

Trabajo Fin de Grado

Ingeniería de las Tecnologías Industriales

Optimización de la fecha de cultivo basada en predicciones por el modelo SARIMA

Autor: Ramón Ángel Simonet Roda

Tutor: José Miguel León Blanco

**Dpto. Organización Industrial y Gestión
de Empresas I**

Escuela Técnica Superior de Ingeniería

Sevilla, 2022



Trabajo Fin de Grado
Ingeniería de las Tecnologías Industriales

Optimización de la fecha de cultivo basada en predicciones por el modelo SARIMA

Autor:

Ramón Ángel Simonet Roda

Tutor:

José Miguel León Blanco

Profesor contratado doctor

Dpto. Organización Industrial y Gestión de Empresas I

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2022

Proyecto Fin de Carrera: Optimización de la fecha de cultivo basada en predicciones por el modelo
SARIMA

Autor: Ramón Ángel Simonet Roda
Tutor: José Miguel León Blanco

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2022

El Secretario del Tribunal

Agradecimientos

Querría agradecer a todas y cada una de las personas que, desde niño, me ayudaron con su cariño a llegar a este momento concreto de mi vida. Quisiera mencionar en especial a José e Irene, por su ayuda constante para llevar adelante este trabajo.

A mis padres, por la vida y su amor. Mis abuelos, por su ternura. Mis hermanas, por su paciencia y enseñanza continua. Mis sobrinas, por traer luz al mundo. Mis amigos (JRHMAVT), por su apoyo y compañía sincera e incondicional. Christos, por sus palabras fraternales de consejo. Ignacio, por ser ejemplo. Andrés, por enseñarme el significado de la amistad. Olaia, por causar el mayor cambio en mi vida y llenarla de FEC.

Ramón Ángel Simonet Roda

Sevilla, 2022

Resumen

El objeto de estudio de este trabajo es la aplicación de un modelo de predicción de precios en busca de la optimización de la fecha en la que se deberían cultivar las plántulas para obtener el mayor beneficio. Para hacer la estimación, debido al carácter estacional de los precios en este sector, se decidió aplicar el modelo SARIMA.

El sector donde se van a aplicar estas herramientas es el cultivo protegido bajo plástico característico de la zona de poniente almeriense y parte de la costa de Granada.

Abstract

The object of study of this work is an application of a price prediction model in search of the optimisation of the date on which a farmer should grow seedlings in order to obtain the highest profit. To make the estimation, due to the seasonal nature of prices in this sector, it was decided to apply the SARIMA model.

The sector where these tools are going to be applied is the protected crop under plastic, characteristic of the western area of Almeria and part of the coast of Granada.

Índice

Agradecimientos	- 7 -
Resumen	- 9 -
Abstract	- 11 -
Índice	- 13 -
Índice de Tablas	- 15 -
Índice de Figuras	- 17 -
1 Objeto y alcance del trabajo	- 19 -
2 Revisión de la literatura	- 21 -
2.1 <i>Predicción de precios</i>	- 21 -
2.1.1 Demanda nivelada	- 22 -
2.1.2 Demanda con tendencia	- 23 -
2.1.3 Demanda estacional	- 25 -
2.2 <i>Duración del proyecto y del análisis del problema</i>	- 30 -
2.3 <i>Herramientas auxiliares</i>	- 31 -
2.3.1 Interpolación mediante Spline	- 31 -
2.3.2 Indicadores estadísticos de error y precisión	- 31 -
3 Caso de estudio	- 33 -
3.1 <i>Contexto del problema</i>	- 33 -
3.2 <i>Datos y variables del problema</i>	- 35 -
3.2.1 Costes	- 35 -
3.2.2 Ingresos	- 36 -
4 Resolución del problema	- 39 -
4.1 <i>Recopilación de datos</i>	- 41 -
4.2 <i>Interpolación para homogeneizar los precios</i>	- 41 -
4.3 <i>Simulación inicial con SARIMA</i>	- 43 -
4.3.1 Simulación y determinación de los parámetros no estacionales de SARIMA	- 43 -
4.3.2 Simulación de los parámetros estacionales de SARIMA	- 44 -
4.4 <i>Elección de los parámetros estacionales</i>	- 45 -
4.5 <i>Control de la previsión</i>	- 46 -
	- 13 -

4.6	<i>Simulación completa</i>	- 46 -
4.7	<i>Interpolación de la predicción</i>	- 47 -
4.8	<i>Optimización de la fecha de cultivo</i>	- 47 -
5	Resultado y discusión	- 51 -
6	líneas futuras de trabajo	- 53 -
	Referencias	- 55 -

ÍNDICE DE TABLAS

- Tabla 1. Representación de p , d y q y diversos métodos, extraído de [2].
- Tabla 2. Resumen de costes recogidos en ‘Análisis de la campaña hortofrutícola Cajamar 2020/2021’.
- Tabla 3. Fechas de cultivo recogidas en el Observatorio de Precios y Mercados de la Consejería de Agricultura, Pesca y Desarrollo Rural de la Junta de Andalucía [10].
- Tabla 4. Duración del periodo de crecimiento basado en información de diferentes empresas del sector.
- Tabla 5. Tabla de previsiones con sus indicadores estadísticos.
- Tabla 6. Tabla de comparación de métodos.
- Tabla 7. Fechas de cultivo óptimas y los parámetros de sus modelos.

ÍNDICE DE FIGURAS

- Figura 1. Esquema de series temporales [1].
- Figura 2. Ejemplo de demanda nivelada para $D=3$.
- Figura 3. Ejemplo de demanda con tendencia para $D=3$ y $p=0.5$.
- Figura 4. Ejemplo de demanda estacional para $D=3$, $p=0.1$ y $f_t=(1;1.1;1.2;0.8;0.9)$.
- Figura 5. Estructura de la red neuronal y modelo de previsión. Extraído de Li et al. (2014)[6].
- Figura 6. Pseudocódigo del modelo híbrido. Extraído de Li et al. (2014)[6].
- Figura 7. Modelo STL-ELM para la predicción de precios extraído de Xiong et al. (2017) [7].
- Figura 8. Esquema de los diferentes lenguajes de programación utilizados.
- Figura 9. Pseudocódigo del proceso.
- Figura 10. Representación de la evolución anual de precios en la campaña.
- Figura 11. Pseudocódigo del proceso de interpolación.
- Figura 12. Representación de la interpolación al pimiento california amarillo.
- Figura 13. Pseudocódigo del proceso para la elección de los parámetros estacionales y no estacionales.
- Figura 14. Elección de los parámetros no estacionales del modelo mediante análisis de AIC y BIC.
- Figura 15. Simulación SARIMA con parámetros $(1,0,1)(1,0,2)$.
- Figura 16. Análisis de las simulaciones para el pimiento california rojo.
- Figura 17. Pseudocódigo del proceso de búsqueda de la fecha óptima de cultivo.
- Figura 17. Gráfica con el cálculo de la fecha optimizada para la sandía negra sin pepitas.
- Figura 18. Gráfica con el cálculo de la fecha optimizada para el pimiento california amarillo.
- Figura 19. Gráfica con el cálculo de la fecha optimizada para el pimiento california rojo.

1 OBJETO Y ALCANCE DEL TRABAJO

En este trabajo se analizará un apartado indispensable en la inversión que realizan los agricultores como es la elección de la fecha de cultivo de sus hortalizas. El contexto donde se basa el estudio es concretamente una región de España que tiene un gran peso dentro la economía local y nacional, el cultivo en invernadero del poniente almeriense y el oeste de la costa granadina. Como ocurre en otros muchos ámbitos económico-industriales, hay una descordinación entre la tecnología vanguardista y las técnicas que se aplican.

El objeto de estudio de este proyecto es utilizar, como herramienta de ayuda para la toma de decisiones del agricultor, una técnica avanzada de predicción de precios. Con ella, se puede adquirir datos sobre como será la evolución de precios de las hortalizas y así poder tener una mayor información técnica que ayude a determinar una fecha en la que plantar el cultivo en cuestión en su invernadero.

En esta zona, se cultivan una grandísima variedad de frutas y hortalizas, que aumenta cada año con la entrada de nuevas especies. Debido a las limitaciones de tiempo propias de un estudio fin de grado como este, se limitó el trabajo a cinco cultivos específicos: Sandía negra sin pepitas, melón piel de sapo y las tres variedades principales de pimiento california (amarillo, verde y rojo).

Respecto a la técnica a utilizar para analizar los precios, se contemplaron en un comienzo herramientas a corto y largo plazo. Aún así, tras hablar con agricultores locales y conocer la gran flexibilidad de plantar cada año un cultivo diferente, se decidió dar un enfoque de corto plazo al estudio. Dentro de las diferentes posibilidades utilizadas en la actualidad por la literatura académica, se contemplaron utilizar diferentes modelos. Finalmente, se decidió elegir el SARIMA por su versatilidad y uso extendido en el ámbito de la investigación y predicción de precios.

2 REVISIÓN DE LA LITERATURA

El análisis de la rentabilidad e idoneidad de una inversión en cualquier ámbito es uno de los temas más estudiados debido a su importancia en cuanto al desarrollo de un proyecto de negocio. A la hora de abordar el problema de inversión que se enfrenta hay varias dudas que se plantean relativos a esta serie de puntos:

- Naturaleza de las variables a tratar.
 - Fijas o variables.
 - Tipo de variabilidad.
- Duración del proyecto y del análisis del problema.

2.1 Predicción de precios

Lo que determinará el éxito o fracaso de una inversión será la evolución de sus variables: los ingresos y los costes. Respecto a la influencia de las variables en el problema, una parte principal del análisis es conocer la evolución futura de las variables. Este es un apartado de investigación donde en las últimas décadas se están haciendo grandes inversiones y avances, debido principalmente a la fuerte irrupción del Machine Learning. En la literatura académica existen multitud de métodos de predicción, dependiendo de la propia naturaleza de los precios.

Como se describe en [1] se pueden distinguir varios tipos de demanda y modelos de predicción asociados a ellos. Se podría resumir, añadiendo algunos métodos más avanzados, en los descritos en este cuadro.

