los valores de AIC y BIC y, además, ya no obtenemos el mismo modelo, por lo tanto, en estos dos casos se usará para la predicción estos nuevos modelos. Para el resto de los datos aplicamos los modelos seleccionados en un primer momento y realizamos la predicción, además, evaluaremos la bondad del ajuste y así comprobaremos que hemos elegido el modelo de manera adecuada y que la metodología se ha aplicado correctamente.

6.5. Fase 4: Aplicación de modelos predictivos

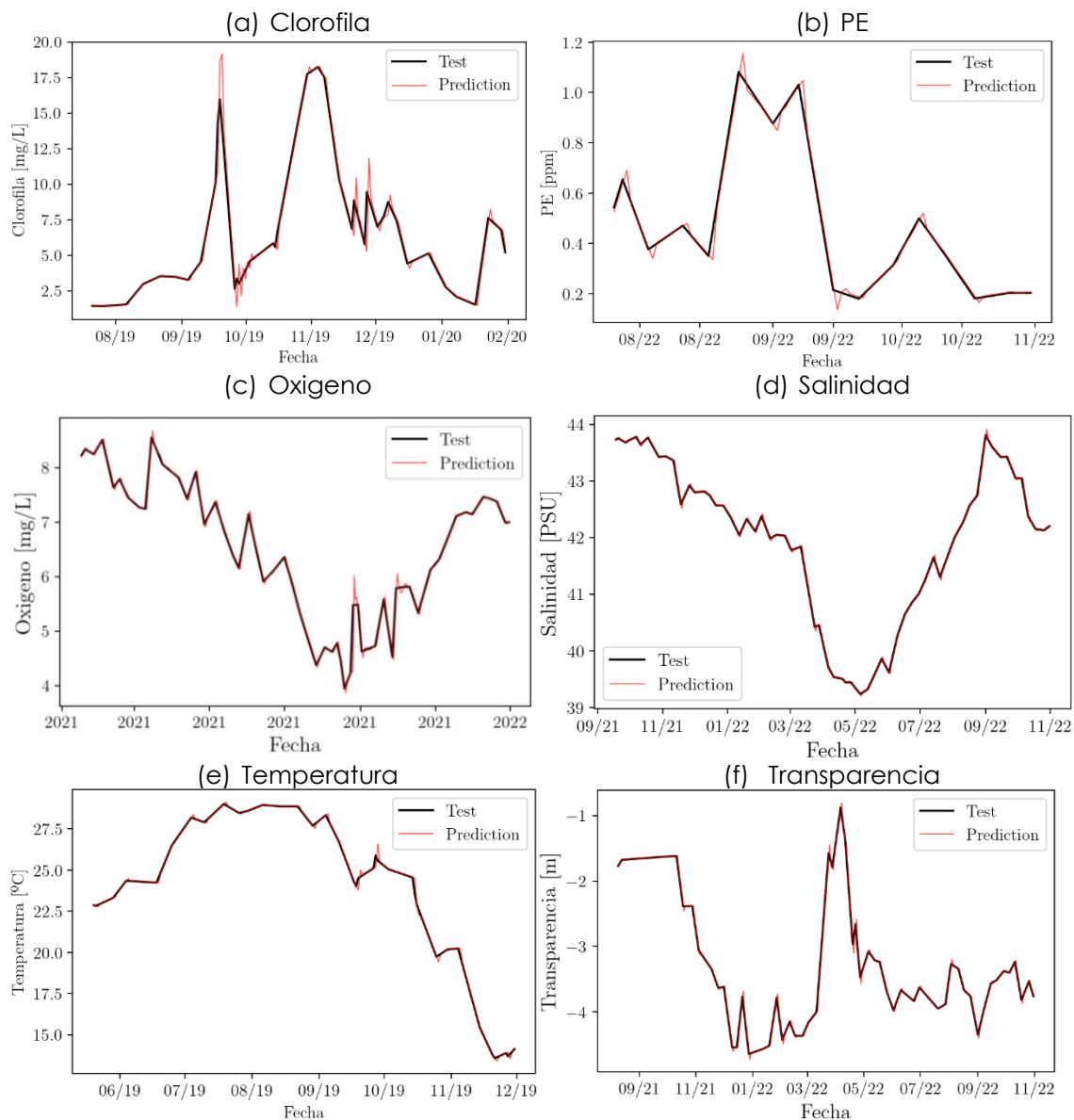


Figura 41. Predicción del 20% de prueba de las series del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)

Es importante recordar que estos modelos se ajustan de manera muy precisa ya que el conjunto de entrenamiento se actualiza en cada instante temporal permitiendo así mejorar las predicciones. El buen ajuste de los resultados que se han obtenido para cada uno de los conjuntos de datos utilizados se muestran, a continuación, mediante las métricas de RMSE, MAE y MAPE.

Conjunto	RMSE	MAE	MAPE
Clorofila	0.741	0.325	0.048
Ficoeritrina (PE)	0.027	0.015	0.034
Oxígeno	0.094	0.046	0.007
Salinidad	0.025	0.015	0.001
Temperatura	0.151	0.061	0.002
Transparencia	0.044	0.027	0.009

Tabla 17. Errores predictivos de las series de tiempo del Mar Menor obtenidos con modelos estadísticos

Estas predicciones se realizan también mediante el modelo Facebook Prophet y las métricas de los errores se muestran en la tabla de a continuación:

Conjunto	RMSE	MAE	MAPE
Clorofila	1.569	1.036	0.685
Ficoeritrina (PE)	0.233	0.179	0.606
Oxígeno	0.393	0.292	0.047
Salinidad	0.394	0.262	0.006
Temperatura	0.707	0.555	0.031
Transparencia	0.690	0.564	0.243

Tabla 18. Errores predictivos de las series de tiempo del Mar Menor obtenidos con el modelo Prophet

Estos resultados, en general, son errores bastante bajos, indicativo de que los modelos se adaptan bien a los conjuntos de datos, como se puede apreciar en la Figura 41. El conjunto de datos con menor error es el de Salinidad para los modelos estadísticos, con un porcentaje de error absoluto medio del 0,1%. Con el modelo Prophet el menor error lo tenemos para la serie de Ficoeritrina, aunque el MAPE en esta es ligeramente mayor que el resto. Podemos confirmar con estos resultados que la metodología seguida y la clasificación de la serie según sus características es adecuada.