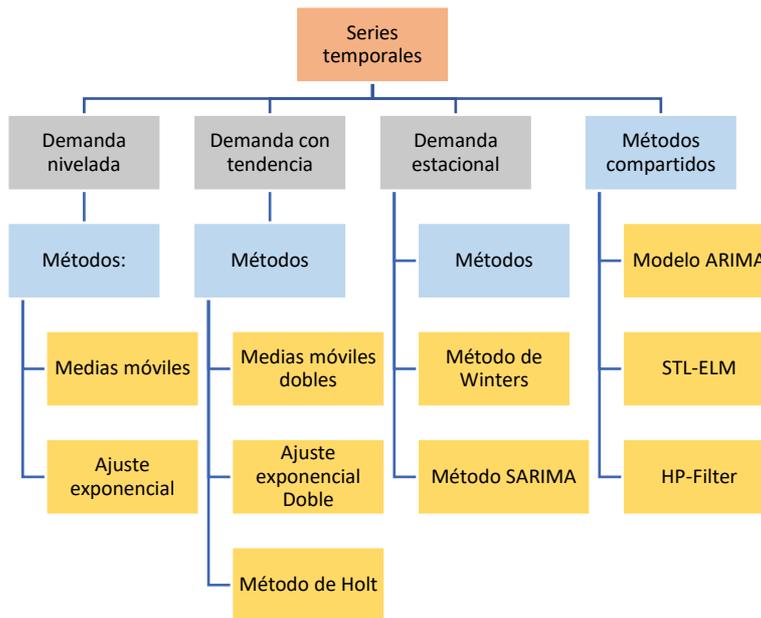


Figura 1. Esquema de series temporales [1]

A continuación, se describirán cada una de las evoluciones y métodos por separado.

2.1.1 Demanda nivelada

Es un tipo de demanda cuya evolución está caracterizada porque varían mínimamente respecto a un valor constante.

$$D_t = D + \varepsilon_t$$

Siendo D la demanda actual y ε_t una componente aleatoria.

La predicción de este tipo de demanda se puede hacer mediante métodos de fácil aplicación como las medias móviles y el ajuste exponencial.

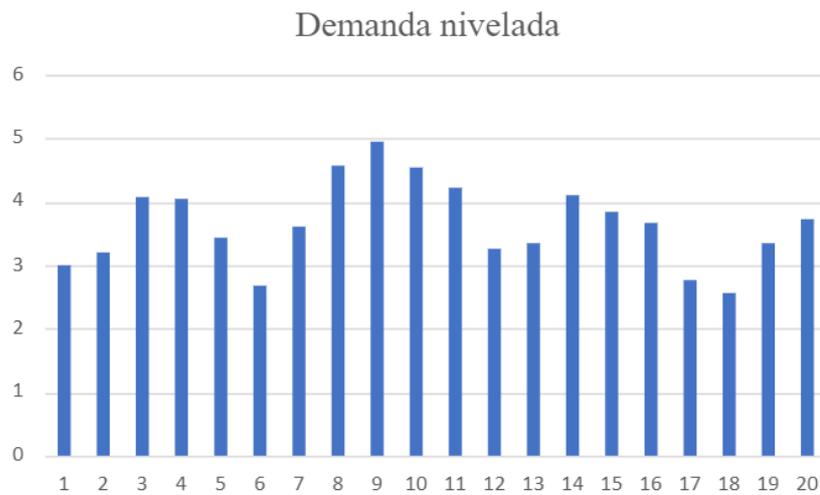


Figura 2. Ejemplo de demanda nivelada para $D=3$

2.1.2 Demanda con tendencia

La demanda con tendencia añade a la nivelada una componente de pendiente que marca si el precio sube o baja.

$$D_t = D + p_\tau + \varepsilon_t$$

Siendo p_τ la componente de crecimiento positivo o negativo.

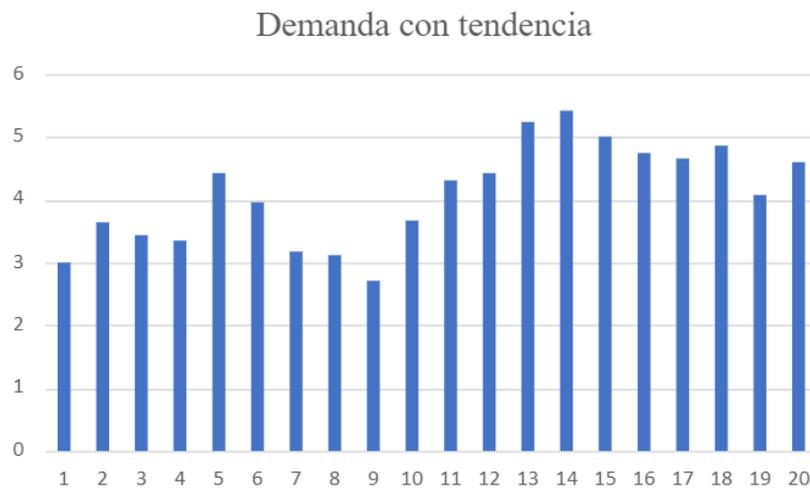


Figura 3. Ejemplo de demanda con tendencia para $D=3$ y $p=0.1$

Para predecir este tipo de evolución de variables existen múltiples modelos como son el de las Medias Móviles Dobles, el Ajuste Exponencial Doble o el Método de Holt. Existen después modelos más complejos, que requieren de programación avanzada para su implementación. Uno de éstos, es el conocido como modelo ARIMA y es del que nace el utilizado para este trabajo, SARIMA.

2.1.2.1 Modelo ARIMA (p,d,q)

Este modelo de predicción de variables sirve tanto para evoluciones niveladas como con tendencia. Sus siglas significan en inglés: *Autoregressive integrated moving average*. Teniendo una serie de datos, ARIMA saca de la base de precios las tendencias y las utiliza para predecir sobre los propios valores.

El modelo tiene tres componentes que dan nombre al modelo (AR-I-MA) y que son descritos en [2]:

- AR: Correspondiente a la autoregresión aplicada en el modelo y representado con la letra p:
 - AR de primer orden: $\hat{Y}_t = \alpha + b_1 Y_{t-1}$
 - AR de segundo orden: $\hat{Y}_t = \alpha + b_1 Y_{t-1} + b_2 Y_{t-2}$
 - AR de n orden: $\hat{Y}_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_n Y_{t-n}$

- MA: Correspondiente a la media móvil aplicada en el modelo y representado con la letra q:
 - MA de primer orden: $\hat{Y}_t = \gamma + d_0 u_t + d_1 u_{t-1}$
 - MA de segundo orden: $\hat{Y}_t = \gamma + d_0 u_t + d_1 u_{t-1} + d_2 u_{t-2}$
 - MA de n orden: $\hat{Y}_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_n Z_{t-n}$

La combinación de ambos modelos corresponde al modelo ARMA, que para parámetros (1,1), se definiría por:

$$\hat{Y}_t = \mu + \phi_1 Y_{t-1} + d_0 u_t + d_1 u_{t-1}$$

- I: Es la parte correspondiente a la integración y sirve para diferenciar las series. Se representa con la letra d.
 - Para d=0: $y_t = Y_t$
 - Para d=1: $y_t = Y_t - Y_{t-1}$
 - Para d=2: $y_t = Y_t - 2Y_{t-1} + Y_{t-2}$

Así, según la elección de parámetros se obtendrán diferentes configuraciones de funcionamiento de la predicción. En [2] se hace una recopilación de diferentes configuraciones:

	p	d	q	Differencing	Method
ARIMA (0, 0, 0)	0	0	0	$y_t = Y_t$	White noise
ARIMA (0, 1, 0)	0	1	0	$y_t = Y_t - Y_{t-1}$	Random walk
ARIMA (0, 2, 0)	0	2	0	$y_t = Y_t - 2Y_{t-1} + Y_{t-2}$	Constant
ARIMA (1, 0, 0)	1	0	0	$\hat{Y}_t = \mu + \phi_1 Y_{t-1} + \varepsilon$	AR(1): First-order regression model
ARIMA (2, 0, 0)	2	0	0	$\hat{Y}_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon$	AR(2): Second-order regression model
ARIMA (1, 1, 0)	1	1	0	$\hat{Y}_t = \mu + Y_{t-1} + \phi_1 (Y_{t-1} - Y_{t-2})$	Differenced first-order autoregressive model
ARIMA (0, 1, 1)	0	1	1	$\hat{Y}_t = Y_{t-1} - \phi_1 \varepsilon_{t-1}$	Simple exponential smoothing
ARIMA (0, 0, 1)	0	0	1	$\hat{Y}_t = \mu_0 + \varepsilon_t - \omega_1 \varepsilon_{t-1}$	MA(1): First-order regression model

Tabla 1. Representación de p, d y q y diversos métodos, extraído de [2].

Por último, es necesario comentar que existen gran cantidad de otros métodos de predicción para este tipo de demanda, relacionados con redes neuronales, que escapan del alcance del trabajo, pero que están teniendo un gran impacto en este mundo de la predicción.

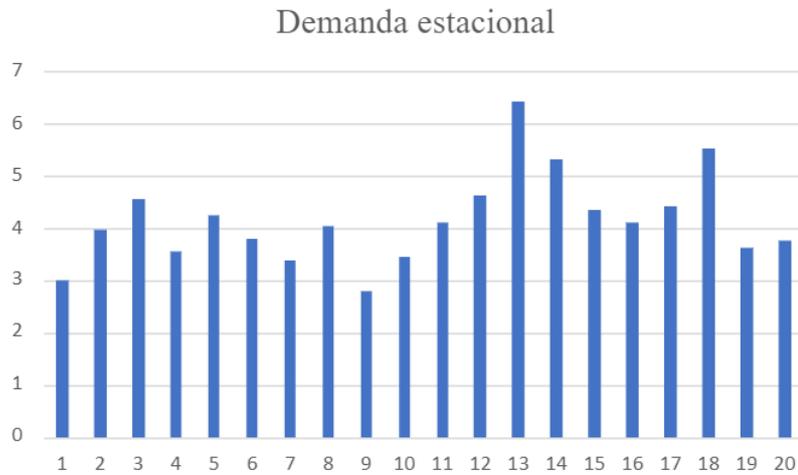
2.1.3 Demanda estacional

Existe una tercera tipología de evolución de variables denominada estacional debido a que sus valores siguen una tendencia concreta cada cierto periodo de tiempo. Esta es una de las series temporales que más se dan en los sectores cuyas producciones se repiten durante una determinada campaña.

Así, se podría hacer un acercamiento a su forma con la ecuación:

$$D_t = (D + p) \cdot f_t + \varepsilon_t$$

Donde la f_t representa un factor multiplicativo $f_t = f_{t+nL}$, siendo L la duración del ciclo estacional.