7. Conclusiones

En este Trabajo de Fin de Estudios (TFE), se ha seguido una metodología sistemática para el análisis de series temporales. A lo largo de este estudio, se han identificado y aplicado una serie de pasos clave que han demostrado ser fundamentales en la obtención de resultados precisos y confiables.

- En primer lugar, se ha destacado la importancia de comprender a fondo la serie temporal, establecer claramente los objetivos del análisis y evaluar los requerimientos de la serie temporal. Esto implica conocer su naturaleza para examinar y comprender tanto la serie como las características de dicha serie. Se han identificado las dos características de las series a las que más hay que prestar atención: estacionalidad y estacionariedad. Esta comprensión detallada de la serie es crucial para reducir los errores y obtener predicciones más precisas. Se concluye que esta parte del proyecto es fundamental para entender los datos y poder prepararlos para los análisis posteriores.
- En segundo lugar, se han clasificado las series según n base a las características anteriores y mediante las métricas de clasificación adecuadas como son el análisis de la autocorrelación y la prueba de Dickey Fuller. Para realizar la clasificación de una serie es importante disponer de unos datos limpios para que la clasificación sea fiable y con datos de calidad. Esto permite una correcta selección del modelo predictivo, construcción y evaluación del modelo predictivo. También se destaca en este trabajo la importancia de tener unas métricas adecuadas para la evaluación de los modelos seleccionados como son las métricas RMSE, MAE y MAPE.
- Por último, mediante las diferentes pruebas realizadas en el desarrollo del trabajo siguiendo la metodología con datos de diferentes tipologías como sintéticos y reales, se ha observado que los modelos estadísticos son eficaces para la predicción a corto plazo cuando se alimentan correctamente. Estos modelos han demostrado un buen rendimiento en la predicción, con errores reducidos al evaluar las predicciones sobre el 20% de la serie utilizado como conjunto de prueba. Por otro lado, Prophet, en general arroja buenos resultados a la hora de predecir las series de tiempo reales al completo.
- En el caso del Mar Menor, en particular, se pueden observar unas series de tiempo que al estar tratados no cuentan con gran cantidad de datos atípicos y no hay datos faltantes, una vez entrenados los datos se han hecho las predicciones mediante modelos estadísticos y con el modelo Facebook Prophet. Los errores de estos modelos predictivos son bastante bajos. Para los modelos estadísticos el menor error lo tenemos en el conjunto de Salinidad con un RMSE de 0.025 y un porcentaje de error absoluto medio del 0,1%. Por otro lado, con el modelo de Facebook Prophet el menor error lo tenemos en la serie de Ficoeritrina con un RMSE de 0.233 y un porcentaje de error absoluto medio del 0.6%.

Con todo, cabe destacar la gran variabilidad de las series temporales, así como los distintos enfoques con los que se puede abordar un conjunto de datos con evolución temporal, remarcando así la importancia de una buena clasificación de cada serie temporal y de seguir una metodología estructurada y clara para el tratamiento de los datos y la aplicación del modelo predictivo.

8. Líneas futuras

A partir de los resultados obtenidos en este proyecto, se pueden establecer áreas de investigación que podrían ampliar y consolidar el análisis en futuros estudios.

- Una de las líneas futuras para este trabajo podría ser la consideración de series temporales multivariantes. Actualmente, el enfoque se ha centrado en el análisis de series temporales univariantes, pero considerar múltiples variables puede proporcionar una perspectiva más completa de las series temporales. Al trabajar con series multivariantes, se pueden identificar interacciones y relaciones entre las variables, lo que puede llevar a una mejor comprensión de los patrones en los datos. La consideración de series temporales multivariantes también abriría la posibilidad de aplicar técnicas avanzadas de análisis de datos, como el análisis de componentes principales. En consecuencia, la inclusión de series temporales multivariantes en futuros estudios podría ampliar el alcance y la aplicabilidad de la metodología propuesta, ofreciendo nuevas oportunidades para el análisis y la toma de decisiones en diversas áreas que involucran series temporales complejas y con múltiples variables interrelacionadas.
- Otra línea futura de investigación es la exploración de algoritmos de inteligencia artificial y machine learning para realizar predicciones a futuro en lugar de utilizar exclusivamente modelos estadísticos como AR, MA o ARMA. Estos algoritmos avanzados pueden ofrecer un enfoque más dinámico, una visión más completa y adaptable para el análisis de series temporales, ya que son capaces de aprender patrones y relaciones complejas en los datos. La aplicación de técnicas de inteligencia artificial, como redes neuronales recurrentes (RNN), máquinas de soporte vectorial (SVM), o redes neuronales convolucionales (CNN), podría permitir una mayor flexibilidad y precisión en la predicción de valores futuros en las series temporales. Por ello, sería interesante en un futuro realizar una comparación de los resultados con otros obtenidos con modelos de inteligencia artificial. Hay que tener en cuenta que los modelos de inteligencia artificial necesitan gran cantidad de datos respecto a los tradicionales, por lo que podría haber una falta de datos usando estos modelos para las series del Mar Menor.