. Ejemplo de demanda estacional para $D=3$, $p=0.1$ y $f_t=(1;1.1;1.2;0.8;0.9)$

Existen diversos modelos predictivos de cara a las evoluciones estacionales, con diferentes grados de dificultad de implementación.

2.1.3.1 Método de Winters

Como se detalla en [1], el algoritmo del método de Winters para la demanda estacional, se compone de los siguientes pasos:

1. Estimación de la componente permanente:

$$S_T = \alpha \cdot \frac{D_T}{F_{T-1}} + (1 - \alpha)[S_{T-1} + p_{T-1}]$$

2. Estimación de la tendencia:

$$p_T = \beta(S_T - S_{T-1}) + (1 - \beta)p_{T-1}$$

3. Estimación del factor estacional:

$$F_T = \gamma \cdot \frac{D_T}{S_T} + (1 - \gamma)F_{T-L}$$

4. Normalización de los factores estacionales:

$$F_t = L \cdot \frac{\bar{F}_t}{\sum_{i=1}^L \bar{F}_i} \text{ para } t = nL + 1, nL + 2, \dots, (n + 1)L$$

5. La previsión para el periodo $T + \tau$ será:

$$D_{T+\tau}(T) = W_{T+\tau} = (S_T + \tau p_\tau)F_{T+\tau-L}$$

Para la selección de los valores iniciales se realiza lo siguiente:

1. Pendiente:

$$p_0 = \frac{1}{L} \sum_{i=1}^L \frac{D_{i+L} - D_i}{L}$$

2. Componente permanente:

$$D_0 = \mu_2 + \frac{(L-1)}{2} p_0$$

3. Componentes estacionales:

$$F_{i-L} = L \cdot \frac{\overline{F_{t-L}}}{\sum_{i=1}^L \overline{F_{t-L}}} \text{ para } i = 1, \dots, L$$

4. Componentes de alisamiento: α, β y γ

2.1.3.2 Modelo SARIMA

El modelo SARIMA es una variante del modelo ARIMA, diseñado para poder tratar series temporales estacionales. Sus siglas representan Seasonal-ARIMA. Al igual que en ARIMA, p, d y q representa respectivamente el orden de la parte autoregresiva, diferencial y de media móvil.

Este nuevo modelo añade a aquel los siguientes parámetros obtenidos de Sánchez de Rivera(2005)[3], Gujarati(2010)[4] y Novales Cinca(1993)[5], de tal forma que queda definido por:

$$ARIMA(p, d, q) \times (P, D, Q)_S$$

Siendo:

- p: Orden AR no estacional
- d: Orden de diferencias no estacional
- q: Orden MA no estacional
- P: Orden AR estacional
- D: Orden de diferencia estacional
- Q: Orden MA estacional
- S: Ventana de tiempo de patron estacional

La ecuación que lo rige, sin operaciones de diferenciación, queda así:

$$\Phi(B^S)\phi(B)(x_t - \mu) = \theta(B^S)\theta(B)\omega_t$$

Las componentes no estacionales serían:

- AR: $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$
- MA: $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$

Mientras que las componentes estacionales quedarían definidas por:

- AR estacional: $\Phi(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_p B^{pS}$
- MA no estacional: $\theta(B^S) = 1 + \theta_1 B^S + \dots + \theta_Q B^{QS}$

2.1.3.3 Otros métodos (H-P filter with ANN, STL-ELM)

Dentro del mundo de las predicciones de series temporales, hay un sector en fuerte desarrollo, que consiste en el uso de redes neuronales y algoritmos complejos de Machine Learning. Mencionaré ahora dos modelos que han sido utilizados para la predicción del precio de cultivos en dos artículos científicos.

Modelo híbrido de red neuronal de HP filter

Es la unión de las técnicas del filtro lineal Hodrick-Prescott (HP) y una red neuronal. Con la ayuda del filtro se puede extraer la tendencia y los componentes cíclicos de las series, mientras que junto a la red neuronal aprenden los patrones de tendencia y estacionales por separado.

En estas figuras del artículo se muestra la estructura de la red neuronal y una visión general del modelo.

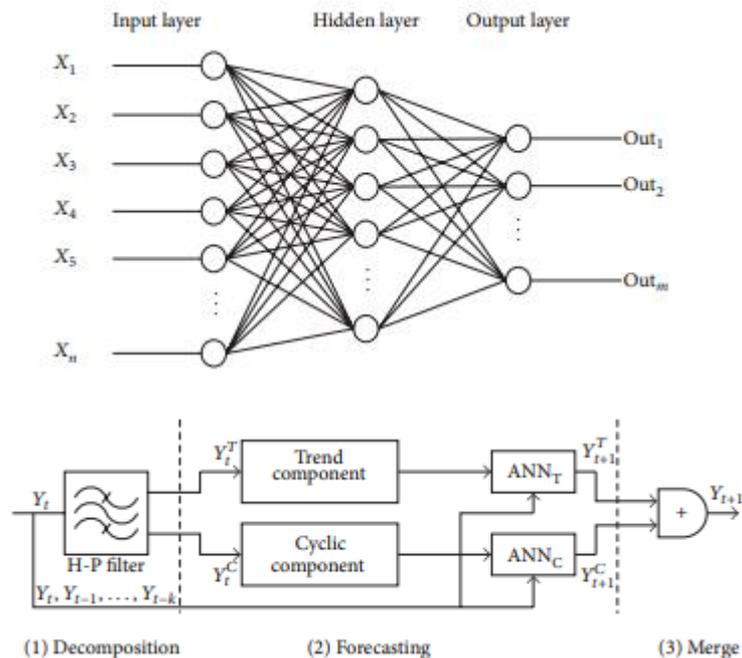


Figura 5. Estructura de la red neuronal y modelo de previsión. Extraído de Li *et al.* (2014)[6]

En la fig.6 se puede observar el pseudocódigo que definiría el modelo híbrido propuesto con su correspondiente diagrama de flujo.

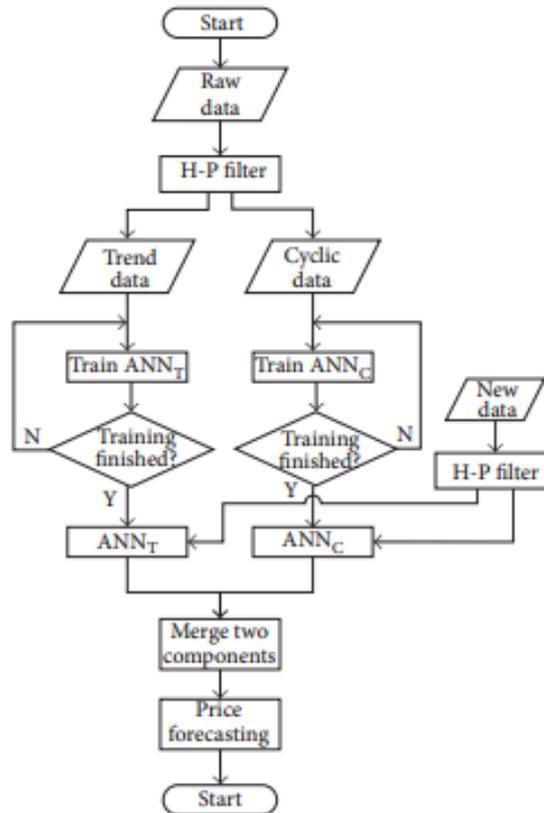


Figura 6. Pseudocódigo del modelo híbrido. Extraído de Li *et al.* (2014)[6].

Modelo híbrido de STL y ELM

Es un modelo híbrido entre la técnica STL (seasonal trend decomposition based on loess) y ELM (Extreme learning machine). Con una primera descomposición de las componentes estacionales, de tendencia y remanentes, se utiliza posteriormente el ELM para predecir las componentes de tendencia y remanente de forma independiente. Se usan además estrategias iteradas para una prevision multietapa.

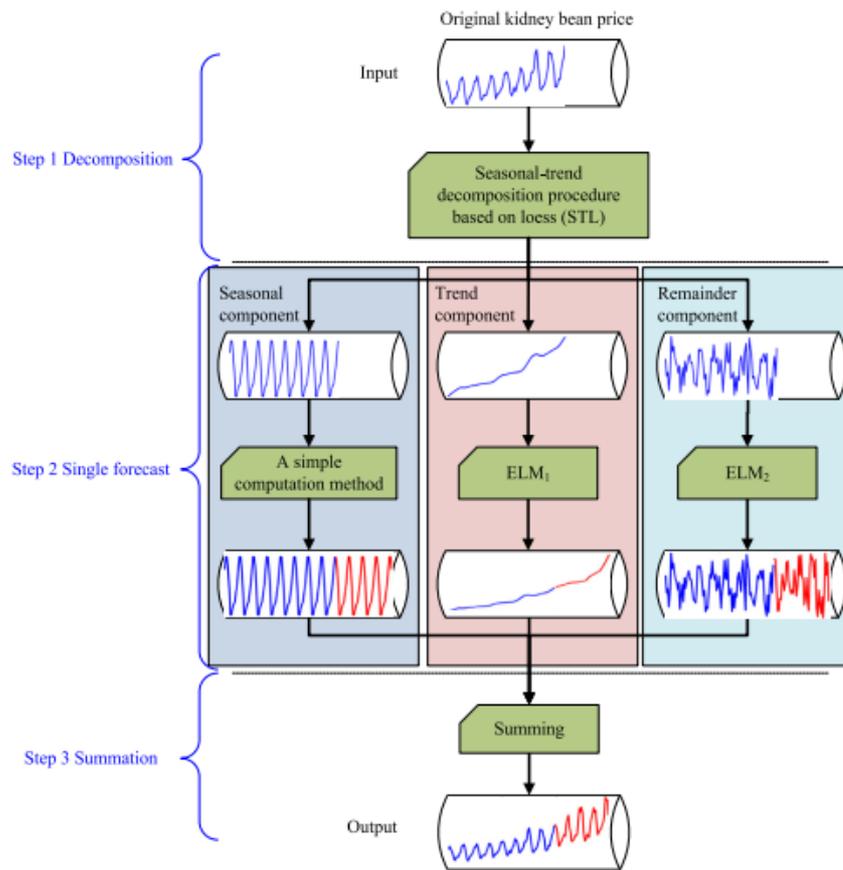


Figura 7. Modelo STL-ELM para la predicción de precios extraído de Xiong *et al.* (2017) [7].