Anexos

A. Aplicación a datos reales (Peña, 2010)

Para apoyar el seguimiento de la metodología expuesta en este TFE se han realizado pruebas teóricas y con datos del Mar Menor como se ha visto anteriormente pero también se han realizado pruebas con conjuntos de datos reales provenientes del libro de series temporales de Daniel Peña (2010). Alguna de estas pruebas y datos se han utilizado como ejemplos en los apartados de este trabajo, como los datos de ozono, de pasajeros de vuelos y los datos de población mayor a 16 años. Por otro lado, en este anexo se podrán encontrar las pruebas completas realizadas para los datos provenientes de este libro que se corresponden a las lluvias de Santiago de Compostela, accidentes laborales y consumo de gasolina.

• Datos lluvias Santiago de Compostela

En el caso de las lluvias de Santiago de Compostela durante 10 años (1988-1998) se observa una serie que en la que no se aprecia tendencia, aunque sí estacionalidad. Esto indica que es una serie estacionaria, y su p-valor, como vemos en la Tabla 25, es menor que 0.05.

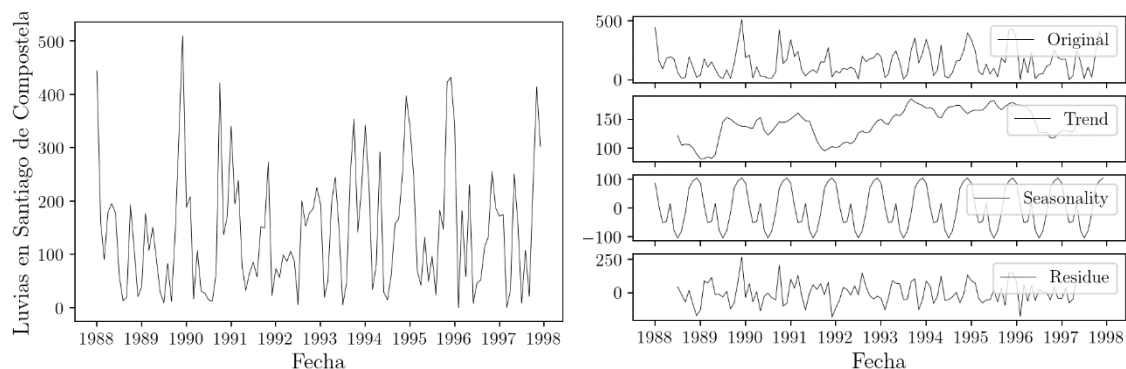


Figura 42. Gráfica de la evolución temporal de lluvias en Santiago de Compostela y su descomposición estacional (Fuente: elaboración propia. Datos: Peña, 2010)

Analizando los retardos en la correlación se observa que el retardo en el periodo de esta serie, que en este caso es 12, no es significativo, por lo que no se aprecia componente de estacionalidad.

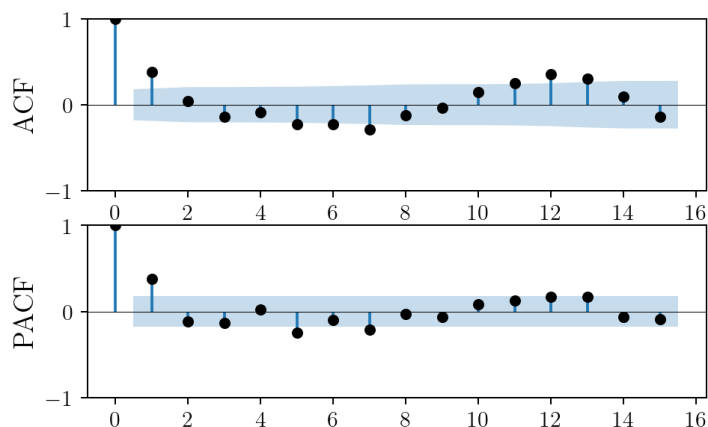


Figura 43. Autocorrelación simple y parcial de la serie de datos de lluvia en Santiago de Compostela (Fuente: elaboración propia. Datos: Peña, 2010)

Según estas premisas y los criterios de información Akaike y Bayesiano que se han estudiado en el trabajo obtenemos que para este conjunto de datos el modelo más adecuado es un ARMA(2,2), obteniendo la siguiente predicción sobre el 20% de la serie:

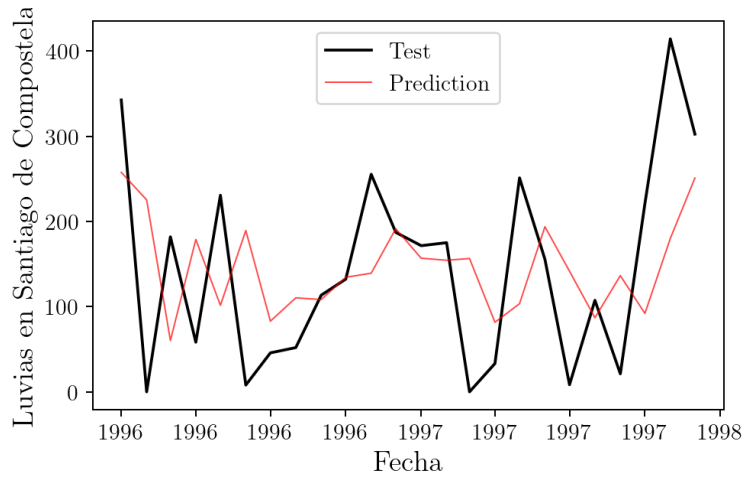


Figura 44. Predicción de los datos de lluvia de Santiago de Compostela utilizando el modelo ARMA (Fuente: elaboración propia. Datos: Peña, 2010)

Este modelo tiene los siguientes errores de predicción, que se presentan en las unidades de la serie de tiempo:

Conjunto	RMSE	MAE
Lluvias en Santiago de Compostela	113.94	91.50

Tabla 19. Errores de predicción para la serie de tiempo de lluvias utilizando el modelo ARMA

Por otro lado, con el modelo Prophet se obtiene la siguiente predicción y errores:

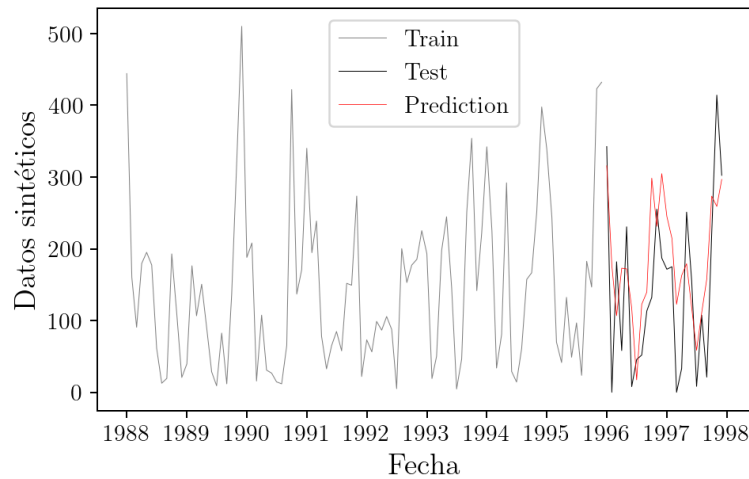


Figura 45. Predicción de los datos de lluvia de Santiago de Compostela utilizando el modelo Prophet (Fuente: elaboración propia. Datos: Peña, 2010)

Conjunto	RMSE	MAE
Lluvias en Santiago de Compostela	93.176	78.078

Tabla 20. Errores de predicción datos de lluvias de Santiago de Compostela utilizando el modelo Prophet

- **Datos de accidentes laborales**

En esta serie de tiempo se puede observar una tendencia creciente clara y además un patrón que se repite por lo que concluimos que es una serie con estacionalidad. Observando el p-valor, confirmamos que la serie no presenta estacionariedad.

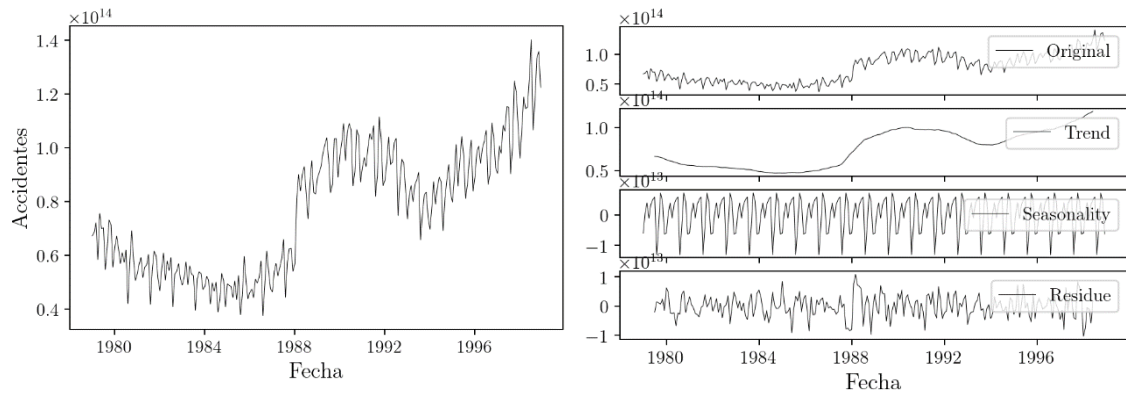


Figura 46. Gráfica de accidentes junto con su descomposición estacional (Fuente: elaboración propia. Datos: Peña, 2010)

Mediante los retardos de la correlación confirmamos que sí existe estacionalidad en este conjunto de datos por lo que el mejor modelo será un modelo SARIMA.

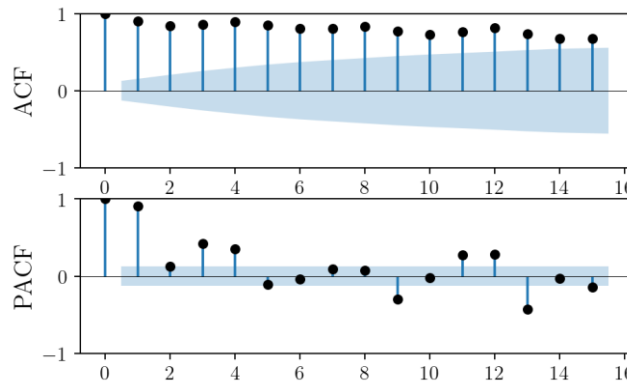


Figura 47. Autocorrelación simple y parcial de la serie de accidentes (Fuente: elaboración propia. Datos: Peña, 2010)

Los parámetros del modelo mediante los criterios de información estudiados serán (2,1,3) y (1,1,1) para la parte estacional, por lo que el modelo se denota como SARIMA(2,1,3)(1,1,1). La predicción de este conjunto de datos y sus errores predictivos se presentan en las consiguientes gráfica y tabla:

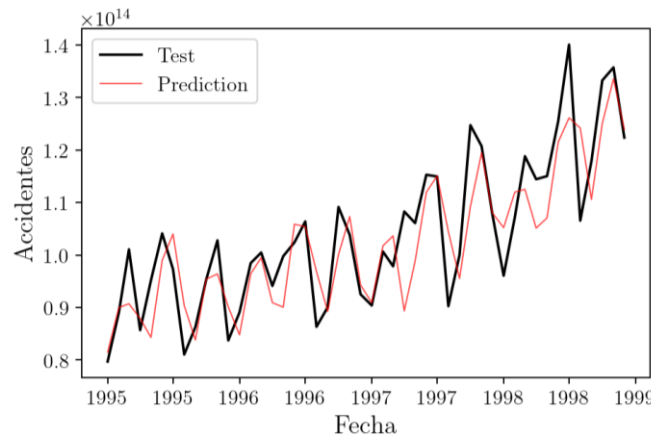


Figura 48. Predicción de la serie de accidentes con el modelo SARIMA (Fuente: elaboración propia. Datos: Peña, 2010)