Debido a que la evolución de los precios es diferente e independiente para cada cultivo, por su naturaleza y a las circunstancias ambientales, económicas y políticas entre otras, se da la realidad de que no hay definido un método mejor absoluto para su predicción. Hay por lo tanto en la literatura académica un gran debate e investigación sobre cuál es la metodología más apropiada para cada producto. Es por ello que hay artículos como el ya mencionado [7] y Yin *et al.* (2020)[8], donde se hacen estudios bibliográficos sobre los modelos utilizados para cada cultivo, teniendo en cuenta sus estacionalidad y variables de entrada. La elección del modelo SARIMA estuvo basada en varios factores. En primer lugar, es uno de los métodos más extendidos en el mundo de la predicción de precios, estando presente en la gran mayoría de artículos académicos referidos a este mundo. Ésto habla muy bien de su fiabilidad y buenos resultados. Además, su versatilidad frente a diferentes tamaños de series es de gran utilidad para su adaptación en cuanto a variaciones del periodo productivo. Por último, debido a la finalidad de este trabajo de hacer una herramienta accesible a los agricultores, este modelo presenta una mayor intuitividad a la hora de ser utilizado y configurado.

2.2 Duración del proyecto y del análisis del problema

Una vez se conoce la naturaleza de las variables se debe determinar si se puede y desea realizar un análisis con una perspectiva a corto o largo plazo.

Para inversiones a largo plazo existen metodologías de análisis muy extendidas como el flujo de fondos descontados o las opciones reales, ya sea mediante simulación Montecarlo, ecuaciones diferenciales parciales o árboles binomiales.

Hay decisiones corto placistas en las que únicamente se hace un balance de considerar los ingresos y costes esperados, con sus probabilidades si es posible, para tomar la decisión de si invertir.

Existen también decisiones, como es el caso de estudio en este Trabajo Fin de Grado, en el que se cruzan las perspectivas a corto y largo plazo. En esta casuística, se debe primero decidir con herramientas a largo plazo acometer o no la inversión, pero una vez acometida, cada cierto tiempo, por ejemplo, un año, se debe reconsiderar con herramientas a corto plazo si continuar o no.

2.3 Herramientas auxiliares

En la realización de este trabajo se han utilizado diferentes herramientas complementarias que fueron necesarias en algunos momentos del desarrollo del trabajo. Aquí se comentarán algunos de ellos.

2.3.1 Interpolación mediante Spline

A veces es necesario hacer una conversión a variable continua de una serie de datos discretos. Para ello se puede hacer uso de una interpolación continua. En la realización de este trabajo se utilizó la librería *Scipy* de Python, con su funcionalidad “*interpolate UnivariateSpline*” que realiza una interpolación con un grado de alisamiento k .

2.3.2 Indicadores estadísticos de error y precisión

Cuando se realiza la predicción de un precio mediante un modelo, es completamente necesario utilizar indicadores que midan el rendimiento de la predicción. Cuando se va a aplicar, se debe guardar un año de datos reales que puedan servir para la comparación con los datos predichos. Los indicadores estadísticos más extendidos entre la literatura científica referidos a la predicciones en series temporales son:

- Error cuadrático medio (RMSE): Sirve para medir el error del conjunto de predicciones, sin importar su signo.

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (p_t - d_t)^2}{N}}$$

Siendo N el número de datos por periodo, p_t los datos predichos y d_t los datos reales.

- Criterio de información de Akaike (AIC): es una estimación de la discrepancia de Kullback-Leibler entre un modelo verdadero y un modelo candidato.

$$AIC = 2 \log L_k + 2k$$

Siendo L_k el máximo valor de la función verosimilitud y k el número de parámetros del modelo.

- Criterio de información Bayesiano (BIC)

$$BIC = k \ln (n) - 2 \log L_k$$

Siendo L_k el máximo valor de la función verosimilitud, k el número de parámetros del modelo y n el tamaño de la muestra.

3 CASO DE ESTUDIO

A la hora de realizar este estudio se contempló aplicar las diferentes técnicas a las diversas regiones de la geografía española. Finalmente, se decidió aplicar el trabajo al poniente almeriense y la costa de Granada, donde sitúan gran parte del cultivo protegido bajo plástico de España. Esta elección se debió a muy diferentes motivos. Por un lado, la agricultura en esta zona es un motor principal económico, del que viven decenas de miles de personas y que sirve de productos a toda Europa. Se suma a ello el hecho de que existan diferentes fundaciones, cooperativas y el propio gobierno estatal, que con una gran labor de recogida de datos y análisis favorecen la realización factible de un proyecto como éste.

Fue, por último, imprescindible, el contacto personal directo que se pudo realizar con diferentes agricultores de la zona, que estuvieron dispuestos a asesorar el proyecto, haciendo que este estudio no fuera un mero trabajo académico alejado de la realidad, sino realista y aplicable. Todos estos motivos impulsaron la decisión de concretar este estudio y región de implementación.

3.1 Contexto del problema

En España la agricultura, ganadería, silvicultura y pesca tiene un peso económico y social muy importante. Siendo el PIB español a precios de mercado durante 2021 de 1.205 billones de euros, la agricultura, ganadería, silvicultura y pesca supuso un valor de 33 mil millones, suponiendo casi un 3% de peso [Datos INE 2021]. Esto lo coloca como el sexto sector más importante económicamente, únicamente por detrás del turismo, construcción, salud, comercio y servicios financieros.

En una provincia como la almeriense, especialmente conocida por su aportación a este sector, el valor es de 2476 millones, que es a su vez el 16,43% del PIB provincial. Estos datos ayudan a entender el volumen e importancia en cuanto a la economía española. [Datos INE 2019]

La cantidad de hectáreas de cultivo invernada en la temporada 2020/2021 fueron 32554ha. Debido a que se realiza en la misma tierra más de un cultivo al año, la superficie total cultivada llegó hasta los 52350 ha.

Siendo el cuarto país de Europa con mayor número de empleos en el sector agrícola [9] , son cientos de miles las personas en España con vinculación y en cierto modo dependencia del mundo del cultivo.

La crisis económica surgida a causa de la pandemia mundial de COVID-19, ha sido muy visibilizada en sectores como el sanitario y de servicios, pero también ha afectado en gran medida al sector agrícola. La volatilidad y la inflación se han convertido en importantes factores que hacen que sea más necesario que nunca poner el ojo en la búsqueda de la optimalidad debido a la creciente subida del precio de la energía y la reciente guerra en Ucrania, que ha afectado gravemente a elementos como los fertilizantes.

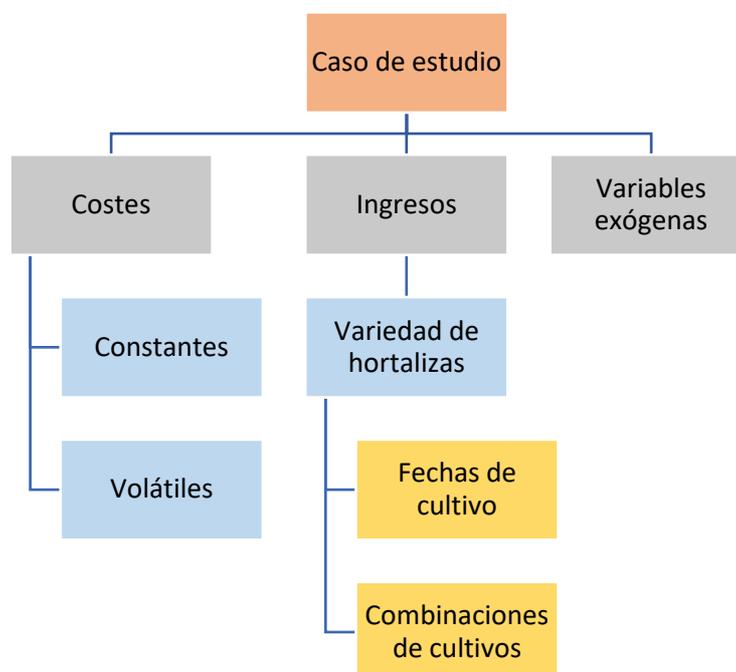
La zona en la que se centrará este problema esta principalmente ocupada por cultivo protegido bajo plástico. El concepto de invernadero de plástico fue desarrollado en los años 60 en esta región y en muy pocas décadas ha demostrado su potencial como economía base.

En la gran mayoría de casos el agricultor dispone de escasas herramientas de cara a la elección de la combinación de cultivos idónea a plantar cada temporada. Siendo éste factor uno de los más determinantes en el global de la inversión, el fin de este estudio es centrar y aplicar técnicas modernas de previsión de precios.

Si se consigue realizar una buena predicción del mercado, se podría conocer la fecha idónea para cultivar y vender las frutas y hortalizas al mayor precio y obtener así unos mayores beneficios que desencadenen en un mejor desarrollo del negocio.

3.2 Datos y variables del problema

Para entender el alcance y estructura del problema a tratar, se considera en nuestro caso la siguiente estructura desagregada de las variables y factores que afectan al cultivo protegido:



3.2.1 Costes

Debido a su peso social y económico, existen diversas fundaciones y cooperativas que recopilan datos y elaboran informes sobre las campañas agrícolas. Una de las principales es la Fundación Cajamar.

Los costes son desagregados en corrientes y amortizados, dependiendo de si son asociados únicamente a esa campaña o son parte de una inversión realizada a largo plazo. Por lo general, la edad media de un invernadero es de 8 años.

Campaña 2020/21		
	Euros	Gastos anuales (%)
Mano de obra	28904	44,8
Semillas y plantones	5935	9,2
Agua	1807	2,8
Fertilizantes	3926	6,1
Fitosanitarios	3510	5,4
Energía	1284	2,0
Servicios	3990	6,2
Transporte	1908	3,0

Otros gastos	1665	2,6
Total de gastos corrientes	51020	79,0
Sustrato/enarenado	2373	3,7
Estructura de invernaderos	5251	8,1
Plástico	3750	5,8
Sistema de riego	805	1,2
Balsa de riego	377	0,6
Otros	981	1,5
Total gastos de amortización	13537	21,0
Total gastos anuales	64557	100

Tabla 2. Resumen de costes recogidos en 'Análisis de la campaña hortofrutícola Cajamar 2020/2021'

Debido a la época actual de gran volatilidad, en la que costes que parecían controlados han disparado sus precios, es necesario hacer una mención especial a aquellos que han sufrido especialmente por factores como la guerra de Ucrania y la inflación.