Conjunto	RMSE	MAE	MAPE
Accidentes	$0.074 \cdot 10^{14}$	$0.057 \cdot 10^{14}$	0.055

Tabla 21. Errores predictivos de la serie de accidentes laborales utilizando el modelo SARIMA

Si modelizamos esta serie con Prophet:

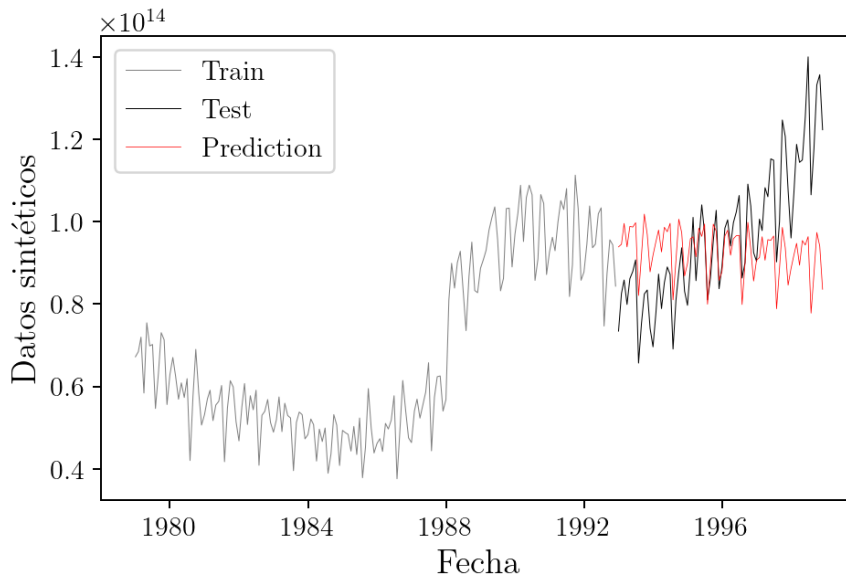


Figura 49. Predicción de la serie de accidentes con Prophet (Fuente: elaboración propia. Datos: Peña, 2010)

Conjunto	RMSE	MAE	MAPE
Accidentes	$0.16 \cdot 10^{14}$	$0.13 \cdot 10^{14}$	0.056

Tabla 22. Errores predictivos con Prophet de la serie de accidentes

- **Datos del consumo de gasolina**

En esta serie, del mismo modo que la anterior, se observan unos datos con unas claras componentes tanto de tendencia como estacionalidad, tanto por la forma de la serie como por los valores del periodo de la serie, observado en la Figura 51, y su p-valor mostrado en la Tabla 25.

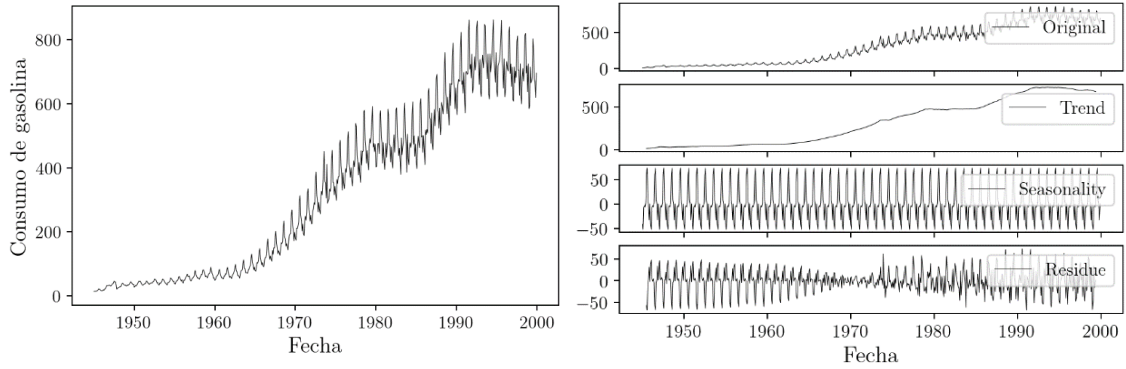


Figura 50. Serie de consumo de gasolina y su descomposición estacional (Fuente: elaboración propia. Datos: Peña, 2010)

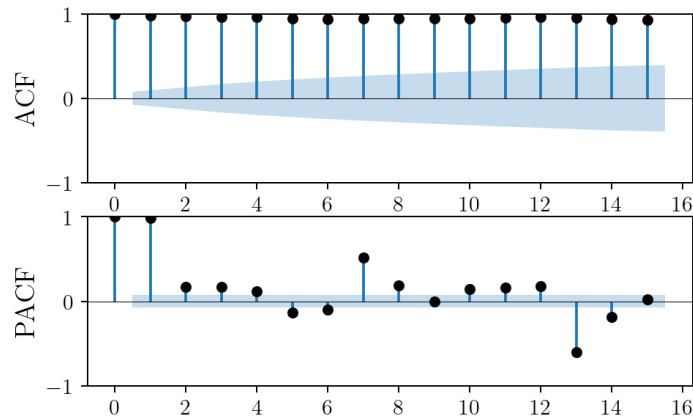


Figura 51. Autocorrelación simple y parcial de la serie de consumo de gasolina (Fuente: elaboración propia. Datos: Peña, 2010)