Por un lado, el precio total de la energía (€/MWh) entre los meses de mayo de 2020 y 2021 fue de media 48.96 mientras que entre 2021 y 2022 ha subido hasta un promedio de 177.39 [10]. Para el sector agrícola la energía no es tan fundamental como podría serlo para el industrial, aún así, es reseñable su influencia.

Otro factor importante generado por estas crisis ha sido para los fertilizantes. Según el Ministerio de Agricultura, los fertilizantes han sufrido una subida del 84.87%. Esto es debido a que Rusia era uno de los principales proveedores de las materias primas con las que se elaboran estos productos.

Por último, es también reseñable el incremento en los precios de los plásticos que cubren las instalaciones, que pueden subir por encima del 60% respecto a su valor anteriormente, según ANAIP.

3.2.2 Ingresos

Los ingresos asociados a un invernadero son los que se reciben por la venta de las frutas y hortalizas. Estos dependen tanto de los precios del día en que se vendan como del rendimiento que esté dando la tierra.

Como ocurre en cualquier geografía, debido a las condiciones concretas de cultivo, existe una variedad limitada de hortalizas y frutas que pueden ser producidas. Con la evolución de la tecnología de semillas, cada vez hay una cantidad mayor de variantes, con unas características especiales, que alteran elementos como la duración después de ser recolectada, sabor, color e incluso semillas. Se están dando estos últimos años también cruces entre especies, que seleccionan las características que desea el mercado. Así, por ejemplo, se eliminaron las pepitas de las sandías o se desarrollaron tomates larga vida.

Los cultivos principales de invernadero en el este de Andalucía y sus meses de producción son los siguientes:

Fechas de producción	
Berenjena negra	Octubre-Junio
Calabacín verde	Octubre-Junio
Judía verde plana	Octubre-Junio
Pepino largo tipo Almería	Junio-October
Pimiento california verde	Septiembre-Junio
Pimiento california amarillo	Septiembre-Junio
Pimiento california rojo	Septiembre-Junio
Tomate rama	Octubre-Junio
Tomate Cherry	Octubre-Junio
Melón piel de sapo	Abril-Julio
Sandía	Abril-Julio

Tabla 3. Fechas de cultivo recogidas en [11]

Debido a los diferentes meses de producción, los agricultores hacen cada temporada la combinación de dos cultivos para así tener un mayor tiempo el invernadero produciendo. Las consecuciones clásicas para cada temporada en esta región geográfica son las siguientes:

- Pimiento-Melón
- Pimiento-Sandía
- Pepino-Sandía
- Calabacín-Calabacín
- Judía verde-Judía verde

La producción de cultivos viene dada por el rendimiento de la tierra ese curso. Este rendimiento también está recogido en los diferentes boletines anuales y sirve para conocer el posible ingreso de la explotación.

Una vez recolectados los productos, por lo general, se transportan a una cooperativa, donde hay una evolución continua de los precios, según dicte el mercado. Estos datos están recogidos en bases estatales como el *Observatorio de Precios y Mercados de la Consejería de Agricultura de la Junta de Andalucía*[10].

La evolución de los precios de las hortalizas y verduras en esta región es conocida por dos características que definirán la resolución del caso de estudio. Por un lado, y como se comentará más adelante,

existe una altísima volatilidad de los precios debido a factores externos como son la temperatura, plagas, introducción de nuevos productos o mercados externos. Por otro lado, también se conoce una cierta estacionalidad de los precios, con tendencias alcistas y bajistas en diferentes puntos del periodo productivo.

Debido a que cada tipo de cultivo presenta un comportamiento diferente en cuanto a evolución de precios y requería un extenso análisis por separado, se decidió acotar el estudio a cinco cultivos. El análisis consistirá en estudiar la idoneidad de las combinaciones pimiento-sandía y pimiento-melón. Para ello se seleccionará entre cada una de las tres variedades de pimiento california (Verde, rojo o amarillo) y entre la sandía negra sin pepitas o el melón piel de sapo.

Los dos principales suministradores de semillas y plántulas (*Bayer* y *Rijk Zwaan*) nos proporcionaron los datos sobre el tiempo que transcurre desde la plantación de las plántulas de los cultivos hasta su recolección.

	Días desde la plantación	Desviación estándar
Sandía negra sin pepitas	90	10
Melón piel de sapo	82	7
Pimiento california amarillo	95	5
Pimiento california rojo	95	5
Pimiento california verde	60	4

Tabla 4. Duración del periodo de crecimiento basado en información de diferentes empresas del sector

La recopilación de todos estos datos sirvió para dos aspectos fundamentales del trabajo. Por un lado ayudaron a conocer en profundidad la naturaleza matemática tan variable de los precios de los cultivos y su consecuencia en la dificultad de los agricultores para predecir. También ha ayudado este estudio para entender la importancia económica-social que supone este sector en miles de personas. Una vez todo definido el modelo se procedió a su resolución.

4 RESOLUCIÓN DEL PROBLEMA

La metodología seguida para la búsqueda de la optimización del cultivo debía lograr conocer la fecha idónea en la que cultivar, tal que se consiguiese que la recolección coincidiese con las fechas de mayores precios en el mercado. Este fue el proceso desarrollado para conseguir determinar la fecha óptima de cultivo. Fue necesario utilizar diferentes lenguajes y softwares para afrontar la resolución.

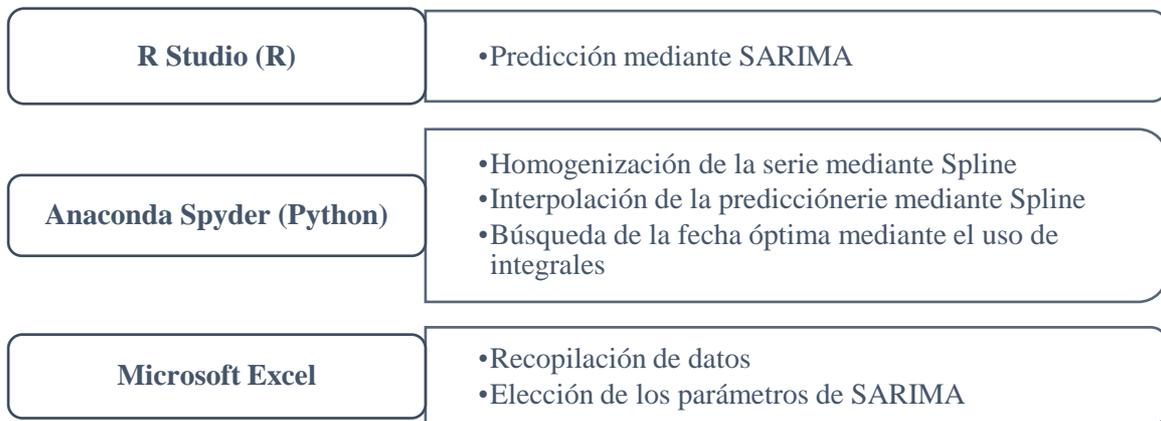


Figura 8. Esquema de los diferentes lenguajes de programación utilizados

El método, con sus diferentes procesos necesarios, se puede ver representado mediante el siguiente pseudocódigo de la figura 9.

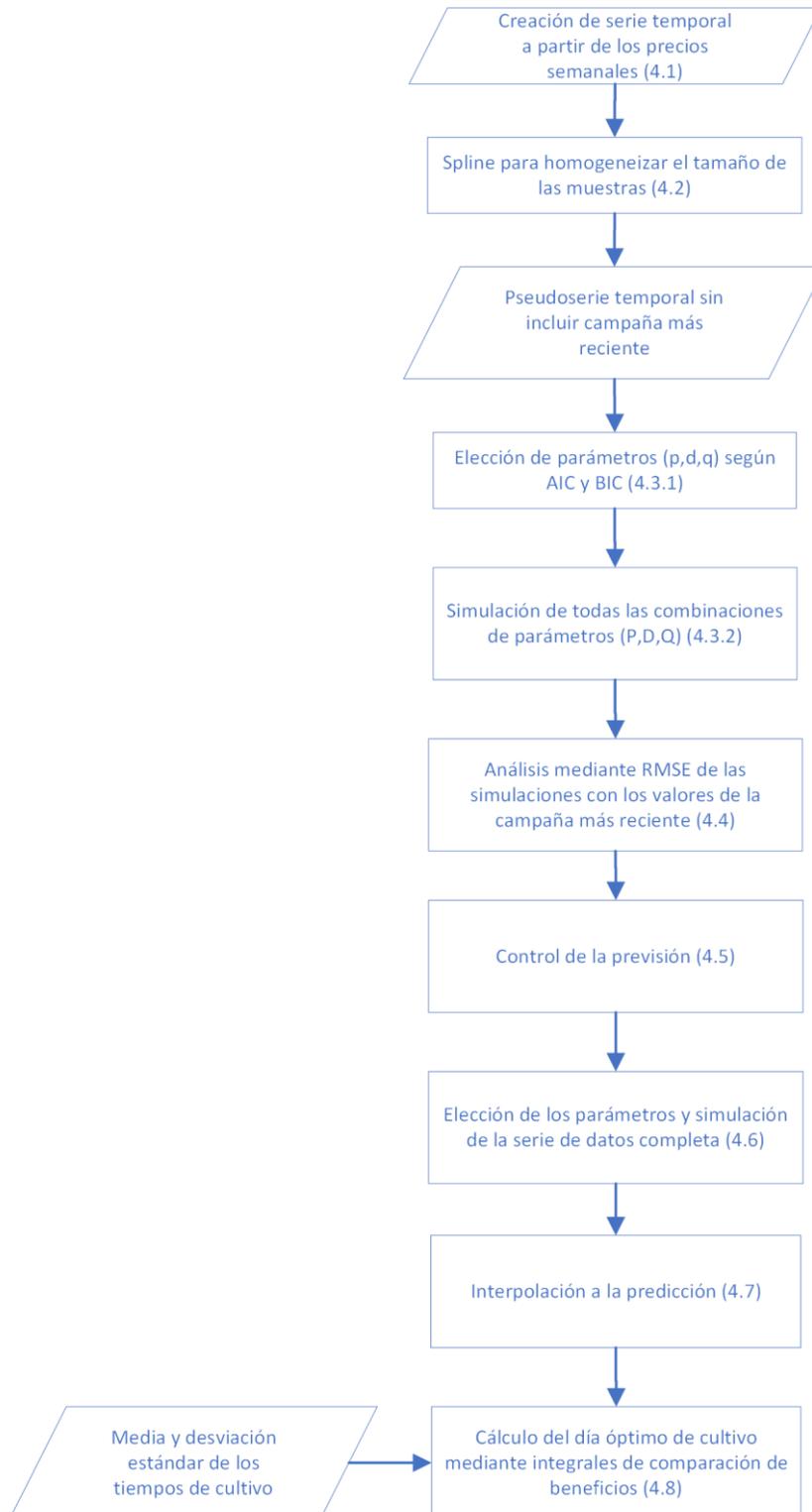


Figura 9. Pseudocódigo del proceso

4.1 Recopilación de datos

Los datos de precios de los que se disponía procedían de la base de datos del observatorio de precios [11]. Aunque muy diversa y completa, ya que contenía datos sobre la gran mayoría de cultivos y variedades de la región, se recogían los precios de forma semanal, lo cual era un primer inconveniente. Como se observa en la Figura 10, se producen en ocasiones grandes picos e irregularidades en la curva de precios debido a los cambios bruscos en la demanda de una semana a otra. También se puede observar ciertos periodos en los que los precios son alcistas respecto a otros momentos.

La base de datos albergaba series desde hacía 20 años, por lo que fue más que suficiente para el estudio. A causa de los ya comentados cambios en el mercado debido a la introducción de nuevos productos, se consideró recopilar únicamente los datos de los últimos cinco años, desde 2017 a 2021, reservándose éste último para las comprobación de bondad del modelo como se describe en el apartado 4.4.

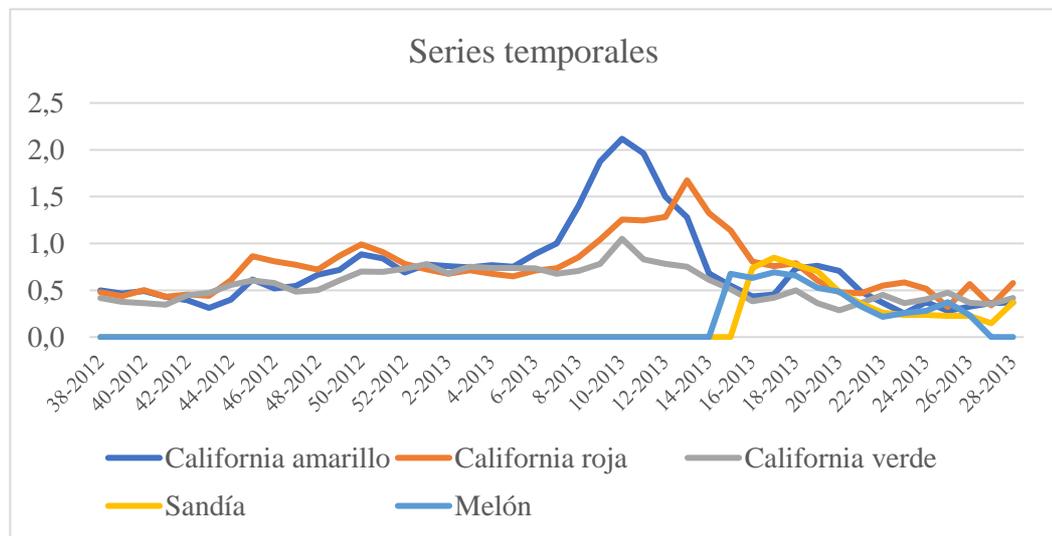


Figura 10. Representación de la evolución anual de precios en la campaña

4.2 Interpolación para homogeneizar los precios

Debido principalmente a los cambios meteorológicos, en este tipo de agricultura las campañas son de duración variable. Sería interesante en futuros estudios hacer una previsión sobre la duración de las campañas teniendo en cuenta conceptos como la temperatura y horas de luz, pero escapa del alcance de estudio de este trabajo académico.

Dada la situación en la que se tienen años con un número de datos diferentes, la base de datos no es uniforme. Este aspecto suponía ciertas dificultades respecto a la aplicación del modelo SARIMA, en concreto debido al parámetro S, que fuerza a que el periodo de la base de datos que sea introducido en el modelo sea constante, por lo que sería incompatible con nuestros datos. Es por esto que se decidió hacer uso de splines.

Así, el número de datos reales por semana de cada temporada se podían interpolar de forma continua para conseguir extraer en número de datos que se quisiese y así poder adaptarlo al parámetro constante S.

Debido a esto se fue, mediante un código en *Python* expuesto en la figura 11, homogenizando independientemente las series de datos de cada uno de los años de la base de precios, obteniéndose finalmente una serie periódica de valores que ya si poder introducir en el modelo SARIMA. Las librerías utilizadas, pertenecientes al lenguaje *Python*, fueron *Scipy*[12] y *Pandas*[13].

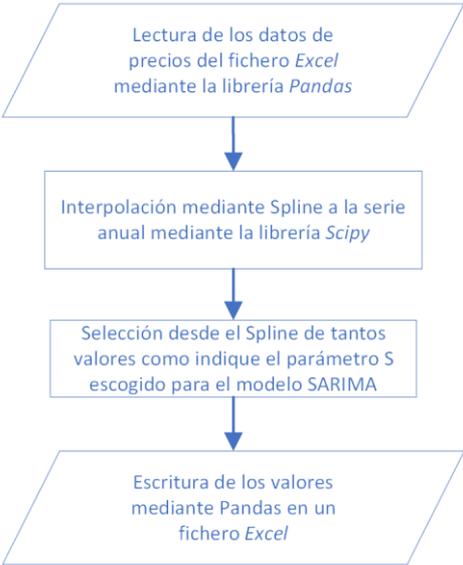


Figura 11. Pseudocódigo del proceso de interpolación

En la Figura 12 se puede observar la interpolación de un año de pimiento california amarillo. En él, la línea azul representa la función discreta de los valores reales, la línea verde fue la interpolación y los puntos son los diferentes valores finales que forman parte de la serie que sería introducida en el modelo.

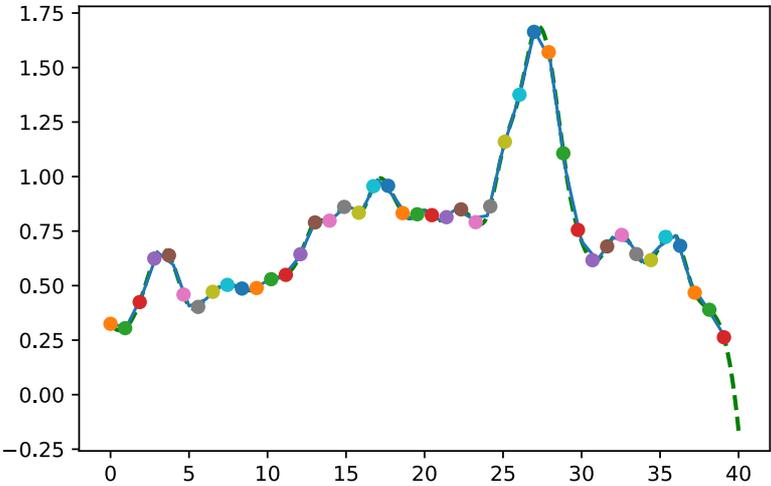


Figura 12. Representación de la interpolación al pimiento california amarillo

4.3 Simulación inicial con SARIMA

4.3.1 Simulación y determinación de los parámetros no estacionales de SARIMA

Una vez preparadas las series temporales que introducir en el modelo, se debían establecer para cada uno de los cultivos los seis parámetros que definen la predicción que hace el modelo SARIMA. Este proceso se decidió hacer en dos etapas, como se muestra en la figura 13. Las librerías utilizadas, pertenecientes al lenguaje *R*, fueron *xlsx*[14] y *forecast*[15].

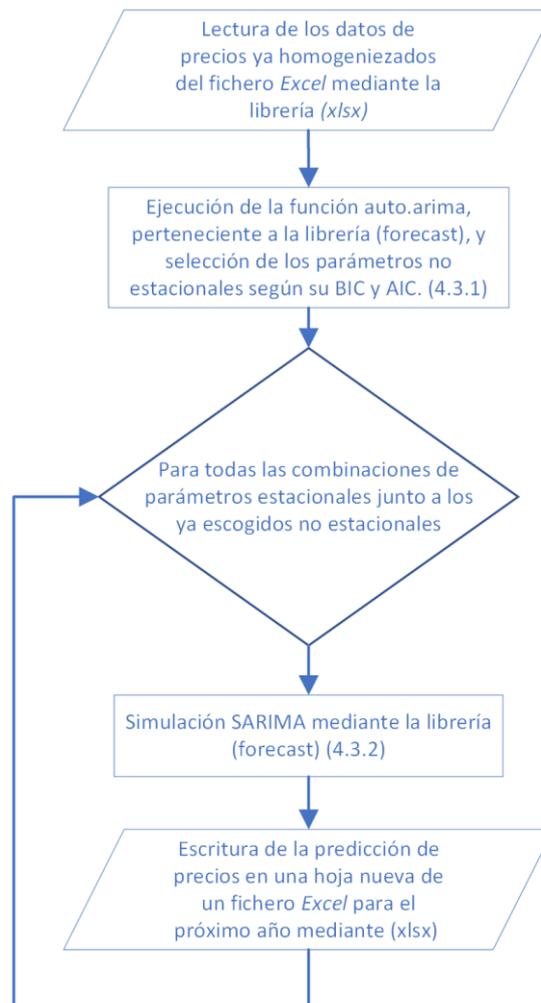


Figura 13. Pseudocódigo del proceso para la elección de los parámetros estacionales y no estacionales

La primera sería para escoger los tres parámetros no estacionales y la segunda para los estacionales. Esta primera etapa se realizó simulando las diferentes combinaciones de valores mediante el comando *auto.arima*, que compara los valores de AIC y BIC de las diferentes combinaciones y muestra el modelo de mejores indicadores. Así, en el ejemplo de la figura 14 se escogería los parámetros no estacionales (1,0,1) de cara a realizar la segunda etapa.

```
Best model: ARIMA(1,0,1)(2,1,0)[43]
Series: y
ARIMA(1,0,1)(2,1,0)[43]
Box Cox transformation: lambda= -0.6851236

Coefficients:
      ar1      ma1      sar1      sar2
      0.6272  0.3965  -0.6121  -0.3015
s.e.  0.0665  0.0813  0.0750  0.0825

sigma^2 = 0.07298: log likelihood = -31.69
AIC=73.38  AICc=73.67  BIC=90.24
```

Figura 14. Elección de los parámetros no estacionales del modelo mediante análisis de AIC y BIC

4.3.2 Simulación de los parámetros estacionales de SARIMA

Una vez definidos p , d y q , se decidió simular la mayor cantidad de escenarios posibles, de cara a escoger una simulación que se adaptase mejor a cada evolución de valores para cada uno de los cultivos. Así, se realizaron 64 simulaciones ($4!$) para cada serie temporal de tal manera: (0,0,0; 0,0,1; 0,1,0; 0,1,1; ...; 3,3,3). Este número de simulaciones estuvo determinado por el gran tiempo computacional de cada simulación cuando el valor de los parámetros aumentaba, habiendo simulaciones que podían tardar días en realizarse.