Así el modelo más adecuado para este conjunto de datos según los criterios de información es el modelo SARIMA(3, 1, 3) (0, 1, 1), y obtenemos la siguiente predicción:

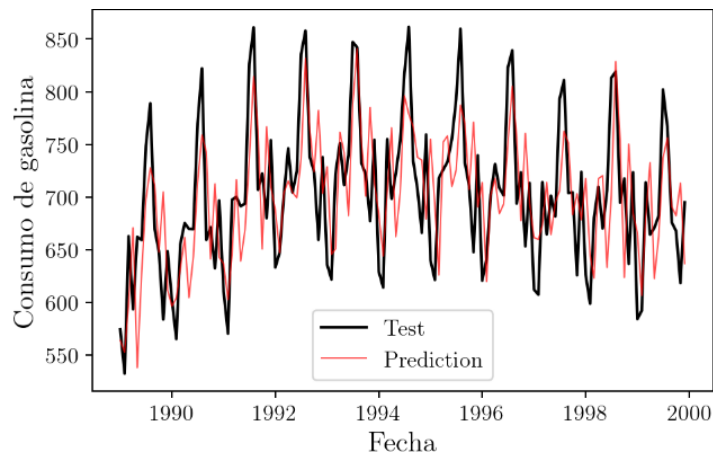


Figura 52. Predicción de la serie de consumo de gasolina con SARIMA (Fuente: elaboración propia. Datos: Peña, 2010)

Los errores de predicción del modelo empleado serán los siguientes:

Conjunto	RMSE	MAE	MAPE
Consumo de gasolina	57.862	48.755	0.07

Tabla 23. Errores predictivos datos del consumo de gasolina

Si este conjunto de datos se entrena con Prophet tenemos:

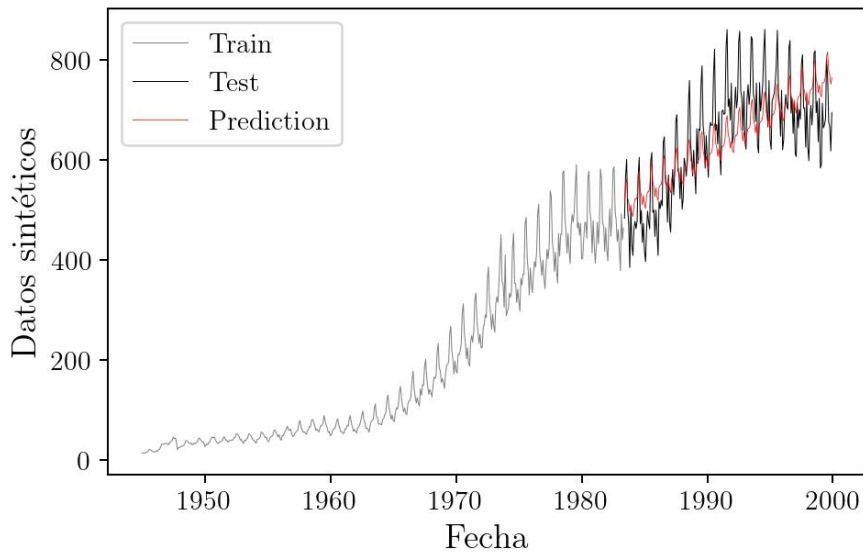


Figura 53. Predicción de la serie de consumo de gasolina con el modelo Prophet (Fuente: elaboración propia. Datos: Peña, 2010)

Conjunto	RMSE	MAE	MAPE
Consumo de gasolina	63.782	52.479	0.210

Tabla 24. Errores predictivos de la serie de consumo de gasolina utilizando el modelo Prophet

Datos	p-valor	Estacionaria
Lluvias en Santiago de Compostela	2.83e-08	Si
Accidentes laborales	0.82	No
Consumo de gasolina	0.85	No

Tabla 25. Valores p del test de Dickey Fuller aplicado a las diferentes series de tiempo y si hay o no estacionariedad