Cada predicción calculada se fue guardando de forma automatizada en hojas de un fichero *Excel*, además de ser representada gráficamente, para su posterior análisis. En la fig. 15 se puede ver una predicción de precios para unos parámetros (1,0,1)(1,0,2).

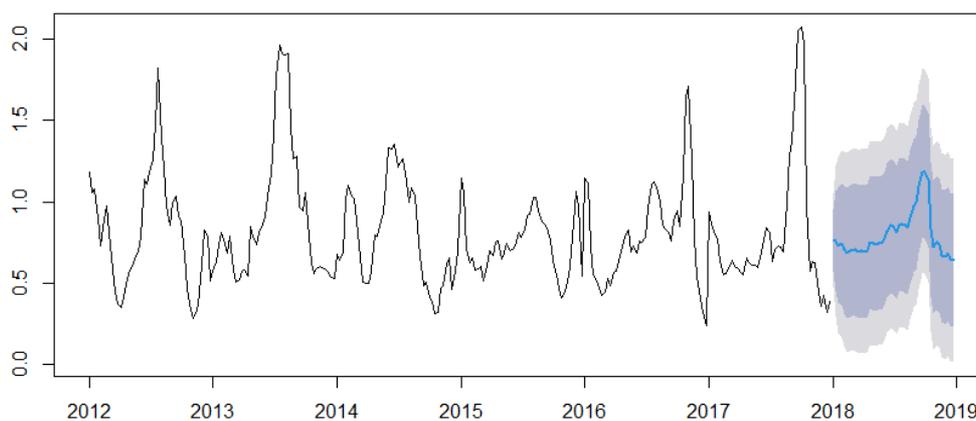


Figura 15. Simulación SARIMA con parámetros (1,0,1)(1,0,2).

4.4 Elección de los parámetros estacionales

Una vez obtenidas todas las $n!$ predicciones, se debían comparar con los valores del año que habíamos guardado. Así, para cada simulación, como se puede observar en la Fig.16, se creó una columna con la diferencia entre predicción y valor real y se calculó el RMSE. A su vez, se realizó mediante un análisis por colores, la comprobación de la bondad de los modelos, figurando de rojo por cada fila el valor más ajustado al real y el verde el más alejado. De este modo se puede observar con facilidad aquellas predicciones más fiables y su comportamiento en diferentes momentos.



Figura 16. Análisis de las simulaciones para el pimiento californiano rojo

Debido a la existencia de semanas en las que los precios son más elevados que el resto, por efectos de la demanda, se consideró necesario hacer también un estudio en esas fechas concretas. Esto es debido a que son las semanas más sensibles, donde es conveniente acertar en la predicción para alcanzar unos mayores ingresos. Por lo tanto se dividió el periodo de recolección de cultivos en dos tramos: bajista y alcista. La primera columna, que contiene los valores reales de mercado, fue también analizada por color y donde tiene un color verde más intenso es donde sus precios son mayores, estableciéndose esa parte como de especial interés. Los valores del RMSE se calcularon tanto en global como por zonas.

Los resultados numéricos para las mejores combinaciones de parámetros para cada uno de los cultivos son los que se muestran en la tabla 5. En ella se representa, el valor promedio del precio del año reservado junto a RSME que cometió la predicción y el porcentaje de error que supone respecto al promedio. Esto está realizado para toda la predicción, para el tramo alcista y para el tramo bajista.

	Sandía negra sin pepitas	Melón piel de sapo	Pimiento californiano amarillo	Pimiento californiano verde	Pimiento californiano rojo
Promedio	0,3694	0,4456	0,8040	0,5928	0,8539
RMSE (global)	0,0933	0,1165	0,4196	0,1110	0,4884
% de error	25,27%	26,14%	52,19%	18,73%	57,20%
Media bajista	0,2529	0,3883	0,6790	0,5867	0,7206
RMSE (bajista)	0,0473	0,0871	0,1530	0,0518	0,1327
% de error	18,72%	22,43%	22,54%	8,84%	18,42%
Media alcista	0,5560	0,6367	1,5580	0,7305	1,9265
RMSE (alza)	0,0934	0,0397	0,6346	0,1468	0,8444
% de error	16,80%	6,25%	40,73%	20,10%	43,83%

Tabla 5. Tabla de previsiones con sus indicadores estadísticos

4.5 Control de la previsión

Para comprobar la valía de los resultados, se realizó un control, haciendo una predicción alternativa con las medias de todos los valores de la primera semana en cada año y así sucesivamente, siguiendo la siguiente ecuación:

$$P_i = \frac{1}{n} \sum_{j=1}^n P_{i-Lj} \quad \forall i \in (1, n)$$

Siendo n el tamaño de años de datos acumulados, L el número de semanas de la muestra y P_i la previsión de precios.

En la tabla 6 se puede observar la comparación entre los valores del porcentaje de error de las predicciones mediante el modelo SARIMA y el método alternativo, de tal modo que se puede observar la clara mejoría que supone este tipo de metodologías a la hora de predecir.

Es interesante y necesario comentar como el modelo de previsión ha sido mejor que el método alternativo para todos los cultivos menos para el pimiento california rojo. Esto debe tomarse desde dos perspectivas, tanto que el modelo aporta información valiosa debido a sus buenos resultados como que no es infalible. De tal modo, es muy recomendable de cara a tomar la decisión de la fecha de inversión, el servirse también de estos métodos más sencillos como una especie de filtrado a las soluciones propuestas por el modelo SARIMA.

	Sandía negra sin pepitas	Melón piel de sapo	Pimiento california amarillo	Pimiento california verde	Pimiento california rojo
% de error de SARIMA	25,27%	26,14%	52,19%	18,73%	57,20%
% de error de método alternativo	28,12%	32,96%	59,49%	19,36%	49,94%
% de error alcista de SARIMA	16,80%	6,25%	40,73%	20,10%	43,83%
% de error alcista de método alternativo	21,73%	18,43%	59,76%	59,76%	22,97%

Tabla 6. Tabla de comparación de métodos

4.6 Simulación completa

Una vez elegidos los parámetros idóneos, se realizó la simulación SARIMA pero esta vez incluyendo ya la serie completa al añadir el último año de datos reales en la base de datos.

4.7 Interpolación de la predicción

Debido a que la predicción realizada por el modelo SARIMA contiene tantos valores como sea el parámetro S de la serie, la predicción calculada está medida en semanas. Siguiendo la finalidad práctica y real de nuestro caso de estudio, en el que la variable de tiempo no es discreta sino que realmente un agricultor puede cultivar el día que desee, lo cual es un carácter continuo, se decidió realizar un spline a la predicción. Ésto es debido a que de ser en semanas, la fecha idónea propuesta tendría un gran error al no considerar la continuidad entre semanas y ofrecer, por ejemplo, invertir en la semana 36 en vez de el día 22 de Octubre.

Por lo tanto, en un proceso similar al del apartado 4.2, se obtuvo de las predicciones seleccionadas, una vez convertidas en Spline, siete datos por cada uno que poseía la predicción, representando éstos a los días de la semana.

4.8 Optimización de la fecha de cultivo

Conocida la media y desviación estándar del tiempo que tarda en crecer cada tipo de cultivo y las series temporales diarias de las predicciones de precios, se podía acometer la búsqueda de las fechas óptimas de cultivo, como se explica en la figura 17.

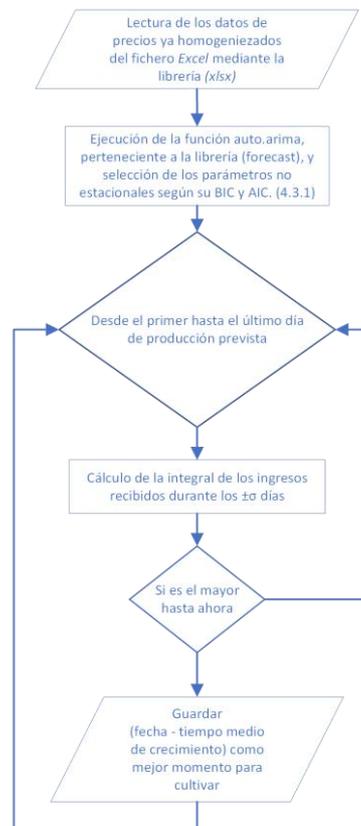


Figura 17. Pseudocódigo del proceso de búsqueda de la fecha óptima de cultivo

Mediante el uso de integrales, se determinó en que periodo era óptimo plantar y recoger los cultivos, haciendo uso de esta fórmula, que calcula los ingresos que se reciben:

$$\int_{x-\sigma}^{x+\sigma} f(x) dx \quad \forall x \in (0, n)$$

Siendo $f(i)$ la función spline a la predicción, σ la desviación estándar de crecimiento del cultivo y n el tamaño de la serie.

Ésto se hizo recorriendo con la integral todo el horizonte de predicción y comparando los sucesivos resultados para determinar el mejor intervalo de recolección y en consecuencia hallar el mejor día de cultivo. Como se puede observar en las figuras 18, 19 y 20, sobre la predicción de precios, representada por la función amarilla, aparecen una serie de puntos de colores, que representan el conjunto de días de recomendada recolección del cultivo. El punto azul muestra la fecha optimizada de plantación calculada para plantar en el invernadero el cultivo.