Bibliografía

- [1] G. Camps i Valls, "Predicción de series temporales," Universidad de Valencia, 2002.
- [2] J. I. Bagnato, "Pronóstico de Series Temporales con Redes Neuronales en Python | Aprende Machine Learning," 2019. <https://www.aprendemachinelarning.com/pronostico-de-series-temporales-con-redes-neuronales-en-python/> (accessed Feb. 21, 2023).
- [3] M. P. González Casimiro, "Análisis de series temporales: Modelos ARIMA," 2009.
- [4] T. Ruppert *et al.*, "Visual-Interactive Preprocessing of Time Series Data," 2012.
- [5] R. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice (3rd ed)," OTexts, 2021. <https://otexts.com/fpp3/index.html> (accessed Mar. 30, 2023).
- [6] J. Guerra Rodríguez, "Fundamentos y variantes de los modelos ARIMA para el análisis de series temporales. Aplicación a la estadística universitaria," Universidad de La Laguna, 2022.
- [7] Á. Hervás Martos, "Análisis de series temporales no estacionarias en RStudio," Universidad de Jaén, 2020.
- [8] E. A. J. El Habib, "Análisis de series temporales. Uso del transporte público en Barcelona," Universidad Politécnica de Cataluña, 2020.
- [9] C. Chatfield, "The Analysis of Time Series: An Introduction, Sixth Edition," *The Analysis of Time Series*, CHAPMAN&HALL/CRC, 2003, doi: 10.4324/9780203491683.
- [10] D. Peña, "Análisis de series temporales," Alianza Editorial, 2010.
- [11] R. H. Shumway and D. S. Stoffer, "Time Series Analysis and Its Applications," Springer Texts in Statistics, 2017, doi: 10.1007/978-3-319-52452-8.
- [12] S. Zanwar, "Autocorrelation Function (ACF) vs Partial Autocorrelation Function (PACF) ," 2023. https://medium.com/@ShankiiZ_/autocorrelation-function-acf-vs-partial-autocorrelation-function-pacf-228f9cf5a81 (accessed Jul. 10, 2023).
- [13] Y. Lyla, "A Quick Start of Time Series Forecasting with a Practical Example using FB Prophet," 2019. <https://towardsdatascience.com/a-quick-start-of-time-series-forecasting-with-a-practical-example-using-fb-prophet-31c4447a2274> (accessed Jun. 11, 2023).
- [14] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, "Time Series Analysis: Forecasting and Control," Wiley Series in Probability and Statistics, 2016.
- [15] J. Alberto Mauricio, "Introducción al análisis de Series Temporales," Universidad Complutense de Madrid, 2007.
- [16] A. G. Becerra, "Time series forecasting using SARIMA and SANN models," Universidad Politécnica de Cataluña, 2021.
- [17] B. Kumar Jha and S. Pande, "Time Series Forecasting Model for Supermarket Sales using FB-Prophet," in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 547–554. doi: 10.1109/ICCMC51019.2021.9418033.
- [18] M. Lambrechts, "When is it justifiable to exclude 'outlier' data points from statistical analyses?" https://www.researchgate.net/post/When_is_it_justifiable_to_exclude_outlier_data_points_from_statistical_analyses (accessed Jul. 11, 2023).
- [19] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 1, no. 1, pp. 73–79, 2011, doi: 10.1002/WIDM.2.
- [20] E. Zvornicanin, "Choosing the best q and p from ACF and PACF plots in ARMA-type modeling," 2022. <https://www.baeldung.com/cs/acf-pacf-plots-arma-modeling> (accessed Feb. 21, 2023).
- [21] P. Martí Sanahuja, "Métricas de evaluación de rendimiento para predicciones de series temporales," 2021. <https://polmartisanahuja.com/metricas-de-evaluacion-de-rendimiento-para-predicciones-de-series-temporales/> (accessed Feb. 21, 2023).
- [22] "Statsmodels - Seasonal decompose." https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal_decompose.html (accessed Jul. 18, 2023).
- [23] "scikit-learn: machine learning in Python — scikit-learn 1.3.0 documentation." <https://scikit-learn.org/stable/index.html> (accessed Jul. 18, 2023).
- [24] "Prophet | Forecasting at scale." <https://facebook.github.io/prophet/> (accessed Jul. 18, 2023).

Lista de figuras

Figura 1. Ejemplo de estacionariedad.....	8
Figura 2. Serie estacional sin tendencia (Fuente: elaboración propia. Datos: Peña, 2010).....	8
Figura 3. Serie con tendencia creciente (Fuente: elaboración propia. Datos: Peña, 2010)	9
Figura 4. Descomposición de la serie del número de pasajeros mensuales (en miles) de las aerolíneas norteamericanas desde 1949 hasta 1960 (Fuente: elaboración propia. Datos: Peña, 2010).....	10
Figura 5. Autocorrelación simple y parcial de un modelo sintético AR(1) (Fuente: elaboración propia)	12
Figura 6. Autocorrelación simple y parcial modelo MA (1) (Fuente: elaboración propia)	13
Figura 7. Autocorrelaciones simple y parcial de modelo ARMA(1,1) (Fuente: elaboración propia)	14
Figura 8. Predicción de serie de pasajeros en vuelos internacionales con Prophet (1949-1960) (Fuente: elaboración propia. Datos: Peña, 2010)	16
Figura 9. Ejemplo de funciones de autocorrelación para un modelo AR(1) (izquierda) y de un modelo MA (1) (derecha) (Fuente: elaboración propia)	17
Figura 10. Ejemplo de las funciones de autocorrelación para un modelo ARMA(1,1) (Fuente: elaboración propia).....	17
Figura 11. Serie de tiempo de Ozono de 1955 a 1972. (Fuente: elaboración propia. Datos: Peña, 2010)	19
Figura 12. Descomposición estacional serie de tiempo de Ozono (Fuente: elaboración propia. Datos: Peña, 2010)	21
Figura 13. A la izquierda la media y desviación de una serie no estacionaria y a la izquierda de la misma serie, pero estacionaria (Fuente: elaboración propia. Datos: Peña, 2010)	21
Figura 14. Autocorrelación simple para una serie estacionaria MA (1) Fuente: elaboración propia).....	22
Figura 15. Autocorrelación parcial serie AR(2) y serie de Ozono (Fuente: elaboración propia)	22
Figura 16. Serie de Ozono original (derecha) y serie diferenciada (izquierda) (Fuente: elaboración propia. Datos: Peña, 2010)	23
Figura 17. Predicción con modelo ARIMA de una serie de tiempo con datos de Ozono (Fuente: elaboración propia. Datos: Peña, 2010)	25
Figura 18. Representación gráfica del modelo AR(1) (Fuente: elaboración propia).....	28
Figura 19. Autocorrelación simple modelo AR(1) (Fuente: elaboración propia).....	29
Figura 20. Predicción modelo teórico AR(1) (Fuente: elaboración propia)	29
Figura 21. Grafica autocorrelaciones modelo MA(2) (Fuente: elaboración propia)	30
Figura 22. Predicción modelo MA(2) teórico (Fuente: elaboración propia)	30
Figura 23. Grafica de autocorrelación de un modelo ARMA(1,1) (Fuente: elaboración propia)	31
Figura 24. Predicción serie temporal ARMA(1,1) (Fuente: elaboración propia)	31
Figura 25. Serie sintética original junto con la predicción con el modelo Prophet (Fuente: elaboración propia)	32
Figura 26. Área de selección empleada para el presente proyecto. (Fuente: SDC – UPCT)	33
Figura 27. Series de tiempo de indicadores del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT) 34	
Figura 28. Descomposición estacional de la serie Clorofila (Fuente: elaboración propia. Datos: SDC - UPCT) 35	
Figura 29. Funciones de autocorrelación simple y parcial para la serie Clorofila (Fuente: elaboración propia. Datos: SDC - UPCT)	35
Figura 30. Descomposición estacional serie PE del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)	36
Figura 31. Autocorrelación simple y parcial para la serie PE (Fuente: elaboración propia. Datos: SDC - UPCT)36	
Figura 32. Descomposición estacional serie de Oxígeno del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)	37
Figura 33. Autocorrelación simple datos de oxígeno del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)	37
Figura 34. Descomposición estacional serie de salinidad (Fuente: elaboración propia. Datos: SDC - UPCT) ...	38
Figura 35. Autocorrelación simple y parcial serie de salinidad (Fuente: elaboración propia. Datos: SDC - UPCT)	38
Figura 36. A la izquierda descomposición estacional, a la izquierda media y desviación típica (Fuente: elaboración propia. Datos: SDC - UPCT)	39
Figura 37. Autocorrelación parcial serie de temperatura (Fuente: elaboración propia. Datos: SDC - UPCT) ...	39
Figura 38. Descomposición estacional serie de transparencia (Fuente: elaboración propia. Datos: SDC - UPCT)	40
Figura 39. Autocorrelación simple serie de temperatura (Fuente: elaboración propia. Datos: SDC - UPCT).....	40
Figura 40. Series de tiempo del Mar Menor diferenciadas (Fuente: elaboración propia. Datos: SDC - UPCT) ..	41
Figura 41. Predicción del 20% de prueba de las series del Mar Menor (Fuente: elaboración propia. Datos: SDC - UPCT)	43
Figura 42. Grafica de lluvias en Santiago de Compostela y su descomposición estacional (Fuente: elaboración propia. Datos: Peña, 2010)	47
Figura 43. Autocorrelación simple y parcial de datos de lluvia en Santiago de Compostela (Fuente: elaboración propia. Datos: Peña, 2010)	47

Figura 44. Predicción de los datos de lluvia de Santiago de Compostela con el modelo ARMA (Fuente: elaboración propia. Datos: Peña, 2010)	48
Figura 45. Predicción de los datos de lluvia de Santiago de Compostela con el modelo Prophet (Fuente: elaboración propia. Datos: Peña, 2010)	48
Figura 46. Gráfica de accidentes junto con su descomposición estacional (Fuente: elaboración propia. Datos: Peña, 2010)	49
Figura 47. Autocorrelación simple y parcial serie de accidentes (Fuente: elaboración propia. Datos: Peña, 2010)	49
Figura 48. Predicción serie de accidentes con el modelo SARIMA (Fuente: elaboración propia. Datos: Peña, 2010)	49
Figura 49. Predicción serie de accidentes con Prophet (Fuente: elaboración propia. Datos: Peña, 2010)	50
Figura 50. Serie de consumo de gasolina y su descomposición estacional (Fuente: elaboración propia. Datos: Peña, 2010)	51
Figura 51. Autocorrelación simple y parcial serie de consumo de gasolina (Fuente: elaboración propia. Datos: Peña, 2010)	51
Figura 52. Predicción serie de consumo de gasolina con SARIMA (Fuente: elaboración propia. Datos: Peña, 2010)	51
Figura 53. Predicción serie de consumo de gasolina con el modelo Prophet (Fuente: elaboración propia. Datos: Peña, 2010)	52

Lista de tablas

Tabla 1. Características de un modelo AR(1)	12
Tabla 2. Media, varianza y autocorrelaciones de un modelo MA(1)	13
Tabla 3. Media, varianza y autocorrelación de un modelo ARMA(1,1)	14
Tabla 4. Resumen de la forma de las funciones de autocorrelación para los modelos MA, AR y ARMA	17
Tabla 5. Fórmulas de los criterios de selección de modelo predictivo	24
Tabla 6. Valor de los criterios de información para un modelo Ozono	24
Tabla 7. Expresiones matemáticas de las métricas de evaluación	26
Tabla 8. Métricas de bondad del modelo de Ozono	26
Tabla 9. Errores predicción del modelo AR(1) teórico, tanto para un conjunto de datos, como la media de 10 conjuntos de datos	29
Tabla 10. Errores de predicción del modelo MA(2) junto con la media de errores y desviación típica de 10 series MA (2)	30
Tabla 11. Errores de predicción medios y desviación típica media de 10 conjuntos ARMA(1, 1)	31
Tabla 12. Errores de predicción del modelo Prophet	32
Tabla 13. Dimensión de datos reales junto con los datos vacíos y atípicos	34
Tabla 14. Valor de p de los conjuntos de datos de Mar Menor	40
Tabla 15. Modelos predictivos para cada conjunto de datos según los criterios AIC y BIC	42
Tabla 16. Modelos predictivos para los conjuntos de oxígeno y salinidad según AIC y BIC	42
Tabla 17. Errores predictivos de las series de tiempo del Mar Menor con modelos estadísticos	43
Tabla 18. Errores predictivos de las series de tiempo del Mar Menor con el modelo Prophet	44
Tabla 19. Errores de predicción serie de tiempo de lluvias modelo ARMA	48
Tabla 20. Errores de predicción datos de lluvias de Santiago de Compostela	48
Tabla 21. Errores predictivos serie de accidentes laborales	50
Tabla 22. Errores predictivos con Prophet de la serie de accidentes	50
Tabla 23. Errores predictivos datos del consumo de gasolina	52
Tabla 24. Errores predictivos serie de consumo de gasolina	52
Tabla 25. Valores p de las diferentes series de tiempo y si estacionariedad	52