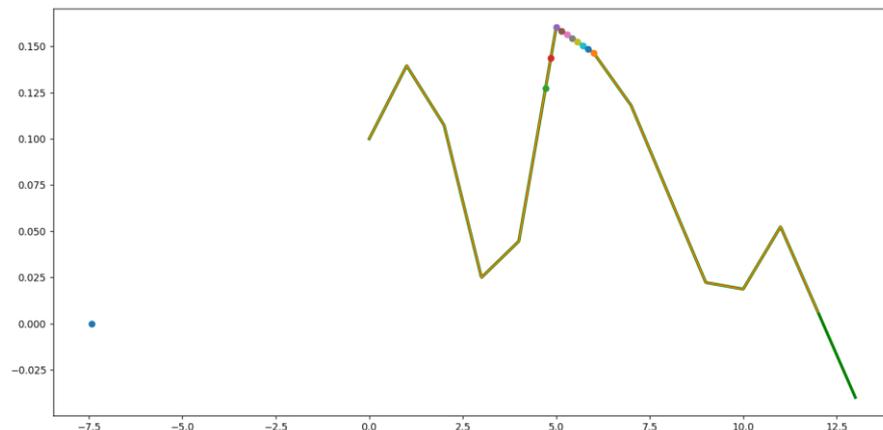


Figura 18. Gráfica con el cálculo de la fecha optimizada para la sandía negra sin pepitas

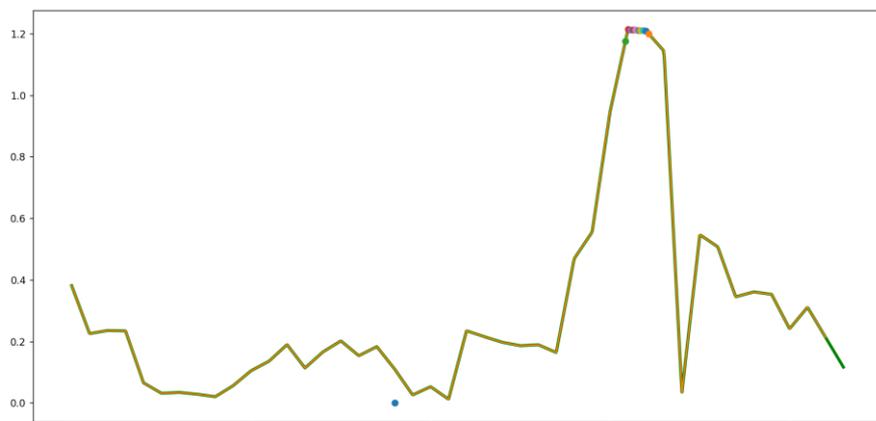


Figura 19. Gráfica con el cálculo de la fecha optimizada para el pimiento california amarillo

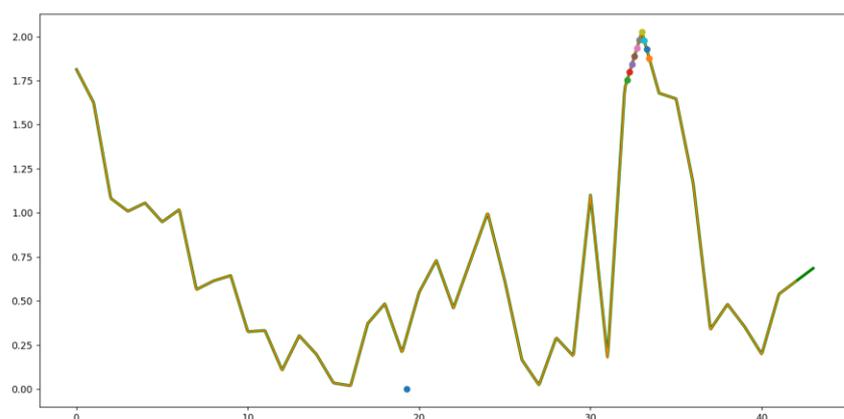


Figura 20. Gráfica con el cálculo de la fecha optimizada para el pimiento california rojo

Las fechas calculadas para la plantación, junto a los parámetros del modelo SARIMA que han sido elegidos, son las que se muestran en la tabla 7.

	Sandía negra sin pepitas	Melón piel de sapo	Pimiento california amarillo	Pimiento california verde	Pimiento california rojo
Fecha de cultivo con riesgo	(8/02/2022)	(19/03/2022)	(20/02/2022)	(26/12/2021)	(2/02/2022)
Parámetros del modelo con riesgo	(2,0,0)(1,0,1)	(1,0,1)(1,2,0)	(1,0,1)(1,3,0)	(1,0,4)(0,1,2)	(1,0,1)(0,3,3)
Fecha de cultivo conservadora	(8/02/2022)	(13/01/2022)	(24/01/2022)	(20/01/2022)	(14/01/2022)
Parámetros del modelo conservador	(2,0,0)(1,0,1)	(1,0,1)(0,0,3)	(1,0,1)(1,0,1)	(1,0,4)(0,1,3)	(1,0,1)(0,1,2)

Tabla 7. Fechas de cultivo óptimas y los parámetros de sus modelos

5 RESULTADO Y DISCUSIÓN

Como se muestra en la tabla 6, y era objetivo de este trabajo, se logró determinar mediante el modelo SARIMA una fechas concretas para cultivar las hortalizas y frutas y así obtener los mayores ingresos. Cuando se hizo el estudio preliminar al trabajo, hablando con agricultores del poniente almeriense de las herramientas utilizadas, fue sorprendente conocer como toda decisión en este ámbito estaba por lo general basada simple y únicamente por la experiencia y métodos sencillos de previsión. Como se observó en la tabla 5, estos métodos no son precisos y pueden llevar a tomar decisiones muy negativas de inversión.

El resultado de este estudio, logrando predicciones precisas en un sector tan estratégico e importante para la economía y vida pública, es muy positivo. La decisión de inversión debe estar siempre influenciada, y más en este siglo tecnológico que tenemos por delante, en metodologías como las desarrolladas o comentadas en este trabajo.

Desde el comienzo de la realización del trabajo han ocurrido ciertos sucesos que han afectado a este sector. Por un lado, la mayoría de los abonos utilizados provienen de Rusia y Ucrania, lo que ha encarecido los precios debido a su conflicto bélico. Ésto se une a la creciente subida de la luz y la inflación, causando grandes aumentos de los costes previstos. Por otro lado, la inestable relación con Argelia tiene consecuencias en la variación de la demanda, afectando a los ingresos.

Es reseñable también comentar cómo el avance de la tecnología genética hace que la entrada de las nuevas variedades de cultivos fuerce una competitividad mayor de precios, afectando también al tiempo de cultivo y resistencia a efectos climatológicos.

Estas variables externas condicionan la temporada y es por ello necesario remarcar el valor de tomar este tipo de decisiones de cultivo tomando en consideración el conjunto de circunstancias del momento junto a las propias herramientas matemáticas.

6 LÍNEAS FUTURAS DE TRABAJO

Siendo el Machine Learning, y en concreto la predicción de series temporales, un mundo en desarrollo e investigación continua, la proyección futura es complicada de determinar a la vez que ilusionante. Por un lado los métodos de redes neuronales parecen los situados en primera línea de cara a lograr las mejores predicciones en un futuro. Sería de interés ampliar este proyecto con la implementación de éstas, pudiendo añadir variables exógenas tan influyentes y complejas como la meteorología e indicadores de la economía global como la inflación y el IPC.

Otro elemento fundamental para líneas futuras sería lograr bases de datos más completas, que recojan, por ejemplo, el precio diario en las cooperativas así como datos más precisos sobre el tiempo de crecimiento de una planta y su relación con la temperatura. También sería de interés el lograr una depuración del código o un servidor que pueda ejecutar una mayor cantidad de simulaciones y así encontrar nuevas combinaciones de parámetros para el modelo SARIMA.

REFERENCIAS

- [1] Onieva Giménez, L., Escudero Santana, A., Cortés Achedad, P., Muñuzuri Sanz, J. y Guadix Martín, J. (2007) Diseño y gestión de sistemas productivos. Dextra.
- [2] Vishwas, B.V. y Patel, A. (2020). Hands-on Time Series Analysis with Python: From Basics to Bleeding Edge Techniques. Apress.
- [3] Peña Sánchez de Rivera, D. (2005). Análisis de series temporales. Alianza Editorial.
- [4] Gujarati, D.N. (2010). Econometría. McGraw-Hill Interamericana de España S.L.
- [5] Novales Cinca, A. (1993). Econometría. McGraw-Hill Interamericana de España S.L.
- [6] Li, Y., Li, C. y Zheng, M. (2014). A Hybrid Neural Network and H-P Filter Model for Short-Term Vegetable Price Forecasting. *Mathematical Problems in Engineering*. 2014(4). 10.1155/2014/135862.
- [7] Xiong, T., Li, C. y Bao, Y. (2017). Seasonal Forecasting of Agricultural Commodity Price Using a Hybrid STL and ELM Method: Evidence from the Vegetable Market in China. *Neurocomputing*. 275. 10.1016/j.neucom.2017.11.053.
- [8] Yin, Helin., Jin, Dong., Gu, Yeong Hyeon., Park, Chang Jin., Han, Sang Keun., Yoo, Seong Joon. (2020). STL-ATTN-LSTM: Vegetable price forecasting using stl and attention mechanism-based LSTM. *Agriculture (Switzerland)*, 10(12), 1-17. doi:10.3390/agriculture10120612
- [9] Parlamento Europeo (02, 12, 2022). Estadísticas sobre la agricultura de la UE: ayudas, empleo, producción. <https://www.europarl.europa.eu/news/es/headlines/society/20211118STO17609/estadisticas-sobre-la-agricultura-de-la-ue-ayudas-empleo-produccion>
- [10] Red Eléctrica de España (02, 11, 2022). Componentes del precio final y energía de cierre (€/MWh | MWh). Sistema eléctrico: Peninsular. <https://www.ree.es/es/datos/mercados/componentes-precio-energia-cierre-desglose>
- [11] Junta de Andalucía (11, 15, 2021). Hortícolas protegidos. Observatorio de Precios y Mercados. Consejería de Agricultura, Pesca y Desarrollo Rural. <https://www.juntadeandalucia.es/agriculturaypesca/observatorio/servlet/FrontController?action=Static&subsector=20&url=subsector.jsp>

[12] Scipy (02,13,2022) <https://docs.scipy.org/doc/scipy/>

[13] Pandas (02,13,2022) <https://pandas.pydata.org/>

[14] xlsx (03, 26, 2022) <https://cran.r-project.org/web/packages/xlsx/index.html>

[15] forecast(03,26,2022) <https://cran.r-project.org/web/packages/forecast/index.html>